A 5-Volume Set

MEDICAL IMAGING SYSTEMS TECHNOLOGY

Modalities

Cornelius T Leondes
editor

Medical Imaging Systems Technology

Modalities

MEDICAL IMAGING SYSTEMS TECHNOLOGY
A 5-Volume Set
Editor: Cornelius T Leondes *(University of California, USA)*

Analysis and Computational Methods
ISBN 981-256-993-6

Modalities
ISBN 981-256-992-8

Methods in General Anatomy
ISBN 981-256-991-X

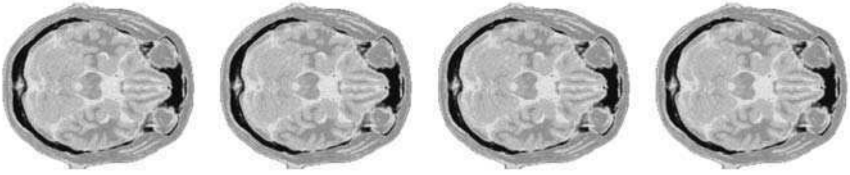
Methods in Diagnosis Optimization
ISBN 981-256-990-1

Methods in Cardiovascular and Brain Systems
ISBN 981-256-989-8

MEDICAL IMAGING SYSTEMS TECHNOLOGY

Modalities

A 5-Volume Set



editor
Cornelius T Leondes
University of California, Los Angeles, USA

World Scientific
NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

MEDICAL IMAGING SYSTEMS TECHNOLOGY
A 5-Volume Set
Modalities

Copyright © 2005 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-256-364-4 (Set)
ISBN 981-256-992-8

Typeset by Stallion Press
Email: enquiries@stallionpress.com

Printed in Singapore.

Preface

Because of the availability of powerful computational techniques, new modality techniques such as Computer-Aided Tomography (CAT), Magnetic Resonance Imaging (MRI) and others, and because of the new techniques of imaging processing (machine vision), the lives of many patients will be saved, and the quality of all our lives improved. This marriage of powerful computer technology and medical imaging has spawned a new and growing generation of young dynamic doctors who hold PhDs in physics and/or computer science, along with their MDs. In addition, technologists and computer scientists, with their superb skills, are also deeply involved in this area of major significance.

This volume covers the subject of medical imaging systems — modalities, by leading contributors on the international scene. This is one of the 5 volumes on medical imaging systems technology, and together they collectively constitute an MRW (Major Reference Work). An MRW is a comprehensive treatment of a subject requiring multiple authors and a number of distinctly-titled and well-integrated volumes. Each volume treats a specific subject area of fundamental importance in medical imaging. The titles of the respective 5 volumes which compose this MRW are:

- Medical Imaging Systems — Analysis & Computational Methods
- Medical Imaging Systems — Modalities
- Medical Imaging Systems — Methods in General Anatomy
- Medical Imaging Systems — Methods in Diagnosis Optimization
- Medical Imaging Systems — Methods in Cardiovascular & Brain Systems

Each volume is self-contained and stands alone for those interested in a specific volume. However, collectively this 5-volume set evidently constitutes the first multi-volume comprehensive reference dedicated to the multi-discipline area of medical imaging.

There are over 130 coauthors of this notable work and they come from 25 countries. The chapters are clearly written, self-contained, readable and comprehensive with helpful guides including introduction, summary, extensive figures and examples with in-depth reference lists. Perhaps the most valuable feature of this work is the breadth and depth of the topics covered.

This volume on "Medical Imaging Systems — Modalities" includes essential subjects like:

(a) Multislice helical computed tomography: Techniques and applications
(b) Techniques in magnetic resonance diffractive imaging and their application
(c) Techniques in 3-D assessment of tracheal stenosis by the means of spiral computed tomography (S-CT) and their application
(d) Edge preserved denoising in magnetic resonance images and their applications
(e) Techniques in X-ray computed tomography in the evaluation of drug release systems and their application
(f) Techniques in treating the bioelectromagnetic source imaging problems and their application
(g) Electrical impedance tomography for imaging and lesion estimation
(h) Single-shot magnetic resonance imaging (MRI) techniques and their applications
(i) Shape based interpolation methods for medical images and their application
(j) Nonparametric pixel appearance probability model using grid quantization for local image information representation
(k) Measurement of the carotid artery stenosis from magnetic resonance angiography

The contributors of this volume clearly reveal the effectiveness of the techniques available and the essential role that they will play in the future. I hope that practitioners, research workers, computer scientists, and students will find this set of volumes to be a unique and significant reference source for years to come.

Contents

Preface v

Chapter 1
Multislice Helical Computed Tomography: Techniques and Applications 1
Patrick J. La Rivière

Chapter 2
Techniques in Magnetic Resonance Diffractive Imaging and Their Application 37
Satoshi Ito and Yoshifumi Yamada

Chapter 3
Techniques in 3D Assessment of Tracheal-Stenosis by the Mean of Spiral Computed Tomography (S-CT) and Their Applications 61
Erich Sorantin, Darius Mohadjer, Franz Lindbichler, Laszlo G. Nyul, Kalman Palagyi and Bernhard Geiger

Chapter 4
Edge Preserved Denoising in Magnetic Resonance Images and Their Applications 81
Paul Bao and Lei Zhang

Chapter 5
Techniques in X-Ray Computed Tomography in the Evaluation of Drug Release Systems and Their Application 105
Agata A. Exner, Jinming Gao and David L. Wilson

Chapter 6
Techniques in Treating the Bioelectromagnetic Source Imaging Problems and Their Application 133
F. Greensite, A. Pullan and G. Huiskamp

Chapter 7
Electrical Impedance Tomography for Imaging and Lesion Estimation 193
Jin Keun Seo, Ohin Kwon and Eung Je Woo

Chapter 8
Single-Shot Magnetic Resonance Imaging (MRI) Techniques and Their Applications 241
Yihong Yang, Hong Gu, Thomas J. Ross, Wang Zhan and Shaolin Yang

Chapter 9
Shape Based Interpolation Methods for Medical Images and Their Application 281
Tong-Yee Lee and Chao-Hung Lin

Chapter 10
Non-Parametric Pixel Appearance Probability Model Using Grid Quantization for Local Image Information Representation 297
Mingzhou Song and Robert M. Haralick

Chapter 11
Measurement of Carotid Artery Stenosis from Magnetic Resonance Angiography 331
Peter J. Yim and J. Kevin Demarco

Index 351

CHAPTER 1

MULTISLICE HELICAL COMPUTED TOMOGRAPHY: TECHNIQUES AND APPLICATIONS

PATRICK J. LA RIVIÈRE
University of Chicago, Department of Radiology and
Committee on Medical Physics, 5841 S. Maryland Ave., MC-1037
Chicago, IL 60637, USA
pjlarivi@midway.uchicago.edu

One of the most significant recent developments in CT technology has been the emergence of multislice systems. In contrast to single-slice helical CT systems, in which the source illuminates a single row of detector elements, the multislice systems feature a two-dimensional array of detector elements, with up to 64 rows in the longitudinal direction in the latest generation of scanners. These scanners radically improve the tradeoffs between imaging time, volume coverage, and longitudinal resolution that constrain any CT study. In this chapter, we discuss the technical background of multislice CT, as well as the most common CT clinical applications and the impact multislice CT has had on them. The new imaging geometry associated with multislice scanners gives rise to new image reconstruction and visualization challenges, and we discuss these at some length. Other topics addressed include sampling and aliasing and dose issues.

Keywords: Computed tomography; helical scan; conebeam tomography; multislice tomography.

1. Introduction

In multislice Computed Tomography (CT), the X-ray source illuminates a curved or flat two-dimensional array of detector elements, as illustrated in Fig. 1. The emergence of multislice scanners has radically improved the tradeoffs between imaging time, volume coverage, and longitudinal resolution that constrain any CT study. The amount of time available for a CT scan is typically limited by the length of time patients can hold their breath — so as to avoid introducing motion artifacts — or by the length of time during which an organ of interest is enhanced in a study employing an injected contrast agent. Given this scan-time limitation, single-slice helical CT scanners often face a bleak tradeoff between longitudinal resolution (along the long axis of the patient) and volume coverage. Covering 30 cm of the thorax in 30 seconds with a single slice scanner with a 0.5 s rotation time requires the use of 5 mm longitudinal slice collimation if the table translation distance per revolution is 5 mm (i.e. the helical pitch is 1) or 2.5 mm longitudinal slice collimation if the table translation distance per revolution is 10 mm (i.e. the helical pitch is 2). Given that CT can reconstruct images with an in-plane pixel size as small as 0.5 mm, there is obviously a mismatch between in-plane and longitudinal resolution in either case.

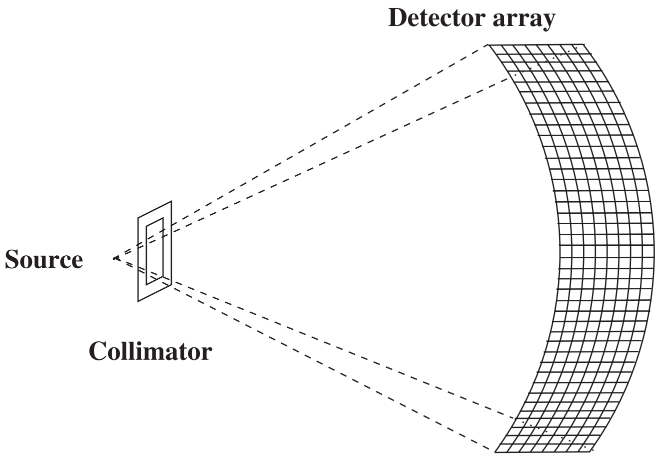


Fig. 1. In multislice helical CT systems, the source illuminates a 2D array of detector elements.

This resolution mismatch has been gradually erased by the various generations of multislice scanners, which increase the volume coverage achievable at any given longitudinal resolution. Indeed, the current generation of 64-slice scanners achieves essentially isotropic 0.5 mm resolution with excellent extended volume coverage. Moreover, the combination of speed and volume coverage attainable by multislice scanners has made it possible to image dynamic processes such as contrast enhancement, perfusion, and even the beating heart over extended areas.

In the remainder of this section we will discuss the historical and technical background to multislice CT, introducing concepts and terminology that will be used throughout the chapter. In Sec. 2 we describe some of the most common CT clinical applications and the impact multislice CT has had on them. The new imaging geometry associated with multislice scanners gives rise to new image reconstruction challenges that are discussed in Sec. 3. Section 4 discusses the challenges posed in managing and visualizing the huge datasets acquired by multislice CT, which reflect the transformation of CT from a slice-based imaging modality, acquiring a finite number of thick slices of the patient, to a true volumetric imaging modality, acquiring a three-dimensional dataset that can be sliced and rendered in a variety of ways. In Sec. 5, we discuss some of the novel sampling and aliasing issues introduced by multislice geometries, and finally in Sec. 6 we address dose issues in multislice CT.

1.1. *Historical development of multislice Computed Tomography*

With roots in classical X-ray tomography and early emission tomography, computed tomography emerged in 1972 with the development of the EMI scanner by Hounsfield.[1] A flurry of academic and commercial research ensued, leading to the development of four distinct "generations" of CT scanners by the end of the decade.[1–3] Comparatively speaking, the 1980s were a quiet time for CT development, with many researchers and manufacturers focusing instead on the emerging technology of magnetic resonance imaging.

CT again became a focus of significant research effort in the 1990s with the emergence of spiral or helical scanning, first implemented by Kalender and his collaborators.[4] This technology was made possible by the development and improvement of slip-ring technology, which allowed the source and detectors of a CT system to revolve continuously, unrestricted by power and data cables. By translating the patient table constantly as the source and detector revolve, one eliminates the "deadtime" spent incrementing the table in the previous step-and-shoot mode. This significantly reduced scan times, allowing for the scanning of extended sections of anatomy during a patient breathhold and also improving patient throughput.

The first multi-row scanner was introduced in 1992 by Elscint, a two-row system dubbed the CT Twin. The other major manufacturers did not introduce multislice CT systems until 1998, when the first four-slice scanners were introduced.[5,6] The pace of development has been nearly exponential since then (see Fig. 2), with

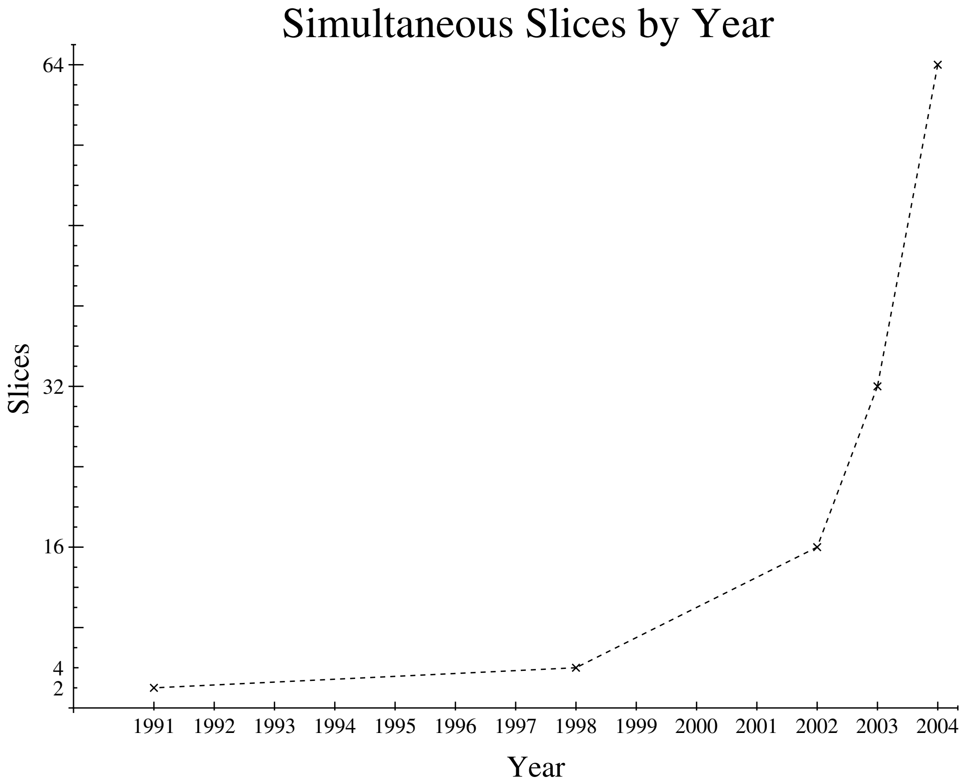


Fig. 2. The increase in the number of slices in CT systems has been essentially exponential in recent years.

64-slice scanners released in 2004. There will surely be an upper limit to this growth, for at some point the additional cost and weight of a still-larger detector will outweigh any additional clinical benefit. Nonetheless, some manufacturers have already demonstrated prototype 256-slice systems with flat-panel detectors that span 12.8 cm longitudinally, which would allow for dynamic imaging of entire organs without patient translation.

1.2. *Multislice CT detector arrays and Data Acquisition Systems*

Naturally, what distinguishes a multislice CT scanner from a single-slice CT scanner is its use of an array detector comprising of multiple rows along the longitudinal axis of the patient, as illustrated schematically in Fig. 1. The detector array is coupled to an electronic Data Acquisition System (DAS) that reads out the detector measurements of each row or, in some cases, of groups of rows clustered electronically into an effective row. The maximum number of longitudinal channels in the DAS determines the maximum number of effective rows in the scanner. This is a very important concept. As illustrated in Fig. 3, all of the original four-slice scanners introduced in 1998 had detector arrays comprising considerably more than four physical rows, but the DAS had only four longitudinal channels and so only four effective rows could be formed.

The different CT manufacturers have taken a variety of approaches to subdividing the detector array into physical rows and regrouping them into effective rows. General Electric (GE) initially favored an equally subdivided detector array. In that manufacturer's four-slice scanners, the array is subdivided into 16 rows of

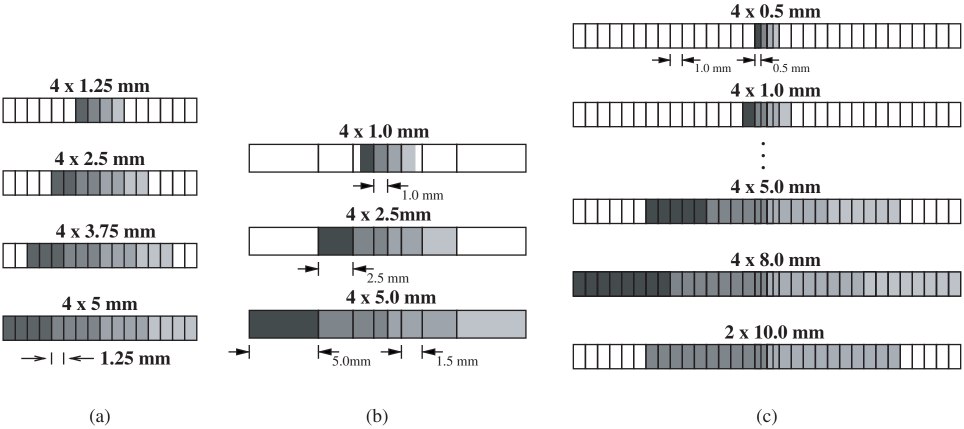


Fig. 3. Longitudinal detector configurations employed in four-slice scanners by various CT manufacturers. Note that all of the detector arrays have more than four physical rows. In practice, the data acquisition system reads out the signals of at most four effective rows, each comprising one or more physical rows clustered together electronically. (a) Equally subdivided array employed by General Electric. (b) Adaptive array employed by Picker/Philips and Siemens. (c) Hybrid array employed by Toshiba.

longitudinal extent 1.25 mm (all dimensions refer to projected dimensions at the system rotation axis), as illustrated in Fig. 3(a). These can be grouped into four effective rows each of extent 1.25 mm, 2.5 mm, 3.75 mm, or 5 mm, as shown. Naturally, prepatient collimation is adjusted to illuminate only the active physical rows so as to minimize radiation dose to the patient. Philips and Siemens have favored an adaptive array, in which the array is subdivided into physical rows of different dimensions, with smaller rows near the center and larger rows near the edges of the detector, as shown in Fig. 3(b). For the four-slice scanners, this allowed three modes of operation: A 4×1.0 mm mode, a 4×2.5 mm mode and a 4×5.0 mm mode. In order to achieve the 4 × 1.0 mm mode with the adaptive array configuration, the prepatient collimation is adjusted so as to illuminate only a portion of the two 1.5 mm rows. The advantages of this configuration relative to the GE array are the slightly higher resolution (1.0 mm versus 1.25 mm) in the highest-resolution mode as well as the slightly higher detection efficiency in the lower-resolution modes, because the adaptive array has fewer non-sensitive gaps separating its smaller number of physical rows than does the equally subdivided array. The disadvantage of this adaptive array is the lack of an intermediate-resolution mode between the 4 × 2.5 mm and 4 × 5.0 mm modes. Finally, as illustrated in Fig. 3(c), Toshiba took a hybrid approach in its four-row scanners, with four central 0.5 mm rows flanked by a total of 30 1.0 mm rows. This larger detector has both an extremely high-resolution mode (4 × 0.5 mm), a number of intermediate modes (4 × 1.0 mm–4 × 5.0 mm), and two lower-resolution, higher volume coverage modes (4 × 8.0 mm and 2 × 10.0 mm).

Variations on these detector configurations, appropriately scaled up, have been employed by manufacturers in their later scanners, although with the soon-to-be-released 64-slice scanners, most of the manufacturers have moved to equally subdivided arrays of 64 0.5 or 0.625 mm physical rows coupled to a 64-channel DAS. Siemens, however, is taking a different approach. The manufacturer has coupled a 64-channel DAS to a detector comprising of 32 0.6 mm rows and 8 1.2 mm rows. However, they are employing a novel X-ray tube with a “flying” focal spot that can shift rapidly between two different longitudinal positions. This allows the system to obtain 64 effective slice measurements at each view angle while illuminating only the central 32 physical rows of the detector, and thus allows for reconstruction of 0.4 mm isotropic voxels.

All existing commercial multislice scanners make use of an array of solid-state detectors, with each element comprising of a scintillator and a photodetector.[7] The scintillator absorbs X-rays through photoelectric interactions and enters a short-lived excited state that resolves itself through the emission of visible or ultraviolet radiation. The photodetector, typically a photodiode, absorbs these emissions and gives rise to an electrical current that is measured by the DAS. These solid-state detectors supplanted the use of high-pressure xenon detectors because of their much higher X-ray detection efficiency as well as the greater ease with which they can be assembled into arrays. However, the solid-state arrays themselves are likely to be supplanted in the near future by integrated flat panel detectors, which will be more practical and economical to manufacture than arrays as the number of rows grows.

1.3. *Multislice helical pitch*

In a single-slice helical CT exam, the pitch is defined as the ratio of the table translation distance Δ to the longitudinal collimation of the detector w, as measured at the rotation axis, i.e.

$$P_{ss} = \frac{\Delta}{w}. \tag{1}$$

Pitches of 1.0–2.0 are typically employed in single-slice CT. In multislice helical CT, there are two competing definitions of helical pitch.[3] Given a system with N effective rows each of width w, as measured at the rotation axis, one could again define pitch as the ratio of the table translation distance to the longitudinal width of each effective row:

$$P_{ms1} = \frac{\Delta}{w}, \tag{2}$$

or instead as the ratio of the table translation distance to the total longitudinal width the effective rows:

$$P_{ms2} = \frac{\Delta}{Nw}. \tag{3}$$

The first definition, P_{ms1}, sometimes called the *row-pitch*, gives an immediate sense of the volume coverage achievable at a given longitudinal resolution that is directly comparable among scanners with different numbers of rows. A four-slice scanner operating at $P_{ms1} = 6$ with 5-mm effective rows is going to cover six times as much distance in a given amount of time as a single-slice scanner operating at $P_{ss} = 1$ with 5-mm collimation. The second definition, P_{ms2}, sometimes called the *beam-pitch*, gives an immediate sense of the relative radiation dose produced by scanners with different numbers of rows, if all other parameters are held constant. A single-slice scanner with $P_{ss} = 1$ will deliver comparable radiation dose to a multislice scanner of any number of rows operating at pitch $P_{ms2} = 1$. All other factors being equal, lower P_{ms2} correspond to higher doses than do higher P_{ms2} because of denser longitudinal sampling. Such denser sampling may, of course, lead to better image quality that justifies the increased dose, and that topic will be addressed further in Sec. 5.

2. Clinical Applications

The increased volume coverage and longitudinal resolution achievable in multislice CT for a given imaging time has greatly improved the quality of a number of clinical CT applications, ranging from CT angiography to lung-cancer screening.

2.1. *CT Angiography*

CT Angiography (CTA) is the clinical application that has benefited most from the imaging speed improvements afforded by multislice scanners.[8] CTA involves the imaging of the vessels of the trunk, brain, or extremities after intravenous injection of an iodinated contrast bolus. Imaging is performed after waiting for the bolus to travel through the heart into the arteries. Because the distribution of the contrast agent changes constantly, the speed of imaging largely determines the resulting image quality. Multislice CT technology even makes it possible to perform bolus tracking: Following the bolus of contrast as it travels through the body, even over the distance from the heart to the feet that is traversed in just a few seconds in a runoff study. A volume rendering of a CTA study conducted on a 40-slice scanner is shown in Fig. 4.

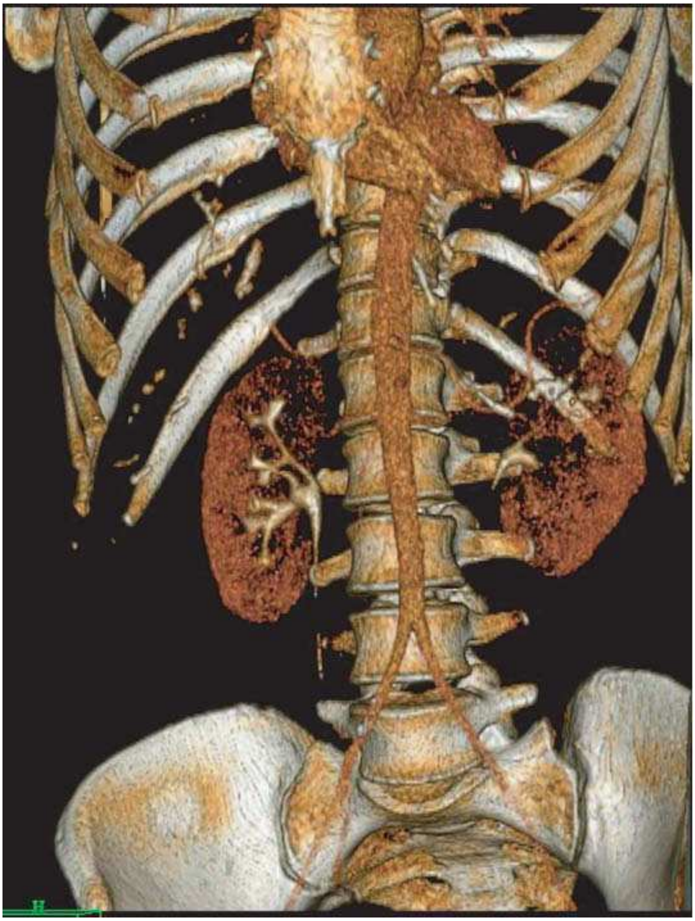


Fig. 4. Volume rendering of a CT angiography study showing bones, vessels, and enhanced organs. The study was performed on a Philips Brilliance 40 CT scanner.

2.2. *Cardiac imaging*

The heart is without a doubt the most difficult organ in the entire body to image. It contracts and twists relentlessly and rapidly — more than 100 times per minute in some patients — and sometimes irregularly. Moreover, the structures of greatest clinical interest, the coronary arteries, are only a few millimeters in diameter. Acquiring a snapshot of the beating heart, with motion frozen and the coronary arteries well resolved, has long been the holy grail of computed tomography imaging. Each new generation of multislice scanner and each increase in gantry rotation speeds has moved us closer to that goal.[9,10]

Early approaches to cardiac CT imaging with helical scanners concentrated mainly on obtaining images of the heart frozen at a single, somewhat extended phase of its cycle — diastole — when motion is minimized. This typically involves using an ElectroCardioGram (ECG) signal acquired in synchrony with the helical imaging to identify those sectors of the measured data that correspond to the extended diastolic phase. So long as each of these sectors spans at least a halfscan angular range (180 degrees plus the fan angle), it is possible to reconstruct an image of the heart at diastole that is relatively free of motion.[3] In single-slice scanners, of course, achieving the volume coverage needed to span the heart during a single breathhold requires some sacrifice in longitudinal resolution. Because of their improved volume coverage, multislice scanners allow these single-phase images to be acquired at much higher longitudinal resolution than do single-slice scanners, providing essentially isotropic resolution in each reconstructed volume.

Even more exciting, the speed and volume coverage of multislice scanners makes it possible to perform multi-phase cardiac imaging, in which one attempts to create snapshot views of the heart at a number of different phases of the cardiac cycle. The end result is a truly four-dimensional dataset: An animation of the beating heart. To reconstruct such multi-phase cardiac images, it is generally necessary to integrate data acquired during several cardiac cycles at a given phase of the heart cycle.[11,12] Figure 5 depicts a volume rendering of a cardiac study acquired on a 40-slice scanner.

In addition to visualizing the coronary arteries, cardiac CT is often employed to produce a calcium score, quantifying the degree of calcification in the coronary arteries, which is believed to be predictive of the severity of coronary disease and the likelihood of future coronary events.[9] The use of this technique as a screening exam is somewhat controversial, however, as some degree of coronary calcification seems to occur in almost all patients as they age, regardless of whether they are genuinely at risk of a heart attack. Thus it is unclear whether the benefits of an invasive procedure in someone with a high calcium score would necessarily outweigh the risks of such a procedure. As far as CT technology, the image-quality demands of calcium scoring are not quite as high as those for images used primarily for visualization of the coronaries, since calcium scoring generally involves averaging CT numbers over some region of interest and is thus tolerant of slightly lower resolution.

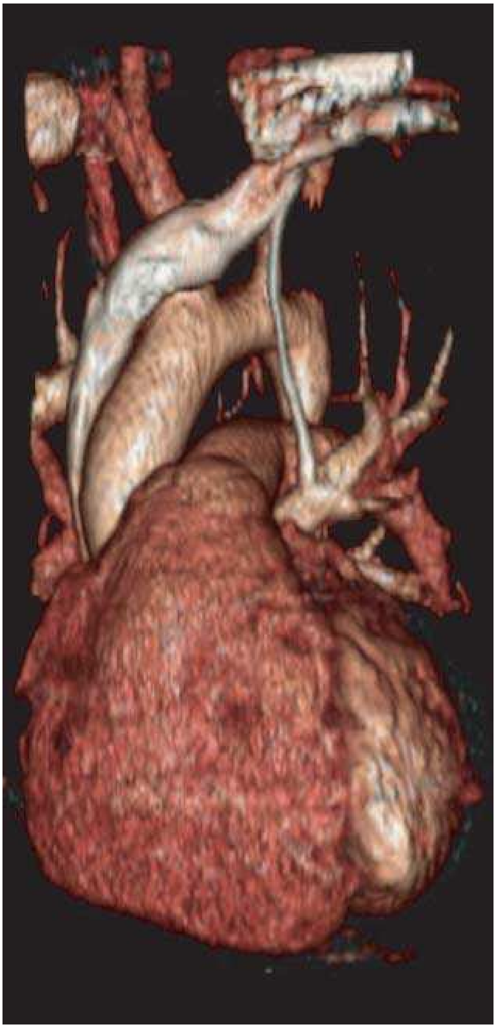


Fig. 5. The beating heart frozen in diastole in a dataset acquired on a Philips Brilliance 40 CT scanner.

2.3. *Multi-phase organ studies*

Iodinated contrast agents are also used to enhance soft-tissue contrast in certain organs such as the liver, spleen, and pancreas.[13,14] In these cases, the speed of multislice helical CT can be leveraged to obtain images of the entire organ at multiple different phases of contrast enhancement. Most commonly, one images one or more of the following four phases:

- Early arterial
- Late arterial

- Early venous
- Late venous

The differential contrast between normal and abnormal tissue may be greater at one of these points than at the others, such that a tumor, for instance, that is essentially invisible in one phase will stand out clearly in another.

2.4. *CT perfusion imaging*

While most contrast-enhanced CT studies serve simply to provide qualitative images of vascular morphology and differential organ enhancement, some aim to determine the quantitative estimates of important physiological parameters. The most common of these "functional CT" applications is perfusion imaging, which seeks to determine blood flow and transit time in organs such as the liver, kidneys, and brain.[15]

CT perfusion studies rely on the linear relationship between CT number and contrast agent concentration. Because of this, a curve plotting change in CT number versus time in a voxel provides much physiological information about the contrast-enhanced blood flowing to that voxel. Two different kinds of model are employed to extract physiological parameters from these enhancement curves: Compartmental models and linear-systems models. In compartmental models, the tissue is modeled as one of series of compartments.[16] So long as the arterial inflow and venous outflow to the compartment are also measured or estimated, it is possible to determine the perfusion of the tissue compartment. In linear-systems models, the tissue of interest is represented as a linear system having some enhancement response to a bolus of contrast agent.[17] By deconvolving the bolus shape from the enhancement curve, one can determine the tissue's inherent response function, and from this extract important parameters such as perfusion or Mean Transit Time (MTT) in the tissue.

Because of the dynamic nature of CT perfusion studies, they are generally performed in cine mode, in which the source and detectors revolve continuously without accompanying patient translation. In single-slice CT, this affords a dynamic view of a single slice of the patient. The great advantage of multislice CT for such studies is that it affords a dynamic view of multiple slices, and possibly even an entire organ for the large cone angle systems under development.

2.5. *CT fluoroscopy*

Just as conventional projection fluoroscopy allows physicians to view an X-ray projection image in real time at high frame rates, CT fluoroscopy allows physicians to view a tomographic slice in real time at high frame rates.[18] The tomographic view can be very useful during complex needle biopsies or catheter insertions. Single-slice CT scanners are, naturally, only able to provide a real-time view of one slice at a time, but multislice scanners can monitor a stack of slices simultaneously, allowing for three-dimensional visualization that can be valuable during invasive procedures amid complex anatomy.

2.6. *Virtual colonoscopy*

Colon cancer is second leading cause of cancer death in the United States. Most colon cancers begin as precancerous polyps; early detection and removal of these polyps is known to reduce the incidence of colon cancer.[19] The primary tool for early polyp detection until recently has been physical colonoscopy, in which an endoscope is inserted into the patient's colon and the resulting images visually examined for evidence of polyps. While effective, the technique is invasive, requiring sedation of the patient, and it carries a risk of morbidity associated with perforation of the bowel.

Improvements in CT technology, in particular the emergence of multislice scanners, have made possible a non-invasive form of "virtual colonoscopy," in which the entire colon is scanned during a single breathhold and the resulting reconstructed volume inspected by a radiologist for evidence of polyps.[19] The bowel may still need to be cleaned of stool prior to imaging, through consumption of a cleansing fluid in the hours before the exam, although the use of stool tagging agents can eliminate the need for this step. The patient is imaged in both prone and supine positions with the bowel insufflated. However, unlike conventional colonoscopy, sedation of the patient is not generally required. The acquired volume is typically viewed through a variety of multiplanar reformats (see Sec. 4) including a flayed, flattened view of the colon,[20] as well as in a fly-through mode that allows the radiologist to navigate through the lumen of the colon obtaining a view comparable to that obtained with an endoscope.[21]

2.7. *Lung cancer screening*

Lung cancer is the leading cause of cancer death in the United States. Yet unlike breast, prostate, and colon cancer, there is not currently a widely accepted screening protocol for lung cancer. Studies of chest radiography and sputum cytology conducted in the 1970s did not reveal a decrease in mortality that justified the expense and complications of widespread lung-cancer screening.[22]

However, interest in lung-cancer screening was reinvigorated in 1999 with a report by Henschke *et al.* on the Early Lung Cancer Action Project (ELCAP), which is studying the use of low-dose helical CT in 1,000 volunteers age 60 or greater with at least a 10 pack-year history of smoking.[23] The authors have thus far concluded that "low-dose CT can greatly improve the likelihood of detection of small non-calcified nodules, and thus of lung cancer at an earlier and potentially more curable stage."

Despite the lack of long-term mortality data for these patients, the preliminary ELCAP results were so promising that they spawned an immediate call by some to launch widespread lung-cancer screening programs using low-dose CT without bothering with further controlled studies. Professional organizations such as the Society of Thoracic Radiology have taken a more cautious approach, saying that existing results are inconclusive with regards to the effect on mortality; they do not advocate CT screening for lung cancer until further controlled studies are performed.[24]

A major National Lung Cancer Screening Trial (NLST) is currently being sponsored by the National Cancer Institute to investigate the efficacy of CT lung cancer screening.

The latest generation of multislice CT scanners allows for the imaging of the entire lung during a patient breathhold with isotropic millimeter or even submillimeter resolution. This should improve the detectability of small, subtle lung nodules that could easily be blurred into the background if imaging were performed with thicker slices, as would be required on a single-slice scanner.

3. Image Reconstruction Challenges

Multislice CT scanners have a conebeam geometry, in which all of the measured projections are oblique relative to the scanner axis; the outer rows of the detector acquire more oblique projections than do the inner rows. This is in contrast to a single-slice scanner in which all the measured projections are perpendicular to the scanner axis. Obliqueness introduces challenges for reconstruction, because in general, standard two-dimensional reconstruction algorithms such as Filtered Backprojection (FBP) cannot be applied directly to the conebeam data. For scanners with four or fewer rows, however, the cone angle is small enough that the obliqueness of the projections can be safely ignored. However, even in four-slice scanners, image reconstruction is not trivial, since when operating in helical mode, the longitudinal sampling patterns of the various rows interlace with one another in complex ways depending on the helical pitch. Algorithms are needed to interpolate among these samples in order to reconstruct slice images. Some of the approaches taken to this interpolation problem for four-slice scanners are discussed in Sec. 3.1 and the more general conebeam reconstruction problem is addressed in Sec. 3.2.

3.1. *Image reconstruction for four-slice CT*

3.1.1. *Sampling in four-slice CT*

In multislice helical CT, the data acquired by an N-row scanner are most naturally parametrized as $p_n^{(\mathrm{h})}(\gamma, \beta')$, where $n = 0, \ldots, N-1$ indexes the detector row number, γ denotes the angle between the projection line and the center line of the fan beam, β' denotes the angular position of the source, and the superscript h serves to remind us that this is a helical dataset. Typically, γ varies from $-\gamma_m$ to γ_m, where $2\gamma_m$ is the fan angle, and β' varies from 0 to $2\pi N_r$, where N_r is the number of revolutions in the scan.

In four-slice CT, ignoring the small cone angle is tantamount to making the multiple-parallel fanbeam assumption. That is, one assumes that the four 1D projections acquired by the detector at any given view angle correspond to four transverse fanbeam projections, parallel to one another and spaced according to the true spacing of the oblique projections as they cross the scanner rotation axis. This is illustrated in Fig. 6.

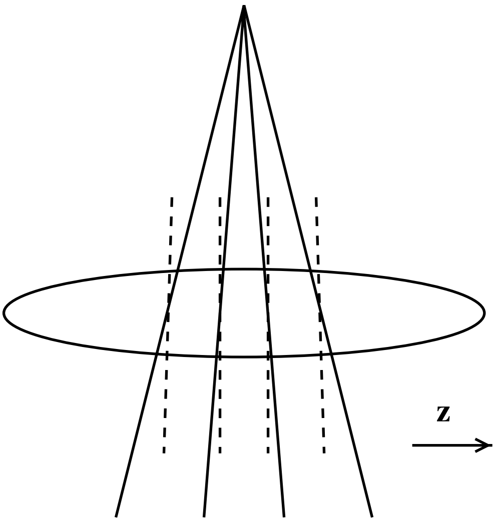


Fig. 6. Illustration of the multiple parallel fanbeam approximation used for image reconstruction in four-slice scanners. This is a side view with the cone angle greatly exaggerated for illustrative purposes. Each solid line represents an oblique fanbeam projection acquired by one of the four effective detector rows. Under the multiple parallel fanbeam approximation, the obliqueness of the projections is ignored, and they are assumed to be the purely transverse projections represented by the dashed lines, which cross the oblique lines at the axis of rotation.

Under this approximation, we can regard the measured helical CT data as providing samples of a three-dimensional (3D) function $p(\gamma, \beta, z)$ in the space $\{\gamma, \beta, z\}$, where γ is as before and where

$$\begin{aligned} \beta &= \beta' \bmod 2\pi \\ z &= \frac{\Delta}{2\pi}\beta'. \end{aligned} \tag{4}$$

Here Δ is the table-translation distance per 2π revolution of the source.

$$\Delta = P_{ms1} w, \tag{5}$$

where P_{ms1} is the row-pitch defined in Eq. (2). The function $p(\gamma, \beta, z)$, if known fully, could be thought of as a continuous stack of 2D fanbeam sinograms.

Figure 7 illustrates the kind of sampling patterns that can arise in multislice helical CT. For a four-slice scanner operating at pitch 6, the samples of $p(\gamma, \beta, z)$ in the $\{\beta, z\}$ subspace for a fixed value of γ (in this case $\gamma = -\pi/12$) lie on the solid lines in the figure. Concentrating solely on the longitudinal sampling, we find that at each γ and β the solid line for detector row n provides *direct samples* at positions

$$z = i\Delta + \frac{\beta}{2\pi}\Delta + nw, \quad i = 0, \ldots, N_r - 1. \tag{6}$$

Of course, fanbeam data acquired over 2π is highly redundant, in the sense that each projection line is measured twice. In helical CT, the second measurement of

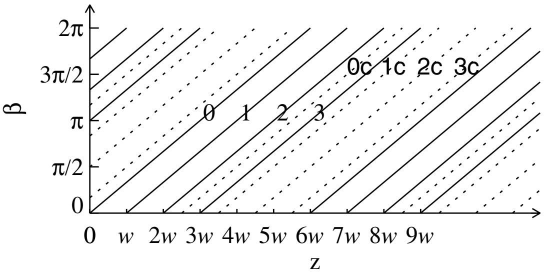


Fig. 7. Illustration of the sampling of $p(\gamma, \beta, z)$ in the $\{\beta, z\}$ subspace for a four-slice scan at pitch 6 and $\gamma = -\pi/12$. The *direct samples* lie on the solid lines and the *complementary samples* lie on the dashed lines. The lines are labeled with the detector row numbers, with the suffix "c" denoting the complementary data. Thus at each pair of γ and β we have eight interlaced sets of longitudinal samples, each with sampling interval $\Delta = 6w$.

a given projection line over the course of a 2π revolution takes place at a different longitudinal position along the object, and thus this redundancy provides an additional set of longitudinal samples. These so-called complementary samples lie on the dashed lines in Fig. 7, which satisfy

$$z = i\Delta + \frac{\beta}{2\pi}\Delta + nw + \Delta\left(\frac{\pi + 2\gamma}{2\pi}\right), \quad i = 0, \ldots, N_r - 1, \tag{7}$$

for detector row n and for each γ and β.

Overall, the relationship between the measured helical data $p_n^{(\mathrm{h})}(\gamma, \beta')$ and the function $p(\gamma, \beta, z)$ can be summarized as

$$\begin{aligned} &p\left(\gamma, \beta, i\Delta + \frac{\beta}{2\pi}\Delta + \alpha_m\right) \\ &\quad = \begin{cases} p_{m/2}^{(\mathrm{h})}(\gamma, \beta + 2\pi i), & m = 0, 2, \ldots, 2N-2 \\ p_{(m-1)/2}^{(\mathrm{h})}(-\gamma, \beta + 2\pi i + \pi + 2\gamma), & m = 1, 3, \ldots, 2N-1, \end{cases} \end{aligned} \tag{8}$$

for $i = 0, \ldots, N_r - 1$, where

$$\alpha_m = \begin{cases} \dfrac{m}{2}\dfrac{\Delta}{P_{ms1}}, & m = 0, 2, \ldots, 2N-2 \\ \dfrac{(m-1)}{2}\dfrac{\Delta}{P_{ms1}} + \Delta\left(\dfrac{\pi + 2\gamma}{2\pi}\right), & m = 1, 3, \ldots, 2N-1. \end{cases} \tag{9}$$

Thus at each pair of γ and β we have $2N$ sets of longitudinal samples, each with sampling interval Δ and with γ-dependent relative offsets α_m. This is known as *interlaced sampling.* The sampling pattern that arises at pitch 3 for a four-slice scanner is particularly simple: The direct samples from the four rows interlace to form a uniform set of samples with sampling interval equal to the longitudinal collimation w. (The zeroth and third rows follow the same trajectory through space and actually provide redundant samples.) The complementary samples also interlace

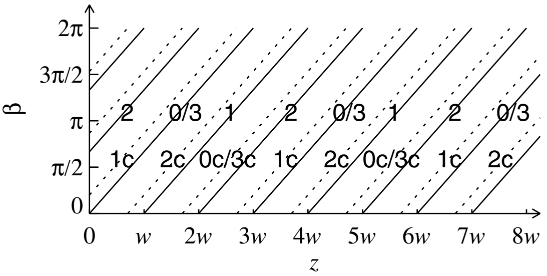


Fig. 8. Illustration of the sampling of $p(\gamma, \beta, z)$ in the $\{\beta, z\}$ subspace for a four-slice scan at pitch 3 and $\gamma = \pi/12$. Direct samples from the four rows interlace to form a uniform set of samples with sampling interval equal to the longitudinal collimation w. (The zeroth and third rows follow the same trajectory through space and actually provide redundant samples.) The complementary samples also interlace uniformly.

uniformly, and are offset from the direct samples by a γ-dependent interval. This is illustrated in Fig. 8.

3.1.2. *The interpolation problem*

In four-slice CT, the goal is to reconstruct images at a number of fixed longitudinal positions z_s. Conceptually, this is achieved by estimating complete fanbeam sinograms $\widehat{p}(\gamma, \beta, z_s)$, for $\gamma \in [-\gamma_m, \gamma_m]$ and $\beta \in [0, 2\pi)$ at a number of fixed values of z_s and then applying Fanbeam Filtered Backprojection (FFBP) to obtain an image. This is tantamount to interpolating samples on vertical lines in the $\{\beta, z\}$ subspace of Fig. 7 for all γ from the measured samples on the diagonal lines.

Specific development of interpolation and reconstruction approaches for four-slice helical CT have been reported by Taguchi and Aradate,[25] Schaller *et al.*,[26] and Hu.[6] Most are based on the use of linear interpolation, possibly with some additional longitudinal filtration. The approach of Schaller *et al.* first performs rebinning of fanbeam data to parallel-beam data prior to performing the interpolations.[26] A version of Hu's algorithm was implemented on General Electric's Lightspeed four-slice scanners. It applies to pitches near 3 and 6, and relies on the use of linear interpolation with some care taken in the handling of redundant samples.[6] This approach also exploits the linearity of the interpolation, filtration and backprojection steps to implement the algorithm in a pipelined fashion, such that each projection is multiplied by an appropriate interpolation and normalization weighting mask and then filtered and backprojected onto appropriate image grids as soon as it is acquired. The reconstructed images then build up correctly as additional projections are measured and processed.

We developed a new projection weighting function for interpolation and reconstruction of multislice helical computed tomography data with the hope of improving longitudinal resolution and reducing longitudinal aliasing in reconstructed volumes.[27] The weighting function is based on the application of the Papoulis generalized sampling theorem to the interlaced longitudinal samples acquired by the multislice scanner. We found that for pitch 3, our approach yielded high-quality images of the 3D Shepp-Logan phantom as well as longitudinal resolution superior to that of Hu's approach. The approach was not as successful at mitigating aliasing as we had hoped, however, because of the limitations of the multiple parallel fan-beam approximation. Sampling and aliasing questions will be discussed in greater detail in Sec. 5.

3.2. *Image reconstruction for conebeam CT*

For systems with more than four slices, it is not possible to ignore the cone angle during reconstruction if one wishes to obtain diagnostic-quality images. The development of true conebeam reconstruction algorithms has been an active area of research for over 20 years, and it is particularly active at present as CT manufacturers introduce scanners with ever larger cone angles requiring ever more accurate algorithms.

The ideal conebeam reconstruction algorithm would be theoretically exact, computationally efficient, and would allow for reconstruction of a Region-Of-Interest (ROI) from the minimal possible dataset. In particular, any practical CT conebeam algorithm must be applicable to the long-object problem, which arises when the imaging detector only views a limited longitudinal section of the patient at a time, as will always be the case in any realistic human CT scanner. On the other hand, if the design of the scanner is such that a more-than-minimal dataset is necessarily acquired, an ideal reconstruction algorithm should also be able to make use of and properly normalize the redundant data, as simply discarding it means that unnecessary radiation dose has been delivered to the patient.

While the mathematical details of conebeam reconstruction are beyond the scope of this chapter, we will introduce a few important geometrical concepts that will aid in the discussion below. The Tam-Danielsson (TD) window for a given projection view refers to the maximum extent of the projection of the source's helical trajectory onto the detector for that view.[28,29] Put differently, the TD window is the area of the detector bounded by the conebeam projections of the upper and lower turns of the helical trajectory bracketing the slice of interest. An example of a TD-window is illustrated in Fig. 9. It is known that a slice can be reconstructed exactly from knowledge of data only within the TD window.[28,29]

A pi-line is a line connecting two points on the source helix separated by less than a 360-degree rotation, as depicted in Fig. 9. It can be shown that each point within the cylinder bounded by the helical trajectory belongs to one and only one pi-line.[28,30] This implies that a region of interest within a patient can be reconstructed, in principle, by computing the reconstruction on the pi-line segments that intersect the ROI.[31,32] Finally, the term overscan refers to the additional angular range the

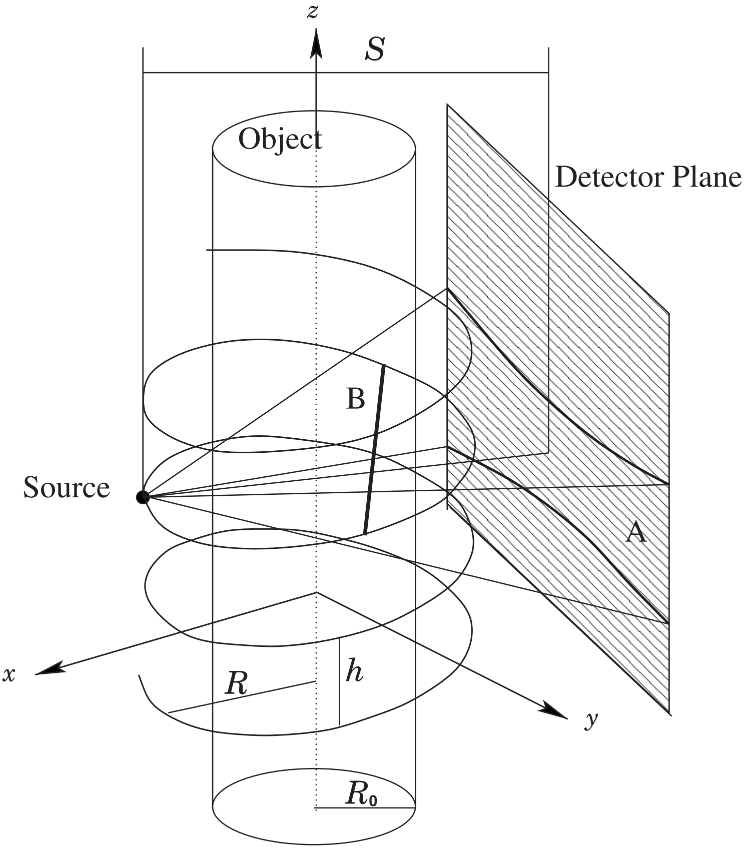


Fig. 9. Geometry for helical conebeam CT. The X-ray source and detector trace a helical trajectory as they revolve around the patient, assumed to lie within the cylinder of radius R_0 illustrated. The area A on the detector between the two curved lines representing the projections of the upper and lower turns of the helix bracketing the source position is referred to as the Tam-Danielson window. The line B illustrated is a pi-line, which is defined as any line connecting two points on the source helix separated by less than a 360-degree rotation.

source must travel beyond the upper and lower transverse planes bounding the ROI in order to allow for reconstruction of the ROI.

We can divide conebeam algorithms into five major families: Rebinning algorithms, Feldkamp-David-Kress (FDK) type algorithms, Radon-transform methods based on exploiting the connection between conebeam data and the 3D Radon transform, exact FBP-style algorithms based on a formula derived by Katsevich, and a new class of BackProjection Filtration (BPF) algorithms. In general, most of the algorithms involve tradeoffs among the ideal characteristics described above — sacrificing exactness for computational efficiency, for example — but recent developments minimize the need for serious tradeoffs and appear to be leading toward algorithms having many of the features described above.

3.2.1. *Rebinning algorithms*

Rebinning algorithms involve rearranging the measured data through changes of coordinates and interpolation into sets of 2D sinograms corresponding, in general, to projections along oblique planes or even hyperplanes.[33–39] These 2D sinograms are used to reconstruct the corresponding oblique planes or hyperplanes and then interpolation is applied among the resulting reconstructions in order to estimate the transverse slices of interest. While it is possible to derive exact rebinning algorithms,[40] these are very computationally demanding and dose inefficient. Most rebinning algorithms are thus approximate, but work quite well for scanners of up to 16 slices.

3.2.2. *FDK-style methods*

In 1984, Feldkamp, Davis, and Kress (FDK) proposed an approximate filtered backprojection style algorithm applicable to circular orbits.[41] Their algorithm had a particularly simple form:

- Perform shift-invariant, one-dimensional filtration of each row of the detector at each projection angle.
- Weight filtered data to compensate for path length and obliquity differences.
- Perform conebeam backprojection.

A number of researchers have extended the FDK algorithm to the case of a helical scan.[42,43] Early work used either fullscan or halfscan helical interpolation, with redundancy handled by use of Parker-like weights in the latter case.[44] Silver derived a general set of weights allowing for arbitrary scan ranges up to 360.[45] In all these cases, a computational drawback is that the data must be refiltered for each slice to be reconstructed because as they are formulated the filtration takes place after the slice-dependent weights are applied. A recent development by Kudo *et al.* addresses this shortcoming by formulating an algorithm in which the weights can be applied after filtration, allowing each projection to be filtered only once.[46]

Like rebinning algorithms, the accuracy of the approximate FDK method breaks down as the cone angle grows.

3.2.3. *Radon-transform based methods*

Grangeat produced one of the most significant advances in conebeam reconstruction in 1991 when he derived a relationship between conebeam data and the derivative of the three-dimensional Radon transform.[47] This allowed the 3D Radon transform along planes intersecting a given source vertex to be estimated from the conebeam projection corresponding to that vertex. The 3D Radon transform is itself readily invertible.[48]

While it represented a major breakthrough in conebeam reconstruction research, the Grangeat algorithm as originally formulated could not be applied to the long object problem. Moreover, it was expressed in such a way as to require measurement of the complete conebeam dataset before processing, which does not allow for projections to be processed sequentially, as they are measured, which is the preferred mode of operation in medical CT.

Over the years, a number of researchers derived algorithms that addressed one or the other of these shortcomings, with sequential processing made possible by FBP-style formulations of the algorithm,[49,50] and the long-object problem handled through approximations that lead to several quasi-exact algorithms, including modified Radon methods of Tam[29] and Kudo and Saito.[51] Finally, a few approaches were presented that allowed both sequential processing and the handling of the long-object problem, including the zero-boundary method of Defrise *et al.*,[30] the virtual circle method of Kudo *et al.*[52] and a local-ROI method of Tam.[53] Overall computational burden and sensitivity to numerical errors among these methods is still high, however.

3.2.4. *Exact FBP-type algorithms*

In 2002, Katsevich made a major contribution to conebeam reconstruction with the derivation of the first exact, FBP-style conebeam reconstruction algorithm, involving only one-dimensional, shift-invariant filtration.[54–56] In general, algorithms derived from this formula are more efficient computationally than are the Radon-transform based algorithms. In its original form, the algorithm was presented in very general terms, without assuming a specific detection geometry. Other researchers have re-expressed the formula in a more practical form in terms of detector coordinates[57,58] and have also used it as the basis for deriving a quasi-exact 3-pi conebeam reconstruction algorithm.[59] Algorithms based upon Katsevich's formula generally require less overscan data than do the Radon-transform based algorithms for exact image reconstruction in long-object problems. However, at each projection angle, all of these existing algorithms require more data than the minimum data within the TD window and thus cannot reconstruct images from data containing transverse truncations.

3.2.5. *BPF algorithms*

Recently Zou and Pan have developed a new strategy that exploits the properties of pi-lines and the TD window for image reconstruction in helical conebeam CT.[31,32] Based upon this strategy they derived two algorithms, one of which is a Backprojection-Filtration algorithm (BPF)[31] and the other a FBP-type algorithm.[32] The FBP algorithm resembles Katsevich's algorithm operationally, but the BPF algorithm differs fundamentally from any existing algorithm. The algorithms require only a minimum scan in handling the long object problem. More importantly, the BPF algorithm can exactly reconstruct ROI images from projections containing transverse truncations, which no other existing algorithm can do. The algorithms provide a useful unified point of view for CT reconstruction, as existing algorithms for circular conebeam, fanbeam, and parallel-beam geometries can be readily derived as special cases of the general FBP-type algorithm.[60]

4. Image Presentation and Data Management Challenges

In the early years of computed tomography, when just a few slice images were acquired in step-and-shoot mode, it was common practice to print the reconstructed images on film and display them on light boxes for radiologists to interpret. When helical CT emerged in the early 1990s, producing studies comprising dozens of slices, this practice became impractical, and radiologists began interpreting most CT studies on computer displays, paging through the slices either manually or in automatic cine mode. The latest generation of multislice scanners, which produce hundreds of slices per study, can make even soft-copy slice-by-slice reading impractical. However, the isotropic resolution of the dataset makes possible whole new modes of presentation, including multiplanar reformats, maximum intensity projections, and volume renderings.[21] In some ways, this requires a very different approach by the radiologist, a shift from passively interpreting a set of preselected images to actively interrogating a massive dataset. A typical CT user interface is shown in Fig. 10.

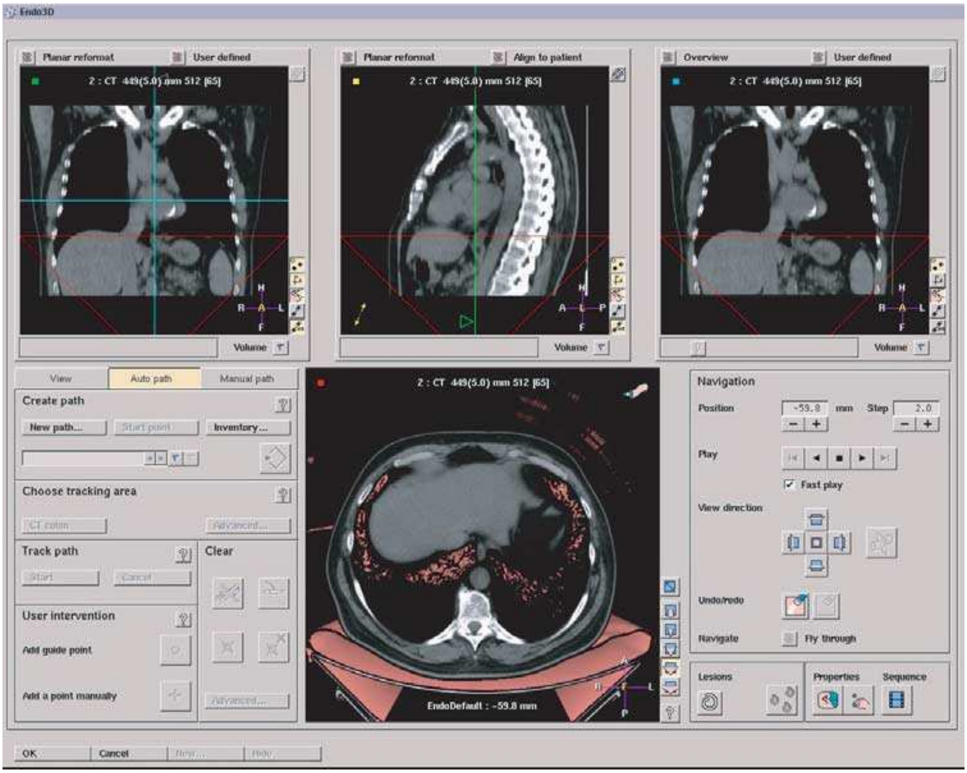
4.1. *Multiplanar reformats*

Given a stack of contiguous or even overlapping CT slice images, it is natural to think of the result as a volume that can be sliced and viewed in a variety of ways. For example, rather than just viewing the slices perpendicular to the long axis of the scanner that are reconstructed directly from the CT data (so-called axial slices), the radiologist could just as easily extract coronal or sagittal slices from the resulting volume. These are called MultiPlanar Reformats (MPRs).[21] In fact, given the isotropic resolution and voxel size of the latest generation of scanners, extracting these other perpendicular planes can involve little more than resampling the reconstructed volume. When resolution is nonisotropic, it may be necessary to interpolate among the reconstructed slices in order to obtain appropriate coronal or sagittal slices.

Of course, interpolation can be used to generate images other than simple coronal or sagittal cuts through a reconstructed volume. An oblique plane may bring more relevant anatomy into a single view, and curved surfaces can be useful for tracking vessels or unwinding tortuous structures such as the colon.[20]

4.2. *Simulated radiographs and Maximum Intensity Projections*

In addition to simply extracting a two-dimensional view from a three-dimensional volume, as in MPRs, one can also reproject the volume, collapsing three dimensions into two to obtain an image suitable for display. If this projection is performed according to the law of exponential attenuation, then one obtains a *simulated radiograph*.[61] If the projection is performed by simply taking the largest attenuation


Endo3D
Planar reformat
User defined
2 : CT 449(5.0) mm 512 [65]
Planar reformat
Align to patient
Overview
User defined
H
R
A
L
F
P
Volume
View
Auto path
Manual path
Create path
New path...
Start point
Inventory...
Choose tracking area
CT colon
Advanced...
Track path
Start
Cancel
Clear
User intervention
Add guide point
Add a point manually
Advanced...
Navigation
Play
View direction
Undo/redo
Navigate
Fly through
Lesions
Properties
Sequence
EndoDefault : -59.8 mm
OK
Cancel
Prev...
Help

Position -59.8 mm Step 2.0

☑ Fast play

Fig. 10. Modern CT workstations offer views of a variety of multiplanar reformats simultaneously, allowing for easier navigation and three-dimensional visualization by the radiologist.

value along each projection line, then one obtains a *Maximum Intensity Projection* (MIP).[62] The latter find widespread use in CTA, as contrast-enhanced vessels will generally be the most attenuating structures in the reconstructed volume and will thus be particularly well visualized on a MIP.

4.3. *Volume and surface rendering*

While MPRs and MIPs represent explicit slicing or collapsing of three-dimensional volumes to obtain two-dimensional images for viewing, techniques such as volume rendering and surface rendering seek to preserve some of the illusion of three-dimensional space in creating two-dimensional view for presentation to radiologists.

In volume rendering, each voxel in a volume is assigned an opacity value based on its pixel value.[63] The volume rendered image is simply a projection of the opacity values, where all projection values above a certain total opacity are assigned the same saturation value. By assigning fairly low opacity values to voxels with attenuation values characteristic of soft tissue, for example, it is possible to produce images that give the illusion of peering deep into the body, seeing bones and contrast-enhanced organs through a translucent matrix of overlying tissues. Figure 11 depicts such a view.

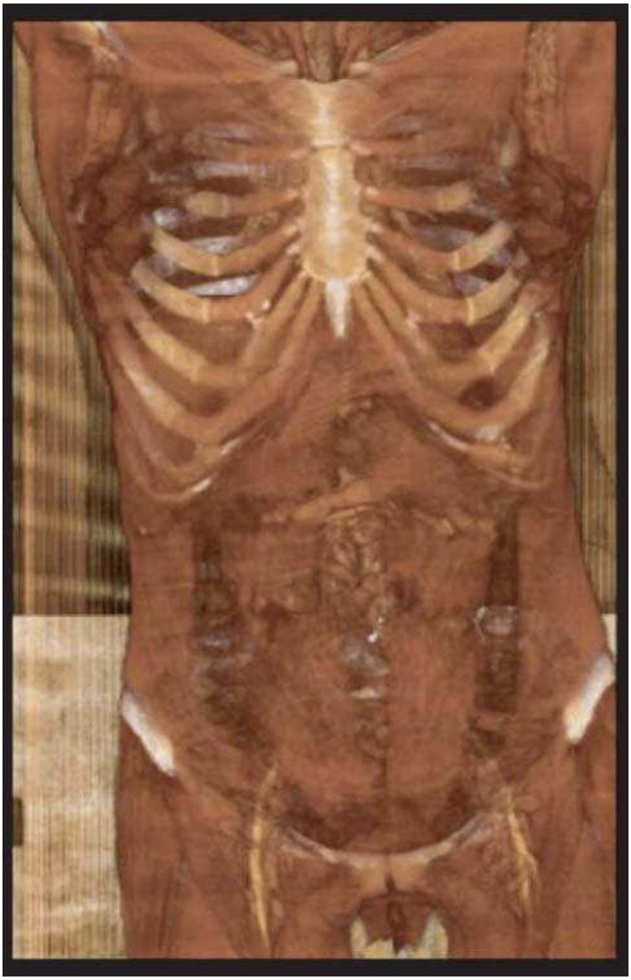


Fig. 11. Volume rendering of a CT dataset depicting primarily muscle and bone.

In surface rendering, one must first segment the anatomy of interest and determine its outer boundary.[64] In the simplest cases, such as the depiction of bone, this can often be achieved through simple pixel value thresholding. By specifying the reflectivity of the surface in question, a virtual light source, and a point of view, a simulated "optical" view can be created to mimic the visual appearance of the anatomy in question.

4.4. *Computer-aided diagnosis*

In order to help radiologists cope with the vast amounts of image data emerging from the latest CT scanners, a number of investigators have begun developing computer-aided diagnosis (CAD) systems to provide automated second readings of

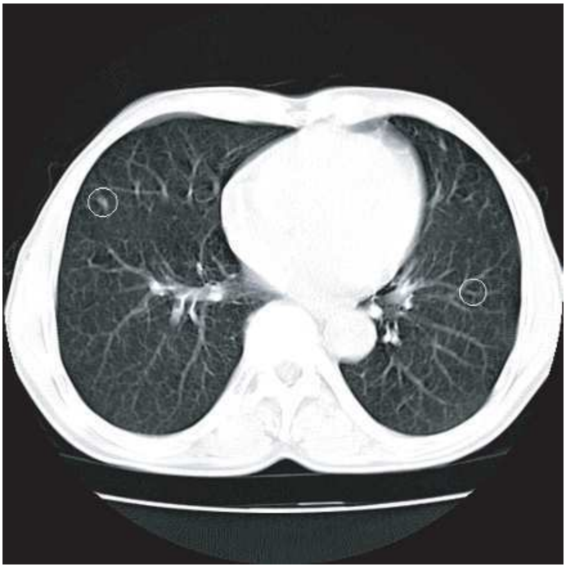


Fig. 12. The output of the University of Chicago computer-aided diagnosis system for a slice of a lung CT scan. The circle on the upper left of the image flags a subtle lung nodule. The circle on the right side of the image is a false positive.

certain kinds of CT studies.[19,65] The goal is not, of course, to replace the radiologist, but simply to flag areas of the image that the computer finds suspicious and that the radiologist may wish to reexamine. This kind of system would be particularly useful for screening studies for lung or colon cancer in which the image interpretation is fairly tedious and task is fairly well circumscribed: Finding a lung nodule or colon polyp. An example of a CT lung image with lesions flagged by a CAD system is shown in Fig. 12.

5. Sampling and Aliasing

CT scanners, like all digital imaging systems, acquire discrete sets of measurements, generally on a regularly sampled grid in some coordinate system. Failure to sample densely enough can lead to reconstructed images containing so-called *aliasing* artifacts, which can take the form of streaks or periodic modulations. In-plane sampling and aliasing effects in conventional, step-and-shoot CT are well understood and have helped inform CT system design. The move to helical scanning produced interesting new sampling patterns in the single-slice case, and Yen *et al.*[66] have shown that the longitudinal sampling engendered by pitches of clinical interest (1 or higher) does not strictly satisfy the sampling condition. This gives rise to spatially variant aliasing effects that are negligible near the isocenter and increase sharply near the edge of the field of view.

We have shown, however, that under certain conditions the second set of longitudinal samples provided by fanbeam redundancy does contain sufficient information to eliminate longitudinal aliasing in single-slice helical CT if exploited correctly.[67] In particular, we have developed a Fourier-based approach to longitudinal interpolation in single-slice helical CT and we have demonstrated that it can, in some cases, reduce or eliminate the aliasing in single-slice helical CT.[67] A more efficient, approximate spatial-domain version of the Fourier-based approach has also been developed.[68]

It is natural, then, to wonder about the aliasing properties that might arise in multislice helical CT reconstructions. This is a vast, rich topic. In the conebeam case, the problem has barely been addressed; one preliminary investigation is presented in Sec. 5.3 below. In the small cone angle regime where the multiple parallel fanbeam approximation is used for reconstruction, the resulting sampling patterns in multislice helical CT are highly pitch-dependent due to the interlacing of measurements from different detector rows, as we have already seen. The pitch dependence of longitudinal sampling in multislice helical CT begs the question of whether some pitches lead to particularly desirable, nonredundant sampling patterns, which is addressed in the next section.

5.1. *Preferred pitch in 4 slice scanners*

For four-slice scanners, the question of the existence of "preferred pitch" has been addressed previously in many ways and often with a different answer. Hu[6] has argued that pitches 3 and 6 are preferred because the bands of projection-ray dependent complementary samples described in Sec. 3.1 are centered between the direct samples and thus produce average sampling intervals equivalent to those in single-slice helical CT operating at pitches 1 and 2, respectively. These pitches thus both lead to a relatively favorable isocenter Slice Sensitivity Profile (SSP), which is the system response to an input that is an impulse in the longitudinal direction and circularly symmetric in the transverse plane.[69] Wang and Vannier[70] performed a "sensitivity analysis" of central detector channel sampling patterns in multislice helical CT and reached a nearly opposite conclusion, arguing that pitch 6 is distinctly suboptimal relative to other nearby pitches and that a pitch slightly less than 3 is to be preferred to pitch 3 itself.

Both of these sampling analyses are limited in their scope, however, in that they only truly apply to the longitudinal sampling for the central detector channel, and thus really only predict performance at the system isocenter. Given the extreme variation in longitudinal sampling patterns between central and peripheral detector channels and the spatially variant nature of aliasing effects, we felt it important to develop a more global analysis accounting for the entirety of the longitudinal sampling that arises at each pitch. We calculated the longitudinal aliasing that arises when imaging a cylindrical phantom similar to that employed by Yen *et al.*,[66] comprising adjacent cylindrical disks of alternating high and low densities, illustrated schematically in Fig. 13. The phantom has a square wave longitudinal profile and aliasing can be easily detected and measured in reconstructed volumes.

Working under the multiple parallel fanbeam approximation, we computed a global contrast-to-aliased noise (CN_aR) figure of merit for each pitch, which we plot in Fig. 14. Higher values of CN_aR are better. As expected, the curve is far from monotonic. Local maxima are evident around pitches 3, 4, 5, and 6. Some complicated variation is evident between pitches 1 and 2, with pitch 2 being a clear minimum in the curve. Other minima occur at half-integer pitches.

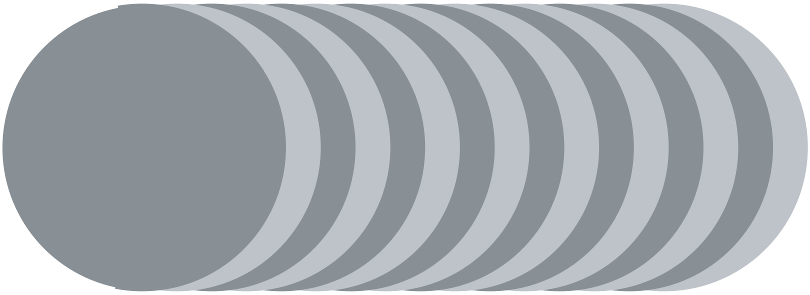


Fig. 13. Schematic illustration of the numerical phantom used in the aliasing studies described in the text. It comprises cylindrical disks of alternating high and low density.

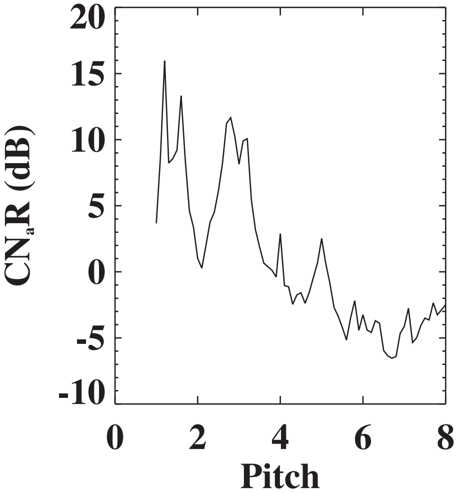


Fig. 14. Contrast-to-aliased-noise ratio (CN_aR), in decibels, versus helical pitch for a 4-slice scanner with longitudinal collimation width 2.5 mm at the isocenter imaging an object with sinusoidal longitudinal variation of frequency $f_0 = 1/3.175\,\text{mm}^{-1}$. Higher values of CN_aR are better. This analysis suggests that pitches near 3, 4, and 5 are optimal and half-integer pitches are suboptimal for four-slice with narrow-kernel interpolation.

The analysis described implicitly blends sampling and interpolation effects. In the results presented straightforward linear interpolation was employed. In practice, straightforward linear interpolation is sometimes avoided because frequent "changeovers" in the pairs of detector rows contributing to a given slice tend to produce artifacts in reconstructed images. In general, either attention is restricted to pitches where linear or quasi-linear approaches can be applied safely[6] or broader, adaptive z-filtering interpolation approaches are employed.[25,26] We do not expect that the use of a different narrow-kernel interpolation approach would fundamentally alter the conclusions of this study. Indeed, a related study by Bricault and Ferretti, in which they introduced the concept of a weighted radiation profile and characterized its width and heterogeneity reached similar conclusions regarding the relative value of different pitches for narrow-kernel interpolation. For wide-kernel interpolation, Bricault has confirmed that Schaller *et al.* are correct in their assertion that no pitches are inherently optimal or preferred.

5.2. *Cone-angle induced aliasing in four-slice scanners*

The study just described and, indeed, all previous studies of four-slice CT sampling and aliasing effects tacitly ignored the small cone angle. In order to examine its effect on aliasing, we studied the differences between the aliasing arising in four-slice CT and that arising in single-slice helical CT in situations when their longitudinal sampling is ostensibly equivalent.

Specifically, we examined longitudinal aliasing properties in four-slice scanners for helical pitches 3 and 6, which have effective longitudinal sampling intervals equivalent to those in single-slice helical CT operating at pitches 1 and 2, respectively. While these equivalences have been supported by comparative studies of slice-sensitivity profiles in single- and multislice helical CT, artifacts have been observed in pitch-3 and pitch-6 multislice images that were not evident in their purported single-slice counterparts.

Our study attributed these differences to aliasing arising in the multislice reconstructions that is not present in the single-slice counterparts. We found that the aliasing has two principal origins: Sampling effects similar to those in the single-slice case and cone-beam effects. The difference between the multislice, pitch-3 and single-slice, pitch-1 results was attributed to the small cone angle in multislice helical CT, which introduces inconsistencies among the measurements of different detector rows. The difference between multislice, pitch-6 and single-slice, pitch-2 results was attributed to a combination of the cone angle and genuine differences in sampling patterns.

To give a sense of the potential image-quality effects of the longitudinal aliasing discussed here, we show in Fig. 15 sagittal slices through two reconstructions of a square-wave phantom of period 3.175 mm. Projections were simulated for single-slice, pitch-1 and multislice, pitch-3 acquisitions. Each image corresponds to half of a sagittal slice through the cylinder, with the rotation axis at the top of the image.

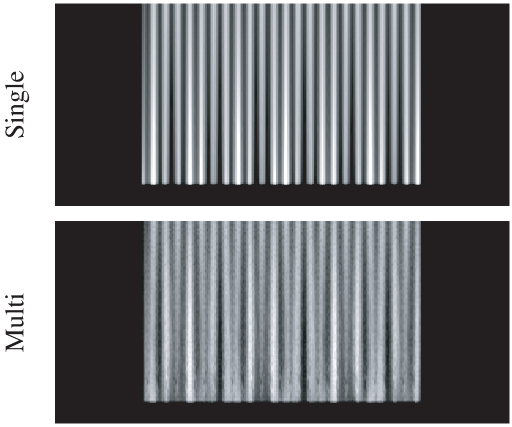


Fig. 15. Sagittal slices through reconstructions of a square-wave phantom of period 3.175 mm for single-slice, pitch-1 and multislice, pitch-3 acquisitions. Each image corresponds to half of a sagittal slice through the cylinder, with the rotation axis at the top of the image. It can be seen that in both cases, aliasing effects are negligible near the isocenter but that aliasing increases with distance from the isocenter.

It can be seen that in both cases, aliasing effects are negligible near the isocenter, where the original square wave pattern of the phantom is well-resolved, but that aliasing increases with distance from the isocenter. The magnitude of the aliasing effect is seen to be worse for the four-slice image than for the single-slice image, as expected from the preceding discussion.

5.3. *Challenges for performing sampling analysis in conebeam scanners*

Conventional multidimensional sampling analysis[71–76] does not extend very naturally to true conebeam CT because the oblique nature of the projections makes it difficult to identify a coordinate system in which the sampling forms a lattice. Simply ignoring the cone angle and analyzing the longitudinal sampling engendered by helical scans is not sound since the effect of the cone angle on aliasing is significant even for four-slice scanners, as discussed above.

One potential approach to characterizing sampling in conebeam CT is to make use of Fourier crosstalk analysis,[77–79] a generalized, object-independent form of aliasing analysis in which one quantifies the transmission of Fourier components through an imaging system as well as the ability to distinguish different Fourier components from the measured data.

In principle, the Fourier crosstalk technique could be used to compare the sampling engendered at different pitches for a given scanner configuration, identifying tradeoffs and possible optimal scan parameters. Unfortunately, the technique is extremely and perhaps impractically computationally demanding for imaging systems with datasets as large as those acquired in CT. In order simply to demonstrate the potential application of Fourier crosstalk in this context, we conducted a preliminary study in which we simply compared three different scanner geometries with ostensibly equivalent volume coverage and average longitudinal sampling: A single-slice scanner operating at pitch 1, a four-slice scanner operating at pitch 3, and a 16-slice scanner operating at pitch 15.

We found that moving from a single-slice to a multislice geometry introduces longitudinal crosstalk characteristic of the longitudinal sampling interval between periods of individual each detector row, and not of the overall interlaced sampling pattern. This was attributed to data inconsistencies caused by the obliqueness of the projections in a multislice/conebeam configuration. However, these preliminary results suggest that the significance of this additional crosstalk actually decreases as the number of detector rows increases, which is an intriguing result that bears further investigation.

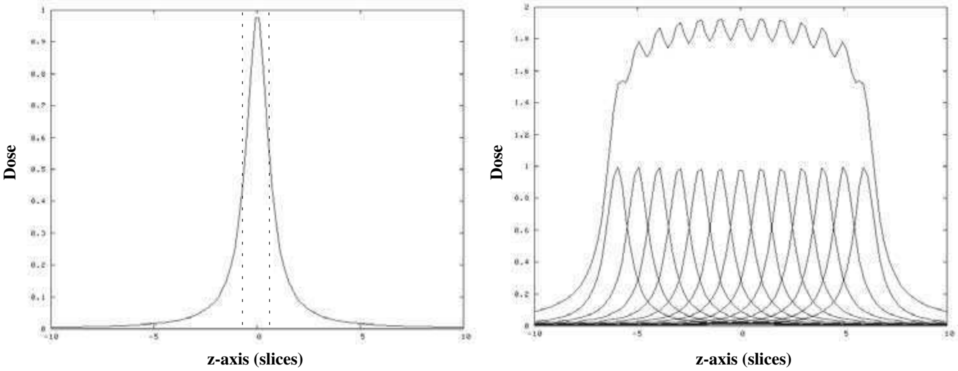
6. Dose

6.1. *Concerns about CT dose*

The new clinical capabilities of CT made possible by the technological innovations discussed in this chapter have spurred explosive growth in the number of CT exams performed annually, which has prompted concern over the associated rise in CT-related radiation dose to the population.[80] Typical CT scans result in doses of 1–10 mSv, and in some cases as high as 30 mSv, which is more than 10 times annual background radiation.[80] With some 60 million CT scans performed in the US,[81] CT is currently the largest source of medical radiation exposure and is second overall only to background radiation in the population as a whole.[81] The number of CT procedures has been growing at 10–15% annually in recent years, and that growth rate is not likely to abate.[80]

6.2. *Dose in single and multislice CT*

While multislice CT has spurred greater use of CT, the nature and distribution of the dose differs only slightly from that arising in single-slice CT. In single-slice, step-and-shoot CT with a 360 degree rotation, the dose delivered in imaging a single slice of a cylindrical phantom is, naturally, circularly symmetric. The dose along the central axis of the cylinder is peaked around the center of the slice being imaged, with a full-width half-maximum approximately equal to the nominal slice width. However, the dose profile plotted in Fig. 16(a) is seen to have fairly substantial extended tails due to the scattering of X-rays out of the slice being imaged. When scanning adjacent slices in step-and-shoot mode, the individual dose profiles simply add to give a multiple-scan dose profile, as shown in Fig. 16(b). Due to the superposition of the tails of the various single-slice profiles, the peak dose values can be considerably higher than the peak value of the single-slice dose profile. Operating a



(a) (b)

Fig. 16. (a) Dose profile in single-slice step-and-shoot CT. The vertical dashed lines denote the nominal slice thickness. The long tails on the dose profile are due to scattering of X-rays out of the slice being imaged. (b) Cumulative dose profile in single-slice step-and-shoot CT after imaging of multiple adjacent slices. Similar dose distributions arise in multislice CT operating in step-and-shoot mode as well as for both types of scanner operating in helical mode.

scanner in helical mode does not fundamentally change these overall dose distributions, certainly not on the central axis of the scanner. At peripheral points in the images volume there may be some modulation of the dose due to the nature of the helical scan.

Most of these basic properties carry over directly to the multislice case, with one significant difference. In single-slice CT, the nominal slice thickness is generally set by the use of pre-patient collimation. Both the umbra and penumbra portions of the transmitted beam are detected and contribute to the reconstructed image. In multislice CT, nominal slice thickness is determined mainly by the longitudinal size and grouping, if any, of the detector rows. Prepatient collimation is used to shape the beam into the desired overall longitudinal extent, but this means that the penumbra will fall either beyond the extent of the detector or onto inactive rows, as illustrated in Fig. 17. The X-rays in the penumbra region thus contribute dose to the patient but do not contribute to the measured data or reconstructed image. If the beam was shaped to allow the penumbra to illuminate the outermost active detector rows, the illumination of the different rows would be very inconsistent and this would lead to artifacts in reconstructed images. Thus, at least on this count, multislice CT is slightly less dose efficient than single-slice CT. However, the magnitude of the differential falls as the longitudinal extent of the beam grows. Thus the already small dose inefficiency of multislice scanners has grown almost negligible as it has become possible to use large beam collimations without sacrificing longitudinal resolution through the use of large, finely subdivided detector arrays.

That said, while the geometric inefficiency of multislice relative to single-slice CT has shrunk in recent years, the temptation to use multislice scanners in ways that increase patient dose has grown. The speed and volume coverage

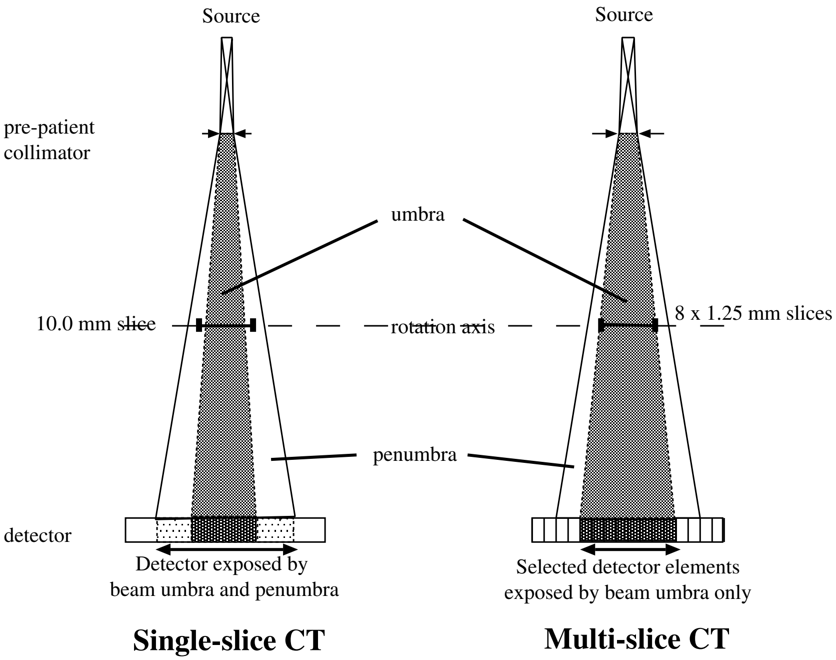

Source
Source
pre-patient
collimator
umbra
10.0 mm ~~slice~~
rotation axis
8 x 1.25 mm slices
penumbra
detector
Detector exposed by
beam umbra and penumbra
Selected detector elements
exposed by beam umbra only
Single-slice CT
Multi-slice CT

Fig. 17. In single-slice CT, both the umbra and penumbra of the beam are detected. The broadening effect of the umbra is accounted for in setting the prepatient collimation, which is set a little narrower than purely geometric considerations would dictate for a given slice size. In multislice CT, only the umbra of the beam is detected, with the prepatient collimation set according to precise geometric specifications to illuminate only the active detector elements of interest. The penumbra falls on inactive detector rows and thus constitutes "wasted" dose.

afforded by multislice scanners means that it is possible to image entire organs with very thin slices that may necessitate increasing X-ray exposure in order to maintain signal-to-noise ratios. It also makes it possible to rescan organs multiple times during a study in order to visualize different phases of contrast enhancement. When considering such studies, the tradeoff between additional dose and the diagnostic information to be gained must always be carefully weighed.

6.3. *Dose reduction strategies*

Due to the rising concerns about CT dose, a number of investigators and manufacturers have developed new techniques and imaging protocols aimed at reducing dose without compromising image quality. One promising approach, already implemented on a number of commercial scanners, is automatic tube current modulation,[82–84] in which the tube output is dynamically modulated based on the thickness of the anatomy expected in a given projection view. For example, when imaging the thorax, fewer X-rays are needed to obtain good measurement statistics when acquiring an anterior-posterior view than when acquiring a lateral view. Manufacturers and investigators are also proposing more individual tailoring of protocols to patient size and shape,[85–90] especially for pediatric patients. Finally, researchers are examining the use of iterative, statistically principled algorithms, either for full-blown image reconstruction[91–102] or for data smoothing,[103] in order to allow high quality images to be obtained from measurements made at lower radiation doses.

Acknowledgments

The author wishes to thank Dr. Michael Vannier of the Department of Radiology at the University of Chicago for reviewing the clinical applications section and providing the clinical images.

References

1. S. Webb, *From the Watching of Shadows: The Origins of Radiological Tomography* (Bristol: IOP Publishing, 1990).
2. W. A. Kalender, *Computed Tomography* (Munich: Publicis MCD Verlag, 2000).
3. J. Hsieh, *Computed Tomography: Principles, Design, Artifacts, and Recent Advances* (Bellingham: SPIE Press, 2003).
4. W. A. Kalender, W. Seissler, E. Klotz and P. Vock, Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation, *Radiology* **176** (1990) 181–183.
5. K. Taguchi and H. Aradate, Algorithm for image reconstruction in multi-slice helical CT, *Med. Phys.* **25** (1998) 550–560.
6. H. Hu, Multi-slice helical CT, *Med. Phys.* **26** (1999) 5–18.
7. J. Hsieh, O. E. Gurman and K. F. King, Investigation of a solid-state detector for advanced computed tomography, *IEEE Trans. Med. Imag.* **19** (2000) 930–940.
8. A. Napoli, D. Fleischmann, F. P. Chan, C. Catalano, J. C. Hellinger, R. Passariello and G. D. Rubin, Computed tomography angiography: State-of-the-art imaging using multidetector-row technology, *J. Computer-Assisted Tomogr.* **28**, Suppl 1 (2004) S32–45.
9. U. J. Schoepf, C. R. Becker, B. M. Ohnesorge and E. K. Yucel, CT of coronary artery disease, *Radiology* **232** (2004) 18–37.
10. P. Schoenhagen, S. S. Halliburton, A. E. Stillman, S. A. Kuzmiak, S. E. Nissen, E. M. Tuzcu and R. D. White, Noninvasive imaging of coronary arteries: Current and future role of multi-detector row CT, *Radiology* **232** (2004) 7–17.
11. M. Kachelreiß and W. A. Kalender, Electrocardiogram-correlated image reconstruction from subsecond spiral computed tomography scans of the heart, *Med. Phys.* **25** (1998) 2417–2431.
12. M. Kachelreiß, S. Ulzheimer and W. A. Kalender, ECG-correlated image reconstruction from subsecond multi-slice spiral CT scans of the heart, *Med. Phys.* **27** (2000) 1881–1902.
13. W. D. Foley, T. A. Mallissee and M. D. Hohenwalter, Multiphase hepatic CT with multirow detector CT scanners, *Am. J. Roentgenol.* **175** (2000) 679–685.
14. S. Fenchel, D. T. Boll, T. R. Fleiter, H. J. Brambs and E. M. Merkle, Multislice helical CT of the pancreas and spleen, *Eur. J. Radiol.* **45**, Suppl 1 (2003) S59–72.
15. K. A. Miles and M. R. Griffiths, Perfusion CT: A worthwhile enhancement, *Br. J. Radiol.* **76** (2003) 220–231.

16. M. Blomley and P. Dawson, Theoretical aspects and physiological models, in *Functional Computed Tomography* (K. Miles, M. Blomley and P. Dawson, eds.), pp. 47–60 (Oxford: Isis Medical Media, 1997).
17. D. G. Nabavi, A. Cenic, R. A. Craen, A. W. Gelb, J. D. Bennett, R. Kozak and T. Y. Lee, CT assessment of cerebral perfusion: Experimental validation and initial clinical experience, *Radiology* **213** (1999) 141–149.
18. J. J. Froelich and H. J. Wagner, CT-fluoroscopy: Tool or gimmick, *Cardiovasc. Intervent. Radiol.* **24** (2001) 297–305.
19. A. H. Dachman, S. Glick and H. Yoshida, Computed tomography colonography and colon cancer screening, *Semin. Roentgenol.* **38** (2003) 54–64.
20. E. G. McFarland, G. Wang, J. A. Brink, D. M. Balfe, J. P. Heiken and M. W. Vannier, Central axis determination and digital unraveling of the colon for spiral CT colography, *Academic Radiology* **4** (1997) 367–373.
21. G. D. Rubin, 3D imaging with MDCT, *Eur. J. Radiol.* **45**, Suppl 1 (2003) S37–41.
22. J. Collins, CT screening for lung cancer: Are we ready yet? *Wisc. Med. Journ.* **101** (2002) 31–34.
23. C. I. Henschke, D. I. McCauley, D. F. Yankelevitz, D. P. Naidich, G. McGuiness, O. S. Miettinen, D. M. Libby, M. W. Pasmantier, J. Koizumi, N. K. Altorki and J. P. Smith, Early lung cancer action project: Overall design and findings from baseline screening, *Lancet* **354** (1999) 99–105.
24. D. R. Aberle, G. Gamsu, C. I. Henshcke, D. P. Naidich and S. J. Swensen, A consensus statement of the Society of Thoracic Radiology: Screening for lung cancer with helical computed tomography, *J. Thoracic Imag.* **16** (2001) 65–68.
25. K. Taguchi and H. Aradate, Algorithm for image reconstruction in multi-slice helical CT, *Med. Phys.* **25** (1998) 550–560.
26. S. Schaller, T. Flohr, K. Klingenbeck, J. Krause, T. Fuchs and W. Kalender, Spiral interpolation algorithms for multi-slice spiral CT part I: Theory, *IEEE Trans. Med. Imag.* **19** (2000) 822–834.
27. P. J. La Rivière and X. Pan, Interlaced interpolation weighting functions for multislice helical CT, *Optical Engineering* **42** (2003) 3461–3470.
28. P. E. Danielsson, P. Edholm and M. Seger, Towards exact 3D-reconstruction for helical cone-beam scanning of long objects. A new detector arrangement and a new completeness condition, in *Proceedings of the 1997 International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine* (D. W. Townsend and P. E. Kinahan, eds.), pp. 141–144 (Pittsburgh, 1997).
29. K. C. Tam, S. Samarasekera and F. Sauer, Exact cone-beam CT with a spiral scan, *Phys. Med. Biol.* **43** (1998) 1015–1024.
30. M. Defrise, F. Noo and H. Kudo, A solution to the long-object problem in helical cone-beam tomography, *Phys. Med. Biol.* **45** (2000) 623–643.
31. Y. Zou and X. Pan, Exact image reconstruction on PI-line from minimum data in helical cone-beam CT, *Phys. Med. Biol.* **49** (2004) 941–959.
32. Y. Zou and X. Pan, Image reconstruction on PI-lines by use of filtered backprojection in helical cone-beam CT, *Phys. Med. Biol.* **49** (2004) 2717–2731.
33. F. Noo, M. Defrise, R. Clackdoyle and H. Kudo, Single-slice rebinning for helical cone-beam CT, *Phys. Med. Biol.* **44** (1999) 561–570.
34. M. Kachelreiß, S. Schaller and W. A. Kalender, Advanced single-slice rebinning in cone-beam spiral CT, *Med. Phys.* **27** (2000) 754–772.
35. M. Defrise, F. Noo and H. Kudo, Rebinning based algorithms for helical cone-beam CT, *Phys. Med. Biol.* **46** (2001) 2911–2937.

36. K. Stierstorfer, T. Flohr and H. Bruder, Segmented multi-plane reconstruction: A novel approximate reconstruction scheme for multi-slice spiral CT, *Phys. Med. Biol.* **47** (2002) 2571–2581.
37. X. Tang, Matched view weighting in tilted-plane-based reconstruction to suppress helical artifacts and optimize noise characteristics, *Med. Phys.* **30** (2003) 2912–2918.
38. M. Defrise, F. Noo and H. Kudo, Improved two-dimensional rebinning of helical cone-beam computerized tomography data using John's equation, *Inv. Prob.* **19** (2003) S41–S54.
39. T. Flohr, B. Ohnesorge, H. Bruder, K. Stierstorfer, J. Simon, C. Suess and S. Schaller, Image reconstruction and performance evaluation for ECG-gated spiral scanning with a 16-slice CT system, *Med. Phys.* **30** (2003) 2650–2662.
40. S. Schaller, F. Noo, F. Sauer, K. C. Tam, G. Lauritsch and T. Flohr, Exact Radon rebinning algorithm for the long object problem in helical cone-beam CT, *IEEE Trans. Med. Imaging* **19** (2000) 361–375.
41. L. A. Feldkamp, L. C. Davis and J. W. Kress, Practical cone-beam algorithm, *J. Opt. Soc. Am. A* **1** (1984) 612–619.
42. X. Yan and R. M. Leahy, Cone-beam tomography with circular, elliptical and spiral orbits, *Phys. Med. Biol.* **37** (1992) 493–506.
43. G. Wang, T.-H. Lin, P. Cheng and D. M. Shinozaki, A general cone-beam reconstruction algorithm, *IEEE Trans. Med. Imaging* **12** (1993) 486–496.
44. D. L. Parker, Optimal short scan convolution reconstruction for fan-beam CT, *Med. Phys.* **9** (1982) 245–257.
45. M. D. Silver, A method for including redundant data in computed tomography, *Med. Phys.* **27** (2000) 773–774.
46. H. Kudo, T. Rodet, F. Noo and M. Defrise, Exact and approximate algorithm for helical cone-beam CT, *Phys. Med. Biol.* **49** (2004) 2913–2931.
47. P. Grangeat, Mathematical framework of cone-beam 3D reconstruction via the first derivative of the radon transform, in *Mathematical Methods in Tomography, Lecture Notes in Mathematics* (G. T. Herman, A. K. Lous and F. Natterer, eds.), pp. 66–97 (Springer-Verlag, Berlin, 1991).
48. M. Y. Chiu, H. H. Barrett and R. G. Simpson, Three-dimensional image reconstruction from projections, *J. Opt. Soc. Am.* **70** (1980) 755–762.
49. H. Kudo and T. Saito, Derivation and implementation of a cone-beam reconstruction algorithm for nonplanar orbits, *IEEE Trans. Med. Imag.* **13** (1994) 195–211.
50. M. Defrise and R. Clack, A cone-beam reconstruction algorithm using shift-variant filtering and cone-beam backprojection, *IEEE Trans. Med. Imag.* **13** (1994) 186–195.
51. H. Kudo and T. Saito, Fast and stable cone-beam filtered backprojection method for non-planar orbits, *Phys. Med. Biol.* **43** (1998) 747–760.
52. H. Kudo, F. Noo and M. Defrise, Quasi-exact filtered backprojection algorithm for long object problem in helical cone-beam tomography, *IEEE Trans. Med. Imaging* **19** (2000) 902–921.
53. K. C. Tam, Exact local regions-of-interest reconstruction in spiral cone-beam filtered backprojection CT: Theory, *Proc. SPIE* **3979** (2000) 506–519.
54. A. Katsevich, Theoretically exact FBP-type inversion algorithm for spiral CT, *SIAM J. Appl. Math.* **62** (2002) 2012–2026.
55. A. Katsevich, Analysis of an exact inversion algorithm for spiral cone-beam CT, *Phys. Med. Biol.* **47** (2002) 2583–2597.
56. A. Katsevich, An improved exact FBP algorithm for spiral CT, *Adv. Appl. Math.* **32** (2004) 681–697.

57. F. Noo, J. Pack and D. Heuscher, Exact helical reconstruction using native cone-beam geometries, *Phys. Med. Biol.* **48** (2003) 3787–3818.
58. Y. Zou and X. Pan, Three-term exact FBP reconstruction in cone-beam helical CT, in *IEEE Medical Imaging Conference Record* (S. D. Metzler, ed.), pp. M10–316 (Portland, Oregan, 2003).
59. C. Bontus, T. Köhler and R. Proksa, A quasiexact reconstruction algorithm for helical CT using a 3-Pi acquisition, *Med. Phys.* **30** (2003) 2493–2502.
60. X. Pan, D. Xia, Y. Zou and L. Yu, A unified analysis of FBP-based algorithms in helical cone-beam and circular cone- and fan-beam scans, *Phys. Med. Biol.* **49** (2004) 4349–4369.
61. S. Napel, G. D. Rubin and R. B. Jeffrey Jr., STS-MIP: A new reconstruction technique for CT of the chest, *J. Computer-Assisted Tomogr.* **17** (1993) 832–838.
62. P. J. Keller, B. P. Drayer, E. K. Fram, K. D. Williams, C. L. Dumoulin and S. P. Souza, MR angiography with two-dimensional acquisition and three-dimensional display, *Radiology* **173** (1989) 527–532.
63. J. K. Udupa, Three-dimensional visualization and analysis methodologies: A current perspective, *Radiographics* **19** (1999) 783–803.
64. M. Magnusson, R. Lenz and P. E. Danielsson, Evaluation of methods for shaded surface display of CT volumes, *Comput. Med. Imaging Graphics* **15** (1991) 247–256.
65. S. G. Armato, M. L. Giger and H. MacMahon, Automated detection of lung nodules in CT scans: Preliminary results, *Med. Phys.* **28** (2001) 1552–1561.
66. S. Y. Yen, C. H. Yan, G. D. Rubin and S. Napel, Longitudinal sampling and aliasing in spiral CT, *IEEE Trans. Med. Imag.* **18** (1999) 43–58.
67. P. J. La Rivière and X. Pan, Fourier-based approach to interpolation in single-slice helical CT, *Med. Phys.* **28** (2001) 381–392.
68. P. J. La Rivière and X. Pan, Anti-aliasing weighting functions for single-slice helical CT, *IEEE Trans. Med. Imag.* **21** (2002) 978–990.
69. A. Polacin, W. A. Kalender and G. Marchal, Evaluation of section sensitivity profiles and image noise in spiral CT, *Radiology* **185** (1992) 29–35.
70. G. Wang and M. W. Vannier, The effect of pitch in multislice spiral/helical CT, *Med. Phys.* **26** (1999) 2648–2653.
71. D. P. Petersen and P. Middleton, Sampling and reconstruction of wave-number limited functions in N-dimensional Euclidean space, *Inf. Control* **5** (1962) 279–323.
72. P. A. Rattey and A. G. Lindgren, Sampling the 2D Radon transform, *IEEE Trans. Acoust. Speech, Signal Process.* **29** (1981) 994–1002.
73. F. Natterer, Sampling in fan beam tomography, *SIAM J. Appl. Math.* **53** (1993) 358–380.
74. A. Faridani, An application of a multidimensional sampling theorem to computed tomography, *Cont. Math.* **113** (1990) 65–80.
75. L. Desbat, Efficient parallel sampling schemes in 3D tomography, *Comptes Rendues de l'Académie des Sciences* **324** (1997) 1193–1197.
76. L. Desbat, S. Roux, P. Grangeat and A. Koenig, Sampling conditions of the 3D fanbeam X-ray transform, in *Proceedings of the VIIth International Conference on Fully 3D Reconstruction in Radiology and Nuclear Medicine* (Y. Bizais, ed.), pp. Mo AM2–2, 2003.
77. H. H. Barrett and H. Gifford, Cone-beam tomography with discrete data sets, *Phys. Med. Biol.* **39** (1994) 451–476.
78. H. H. Barrett, J. L. Denny, R. F. Wagner and K. J. Myers, Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance, *J. Opt. Soc. Am.* **12** (1995) 834–852.

79. H. C. Gifford, *Theory and Application of Fourier Crosstalk: An Evaluator for Digital-System Design*. PhD thesis (The University of Arizona, Tucson, 1997).
80. D. P. Frush and K. Applegate, Computed tomography and radiation: Understanding the issues, *J. Am. Coll. Radiol.* **1** (2004) 113–119.
81. O. W. Linton and F. A. Mettler, National conference on dose reduction in computed tomography, emphasis on pediatrics, *AJR* **181** (2003) 321–329.
82. W. A. Kalender, H. Wolf, C. Suess, M. Gies, H. Greess and W. A. Bautz, Dose reduction in CT by on-line tube current control: Principles and validation on phantoms and cadavers, *Eur Radiol.* **9** (1999) 323–328.
83. H. Greess, H. Wolf, U. Baum, M. Lell, M. Pirkl, W. Kalender and W. A. Bautz, Dose reduction in computed tomography by attenuation-based on-line modulation of tube current: Evaluation of six anatomical regions, *Eur Radiol.* **10** (2000) 391–394.
84. D. Tack, V. D. Maertelaer and P. A. Gevenois, Dose reduction in multidetector CT using attenuation-based online tube current modulation, *AJR* **181** (2003) 331–334.
85. J. M. Boone, E. M. Geraghty, J. A. Seibert and S. L. Wootton-Gorges, Dose reduction in pediatric CT: A rational approach, *Radiology* **228** (2003) 352–360.
86. M. K. Kalra, S. Prasad, S. Saini, M. A. Blake, J. Varghese, E. F. H. Ef, J. H. Thrall and J. T. Rhea, Clinical comparison of standard-dose and 50% reduced-dose abdominal CT: Effect on image quality, *AJR* **179** (2002) 1101–1106.
87. I. R. Kamel, R. J. Hernandez, J. E. Martin, A. E. Schlesinger, L. T. Niklason and K. E. Guire, Radiation dose reduction in CT of the pediatric pelvis, *Radiology* **190** (1994) 683–687.
88. M. E. Mullins, M. H. Lev, P. Bove, C. E. O'Reilly, S. Saini, J. T. Rhea, J. H. Thrall, G. J. Hunter, L. M. Hamberg and R. G. Gonzalez, Comparison of image quality between conventional and low-dose nonenhanced head CT, *Am. J. Neuroradiol.* **25** (2004) 533–538.
89. S. R. Prasad, C. Wittram, J. A. Shepard, T. McLoud and J. Rhea, Standard-dose and 50%-reduced-dose chest CT: Comparing the effect on image quality, *AJR* **179** (2002) 461–465.
90. R. van Gelder, H. W. Venema, I. W. Serlie, C. Y. Nio, R. M. Determann, C. A. Tipker, F. M. Vos, A. S. Glas, J. F. Bartelsman, P. M. Bossuyt, J. Lameris and J. Stoker, CT colonography at different radiation dose levels: Feasibility of dose reduction, *Radiology* **224** (2002) 25–33.
91. S. H. Manglos, G. M. Gange, A. Krol, F. D. Thomas and R. Narayanaswamy, Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT, *Phys. Med. Biol.* **40** (1995) 1225–1241.
92. C. Kamphuis and F. J. Beekman, Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm, *IEEE Trans. Med. Imag.* **17** (1998) 1101–1105.
93. J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens and L. Mortelmans, Iterative reconstruction for helical CT: A simulation study, *Phys. Med. Biol.* **43** (1998) 729–737.
94. B. DeMan, J. Nuyts, P. Dupont, G. Marchal and P. Suetens, Reduction of metal streak artifacts in X-ray computed tomography using a transmission maximum a posteriori algorithm, *IEEE Trans. Nucl. Sci.* **47** (2000) 977–981.
95. P. Sukovic and N. H. Clinthorne, Penalized weighted least-squares image reconstruction in single and dual energy X-ray computed tomography, *IEEE Trans. Med. Imag.* **19** (2000) 1075–1081.
96. F. J. Beekman and C. Kamphuis, Ordered subset reconstruction for X-ray CT, *Phys. Med. Biol.* **46** (2001) 1835–1855.

97. B. De Man, J. Nuyts, P. Dupont, G. Marchal and P. Suetens, An iterative maximum-likelihood polychromatic algorithm for CT, *IEEE Trans. Med. Imag.* **20** (2001) 999–1008.
98. I. A. Elbakri and J. A. Fessler, Statistical image reconstruction for polyenergetic X-ray computed tomography, *IEEE Trans. Med. Imag.* **21** (2002) 89–99.
99. J. F. Williamson, B. R. Whiting, J. Benac, R. J. Murphy, G. J. Blaine, J. A. O'Sullivan, D. G. Politte and D. L. Snyder, Prospects for quantitative computed tomography imaging in the presence of foreign metal bodies using statistical image reconstruction, *Med. Phys.* **29** (2002) 2404–2418.
100. I. A. Elbakri and J. A. Fessler, Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation, *Phys. Med. Biol.* **48** (2003) 2543–2578.
101. I. A. Elbakri and J. A. Fessler, Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography, in *Proc. SPIE* **5032** (2003) 1839–1850.
102. I. A. Elbakri, *X-ray CT Image Reconstruction Using Statistical Methods*, PhD thesis (The University of Michigan, 2003).
103. P. J. La Rivière, Penalized-likelihood sinogram smoothing for low-dose CT, *Med. Phys.* **32** (2005), 1676–1683.

CHAPTER 2

TECHNIQUES IN MAGNETIC RESONANCE DIFFRACTIVE IMAGING AND THEIR APPLICATION

SATOSHI ITO* and YOSHIFUMI YAMADA
Department of Information and Control Systems Science
Graduate School of Engineering, Utsunomiya University
7-1-2 Yoto, Utsunomiya, 321-8585 Japan
**itohst@is.utsunomiya-u.ac.jp*

Magnetic resonance diffractive imaging is proposed as a new approach to Magnetic Resonance (MR) angiography. The expression of the Nuclear Magnetic Resonance signal is similar to the equation for the Fresnel diffraction of a 3D object in light or sound waves. The proposed technique offers the possibility of fast angiographic imaging and the online reconstruction of 3D volumetric images using the holographic technique. Static imaging experiments using an ultra-low-field MRI system are performed to verify the feasibility of the technique. It is shown that the images focused on an arbitrary plane can be reconstructed from data scanned in two-dimensions, even though blurred image data is superimposed on the image. Moreover, the 3D image can be observed in a coherent optical imaging system. This study demonstrates the possibility of the proposed method as a fast imaging technique for MR angiography.

Keywords: Magnetic resonance (MR); holography; Fresnel transform.

1. Introduction

Angiographic imaging is widely used for clinical diagnosis. However, the use of Magnetic Resonance (MR) angiography to collect three-dimensional data is a time consuming process. Decreasing the data acquisition time will therefore reduce the burden on patients and increase patient throughput.

In order to reduce the data acquisition time, a feasibility study on the use of magnetic resonance diffractive imaging, in which images containing depth information are derived from data scanned in two-dimensions, is performed. Although Nuclear Magnetic Resonance (NMR) phenomena do not demonstrate the characteristics of wave motion, they approximate the description of the wavefront in the Fresnel diffractive region by using nonlinear field gradients. We have studied an NMR imaging technique in which the NMR signal exhibits the form of the Fresnel integral equation of 2D-objects by scanning a nonlinear quadratic field gradient over the imaging region in two-dimensions.[1,2] We call this technique the "Fresnel transform imaging technique". The NMR signal has the equation similar to the wavefront equation, so MR images can be obtained using both the numerical reconstruction and the holographic reconstruction methods.[3,4] We have applied the holographic reconstruction method to the online reconstruction of 2D MR images.[3–6]

Diffractive tomography is frequently used to obtain the two-dimensional and three-dimensional distributions of objects based on the measurement of diffracted waves.[7] It has been applied to optical microscopes[8–10] and techniques with supersonic waves.[11–13]

The proposed technique offers the following advantages over conventional imaging for obtaining angiographic images:

(i) The capability to obtain an image focused on an arbitrary plane to derive depth-related information from two-dimensional image data which were obtained in the two-dimensional imaging.
(ii) The capability to reconstruct 3D holographic images from a hologram produced by the NMR signal. Thus, the real-time reconstruction of 3D images is possible in principle.

To confirm the effectiveness of the proposed technique, we performed an experiment to obtain the signal in MR diffractive imaging, and image reconstruction was performed either optically on an optical bench or digitally using a computer. Since NMR signals with a high Signal-to-Noise Ratio (SNR) could not be obtained with ultra-low-field MRI, static imaging using a tube model was performed. The practical application of the proposed technique is taken into consideration.

2. NMR Signal in MR Diffractive Imaging

In the NMR Fresnel transform imaging technique, the NMR signal has the form similar to the wavefront of 2D objects.[1,2] Considering the difference in diffraction equations between 2D and 3D objects is helpful in the design of magnetic fields gradients, which are required to obtain the NMR signal in the proposed imaging technique.

2.1. *Wavefront equation of 3D objects*

The diffractive wavefront scattered from an 3D object $g(x_0, y_0, z_0)$ in the Fresnel region, $u(x_i, y_i)|_{z=z_i}$, is described by Eq. (1)[14]

$$u(x_i, y_i)|_{z=z_i} = \frac{1}{j\lambda(z_i - z_0)} \exp\{jk(z_i - z_0)\} \cdot \iiint_{-\infty}^{\infty} g(x_0, y_0, z_0) \exp\left\{jk\frac{(x_i - x_0)^2 + (y_i - y_0)^2}{2(z_i - z_0)}\right\} dx_0\, dy_0\, dz_0, \tag{1}$$

where λ is the wavelength of the light source, k is the wave number and z_i represents the distance from the center of coordinates (x_0, y_0, z_0) to the screen. Equation (1) is rewritten by the convolution equation as follows:

$$u(x_i, y_i)|_{z=z_i} = g(x_i, y_i, z_i) * f(x_i, y_i, z_i), \tag{2}$$

where $f(x_i, y_i, z_i)$ is written as

$$f(x_i, y_i, z_i) = \frac{1}{j\lambda z_i} \exp\left\{jk\left[z_i + (x_i^2 + y_i^2)/2z_i\right]\right\}. \tag{3}$$

The function $f(x_i, y_i, z_i)$ is called the "point-spread function".

2.2. *NMR signal equation*

In this section, we examine the similarity between the equations of the NMR signal and those of the three-dimensional diffracted wavefront of Eq. (1). In NMR Fresnel transform imaging, the Fresnel-transformed signal can be obtained by scanning a quadratic field gradient over the imaging region.[1,2] The quadratic field and the NMR signal is given as follows:

$$\Delta B = b\left\{(x' - x)^2 + (y' - y)^2\right\}, \tag{4}$$

$$v(x', y') = \iint_{-\infty}^{\infty} \rho(x, y) \exp\left\{-j\gamma b\tau\left[(x' - x)^2 + (y' - y)^2\right]\right\} dx\, dy, \tag{5}$$

where b is a coefficient of a quadratic field gradient and (x', y') are the coordinates of the center of the quadratic field gradient, which can be fixed in an arbitrary location by the external field, $\rho(x, y)$ represents the spin density distribution in the subject, γ is the magnetogyric ratio and τ represents impressed time. A detailed explanation of the scanning of a quadratic field is given in Sec. 4.2.

To obtain NMR signals with expressions identical to Eq. (1), the coefficient of the quadratic field must be varied as a function of $1/z$, which makes the design of the magnetic field a difficult process. Therefore, we attempted to simplify the design of the magnetic fields and approximated the function $k/2(z_i - z_0)$ in Eq. (1) as the first order equation $(1 + \alpha z)$. Based on the condition above, we adopted a quadratic field gradient whose coefficient increases by the increment factor α in the z-direction, as shown in Eq. (6).

$$\Delta B = b(1 + \alpha z)\left\{(x' - x)^2 + (y' - y)^2\right\}. \tag{6}$$

We can obtain the target equation for the NMR signal written in Eq. (7) by scanning the quadratic field gradient in the x–y plane as in the Fresnel transform imaging technique.

$$v(x', y') = P \iiint_{-\infty}^{\infty} \rho(x, y, z) \exp\left\{-j\gamma b\tau(1 + \alpha z)\left[(x' - x)^2 + (y' - y)^2\right]\right\} dx\, dy\, dz, \tag{7}$$

where $\rho(x, y, z)$ represents the 3D-distribution of the spin density in the subject and P is a constant. In order to obtain a signal that approximates the form of Eq. (7), we used four-types of magnetic fields written as Eqs. (8) to (11) when the phase encoding and read-out directions are set in the x- and y-directions, respectively.

$$(a) \quad \text{quadratic field: } \Delta b(x, y, z) = b\,(1 + \alpha z)\left(x^2 + y^2\right); \tag{8}$$

$$(b) \quad \text{scanning field: } \Delta b_{0x}(z) = \sqrt{1 + \alpha z}\, b_{0x}; \tag{9}$$

$$(c)\quad \text{sweeping field: } \Delta b_{0y}(z) = (1+\alpha z)\, b_{0y}; \tag{10}$$

$$(d)\quad \text{field gradient: } \Delta G_{yz}(y,z) = (1+\alpha z)\, G_y y. \tag{11}$$

Equation (8) is a quadratic field gradient whose coefficient is varied in the z-direction. The center position of the quadratic field can be moved to an arbitrary position (x'), as shown in Eq. (6), by simultaneous application of the scanning field written as Eq. (9). Equation (10) is a field gradient in the z-direction under the presence of a field offset of b_{0y}, and Eq. (11) is a field gradient in the y-direction whose coefficient increments by α in the z-direction. The field gradient of Eq. (9) can be rewritten as $G_y y + G_{yz} yz$, thus Eq. (9) can be generated by field gradients for the y and yz directions. The fields shown in Eqs. (10) and (11) are required to make the signal equation having the Fresnel transform form into the read-out direction (y-direction).

Figure 1 shows a pulse sequence for MR diffractive imaging using the 3D time-of-flight method. The first step is to saturate the slab region in which the vessels are to be imaged by a pre pulse under the presence of the volume-selective gradient G_{sel}. A spoiling gradient, G_{sp}, is then applied to eliminate transverse magnetization in that region. In the second step, an RF $\theta°$ pulse is applied to excite the spins contained in the unsaturated inflow from adjacent slabs. In the third step, a quadratic field gradient, Δb, and a scanning field, Δb_{0x}, are applied simultaneously in the phase encoding direction for a period τ. The scanning field is applied to move the center of the quadratic field in the x- (or y-) direction, so that when the amount of scanning is set to be x', the expression of the quadratic field after scanning is

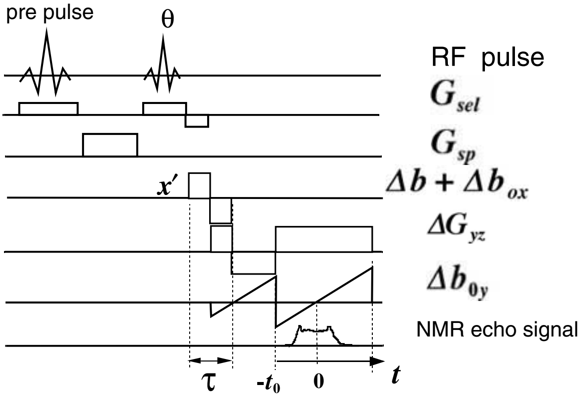


Fig. 1. Pulse sequence of MR diffractive imaging using the time-of-flight method. First, pre pulses saturate the slab in which the vessels are imaged under the presence of volume-selective gradient, G_{sel}. The spoiling gradient, G_{sp}, is then applied to eliminate transverse magnetization in that region. In the second step, the RF $\theta°$ pulse is applied to excite the spins contained in the unsaturated inflow from adjacent slabs. In the third step, a quadratic field gradient, Δb, and a scanning field, Δb_{0x}, are applied simultaneously in the phase encoding direction for a period τ period. Finally, a sweeping field, Δb_{0y}, and a field gradient, ΔG_{yz}, are applied simultaneously, by which the quadratic field is scanned in the y-direction.

$\Delta b = b(1+\alpha z)\left\{(x'-x)^2+y^2\right\}$. Finally, the sweeping field, Δb_{0y}, and the field gradient, ΔG_{yz}, are applied simultaneously, by which the quadratic field is equivalently scanned in the y-direction. The echo signal obtained in this sequence can be written as Eq. (12) when the origin of the time is set at the center of the NMR echo signal and the inversion time of ΔG_{yz} is set as $-t_0$. The NMR echo signal has a form that is significantly different to the signal obtained by conventional Fourier-transform based imaging.

$$\begin{aligned} v(t,y') &= \exp(-j\gamma b_0 t)\iiint_{-\infty}^{\infty}\rho(x,y,z)\exp\left\{-j\gamma b\tau\,(1+\alpha z)\left[(x'-x)^2+y^2\right]\right\} \\ &\quad\cdot\exp\left\{-j\gamma\,(1+\alpha z)G_y t+\gamma\int_{-t_0}^{t}(1+\alpha z)b_{0y}t dt\right\}dx\,dy\,dz \\ &= \exp\left\{-j\gamma b_0 t+\frac{\gamma b_{0y}}{2}{t_0}^2\right\} \\ &\quad\cdot\iiint_{-\infty}^{\infty}\rho(x,y,z)\exp\left\{-j\gamma b\tau(1+\alpha z)\left[(x'-x)^2+\left(y+\frac{G_y t}{2b\tau}\right)^2\right]\right\} \\ &\quad\cdot\exp\left\{j\gamma t^2(1+\alpha z)\left(\frac{{G_y}^2}{4b\tau}-\frac{b_{0y}}{2}\right)\right\}dx\,dy\,dz, \end{aligned} \tag{12}$$

where b_0 is a field offset for the static field from the on resonance condition. When the parameters are set to Eq. (13) and the variables are transformed, Eq. (12) can be rewritten as Eq. (14), and hence Eq. (7) is obtained

$$\frac{{G_y}^2}{4b\tau}=\frac{b_{0y}}{2},\quad y'=-\frac{G_y t}{2b\tau}, \tag{13}$$

$$\begin{aligned} v(x',y') &= P\exp(j\beta y')\iiint_{-\infty}^{\infty}\rho(x,y,z) \\ &\qquad\cdot\exp\left\{-j\gamma b\tau\,(1+\alpha z)\left[(x'-x)^2+(y'-y)^2\right]\right\}dx\,dy\,dz, \end{aligned} \tag{14}$$

where P and β are set as follows:

$$P=e^{-j\frac{\gamma b_{0y}}{2}{t_0}^2},\quad \beta=\frac{2\gamma b_0 b\tau}{G_y}. \tag{15}$$

3. Image Reconstruction

3.1. *Numerical reconstruction*

Two methods can be employed to reconstruct images from the Fresnel integral equation. One is a technique using the inverse Fourier transformation once after multiplying the quadratic phase term. The other method solves the convolution integral by inverse filtering.[2] In this paper, we will discuss the method for reconstructing images by the inverse filtering, in which the pixel width is not changed by the imaging parameters depending on the focal plane.

Let $\mathcal{F}_{xy}$ denote the Fourier transformation with respect to x and y. When $\mathcal{F}_{xy}$ is applied to Eq. (14), the following equation is obtained.

$$\begin{aligned}\mathcal{F}_{xy}[v(x',y')] = P\exp(j\beta y')\int_{-\infty}^{\infty}\exp\left(-j\frac{\pi}{2}\right)\frac{\pi}{\gamma b\tau(1+\alpha z)}\\ \cdot R(k_x,k_y,z)\exp\left\{j\frac{{k_x}^2+{k_y}^2}{4\gamma b\tau(1+\alpha z)}\right\}dz,\end{aligned} \tag{16}$$

where $R(k_x,k_y,z)$ denotes the Fourier transform of $\rho(x,y,z)$ with respect to the x and y coordinate system. Images focused on an arbitrary z' plane are obtained by multiplying the inverse function of the modulation transfer function on the z' plane to $\mathcal{F}_{xy}\{v(x',y')\}$, and taking the inverse Fourier transformation. Let $\rho'(x',y',z')$ denote the image focused on plane z', where $\rho'(x',y',z')$ is given by the following equation:

$$\begin{aligned}\rho'(x',y',z') &= \frac{1}{P}\exp\left\{j\left(\frac{\pi}{2}-\beta y'\right)\right\}\frac{\gamma b\tau(1+\alpha z')}{\pi}\\ &\quad\cdot\mathcal{F}_{xy}^{-1}\left[\mathcal{F}_{xy}[v(x',y')]\exp\left\{-j\frac{{k_x}^2+{k_y}^2}{4\gamma b\tau(1+\alpha z')}\right\}\right]\\ &= \frac{1}{P}\exp\left\{j\left(\frac{\pi}{2}-\beta y'\right)\right\}\mathcal{F}_{xy}^{-1}\left[\int_{-\infty}^{\infty}\frac{1+\alpha z'}{1+\alpha z}\right.\\ &\quad\left.\cdot R(k_x,k_y,z)\exp\left\{j\frac{\alpha(z-z')({k_x}^2+{k_y}^2)}{4\gamma b\tau(1+\alpha z)(1+\alpha z')}\right\}dz\right].\end{aligned} \tag{17}$$

We can obtain images focused on an arbitrary plane by substituting the coordinates of z' into Eq. (17).

The space described by Eq. (17) is considered to be identical to the signal space, because the images are reconstructed by inverse filtering. Let the sampling step of the NMR signal be $\Delta x'$ (and $\Delta y'$), and the pixel width of the reconstructed image be Δx_w (and Δy_w). The relation $\Delta x_w \approx \Delta x'$ holds, because the space described by Eq. (17) is obtained using the Fourier-transform twice in the inverse filtering algorithm (here, the inverse Fourier transform is considered as a Fourier transform in a broad sense).

Let the number of sampling points and the spatial resolution be N, Δx (and Δy), respectively. Then, the spatial resolution in the x-(y-)directions in the inverse filtering becomes either $\Delta x'$ ($\Delta y'$) or $\Delta x = \pi/\{\gamma b\tau(1+\alpha z)N\Delta x'\}$ whichever is greater. A detailed description of the spatial resolution in the Fresnel-transformed signal can be found in Yamada *et al.*[1,2]

$$\text{(i)}\quad \text{if}\quad \Delta x' \le \sqrt{\frac{\pi}{\gamma b\tau(1+\alpha z')N}},\quad \text{then } \Delta x \simeq \frac{\pi}{\gamma b\tau(1+\alpha z)N\Delta x'}, \tag{18}$$

$$\text{(ii)}\quad \text{if}\quad \Delta x' \ge \sqrt{\frac{\pi}{\gamma b\tau(1+\alpha z')N}},\quad \text{then } \Delta x \simeq \Delta x'. \tag{19}$$

3.2. *Holographic reconstruction*

The similarity between the equations of the NMR signal obtained by MR diffractive imaging and the wavefront of light waves implies that holographic reconstruction of MR images is possible by recording the NMR signals as holograms and reconstructing the images in a coherent optical imaging system. Turner proposed the holographic reconstruction of MR images and reconstructed images using the Fourier hologram produced by NMR signals obtained in the Fourier transform imaging in 1985.[15] The procedure to obtain MR images by the holographic method is briefly discussed in the following section.

3.2.1. *Principle of holography*

In holography, a reference light is superimposed on an object light $O(x_i, y_i)$ scattered from an object.[14,16] Assuming the case when the wavefront of the reference plane-wave light $R(x_i, y_i) = R_0 \exp(j\beta_{opt} x_i)$ is illuminated parallel to the x-axis, the intensity of the interferometric fringe, $H(x_i, y_i)_{opt}$, is given by the following relation:

$$\begin{aligned} H(x_i, y_i)_{opt} &= |O(x_i, y_i) + R(x_i, y_i)|^2 \\ &= |O(x_i, y_i)|^2 + |R_0|^2 + O(x_i, y_i) R_0 \exp(j\beta_{opt} x_i) \\ &\quad + O^*(x_i, y_i) R_0 \exp(-j\beta_{opt} x_i) \qquad (20) \\ &= |O(x_i, y_i)|^2 + |R_0|^2 + 2\,\mathrm{Re}[O(x_i, y_i) R_0 \exp\{j\beta_{opt} x_i\}]\,, \qquad (21) \end{aligned}$$

where the $*$ denotes the complex-conjugate, R_0 is the amplitude of the reference light and β_{opt} is a parameter determined from the wavelength of the light and the angle between the reference light and the object light. The first and second terms in Eq. (20) correspond to the intensities of the object light and the reference light, and the third and fourth terms in Eq. (20) and the third term in Eq. (21) contain information pertaining to the object light and its conjugate light, respectively. In conventional off-axis Fresnel holography, the first and second terms are not used in the formation of the image, so they can be replaced by a constant K as in Eq. (22)

$$H(x_i, y_i)_{opt} \approx K + 2\,\mathrm{Re}[O(x_i, y_i) R_0 \exp(j\beta_{opt} x_i)]\,. \qquad (22)$$

3.2.2. *NMR hologram*

The NMR signal in Eq. (14) can be rewritten as

$$v(x', y') = P \exp(j\beta y') v'(x', y'), \qquad (23)$$

where

$$v'(x', y') = \iiint_{-\infty}^{\infty} \rho(x, y, z) \exp\{-j\gamma b\tau(1 + \alpha z)[(x' - x)^2 + (y' - y)^2]\} dx\, dy\, dz. \qquad (24)$$

Comparing Eqs. (23) and (22), we find that $v'(x', y')$ corresponds to the object light $O(x_i, y_i)$ and $\exp(-j\beta y')$, giving an amplitude modulation to $v'(x', y')$, corresponds to the illumination by the reference plane-wave light in off-axis holography. Therefore, the NMR hologram is produced by superimposing the constant K' for the real part of the signal[15]

$$H_{NMR} = K' + \mathrm{Re}[v(k_x, k_y)], \tag{25}$$

where the value of constant K' should be of sufficient magnitude such that H_{NMR} is always positive.

The reconstruction of images from the NMR hologram is achieved by illuminating the hologram using a plane-wave laser light. Fresnel holograms can reproduce convergent and divergent light waves as well as direct transmitting light waves. A convergent light wave reconstructs an image at a certain focal length. Let us denote the focal length of the hologram, f_{hlg}, as the distance from the hologram to the plane in which the object image on the $z = 0$ plane is to be focused. f_{hlg} is given based on the Fresnel zone plate theory when the scan width of the quadratic field gradient is set to $x_{smax}(=N\Delta x')$, and the size of the hologram on the liquid crystal panel is L_h

$$f_{hlg} = \frac{\pi {L_h}^2}{\lambda\gamma b\tau {x_{smax}}^2}. \tag{26}$$

A divergent light wave is the reproduced light wave scattered from the object, enabling us to view the 3D image of the object through the hologram. Here, the 3D image is viewed at a distance from the hologram calculated by Eq. (26), and the size of the reproduced image approximates the size of the hologram's L_h. When the size of the hologram is small, the size of the reconstructed image becomes too small to observe without the aid of specialist equipment. Therefore, we have attempted to enlarge the size of the reconstructed image by increasing the size of the data matrix of the hologram. However, the focal length of the hologram is enlarged by the square of L_h by Eq. (26). A long focal length reduces the perspective because the reconstructed images are observed at a considerable distance from the view point. To solve this problem, we shortened the focal length of the hologram by introducing computerized post processing.

Reducing the focal length of the hologram is similar to the process of image reconstruction by inverse filtering. In the inverse filtering algorithm, the complex conjugate of the Fourier transform of the point spread function, which is called the "modulation transfer function", is multiplied in the Fourier transformed domain of the NMR signal such that blurring in the image is eliminated. The Short Focal Length Hologram (SFH) procedure is equivalent to an original image by back-propagating the diffracted wavefront of the object in optical holography. Since the back-propagation of the wavefront produced by NMR holography requires for a considerable distance, a portion of the back-propagation is executed numerically in the inverse filtering algorithm and the remainder is executed optically. The $\gamma b\tau$ value of the NMR signal is enlarged by this computer-assisted back propagation, resulting in a shortening of the focal length of the hologram. In SFH, the coefficient s, which approximates the value of $\gamma b\tau$ depending on the effect of SFH, is used. Let the pseudo-enlarged $\gamma b\tau$ be $\gamma b\tau_{large}$. The NMR signal obtained by SFH is written as

$$\begin{aligned}
v(x', y') &= P \exp\left\{ -j\left(\frac{\pi}{2} - \beta y'\right)\right\} \mathcal{F}_{xy}^{-1}\left[\int_{-\infty}^{\infty} \frac{s}{\gamma b\tau(1+\alpha z)}\right. \\
&\qquad \left. \cdot R(k_x, k_y, z) \exp\left\{ -\frac{{k_x}^2 + {k_y}^2}{4\gamma b\tau} + \frac{{k_x}^2 + {k_y}^2}{4s}\right\} dz\right] \\
&= P \exp\left\{ -j\left(\frac{\pi}{2} - \beta y'\right)\right\} \mathcal{F}_{xy}^{-1}\left[\int_{-\infty}^{\infty} \frac{s}{\gamma b\tau(1+\alpha z)}\right. \\
&\qquad \left. \cdot R(k_x, k_y, z) \exp\left\{ j\frac{[s - \gamma b\tau\,(1+\alpha z)]\,({k_x}^2 + {k_y}^2)}{4s\gamma b\tau\,(1+\alpha z)}\right\} dz\right] \\
&= P \exp(j\beta y') \iiint_{-\infty}^{\infty} \rho\,(x, y, z)\, \frac{\gamma b\tau_{large}(z)}{\gamma b\tau} \\
&\qquad \cdot \exp\left\{ -j\gamma b\tau_{large}(z)\,(1+\alpha z)\left[(x'-x)^2 + (y'-y)^2\right]\right\} dx\, dy\, dz, \qquad (27)
\end{aligned}$$

where

$$\gamma b\tau_{large}(z) = \frac{s}{s - \gamma b\tau(1+\alpha z)}\gamma b\tau. \qquad (28)$$

Equation (28) implies that parameter $\gamma b\tau_{large}$ becomes a value $s/\left\{s - \gamma b\tau\,(1+\alpha z)\right\}$ times $\gamma b\tau$. For instance, if s is set to be $1.1 \times \gamma b\tau$, $\gamma b\tau_{large}$ on the plane $z = 0$ becomes 11 times as large as $\gamma b\tau$, resulting in a shortening of the focal length by 1/11 from Eq. (26).

Along with SFH, the difference in the focal length in front of and behind the object, i.e. the size of the object in the z-direction, is taken into consideration, which is important in stereo viewing applications. Figure 2 illustrates the shortening of the focal length. Let z_f and z_b be the coordinates of the front and rear side of the object. Since the focal length is a function of $\gamma b\tau$ and z from Eqs. (27) and (28), it can be written as $f_{hlg}(z, \gamma b\tau)$. Thus, $f_{hlg}(z_f, \gamma b\tau) - f_{hlg}(z_b, \gamma b\tau)$ is calculated as

$$\begin{aligned}
f_{hlg}(z_f, \gamma b\tau) - f_{hlg}(z_b, \gamma b\tau) &= \frac{\pi {L_h}^2}{\lambda\gamma b\tau x_{smax}^2}\left\{\frac{1}{(1+\alpha z_f)} - \frac{1}{(1+\alpha z_b)}\right\} \\
&= \frac{\pi {L_h}^2 \alpha(z_b - z_f)}{\lambda\gamma b\tau x_{smax}^2(1+\alpha z_f)(1+\alpha z_b)}. \qquad (29)
\end{aligned}$$

In the case when SFH is applied, $f_{hlg}(z_f, \gamma b\tau_{large}) - f_{hlg}(z_b, \gamma b\tau_{large})$ is calculated as

$$\begin{aligned}
&f_{hlg}(z_f, \gamma b\tau_{large}) - f_{hlg}(z_b, \gamma b\tau_{large}) \\
&\quad = \frac{\pi {L_h}^2}{\lambda\gamma b\tau x_{smax}^2}\left\{\frac{1}{\gamma b\tau_{large}(z_f)(1+\alpha z_f)} - \frac{1}{\gamma b\tau_{large}(z_b)(1+\alpha z_b)}\right\} \\
&\quad = \frac{\pi {L_h}^2 \alpha(z_b - z_f)}{\lambda\gamma b\tau x_{smax}^2(1+\alpha z_f)(1+\alpha z_b)} \\
&\quad = f_{hlg}(z_f, \gamma b\tau) - f_{hlg}(z_b, \gamma b\tau). \qquad (30)
\end{aligned}$$

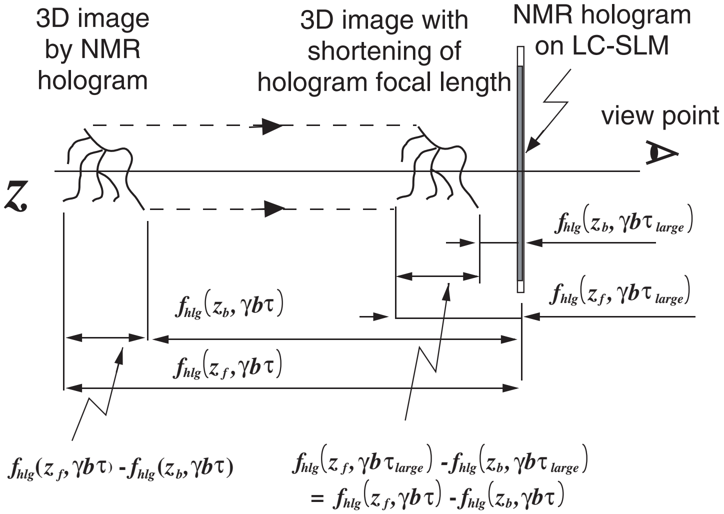

3D image by NMR hologram
3D image with shortening of hologram focal length
NMR hologram on LC-SLM
view point
z
$f_{hlg}(z_b, \gamma b\tau_{large})$
$f_{hlg}(z_f, \gamma b\tau_{large})$
$f_{hlg}(z_b, \gamma b\tau)$
$f_{hlg}(z_f, \gamma b\tau)$
$f_{hlg}(z_f, \gamma b\tau) - f_{hlg}(z_b, \gamma b\tau)$
$f_{hlg}(z_f, \gamma b\tau_{large}) - f_{hlg}(z_b, \gamma b\tau_{large})$
$= f_{hlg}(z_f, \gamma b\tau) - f_{hlg}(z_b, \gamma b\tau)$

Fig. 2. Focal length of hologram can be shortened by enlarging the $\gamma b\tau$ value numerically without changing the difference in the focal length in front of and behind the object.

Equation (30) indicates that the distance is independent of the application of SFH, but is enlarged by the square of size of the hologram L_h. Therefore, the size of the hologram should only be enlarged by a factor by which the distance, $f_{hlg}(z_f, \gamma b\tau) - f_{hlg}(z_b, \gamma b\tau)$, does not become too large.

4. Design of Quadratic and Scanning Fields

4.1. *Quadratic field gradient*

The static field B_0 is defined along the z axis. Consider a rectangular prism coil located in B_0 such that the center axis, the long axis, of the coil is parallel to B_0 as shown in Fig. 3(a). A quadratic field gradient can be generated when the current supplied to the prism coil, flowing parallel in Fig. 3(a), is made to travel in the reverse direction.[4] Such currents can be produced by combining four sections of rectangular coils as shown in Fig. 3(a). Let $M_L I_L$ be the ampere turns supplied by the current, r be the distance from the z-axis line to each current, and μ_0 be the permeability of free space. The field components in the x- and y-direction, B_{x1} and B_{y1}, are produced as shown in Fig. 3(a)

$$B_{x1} \approx -\frac{2\mu_0 M_L I_L}{\pi {r_0}^2} x, \tag{31}$$

$$B_{y1} \approx \frac{2\mu_0 M_L I_L}{\pi {r_0}^2} y. \tag{32}$$

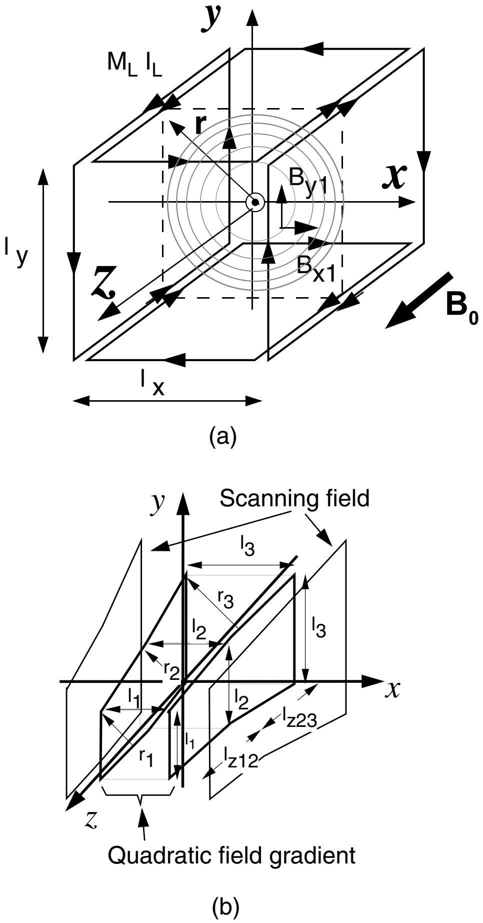


Fig. 3. Coil configuration for generating the quadratic field gradient; (a) Rectangular prism coil is located in static field B_0 such that the center axis of the coil is parallel to the direction of B_0. $M_L I_L$ is the ampere turns supplied by the current, and r is the distance from the z-axis to each current. When a current is supplied in the direction shown, the coil generates a field having components B_{x1} and B_{y1} in the x- and y-directions, respectively. The field intensity deviation of the field from the static field intensity B_0 is a quadratic function and the contour map is a cylindrical shape extending in the z-direction. (b) Configuration of coil generating a quadratic field gradient and scanning field. The coefficient of the quadratic field is varied in the z-direction by a factor α. Movement of the center position of the quadratic field is performed by supplying a uniform field by the scanning coil pair.

The resultant field by the sum of B_0 is rewritten as Eq. (33):

$$\begin{aligned} B &= \sqrt{\left(B_0^2 + {B_{x1}}^2 + {B_{y1}}^2\right)} \\ &= B_0\left[1 + \frac{{B_{x1}}^2 + {B_{y1}}^2}{2B_0^2} + \cdots\right] \\ &\approx B_0 + b(x^2 + y^2), \end{aligned} \tag{33}$$

where the coefficient b is defined as follows:

$$b = \frac{2{\mu_0}^2(M_L I_L)^2}{\pi^2 B_0 r^4}. \tag{34}$$

In order to vary the coefficient b of the quadratic field in the z-direction by coefficient α, the radius r must to satisfy the following equation, where r_0 is the radius of r at $z = 0$.

$$\frac{1}{\left(\frac{r}{r_0}\right)^4} = (1 + \alpha z). \tag{35}$$

Equation (36) is obtained by rearranging the z and r relation. Based on this equation, the orbit of the current must be in inverse proportion to the biquadratic power of coil section radius of r:

$$z = \frac{1}{\alpha}\left[\left(\frac{r_0}{r}\right)^4 - 1\right]. \tag{36}$$

In this case, the coefficient of the quadratic field in Eq. (34) in the z-direction is varied according to the following equation:

$$\begin{aligned} b(z) &= \frac{2{\mu_0}^2(M_L I_L)^2}{\pi^2 B_0 r^4 {r_0}^4 \{1/(1+\alpha z)\}} \\ &= (1 + \alpha z)b. \end{aligned} \tag{37}$$

Passing an electrical current through a line described by Eq. (36) is difficult. Therefore, production of the field given by Eq. (37) is realized by approximating the current through a curved line by a combination of currents on the straight line. This was designed based on the use of a current inversing method,[17] which revolves the polarity of field pulse for phase encoding, shown in the pulse sequence of Fig. 1. This method can produce an ideal quadratic field by removing unnecessary field components in the z-direction due to linear currents, passing in a direction perpendicular to the static field. As α becomes large, the image out of the focal plane is significantly blurred and the judgement as to whether the region of interest is on the focal plane is simplified. However, designing such coils is a difficult process. Considering blurring effects around the focal plane as well as the ease of design, the coefficient α is set to be 0.033/cm. After calculation based on infinite straight wire current theory, a satisfactory field coefficient can be obtained using two sections of straight coils. Figure 3(b) shows the coil configuration. Figure 4 shows the results of measuring and calculating the quadratic gradient when design parameters of the generating coil are as follows: $l_{z12} = l_{z23} = 25\,\text{cm}$, $l_2 = 8.8$, $l_1 : l_2 : l_3 = 1 : 1.25 : 1.6$.

4.2. *Scanning the quadratic field gradient*

The position of the center of the quadratic field described in Sec. 4.1 can be moved by superimposing an additional field component in the x- or y-directions. Here, movement in the x-direction is taken into consideration. In scanning the quadratic

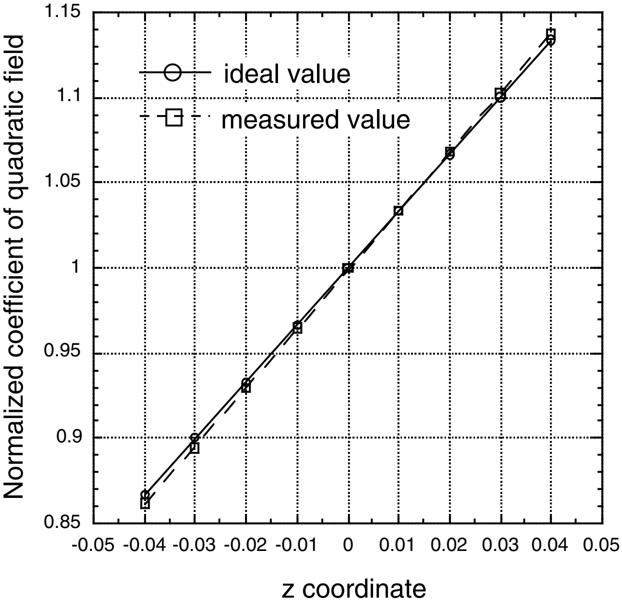


Fig. 4. Variation of normalized coefficient of the quadratic field gradient in the z-direction.

field of Eq. (37), the field component in the z-direction is not used, which is commonly used in Fourier transform imaging, but the field component in the x- or y-direction is used. In the case in which the field component in the x-direction, B_{x0}, is superimposed on B_{x1}, the total x component, B_{xt}, is given by the following equation:

$$\begin{aligned} B_{xt} &= B_{x0} + B_{x1} \approx B_{x0} - \frac{2\mu_0 M_L I_L}{\pi {r_0}^2} x \\ &\approx \frac{2\mu_0 M_L I_L}{\pi {r_0}^2} (x' - x) . \end{aligned} \tag{38}$$

The position x' at which B_{xt} becomes zero is given by;

$$x' = \frac{\pi {r_0}^2}{2\mu_0 M_L I_L} B_{x0} . \tag{39}$$

Equations (33) and (38) show that the center position of the quadratic field is moved in the x-direction by x'. Since the coefficient of a quadratic field, b, is proportional to the square of B_{xt}, by Eqs. (31) and (34), the strength of the scanning field should be varied as a function of $\sqrt{1+\alpha z}$ in the z-direction, being uniform in the x–y plane.

4.3. *Sweeping field gradient*

A sweeping field is the field that is uniform in the x- and y-directions but can be transformed linearly in the z-direction. When Eq. (10) is arranged, Eq. (41) can be represented by the combination of a uniform field and a gradient field of $G_{\alpha z}$ in the z direction.

$$\Delta b_{0y}(z) = (1 + \alpha z)\, b_{0y}, \tag{40}$$

$$= b_{0y} + G_{\alpha z} z, \tag{41}$$

where $G_{\alpha z}$ is set to be αb_{0y}. A uniform field, b_{0y} is generated by Helmholtz coil, while a gradient field $G_{\alpha z}$ is produced by circular coil pair called Maxwell Pair.[19]

4.4. *Linear field gradient*

Arranging Eq. (11) to Eq. (43), the field gradient given by Eq. (11) can be produced by the G_y gradient and the $G_y z$ field gradient.

$$\Delta G_{yz}(y, z) = (1 + \alpha z)\, G_y y, \tag{42}$$

$$= G_y y + G_{yz} yz. \tag{43}$$

Field gradient coil, which forms a field gradient by straight current passing in vertical direction to the B_0 field,[20] was used for generating G_y gradient. And, G_{yz} was designed using the Anderson's equation designed for shimming the main static field.[21] Figure 5 shows photographs of the coil system designed.

5. Experiment

5.1. *Experimental conditions*

The MRI system used in the experiments generates a static magnetic field of $B_0 = 0.0183\,\mathrm{T}$ by the solenoid coil (the resonant frequency is 779 kHz). Since the strength of the main static field of the MRI system is low and the SNR of the NMR signal is small, we improved the SNR of images by averaging the signal over several dozen observations. For that reason, we have not performed flow

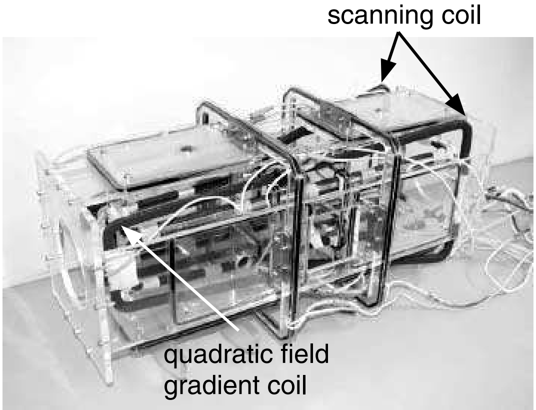


Fig. 5. Coil system for MR angiography diffractive imaging.

imaging experiments; only static imaging experiments to verify the feasibility of the proposed imaging technique were performed. The parameters of the experiments were as follows: $\gamma b\tau = 1.49\,\mathrm{rad/cm^2}$, the repetition time for the pulse sequence, $\mathrm{TR} = 300\,\mathrm{msec}$, scan step of the quadratic field, $\Delta x' = \Delta y' = 0.2\,\mathrm{cm}$, and the data matrix of the NMR signal was a 64×64 matrix. The RF flip angle of the pre and excitation pulses in Fig. 1 were 90°, and the slice selection pulse, G_{sel}, and the spoiling gradient pulse, G_{sp}, were not utilized. The experimental arrangement used for online reconstruction of 3D angiographic images is illustrated in Fig. 6. NMR signal detected by a receiver coil was amplified and converted to complex data by AMP and PSD, respectively. The signal was then converted to digital data by ADC and stored in the frame memory of the computer. The NMR hologram data were computed from the NMR signal data, and transferred to the Liquid Crystal-Spatial Light Modulator (LC-SLM) driver. The LC-SLM was a liquid-crystal panel for the red color display of a video projector XP700X (Fuji Xerox). There were 1024×768 pixels in a display area of $26 \times 21\,\mathrm{mm}$ (pixel size $25\,\mu\mathrm{m} \times 24\,\mu\mathrm{m}$). The LC-SLM was an active matrix type and had a polycrystalline Si TFT circuit in each pixel. Coherent plane wave laser light illuminated the hologram on the LC-SLM

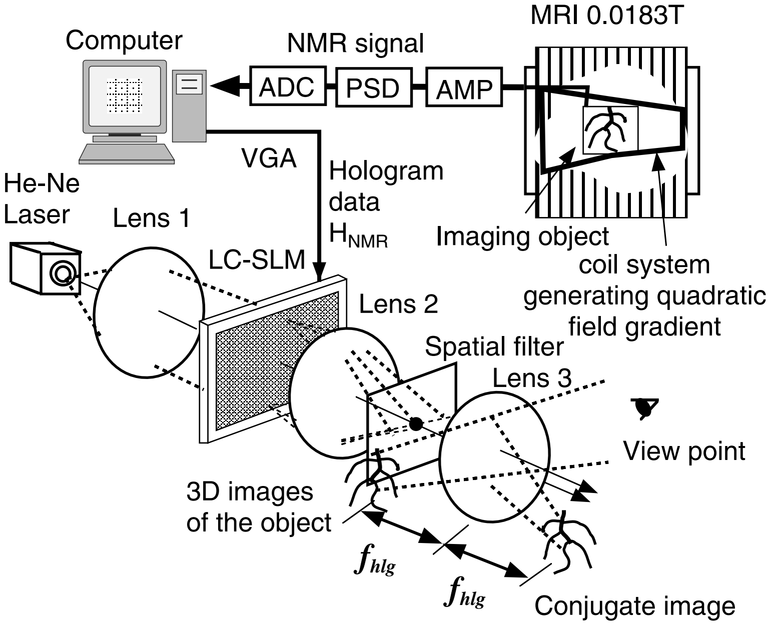

MRI 0.0183T
Computer
NMR signal
ADC
PSD
AMP
VGA
Hologram data H_{NMR}
He-Ne Laser
Lens 1
LC-SLM
Imaging object
coil system generating quadratic field gradient
Lens 2
Spatial filter
Lens 3
View point
3D images of the object
f_{hlg}
f_{hlg}
Conjugate image

Fig. 6. Schematic of the experimental arrangement for online 3D reconstruction of MR angiographic images. An NMR signal detected by a receiver coil is amplified and converted to complex data by AMP and PSD, respectively. The signal is then converted to digital data by ADC and stored in the frame memory of the computer. The NMR hologram data is computed from the NMR signal data, and transferred to the LC-SLM driver. Coherent plane wave laser light is used to illuminate the hologram on the LC-SLM and the amplitude of the input light is modulated by the hologram. The 3D images are viewed through the hologram.

and the amplitude of the input light was modulated by the hologram. A linearly polarized He-Ne laser (4.5 mW, 632.8 nm) was used as the light source. The zeroth-order diffracted light wave does not contribute to image formation and the light may be superimposed on the reproduced wavefront light of the object, so it was removed using an optical filter in the spatial frequency domain as shown in Fig. 6. 3D images of the angiographic image were viewed through the hologram.

5.2. *Results*

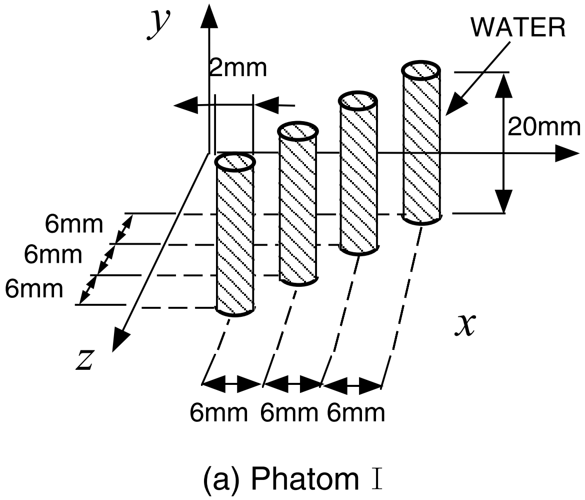
5.2.1. *Reconstructing images focused on an arbirary plane*

Preliminary experiments were performed in order to determine if images focused on an arbirary plane could be obtained from the signal scanned in two-dimensions. To improve the SNR of the NMR signal, 64-times averaging of the NMR signal was performed. Figure 7(a) shows the phantom (phantom I) used in the experiment, where water poles were placed at equal intervals in the x- and z-directions. Images were reconstructed numerically based on the reconstruction equation of Eq. (17). Figures 8(a)–8(d) show the numerically reconstructed images focused on the plane from $z' = 9\,\text{mm}$ to $z' = -9\,\text{mm}$ with a 6 mm spacing. The reconstructed images show that the water pole on a focal plane was imaged clearly. Conversely, the definition of the water pole out of the focal plane was reduced in proportion to the distance from the pole to the focal plane. These features provide an indication of the depth of the imaging object. The results of the preliminary experiments indicate that we can obtain images focused on an arbitrary plane by a signal scanned in two-dimensions.

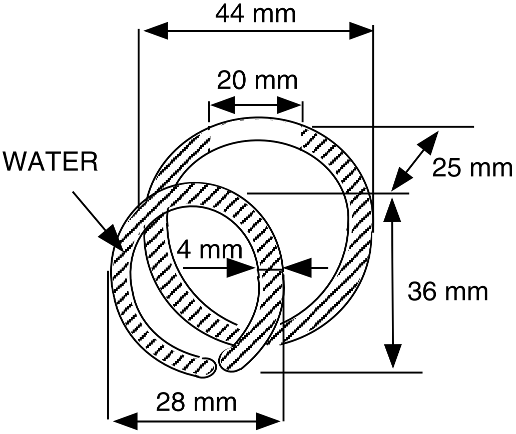
5.2.2. *Imaging experiments with 3D tube*

Image reconstruction experiments using an enhanced model were performed. The phantom II with a 4 mm diameter tube including water shown in Fig. 7(b) was used in the experiments. Numerical and holographic reconstructions were performed. Numerically reconstructed images obtained by adjusting the focal plane z' from $z' = 12\,\text{mm}$ to $z' = -12\,\text{mm}$ with an 8 mm spacing are shown in Figs. 9(a)–9(d). Holographically reconstructed images focused on the same planes are shown in Figs. 9(a-o), (b-o), (c-o) and (d-o), respectively. Figure 9(e) shows the hologram obtained from NMR signal data applying a SFH after 6-times enlargement (384 pixels). Amplitude modulation in Eq. (23) and superimposition of an appropriate constant K' in Eq. (25) were performed by numerical calculation using a computer. Without SFH, the enlarged hologram had a focal length of 193 cm under the following conditions: $\gamma b\tau = 1.49\,\text{rad/cm}^2$, $x_{s\,\max} = 12.8\,\text{cm}$, and $L_h = 9.8(1.63 \times 6)\,\text{mm}$, With the SFH using $s = 1.6\,\text{rad/cm}^2$, the focal length of the hologram was reduced to 13.4 cm, thereby enabling close observation of the 3D images.

In the numerical reconstructed images, we can obtain a 3D-distribution of the propagation wavefront by calculating images at small intervals in the z-direction.



(a) Phatom I



(b) Phatom II

Fig. 7. Phantoms used in the experiments: (a) phantom I, (b) phantom II.

Figure 10 shows the 3D-distribution of the propagation wavefront of phantom II. Image A shows the projectional view of the image, and B and C show the cross-sectional views of the wavefront in the $x = x'$ and $y = y'$ planes, respectively. Arrows p_a, p_b and p_c denote the focal points of regions a, b and c, where the diffracted wavefront becomes the narrowest, having the highest intensity in the z-direction. This wavefront information provides the spatial distribution of the object in the z-direction.

There are some regions in which the amplitude of the focal plane images were emphasized by the superimposition of blurred image components outside the focal

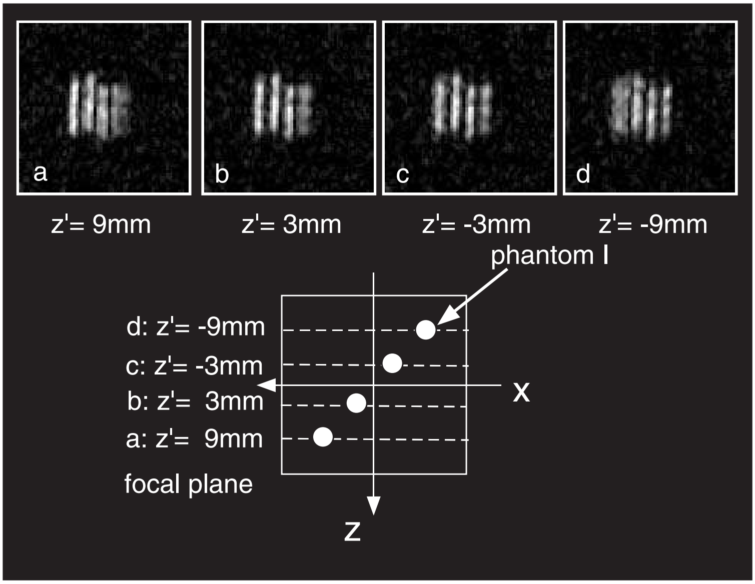


Fig. 8. Four reconstructed images of phantom I, each focused from (a) to (d) with 6-mm spacing. The water pole on the focal plane is clearly imaged.

plane. Such regions have a higher intensity and are suspected to contain diseased vessels. In this case, it may difficult to distinguish between normal and diseased regions. In order to reduce these interference problems, another experiment in which the view point was varied was performed. Figures 11(a)–11(d) show the results of using the phantom II in different locations to the case of Figs. 9(a-o), (b-o), (c-o) and (d-o) show the 3D images focused on the same planes, by adjusting the focal length of the CCD camera. Figure 8(e) shows the hologram produced by the NMR signal. Since the focal length of the hologram was shortened to 13.4 cm, the amplitude of the hologram roughly represented the distribution of the phantom structure. The manner in which the interference is changed and the absence of higher intensity regions in the reconstructed images enable us to conclude that the suspected region is not diseased.

6. Discussion

In general, MR flow imaging is based on either the Time-Of-Flight (TOF) or the Phase Contrast (PC) methods. The TOF method is based on the enhancement of signal intensity by the spins entering the slab between successive excitations. Since the RF excited region in the proposed imaging technique is similar to that in the 3D TOF method, we can conclude that TOF method is applicable to MR diffractive imaging. PC sequences contain incremental bipolar gradients on x-, y-, or all 3-axes to impart phase shifts in proportion to flow speed along each axis.

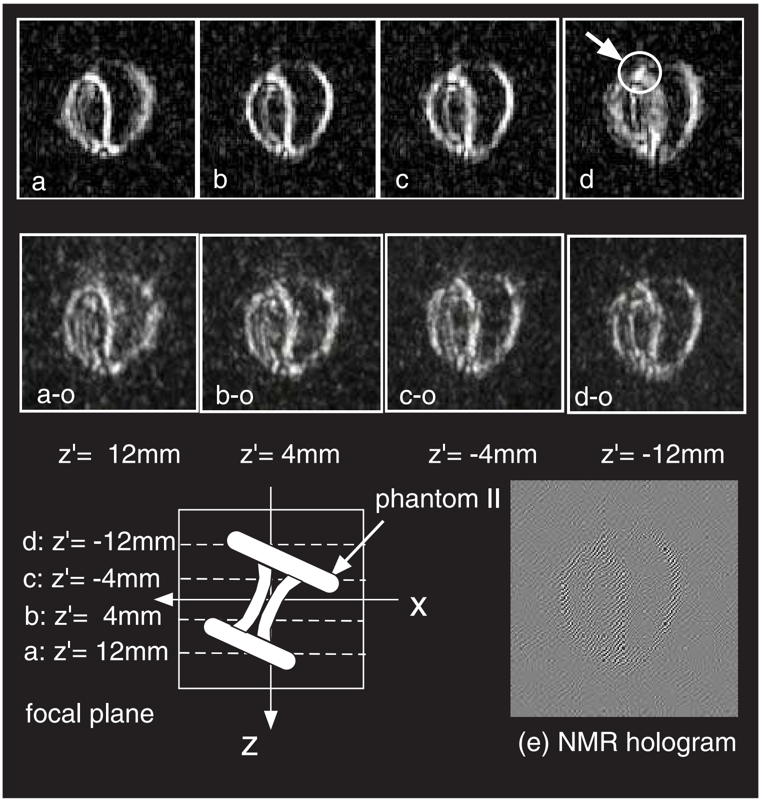


Fig. 9. Reconstructed images of phantom II by numerical and holographic methods: (a)–(d) show the numerically reconstructed images focused on planes from $z' = 12\,\text{mm}$ to $z' = -12\,\text{mm}$ at a spacing of 8 mm. (a-o)–(d-o) show the 3D images obtained using the holographic method, each focused on the same planes as (a)–(d) by adjusting the focal length of the CCD camera. (e) shows the hologram produced by the NMR signal using the proposed imaging technique. The focal length of the hologram is reduced to 13.4 cm.

Typically, a reference scan is used to remove effects from the phase shifts that are unrelated to flow. The PC method is also applicable to MR diffractive imaging in principle, however, it must be noted that the phase of the reconstructed image is shifted not only by spin movement but also by the deblurring effects of the out-of-focus plane components. In MR diffractive imaging, reconstructed images are considered as the sum of the sliced image components in the z-direction, which consist of an in-focus and out-of-focus plane components. The phase of out-of-focus images is shifted by the deblurring effect as shown in the diffractive equation. Therefore, when calculating the phase shifts relating to the spin movement, the reference image must be focused on the same plane as the flow-encoded image for the depth-direction such that phase shifts due to the out-of-focus plane components are removed.

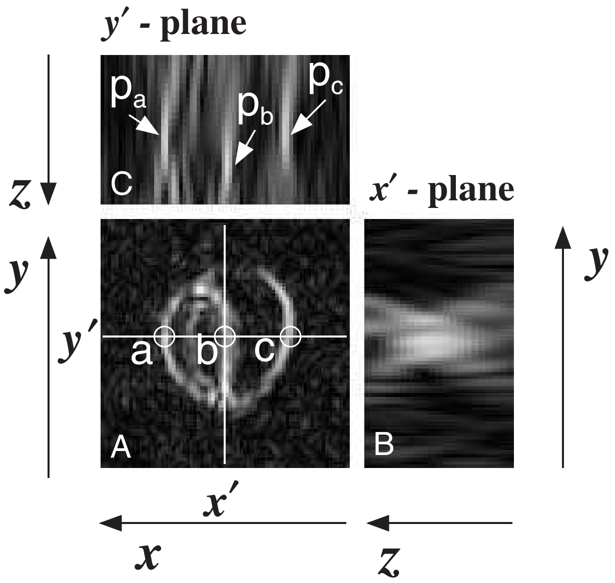


Fig. 10. Three-dimensional distribution of the propagation wavefront calculated using the numerical method. Image A shows the projectional view of the image, and B and C show the cross-sectional views of the wavefront on the $x = x'$ and $y = y'$ planes, respectively. The arrows show the focal point of the region a, b, c on the image A, where the diffracted wavefront is minimized and exhibits the highest intensity in the z-direction.

Image reconstruction from the diffracted wavefront can be performed either optically on an optical bench or digitally using a computer. Both methods have advantages and disadvantages. The former can be used to reconstruct 3D images that can be viewed without specialist equipment. However, the image SNR is inferior to the computed image due to speckle noise arising during reconstruction. In contrast, in the latter method, images free from speckle noise can be obtained and it is easier to identify the 3D-distribution of the object by varying the focal plane of the images. Moreover, improvement of the image SNR or enhancement of the infocal image may be performed using image processing techniques. At this time, we advocate supplementing the reading of the computed images with NMR holograms.

A 3D holographic display system for medical imaging has already been developed and feasibility studies have been performed.[18] It was found that holographically reconstructed 3D images detected subtle anatomical features that were hidden by overlapping structures in radiographs and 3D CT images. However, the system requires several dozen minutes and significant computation capability to produce each 3D-hologram. Conversely, in MR diffractive imaging, the hologram data is obtained from the NMR signal, since the NMR signal contains a description of the wavefront recorded on the hologram. In principle, real-time holographic reconstruction of 3D images is possible when NMR signals with a high SNR are transformed into holograms as soon as they are acquired.

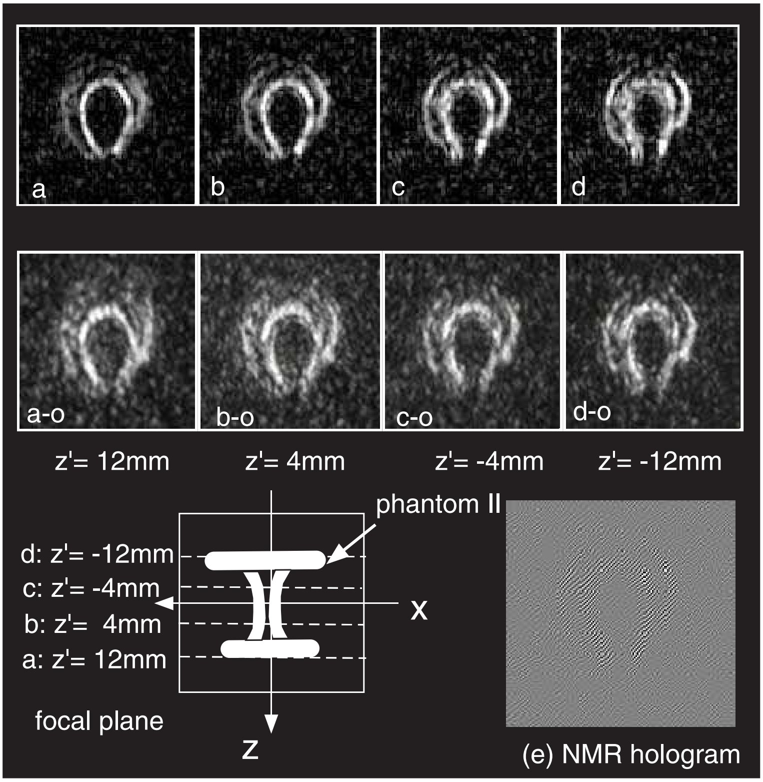


Fig. 11. Reconstructed images of phantom II viewed from different angles: (a)–(d) show the numerically reconstructed images focused on planes from $z' = 12\,\text{mm}$ to $z' = -12\,\text{mm}$ at a spacing of 8 mm. (a-o)–(d-o) show the 3D images obtained using the holographic method, each focused on the same planes as (a)–(d), by adjusting the focal length of the CCD camera. (e) shows the hologram produced by the NMR signal using the proposed imaging technique. Since the interference differs to that in Fig. 9, these images provide supplemental information to aid in the diagnosis.

As the LC-SLM used in the experiment had a 1.3 inch-LCD panel designed for video projection, the size of the reconstructed image was limited to be less than the size of the LC panel. If the LC panel is replaced by a large LCD panel such as those used for laptop computers, larger-sized images would be observed. The clinical significance of the images are significantly improved if they are viewed in real-time.

7. Conclusion

Magnetic resonance diffractive imaging is proposed and demonstrated as a new approach to MR angiography. The technique has two distinct advantages over existing techniques; the possibility of fast angiographic imaging and the online reconstruction of 3D volumetric images in the holographic technique. To verify the feasibility of this imaging method, static imaging experiments using an ultra-low-field MRI were performed. Images were reconstructed using the numerical and holographic methods. In the numerical method, images focused on an arbitrary plane in the depth direction could be reconstructed, which is helpful in the identification of the 3D-distribution of the object. In the holographic method, 3D images of the phantom were viewed through a hologram produced by an NMR signal. In addition, it was shown that acquiring images from different angles is useful in order to obtain supplemental information for recognizing the spatial distribution of the object. The experimental results show that the proposed imaging technique offers the possibility of fast angiographic imaging and enables the online reconstruction of 3D images of vessels. A 0.2 T MRI system is under development for performing flow imaging with high-SNR signals.

Acknowledgments

This research was partly supported by a Grant-in-Aid for Encouragement of Young Scientists from the Ministry of Education, Science, Sports and Culture, Japan and Nakatani Electronic Measuring Technology Association of Japan.

References

1. Y. Yamada, K. Tanaka and Z. Abe, NMR reconstruction imaging using a scannable nonlinear field gradient, in *Proc. of 14th Int. Conf. on MBE*, 17, 25 (1985) 834–835.
2. Y. Yamada, K. Tanaka and Z. Abe, NMR Fresnel transform imaging technique using a quadratic nonlinear field gradient, *Rev. Sci. Instrum.* **63** (1992) 5348–5358.
3. S. Ito, O. Sato, Y. Yamada and Y. Kamimura, Online holographic reconstruction of NMR images by means of a liquid crystal spatial light modulator, in *Proc. IEEE Int. Conf. on Image Processing* 96, **III**, Lausanne, Switzerland, pp. 531–534, 1996.
4. S. Ito, Y. Yamada and Y. Kamimura, Real-time holographic reconstruction of NMR images in Fresnel transform imaging technique, in *Proc. of IEEE Int. Conf. on Engineering in Medicine and Biology Society*, 2.2.1-e, Chicago IL, USA, pp. 467–469, 1997.
5. S. Ito, Y. Yamada and Y. Kamimura, Real-time holographic reconstruction of NMR images using a liquid crystal-spatial light modulator, *Systems and Computers in Japan* **31** (2000) 70–80.
6. S. Ito, Y. Kamimura and Y. Yamada, Holographic running reconstruction of MR images in phase scrambled Fourier imaging technique, in *Proc. 7th International Society for Magnetic Resonance in Medicine*, **1** Philadelphia, USA, p. 97, 1999.
7. E. Wolf, Three-dimensional structure determination of semi-transparent objects from holographic data, *Optic Comm.* **1** (1969) 153–156.
8. S. Kawata, O. Nakamura and T. Noda, Laser computed-tomography microscope, *Appl. Opt.* **29** (1990) 3805–3809.
9. T. Noda, S. Kawata and S. Minami, Three-dimensional phase-contrast imaging by a computed-tomography microscope, *Appl. Opt.* **31** (1992) 670–674.
10. D. A. Agard, Y. Hiraoka and P. Shaw, Fluorescence microscopy in three dimensions, *Methods in Cell Biology*, **30** (New York: Academic Press, 1989), pp. 353–377.
11. C. K. Avinash and M. Slaney, Principles of computerized tomographic imaging (New York: IEEE Press, 1988).

12. R. K. Mueller, M. Kaveh and G. Wade, A new approach to acoustic tomography and applications to ultrasonics, in *Proc. IEEE* **67** (1979) 567–587.
13. A. J. Devancy, A filtered backpropagational algorithm for diffraction tomography, *Ultrason, Imaging* **4** (1982) 336–350.
14. G. W. Stroke, An introduction to coherent optics and holography (New York: Academic Press, 1969).
15. R. Turner, Optical reconstruction of NMR images, *J. Phys. E. Sci. Instrum.* **18** (1985) 875–878.
16. D. Gabor, A new microscopic principle, *Nature* **161** (1948) 777–778.
17. N. Kanzaki, S. Tajima, S. Ito, Y. Kamimura and Y. Yamada, Reconstruction image due to defectiveness of characteristic field in NMR Fresnel transform imaging, *IEICE Transactions on Information and Systems*, **J81-D-II** (1998) 2867–2874.
18. D. D. Robertson, C. J. Sutherland and B. W. Chan, Depiction of pelvic fractures using 3D volumetric holography: Comparison of plain X-ray and CT, *J. Comp. Assist. Tomogr.* **19** (1995) 967–974.
19. J. Rodney and M. Vaugham, Representation of axisymmetric magnetic fields in computer programs, *IEEE Trans. Ed.* **ED-19**, 2 (1972) 44.
20. P. Mansfield and P. G. Morris, NMR imaging in Biomedicine, in *Advance in Magnetic Resonance Suppl.* 2, (New York: Academic Press, 1982).
21. W. A. Anderson, Electric current shims for correcting magnetic fields, *Rev. Sci. Instrum.* **32** (1961) 3.

This page intentionally left blank

CHAPTER 3

TECHNIQUES IN 3D ASSESSMENT OF TRACHEAL-STENOSIS BY THE MEAN OF SPIRAL COMPUTED TOMOGRAPHY (S-CT) AND THEIR APPLICATIONS

ERICH SORANTIN, DARIUS MOHADJER and FRANZ LINDBICHLER
Section of Digital Information and Image Processing
Department of Radiology, Univ. Hospital Graz
Auenbruggerplatz 34, A-8036 Graz, Austria

LASZLO G. NYUL and KALMAN PALAGYI
Department of Applied Informatics
Josefz Attila University Szeged, Arpad ter 2
H-6720 Szeged, Hungary

BERNHARD GEIGER
Siemens Corporate Research Princeton Inc.
755 College Road East Princeton
NJ 8540, USA

Endotracheal intubation is the most common cause of Laryngo-Tracheal Stenoses (LTS), followed by trauma and prior airway surgery.[1–3] In rare cases LTS may have resulted also from inhalation injuries, gastro-esophageal reflux disease, neoplasia and autoimmune diseases like Wegeners granulomatosis or relapsing polychondritis.[1,4] In pediatric patients vascular compression of the trachea is a common cause of tracheal indentations.[5] Clinical management of these conditions requires information on localization, grade, length and dynamics of the stenosis. Exact LTS information is necessary, since stenoses with a length less than 1.0 cm can be treated by an endoscopic surgery.[6,7] Besides Fiberoptic Endoscopy (FE), which represents the gold standard for airway evaluation, imaging modalities like conventional radiography, fluoroscopy, tracheal tomograms, Magnetic Resonance Imaging (MRI) and above all Spiral Computed Tomography (S-CT) are an essential part of the clinical work.[1,8] S-CT and the recent introduction of multislice imaging allows volumetric data acquisition of the Laryngo–Tracheal Tract (LTT) during a short time span. Decreased motion artifacts and increased spatial resolution form the basis for high quality post processing.[9,10] The improved performance of today's workstations permits the use of sophisticated post processing algorithms even on standard hardware like personal computers. Thus real time 3D display and virtual endoscopic views (virtual endoscopy) are just one mouse click away. Other algorithms compute the medial axis of tubular structures like airways or vessels in 3D, which can be used for the calculation of 3D cross sectional profiles for better demonstration of caliber changes.[11] Thus display of S-CT axial source images is moving rapidly to 3D display. Moreover, established network connections within and between institutions allows telemedical cooperation. Web technologies offer an easy to use way for information exchange. The objective of this paper is to present an overview on 3D display and quantification of LTS as well as to provide information how these results can be presented and shared with the referring physicians on the hospitals computer network. This article is structured in seven parts; namely: S-CT data acquisition for LTS imaging; selected 3D image post processing

algorithms; 3D display; Virtual endoscopy; Objective LTS degree and length estimation using LTT 3D — cross-sectional profiles; Intranet applications; and a conclusion is drawn in the final section.

Keywords: Endotracheal intubation; LTS; Spiral Computed Tomography.

1. S-CT Data Acquisition for Imaging of LTS

1.1. *Positioning*

All CT studies should be performed in helical mode. The patient is scanned supine and special care has to taken for head positioning in order to faciliate comparable results for follow-up studies. In our institution we prefer a head position, where the vertical beam of the CT machine laser positioning tool targets the lateral orbital angle and the tragus cartilage. This procedure puts the head in a neutral position and is easy to reproduce.

1.2. *Scan range*

Helical scanning is performed from the caudal mastoid border to the tracheal bifurcation.

1.3. *Scan parameters*

Modern multislice scanners now allow a beam collimation of about 0.75 mm to 1.25 mm. In order to keep the radiation dose as low as possible, the highest pitch levels, as recommended by the manufacturer, are preferable and the lowest achievable tube current should be used. At reconstruction an overlap of 50% to 70% should be selected. Thus a whole study ends up with about 200 to 300 axial source images.

1.4. *Intravenous contrast medium injection*

For LTT evaluation there is usually no need for intravenous (i.v.) contrast medium injection. Exceptions are oncologic staging investigations, especially laryngeal cancers, and the suspicion of vascular anomalies in children. Whenever i.v. contrast injection is necessary the usuage of an power injector is recommended whereby the flow rate should be set to 3 ml/s and the bolus tracking systems used.

2. Selected 3D Image Post Processing Algorithms

2.1. *Airway segmentation*

Segmentation can be defined as procedure to define the boundaries of the "Organ of Desire" or more general speaking "Regions of Interest" (ROI's). The segmentations result in a binary mask image (volume), where the value "1" represents pixels belonging to the segmented ROI's. All other pixels are set to the value "0". An ideal algorithm should be as much as possible free from the operators in influence and fast. Several algorithms can be used — from simple manual tracing and region growing to more sophisticated ones like the "Fuzzy Connectedness" algorithm — the later ones will be presented in more detail below.

2.1.1. *Region growing*

Due to the air content of the Laryngo Tracheal Tract (LTT) the attenuation coefficients are well below −150 Hounsfield Units.[12] Axial S-CT slices are displayed at lung window settings (center/width −600/1200 Hounsfield Units). By starting from a seed point inside the trachea, neighboring voxels are added if their attenuation coefficients are below a specified threshold (usually less than −150 Hounsfield Units).[12] Due to the partial volume effects boundary pixels, respective voxels exhibit a variation of the attenuation coefficients. Thus organ boundaries are frequently not closed on axial source images. Therefore, by choosing a fixed threshold value the segmented contour is kept either to "close" function or "bleeding" occurs (meaning the extension of the segmented region to unwarranted dimensions). The chosen threshold can be adapted on slice-to-slice basis to the individual S-CT patient data.[13] Next, the middle of the segmented region is determined and projected to the next slice, where it serves as a new seed point. The entire process is repeated until the whole upper respiratory tract has been defined and each step is performed under the control of the operator. This approach inherits the disadvantage of being highly operator dependent.

2.1.2. *Fuzzy connectedness*

"Fuzzy Connectedness" captures the image inherent fuzziness as well as the spatial-coherence of the voxels in a well defined manner.[14,15] In case of LTT, air has a well defined range of Hounsfield units. Therefore the parameters, needed for the definition of the fuzzy affinities, can be set once and used for all studies without a per-study training. On one or more axial slices the operator selects by a mouse click a "seed point" within the LTT center for seeding the fuzzy connected objects. Since absolute "Fuzzy Connectedness" is used and a single object is segmented, the uncertain boundary regions (due to partial volume effects) are not captured by the fuzzy objects. Thus the resulting segmented LTT is uniformly smaller than the physicians expectations. Hence a 3D dilation using a $3 \times 3 \times 3$ structuring element is applied to the segmented fuzzy connected object. Finaly the operator controlls the results of segmentation on the computer screen at fixed window settings (center −600 Hounsfield units, width 1200 Hounsfield units), where the segmented LTT boundaries are outlined on original axial slices (Fig. 1). According to our own investigations the "Fuzzy Connectedness" algorithm was applied for LTS segmentation in 36 patients and 18 normal controls. On average 3.9 slices had to be edited in patients and 2.4 in the normal controls — this difference was found

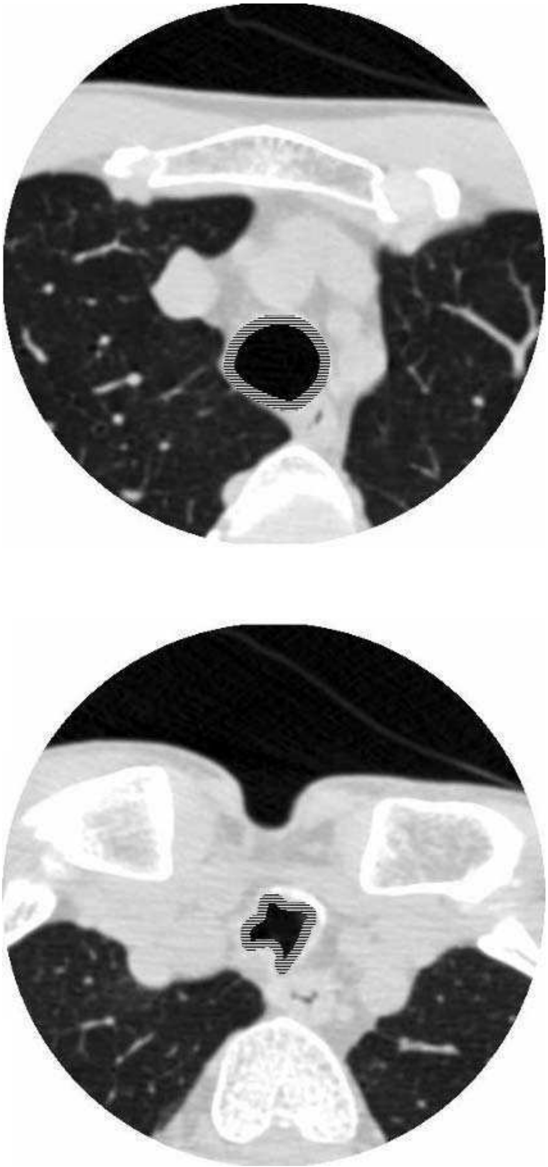


Fig. 1. S-CT, LTT segmentation based on fuzzy connectedness: The dashed line, superimposed on the original axial S-CT image outlines the segmentation results of a normal (upper part) and a pathologic tracheal part (lower part).

to be statistically insignificant ($p = 0.06$). In 90% of all cases only less than five slices had to be corrected manually. In another study the accuracy and precision of three different approaches for LTS segmentation were investigated, namely: Free hand tracing, manual tracing augmented by splines and the "Fuzzy Connectedness" algorithm.[16] Segmentation results revealed no statistically significant differences regarding accuracy and precision, but the “Fuzzy Connectedness” algorithm proved to be the fastest, just lasting 15–20 seconds for a complete run of LTT segmentation. Therefore the conclusion can be drawn, that the “Fuzzy Connectedness” algorithm represents a fast and robust tool for LTT segmentation.

2.2. *Computation of the LTT skeleton*

The notion of skeleton was introduced by Blum as a region-based shape feature/descriptor which summarizes the general form of objects/shapes.[17] A very illustrative definition of the skeleton is given using the prairie-fire analogy: The object boundary is set on fire and the skeleton is formed by the loci where the fire fronts meet and quench each other. This definition can be naturally extended to any dimension. The thinning process is a frequently used method for producing an approximation to the skeleton in a topology-preserving way.[18] It is based on digital simulation of the fire front propagation: Border points (i.e. 1’s that are “adjacent” to 0’s) of a binary object that satisfy certain topological and geometrical constraints are deleted in iteration steps. The entire process is repeated until only a reasonable approximation to the skeleton is left. In 3D, there are two major types of tinning: These algorithms produce either the medial surface of an object (by preserving surface endpoints) or can extract the medial lines of an object (by preserving line end-points).[19] Figure 2 illustrates the results of the two types of thinning approaches, whereas Fig. 2(c) demonstrates that extracting medial lines yields the medical relevant information of the medial axis for “tubular” objects like LTT and blood vessels. Most of the existing thinning algorithms are parallel, since the fire front propagation is by nature parallel, meaning that all border points satisfying the deletion condition of the actual phase of the process are simultaneously deleted. A recently published parallel 3D 6-subiteration directional thinning algorithm[20] is applied to extract the medial lines from the segmented LTT. Figure 3 exhibits an

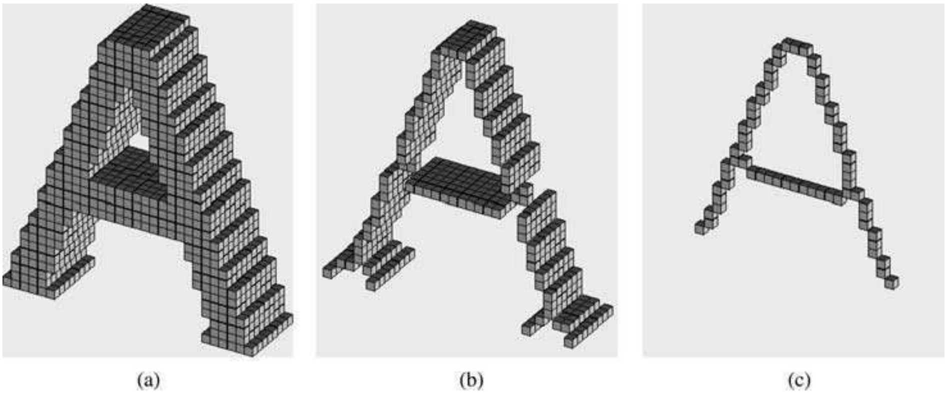


Fig. 2. Example of the two types of 3D thinning. Part (a) depicts the character “A” as a 3D synthetic object, (b) its medial surface, and (c) and its medial lines. The medial lines approximate the center line of the object, which is the medical relevant information.

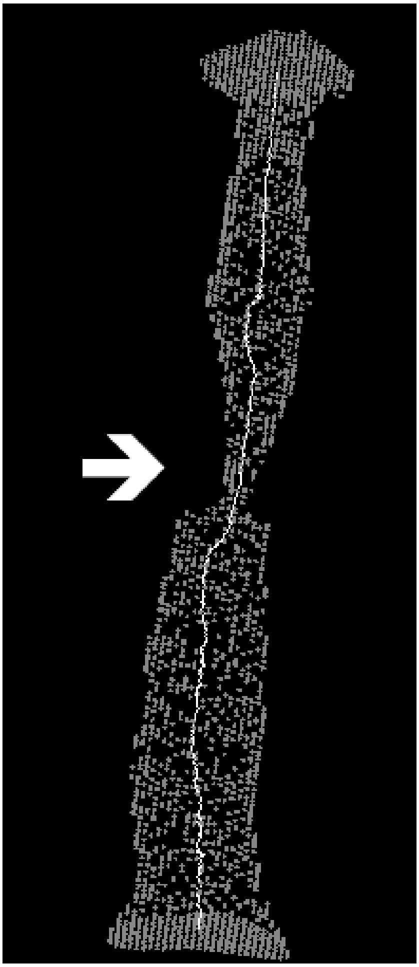


Fig. 3. 3D tracheal model from a patient suffering from a tracheal stenosis. Site of the stenosis is indicated by the white arrow. The tracheal medial axis is shown as bright line in the center.

example of the computed skeleton in a patient suffering from a tracheal stenosis. Since skeletonization is rather sensitive to "boundary noise" (i.e. the roughness of the boundary of the object), the skeleton can contain several unwanted/parasitic segments. To overcome this problem, pruning of the resulted skeleton (as a post-processing step) is performed. Since the extracted LTT medial axis exhibits local variations, thus making a following smoothing according Casteljau's algorithm is necessary.[21]

2.3. Calculation of the LTT 3D cross sectional profile along the LTT medial axis

Along the extracted LTT medial axis the orthogonal cross sectional area is computed using the interpolated data volume, which represents the segmented LTT.

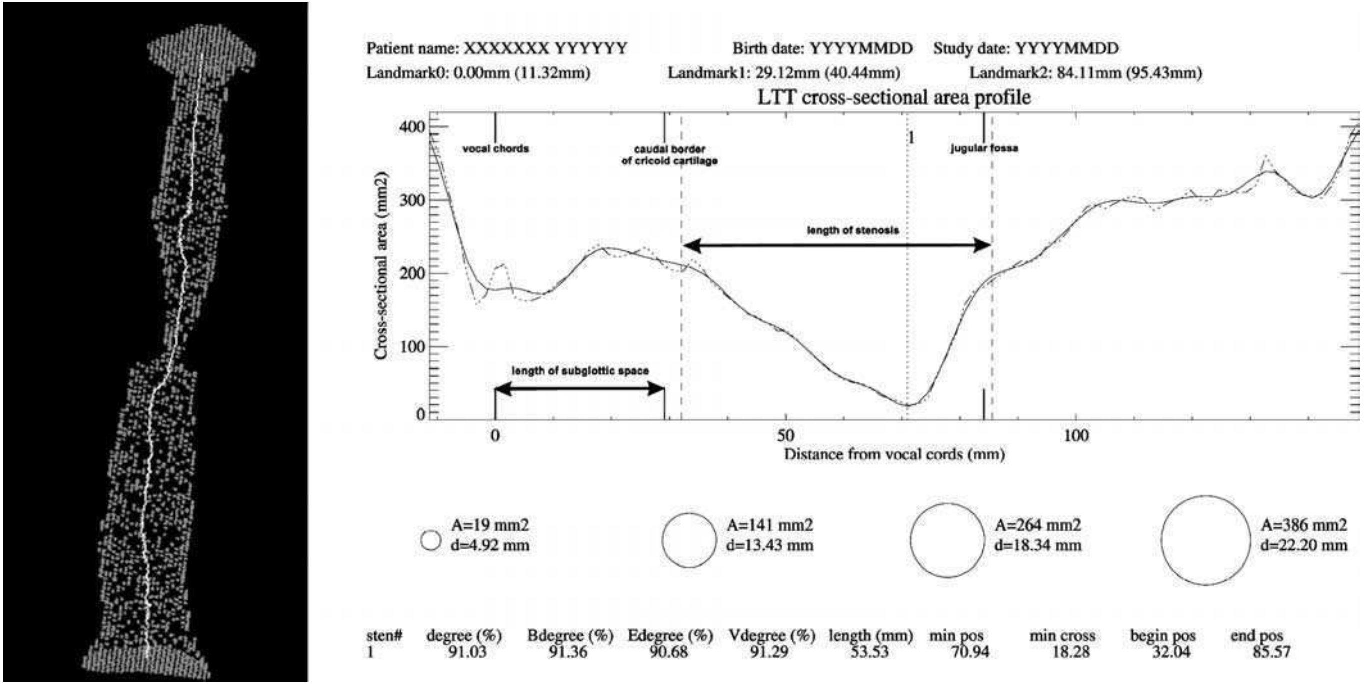
Thereafter the cross sectional areas are plotted against the positions of the LTT medial axis. Localization of vocal cords on the LTT medial axis is chosen as the zero position. Positions on the LTT medial axis lower to the vocal cords are marked positively and those above negatively. LTS regions can be detected as a decrease of the cross sectional area. In order to facilitate anatomic cross reference between endoscopy, imaging and LTT 3D cross sectional charts, the positions of the vocal cords, caudal border of cricoid (end of subglottic space) and jugular fossa are automatically marked as vertical bars on the line charts too. Additionally, for better illustration of the plotted cross sectional areas, four circles are drawn on the chart in real size, whose areas cover the range between the 5th and 95th percentile of the calculated cross-sectional areas. Figure 4 shows a 3D model of a trachea in a patient suffering from a tracheal stenosis as well as the corresponding cross sectional profile.

3. 3D Display

In medical literature it is well known that the mental reconstruction process of multiple, transaxial sections may fail in patients with tracheobronchial deformities.[22] On axial CT the shape of tubular structures like airways depends on the angle between the structure itself and the slice plane: If this angle is 90 degrees to the longitudinal axis of the tubular structure, the true cross section, e.g. a circle, is displayed. If this angle is oblique the displayed shape will change, e.g. it will get more elliptic. Therefore the true caliber and shape in 3D cannot be determined on axial slices alone. 3D reconstructions help to avoid this problem. Remy–Jardin published a paper regarding the comparison between reading axial slices alone and volume rendered transparent bronchographic images. It was found that the volume rendered bronchographic 3D reconstructions were superior to reading axial slices alone. 3D display allows one to demonstrate clinical colleagues CT findings in an easy and impressive way. It addition, they are useful for follow-up investigation, since the underlying pathoanatomy is displayed on just a few views. In medical image processing mainly “Multiplanar Reformation”, “Surface Rendering” and “Volume Rendering” are used — these techniques will be described in the following sections.

3.1. *MultiPlanar Reformation (MPR)*

For computing MPRs all source images are stacked in order to built up a volume. Care has to be taken in order to keep the correct aspect ratio — either interpolation is used or the volume is stretched according the ratio between the x-, y- and z-voxelsize. MPRs are well known for the assessment of spinal CT investigations. Similar, sagittal and coronal MPR enable additional LTS views. In the case of a buckled LTT it is not possible to display the LTT in one view, but this can be achieved using curved MPR’s. In addition, thick MPR’s (MPR with a thickness of more than one pixel) can be used for better overview (Fig. 5).

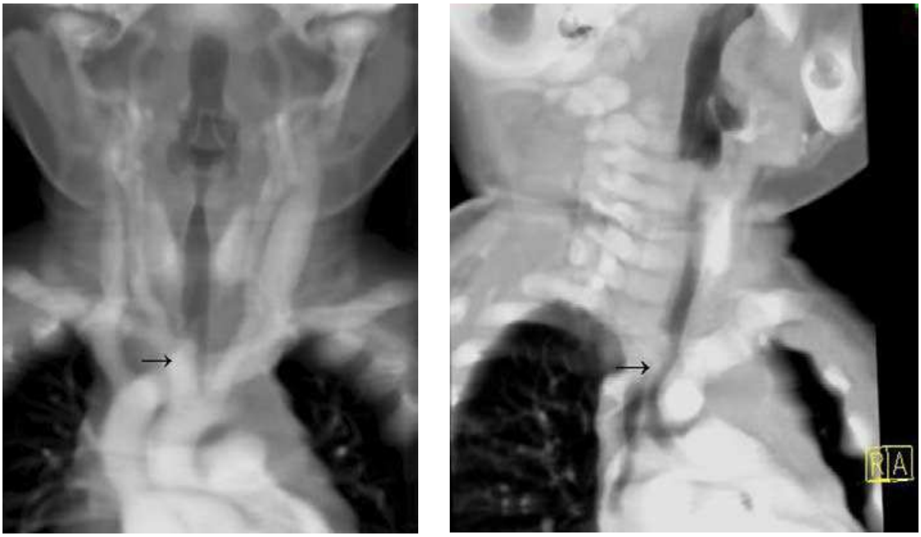


Patient name: XXXXXXX YYYYYY Birth date: YYYYMMDD Study date: YYYYMMDD

Landmark0: 0.00mm (11.32mm) Landmark1: 29.12mm (40.44mm) Landmark2: 84.11mm (95.43mm)

sten#	degree (%)	Bdegree (%)	Edegree (%)	Vdegree (%)	length (mm)	min pos	min cross	begin pos	end pos
1	91.03	91.36	90.68	91.29	53.53	70.94	18.28	32.04	85.57

Fig. 4. LTT 3D cross sectional chart from a patient suffering from tracheal stenosis (shown on left side). The short vertical bars represent the three chosen anatomic landmarks. Distal the caudal border of the cricoid cartilage a drop of the cross sectional area can be seen, representing the stenotic segment. For better anatomic cross reference three landmarks are displayed as short vertical bars — position of vocal cords (indicates the chosen zero point), caudal border of cricoid cartilage and jugular fossa. Distal the caudal border of the criocoid cartilage a drop of the cross sectional area can be seen, thus representing the stenotic segment. For better correlation between caliber changes on 3D — cross sectional chart and the "Real World" there are four circles drawn in real size below the chart, covering the range of the plotted cross sectional areas. In addition to the table with the quantification results shown the following abbreviations are used: Degree, represents the degree of caliber changes in percent; length, represents the length of the LTT caliber changes, calculated by the difference between the end pos and the begin pos (see below); min pos is the position of the minimal cross sectional area on the 3D cross sectional chart (equal to distance to vocal cords); min cross is the absolute value of the minimum cross sectional area at min pos; begin pos is the starting point of local airway caliber change — a deflection of the start of the 3D cross sectional profile; end pos is the end point of local airway caliber change — similar to begin pos.



(a) (b)

Fig. 5. S-CT, thick MPR reconstruction: Tracheal compression caused by the brachiocephalic artery (marked by black arrow). (a) shows an ap-view, and (b) a lateral view. On both views the tracheal compression can be percepted.

3.2. *Surface rendering (Synomina: Shaded surface display, iso-contour rendering)*

After segmentation the contours of the ROIs can be extracted easily. Using the morphological operation of erosion, one layer of the segmented surface voxels can be "peeled off".[23] The difference between the resulting data volume and the original volume are just the surface voxels. Therefore once the data volume has been subtracted from the original volume only the contours will remain. In order to obtain polyhedral surface models these contours are converted into 3D models using the Delaunay triangulation method.[24] Figure 6 demonstrates the individual steps for generating these models. These surface models can be rendered at interactive speed at a workstation. Properties such as color and transparency may be manipulated. In addition, real time rotation, panning and zooming are possible. By conversion of such 3D models according the standards of the Virtual Reality Modeling Language 2.0 (VRML) they can even be visualized on a Personal Computer using a standard web browser.[25] The main disadvantage of "Surface Rendering" is the necessity of segmentation. The procedure works well only in areas of high contrast — e.g. 3D — display of bones based on native CT scans, or air filled organs like the LTT. As already mentioned in Sec. 2.1, due to partial volume effects the boundaries of biological structures like the LTT are often not closed, thus leading to "bleeding". Consecutively, (sometimes exhausting) editing of the segmented boundaries are necessary. Moreover, only the prior segmented organ boundaries can

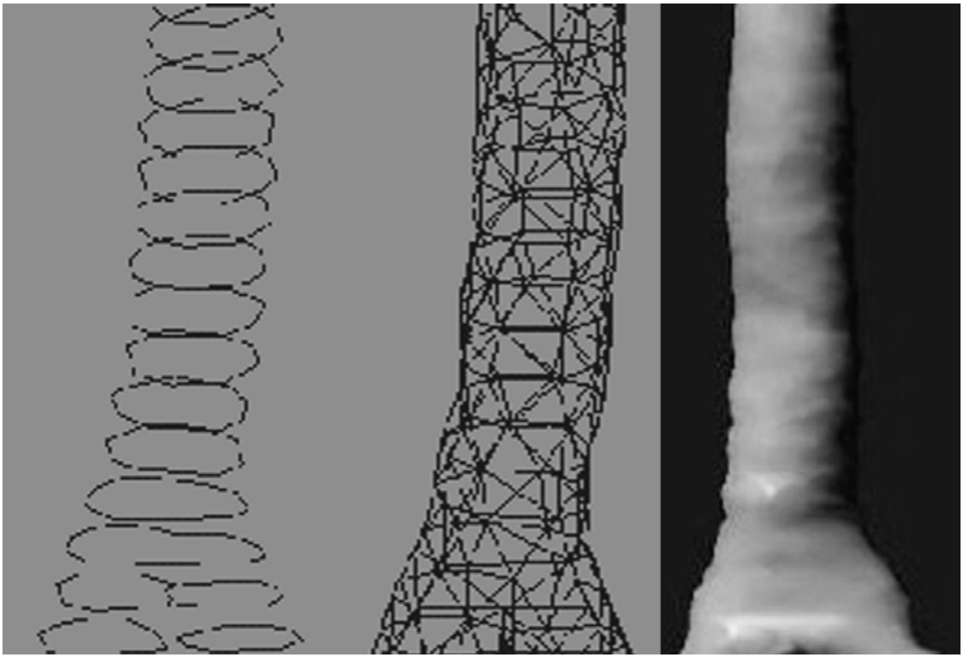


Fig. 6. The individual steps for creating a polyhedral surface model are demonstrated: (a) The LTT contours after segmentation, (b) result after triangulation and (c) the final result after artificial coloring.

be displayed in 3D. Thus for display just about 10% of all available source data are used. Therefore anatomical information is limited to the boundaries of the segmented organs. Due to its undemanding computational needs "Surface Rendering" was one of the earliest techniques for 3D display.

3.3. *Volume rendering (Synonym: Percentage rendering)*

A completely different approach is used by the volume rendering algorithm. No segmentation is needed using this 3D reconstruction method. Basically, an "opacity curve" is constructed for a given data volume. This curve assigns every grey value a particular opacity, ranging from 0% (= completely transparent) to 100% (= completely opaque). By manipulating the opacity curve different anatomical structures can be displayed. A virtual ray is sent through the data volume. All grey values along the ray are collected and their opacity is changed according the chosen shape of the opacity curve. Details of the algorithm have been published elsewhere.[9] In order to achieve high quality 3D reconstructions with multiorgan display the desired organ systems need good contrast to the surrounding anatomy. As mentioned above, whenever vessels are of interest, intravenous contrast administration by a power injector is mandatory. Volume rendering offers several challenges for the radiologist in presentation of S-CT. The most important step is the

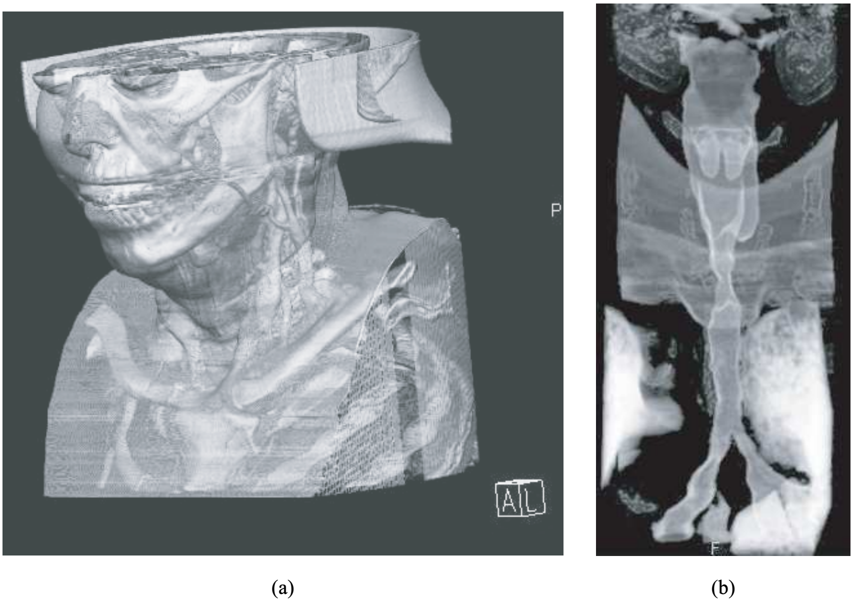


Fig. 7. S-CT after iv. contrast injection, multiorgan display: "Volume Rendered" 3D reconstruction of the neck and upper chest in a patient suffering from a tracheal carcinoma where a tracheal stent was also implanted: (a) View from outside, (b) result after changing the opacity curve to a bronchographic appearance of the LTT. The implanted stent as well as the laryngeal cartilage are depicted by the dark areas.

adjustment of the opacity curve. This can be tricky sometimes especially at low contrast states. By using different colors for the airways and displayed surrounding anatomy, topographic relationships can be displayed clearly (Fig. 7). The LTT as well as the surrounding anatomy can be shown in a comprehensive way.

3.4. *Hybrid rendering*

The combination of "Surface Rendering" and "Volume Rendering" is called "Hybrid Rendering". This makes sense, if in low contrast situations (e.g. tumors) the pathoanatomy cannot be displayed by "Volume Rendering" alone. Prior segmentation (even using manual tracing) allows to define the boundaries of low contrast structures. By displaying both, the segmented boundaries by "Surface Rendering" and the surrounding anatomy by "Volume Rendering", usually a sufficient 3D representation, can be achieved.

4. Virtual Endoscopy

Virtual Endoscopy (VE) was defined as a method that creates visualizations from 3D medical image scans similar to those produced by fiberoptic endoscopy.[26] There are many synonyms of VE especially concerning the gastro-intestinal tract: CT-based virtual endoscopy, virtual colonoscopy, CT colography, three dimensional S-CT pneumocolon, 3D colonography, to name a few.[27] Rogers has suggested a new policy in naming VE images, mainly attaching the suffix-graphy to the organ system rendered. In order to indicate which imaging modality was used, a prefix is formed from that modality e.g. CT-tracheabronchography (CT-TB), CT-colonography.[27] The generation of these virtual views from inner body surfaces is based on surface — and volume rendering algorithms as described above. Additionally perspective is used, i.e. objects closer to the virtual camera will appear larger than objects of the same size farther away from the virtual camera (Fig. 8). It is the same effect as looking down from a skyscraper: A person just beneath us appears properly sized whereas people down in the street appear tiny.

Fiberoptic Tracheoscopy (FTB) enables the inspection of the airway surface including mucosal changes as well as the dynamics. Information regarding the surrounding anatomy is limited to the perception of abnormal shapes or vessel

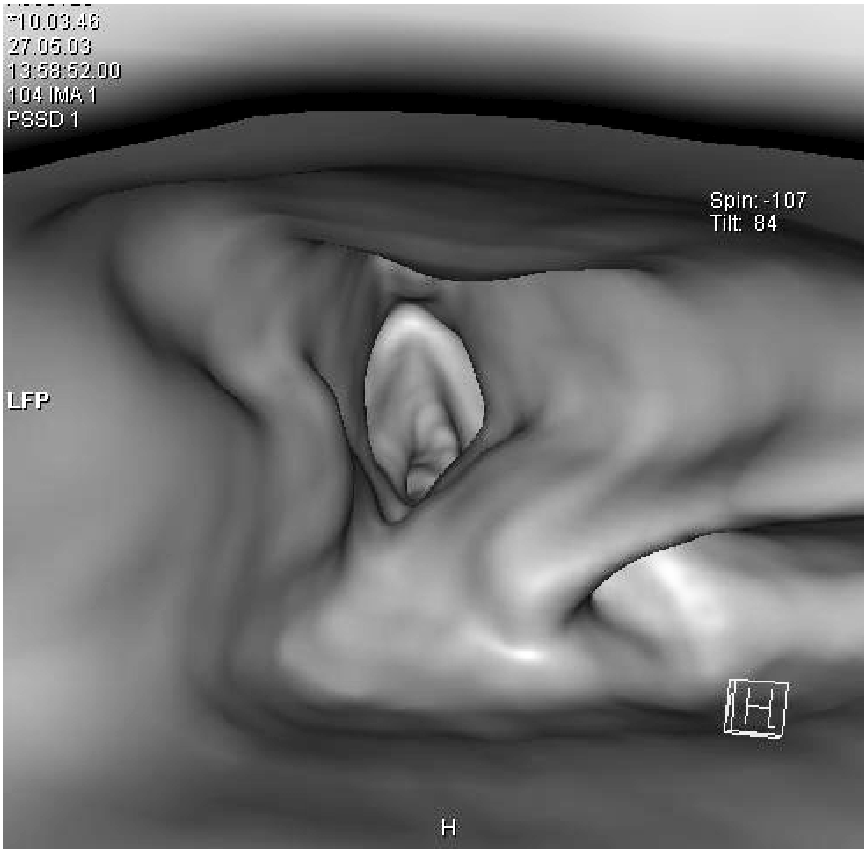


Fig. 8. Virtual endoscopic view of the larynx entrance based on S-CT.

pulsations. S-CT on the other hand, due to its excellent spatial and contrast resolution, provides information on intra- and extraluminal anatomy, but visualization of mucosal changes is not possible. As mentioned above, the shape of the trachea or bronchus depends on the angle between the axis of the trachea or bronchus and the CT slice plane. In a normal individual, where the trachea is just a little bent to the right and slightly angulated from superior anterior to caudal posterior, it can be expected that the axial S-CT slice will characterize the shape of the airways properly. But this may not be assumed in pathological cases, where the trachea can be buckled in any direction. Therefore LTS characterization on axial slices can be a troublesome topic. But CT-TB enables the radiologist to navigate through the airways in any direction interactively, to inspect every part from different views and to assess changes in diameter and shape directly, similar to FTB. In addition, findings of CT-TB can be easily compared to those of FTB. This is supported by a study of McAdams, who compared findings on axial CT slices and those of CT-TB in lung transplant recipients regarding the length and degree of airway stenosis.[28] They concluded that CT-TB was more accurate than axial CT for diagnosis of clinical relevant stenosis. A study of our institution revealed, that by comparing findings of axial CT slices (including MPR) with axial CT slice, MPR and virtual endoscopy, there was a clear advantage of using all three display modes and a reduction of false negative findings.[13] At FTB, the film documentation of a patients CT will be on the light box and the endoscopist tries to match the information from FTB with that of CT in his/her mind.

Computer simulations can help in this situation. Since the underlying S-CT slices of CT-TB contain information about the surrounding anatomy too, this can be exploited by displaying additional views. As shown in Fig. 9 the global view of the airways, the axial S-CT slice and the virtual endoscopic view can be displayed simultaneously. The position of the virtual camera is marked on all views in order to establish anatomical cross reference. This display allows one to study the topographic relationships of a patient's anatomy in a comprehensive way. In addition, this kind of display is a promising tool for teaching students and residents. There are even more advantages of CT-TB. At planning for a transbronchial biopsy, the best suited place for sampling can be chosen interactively.[29] Potential hazards of injuring vessels or other vital structures can be simulated without any danger to the patient. If transparent rendering of the tracheal wall is used, the extraluminal anatomy can be inspected within the 3D shape. Airways that cannot be explored by FTB, can be passed with the virtual endoscope and virtual, retrograde endoscopic views can be computed as well (Fig. 10). This is far beyond the possibilities of FTB. Moreover, for imaging of caliber changes during the respiratory cycle dynamic CT has to be performed, which is not undertaken routinely at every institution due to the inherited increased radiation. Therefore both, FTB and S-CT including CT-TB, are not competitive but complementary.

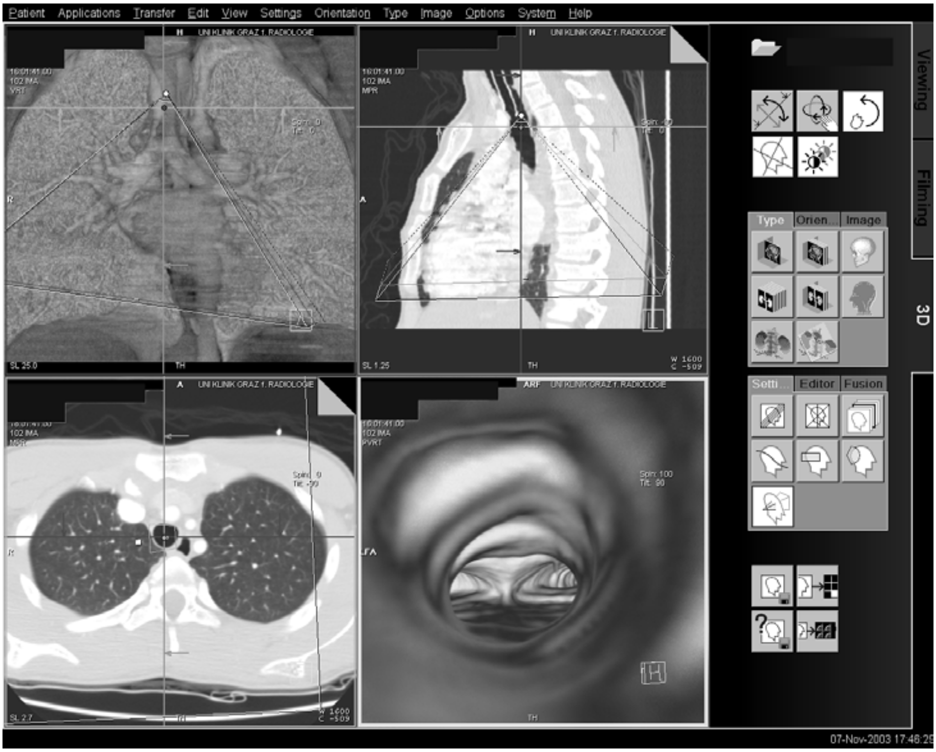
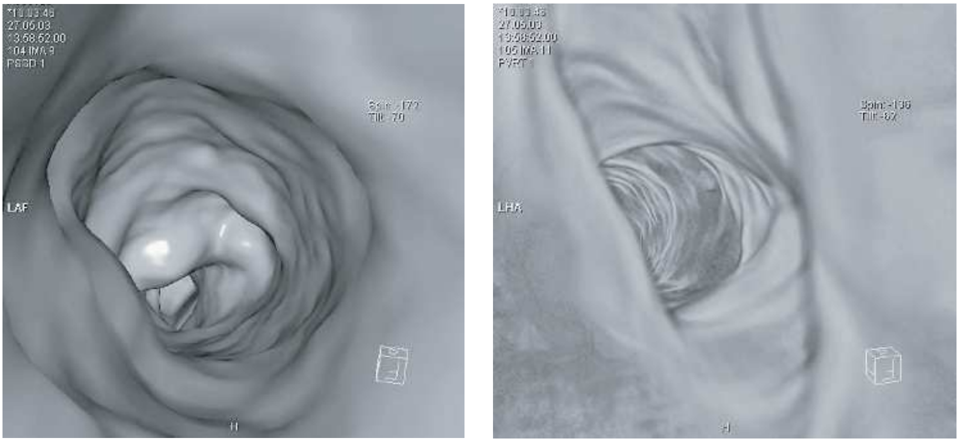

Patient Applications Transfer Edit View Settings Orientation Type Image Options System Help
UNI KLINIK GRAZ 1. RADIOLOGIE
VRT
MPR
PVRT
SL 25.0
SL 1.25
SL 2.7
TH
W 1600
C -509
Viewing
Filming
3D
Type
Orien...
Image
Setti...
Editor
Fusion
07-Nov-2003 17:46:29

Fig. 9. "Virtual Endoscopy" — composition of the workstation screen, which is divided into four parts: Left upper part displays a "Volume Rendered" semitransparent 3D reconstruction of the chest, the right upper part a sagittal MPR, whereas in the left lower part the axial slice in lung window settings is shown. On all three parts there is superposition of the position and opening angle of the virtual camera. On the right lower quadrant the virtual endoscopic view using "Volume Rendering" is shown.

5. Objective LTS Degree and Length Estimation Using LTT 3D — Tracheal Cross Sectional Profiles

Clinical management of patients suffering from LTS is based on FTB and imaging modalities. Based on FE there are several classifications for LTS, but they are either not practicable or do not predict the clinical course.[1,30] Moreover, at FTB the estimation of the length and degree in LTS is regarded to be operator dependent.[31] Unfortunately imaging modalities have their inherent weak points too. Conventional radiographs allow one to estimate the sagittal and transverse diameter of the airways. For assessment of LTS, where the trachea may be shaped asymmetrically, sagittal and transversal diameters do not characterize the shape of the airways properly. This was confirmed by Huber *et al.* who investigated different methods including radiographs of the neck, tracheoscopy, direct surgical measurements as well as necropsy measurements for assessment of tracheal stenosis.[32] They concluded that accurate measurements of tracheal stenosis cannot be done by radiographs and



(a) (b)

Fig. 10. Virtual retrograde, endoscopic view of a tracheal stenosis (same patient as in Fig. 7), the cube in the right lower corner indicates the viewing position — the capital letter "F" stands for viewing direction from feet, "P" for posterior — the dorsal tracheal wall is in the upper part of the image, the ventral wall on the lower part. (a) "Surface Rendering": The narrowed lumen at the caudol border of the stent can be seen, (b) "Volume Rendering": The tracheal wall is semitransparent and the dark areas represent the implanted tracheal stent.

tracheoscopy. Using conventional CT or electron beam CT, changes in the cross sectional area of the upper respiratory tract had been reported for healthy volunteers and for the evaluation of chronic airway obstruction in children.[33,34] In both studies the cross sectional area was determined on axial slices alone. As mentioned already, in the case of a LTS, where the trachea may not be straight but buckled in any direction, the measured cross sectional area on axial slices will not be a reliable characterization of the lumen as it will for healthy individuals. Own investigations yielded, that estimating LTS degree on axial S-CT slices and MPR alone are burdened by a high interobserver error — on average 43.3%, up to a maximum of 141.2%.[35] For the same reason the length of a stenosis cannot be estimated by just calculating the difference of slice positions. Curved multiplanar reformation could be used for length measurements, thus making it necessary to draw the medial axis on sagittal or coronal reconstructions. In case of LTS with a buckled trachea, this is a difficult task and the resulting medial axis will be again operator dependent. In addition, not all vendors of medical workstations are capable of obtaining length measurements from curved multiplanar reformations. 3D reconstructions, as described in Sec. 3, helps to display the 3D shape and extension of the airways but the length and degree of LTS have to been estimated visually and semi quantitative by the reporting radiologist. Although the authors has no scientific evidence, it is their belief, that this visual assessments will suffer from an similar interoperator error as FTB. A potential solution for these problems is the calculation of

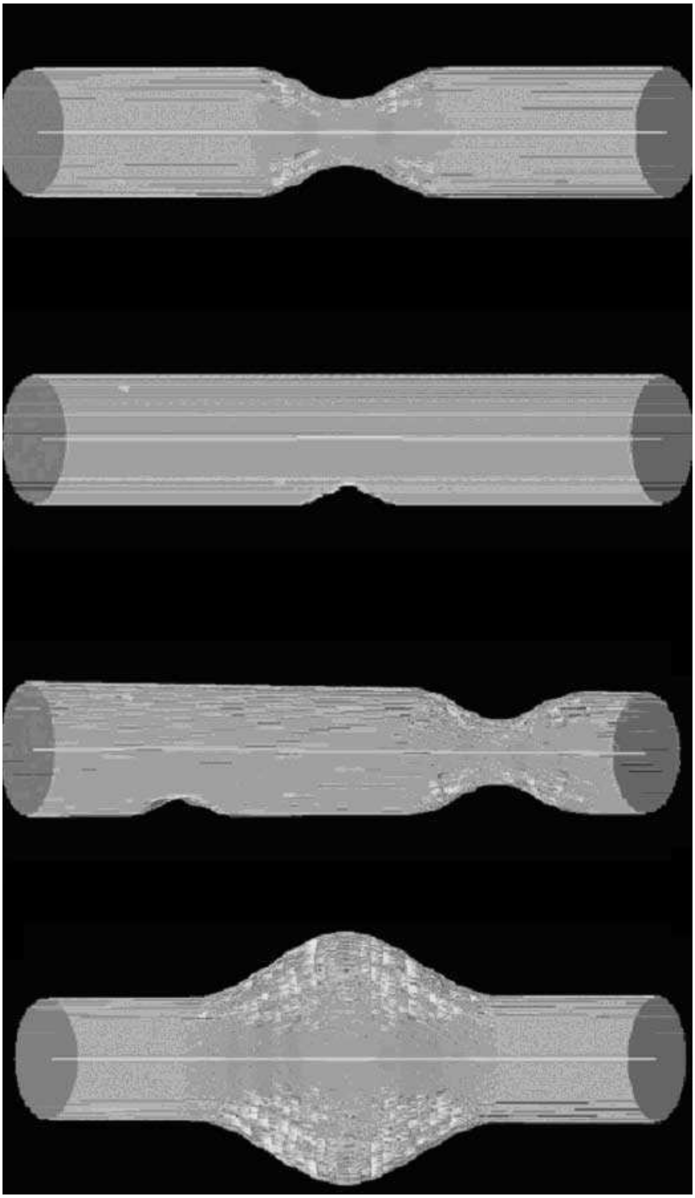


Fig. 11. VGB 1–4 for assessing accuracy and precision of the 3D cross-sectional profile. They consisted of tubes with symmetric and asymmetric narrowings (upper three) and an intersection of a sphere with a tube (bottom).

the S-CT based LTT 3D cross-sectional profile using a skeletonization algorithm[11] as described in Sec. 2. The (marked) drop of the cross sectional area outlines the stenosis. By subtraction of the end position from the start one, the true length in 3D can be calculated along the LTT medial axis. Therefore, the tracheal caliber

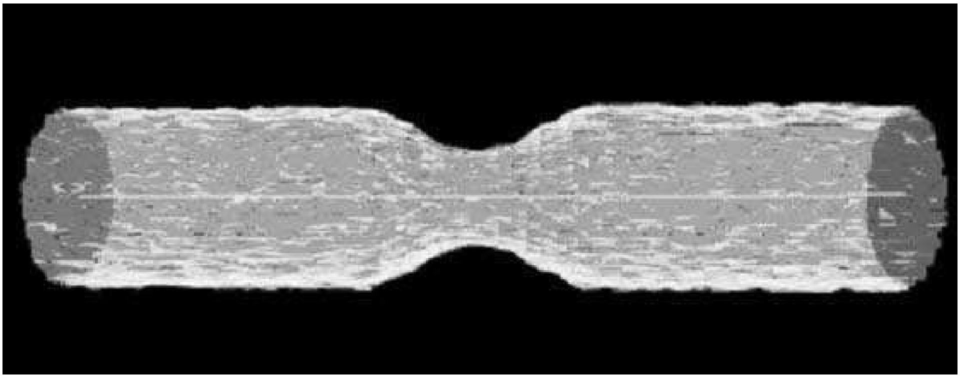


Fig. 12. VGB with added surface noise.

changes can be displayed on this charts in a quantitative way and length and degree of LTS can be determined, as described in Sec. 2. Accuracy and precision of the 3D cross sectional profiles were evaluated using virtual geometric bodies (VGB) containing narrowed and expanded areas (Fig. 11). These VGBs were generated mathematically in 3D as binary data volumes and for symmetrical caliber changes the true 3D cross sectional profile was known as well as the degree and length of caliber changes. Furthermore, in order to simulate (the always existing) noise in the CT slices, VGBs with a "rough" surface were computed too (Fig. 12).[11] Comparison of the true and the computed 3D cross sectional profile by linear regression yielded excellent results (Fig. 13).

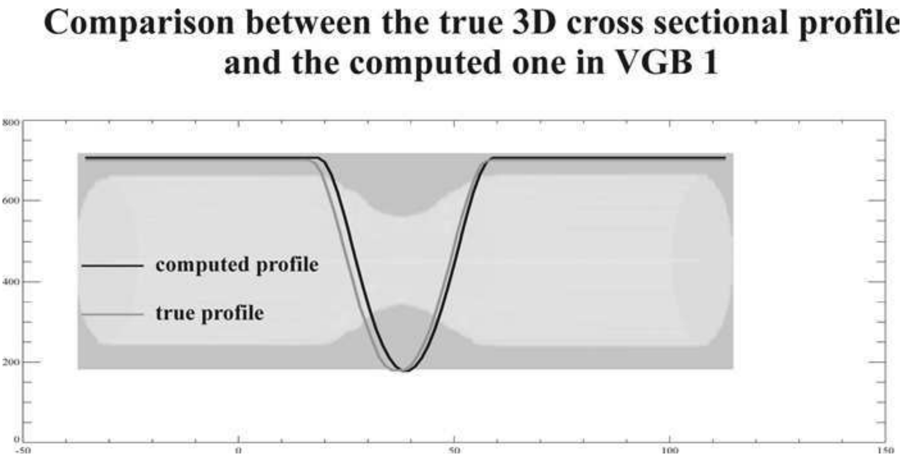


Fig. 13. Correlation between the true 3D cross sectional profile (gray line) and computed one, obtained by skeletonization (black line) in the VGB as displayed in Fig. 11. Both 3D cross sectional profiles are nearly identical and linear regression revealed an excellent correlated between both ($p < 0.005$).

6. Intranet Applications

As described in the preceding sections there are several possibilities for imaging of LTS. Unfortunately, the above described techniques for LTS imaging and assessment involve different computer platforms, operating systems and software, not all being embedded in an existing Picture Archiving and Communications System (PACS). At the Department of Radiology Graz/Austria an intranet application was developed, which allows one to collect all images and data of a particular patient using standard network computer interfaces (e.g. ftp, network drives, samba clients). Using the scripting language PHP, the application is fully independent of operating system.[36] Since the software runs in a web browser window, it can be started from any computer connected to the hospitals network by just using a web browser. The collected data are processed automatically and displayed as web pages on the hospitals intranet. Multimedia content, like digital videos from virtual endoscopy, can be viewed by using free software like media players, which can be downloaded from the internet for almost all operating systems. Therefore the referring physicians can observe these videos even on his/her desktop computer at no extra costs.

On the internet, web server software is free available for almost all computer and operating systems, the most utilized being the Apache software.[37] Therefore, as long as a hospital's network exists, almost every computer can be turned in a web server without spending additional money for infrastructure. Moreover, by using a SQL database (e.g. MySQL) for patient data administration, automated web server administration systems can be programmed.[38] At our institution such software keeps the list of accessible patient data as low as reasonable, which more or less means, not to display unwanted patients. Algorithms like "First in first out" will prevent the display of a particular patient at a prior chosen watermark — e.g. all patients data generated more than three months ago will not be shown. Cash algorithms represent a better solution. The software will check the date of the patient's data last assessment and remove those who were not opened for a previously fixated time span. Therefore only the needed patients are kept dynamically. Of course, all patient data, not saved to the PACS System, have to be archived according the national laws and attention has to be paid to security items.

7. Conclusion

As described in the previous sections, post processing of S-CT offers challenging possibilities for radiology including LTS imaging and assessment. The inherent information of S-CT can be displayed in different ways to the referring clinician in order to facilitate optimal patient management. Modern information technology and affordable computer hardware allow for new ways of data exchange and interaction between hospital departments. By using internet technology, asynchronous, interactive access to results of imaging and post processing can be provided. For imaging of LTS different facet's can be shown in order to provide a road map for therapeutical decisions.

References

1. M. Couray and R. Ossoff, Laryngeal stenosis: A review of staging, treatment and current research, *Current Opinion in Otolaryngology, Head and Neck Surgery* **6** (1998) 407–410.
2. H. Grillo, D. Dnonahue, D. Mathiesen, J. Wain and C. Wright, Postintubation tracheal stenosis: Treatment and results, *J. Thorax Cardiovasc. Surg.* **109** (1995) 486–492.
3. C. Lano, J. Duncavage, L. Reinisch, R. Ossoff, M. Couray and J. Netterville, Laryngotracheal reconstruction in the adult: A ten year experience, *Ann. Otol. Rhinol. Laryngol.* **107** (1998) 92–96.
4. H. Spraggs and P. Tostevin, Management of laryngotracheobronchial sequelae and complications of relapsing polychondritis, *Laryngoscope* **107** (1997) 936–941.
5. W. Berdon, V. Condon, G. Currarino, C. Fitz, J. Leonidas, B. Parker, T. Slovis and B. Wood, *Caffey's Pediatric X-Ray Diagnosis*, (Mosby, 1993).
6. J. Coleman, J. VanDuyne and R. Ossoff, Laser treatment of lower airway stenosis, *Otolaryngol. Clin. North Am.* **28** (1995) 771–783.
7. R. Ossof, G. Tucker, J. Duncavage and R. Toohill, Efficacy of bronchoscopic carbon dioxide laser surgery for benign strictures of the trachea, *Laryngoscope* **95** (1985) 1220–1223.
8. J. Czaja and T. McCaffrey, Acoustic measurement of subglottic stenosis, *Ann. Otol. Rhinol. Laryngol.* **105** (1996) 504–509.
9. B. Kuszyk, D. Heath, D. Bliss and E. Fishman, Skeletal 3D-CT: Advantages of volume rendering over surface rendering, *Skeletal Radiol.* **25** (1996) 207–214.
10. A. Zeiberg, P. Silverman, R. Sessions, T. Troost, W. Davros and R. Zeman, Helical (spiral) ct of the upper airway with three-dimensional imaging: Technique and clinical assessment, *AJR* **166** (1996) 293–299.
11. E. Sorantin, C. Halmai, B. Erdöhelyi, K. Palágyi, L. G. Nyúl, K. Ollé, B. Geiger, F. Lindbichler, G. Friedrich and K. Kiesler, Spiral-CT based assessment of tracheal stenoses using 3D-skeletonization, *IEEE TMI* **21** (2002) 263–273.
12. M. Lacrosse, J. Trigaux, B. van Beers and P. Weynants, 3D spiral CT of the tacheobronchial tree, *J. Comput. Assist. Tomogra.* **19**, 3 (1995) 341–347.
13. E. Sorantin, B. Geiger, F. Lindbichler, E. Eber and G. Schimpl, CT-based virtual tracheobronchoscopy in children — Comparison with axial CT and multiplanar reconstruction: Preliminary results, *Pediatric Radiol.* **32** (2002) 8–15.
14. J. Udupa and S. Samarasekera, Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation, *Graphical Models and Image Processing* **58**, 3 (1996) 246–261.
15. J. Udupa, D. Odhner, S. Samarasekera, R. Goncalves and K. Lyer, 3DVIEWNIX: An open, transportable, multidimensional, multimodality, multiparametric imaging software system, *SPIE Proceedings* **2164** (1994) 58–73.
16. D. Mohadjer, Analyse der Einflußfaktoren und Reproduzierbarkeit von Querschnittsprofilen zur Charakterisierung von Trachealllumina, *PhD thesis* (University–Hospital Graz, Karl Franzens University, Austria, 2001).
17. H. Blum, A transformation for extracting new descriptors of shape, *Symposium on Models for the Perception of Speech and Visual Form* (1964).
18. T. Kong and A. Rosenfeld, Digital topology: Introduction and survey, *Computer Vision, Graphics, and Image Processing* **48** (1989) 357–393.
19. K. Palágyi and A. Kuba, A parallel 3D 12-subiteration thinning algorithm, *Graphical Models and Image Processing* **61** (1999) 199–221.
20. K. Palágyi and A. Kuba, A 3D 6-subiteration thinning algorithm for extracting medial lines, *Pattern Recognition Letters* **19** (1998) 613–627.

21. G. F. Farin, *Curves and Surfaces for Computer Aided Geometric Design: A Practical Guide* (London Academic Press, 1990).
22. M. Remy-Jardin, J. Remy, D. Artaud, M. Fribourg and A. Duhamel, Volume rendering of tracheobronchial tree: Clinical evaluation of bronchographic images, *Radiology* **208** (1998) 761–770.
23. C. Gonzalez and R. Woods, *Morphology* (Addison-Wesley-Publishing Company, 1993).
24. J. Boissonnat and B. Geiger, Three dimensional reconstruction of complex shapes based on the (delaunay) triangulation. In R. Acharya and D. Goldgof, eds., *Biomedical Image Processing and Biomedical Visualization, SPIE Proceedings, San Jose, CA*, pp. 964–975 (1993).
25. R. Carey and G. Bell, *The Annotated VRML 2.0 Reference Manual* (Addison-Wesley Developers Press, 1996).
26. D. Blezek and R. Robb, Evaluating virtual endoscopy for clinical use, *Journal of Digital Imaging* **3**, Suppl 1 (1997) 51–55.
27. L. Rogers, A day in the court of lexicon: Virtual endoscopy, *AJR* **171** (1998) 1185.
28. H. McAdams, S. Palmer, J. Erasmus, E. Patz, J. Connolly, P. Goodman, D. Delong and V. Tapson, Bronchial anastomotic complications in lung transplant recipients: Virtual bronchoscopy for non-invasive assessment, *Radiology* **209** (1998) 689–695.
29. G. Rubin, C. Beaulieu and V. Argiro, Perspective volume rendering of CT and MR images: Applications for endoscopic imaging, *Radiology* **199** (1996) 321–330.
30. T. McCaffrey and J. Czaja, Classification of laryngeal stenosis, *Laryngoscope* **102** (1992) 1335–1340.
31. B. Jewett, R. Cook, K. Johnson, T. Logan, K. Rosbe, S. Mukherji and W. Shockley, Subglottic stenosis: Correlation between computed tomography and bronchoscopy, *Ann. Otol. Rhinol. Laryngol.* **108** (1999) 837–841.
32. M. Huber, R. Henderson, S. Finn-Bodner, D. Macinitre, J. Wrigh and G. Hankes, Assessment of current techniques for determining tracheal luminal stenosis in dogs, *AJVR* **10** (1997) 1051–1054.
33. E. Stern, C. Graham, R. Webb and G. Gamsu, Normal trachea during forced expiration: Dynamic CT measurements, *Radiology* **187** (1993) 27–31.
34. E. Frey, W. Smith, S. Grandgeorge, P. McCray and J. Wagener, Jr., Chronic airway obstruction in children: Evaluation with cine-CT, *AJR* **148** (1987) 347–352.
35. E. Sorantin (unpublished data).
36. http://www.php.net.
37. http://www.apache.com.
38. http://www.mysql.com.

CHAPTER 4

EDGE PRESERVED DENOISING IN MAGNETIC RESONANCE IMAGES AND THEIR APPLICATIONS

PAUL BAO
School of Computer Engineering
The Nanyang Technological University Singapore

LEI ZHANG
Department of Electrical & Computer Engineering
McMaster University, Hamilton, Canada

Edge-preserving image enhancement and noise removal are of great interest in medical imaging. This chapter describes schemes for noise suppression of magnetic resonance images using wavelet multiscale thresholding. To sufficiently exploit the wavelet interscale dependencies, we multiply the adjacent wavelet subbands of a Canny-edge-detector-like dyadic wavelet to form a multiscale product function where the significant features in images evolving with high magnitude across wavelet scales are amplified while noises are deteriorated, which facilitates an easy differentiation of edge structures from noise. Thereafter an adaptive threshold is calculated and imposed on the products, instead directly on the wavelet coefficients, to identify important features. Experiments show that the proposed scheme outperforms other wavelet-thresholding denoising methods in suppressing noise and preserving edges.

Keywords: Edge preserved denoising; magnetic resonance; wavelet.

1. Introduction

Magnetic Resonance Imaging (MRI) is a compelling diagnostic technique. The incorporated noise during image acquisition however degrades the human interpretation, or computer-aided analysis of the features in the MRI images. Time averaging of image sequences aimed at improving the Signal to Noise Ratio (SNR) would lead to additional acquisition time and reduce the temporal resolution. Therefore denoising instead should be performed to enhance the image quality aiming for more accurate diagnosis.

Wavelet transform[1–5] based noise reduction schemes, many of which have appeared in literature[18–23] during the last two decades, proved to be very effective in noise removal and feature preserving. Most of wavelet based denoising schemes consider the incorporated noise as additive Gaussian white. MRI magnitude images are however usually modeled by a *Rician* distribution[9,10] and the so-called *Rician* noise (the error between the underlying image intensities and the measurement data) is locally signal dependent. The *Rician* noise distribution can be well approximated by a Gaussian in bright (high SNR) regions while a Rayleigh distribution is more appropriate in dark (low SNR) regions.

During the past decade, a number of wavelet-based image denoising and enhancement schemes specifically designed for medical images have been proposed. A wavelet-based Wiener-filter-like denoising method accounting for the *Rician* noise was proposed by Nowak,[8] where the magnitude MRI image is squared and the square of the *Rician* random variable is modeled by a scaled non-central chi-square distribution. Although the noise in magnitude MRI images are *Rician*, the additive Gaussian white noise assumption holds for each component of the complex MRI data in k-space. Zhong *et al.*[27] presented a wavelet denoising algorithm using Sub-domain Noise Detection (SND) and Adaptive Least Squared Error Clustering (ALSEC). This algorithm is particularly effective in denoising poor SNR medical images. Ouda *et al.*[30] proposed an adaptive signal-preserving technique for noise suppression in functional magnetic resonance imaging data based on spectrum subtraction. The method estimates a model for the power spectrum of random noise from the acquired data. Pizurica *et al.*[16] proposed wavelet noise filtering method adaptive to various types of image noise and to the preference of the medical expert. The algorithm defines a parameter aimed at balancing the preservation of expert-dependent relevant details against the degree of noise reduction and classifies the coefficients based on the correlation of significant image features across the resolution scales in estimating the statistical distributions of the coefficients that represent useful image features and noise. Assuming that noise in *no signal* regions of magnetic resonance magnitude images is Rayleigh distributed, Wu *et al.*[31] presented an effective wavelet-based denoising technique which can work as a standalone denoising procedure or couple with existing denoising algorithms to enhance their effectiveness. Lysaker[32] presented an image smoothing method for medical magnetic resonance images based on a fourth-order partial differential equation model. The method has demonstrated good noise suppression at poor signal-to-noise ratio.

Denoising can be applied to the real and imaginary channels respectively rather than to the magnitude images. It has been shown that wavelet denoising techniques will yield better edge preservation if performed on the raw real and imaginary images prior to rectification.[11–13] Wood[28] showed while magnitude and complex denoising both significantly improved SNR, SBR, and CNR, complex denoising yielded sharper edge resolution and feature extraction. In view of this, the additive Gaussian white noise model is adopted in this chapter.

With the denoising applied on the real and imaginary channels prior to rectification, the MRI image denoising becomes more or less a generic image denoising issue where denoising schemes are well researched and richly available. Among the wavelet-based noise reduction techniques, nonlinear thresholding is simple yet very effective. In his innovative work,[19] Donoho showed that the *Universal* threshold $t = \sigma\sqrt{2\log N}$ is asymptotically optimal in the *minimax* sense, where σ is the standard deviation of additive white noise and N is the sample length. However, it is well known that the *Universal* threshold over-smoothes images. Donoho improved his work[20] using the *SURE* threshold. It is subband adaptive and is derived by minimizing Stein's unbiased risk estimator. Recently, by modeling the wavelet coefficients within each subband as *i.i.d* random variables with generalized Gaussian Distribution (GGD), Chang *et al.*[23] proposed the *BayesShrink* scheme. The *BayesShrink* threshold is also subband dependent and yields better results than the *SURE* threshold. The thresholds mentioned above are based on orthogonal wavelets and are *soft*, implying that the input w is shrunk to zero by an amount of threshold t. In Ref. 22 Pan *et al.* presented a *hard* threshold with a non-orthogonal wavelet expansion. Denoting the standard deviation of noise at the jth wavelet scale by σ_j, Pan *et al.* imposed $t_j = c\sigma_j$, where σ_j is the standard deviation of noise at the jth scale and $c \in [3, 4]$ is a constant, to identify significant structures. The word *hard* implies that the input w is preserved if it is greater than the threshold; otherwise it is set to zero. The factor c is a constant and Pan *et al.* set it around 3 since the values of Gaussian distributed noise are, in high probability, within three times its standard deviation. The *soft BayesShrink* and Pan's *hard thresholding* are used for comparison in the sequel.

There exist dependencies between wavelet coefficients. In Ref. 7, Crouse *et al.* used the Hidden Markov Tree (HMT) models to characterize the joint statistics of wavelet coefficients across scales. In the noise reduction technique of Pizurica,[16] the interscale correlation information is exploited to classify the wavelet coefficients. The preliminary classification is then used to estimate the distribution of a coefficient to decide if it is a feature. If a coefficient at a coarser scale has small magnitude, its descendant coefficients at finer scales are also likely to be small. Shapiro exploited this property to develop the well-known embedded zerotree wavelet coder.[6] Conversely, if a wavelet coefficient produced by a true signal is of large magnitude at a finer scale, its parents at coarser scales are likely to be large as well. However for those coefficients caused by noise, the magnitudes will decay rapidly along the scales. With this observation, Xu *et al.*[18] multiplied the adjacent wavelet scales to sharpen the important structures while weakening noise. They developed a spatially selective filtering technique by iteratively selecting edge pixels in the multiscale products. Sadler and Swami[24] analyzed the multiscale products and applied them to step detection and estimation. Both Xu's and Sadler's works are implemented with a dyadic wavelet constructed by Mallat and Zhong.[4] The so-called MZ wavelet is a compactly supported quadratic spline function that approximates the first derivative of Gaussian. The corresponding dyadic wavelet transform is equivalent to the Canny edge detection[17] and characterizes the instantaneous features in a signal well. The MZ wavelet is also employed in this chapter.

Wavelet thresholding is simple and efficient but takes little or no advantage of the dependency information between wavelet scales. In this chapter, we present a multiscale thresholding scheme to incorporate the merits of interscale dependencies into the thresholding technique for denoising. Two adjacent wavelet subbands are multiplied to amplify the significant features and dilute noise. In contrast to other schemes, we apply thresholding to the multiscale products instead of the wavelet coefficients. As we will show, the proposed multiscale products thresholding can effectively distinguish edge structures from noise. The variance of noise needs to be estimated to implement the denoising scheme. A new noise level estimator is also proposed in this chapter.

The chapter is organized as follows. Section 2 discusses the wavelet multiscale products. Section 3 describes the thresholding scheme. An image adaptive threshold imposed on the multiscale products is calculated to identify the significant structures. Experiments are given in Sec. 4 in comparison with some existing wavelet thresholding schemes. The chapter is concluded in Sec. 5.

2. Wavelet Multiscale Products

2.1. *Dyadic wavelet transform as a multiscale edge detector*

A wavelet transform represents a signal f as a linear combination of elementary atoms that appear at different resolutions. It is computed by convoluting the input signal with dilated wavelet filters recursively. More details about the theory of wavelets and their applications in signal processing can be found in Daubechies,[1] Meyer,[2] Mallat[3,4] and Vetterli.[5]

We denote by $\xi_s(x)$ the dilation of a function $\xi(x)$ by a scale factor s:

$$\xi_s(x) = \xi(x/s)/s. \tag{1}$$

Suppose function $\psi(x)$ satisfies the requirements to be a wavelet. The continuous wavelet transform of any measurable and square-integrable function $f(x)$, $f \in L^2(R)$, at scale s and position x is defined as

$$W_s f(x) = f * \psi_s(x), \tag{2}$$

where the symbol $*$ denotes the convolution operation.

The wavelet transform can be designed as a multiscale edge detector to enhance the signal's instantaneous features.[4] Suppose that $\theta(x)$ is a differentiable smooth function whose integral is equal to 1 and that it converges to 0 at infinity. Lets define $\psi(x)$ as the first order derivative of $\theta(x)$

$$\psi(x) = d\theta(x)/dx, \tag{3}$$

then $W_s f(x)$ can be written as

$$W_s f(x) = f * \left(s\frac{d\theta_s}{dx}\right)(x) = s\frac{d}{dx}(f * \theta_s)(x). \tag{4}$$

It can be seen that the wavelet transform $W_s f(x)$ is the first derivative of $f(x)$ smoothed by $\theta_s(x)$. In particular, when $\theta(x)$ is a Gaussian function, the local extrema determination in $W_s f(x)$ is equivalent to the well-known Canny edge detection.[17]

The Canny edge detector-like wavelet transform can be extended to 2D images. Suppose $\theta(x, y)$ is a 2D differentiable smooth function whose integral is equal to 1 and converges to 0 at infinity. For example $\theta(x, y)$ could be the tensor product of 1D smooth functions: $\theta(x, y) = \theta(x) \cdot \theta(y)$. We define the two wavelets $\psi^x(x, y)$ and $\psi^y(x, y)$ at horizontal and vertical directions as:

$$\psi^x(x, y) = \partial\theta(x, y)/\partial x, \quad \psi^y(x, y) = \partial\theta(x, y)/\partial y. \tag{5}$$

The dilation of any 2D function $\xi(x, y)$ by scale s can be therefore denoted by

$$\xi_s(x, y) = s^{-2}\xi(x/s, y/s). \tag{6}$$

Suppose $f(x, y)$ is a 2D measurable and square-integrable function such that $f \in L^2(R^2)$. The wavelet transform of $f(x, y)$ at scale s and position (x, y) has two components

$$W_s^x f(x, y) = f * \psi_s^x(x, y) \quad \text{and} \quad W_s^y f(x, y) = f * \psi_s^y(x, y). \tag{7}$$

Similarly to (4) these two components can be rewritten as

$$\begin{aligned} W_s^x f(x, y) &= s\frac{\partial}{\partial x}(f * \theta_s)(x, y), \\ W_s^y f(x, y) &= s\frac{\partial}{\partial y}(f * \theta_s)(x, y). \end{aligned} \tag{8}$$

In the case when $\theta(x, y)$ is a Gaussian function, detecting the local extrema form $\begin{pmatrix} W_s^x f(x, y) \\ W_s^y f(x, y) \end{pmatrix}$ is equivalent to the Canny edge detection.

For the purpose of fast numerical implementation, we restrict the scale s to vary along the dyadic sequence $(2^j)_{j \in Z}$. For simplicity, we denote by $\xi_j(x)$ (no confusion with $\xi_s(x)$ in Eq. (1)) the dilation of function $\xi(x)$ by 2^j, then

$$\xi_j(x) = \xi(x/2^j)/2^j. \tag{9}$$

The dyadic wavelet transform (DWT) of $f(x)$ at dyadic scale 2^j and position x is

$$W_j f(x) = f * \psi_j(x). \tag{10}$$

The function $f(x)$ can be recovered from its DWT by

$$f(x) = \sum_{j=-\infty}^{\infty} W_j f * \chi_j(x), \tag{11}$$

where $\chi(x)$ is any reconstructing wavelet whose Fourier transform satisfies[4]

$$\sum_{j=-\infty}^{\infty} \hat{\psi}(2^j\omega)\hat{\chi}(2^j\omega) = 1. \tag{12}$$

The wavelet used in this chapter is the MZ wavelet constructed by Mallat and Zhong.[4] The associated smooth function $\theta(x)$ is a cubic spline, which closely approximates a Gaussian function. The wavelet $\psi(x)$ is a quadratic spline that approximates the first derivative of Gaussian. Thus the DWT behaves like a Canny edge detector. Section 2.2 shows the functions $\theta(x)$ and $\psi(x)$, and illustrates the discrete decomposition algorithms of the 1D and 2D DWT. Details about the derivation of the MZ wavelet can be found in Ref. 4.

2.2. *Wavelet and the discrete decomposition algorithm*

Mallat and Zhong[4] defined a class of wavelets that can be used in the implementation of dyadic wavelet transform. The Fourier Transform (FT) of the wavelet $\psi(x)$ is

$$\hat{\psi}(\omega) = i\omega\left(\frac{\sin(\omega/4)}{\omega/4}\right)^4. \tag{13}$$

Therefore the FT of its associated smooth function $\theta(x)$, the primitive of $\psi(x)$, is

$$\hat{\theta}(\omega) = \left(\frac{\sin(\omega/4)}{\omega/4}\right)^4. \tag{14}$$

The $\theta(x)$ is a cubic spline whose integral is equal to 1 and $\psi(x)$ is a quadratic spline. In Fig. 1, they are plotted and compared with a Gaussian function and its first derivative. It is noticed that the $\theta(x)$ approximates closely to the *Gaussian* function. The wavelet transform behaves like a Canny edge detector.[17]

The discrete decomposition algorithms of 1D and 2D dyadic wavelet transform are illustrated in Fig. 2. Filter H_j (G_j) is the 2^j scale dilation of H_0 (G_0) (putting $2^j - 1$ zeros between each of the coefficients of H_0 (G_0)). H_j' (G_j') is the transpose of H_j (G_j). The coefficients of filters H_0 and G_0 are available in Ref. 4. Suppose the input signal $S_0 f$ has N samples, then at each scale 2^j the wavelet coefficients $S_j f$ and $W_j f$ also have N samples. There are at most $\log_2 N$ scales and the complexity of the decomposition algorithm is $O(N \log_2 N)$.

It should be noted that in the discrete implementation, at each scale the wavelet coefficients should be sampled with a constant shift. For a 1D signal, we denote the discrete sample sequence by

$$dW_j f(n) = W_j f(n + w_j), \tag{15}$$

where the shift variable w_j is

$$w_j = 1, 2, \ldots, 2^{j-1}. \tag{16}$$

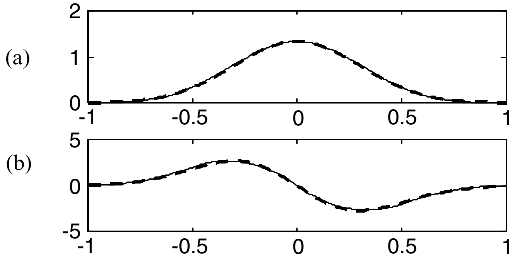
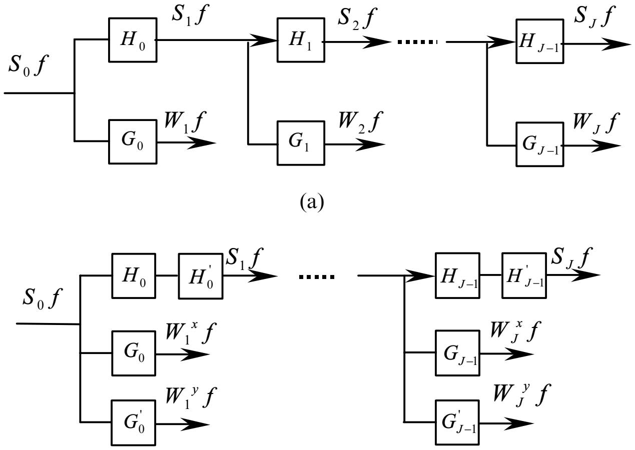


Fig. 1. (a) The smooth function $\theta(x)$ (solid) and a *Gaussian* function (dashed). (b) Wavelet $\psi(x)$ (solid) and the first derivative of the *Gaussian* in (a) (dashed).



(a)

(b)

Fig. 2. The discrete decomposition algorithms of (a) 1D Dyadic Wavelet Transform (DWT) and (b) 2D DWT, where filter H_j (G_j) is the 2^j dilation of H_0 (G_0) (putting $2^j - 1$ zeros between each of coefficients of H_0 (G_0)) and H'_j (G'_j) is the transpose of H_j (G_j).

For a 2D image, there are two sample sequences obtained in horizontal and vertical directions:

$$\begin{aligned} dW_j^x f(n,m) &= W_j^x f(n + w_j^{x,n}, m + w_j^{x,m}), \\ dW_j^y f(n,m) &= W_j^y f(n + w_j^{y,n}, m + w_j^{y,m}), \end{aligned} \tag{17}$$

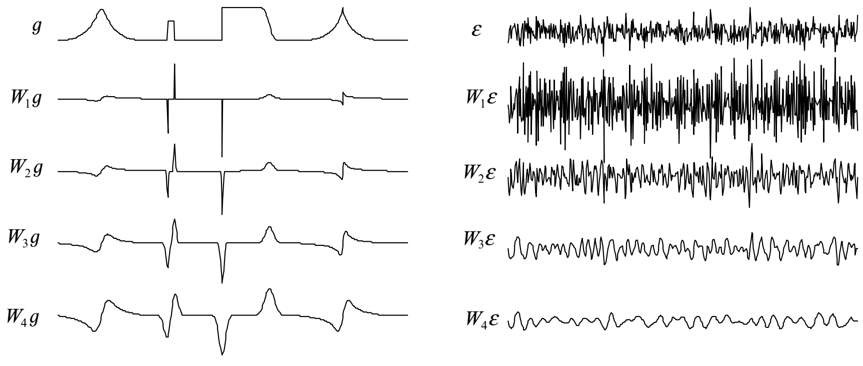
where the sample shifts are

$$\begin{aligned} w_j^{x,n} &= 1, 2, \ldots, 2^{j-1}, \quad w_j^{x,m} = 0, 1, 2, 4, \ldots, 2^{j-2}, \\ w_j^{y,n} &= 0, 1, 2, 4, \ldots, 2^{j-2}, \quad w_j^{y,m} = 1, 2, \ldots, 2^{j-1}. \end{aligned} \tag{18}$$

2.3. *Multiscale products*

Signals and noise behave very differently in the wavelet transform domain. The evolution of singularities and noise across wavelet scales were analyzed by Mallat *et al.*[3,4] using the mathematical concept of the Lipschitz regularity. Singularities are more regular than noise and have higher Lipschitz regularities. For example, the Lipschitz regularity of a step edge is 0. If a structure is smoother than the step, it will have positive Lipschitz regularity. Otherwise it can be considered having negative Lipschitz regularity. The Lipschitz regularity of the Dirac function is equal to -1. White noise is almost singular everywhere and has a uniform Lipschitz regularity that is equal to $-1/2$.

Meyer[2] presented a theorem to relate the evolution of the wavelet transform magnitude with the signal's Lipschitz regularity. A function $f(x)$ is uniformly Lipschitz α $(0 < \alpha < 1)$ over interval $[a,b]$ if and only if there exists a constant



(a) (b)

Fig. 3. (a) The Dyadic Wavelet Transform (DWT) of a test signal g at the first four scales. (b) The DWT of a sequence of *Gaussian* white noise ε at the first four scales.

$K > 0$ such that for all $x \in [a, b]$, the wavelet transform satisfies

$$|W_j f(x)| \leq K(2^j)^{\alpha}. \tag{19}$$

The above equation implies that the wavelet transform magnitudes increase for positive α with increasing scales. Contrarily, wavelet transform magnitudes decrease for negative Lipschitz regularities with increasing scales. In Fig. 3 the DWT at the first four scales of a test signal g, and a sequence of Gaussian white noise ε, are illustrated. Notice that the signal singularities evolve across scales with observable peaks while noise decays rapidly along scales. As illustrated in Fig. 3(b), Mallat and Hwang[3] observed that, for Gaussian white noise, the average number of local maxima at scale 2^{j+1} is half of that at scale 2^j.

With the observation of Fig. 3, we can imagine that multiplying the DWT at adjacent scales would amplify edge structures and dilute noise. This favorite property has been exploited by Xu *et al.*[18] and Sadler[24] in noise reduction and step detection. In this chapter, we define the multiscale products of $W_j f$ as

$$P_j f(x) = \prod_{i=-k1}^{k2} W_{j+i} f(x), \tag{20}$$

where $k1$ and $k2$ are non-negative integers.

The support of an isolated edge will increase by a factor of two across scale and the neighboring edges will interfere with each other at coarse scales (Fig. 3). So in practice it is sufficient to implement the multiplication at two adjacent scales. Let $k1 = 0$ and $k2 = 1$, then we calculate the DWT scale products as:

$$P_j f(x) = W_j f(x) \cdot W_{j+1} f(x). \tag{21}$$

Similarly for 2D images, the multiscale products have two components:

$$\begin{aligned} P_j^x f(x,y) &= W_j^x f(x,y) \cdot W_{j+1}^x f(x,y), \\ P_j^y f(x,y) &= W_j^y f(x,y) \cdot W_{j+1}^y f(x,y). \end{aligned} \tag{22}$$

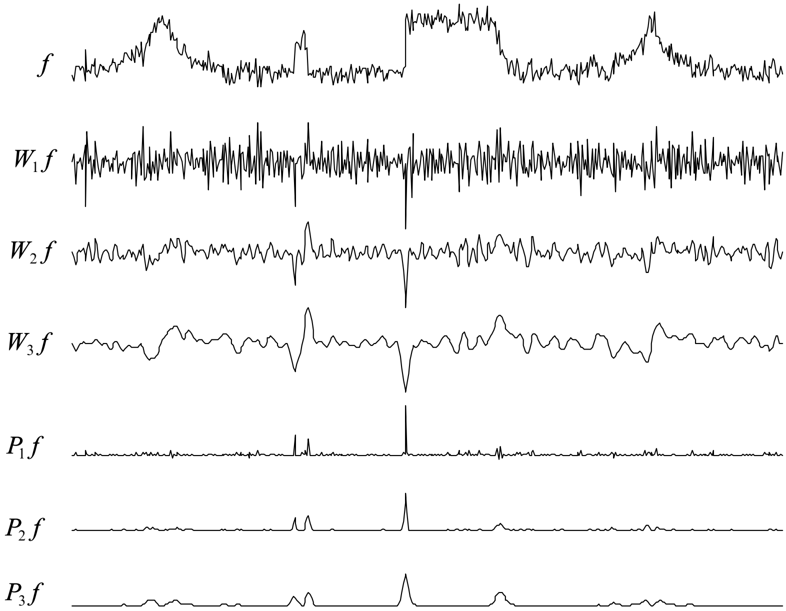


Fig. 4. The DWT and multiscale products of a noisy test signal at the first three scales.

In Fig. 4 the DWT and multiscale products of a noisy test signal $f = g + \varepsilon$ are illustrated. Though the wavelet transform coefficients of the original signal g are immersed into noise at fine scales, they are enhanced in the scale products P_jf. The significant features of g are more distinguishable in P_jf than in W_jf.

3. Adaptive Multiscale Products Thresholding

3.1. *The thresholding scheme*

Wavelet-based thresholding techniques have proved to be effective in denoising.[18–23] Non-significant wavelet coefficients below a preset threshold value are discarded as noise and the image is reconstructed from the remaining significant coefficients. Compared with the linear denoising methods that blur images as well as smoothing noise, the nonlinear wavelet thresholding schemes preserve image singularities better.

In general, thresholds are classified into *soft* and *hard*. The *soft* thresholds shrink the input wavelet coefficient w to zero by an amount t, i.e. $\eta_t(w) = \mathrm{sgn}(w) \cdot \max(0, |w| - t)$. Contrarily, the *hard* thresholds preserve the input coefficient if it is greater than the threshold, i.e. $\eta_t(w) = w \cdot \mathbf{1}\{|w| > t\}$. Naturally, the determination of the threshold value is extremely critical to the threshold-based algorithms. We denote by $f = g + \varepsilon$ the measurements of image g corrupted by Gaussian white noise $\varepsilon \sim N(0, \sigma^2)$. Donoho *et al.*[19] presented the *Universal* threshold $t = \sigma\sqrt{2\log N}$ in his well-known *Wavelet Shrinkage* scheme. Chang *et al.*[23]

presented the *BayesShrink* threshold $t = \sigma^2/\sigma_{W_j g}$, where $\sigma_{W_j g}$ is the image standard deviation at the jth wavelet scale. The above three thresholds are *soft* and are derived from orthogonal wavelet bases. A *hard* threshold that could be applied to non-orthogonal wavelet transforms was proposed by Pan *et al.*[22]

All the above wavelet thresholding schemes impose the threshold directly on wavelet coefficients. They do not exploit the dependencies that exist between adjacent wavelet scales. From Fig. 4 it can be noticed that, at finer scales if the threshold t applied to $W_j f$ is relatively sizeable, some edge structures may be suppressed as noise. Otherwise, if t is relatively small, many noisy pixels would be undesirably preserved. However, in the multiscale products $P_j f$ it can be seen that the significant structures are strengthened while the noise is weakened. $P_j f$ results in a more effective discrimination between edges and noise than $W_j f$. With such observations Xu *et al.*[18] and Sadler *et al.*[24] have exploited multiscale products in denoising and step estimation.

In this chapter, we propose a new denoising scheme, the *adaptive multiscale products thresholding*, to merge the merits of the thresholding technique and wavelet interscale dependencies. A significant wavelet coefficient $\hat{W}_j^d f(x, y)$, where $d = x, y$ indicates x or y dimension, is identified if its corresponding multiscale products value $P_j^d f(x, y)$ is greater than an adaptive threshold $t_p^d(j)$. The algorithm is summarized as follows:

(i) Compute the DWT of input image f up to J scales.
(ii) Calculate the multiscale products $P_j^d f$ and preset the thresholds $t_p^d(j)$. Then threshold the wavelet coefficients by:

$$\hat{W}_j^d f(x,y) = \begin{cases} W_j^d f(x,y) & P_j^d f(x,y) \geq t_p^d(j) \\ 0 & P_j^d f(x,y) < t_p^d(j) \end{cases}, \quad j = 1, \ldots, J; \;\; d = x, y \tag{23}$$

(iii) Recover the image from the thresholded wavelet coefficients $\hat{W}_j^x f(x, y)$ and $\hat{W}_j^y f(x, y)$.

3.2. *Determination of the threshold*

Since a wavelet transform is a linear transform, the DWT of a noisy image $f = g + \varepsilon$ can be written as

$$W_j^d f = W_j^d g + W_j^d \varepsilon, \tag{24}$$

where $W_j^d g$ is the DWT of original image g and $W_j^d \varepsilon$ is the DWT of additive noise ε. For convenience, we denote

$$Z_j^d = P_j^d f = W_j^d f \cdot W_{j+1}^d f. \tag{25}$$

Due to the high dependencies existing between $W_j^d f$ and $W_{j+1}^d f$, the histograms of Z_j^d will have a heavy positive tail (See Fig. 6). A proper threshold $t_p^d(j)$ can be determined and imposed on Z_j^d to eliminate the highly noise corrupted pixels and identify the significant image structures.

Suppose that the input image is Gaussian white noise ε and it is an ergodic stationary process. For the convenience of expression, we denote the DWT of ε by

$$U_j^*(x,y) = W_j^*\varepsilon(x,y) = \varepsilon * \psi_j^*(x,y), \quad * = x, y, \tag{26}$$

where U_j^* is a Gaussian colored noise process and its standard deviation is

$$\sigma_j = \|\psi_j\|\,\sigma, \tag{27}$$

where norm $\|\psi_j\| = \sqrt{\iint \psi_j^2(x,y)dx\,dy}$. We do not use the superscript "d" in Eqs. (3–5) because the norm values $\|\psi_j^x\|$ and $\|\psi_j^y\|$ are the same.

U_j^d and U_{j+1}^d are jointly Gaussian distributed with probability density function (pdf)[26]

$$p(u_j, u_{j+1}) = \frac{1}{2\pi\sigma_j\sigma_{j+1}\sqrt{1-\rho_{j+1,j}^2}} e^{-\frac{1}{2(1-\rho_{j+1,j}^2)}\left[\frac{u_j^2}{\sigma_j^2} - \frac{2\rho_{j+1,j}u_ju_{j+1}}{\sigma_j\sigma_{j+1}} + \frac{u_{j+1}^2}{\sigma_{j+1}^2}\right]}, \tag{28}$$

where the correlation coefficient $\rho_{j+1,j}$ of U_j^d and U_{j+1}^d is

$$\rho_{j+1,j} = \frac{\iint \psi_j(x,y)\cdot\psi_{j+1}(x,y)dx\,dy}{\sqrt{\iint \psi_j^2(x,y)dx\,dy \cdot \iint \psi_{j+1}^2(x,y)dx\,dy}}. \tag{29}$$

The values of σ_j and $\rho_{j+1,j}$ in the discrete implementation are listed in Table 1 by setting $\sigma = 1$.

We denote the scale products of U_j^d and U_{j+1}^d by

$$V_j^d = U_j^d \cdot U_{j+1}^d. \tag{30}$$

Then the pdf of V_j^d will have the following form[25]

$$p(v_j) = \frac{1}{\pi\Gamma(1/2)\sigma_j\sigma_{j+1}\sqrt{1-\rho_{j+1,j}^2}} e^{\frac{\rho_{j+1,j}v_j}{(1-\rho_{j+1,j}^2)\sigma_j\sigma_{j+1}}} K_0\left(\frac{|v_j|}{(1-\rho_{j+1,j}^2)\sigma_j\sigma_{j+1}}\right), \tag{31}$$

where $\Gamma(t) = \int_0^\infty e^{-u}u^{t-1}du$ is the Gamma function and K_0 modified Bessel function of the second kind with order zero. When $\rho_{j+1,j}$ is positive, $p(v_j)$ is right skewed. In Fig. 5, the theoretical pdf's $p(v_1)$ and $p(v_2)$ are plotted by setting $\sigma = 10$. Notice that $p(v_2)$ is more positively tailed than $p(v_1)$ because $\rho_{3,2}$ is higher than $\rho_{2,1}$.

Table 1. Noise standard deviation and correlation coefficient values of the MZ wavelet in discrete implementation at scale 2^j. The input noise is assumed to be unit *Gaussian* white.

j	1	2	3	4
σ_j	2.8284	0.7395	0.3173	0.1531
$\rho_{j+1,j}$	0.3586	0.5504	0.5957	0.6063

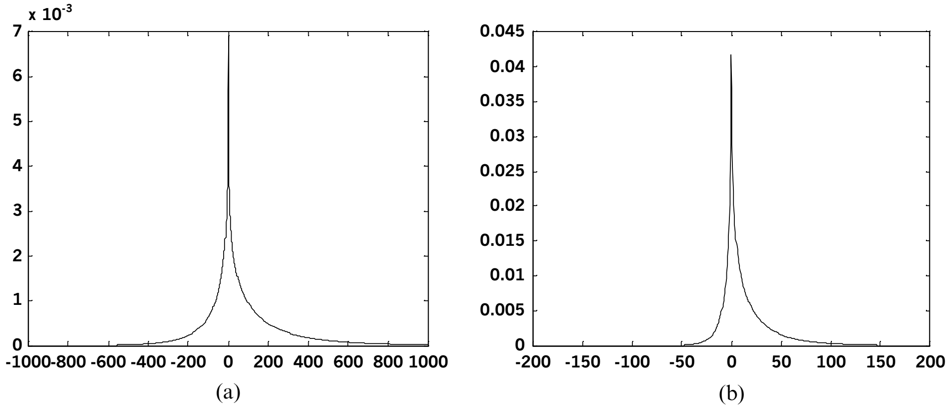


Fig. 5. The theoretical pdf of the multiscale products V_j for *Gaussian* white noise ($\sigma = 10$). (a) At the first scale $j = 1$. (b) At the second scale $j = 2$.

In applications the wavelet coefficient obtained is $W_j^d f$, which is the sum of noiseless coefficient $W_j^d g$ and noise $W_j^d \varepsilon$. Since white noise is singular almost everywhere, at fine scales $W_j^d \varepsilon$ will be predominant in $W_j^d f$ except for some significant features to be preserved (Refer to Fig. 4). In Fig. 6(a), the histograms of Z_j^x at the first three scales are plotted for the noiseless image *Lena* (i.e. $Z_j^x = W_j^x g \cdot W_{j+1}^x g$). Figure 6(b) shows the histograms of Z_j^x when the input is Gaussian white noise with zero mean and standard deviation $\sigma = 30$ (i.e. $Z_j^x = W_j^x \varepsilon \cdot W_{j+1}^x \varepsilon$). In Fig. 6(c), the histograms of Z_j^x, where the input is the noisy *Lena* ($\sigma = 30$, SNR $= 12.93\,\text{dB}$), are shown (i.e. $Z_j^x = W_j^x f \cdot W_{j+1}^x f$). It is noticeable that at scales 2^1 and 2^2, the corresponding histograms in Figs. 6(b) and (c) are very similar. This is because the energy of noise in these subbands is relatively high. At coarse scales the energy of the image increases but that of noise decreases rapidly. The histograms of $Z_j^x = W_j^x f \cdot W_{j+1}^x f$ will be close to those of $Z_j^x = W_j^x g \cdot W_{j+1}^x g$ step by step.

The standard deviation of V_j^d is[26]

$$\kappa_j = \sqrt{E[v_j^2]} = \sqrt{E[u_j^2 u_{j+1}^2]} = \sqrt{1 + 2\rho_{j+1,j}^2} \cdot \sigma_j \sigma_{j+1}. \tag{32}$$

In Table 2 we compute the values of probability

$$\Pr_j(c) = P\{v_j \le c \cdot \kappa_j\}, \tag{33}$$

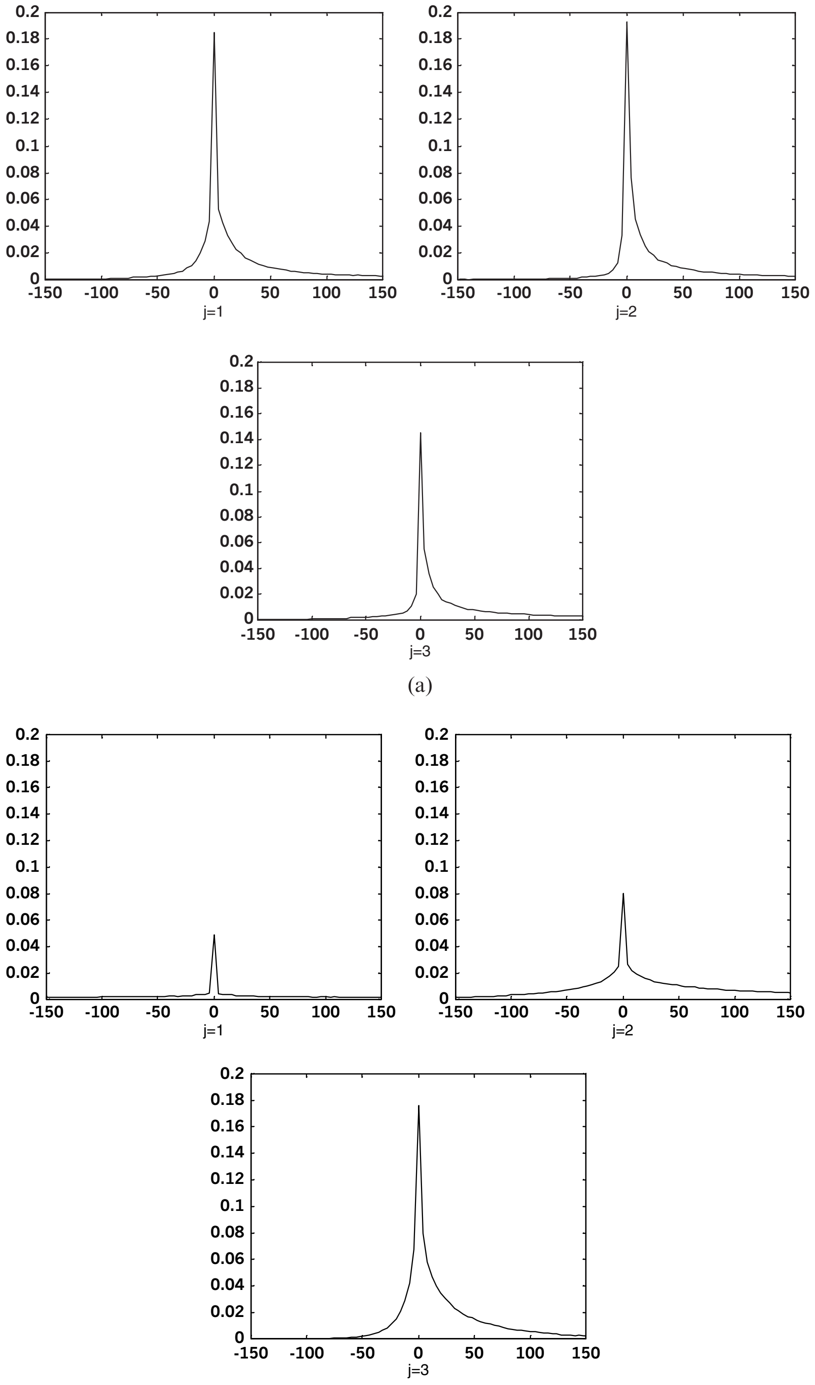
where constant c varies from 1 to 5 by step length 1. Notice that when $c \ge 5$, the probability $\Pr_j(c) \to 1$, implying that $5\kappa_j$ will suppress most of the data in V_j.

We denote

$$\mu_f^d(j) = \mathrm{E}[Z_j^d], \quad \mu_\varepsilon^d(j) = \mathrm{E}[V_j^d], \quad \mu_g^d(j) = \mathrm{E}[W_j^d g \cdot W_{j+1}^d g]. \tag{34}$$

Since the noise ε is independent of the noiseless image g, it can be derived that

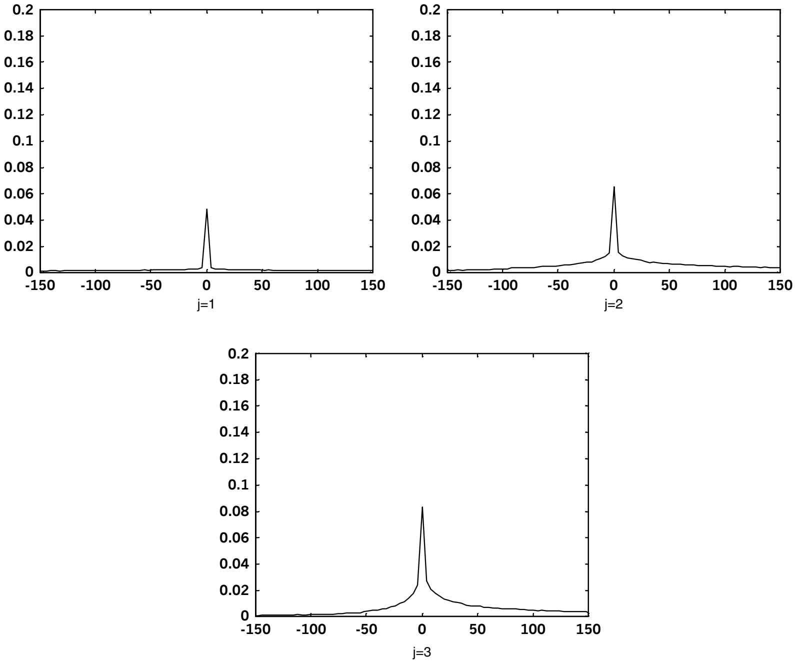
$$\mu_g^d(j) = \mu_f^d(j) - \mu_\varepsilon^d(j), \tag{35}$$



(a)

(b)

Fig. 6. The histograms of the multiscale products $P_j^x f$ at the first three scales when f is (a) Noiseless image *Lena*; (b) Gaussian white noise; (c) Noisy image *Lena*.



(c)

Fig. 6. (*Continued*)

Table 2. The values of probability $\mathrm{Pr}_j(c) = \mathrm{P}\{v_j \leq c \cdot \kappa_j\}$ for $j = 1$ and $j = 2$.

c	1	2	3	4	5
$\mathrm{Pr}_1(c)$	0.8445	0.9456	0.9799	0.9927	0.9976
$\mathrm{Pr}_2(c)$	0.8291	0.9396	0.9774	0.9915	0.9971

and

$$\mu_\varepsilon^d(j) = \rho_{j+1,j}\sigma_j\sigma_{j+1} . \tag{36}$$

The ratio $\mu_\varepsilon^d(j)/\mu_g^d(j)$ is a measurement for the intensity of noise against signal in the multiscale products Z_j^d. This ratio can be used to adjust the threshold $t_p^d(j)$ imposed on Z_j^d. We set the multiscale products threshold as

$$t_p^*(j) = 5\kappa_j\left(1 + \frac{\mu_\varepsilon^*(j)}{\mu_g^*(j)}\right) . \tag{37}$$

The adaptive threshold $t_p^d(j)$ is intuitive and effective. When noise is much stronger compared with the image (i.e. at fine scales), the ratio $\mu_\varepsilon^d(j)/\mu_g^d(j)$ is high. Therefore the threshold $t_p^d(j)$ becomes sufficiently large to suppress the overwhelming noise. When the image is dominative (i.e. at coarse scales or when additive

Table 3. The ratios of $\mu_\varepsilon^x(j)/\mu_g^x(j)$ at the first two scales for image Lena on different noise levels σ.

σ	10	20	30	40
$\mu_\varepsilon^x(1)/\mu_g^x(1)$	0.27	1.18	3.20	7.33
$\mu_\varepsilon^x(2)/\mu_g^x(2)$	0.02	0.10	0.22	0.40

noise is low), the ratio $\mu_\varepsilon^d(j)/\mu_g^d(j)$ is small and the threshold $t_p^d(j)$ is at an appropriate level to preserve the image instantaneous features while removing noise. In Table 3, we gave some values of $\mu_\varepsilon^d(j)/\mu_g^d(j)$ for image Lena with noise levels $\sigma = 10, 20, 30, 40$, respectively.

3.3. *Noise level estimation*

The standard deviation of additive Gaussian white noise, σ, should be estimated to implement the denoising scheme. A popular noise level estimator has been proposed by Donoho.[19] The Median Absolute Value (MAV) of the wavelet coefficients at the finest scale is first calculated and the standard deviation of noise is then estimated as MAV/0.6745. The MAV estimator is inaccurate for those images containing massive fine structures.

We propose a new noise level estimation method as follows. We compute the Orthogonal Wavelet Transform (OWT) of the noisy image at the finest scale and denote by W the wavelet coefficients in the diagonal direction. Because OWT is a unitary transform, at each wavelet scale the noise standard deviation is equal to σ. Thus the variance of W is

$$\sigma_f^2 = E[W^2] = \sigma_g^2 + \sigma^2, \tag{38}$$

where σ_g is the standard deviation of the wavelet coefficients of the noiseless image.

Suppose N is a zero mean Gaussian process with standard deviation σ_N, we divide it into two parts. The first part N_a consists of points that $|N(\cdot)| > \sigma_N$ and the second part N_b consists of points where $|N(\cdot)| \le \sigma_N$. Let $\sigma_a = \sqrt{E[N_a^2]}$ and $\sigma_b = \sqrt{E[N_b^2]}$, we have

$$\sigma_a^2 = 2\int_{\sigma_N}^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma_N(1-2erf(1))} e^{-x^2/(2\sigma_N^2)}dx, \tag{39}$$

$$\sigma_b^2 = \int_0^{\sigma_N} \frac{x^2}{\sqrt{2\pi}\sigma_N erf(1)} e^{-x^2/(2\sigma_N^2)}dx, \tag{40}$$

where $erf(t) = \frac{1}{\sqrt{2\pi}}\int_0^t e^{-x^2/2}\,dx$ is the error function. The ratio of σ_a to σ_b is independent of σ_N and we can calculate that $\sigma_a/\sigma_b \approx 2.945$.

Next we split W into two parts, W_a such that $|W_a(\cdot)| > \sigma_f$ and W_b such that $|W_b(\cdot)| \le \sigma_f$. Let $\sigma_w^a = \sqrt{E[W_a^2]}$ and $\sigma_w^b = \sqrt{E[W_b^2]}$. Generally the noise energy is concentrated on W_b and σ_w^b can be considered as an approximation of the noise level σ. We define

$$\hat{\sigma}_g = \sqrt{(\sigma_w^a)^2 - 2.9^2(\sigma_w^b)^2}. \tag{41}$$

If W is produced totally by noise, obviously $\hat{\sigma}_g$ will be equal to zero. The more image details involved in W_a, the greater the value of $\hat{\sigma}_g$. In this case $\hat{\sigma}_g$ can be seen as an approximated estimation of σ_g. We then take

$$r = \hat{\sigma}_g / \sigma_w^b \tag{42}$$

as a measure of signal-to-noise-ratio at the finest scale. Finally the noise level can be estimated as

$$\hat{\sigma} = \sigma_f \big/ \sqrt{1 + r^2}. \tag{43}$$

We denote the MAV noise estimator of Donoho by

$$\hat{\sigma}_d = \text{Median}(|W|)/0.6745. \tag{44}$$

The Monte Carlo experimental results using the two estimators are listed in Table 4. The test images employed are *Lena*, *Sailboat*, *Goldhill*, *Bridge* and *Baboon*. We added the Gaussian white noise with different standard deviation σ to each of them. The Daubechies wavelet[1] with four vanishing moments is used for the OWT. It can be observed that the proposed method generally outperforms Donoho's MAV estimation scheme. It performs especially well for the image *Baboon* that contains massive fine structures.

4. Experiments

In this section, the performances by the proposed scheme on some MRI images are compared with those of the *soft* thresholding scheme *BayesShrink* of Chang *et al.*[23] and the *hard* thresholding scheme of Pan *et al.*[22] For convenience, we refer the

Table 4. Noise level estimation results. *Gaussian* white noise with standard deviation σ is added to five test images. $\hat{\sigma}_d$ is the estimation by Donoho's median method and $\hat{\sigma}$ is the estimation by our scheme.

σ		5	10	15	20	25	30	35	40
Lena	$\hat{\sigma}_d$	6.15	11.02	15.87	20.82	25.68	30.56	35.54	40.52
	$\hat{\sigma}$	3.20	8.37	14.30	19.66	24.13	30.21	35.31	40.56
Sailboat	$\hat{\sigma}_d$	7.84	12.16	16.78	21.47	26.20	31.13	35.94	40.86
	$\hat{\sigma}$	4.06	8.78	14.57	19.88	25.14	30.71	35.83	40.80
Goldhill	$\hat{\sigma}_d$	7.49	11.79	16.46	21.18	25.94	30.84	35.69	40.67
	$\hat{\sigma}$	4.22	9.02	14.70	20.13	25.34	30.42	35.21	40.78
Bridge	$\hat{\sigma}_d$	8.46	12.73	17.17	21.84	26.54	31.29	36.19	40.98
	$\hat{\sigma}$	4.89	9.58	15.10	20.51	25.82	30.95	35.48	40.94
Baboon	$\hat{\sigma}_d$	14.58	18.17	22.29	26.65	30.93	35.34	39.90	44.51
	$\hat{\sigma}$	7.05	9.66	13.61	18.66	24.01	29.53	35.25	41.37

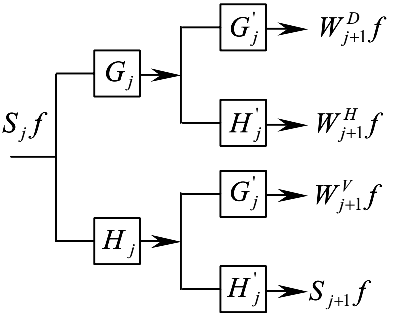


Fig. 7. One stage transform structure of the 2D Overcomplete Wavelet Expansion (OWE). Filter F_j is the 2^j scale dilation of F_0 (putting $2^j - 1$ zeros between each of the coefficients of F_0) and F'_j is the transpose of F_j. $W_j^H f$, $W_j^V f$ and $W_j^D f$ are the wavelet coefficients at the horizontal, vertical and diagonal directions.

two methods as STH and HTH, respectively. It is well observed that thresholding with the OWT produces unpleasant Gibbs-like edge artifacts.[21] Thus we implement the two schemes with the over-complete wavelet expansion (OWE). The one stage transform of OWE is illustrated in Fig. 7. The resultant denoising by thresholding with the OWE can be interpreted as the average of the circularly shifted denoising outcomes by the OWT. The residual noise is better smoothed and the artifacts are attenuated. The wavelet employed in the STH and HTH schemes is the compactly supported orthogonal wavelet of Daubechies with four vanishing moments.[1] The constant c appearing in the threshold of the scheme HTH is set at 3.1. The proposed scheme is referred as MPTH. The MRI images in our experiments are 512 × 512 in size and the decomposition level is 4.

To evaluate the medical image quality, we compute the Mean-to-Standard-deviation-Ratio (MSR)[14,15] in a Desired Region Of Interest (DROI):

$$\mathrm{MSR} = \frac{\mu_d}{\sigma_d}, \tag{45}$$

where μ_d and σ_d are the mean and the standard deviation computed in the DROI. The Contrast to Noise Ratio (CNR) is also an important quality measurement for medical image interpretation. It is defined as

$$\mathrm{CNR} = \frac{|\mu_d - \mu_u|}{\sqrt{0.5(\sigma_d^2 + \sigma_u^2)}}, \tag{46}$$

where μ_u and σ_u are the mean and the standard deviation computed in an Undesired Region Of Interest (UROI) such as a window or background. Both the MSR and CNR measurements are proportional to the medical image quality.

In the k-space, the raw MRI data is denoted by[8–10]

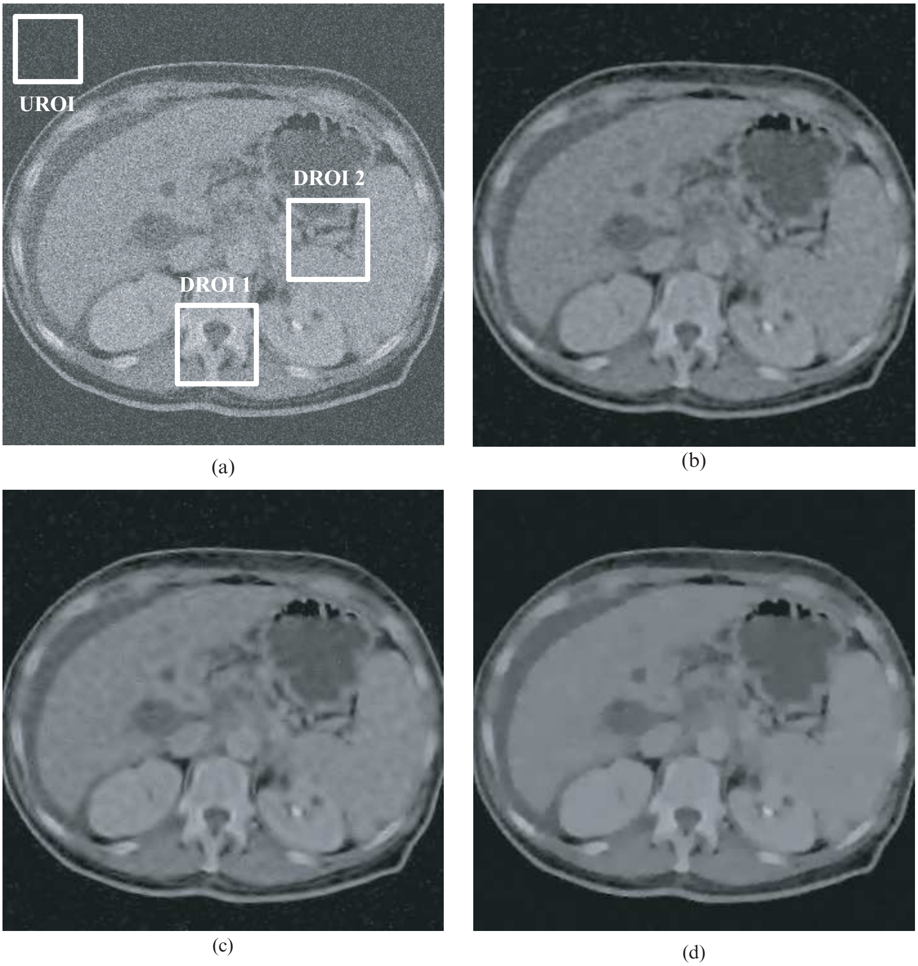
$$F(\mu, \nu) = G(\mu, \nu) + \xi(\mu, \nu), \tag{47}$$

where $G(\mu, \nu)$ is the underlying signal and $\xi(\mu, \nu)$ is a complex Gaussian white noise. By computing the Fourier transform (FT) of $F(\mu, \nu)$ in the complex image domain, we have

$$f(x, y) = g(x, y) + \varepsilon(x, y). \tag{48}$$

$\varepsilon(x, y)$ is again a complex Gaussian white noise due to the unitarity of the FT. The denoising schemes could be applied to each of the real and imaginary components of $f(x, y)$. For visual inspection, the moduli of the complex data $f(x, y)$ are shown as the magnitude image.

Figure 8(a) is a noisy MRI image *Liver*. The DROI and UROI used for calculating the MSR and CNR indexes are highlighted. Denoised images by the three schemes are illustrated in Figs. 8(b)–(d), respectively and the MSR and CNR values



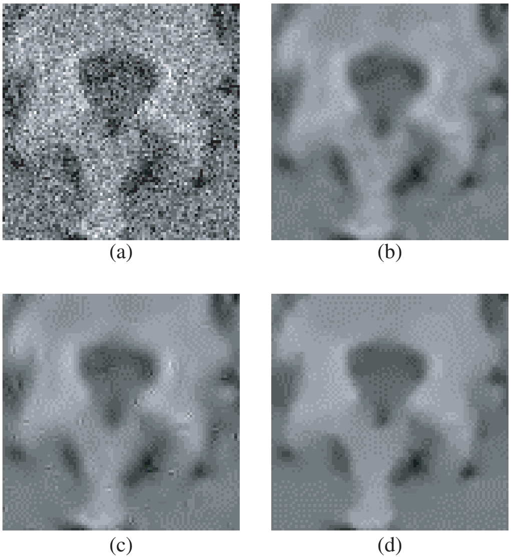
(a) (b) (c) (d)

Fig. 8. Experiments on MRI image *Liver*. The Desired Region Of Interest (DROI) and Undesired Region Of Interest (UROI) used to compute the MSR and CNR indexes (listed in Table 5) are highlighted. (a) The noisy image. (b) Estimated by the STH. (c) Estimated by the HTH. (d) Estimated by the presented MPTH.

Table 5. The MSR and CNR results of MRI images *Liver* by the three schemes.

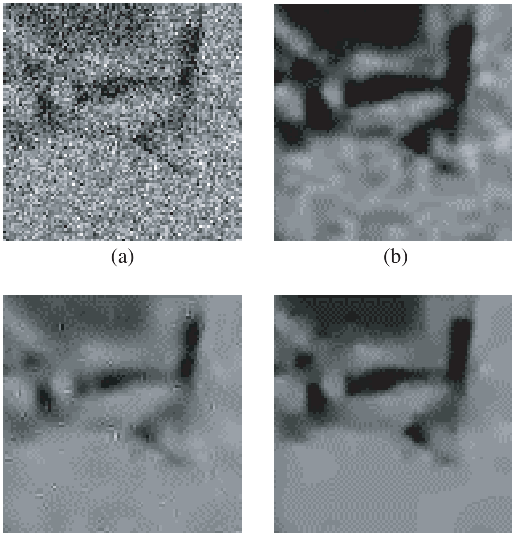
Method	DROI 1		DROI 2	
	CNR	MSR	CNR	CNR
Original	2.61	2.61	2.58	2.45
STH	4.08	5.53	4.55	6.12
HTH	4.10	5.58	4.60	6.20
MPTH	4.31	5.98	4.85	6.72

are listed in Table 5. The presented algorithm MPTH achieves the highest quantity measurements. Notice that the denoised image by the STH contains a few stains and the result by the HTH retains much noise. (If the threshold of the HTH is set higher to suppress noise, the estimated image would be over-smoothed.) The MPTH preserves edges better and yet effectively removes noise. Zoom-in images of DROI 1 and DROI 2 are illustrated in Figs. 9 and 10. Another experiment on an MRI image *Spine* is illustrated in Figs. 11–13. The MSR and CNR measurements are listed in Table 6. The results showed in Figs. 11(b) and (c) appear to be veiled by the residual noise. Zoom-in images of the two DROI are illustrated in Figs. 12 and 13. Although some *stings* (discontinuities) appeared in Figs. 12(d) and 13(d), they are almost edge points detected from the multiscale products, but discarded by the other two schemes.


(a) (b)

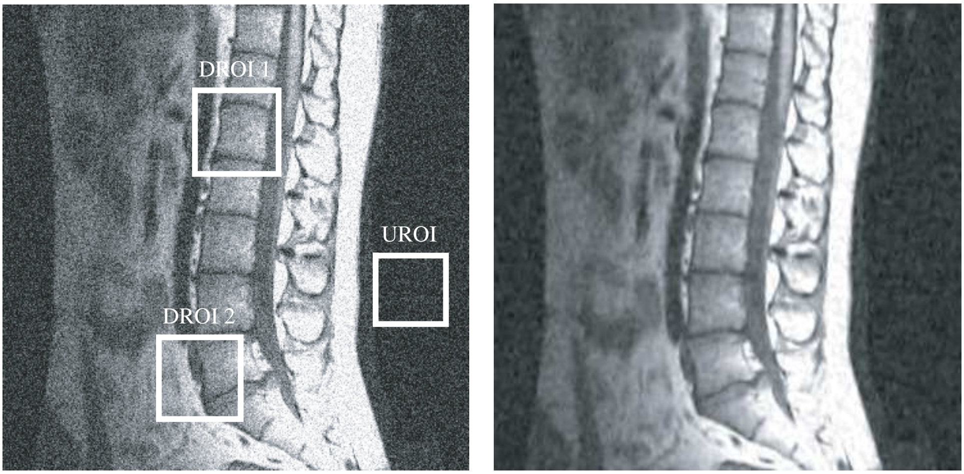
(c) (d)

Fig. 9. Zoom in of the DROI 1. (a) The noisy image. (b) Estimated by the STH. (c) Estimated by the HTH. (d) Estimated by the presented MPTH.



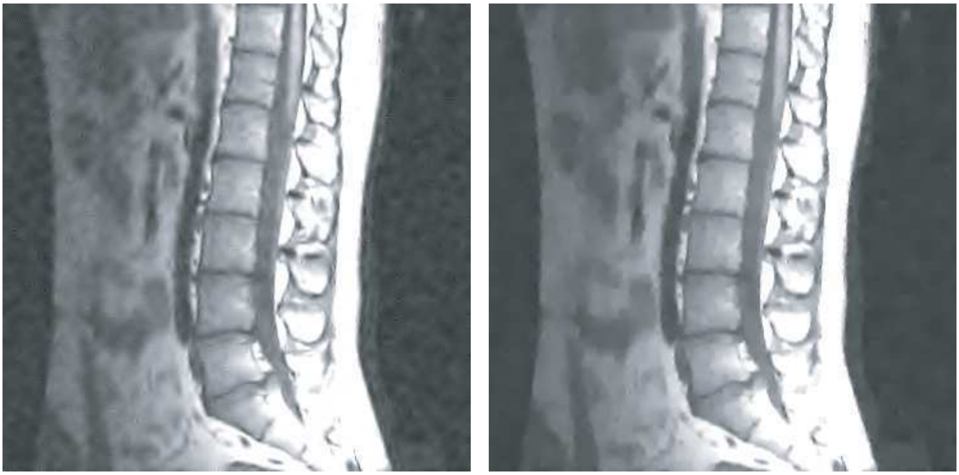
(a) (b) (c) (d)

Fig. 10. Zoom in of the DROI 2. (a) The noisy image. (b) Estimated by the STH. (c) Estimated by the HTH. (d) Estimated by the presented MPTH.



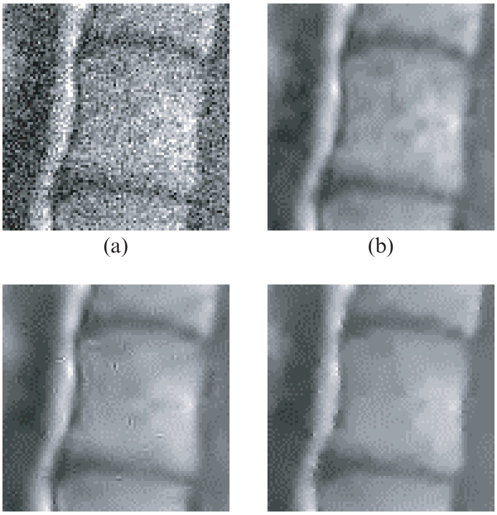
(a) (b)

Fig. 11. Experiments on MRI image *Spine*. The Desired Region Of Interest (DROI) and Undesired Region Of Interest (UROI) used to compute the MSR and CNR indexes (listed in Table 6) are highlighted. (a) The noisy image. (b) Estimated by the STH. (c) Estimated by the HTH. (d) Estimated by the presented MPTH.



(c) (d)

Fig. 11. (*Continued*)



(a) (b)

(c) (d)

Fig. 12. Zoom in of the DROI 1. (a) The noisy image. (b) Estimated by the STH. (c) Estimated by the HTH. (d) Estimated by the presented MPTH.

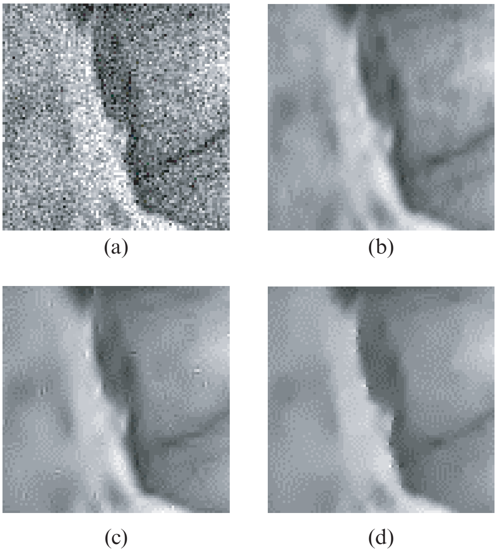


Fig. 13. Zoom in of the DROI 2. (a) The noisy image. (b) Estimated by the STH. (c) Estimated by the HTH. (d) Estimated by the presented MPTH.

Table 6. The MSR and CNR results of MRI images *Spine* by the three schemes.

Method	DROI 1		DROI 2	
	CNR	MSR	CNR	CNR
Original	2.26	1.79	2.82	2.34
STH	2.96	2.75	3.99	3.86
HTH	2.98	2.78	4.02	3.91
MPTH	3.10	2.91	4.17	4.06

5. Conclusion

This chapter describes an MRI image denoising scheme using an adaptive wavelet thresholding technique. Unlike many traditional schemes that directly threshold the wavelet coefficients, the proposed scheme multiplies the adjacent wavelet subbands to amplify the significant features and then applies the thresholding to the multiscale products to better differentiate edge structures from noise. The distribution of the products was analyzed and an adaptive threshold was formulated with a reasonable value to remove most of the noise. Experiments on the MRI images show that the proposed scheme not only achieves high MSR and CNR measurements but also preserves more edge features.

References

1. I. Daubechies, *Ten Lectures on Wavelets* (Philadelphia, PA: SIAM, 1992).
2. Y. Meyer, *Wavelets and Operators* (Cambridge, Cambridge University Press, 1992).

3. S. Mallat and W. L. Hwang, Singularity detection and processing with wavelets, *IEEE Trans. on Information Theory* **32** (1992) 617–643.
4. S. Mallat and S. Zhong, Characterization of signals from multiscale edges, *IEEE Trans. Pattern Analysis and Machine Intelligence* **14** (1992) 710–732.
5. M. Vetterli and C. Herley, Wavelet and filter banks: Theory and design, *IEEE Trans. Signal Processing* **40** (1992) 2207–2232.
6. J. M. Shapiro, Embedded image coding using zerostrees of wavelet coefficients, *IEEE Trans. Signal Processing* **41**, 12 (1993) 3445–3462.
7. M. S. Crouse, R. D. Nowak and R. G. Baraniuk, Wavelet-based statistical signal processing using hidden Markov models, *IEEE Trans. Signal Processing* **46** (1998) 886–902.
8. R. D. Nowak, Wavelet-based Rician noise removal for magnetic resonance, *IEEE Trans. Image Processing* **8** (1999) 1408–1419.
9. A. Macovski, Noise in MRI, *Magn. Reson. Med.* **36** (1996) 494–497.
10. H. Gudbjartsson and S. Patz, The Rician distribution of noisy MRI data, *Magn. Reson. Med.* **34** (1995) 910–914.
11. J. C. Wood and K. M. Johnson, Wavelet packet denoising of magnetic resonance images: Importance of Rician noise at low SNR, *Magn. Reson. Med.* **41** (1999) 631–635.
12. S. Zaroubi and G. Goelman, Complex denoising of MR data via wavelet analysis: Application for functional MRI, *Magnetic Resonance Imaging* **18** (2000) 59–68.
13. M. E. Alexander, A wavelet-based method for improving signal-to-noise ratio and contrast in MR images, *Magnetic Resonance Imaging* **18** (2000) 169–180.
14. H. Soltanian-Zadeh, J. P. Windham and A. E. Yagle, A multidimensional nonlinear edge-preserving filter for magnetic resonance image restoration, *IEEE Trans. Image Processing* **4** (1995) 147–161.
15. G. Cincotti, G. Loi and M. Pappalardo, Frequency decomposition and compounding of ultrasound medical images with wavelet packets, *IEEE Trans. Medical Imaging* **20** (2001) 764–771.
16. A. Pizurica, W. Philips, I. Lemahieu and M. Acheroy, A versatile wavelet domain noise filtration technique for medical imaging, *IEEE Trans. Medical Imaging* (accepted).
17. J. Canny, A computational approach to edge detection, *IEEE Trans. PAMI* **PAMI-8** (1986) 679–697.
18. Y. Xu, J. B. Weaver, D. M. Healy, Jr. and J. Lu, Wavelet transform domain filters: A spatially selective noise filtration technique, *IEEE Trans. Image Processing* **3** (1994) 747–758.
19. D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, *Biometrika* **81** (1994) 425–455.
20. D. L. Donoho and I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *J. of the American Statistical Assoc.* **90** (1995) 1200–1224.
21. R. R. Coifman and D. L. Donoho, Translation-invariant denoising, in *Wavelet and Statistics*, A. Antoniadis and G. Oppenheim (Eds.), (Berlin, Germany: Springer-Verlag, 1995).
22. Q. Pan and L. Zhang, Two denoising methods by wavelet transform, *IEEE Trans. Signal Processing* **47** (1999) 3401–3406.
23. S. G. Chang and B. Yu, Adaptive wavelet thresholding for image denoising and compression, *IEEE Trans. Image Processing* **9** (2000) 1532–1546.
24. B. M. Sadler and A. Swami, Analysis of multiscale products for step detection and estimation, *IEEE Trans. Information Theory* **45** (1999) 1043–1051.
25. K. S. Miller, *Multidimensional Gaussian Distributions* (New York: Wiley, 1964).

26. J. K. Patel and C. B. Read, *Handbook of The Normal Distribution* (Marcel Dekker Inc., New York, 1982).
27. H. Zhong and M. Chong, A wavelet-denoising approach for removing background noise in medical images, *Information, Proceedings of 1997 International Conference on Communications and Signal Processing*, pp. 980–983.
28. J. C. Wood and K. Johnson, Wavelet-denoising of complex magnetic resonance images, *Proceedings of the IEEE 24th Annual Bioengineering Conference*, 1998, Northeast, pp. 32–34.
29. J. Lin, A. F. Laine and S. R. Bergmann, Improving PET-based physiological quantification through methods of wavelet denoising, *IEEE Transactions on Biomedical Engineering* **48** (2001) 202–212.
30. B. K. Ouda, B. M. Tawfik, A. Youssef and Y. M. Kadah, Adaptive denoising technique for robust analysis of functional magnetic resonance imaging data, *Proceedings of the 23rd Annual International Conference of the IEEE* **3** (2001) 2476–2479.
31. Z. Q. Wu, J. A. Ware and J. Jiang, Wavelet-based Rayleigh background removal in MRI, *Electronics Letters* **39** (2003) 603–604.
32. M. Lysaker, A. Lundervold and X. C. Tai, Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time, *IEEE Transactions on Image Processing* **12** (2003) 1579–1590.

CHAPTER 5

TECHNIQUES IN X-RAY COMPUTED TOMOGRAPHY IN THE EVALUATION OF DRUG RELEASE SYSTEMS AND THEIR APPLICATION

AGATA A. EXNER*, JINMING GAO and DAVID L. WILSON
Departments of Radiology and Biomedical Engineering, Case Western Reserve University
11100 Euclid Avenue, Cleveland, OH 44106-5056, USA
**agata.exner@cwru.edu*

Advancement of site-specific drug delivery systems has been hampered by the shortage of direct techniques for sampling and analysis of drug concentrations at the site of action. Because the potential target is often deeply imbedded within an organ, the local concentration of a drug delivered via implantable devices and its movement in tissue and vasculature (regarded as one significant aspect of pharmacokinetics) are often difficult to obtain with traditional methods like plasma and urine analysis, and alternative methods are highly sought after. Imaging is a natural candidate for this application. Thanks to modern technology, clinical imaging modalities with high spatial and temporal resolution combined with exceptional sensitivity and ease of use have greatly impacted the current diagnosis and treatment of diseases. Although all types of modalities are excellent candidates for pharmacokinetic imaging, X-ray Computed Tomography (CT) possesses a unique combination of speed, resolution and inherent simplicity in quantitative analysis that makes it an attractive choice. This chapter provides a general overview of CT methods in pharmacokinetic imaging with a specific application to the development and evaluation of a drug release system for local chemotherapy of liver tumors.

Keywords: X-ray computed tomography; site-specific drug release system; pharmacokinetic imaging.

1. Introduction

Direct visualization of physiological events has only become feasible within the past 50 years with the advent of functional imaging technology. Techniques that previously allowed us only to take a snapshot of the anatomy have evolved into incredible devices capable of providing an often quantitative glimpse into the actively changing inner workings of our bodies. Tumor metabolism, blood flow, memory and thought are only a few of the numerous processes that have been explored within the field of functional imaging. Recently, the same techniques typically used to monitor the baseline physiology and pathology have also been applied to the research and development of new drugs and Drug Delivery Systems (DDS). Most commonly utilized modalities are PET and SPECT with some work being done with MRI, ultrasound and CT, although a surprisingly small fraction has been devoted to the latter. Our team has focused on exploration of CT in evaluation of drug release systems and their efficacy in the treatment of disease, in particular in treatment of cancer. The following pages review some basic principles of functional CT imaging and how it can be utilized in the field of drug delivery.

Sections of this chapter are dedicated to describing the development of CT imaging protocols, optimization of image analysis methods, *in vitro/in vivo* validation of the image data, and functional application of the concepts to a clinically applicable problem. The final section discusses encountered and potential problems and offers perspectives for the future of CT in drug development.

1.1. *Minimally invasive treatment*

The concepts of site-specific drug delivery can be applied to address a multitude of clinical problems. One of the most difficult undertakings is the development of local chemotherapy for treatment of malignancies, the successful treatment and cure of which has been the driving force for medical research for hundreds of years. Despite the wide range of options available today (molecular-targeting agents, minimally invasive tumor ablation, and targeted radiation delivered through a radioactive seed, or the gamma knife, to name just a few), chemotherapy remains the staple, yet the most systemically debilitating, approach. The stress from systemic chemotherapy is so great that, in fact, some patients refuse to or cannot undergo treatment.

In the last two decades, cancer therapy has taken a minimally invasive approach, where the stress on the body is lessened but the outcome is equal to or superior to the widely accepted surgical procedures. This treatment is most useful for cancers that exhibit solid, localized tumors — ones that have not spread, or metastasized, throughout the body. Included in this group are: (i) chemoembolization, where a chemical mixed with a gelling agent is injected directly into the tumor blood supply to block it; (ii) ablation, where the tumor is either killed by local application of heat or cold; and (iii) local chemotherapy, where an anticancer agent is placed directly into the tumor and released immediately or over time to destroy the cancer cells. Because of their pertinence to the rest of this work, only thermal ablation and local chemotherapy will be discussed in detail.

1.1.1. *Radiofrequency ablation in treatment of liver cancer*

Image-guided RadioFrequency Ablation (RFA) has emerged as the cutting-edge treatment of liver, pancreas, prostate and even lung cancer.[1–10] The main goal of RFA is the induction of thermal coagulative necrosis in the tumor volume to kill malignant cells while leaving the adjacent healthy tissue unharmed. To carry out the procedure, a needle electrode is inserted under image guidance directly into the center of the tumor. To complete the circuit the electrode is connected with an RF generator, and a large reference electrode pad is placed on the outside of the body. The generator produces a voltage between the two electrodes causing ions in tissue to oscillate, and the area is heated by a resistive energy loss.[11–13] The temperature at which the tissue undergoes cellular destruction ranges between 60–100°C. At this range, the proteins in a cell coagulate, and damage occurs to cytosolic and mitochondrial enzymes and nucleic acid-histone protein complexes. This irreversible damage leads to coagulation necrosis within several days. At temperatures exceeding 105°C, tissue charring and vaporization occur and limit the transfer of energy necessary for a successful ablation.[11–13]

Multiple studies have examined the efficacy of RF ablation in the treatment of HepatoCellular Carcinomas (HCC).[13–21] Rossi *et al.* reported on the ablation of HCC in 23 patients with HCC up to 3.5 cm in diameter. A complete response was seen in all tumors by this group. Livraghi *et al.* compared Percutaneous Ethanol Injection (PEI) to RF ablation in 86 patients. Here, a complete response was seen in 90% of patients undergoing RFA compared to 80% treated with PEI. Lencioni *et al.* studied 80 patients, out of which 87% exhibited long-term tumor control. Livraghi *et al.* also studied 114 patients with medium (3–5 cm) and large (5.1–9 cm) HCC. The results show complete necrosis attained in 47.6% and nearly complete necrosis in 31.7% of the tumors. Solbiati *et al.* report on the RF ablation of 109 patients with 172 hepatic metastases (< 4 cm) from colorectal cancer.[17,18] Local control was achieved in 70.4% of the lesions, with local recurrence occurring in the remaining 29.6%. New metastases developed in 50.4% of the patients after 12 months. The authors report 2 and 3-year survival rates to be 67% and 33% respectively, and a zero mortality rate resulting from the procedure.[17,18] Curley *et al.* report on treatment and 19-month follow up of 110 patients with cirrhosis and unresectable HCC.[19] They report local tumor recurrence at the ablation site of 3.6% of the cases, and a new tumor growth or extrahepatic metastasis in 45.5% of the patients.[20]

The limiting factor in the success of RFA is the size of the ablation volume. The monopolar electrode configuration yields an ablation area of no more than 2 cm in diameter, but electrode configuration alterations have lead to an increase of the volume upwards to 5.3 cm with a cluster electrode configuration.[11–13] Several physiological factors limit the efficacy of thermal ablation *in vivo*. The presence of extensive vasculature in the vicinity of the ablation site leads to perfusion mediated tissue cooling, which is by far the most significant reason for incomplete ablation. To assure complete ablation, a margin of healthy tissue surrounding the tumor (~ 0.5 cm) may be ablated.[11–13] An alternative strategy may be the local release of chemotherapeutic agents to supplement the RFA treatment and eliminate any uncertainty regarding completeness of the procedure.

1.2. *Site-specific drug release systems*

Site-specific drug delivery is a powerful technique that has the potential to become a mainstream alternative to conventional systemic chemotherapy. This mode of drug administration can elevate drug concentration at the target to a concentration significantly higher than is possible with intravenous injection or oral delivery site, while limiting systemic effects of toxic drugs; intratumoral drug delivery is one example of this approach. Advantages of site-specific release include reduced drug dose, sustained delivery of drug within the therapeutic window, protection of drugs with a narrow therapeutic index or short half-life, increased comfort, improved patient compliance, and a decreased number of treatments.

Although superficially a simple concept, advanced development of site-specific delivery systems has not progressed as quickly as expected due to a number of reasons, among which lack of accurate local pharmacokinetic analysis techniques is a primary factor. The goal of pharmacokinetics, the study of drug absorption, distribution, metabolism and excretion of a drug, is to determine the changes in drug concentration at the site of action with time. Because conventional analysis of drug concentration in plasma and urine does not directly reflect drug concentration at the site of action in a local therapy scheme, a more direct visualization and analysis method is desirable in order to accurately characterize the pertinent properties of any such device.

1.3. *Non-invasive pharmacokinetic imaging*

Understanding the clinical PharmacoKinetics (PK) of drugs in tissue is essential for the design of a successful local drug delivery device. In order for a local therapy to attain maximum efficacy, it must deliver the therapeutic dose of a drug to a nearby target area without substantially affecting the normal tissues. The drug released from the implant must travel a significant distance and reach the site of action in its active state, and its ability to do so will depend highly on the properties of the drug as well as on the property of the tissue. Currently, the local drug release and transport properties in tissues are not known but need to be quantified for a successful development of a local drug delivery system. This is true especially in complex tissue environments such as thermally ablated tissue. Since the viable and ablated tissue structure varies drastically, with the ablated region lacking viable cells and vasculature, we predict that this difference will be imposed on the local drug release profile and drug kinetics.

Non-invasive pharmacokinetic imaging is particularly attractive in cases where *in vitro* (or test-tube) dissolution characterization of a drug delivery device does not correspond well to the *in vivo* system, such as in the case of tumor tissue after RadioFrequency Ablation (RFA) treatment, where the physiology has been severely altered from the baseline of normal tissue. A number of studies have shown that the vasculature of the ablated tissue is completely destroyed, leading to a low-perfusion environment that can have a significant impact on drug delivery to the area.[22–26] Because one of our initiatives is the development of a local chemotherapy that can be used in conjunction with RFA to control any local tumor recurrence after this treatment, the data gathered with this pharmacokinetic imaging technique will be invaluable in future development of this approach. The combination of CT imaging and quantitative image analysis can potentially provide a direct method for the evaluation of the transport of drugs released from a localized, controlled release device *in vivo*.

Conventional clinical pharmacokinetic studies require the collection of blood, urine or tissue samples to determine the tissue distribution, clearance rate, and even efficacy of the treatment. These procedures are invasive and therefore are not compliant with patients, especially when multiple sampling is called for. Furthermore, quantitative interpretation of this data can be complicated by issues ranging from inter-individual variability to uncertainty in the technical procedure, such as incomplete extraction of drugs from tissue.

Because it addresses many of the issues associated with development of drug release systems, non-invasive pharmacokinetic imaging is quickly becoming a vital tool to facilitate the development of new drugs and drug delivery methods. Diagnostic imaging techniques such as Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), and Magnetic Resonance Imaging (MRI) have been used extensively to study drug distribution and metabolism in patients with various types of cancer.[27–31] However, limitations such as low spatial resolution, radioactive half-life restrictions, long image acquisition time, and the need for radiopharmaceutical production, restrict their applications in pharmacokinetic monitoring.[28,29]

1.3.1. *Notable pharmacokinetic parameters*

Local and regional drug delivery presents a unique challenge for the study of drug PharmacoKinetics (PK) to characterize the drug concentration at its site of action. Traditionally, measurement of drug concentration at the site of action is not practical and, alternatively, the drug concentration in plasma and urine — two easily tested bodily fluids — are used to describe the clinical pharmacokinetics. A result of years of work, PK models have been created to relate fluid drug levels to rates of absorption, clearance and ultimately tissue distribution of the drug. In local drug delivery, however, the traditional PK models do not apply. The nature of the local therapy, an “inside-out” process, is such that the levels of drug in the fluids no longer adequately describe the movement of the drug in the body. Because of this, study of local pharmacokinetics becomes necessary. Here, the crucial parameter is the concentration of drug in the tissue. Other parameters that need to be studied are the maximum drug penetration distance which represents the depth of tissue penetration of a predetermined therapeutic drug concentration, the change of concentration with distance from implant with time, and the change in concentration with time at a particular distance. The Area Under the Concentration-time curve (AUC) measurements can also be made at the site of action (the radius at which the drug is needed). Importantly, the methods of directly obtaining these measurements are limited to either tissue biopsies or non-invasive monitoring.

1.3.2. *Available imaging technology*

To circumvent limitations associated with standard PK studies, significant research efforts have been focused on the development of non-invasive imaging techniques to characterize the drug movement in the target tissue *in vivo*.[28] Such imaging techniques include gamma-emission imaging (single photon),[29] Positron-Emission Tomography (PET, dual photons),[29] and Magnetic Resonance Imaging (MRI).[32] Gamma imaging requires the use of radionuclides that emit high-energy gamma irradiation (typically 60–600 keV) while PET detects positron annihilation radiation (511 keV). These two methods provide the most sensitive measurement of radiolabeled drugs at their therapeutic concentrations. The drawback of these two methods is their low spatial resolution — for example, the clinical gamma imaging and PET instruments only provide resolutions at 5 and 10 mm, respectively.[31] This low resolution is not adequate to provide an accurate assessment of the therapeutic margin of drugs at the tumor tissue. Magnetic Resonance Imaging (MRI) is another non-invasive technique to study pharmacokinetics.[34] The MRI detects the magnetization changes of nuclei that have net angular momentum such as ^{1}H, ^{13}C and ^{19}F. Examples of drugs studied by MRI include 5-fluorouracil (^{19}F), ^{13}C-labelled temozolomide and ifosfamide (^{31}P).[32]

Although clinical MRI instruments can achieve better spatial resolution (2 mm) than PET, its main limitation is the low sensitivity that makes it difficult to detect drugs at a therapeutic level.[32] In the proposed research, we will introduce computed tomography as a new imaging modality to study the pharmacokinetics of platinum-containing drugs *in vivo*. Compared to other non-invasive imaging techniques, CT has the advantage to combine superb spatial resolution with high sensitivity without the necessity to radiolabel selected drugs.

1.3.3. *X-ray CT in pharmacokinetic imaging*

The CT instrumentation has evolved at an astounding pace to revolutionize the diagnosis and treatment of cancer. Technological advancement has greatly improved the speed, and accuracy of CT in medical applications. For example, the Phillips Mx8000 system, the CT scanner described in this research, can achieve high temporal resolution (0.5 secs/rotation), spatial resolution ($0.3 \times 0.3 \times 0.5\,\text{mm}^3$) and sensitivity. Spatial resolution and sensitivity are especially important for pharmacokinetic studies as described in this application.

The physical basis of CT contrast originates from the different abilities of materials to attenuate the X-ray photons in a beam of radiation.[33–35] There are three mechanisms that cause the attenuation of X-ray photons: Raleigh scattering, Compton scattering and photoelectric absorption. In typical CT analysis, the maximum energy of the X-ray beam (kVp) is set at 120 keV and the effective energy is approximately 80 keV for the Phillips scanner. At this energy, Raleigh scattering has relatively little contribution to the attenuation of the X-rays. Instead, the attenuation of primary photons in soft tissues such as liver (composed mostly of C, H, N, O elements) is primarily due to Compton scattering. This interaction occurs between X-ray photons and outer shell electrons, in which the electron is ejected from the atom and the primary photon is scattered with reduction in energy. In comparison, for CT contrast agent (e.g. iohexol) or platinum-containing drugs, the attenuation is mostly caused by photoelectric absorption of primary photons by the heavy elements such as iodine (atomic number $Z = 53$) or platinum ($Z = 78$). In photoelectric absorption, the photon is completely absorbed and an inner shell electron (photoelectron) is ejected from the atom. The CT contrast of different materials can be characterized and quantified by their linear attenuation coefficients (μ, unit: cm^{-1}) or mass attenuation coefficients (μ_m, cm^2/g). The latter is obtained by normalizing the value of μ for density ($\mathrm{g/cm}^3$) to correct the influence of material density on X-ray attenuation. At 80 keV, the values of μ_m for carbon, iodine and platinum are 0.16, 3.51 and 8.73 cm^2/g, respectively.[36] The difference in attenuation coefficients provides the physical basis of CT contrast for platinum-containing drugs over background liver tissue.

CT has several advantages for pharmacokinetic studies. It allows non-invasive, high resolution, fast, accurate, repeated measurements on a single animal. Because the animal can be used as its own control, inter-animal variability is eliminated. With proper testing and validation, CT can potentially be used without additional manipulation to study the *in vitro* and, more importantly, *in vivo* properties of local drug release systems and their behavior in a target tissue environment.

1.4. *Chapter overview*

The initial direction of the CT development focused on the establishment of the image acquisition and image analysis techniques. First, the image acquisition parameters (e.g. voltage, current, slice thickness, rotation time) of the CT scanner were evaluated and optimized to achieve the best sensitivity and resolution, in essence to maximize the signal to noise ratio, for the detection of the active agent in tissue. The sensitivity limit for iohexol (a CT contrast agent used in some preliminary experiments) and carboplatin was determined. Second, an *in vitro* model consisting of gelatin gel as a tissue mimic and iohexol as a drug mimic was used to validate the data obtained by CT analysis. Third, the same validation was carried out in a rabbit model to test verify the procedure in a living, breathing, animal. Fourth, the procedure was applied in an animal model to non-invasively examine the release of iohexol from the implants and the effect of RF ablation on this process. The optimization and *in vitro* — *in vivo* validation are described in Sec. 2, and the *in vivo* evaluation of drug release is described in Sec. 3. Finally, the technique was applied to determine the tissue penetration and distribution of carboplatin, a platinum-containing anticancer agent, in normal liver and livers affected by radiofrequency ablation. Two schemes of image analysis were applied to the data, and the image-assisted findings were correlated with typical chemical evaluation of tissue platinum levels. This is described in Sec. 4.

2. Optimization and *In Vitro/In Vivo* Validation of CT in Imaging of Drug Release Systems

Initial excitement about the potential of using computed tomography as a tool in the development on evaluation of local drug release systems originated due to the inherently simple linear relationship between X-ray attenuation and concentration of any radiopaque atom. By exploiting this relationship, CT can provide us with the desired pharmacokinetic parameters such as concentration of drug at the target site, local tissue permeation and diffusion into tissue, and information such as the rate of drug release from implantable devices. In the simplest sense, analyzing the straightforward changes in contrast density with time in a Region Of Interest (ROI) should provide us with quantitative, direct data that cannot be obtained by any other experimental means. No longer does the drug concentration at the target site have to be inferred from plasma and urine concentrations. Imaging can provide us with direct measurement of this parameter.

Monitoring the principal characteristics of a drug release system under unperturbed physiological conditions is important for the rational design and development of a device with optimal drug dosage, release rate and duration. These parameters are essential for achieving a safe and sufficient local therapy in a specific tissue environment. In this section we summarize the development and application of the CT method in examining the release kinetics of an agent from a polymer implant into rabbit livers. The image acquisition technique, imaging parameters to maximize the sensitivity and resolution, and image analysis procedures for quantitative measurement are established, and the method is validated in a gelatin model system as well as in rabbit livers *in vivo.*

2.1. *Optimization of CT image acquisition*

Gelatin gels containing iohexol concentrations from 0.01 to 10 mg/ml were used as imaging phantoms to evaluate the sensitivity of detection under various image acquisition conditions. The influence of the CT parameters such as peak kilo-voltage of the X-ray beam (kVp), current-time (mAs), and slice thickness were investigated. These parameters were examined systematically, with each set of data being analyzed and optimized before proceeding to the next study. The other parameters were kept constant: Helical scan at 0.5 mm pitch, 1.5 sec rotation time, B filter, high resolution, 160 mm Field Of View (FOV) and 512 × 512 pixel matrix.

For the optimization studies, the middle 4 slices, (or the middle 2 in the 2.5 mm case) were analyzed in order to account for the partial volume effect, where the data from the top and bottom edges of the plate may be averaged with the air, and incorrect pixel values would be obtained. The analysis consisted of selecting a circular Region Of Interest (ROI) of 120–125 pixels and determining the values along the perimeter. The average and standard deviation of the CT intensity in Hounsfield Units (HU) of these pixels were calculated. Next, the average CT intensity was plotted versus iohexol concentration and the slope of the line was determined. This slope was then used to convert the standard deviation in

Hounsfield Units (σ_{HU}) to standard deviation in concentration (σ_{C}). Then, the sensitivity was calculated by taking the ratio of the noise to the concentration (σ_{C}/C). The value of σ_{C}/C was then plotted versus the iohexol concentration and a power curve ($y = ax^b$) was fitted to each set of data to estimate the sensitivity cutoff. We defined the concentration at which the signal is equal to noise ($\sigma_C/C = 1$) as the sensitivity limit of iohexol detection by CT for each set of parameters.

Table 1 summarizes the results from the CT optimization and sensitivity studies. We found that increasing the current-time, peak kilo-voltage, and/or the slice thickness increases the sensitivity and improves the limit of iohexol detection. Analysis of the gelatin phantoms indicates that the sensitivity ranges from 0.18 to 0.30 mg/ml iohexol for the evaluated parameters. The sensitivity curves for these two extremes are shown in Fig. 1. The lowest concentration of iohexol that exceeds the quantum noise is 0.18 mg/ml. However, because the slice thickness needs to be 2.5 mm to achieve this limit, this set of parameters compromised the resolution for our application and will not be considered. The lowest sensitivity attained with a practical set of parameters is 0.21 mg iohexol/ml at 600 mAs, 120 kVp, and 1 mm slice thickness. These parameters were chosen as the most favorable image acquisition conditions and were used in all subsequent studies.

The sensitivity limit of CT is highly dependent on image noise and can be optimized by maximizing the signal to noise ratio. The image acquisition parameters that play a crucial role in this effect are the mAs, kVp, slice thickness and the reconstruction algorithm. The mAs is directly related to the number of photons emitted in an X-ray beam, and therefore it is inversely correlated to the square root of the noise of an image. The kVp affects the number of photons as well as the energy of the X-ray beam. Higher energy X-rays penetrate an object to a greater degree, while lower energy X-rays are attenuated more.[34,35] Increasing the energy of the beam, thus, increases the amount of signal reaching the detector, and thus increases the sensitivity. The slice thickness, which is affected by beam collimation, is also an important factor in the CT signal. Narrowing the collimation reduces the voxel size, thus decreasing the number of photons per voxel. This leads to an increase in quantum noise. In our application, it is important to achieve a

Table 1. Dependence of CT detection limit on image acquisition parameters. (adapted from A. Szymanski-Exner *et al.*, *J. Pharm. Sci., 2003*)

mAs	kVp	Slice (mm)	Detection limit (mg/ml)
600	120	1	0.21
400	120	1	0.23
200	120	1	0.30
400	140	1	0.24
400	90	1	0.22
600	120	2.5	0.18

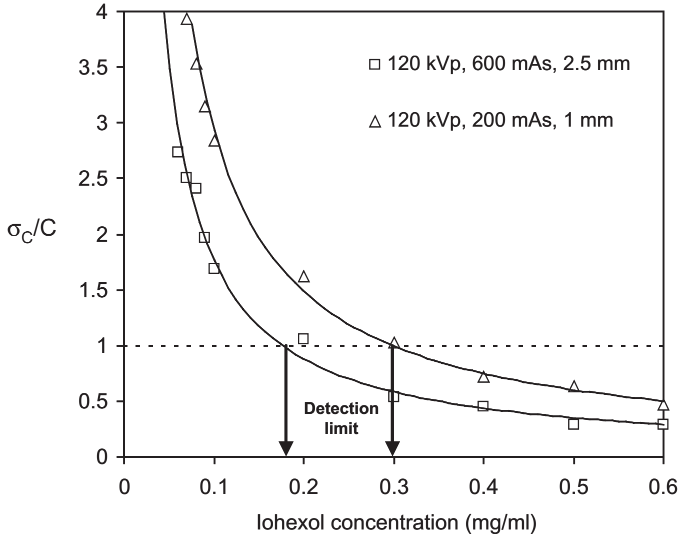


Fig. 1. Representative sensitivity plots. Shown are plots from two image acquisition conditions exhibiting high (□) and low (△) sensitivity with cutoff values at 0.18 and 0.30 mg/ml iohexol, respectively (adapted from A. Szymanski-Exner *et al.*, *J. Pharm. Sci., 2003*).

maximum sensitivity limit without severely compromising resolution. Because a trade-off exists between the two, the optimization studies were a necessary step to determine the optimal parameter combination (e.g. 600 mAs, 120 kVp, 1 mm, as shown in Table 1). In a related note, since the k-edge of Pt is 78.5 keV, it is possible that the optimal parameters and maximum sensitivity will actually occur at a voltage much lower than those tested. Little change was noted in the average signal level with increasing kVp, most likely indicating that the k-edge is currently some distance away from the range of voltages tested.

2.2. *In vitro/In vivo correlation*

The model drug release system used for the validation consisted of a biodegradable polymer matrix entrapping a CT contrast agent, iohexol. The purpose for such a model system is twofold: It allows for accurate assessment of the CT monitoring method in pharmacokinetic studies while providing valuable insight regarding the local drug release kinetics in the liver. Polymer implants were fabricated by compression-heat molding from PLGA, poly(D,L-lactide-*co*-glycolide), a commonly utilized, FDA compliant, biodegradable polymer.[37]

Gelatin phantoms were used as a tissue-mimicking environment to determine the accuracy of the CT method in monitoring the release of iohexol from the implants. Implants were inserted periodically over six days into 10% gelatin gels and iohexol was allowed to be released from the implant and diffuse into the matrix. The tissue phantoms were imaged using optimized parameters, and the implants were immediately removed from the gels after the CT scan. The iohexol remaining in the implants was extracted into PBS over six days and analyzed by UV-Vis spectrophotometry. The average CT intensity in each implant was determined from image analysis of circular ROIs (5 pixel diameter) averaged along the length of each implant. A background value for the polymer was subtracted from the average ROI measurement to obtain the signal from iohexol.

The release of iohexol was assessed by calculating the contrast change with time on the images and correlating this value to the UV-Vis measurement of extracted iohexol. These data were standardized to the values at $t = 0$. For the CT data, the average intensity in HU for the implants scanned immediately after implantation into gels ($t = 0$) was used to calculate the percentage of release at other time points. For the comparable chemical analysis, the same implants were removed from the gels, and the extracted iohexol concentration was measured with UV-Vis and used as the standardization value for the other time points.

The relative contrast decrease calculated from the CT images over time can be directly related to the concentration of iohexol in the explanted implants at each time point. The correlation between these two modes of analysis was excellent in the *in vitro* gelatin system, as shown in Fig. 2. The largest deviation between CT and UV-Vis data was 12%. Figure 2 insert shows the linear correlation ($R^2 = 0.99$) of both analysis methods. This result proves conclusively that CT is capable of accurately determining the concentration of an active agent in a model system. Furthermore, it should be noted that the study was performed in duplicate ($n = 2$) with excellent data reproducibility (see error bars in Fig. 2).

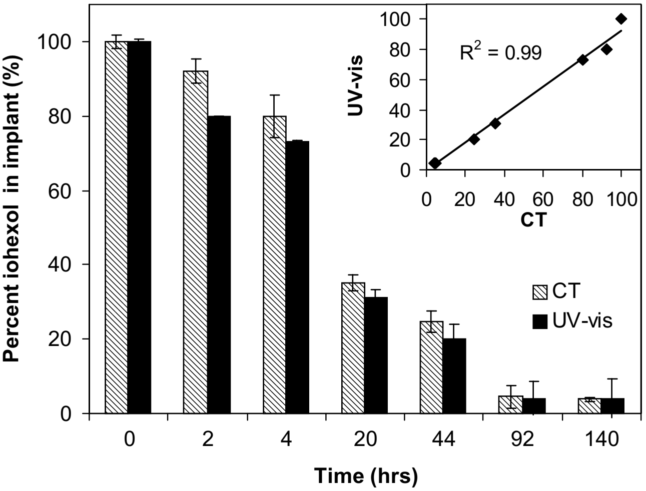


Fig. 2. Comparison of iohexol release profiles in gelatin gels from CT method and UV-Vis analysis. Figure 3 insert shows the linear correlation of the two methods (adapted from A. Szymanski-Exner *et al.*, *J. Pharm. Sci., 2003*).

The relationship between iohexol concentration and CT intensity in Hounsfield units (which is reflective of the X-ray attenuation by iodine) was also found to be linear, as expected. A plot of CT intensity versus iohexol concentration in the implants was constructed based on the experimental data. The slope for this relationship is 8.46 HU/(mg/ml). This value was used in the following *in vivo* studies to convert the image data to iohexol concentration.

2.3. *In vivo validation and CT monitoring of iohexol release*

Imagining parameters obtained from the optimization studies (600 mAs, 120 kVp, 1 mm slice thickness) were used to monitor the *in vivo* release of iohexol. Implants were placed into the livers of New Zealand White rabbits. The animals were imaged independently at 3 time points (1, 4 and 24 hrs) using the optimized scan parameters. The rabbits were then sacrificed and the scan was repeated. The implants were removed and iohexol was extracted into PBS and analyzed with UV-Vis spectrometry. The CT images were processed as described below.

2.3.1. *Image processing and analysis of iohexol release in vivo*

A three-dimensional registration method was utilized to spatially align serial CT volumes. A detailed description of this method was described previously.[38–40] This approach uses line paths and point landmarks to obtain volume registration. After registration, the image volume was re-sliced so that the orientation of the slices was exactly perpendicular to long axis of the implant. Based on phantom simulations of manual localization error and implant orientation typical of our *in vivo* experiments, we predicted a voxel displacement registration error of less than 0.5 mm. The new image slices were analyzed for the relative change of intensity within the implanted rod. The average concentration of iohexol remaining within the implant was estimated by calculating average pixel values in circular ROIs within the implant body (5 pixels diameter) and subtracting the average base polymer attenuation level (in gelatin) from the average data. The conversion value of 8.46 HU/(mg/ml) obtained from gelatin studies was used to calculate the iohexol concentration.

2.3.2. *Correction factor for respiratory motion*

Several experiments were carried out to examine the effect of motion artifacts on the accuracy of the concentration measurements obtained from image data analysis. We suspected that the lengthy 1.5-second rotation time used for minimizing noise and maximizing sensitivity during image acquisition might result in motion blur and artifact that would limit the quantitative concentration measurement. To examine the extent of motion artifact, we compared CT measurements of iohexol concentration in studies with movement (live animal) and without movement (following animal sacrifice) at three time points. A correction factor was calculated

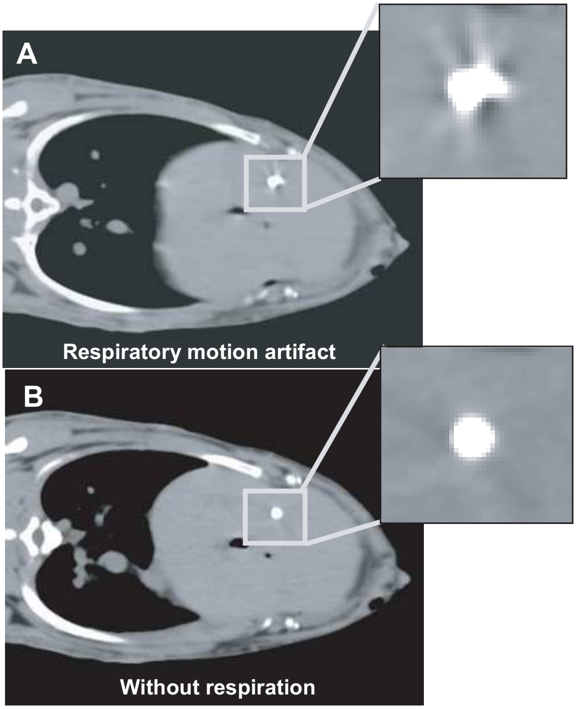


Fig. 3. CT images with and without respiratory motion artifact. A. Representative cross-sectional slice of the millirod in a live animal, B. The same slice following animal sacrifice (adapted from A. Szymanski–Exner *et al.*, *J. Pharm. Sci., 2003*).

as the slope of a linear fit between data without motion versus with respiratory motion. This factor was applied to all other *in vivo* concentration measurements to correct motion artifact in living, breathing animals.

The validity of the method was tested *in vivo* by directly monitoring the release of iohexol in rabbit livers at discrete time points. Figure 3 shows two CT slices of the same iohexol implant in a rabbit liver taken before and after sacrifice. It is evident that the respiratory motion plays a significant role in the quantitative image analysis at the current rotation time (1.5 sec) chosen for this study. Iohexol concentrations calculated from images taken after sacrifice (no motion) were significantly higher than the concentrations calculated from breathing animals. When plotted against each other, the slope of a line fit to this data using linear regression was 1.4 ± 0.2. We used this value as the correction factor to calculate the *in vivo* release data from moving, breathing rabbits. This correction factor was further validated during ongoing experiments with excellent results in multiple rabbits. The correction is necessary to improve the accuracy of concentration measurement in live rabbits but will vary depending on the experimental conditions (e.g. the breathing rate of the animal, and CT rotation time).

2.4. *Summary*

As evident by the data, the initial validation of CT in quantifying drug concentration in polymer release systems as well as evaluating the release profiles of drug from the implants proved to be quite successful and instigated further exploration of the technique, as described next.

3. CT Evaluation of *In vivo* Drug Release from Polymer Delivery Devices

The CT method was next applied to evaluate the release kinetics of an agent from a polymer drug release system in normal livers and livers with a pathologic environment induced by radiofrequency ablation. The tissue damage inflicted by the ablation procedure has a significant effect on the drug release rate from the implant, and thus must be taken into account when designing the therapeutic system for its clinical application.

Image acquisition parameters obtained from optimization studies were used to monitor the *in vivo* release kinetics of iohexol in rabbits with and without ablation of the liver tissue. Each rabbit was scanned at 1, 4, 24 and 48 hours following RFA treatment and implant placement. Concurrently, identical implants were placed in physiological buffer and the drug release was evaluated *in vitro*. The measurement of *in vivo* agent release from the image data was carried out by taking the average pixel value measurements in circular regions of interest along the entire volume of the rod. The amount of iohexol remaining in the implant was calculated by averaging the pixel values from the regions of interest within the implant matrix, subtracting the average base polymer attenuation levels from the average data, and using a conversion factor of 8.46 HU/(mg/ml) to calculate concentration values. The correction for respiratory motion was also applied to the data. The 50% release time was calculated from exponential decay curves fit to the animal data and for the *in vitro* data by linear extrapolation. The average release rate during each sampling interval was calculated by determining the slope of a line between each interval (HU/(mg/ml)).

The CT intensity was converted to iohexol concentration to permit quantitative measurement of iohexol release *in vivo*. The iohexol concentration in the ablated liver implants is between 16 to 30% higher than that in the normal liver implants during the first hour and 15.2 to 81.7% higher in 12/16 time points. When examining the average measurements in the normal liver environment, approximately 20% less iohexol remains in the polymer after the first hour. This discrepancy increases to 43% after 48 hrs. The 50% release time ($t_{1/2}$) differs significantly ($p = 0.046$) and is 1.7 times faster in normal liver (12.1 ± 5.4 hrs) as compared to ablated liver (20.6 ± 5.9 hrs).

The *in vitro* release data were compared to those *in vivo* to evaluate the adequacy of the *in vitro* model in representing the physiological system (Fig. 4). The

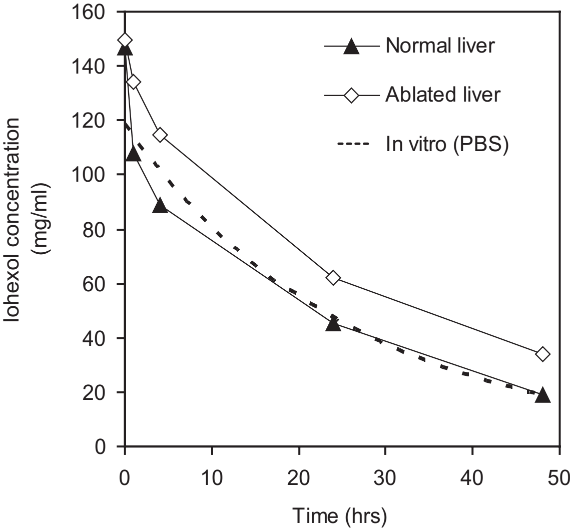


Fig. 4. Cumulative release of iohexol from PLGA implants in normal and ablated liver *in vivo* over 48 hours. The *in vitro* release data in PBS at 37°C is also shown for comparison.

release of iohexol *in vitro* is comparable to that in normal livers, as evident by similar $t_{1/2}$ values. This indicates that the local release of an agent in normal livers is equivalent to, and therefore may be predicted by, the release in PBS. More specifically, the *in vitro* release predicts a $t_{1/2}$ of 10.1 ± 1.2 hrs for normal livers and correlates well to the 12.1 ± 5.4 hrs calculated from CT data. The value of $t_{1/2}$ is two times longer in the ablated tissue (20.6 ± 5.9 hrs) than that indicated by the *in vitro* model.

The effect of tissue environment on drug release kinetics presents a unique challenge for the development of local drug delivery systems. Each tissue environment has a distinctive set of physiological parameters that affect the local release of a drug from the implant, and these effects must be evaluated before a successful local release system is produced. Here, we examined the influence of thermal ablation of the liver on the local drug release kinetics. Tissue damage caused by RF ablation is a chief factor in the potential clinical application of our drug delivery device.

Non-invasive examination of *in vivo* iohexol release from the implants showed significantly slower release kinetics in ablated livers as compared to normal livers. We believe these results reflect the different drug transport processes at the two implantation sites. In normal livers, drug transport consists of interstitial diffusion, cellular uptake, and drug clearance by perfusion processes.[41,42] If iohexol, an extracellular contrast agent, is released into viable liver tissue, diffusion and perfusion are the only two processes taking place. In non-ablated livers, the clearance of iohexol from the tissue/implant interface is fast due to the high blood perfusion rate. Consequently, the release rate in normal livers is faster due to the higher concentration gradient at this interface.

The unique local environment of the ablated tissue requires special consideration for the development of a suitable local therapy. Test tube studies are widely accepted as a sufficient method to characterize the release kinetics of an agent from a drug delivery device. However, the results suggest that precaution needs to be taken when *in vitro* release data is used to predict the release properties *in vivo*. The discrepancy in release kinetics between PBS and ablated liver demonstrates the limitations of the PBS system in approximating drug release in ablated livers and the importance of non-invasive imaging techniques in pharmacokinetic studies.

These results clearly demonstrate the immense potential of using computed tomography in monitoring local pharmacokinetics and confirm that CT monitoring provides physiologically relevant data that may not be otherwise directly observed. Because of its non-invasive nature, the technique can examine release in the same animal over time and minimize the extensive sample collection and processing required with a large animal group compared to conventional pharmacokinetic methods.

4. CT Methods for Evaluation of Tissue Penetration and Distribution of Drugs Delivered with Local Drug Release Systems

To validate the efficacy and demonstrate the feasibility of any localized drug delivery system, questions regarding the dynamics of drug distribution in tumor tissues and therapeutic margin need to be answered. In contrast to Sec. 3, where only the release of a model drug from the implant matrix was evaluated with CT, this section describes the effectiveness of CT image analysis techniques for the evaluation of the tissue penetration and distribution of platinated drugs delivered to the target site via an implanted sustained-release system. Two distinct methods of image analysis were employed in the project. In Method 1, registration and reformatting was explored for the straightforward, direct comparison of single image slices acquired at different times.[40] In Method 2, a more robust, volume analysis was utilized to obtain related measurements of drug levels in tissue.[43] As a practical example, the differences in local pharmacokinetics of normal and ablated liver tissue were explored. The methods were evaluated in a rabbit (Method 1) and rat (Method 2) model. The implants were placed in the liver, either with or without prior treatment with radiofrequency ablation (which induces coagulative necrosis). The animals were imaged with the optimal scan parameters, and the images were analyzed with either technique.

4.1. *Calibration of CT intensity to drug concentration*

A common thread in both methods was the determination of a calibration curve for the correlation of CT intensity in Hounsfield Units (HU) to the concentration of carboplatin suspended in various gelatin imaging phantoms. To determine the CT intensity for each carboplatin concentration, a circular Region Of Interest (ROI) was manually defined in the four central slices in each volume, and the average HU value along the perimeter of this region was calculated. The average HU values were plotted against carboplatin concentration, and linear regression was used to determine the parameters of a linear fit. Linear regression resulted in a statistically significant fit ($p < 0.05$) with a slope of 15.97 HU/(mg/mL). This corresponds to a sensitivity of 1 HU of CT signal for every 62.6 μg/mL increase in drug concentration.

4.2. *Evaluation of drug pharmacokinetics in liver tissue by Method 1*

4.2.1. *Registration and reformatting of CT volumes*

Serial CT slices were first registered and reformatted in order to spatially examine drug distributions at different times. The method used line paths, and optional point landmarks, to obtain volume registration.[38] First, the intersection of the cylindrical implants with every applicable slice was manually localized. Next, an optimum set of rigid body registration parameters was computed by iterative minimization of an objective function representing the mean Euclidean distance between correspondence points along the implant paths. The function value provided an estimate of registration error in the region of interest. In addition, based on phantom simulations of manual localization error and implant orientation typical of *in vivo* experiments, a voxel displacement registration error of < 0.5 mm was predicted.[38] Finally, all volumes were transformed so that the implant is perpendicular to the slice plane, with voxels at the original in-plane resolution. Image volumes from three time points were registered to allow accurate comparison between image sets.

The registration and reformatting process is crucial to the accurate analysis of drug pharmacokinetics. Registration allows a number of time points to be compared even though the animal has moved between imaging sessions, while reformatting provides perfect axial slices through the implant, decreasing partial volume effects and allowing an evaluation of the symmetry of drug transport.

4.2.2. *Background subtraction and circumferential averaging of radial plots*

To quantitatively evaluate the release of carboplatin, the image background was subtracted from each registered image set, leaving only signal produced by the presence of carboplatin. For each image, a background subtraction mask was created that consisted of three, manually segmented, homogeneous regions (implant, ablated tissue, and normal tissue). To obtain the value for normal tissue, a homogeneous region of liver background tissue was manually selected and the mean and standard deviation of the pixel gray levels were computed. The statistics for ablated tissue regions and the polymer implant were measured in a similar fashion. Ablation and implant boundaries were defined for each registered image set to create the subtraction mask. The subtracted image was displayed to verify the quality of the subtraction.

An averaging algorithm was created to evaluate drug transport out of the implant and into the surrounding tissue. The registration and reformatting process provides a series of axial slices through each implant, and the subtraction process resulted in images where the entire signal is produced by the presence of carboplatin. The maximum pixel in the ROI surrounding each rod was defined as that rod's center. From this center point, a series of 360 equally spaced radial lines were drawn and the radial profiles from all 360 lines in each of three slices at the center of each registered implant volume were averaged together. The previously determined relationship between CT signal and drug concentration was then applied to the average profile, converting the CT intensity in HU to the number of drug molecules present in each pixel over time, revealing the release and distribution of carboplatin out of the implant. Accurate background subtraction coupled with circumferential averaging provides previously unknown quantitative information about the concentration of drug at various spatial locations.

Figure 5 shows that drug is transported away from the implant in all directions. This suggests that an appropriate method for measuring the typical transport is to circumferentially average over a number of radial lines extending from the center of each implant. The normal implants, without ablation, show a rapid decrease in carboplatin both surrounding and inside the rod. Approximately 50% of the drug in the rod is lost between 1 hour and 24 hours. With ablation, most of the drug is retained from 1 hour to 24 hours. This is most likely the result of a destruction of local vasculature due to the ablation procedure, as supported by histological

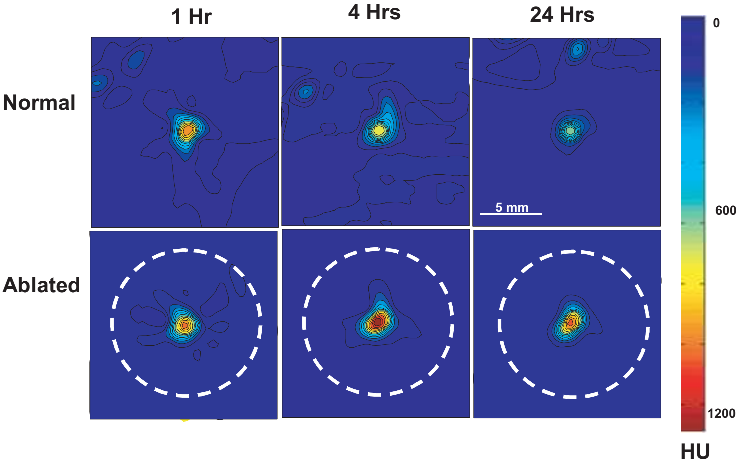


Fig. 5. Contour plots showing the spatial distribution of carboplatin in normal and ablated tissue after subtraction of a background mask. Carboplatin distribution is more extensive in ablated tissue at 4 and 24 hours, while at 1 hour the spread of drug is about the same in normal and ablated tissue.

analysis.[25] In all cases, about 4 mm from the center of the implanted implant, the CT images no longer show measurable drug concentrations.

4.2.3. *Regional drug distribution through a pillbox analysis*

The local distribution of carboplatin was also evaluated by calculating the number of drug molecules present in a series of virtual “pillboxes”, or cylindrical volumes defined to enclose a relevant structure such as the implant or a tumor. In our analysis, three pillboxes were defined: Implant, virtual tumor and tissue. A fourth compartment, drug washout, was set to be the total drug dosage minus the total number of drug molecules that can be counted in the image. To determine the number of drug molecules in a particular Volume Of Interest (VOI), we integrated over the pillbox and subtracted the number of molecules that can be committed to other enclosed regions. The initial number of molecules at the time of implantation was determined by measuring a control rod placed outside the rabbit. Drug washout provided a measure of vascular clearance as a result of blood perfusion.

Figure 6 shows the number of molecules in the rod, virtual tumor, and surrounding tissue, as well as other drug molecules that have been washed out of the region. As shown in the radial plots, the amount of drug inside the rod and in the tumor pillbox is considerably higher in the ablated tissue than in normal tissue, while the tissue pillbox shows no appreciable drug concentration because it is either lower than the sensitivity of the technique or contained within the noise of the liver background. The amount of drug washed out of the treatment site is much higher in normal tissue at all time points, most likely due to the presence of local vasculature. Importantly, the amount of drug inside the rod and tumor pillboxes after 24 hours is 25% higher in ablated as opposed to normal tissue. The pillbox analysis gives insight into the bulk location of drug at a given time and could be easily expanded

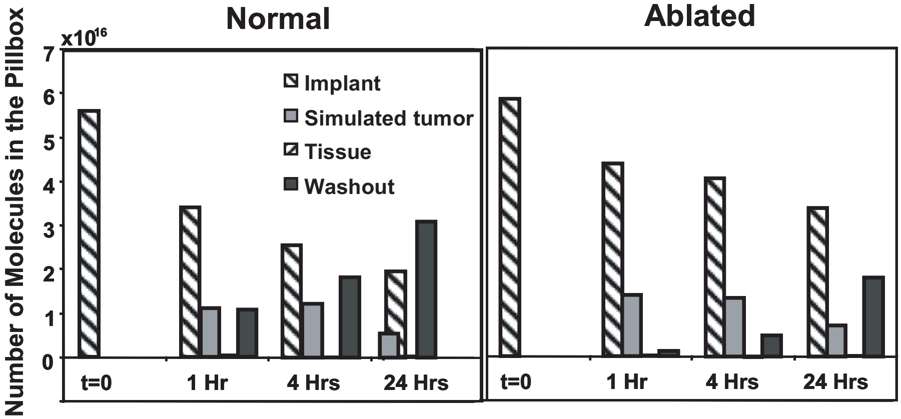


Fig. 6. Pillbox analysis of drug moelcules in tissue and implant throughout the course of the experiment.

to look at the movement of drug to other areas of interest or anatomical locations, such as the kidney or bladder.

Although this analysis technique successfully demonstrated the proof-of-concept of CT is evaluating local tissue pharmacokinetics, it also revealed a number of shortcomings that we attempted to correct with a second, more comprehensive technique. The following section describes the second method of data analysis.

4.3. *Evaluation of drug pharmacokinetics in liver tissue by Method 2*

CT data collected in the same manner as above CT data underwent extensive processing to evaluate the local pharmacokinetics of carboplatin. An analysis regimen was developed to isolate tissue and implant levels of carboplatin from the other components in each image volume. The method begins with background subtraction of tissue and polymer attenuation, followed by circumferential averaging of the data about each cylindrical implant. Next an inherent blur correction is applied and the analysis is concluded with a separate regional distribution analyses in the implant and tissue volumes. A detailed description of this analysis is provided below.

4.3.1. *Sampling of CT volumes and circumferential averaging of radial profiles*

A three-dimensional sampling method was used to spatially examine drug distributions at an orientation exactly perpendicular to the long axis of the cylindrical polymer implant. First, the center of the implant was manually localized. Next, to accurately determine the precise orientation of the implant in the CT volume a best-fit line was computed in three dimensions using the least squares algorithm. To determine HU values from the images, a trilinear interpolation was used to obtain the signal intensity at points along concentric circles perpendicular to the 3D best-fit line.

To evaluate drug transport from the implant into surrounding tissue, points in the 3D image space were circumferentially averaged to yield a single 1D radial profile for each implant. Circumferential samples were obtained with an angular interval of five degrees and a radial interval of one half of the in-plane pixel width (0.147 mm). Sampled points were binned by radial distance and averaged. The measurements were then compared between slices to establish differences in release properties along the length of the implant. The profiles were further averaged along the z-axis to yield one average profile of drug concentration as a function of distance from the center of the implant. This facilitated direct comparison of the CT and chemical analysis data.

4.3.2. *Accounting for blurring of the implant by the CT imaging system*

To account for blurring of the implant by the imaging system, we assumed a linear response to drug in the implant. We placed a carboplatin implant in a gelatin phantom and immediately imaged it. This gave the rod image at $t = 0$, and the measured, circumferentially averaged profile was designated as the background profile due to blurring of a drug-containing rod in drug-free surrounding tissue. For each subsequent experimental profile, this profile was linearly scaled to the peak signal of the implant as averaged over the innermost half of the implant (radius $< 0.4\,\text{mm}$). The background could then be removed from subsequent measurements to obtain the drug in the tissue.

4.3.3. *Calculating carboplatin contained within the implant body and surrounding tissue*

To determine the average carboplatin concentration inside the implant body, we determined the HU values of the background rod profile that had been scaled to the experimental profile. Those points nearest the edge of the implant (0.4–0.8 mm from the rod center) were excluded to minimize the effects of drug inhomogeneity within the implant. To account for X-ray attenuation from the polymer component of the rod, the HU value for a PLGA polymer implant without drug was subtracted from each HU value within the drug-containing implants. This correction was typically less than 10% of the total. The resulting difference was converted to carboplatin concentration using the *in vitro* rod conversion value determined as described above. Volume integration of these spatially varying concentrations for each profile allowed calculation of the drug mass present in each slice of the implant, which could then be divided by the volume to yield an average drug concentration. A mean concentration of carboplatin for the entire rod was determined by further averaging the carboplatin concentrations along the length of the device.

Carboplatin concentrations outside of the implant were determined by a similar method. First, we subtracted the background image of the rod that had been linearly scaled to the experimental profile. These HU values were then converted to carboplatin concentration by subtracting a background corresponding to either healthy or ablated liver tissue and using the *in vitro* conversion value determined in the gelatin phantom. Figure 7 shows the algorithm used to calculate carboplatin concentrations, while Fig. 8 shows the distribution of carboplatin in liver tissue as calculated by Method 2.

In the present method of image processing, sampling, background subtraction and a precise calculation of the drug concentration to HU conversion factor are all essential for achieving the most accurate quantitative description of the system. The fitting of a line to the center of the cylindrical implant allows sampling of data at an orientation exactly perpendicular to the long axis of the implant, optimizing the circumferential averaging process and allowing simple ROI analysis. Contrary to Method 1, this technique circumferentially averages the signal from a control image set, and then is linearly scaled and subtracted from the experimental profiles to determine the carboplatin concentrations both within the implants and in the surrounding tissue.

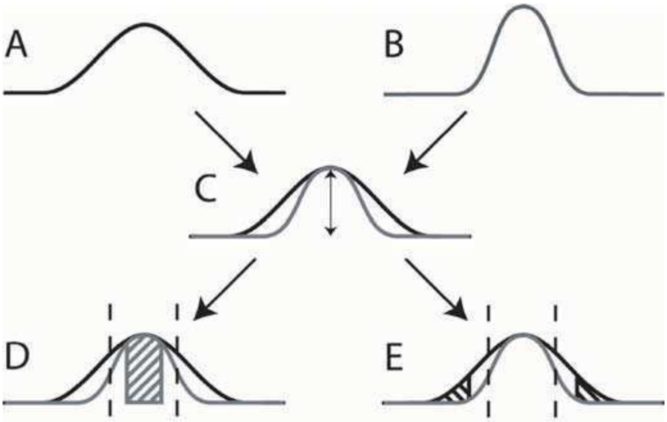


Fig. 7. Schematic of algorithm used to determine carboplatin concentrations in Method 2. After obtaining an experimental profile (A), the background profile generated by a rod in gelatin at time = 0 (B) was scaled to the height of the experimental profile (C). After scaling, the concentration value within the implant was determined by averaging the scaled values within 0.4 mm of the implant center (hashed region, D). The value in tissue was determined by subtracting the difference between the two profiles hashed region, E). The dashed lines (D and E) show approximately where the implant/tissue interface would be located.

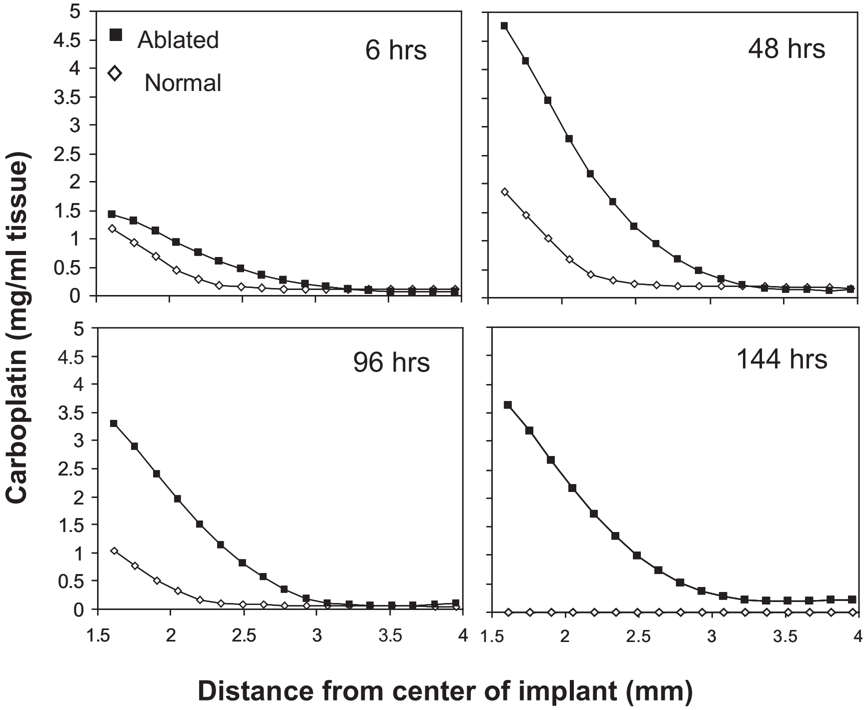


Fig. 8. Carboplatin distribution in normal (◇) and ablated (■) liver tissue measured by CT and Method 2 processing.

4.4. *Correlation of CT data with chemical analysis*

The image data from Method 2 was compared to Atomic Absorption Spectroscopy (AAS) analysis of carboplatin content in the liver tissue. AAS is a widely accepted method of analyzing tissue platinum levels. From this analysis we found that the CT data overestimates the drug content determined by the AAS data predominantly between 1–2 mm and underestimates it as the distance from the implant increased. Figure 9 is an example of this correlation and describes the quantitative change in drug concentration as a function of time for a tissue volume. In the ablated liver, the concentration of carboplatin is sufficiently high and can be detected by CT with fair correlation to the actual concentration (from AAS). However, when the drug concentration decreases, the accuracy of the CT detection diminishes. At distances away from the implant, the CT profile data describes the retention and penetration of the carboplatin as more extensive in the ablated liver, but it is not accurate at the low drug concentrations detected by AAS (quantification limit 25 ppb). The quantum noise combined with background subtraction lead to negative concentration values which were regarded as zero.

This is the effect of insufficient sensitivity in the CT imaging and is most likely a result of two factors: (i) the inherent blur artifact associated with small bright objects (e.g. implants) placed in a larger and less-attenuating matrix (e.g. liver tissue), and (ii) the beam hardening artifact resulting from the high contrast difference between the implant and the tissue. Both of the artifacts are clearly visible in the CT images, and although found in the normal liver, the effects are more

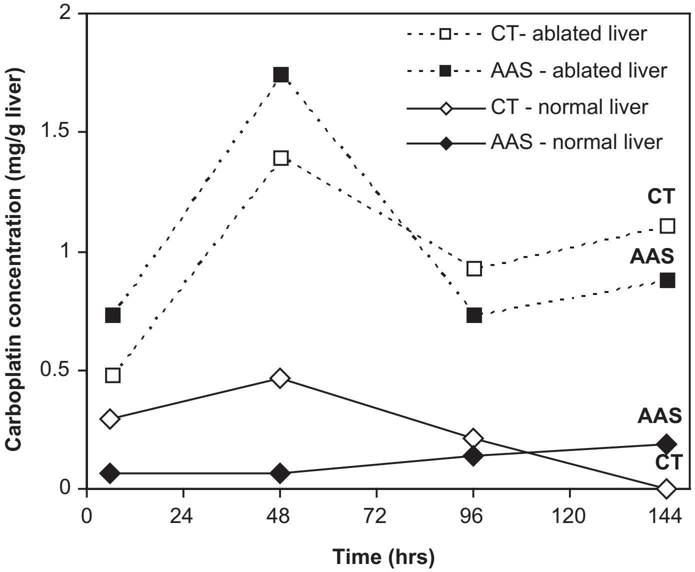


Fig. 9. Change with time of average carboplatin concentration in normal and ablated liver tissue volumes as determined by CT(◇, □) and AAS(■, ◆). Data represents mean ± SD.

notable in the ablated tissue (which has gray levels approximately 30 HU lower than the normal liver). It is also evident that the greatest overestimation occurs at distances closest to the implant boundary and lessens with increasing radius as well as with time. The less than optimal correlation can also be attributed in part to the difference in spatial resolution of the two methods. AAS sampling occurred at a frequency of 2 mm compared to the sub-mm sampling of the CT.

4.5. *Summary*

Section 4 presented two possible methods for analysis of CT data collected from the evaluation of a local drug release system. Method 1 utilizes a single slice approach, where a direct comparison of single image slices acquired at different times was carried out by a circumferential averaging algorithm. This method also utilized a volume or "pillbox" analysis to gain a more comprehensive understanding of the local drug behavior. In Method 2, a volume analysis was utilized throughout the process to obtain the same data as from Method 1. The primary advantage of this method is that the manual slice by slice registration and reformatting of the images is no longer required. Furthermore, the method recognizes and attempts to correct for the blurring artifact, to provide a more precise measurement. Both methods utilize a similar conversion from HU to drug concentration, and both fall short of the desired sensitivity for such an analysis technique. The bottom line is that although the analysis can provide us with accurate measurements of drug movement within the implant matrix, the signal to noise ratio is insufficient to detect drugs at the lower limits of their active concentrations within the tissue. Despite the evolution of the analysis process, a number of improvements can still be incorporated into the processing techniques to optimize these measurements. Some of these limitations and ideas for technique refinement are discussed further in the following section.

5. Potential Limitations of CT in Drug Monitoring

Before the CT tracking method can be reliably used for in depth pharmacokinetic analysis, several concerns must be addressed. The primary issue involves applicability of the technique to therapeutic drugs, which are limited to chemicals containing a heavy element that will attenuate X-rays. Fortunately, there are currently 25 platinum-containing anticancer agents in clinical trials, and four have been approved for clinical use.[44,45] Thus the CT method is directly applicable for studying the local pharmacokinetics of clinically used drugs, such as carboplatin and cisplatin. For most other drugs, covalent attachment of a heavy (but not radioactive) element may be necessary to generate CT contrast. However, special precaution needs to be taken to ensure the addition of such an element does not affect the drug pharmacokinetics and mechanism of action. Another concern, which we have touched upon in this work, is the respiratory motion artifact. The correction factor used here is an interim solution to a crucial issue affecting the accuracy of the quantitative data as well as the sensitivity of the method. The tradeoff between longer scan times to improve sensitivity and motion reduction to prevent artifact requires further investigation, particularly in examining tissue penetration and distribution of a drug in animals as well as in humans. Respiratory gating may offer a superior resolution to this dilemma. Other potential improvements can also be made to optimize the technique. For example, blurring, which most likely is responsible for the overestimation of drug concentration adjacent to the implant/tissue boundary, can be diminished by changing the reconstruction filter to obtain a more accurate drug release profile.

Although it is clear that the detection limit of the current CT technique (which is highly dependent on image noise) is too low to adequately describe the *absolute* drug levels present in the tissue, the clear benefits of the CT monitoring lie instead in the analysis of release from the implants and in the *relative* comparisons of drug release and tissue distribution in different systems. Already it can be seen that the technique does detect a difference between the normal and ablated liver distribution (as seen in the maximum penetration distance analysis). However, although one can track the amount of drug in the implant, the accurate quantification of absolute drug concentration and spatial distribution in the tissue is not realistic at this time.

Overall, the method provides a simple and convincing approach for monitoring *in vivo* drug release in agents able to attenuate X-rays, but additional work is required for the technique to be the sole assessment of tissue distribution (by radial profiles) at low drug concentrations. Evidence is nevertheless compelling that the method is quite capable in evaluating the release of an agent from a small drug delivery device and non-invasively acquiring data unavailable by any other means. The results of this work could potentially reduce inter-animal variability common in *in vivo* evaluations of local therapies and could significantly reduce the number of animals required to carry out these studies.

6. Conclusions

The current studies demonstrate the feasibility of computed tomography in non-invasive drug monitoring and offer a glimpse into the immense potential of this powerful technique. The CT method has a number of unique advantages that make it an ideal modality for this application. Primarily, CT has a superb combination of sensitivity and resolution as compared to other imagining modalities. Moreover, CT allows a rapid collection of many high-resolution images (less than 20 seconds per animal scan) and can minimize the variability of release kinetics due to long scanning time (e.g. MRI scans often take significantly longer). Furthermore, it eliminates the need for extensive tissue collection and requires fewer animals than more traditional pharmacokinetic analysis. This non-invasive nature, in turn, minimizes animal variability. With these advantages, the potential of CT in advancing the development and clinical applications of drug release systems is unparalleled and should be pursued by investigators requiring non-invasive pharmacokinetic monitoring in their research quest.

Acknowledgments

The authors would like to thank Drs. John Haaga, Uri Shreter, Kyle Salem, and Nick Stowe along with Mr. Les Ciancibello, and Brent Weinberg for their assistance with this project. Funding for this work was provided by the Whitaker Foundation and NIH (CA 93993 and CA 90696).

References

1. G. J. Becker, *Radiology* **220** (2001) 281.
2. C. E. Ray, Jr, *Am. Fam. Physician* **62** (2000) 102.
3. T. J. Vogl, P. K. Muller, M. G. Mack, R. Straub, K. Engelmann and P. Neuhaus, *Eur. Radiol.* **9** (1999) 675.
4. M. L. Montgomery and J. P. Sullivan, *Postgrad Med.* **109** (2001) 93.
5. S. Rossi, M. Di Stasi, E. Buscarini, P. Quaretti, F. Garbagnati, L. Squassante, C. T. Paties, D. E. Silverman and L. Buscarini, *AJR Am. J. Roentgenol.* **167** (1996) 759.
6. S. Rossi, E. Buscarini, F. Garbagnati, M. Di Stasi, P. Quaretti, M. Rago, A. Zangrandi, S. Andreola, D. Silverman and L. Buscarini, *AJR Am. J. Roentgenol.* **170** (1998) 1015.
7. L. R. Jiao, P. D. Hansen, R. Havlik, R. R. Mitry, M. Pignatelli and N. Habib, *Am. J. Surg.* **177** (1999) 303.
8. J. S. Lewin, C. F. Connell, J. L. Duerk, Y. C. Chung, M. E. Clampitt, J. Spisak, S. G. Gazelle and J. R. Haaga, *J. Magn. Reson. Imaging.* **8** (1998) 40.
9. G. D. Dodd III, M. C. Soulen, R. A. Kane, T. Livraghi, W. R. Lees, Y. Yamashita, A. R. Gillams, O. I. Karahan and H. Rhim, *Radiographics* **20** (2000) 9.
10. H. Rhim and G. D. Dodd III, *J. Clin. Ultrasound* **27** (1999) 221.
11. S. N. Goldberg, G. S. Gazelle, C. C. Compton, P. R. Mueller and K. K. Tanabe, *Cancer* **88** (2000) 2452.
12. S. N. Goldberg and G. S. Gazelle, *Hepatogastroenterology* **48** (2001) 359.
13. S. N. Goldberg, *Eur. J. Ultrasound* **13** (2001) 129.
14. T. Livraghi, S. N. Goldberg, S. Lazzaroni, F. Meloni, L. Solbiati and G. S. Gazelle, *Radiology* **210** (1999) 655.
15. R. Lencioni, O. Goletti, N. Armillotta, A. Paolicchi, M. Moretti, D. Cioni, F. Donati, A. Cicorelli, A. Ricci and M. Carrai, *Eur. Radiol.* **8** (1998) 1205.
16. T. Livraghi, S. N. Goldberg, S. Lazzaroni, F. Meloni, T. Terace, L. Solbiati and G. S. Gazelle, *Radiology* **214** (2000) 761.
17. L. Solbiati, T. Livraghi, S. N. Goldberg, T. Ierace, F. Meloni, M. Dellanoce, L. Cova, E. F. Halpern and G. S. Gazelle, *Radiology* **221** (2001) 159.
18. L. Solbiati, T. Ierace, M. Tonolini, V. Osti and L. Cova, *Eur. J. Ultrasound* **13** (2001) 149.
19. J. M. Llovet, R. Vilana, C. Bru, L. Bianchi, J. M. Salmeron, L. Boix, S. Ganau, M. Sala, M. Pages and C. Ayuso, *Hepatology* **33** (2001) 1124.
20. S. A. Curley, F. Izzo, L. M. Ellis, N. Vauthey and P. Vallone, *Ann. Surg.* **232** (2000) 381.

21. S. N. Goldberg and L. Solbiati, *Hepatology* **34** (2001) 609.
22. M. Ahmed, W. E. Monsky, G. Girnun, A. Lukyanov, G. D'Ippolito, J. B. Kruskal, K. E. Stuart, V. P. Torchilin and S. N. Goldberg, *Cancer Res.* **63** (2003) 6327.
23. W. L. Monsky, J. B. Kruskal, A. N. Lukyanov, G. D. Girnun, M. Ahmed, G. S. Gazelle, J. C. Huertas, K. E. Stuart, V. P. Torchilin and S. N. Goldberg, *Radiology* **224** (2002) 823.
24. F. Qian, N. Stowe, G. M. Saidel and J. Gao, *J. Pharm. Res.* **21** (2004) 394.
25. J. Gao, F. Qian, A. Szymanski-Exner, N. Stowe and J. Haaga, *J. Biomed. Mater. Res.* **62** (2002) 308.
26. A. Szymanski-Exner, N. T. Stowe, R. S. Lazebnik, K. Salem, D. L. Wilson, J. R. Haaga and J. Gao, *J. Control Release* **83** (2002) 415.
27. R. E. Port and W. Wolf, *Invest New Drugs* **21** (2003) 157.
28. M. Singh and V. Waluch, *Adv. Drug Deliv. Rev.* **41** (2000) 7.
29. A. Bhatnagar, R. Hustinx and A. Alavi, *Adv. Drug Deliv. Rev.* **41** (2000) 41.
30. J. S. Taylor and W. E. Reddick, *Adv. Drug Deliv. Rev.* **41** (2000) 91.
31. T. T. Nguyen, Y. S. Pannu, C. Sung, R. L. Dedrick, S. Walbridge, M. W. Brechbiel, K. Garmestani, M. Beitzel, A. T. Yordanov and E. H. Oldfield, *J. Neurosurg.* **98** (2003) 584.
32. J. R. Griffiths and J. D. Glickson, *Adv. Drug Deliv. Rev.* **41** (2000) 74.
33. D. J. Dowsett, P. A. Kenny and R. E. Johnston, *The Physics of Diagnostic Imaging* (Chapman & Hall Medical, London, 1998).
34. J. T. Bushberg, J. A. Seibert, E. M. Leidholdt and J. M. Boone, *The Essential Physics of Medical Imaging* (Williams & Wilkins, Baltimore, 1994).
35. T. S. Curry, J. E. Dowdey and R. C. Murry, *Christensen's Physics of Diagnostic Radiology* (Lippincott Williams & Wilkins, Philadelphia, 1990).
36. J. H. Hubbell and J. H. Scofield, *X-Ray Attenuation Cross Sections for Energies 100 eV to 100 keV and Elements* $Z = 1$ *to* $Z = 92$*; Atomic Data and Nuclear Data Tables* **38** (1988).
37. F. Qian, A. Szymanski and J. Gao, *J. Biomed. Mater. Res.* **55** (2001) 512.
38. R. S. Lazebnik, T. L. Lancaster, M. S. Breen, J. S. Lewin and D. L. Wilson, *IEEE Transactions in Medical Imaging*, in press (2002).
39. R. S. Lazebnik, T. L. Lancaster, M. S. Breen, S. G. Nour, J. S. Lewin and D. L. Wilson, *Proceedings of SPIE Medical Imaging: Visualization, Display and Image-Guided Procedures* **4681** (2002) 39.
40. K. A. Salem, A. Szymanski-Exner, R. S. Lazebnik, M. S. Breen, J. Gao and D. L. Wilson, *IEEE Trans Med Imaging* **21** (2002) 1310.
41. R. Jain, N. H. Shah, A. W. Malick and C. T. Rhodes, *Drug Dev. Ind. Pharm.* **24** (1998) 703.
42. J. L. Au, S. H. Jang, J. Zheng, C. Chen, S. Song, L. Hu and M. G. Wientjes, *J. Control Release* **74** (2001) 31.
43. A. A. Exner, B. Weinberg, N. T. Stowe, A. Gallacher, D. L. Wilson, J. R. Haaga and J. Gao, *Acad. Radiol.* **11** (2004) 1326.
44. M. A. Jakupec, M. Galanski and B. K. Keppler, *Rev. Physiol. Biochem. Pharmacol.* **146** (2003) 1.
45. D. Lebwohl and R. Canetta, *Eur. J. Cancer* **34** (1998) 1522.

This page intentionally left blank

CHAPTER 6

TECHNIQUES IN TREATING THE BIOELECTROMAGNETIC SOURCE IMAGING PROBLEMS AND THEIR APPLICATION

F. GREENSITE
Department of Radiological Sciences
University of California-Irvine, CA, USA
fredg@uci.edu

A. PULLAN
Department of Engineering Science
University of Auckland, New Zealand
a.pullan@auckland.ac.nz

G. HUISKAMP
Department of Clinical Neurophysiology
University Medical Center, Utrecht, The Netherlands
ghuiskam@neuro.azu.nl

We present the theory and application of non-invasive or minimally invasive imaging of bioelectromagnetic sources. We are not concerned with imaging secondary effects of the sources (that might be used to infer their location), such as changes in tissue oxygenation, blood flow, or glucose utilization (as with BOLD or dynamic gadolinium functional MRI strategies, or PET). Thus, we are directly concerned with providing images of source currents or source potentials. These sources reside in "excitable tissues", such as brain, heart, gut, and skeletal muscle.

While we will not be conducting a comprehensive literature survey, this chapter is intended to provide an overview of the general approaches and trends in bioelectromagnetic imaging that presently characterize the field. Within reason, we attempt this by considering the relevant physiology, physics, engineering, and mathematics that in concert allow a coherent understanding of the present state of affairs. Accordingly, this chapter's sections are arranged as follows:

- General Principles
  - Modeling of excitable tissue
  - Physics of bioelectromagnetism
  - Engineering issues: Signal acquisition
  - Mathematical methods

- Source Categories
  - Brain: Source physiology, dipole localization, imaging formulations, linear versus nonlinear problems
  - Heart: Source physiology, endocardial, epicardial, and transmembrane potential imaging
  - Smooth and Skeletal muscle: Source physiology, inverse problems related to electromyography and electrogastrography.

Keywords: Bioelectromagnetic source imaging; heart; skeletal muscle; gastrointestinal smooth muscle.

GENERAL PRINCIPLES

1. Modeling of Excitable Tissue

1.1. *Ionic currents*

The sources that we ultimately seek to image arise from various ionic currents that control the activity of the cells within the excitable tissues. These ionic currents result from the transfer of ions through the cell membrane between the intra-cellular and extra-cellular fluid.

In general the cell-membrane will be permeable only to certain ions in certain directions. Typically, the intra-cellular concentration for an ion will be different from the extra-cellular concentration for the ion. The concentrations of common ions in the different cell spaces for different tissues is given in Tables 1 and 2.

The different concentrations set up an electro-chemical potential across the cell membrane. This potential is given by the Nernst equation

$$E = \frac{RT}{ZF} \log_e \frac{[C]_o}{[C]_i}, \tag{1}$$

where R is the gas constant, T the absolute temperature, Z the valency of the ion involved and F is Faraday's constant. At 37°C $\frac{RT}{F}$ is approximately 26 mV so a 10 fold change in ion concentration corresponds to approximately a 60 mV change in potential.

Now from Ohm's law the ionic current through a membrane for say, potassium, will be given by

$$\begin{aligned} i_K &= \frac{1}{R}\Delta V \\ &= g_K(V_m - E_K), \end{aligned}$$

Table 1. Intra- and Extra-cellular ion concentrations and equilibrium potentials for mammalian skeletal muscle.

	Intra-cellular concentration (mM)	Extra-cellular concentration (mM)	Equilibrium potential (mV)
Na^+	12	145	$E_{Na^+} = +65$
K^+	155	4	$E_{K^+} = -95$
Ca^{2+}	$<10^{-7}$	1.5	$E_{Ca^{2+}} = +128$
Cl^-	4.2	123	$E_{Cl^-} = -88$

Table 2. Intra- and Extra-cellular ion concentrations and equilibrium potentials for mammalian cardiac muscle.

	Intra-cellular concentration (mM)	Extra-cellular concentration (mM)	Equilibrium potential (mV)
Na^+	5–34	140	$E_{Na^+} = +51$
K^+	104–180	5.4	$E_{K^+} = -85$
Ca^{2+}	0.0003–0.001	3	$E_{Ca^{2+}} = +133$
Cl^-	8–79	100	$E_{Cl^-} = -22$

where g_K is the membrane conductance for potassium ions and V_m is the transmembrane potential (i.e. the difference in potential between the intra-cellular and extra-cellular spaces).

It should be noted that by definition an outward current (from intra- to extracellular space) is defined as a positive current. It should also be noted that electrical current flows in the direction that the positive charge carriers move. Hence the potassium current (in which the positive K^+ ions move from inside the cell to outside the cell) is a positive current. As the transmembrane potential increases above E_K the membrane current changes from a negative (inward) current to a positive (outward) current. The point E_K is hence known as the reversal potential as it is at this point that the current reverses its direction.

The membrane conductance for an ion depends on the number of conducting channels and their properties or gating. In 1952 Alan Hodgkin and Andrew Huxley performed voltage clamp experiments on giant squid axons.[64] From their experiments, they determined the conductances and dynamic gating mechanisms controlling the movement of potassium and sodium ions across the cell membrane. The work of Hodgkin and Huxley (for which they were awarded the 1963 Noble prize for Physiology or Medicine) can be summarized in the Hodgkin–Huxley equations for the action potential.

$$\begin{aligned}
\frac{1}{r_a}\frac{\partial^2 V_m}{\partial x^2} &= C\frac{\partial V_m}{\partial t} + i_K + i_{Na} \\
i_K &= n^4\overline{g_K}(V_m - E_K) \\
i_{Na} &= m^3 h\overline{g_{Na}}(V_m - E_{Na}) \\
\frac{dn}{dt} &= \alpha_n(1-n) + \beta_n n \\
\frac{dm}{dt} &= \alpha_m(1-m) + \beta_m m \\
\frac{dh}{dt} &= \alpha_h(1-h) + \beta_h h,
\end{aligned} \tag{2}$$

where $\alpha_n, \beta_n, \alpha_m, \beta_m, \alpha_h$ and β_h are all voltage dependent rate constants, x is the distance along the axon, r_a is the resistance of the axon and t is time. The variables n, m and h are the names given to the various ion channel gates associated with the conductances of the K^+ and Na^+ ions.

Putting the Na^+ and K^+ channels together we obtain the nerve action potential shown in Fig. 1.

Action potentials are propagated by the following basic mechanism: When the membrane potential of one end of the nerve is raised an axial current along the nerve from the high potential end to the low potential end is generated in accordance with Ohm's law. This current raises the membrane potential (depolarises) to a point whereby the sodium channels open. The point is known as the threshold voltage. Once the sodium channels open the sodium ions flood in and cause an inward current which raises the membrane potential even higher. This generates

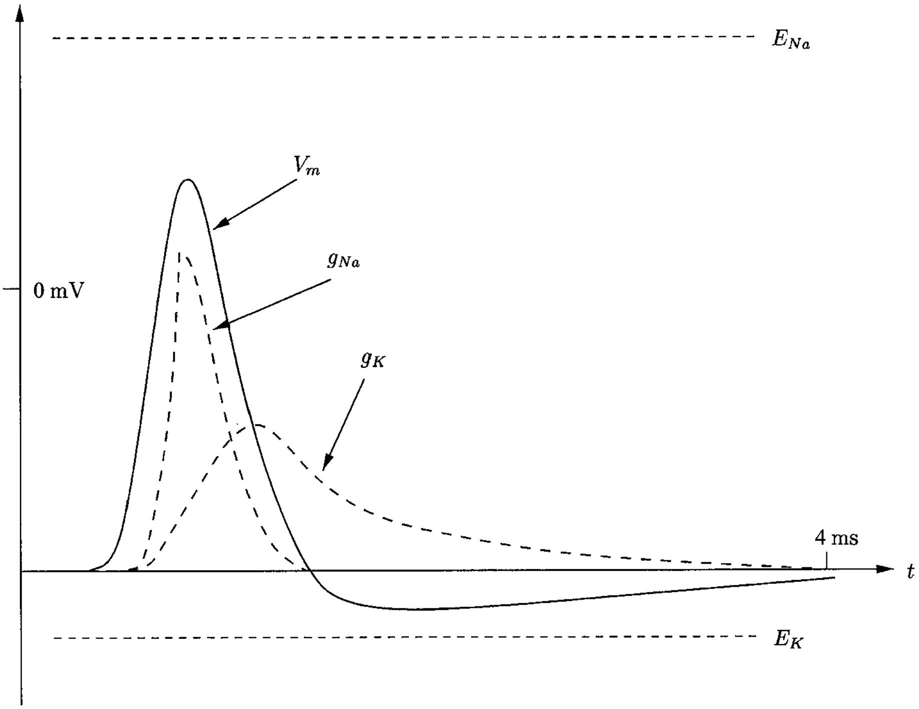


Fig. 1. The squid action potential with the sodium and potassium channel conductances.

further axial current and a resultant action potential is hence propagated down the nerve. It should be noted that this process requires ion pumps in the membrane to restore the intra- and extra-cellular ion concentrations to their resting levels. These pumps require energy (ATP — adenosine triphosphate) to operate. Until the ionic concentrations in an area of nerve have been returned to their resting states an action potential will not propagate in this area of nerve. This period after an action potential has passed when the nerve cannot be activated is called the refractory period.

In contrast to neuronal tissue, muscle is excitable *and* contractile. There are several types of muscle in the human body namely cardiac, skeletal and smooth muscle. Muscle cell membranes are more complicated than that of the squid axon. In addition to the sodium and potassium channels muscle cells also have calcium channels, whose primary role is contraction. When muscle is stimulated appropriately, it generates an action potential, and Ca^{+2} is released resulting in a contraction. The details associated with action potential activity and Ca^{+2} release are different for each muscle type. As we will see, these details affect not only the way we attempt to non-invasively image the underlying electrical activity, but also the type of activity we can hope to detect. Relevant aspects of neuronal, cardiac muscle, skeletal muscle, and smooth muscle structure and physiology with be presented in later sections.

1.2. *Models of the electrical activity of non-neural cells*

1.2.1. *Cardiac cell models*

Most cardiac cell models contain components that are based on the Hodgkin–Huxley model of the squid axon. Since that model first appeared, the number of different cell models has grown rapidly as more has become known about the different subcellular mechanisms responsible for the functioning of the cell. Some of the better known models include the DiFrancesco–Noble model of the Purkinje fiber cell,[31] the Beeler–Reuter ventricular cell model,[10] the Luo–Rudy mammalian ventricular cell models (LRI and LRII)[84,85] and the Noble model of the guinea pig ventricular cell.[103] Figure 2 is a schematic representation of the ionic current, pumps and exchangers that are captured in the LR-II model. An action potential generated from this model shown in Fig. 3.

The level of complexity inherent in these cellular models has also increased significantly, with 2 currents (I_{Na}, I_K) and 3 gating variables $(m, h$ and $n)$ required in the original Hodgkin–Huxley model, 4 currents $(I_{Na}, I_s, I_{x1}, I_{K1})$ and 6 gating variables $(m, h, j, d, f, x1)$ in the Beeler–Reuter model, 12 currents $(I_f, I_K, I_{K1}, I_{t0}, I_{bNa}, I_{bCa}, I_p, I_{NaCa}, I_{Na}, I_{Caf}, I_{Cas}, I_{pulse})$ and 7 gating variables (y, x, r, m, h, d, f) in the model of Difrancesco–Noble and the most recent version of the Luo–Rudy II model containing 14 currents $(I_{Na}, I_{Ca}, I_{CaNa}, I_{CaK}, I_K, I_{NaCa}, I_{K1}, I_{Kp}, I_{p(Ca)}, I_{Nab}, I_{Cab}, I_{NaK}, I_{nsNa}, I_{nsK})$, 4 fluxes $(I_{rel}, I_{up}, I_{leak}, I_{tr})$ and 11 gating variables $(m, h, j, d, f, f_{Ca}, X, X_i, K_1, K_p, f_{NaK})$.

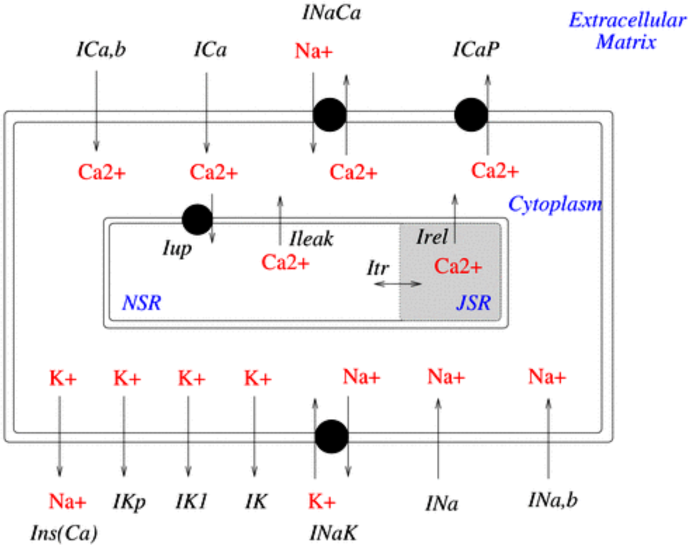


Fig. 2. A schematic diagram describing the ionic currents, pumps and exchangers that are captured in the LR-II model. The intracellular compartment is the Sarcoplasmic Reticulum (SR), which is divided into the two subcompartments, the Network SR (NSR) and the Junctional SR (JSR). Ca^{2+} buffers are present in both the cytoplasm and the JSR.

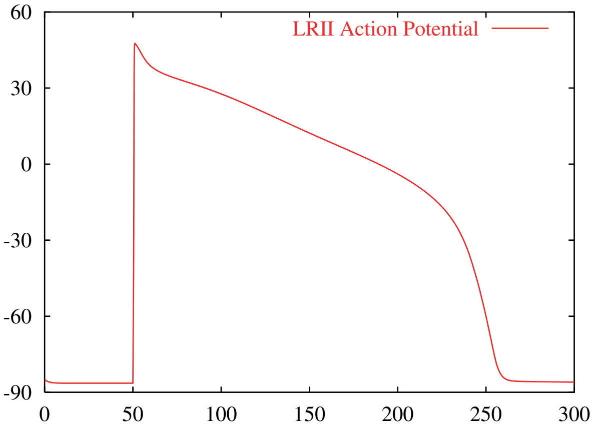


Fig. 3. An action potential simulated using the LR-II model.

The models mentioned above are all biophysically based, i.e. they attempt to describe the action potential by modeling the biophysics of the subcellular processes. Alternative non biophysical cell models have been developed which are typically much simpler and computationally more efficient. These types of models simulate the basic properties of an action potential without attempting to address the underlying processes. In such models the variables typically have no physical meaning. The most widely known of these models is the Fitzhugh–Nagumo model.[34,98] The Fitzhugh–Nagumo model is based on a cubic excitation model but also includes a recovery variable so both depolarisation and repolarisation may be modeled.

1.2.2. *Skeletal muscle cell models*

In contrast to cardiac action potential models, there are relatively few biophysically based models of the skeletal muscle action potential. The skeletal muscle cell models are also not yet as complex as the cardiac cell models. Adrian and co-workers were the pioneers in formulating biophysical models for skeletal muscle electrical activity.[1,2] Among the more recent models are the ones published by Henneberg and Roberge[62] and Wallinga *et al.*[145] These models take into account up to five different ionic currents (see Fig. 4).

Non-biophysically based models of the skeletal muscle action potential exist, for example that of Rosenfalck.[121]

1.2.3. *Gastrointestinal smooth muscle cell models*

To date no detailed biophysically based models exist that explicitly and separately represent the electrical activity of both smooth muscle and the Interstitial Cells of Cajal (ICCs). The most detailed biophysically based model to date is that of Miftakhov *et al.*[93] but this model makes no distinction between smooth muscle cells and ICCs, making it unsuitable for an investigation into the underlying mechanisms behind ECA initiation and propagation. The most advanced model that

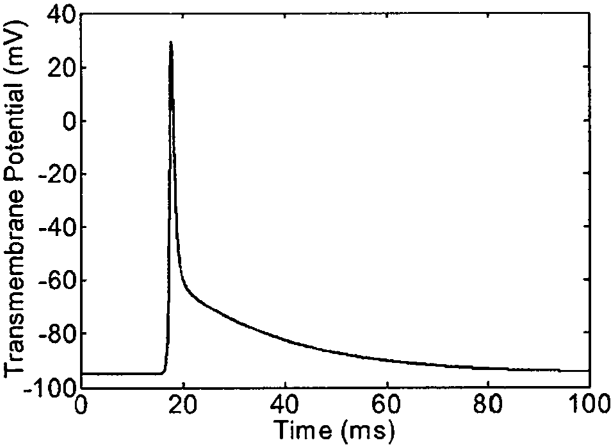


Fig. 4. Action potential generated using the Adrian *et al.* model (1973)[2] of frog sartorius muscle electrical activity.

does differentiate between the cell types appears to be that of Aliev *et al.*[4] in which modified Fitzhugh–Nagumo equations were used to describe both smooth muscle and ICC.

1.3. *Tissue and whole organ models*

Models of the muscle cell electrical activity can be incorporated into larger-scale continuum based models. The excitable tissue is treated as a syncytium and at any given point there is assumed to exist both intracellular and extracellular space. This coexistence of extracellular and intracellular information at every point in space is modeled by the well-known bidomain equations. The equations for the model are given below

$$\nabla \cdot (\sigma_i \nabla V_m) = -\nabla \cdot ((\sigma_i + \sigma_e)\nabla \phi_e), \tag{3}$$

$$\nabla \cdot (\sigma_i \nabla V_m) + \nabla \cdot (\sigma_i \nabla \phi_e) = A_m \left(C_m \frac{\partial V_m}{\partial t} + I_{ion} \right). \tag{4}$$

Here ϕ_e is the extracellular potential, σ_i and σ_e are the intra- and extra-cellular conductivities respectively, A_m is the surface to volume ratio of the cell membrane, C_m is a membrane capacitance per unit area, and I_{ion} is the appropriate cellular level model.

2. Physics

2.1. *Basic equations*

Bioelectric currents and the accompanying fields are electromagnetic phenomena and thus are ruled by Maxwell's equations[73]:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{5}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}, \tag{6}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{7}$$

$$\nabla \cdot \mathbf{D} = \rho, \tag{8}$$

where **E** is the electric field, **H** the magnetic field, **J** the current density, **B** the magnetic induction, **D** electric displacement and ρ charge density. The properties of the medium in which these are defined enter into the additional "constitutive" equations (assumed to be linear[115]):

$$\mathbf{D} = \epsilon \mathbf{E}, \quad \mathbf{B} = \mu \mathbf{H}, \quad \mathbf{J} = \sigma \mathbf{E} + \mathbf{J}_s. \tag{9}$$

In the problem addressed, the medium is a bounded volume *conductor*. The permeability μ is taken to be the value in a vacuum, μ_0. Macroscopically the medium can be assumed to be electrically neutral, and a separation of charges cannot occur. Microscopically, there is a charge built-up over capacitance of the cell membranes, which only in a non-steady state can produce a net current across the membrane, often called the *impressed* current, $\mathbf{J_s}$, resulting in a passive return current $\sigma\mathbf{E}$ in the volume conductor. Thus the main constitutive relation is a generalization of Ohm's law.

2.2. *The quasi-static approximation*

One important simplification can be made for sources and fields in biological media. In the so-called *quasi-static* approximation, it is assumed that there is no frequency dependence of the conductivity and over the length scales at hand no phase shifts occur. All changes in the fields are instantaneous with changes in the membrane current sources.[115] Thus, the partial derivatives in Eqs. (5) and (6) can be set to 0. As a result, since $\nabla \times \mathbf{E} = 0$, the electric field can be expressed by a scalar potential

$$\mathbf{E} = -\nabla \Phi, \tag{10}$$

and, taking the divergence of Eq. (6):

$$\nabla \cdot \mathbf{J} = 0, \tag{11}$$

and of the left- and right-hand side of third equation in (9) (substituting (10) and (11)) we get:

$$\nabla \cdot (\sigma \nabla \Phi) = \nabla \cdot \mathbf{J}_s = -I_v, \tag{12}$$

which is Poisson's equation, with I_v the impressed volume current density. σ is kept within the gradient operator here to allow for an anisotropic conductivity, in which case it is a tensor rather than a scalar. Note that in the volume outside the region where sources exist the potential is defined by Laplace's equation:

$$\nabla \cdot (\sigma \nabla \Phi) = 0.$$

Similarly, since $\nabla \cdot \mathbf{B} = 0$, the magnetic field can be expressed in terms of a vector potential

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{13}$$

Under quasi-static conditions, the following gauge is used[54]

$$\nabla \cdot \mathbf{A} + \mu_0 \sigma \Phi = 0. \tag{14}$$

In view of $\mathbf{H} = (1/\mu_0)\mathbf{B}$, substituting (13) into (6) leads to

$$\begin{aligned} \nabla(\nabla \cdot \mathbf{A}) - \nabla \times (\nabla \times \mathbf{A}) &= \mu_0 \mathbf{J}, \\ -\mu_0 \sigma \nabla \Phi - \nabla^2 \mathbf{A} &= \mu_0 \mathbf{J}, \\ \mu_0 \sigma \mathbf{E} - \nabla^2 \mathbf{A} &= \mu_0 \mathbf{J}, \\ \mu_0 (\mathbf{J} - \mathbf{J}_s) - \nabla^2 \mathbf{A} &= \mu_0 \mathbf{J}, \end{aligned}$$

where use is made of the constitutive equations (9). So for the vector potential we are left with a vector form of Poisson's equation:

$$\nabla^2 \mathbf{A} = -\mu_0 \mathbf{J}_s. \tag{15}$$

2.3. *Solutions in infinite homogeneous media*

2.3.1. *Electrical potential*

For the electric potential at observation point $\mathbf{r}$, the solution of (12) in the case of an infinite homogeneous isotropic medium is

$$\Phi_\infty(\mathbf{r}) = \frac{1}{4\pi\sigma} \int_V \frac{-\nabla' \cdot \mathbf{J}_s(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, dV' \tag{16}$$

$$= \frac{1}{4\pi\sigma} \int_V \frac{I_v(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, dV', \tag{17}$$

where the integration is over the volume V in which sources exist. Different source models may lead to different specifications of I_v or $\nabla \cdot \mathbf{J}_s$, e.g. by formulating equivalent sources on the surface bounding the volume V. This will be addressed in the sections about specific applications. There is one model that is of particular interest in the general bioelectromagnetic source imaging problem: The equivalent current dipole.

Consider the Taylor's series expansion of $|\mathbf{r} - \mathbf{r}'|^{-1}$ with respect to $\mathbf{r}'$:

$$\begin{aligned} \Phi(\mathbf{r}) &= \frac{1}{4\pi\sigma} \int_V \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \left[x' \frac{\partial}{\partial x} + y' \frac{\partial}{\partial y} + z' \frac{\partial}{\partial z} \right] \left(\frac{1}{r} \right) I_v(\mathbf{r}') dV' \\ &= \sum_{n=0}^{\infty} \Phi_n(\mathbf{r}). \end{aligned}$$

The integral in the first term in this expansion ($n = 0$),

$$\Phi_0(\mathbf{r}) = \frac{1}{4\pi\sigma r} \int_V I_v(\mathbf{r}') dV',$$

is the total current in the volume which for a biological source will always be zero. The second term is:

$$\Phi_1(\mathbf{r}) = -\frac{1}{4\pi\sigma}\nabla\left(\frac{1}{r}\right)\cdot\int_V \mathbf{r}' I_v(\mathbf{r}')dV'. \tag{18}$$

This represents a *dipolar* field with dipole components defined by the integral, which drops off as r^{-2}. The third term, which will not be evaluated here (cf.[54]) constitutes a *quadrupolar* field and drops off as r^{-3}. Since for many applications the points at which Φ will be evaluated will be relatively distant from the sources — compared to the extent of the sources — the contribution of quadrupolar or higher order fields will be relatively small, and effectively the sources can be described by (a summation of) equivalent current dipoles.

2.3.2. *Magnetic field*

For the vector potential $\mathbf{A}$, the solution of (15) in the case of an infinite homogeneous isotropic medium is:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi}\int_V \frac{\mathbf{J}_s(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dV'.$$

With $\mathbf{B} = \nabla\times\mathbf{A}$ and some elementary calculus we get for the magnetic field

$$\mathbf{B}_\infty(\mathbf{r}) = \frac{\mu}{4\pi}\int_V \frac{\nabla'\times\mathbf{J}_s(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}dV'. \tag{19}$$

For the multipole expansion of the magnetic field we refer to Gulrajani.[54] As in the case of the electric potential, remote from the sources the magnetic field will be dominated by the field of an equivalent current dipole as well. The issue of (the existence of) pure magnetic dipoles in the bioelectromagnetic source imaging problem will not be addressed here.

2.4. *Solutions in finite inhomogeneous media*

Solutions presented in the previous section are valid for unbounded, homogeneous media only. In the more realistic setting of a volume V of conductivity σ bounded by a surface S, analytical solutions can be found for Eqs. (12) and (15) if these surfaces and volumes have relatively simple shapes, e.g. spheres. For the solution in generally shaped volumes numerical techniques have to be used. One such technique, the Finite Element Method (FEM), proceeds by addressing (12) and the boundary condition $\nabla\Phi\cdot\hat{n}$ on S (indicating that no net current can leave the bounding surface S) through a variational principle or Galerkin procedure (see the Mathematical Methods section). The other technique, the Boundary Element Method (BEM), makes use of integral formulations of (12), which are the subject of this section. An isotropic — scalar — conductivity σ is assumed here.

2.4.1. *Electrical potential*

Consider Green's second identity[73]:

$$\int_V \sigma(\Psi\nabla^2\Phi - \Phi\nabla^2\Psi)dV = \int_S \sigma(\Psi\nabla\Phi - \Phi\nabla\Psi)\cdot d\mathbf{S} \tag{20}$$

with Φ and Ψ well-behaved scalar functions ($\Phi(\mathbf{r})$ will be electrical potential at observation point $\mathbf{r}$ as before). With the choice of $\Psi = \frac{1}{|\mathbf{r}-\mathbf{r}'|}$ (and thus $\nabla^2\Psi = -4\pi\delta(\mathbf{r}-\mathbf{r}')$) we get:

$$\int_V \sigma\left[\frac{\nabla'^2\Phi}{|\mathbf{r}-\mathbf{r}'|} + 4\pi\Phi(\mathbf{r})\delta(\mathbf{r}-\mathbf{r}')\right]dV' = \int_S \sigma\left[\frac{\nabla'\Phi}{|\mathbf{r}-\mathbf{r}'|} - \Phi(\mathbf{r})\nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right)\right]\cdot\, d\mathbf{S}'.$$

Inserting (12) and boundary condition $\nabla\Phi\cdot d\mathbf{S} = 0$, we obtain

$$\Phi(\mathbf{r}) = -\frac{1}{4\pi\sigma}\int_{V_s}\frac{\nabla'\cdot\mathbf{J}_s(\mathbf{r}'))}{|\mathbf{r}-\mathbf{r}'|}\,dV' - \frac{1}{4\pi}\int_S \Phi(\mathbf{r}')\nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right)\cdot\, d\mathbf{S}', \tag{21}$$

where the volume integral can be restricted to the source region V_s. Note that the term with this volume integral is the infinite homogeneous medium solution (16) for the potential $\Phi_\infty(\mathbf{r})$. The second, surface-integral usually is interpreted as the term representing *secondary sources* that correct for the bounded nature of the volume conductor. The interpretation of

$$\nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right)\cdot\, d\mathbf{S}' = \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\cdot d\mathbf{S}' = d\omega(\mathbf{r},\mathbf{r}') \tag{22}$$

as the solid angle of the surface element $d\mathbf{S}$ at $\mathbf{r}'$ as observed from $\mathbf{r}$ shows that these secondary sources are equivalent to a dipole layer density on S with orientation normal to the surface at $\mathbf{r}'$ and strength proportional to $\Phi(\mathbf{r}')$.[73] The solid angle interpretation is at the core of many BEM implementations.

Generalization of (21) to a piecewise homogeneous volume conductor with N (multiply) contained internal surfaces S_i with conductivities σ_i for the potentials $\Phi(\mathbf{r})$ inside and on the bounding surface S_0 reads:

$$\Phi(\mathbf{r}) = \frac{\sigma_s}{\sigma_r}\Phi_\infty(\mathbf{r}) - \frac{1}{4\pi}\sum_{i=0}^{N}\frac{\sigma_i^- - \sigma_i^+}{\sigma_r}\int_{S_i}\Phi(\mathbf{r}')\frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\cdot d\mathbf{S}', \tag{23}$$

with

σ_s : The conductivity in the surface containing the sources,
σ_r : The conductivity at the observation point,
σ_i^+ : The conductivity just outside surface S_i,
σ_i^- : The conductivity just inside surface S_i.

Here the additional boundary condition that the normal component of the current density at internal surfaces S_i is continuous is used.

2.4.2. *Source model free solutions for the electric potential*

If a surface S_s bounding the volume containing all active sources can be defined, then in Green's identity (20) the source-free volume can be used, for which Laplace's equation $\nabla \cdot (\sigma \nabla \Phi) = 0$ holds. The equivalent surface integral is then over both the outer surface S_0 and the source-bounding surface S_s (with the appropriate sign). Again taking $\Psi = \frac{1}{|\mathbf{r}-\mathbf{r}'|}$ and boundary condition $\nabla\Phi \cdot d\mathbf{S_0} = 0$ we obtain:

$$\begin{aligned}\Phi(\mathbf{r}) = &-\frac{1}{4\pi}\int_{S_s} \frac{1}{|\mathbf{r}-\mathbf{r}'|}\nabla'\Phi(\mathbf{r}') \cdot d\mathbf{S}' \\ &+\frac{1}{4\pi}\int_{S_s} \Phi(\mathbf{r}')\nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) \cdot d\mathbf{S}' \\ &-\frac{1}{4\pi}\int_{S_0} \Phi(\mathbf{r}')\nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) \cdot d\mathbf{S}'. \end{aligned} \tag{24}$$

By evaluating $\Phi(\mathbf{r})$ for $\mathbf{r}$ on S_0 and S_s separately, one derives a set of equations implicitly relating potentials on S_0 and on S_s to potentials and gradients of potentials on S_s. This set of equations can be solved to obtain an explicit relation between potentials on S_0 and ${S_s}^8$:

$$\Phi_0(\mathbf{r}) = \int_{S_s} T(\mathbf{r},\mathbf{r}'))\Phi_s(\mathbf{r}')dS. \tag{25}$$

2.4.3. *Magnetic field*

For the solution for the magnetic field inside bounded piecewise homogeneous volume conductors, it is convenient to consider the (continuous form of the) Biot–Savard law:

$$\mathbf{B}(\mathbf{r}) = \frac{\mu}{4\pi}\int_V \mathbf{J}(\mathbf{r}') \times \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\, dV'. \tag{26}$$

Here $\mathbf{J}$ is the total current density flowing in the bounded inhomogeneous volume V, $\mathbf{B}$ is the magnetic field *outside* V, the permeability μ is the same inside and outside V. With $\nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) = \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}$, inserting (9) (i.e. $\mathbf{J} = -\sigma\nabla\Phi + \mathbf{J}_s$), and splitting V into its N piecewise homogeneous constituents, we get

$$\begin{aligned}\mathbf{B}(\mathbf{r}) = &\frac{\mu}{4\pi}\int_V \mathbf{J}_s \times \nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) dV' - \frac{\mu}{4\pi}\sum_{i=0}^{N}\int_{V_j} \sigma_i \nabla'\Phi(\mathbf{r}') \\ &\times \nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) dV'.\end{aligned}$$

The first term on the right hand side is equivalent to the expression for the magnetic field due to the impressed currents $\mathbf{J}_s$ in an infinite homogeneous medium $\mathbf{B}_\infty(\mathbf{r})$.[54] Using the vector identity $\nabla \times (\Phi\mathbf{a}) = \nabla\Phi \times \mathbf{a} + \Phi\nabla\times\mathbf{a}$, and noting that $\nabla\times\nabla\Psi = 0$

for any Ψ, the above expression reduces to

$$\mathbf{B}(\mathbf{r}) = \mathbf{B}_\infty(\mathbf{r}) - \frac{\mu}{4\pi}\sum_{i=0}^{N}\int_{V_i}\sigma_i\nabla' \times \Phi(\mathbf{r}')\nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) dV. \tag{27}$$

We can now transform each volume integral to the proper surface integral using the identity $\int_v \nabla \times \mathbf{a}\, dv = -\int_s \mathbf{a} \times d\mathbf{s}$ (and again expressing $\nabla'\left(\frac{1}{|\mathbf{r}-\mathbf{r}'|}\right) = \frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}\Big)$:

$$\mathbf{B}(\mathbf{r}) = \mathbf{B}_\infty(\mathbf{r}) - \frac{\mu}{4\pi}\sum_{i=0}^{N}(\sigma_i^- - \sigma_i^+)\int_{S_i}\Phi(\mathbf{r}')\frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3} \times d\mathbf{S}. \tag{28}$$

Thus, the magnetic field can be expressed as the sum of the contributions due to *impressed* sources in a infinite homogeneous medium and that of *secondary* sources that are the same as those for the electrical potential (cf. Eq. (23)). Two important properties of the magnetic field due to current sources in bounded volume conductors that can be derived from this are[56]:

- In a spherically symmetric conductor the second term on the right-hand-side of (28) vanishes: *The magnetic field is determined by the impressed currents only.*
- In this situation the radial component of the impressed current produces no magnetic field. *Magnetic field measurements will reflect tangentially oriented dipolar source activity only.*

3. Signal Acquisition

A necessary prerequisite in source imaging is the acquisition of the bioelectric signals originating from the sources themselves. This is largely an engineering problem, complicated to a degree by the fact that one is dealing with human subjects which can impose certain practical, logistical and ethical limitations. Fundamentally, there are two different types of signal that one might seek to measure, namely an electrical signal (in the micro- to milli-voltage range) or a magnetic signal (typically in the pico- to nano-Tesla range). We discuss each of these in turn.

3.1. *Recording electrical signals*

Generally one is seeking to record many channels of the bioelectrical signal in the nearest neighborhood of the actual sources. All applications therefore require appropriate hardware and sensors/electrodes to achieve this. The question of how many channels is somewhat application dependent, but is also an open issue. Whilst, for instance, it has been shown[86,50] that even with optimal recording conditions, no more than 20 or so independent pieces of information are present when recording densely sampled ECGs (known as body surface potential mapping) there is still the question of where are the optimal locations to obtain this information, in both normal and diseased conditions. Attempts have certainly been made to answer this question.[86] It is clear that in the cardiac case, these recording sites on the body surface are not in the same place in all situations, and current researchers using densely sampled body surface ECGs for source localization have a range of sampling sizes and locations (e.g. 64[66,128], 256[99], 384[119]). In the brain field, the issue is also still open, with ranges from 3 to 100s of electrodes used for acquisition. For non-invasive recording of gastric smooth muscle the standard Electrogastrogram contains three leads, while for skeletal muscle no standard exists and recordings are made from a range of different setups.

In all applications, the recording devices must be capable of multichannel recording. With a low channel count, many potential devices can deliver this, but for higher channel counts, more specialized hardware is required with many of these systems being developed in the research labs of various Universities, some of which are now being sold commercially.[13,140] Sampling frequency is also an issue to be considered. For cardiac applications, typically 0.05 Hz to 500 Hz is the recording band. Due to the Nyquist sampling theorem, this means that upper frequency limits of 1000 Hz or higher are required of the device. Recordings from neural or smooth muscle activity typically have much lower frequencies — for instance the slow wave of the gastric smooth muscle is normally at 0.05 Hz.

Efforts should also be made to ensure the highest possible signal quality. Shielded cables, appropriate grounding, active noise cancellation (driven shields, right leg drive) etc. all aid in this. Vitally important is the recording sensor itself, which can be either active or passive. Passive electrodes are certainly the most common and to minimize recording noise, the use of conductive gel and other skin preparation is desirable. Skin preparation typically involves washing with alcohol or other appropriate liquid to remove any skin oils and dead skin, and gel also aids in this. Sometimes shaving is also required (if practical). The aim of these measures is to decrease the skin impedance, which results in the electrode/skin interface being less susceptible to noise. Ideally 5 kΩ or less is desirable for ECG. However, with EEG recordings with an unshaven scalp this can be unrealistically low — and values an order of magnitude higher can be experienced.

Active electrodes contain amplification within the electrode itself, which can help reduce the electrode/tissue noise problems. Many use such electrodes dry (gelless) and it is a claimed feature of active electrodes, although others still recommend the use of gel.[14]

The ultimate aim here is to acquire as many channels as the user feels is necessary to properly characterize (and capture the details of) the sources being studied with signals of the highest possible quality. At best one is dealing with signals in the mV range (intracardiac signals can be tens of millivolts, ECG signals a few millivolts) and often lower (EEG signals are tens of microvolts). A measure of the size of the signal relative to background noise is given by the Signal to Noise Ratio (SNR). The ratio is usually measured in decibels (dB). If the incoming signal strength is V_s, and the noise level (in the same units as V_s) is V_n, then the signal-to-noise ratio (in decibels) is given by the formula $S/N = 20\log_{10}(V_s/V_n)$. If $V_s = V_n$, then the SNR $= 0$. In this situation, the signal borders on unreadable, because the noise level severely competes with it. Ideally, V_s is greater than V_n, so the SNR is positive. As an example, suppose that $V_s = 10.0\,\text{mV}$ and $V_n = 1.00\,\text{mV}$. Then the $\text{SNR} = 20\log_{10}(10.0) = 20.0\,\text{dB}$ which results in the signal being clearly readable. If the signal is much weaker but still above the noise — say 1.30 mV — then the $\text{SNR} = 20\log_{10}(1.30) = 2.28\,\text{dB}$ which is a marginal situation.

In optimal recording conditions, noise levels can be as low as a few μV and 10–20 μV is often considered good. For ECG signals this results in high SNRs, but for EEG and other types of activity, the signal size is comparable to this level of noise. Signal processing techniques are often then used (e.g. beat averaging) to improve signal quality, although this is not appropriate for one-off type events. For inverse studies, one must know where the electrodes are. This can be achieved by using standard arrangements of electrodes and/or obtaining structural images of the patient with the electrodes (or other markers) in place.

3.2. *Recording magnetic signals*

The changing electrical potential of the bioelectric source creates both external electrical and magnetic fields. With the advent of Superconducting QUantum Interference Devices (SQUIDs) the recording of the external magnetic fields has become a reality. This provides a non-contact way of recording activity generated by the bioelectric source. The magnetic field generated by a single neuron is almost negligible; thus, when several thousands of nearby cells are synchronously active, the summated extracranial magnetic field typically achieves a magnitude of only a few hundred femto Tesla (1 femto Tesla = 10^{-15} Tesla). Even the strongest neuromagnetic signals, those associated with epileptic spikes, are only a thousand femto Tesla (i.e. in the order of 10^{-13} Tesla in magnitude). This is still more than one billion times smaller than the earth's steady magnetic field and the noise fields generated by even distant moving metal objects (e.g. cars and elevators) and power lines. To reduce the amount of magnetic noise reaching the biomagnetometer, the system is operated in a magnetic and radiofrequency shielded room made of mu-metal and aluminum. The recording dewar contains magnetic detection coils which are continuously bathed in liquid helium to superconducting temperatures of $-269°\text{C}$ (4.2 K). Some people have reported the uses of SQUIDs in unshielded rooms[22] these SQUIDs having been designed with special noise cancellation devices in built.

The number of channels to use is, again, an open question — and channel counts continually increase. There are MEG recording systems commercially available with more than 300 channels offering sampling at up to 8 kHz, together with the ability to simultaneously record 64 EEG signals as well.[101]

4. Mathematical Methods

For all imaging problems under consideration, there is an initially unknown spatiotemporally elaborated bioelectromagnetic source distribution $g(x,t)$ (a vector or scalar valued function), where x is a spatial variable, and t is the temporal variable. Remote sensing provides data in the form of $h(y,t)$, where y varies over spatial locations accessible to sensor placement. There are both deterministic and stochastic components in the relationship between g and h, generally expressed as in the additive model

$$h = Fg + \nu, \tag{29}$$

where F is some operator and ν is noise.

- The *Forward Problem* is to determine (or estimate) h given g. In practical terms, this is equivalent to constructing F, based on physical and geometrical considerations inherent in the setting.
- The *Inverse Problem* is to determine (or estimate) g given h. Note that this presupposes F is given — i.e. the Forward Problem has already been treated.

Since F (as provided by the forward problem treatment) is undoubtedly noise-corrupted, noise ν on the right-hand-side of (29) cannot be considered entirely independent of source g — a complication that is difficult to deal with or even quantify, and is often ignored.

The basic relationship between bioelectric sources and their generated electrical potentials is contained in the field equation presented in the prior section, that we briefly recap in this paragraph: Calculations and measurements demonstrate that capacitive, inductive, and electromagnetic propagative effects in the body are negligible in this problem. Hence, current density $\mathbf{j}$ at any location in the body is the sum of Ohmic and source terms,

$$\mathbf{j} = \sigma \mathbf{E} + \mathbf{j}_s, \tag{30}$$

where σ is the tissue conductivity tensor (which is independent of field strength for frequencies important in biological sources), $\mathbf{E}$ is electric field, and $\mathbf{j}_s$ is "impressed" (or source) current — referrable to the transmembrane currents in excitable tissue. Charge conservation (the continuity condition) under these quasi-electrostatic conditions requires that the current be divergenceless. Hence,

$$\nabla \cdot \mathbf{j} = \nabla \cdot (\sigma \mathbf{E}) + \nabla \cdot \mathbf{j}_s = 0.$$

Quasi-electrostatic conditions also imply that the electric field $\mathbf{E}$ is the (negative) gradient of a potential ϕ. Thus, the above can be taken to be a Poisson equation,

$$L[\phi] \equiv \nabla \cdot (\sigma \nabla \phi) = \nabla \cdot \mathbf{j}_s, \tag{31}$$

where the linear operator $L[\cdot]$ is understood to include the context of the applicable boundary conditions, e.g. the vanishing of the normal component of current density on the body surface (the air around the body is insulating). An analogous expression can be derived relevant to magnetic field data.

Evidently, the Forward Problem is to "invert" the linear operator $L[\cdot]$ so as to compute potential ϕ given the source $\nabla \cdot \mathbf{j}_s$ (a typical problem in the theory of partial differential equations).

Note that knowledge of $L[\cdot]$ requires knowledge of the the spatial variation of conductivity σ over the complicated body geometry. This is most pronounced in the case of the heart, which (as noted earlier) is composed of spiraling fibers, transversely linked to each other via low resistance "gap junctions". Accordingly, conductivity is markedly different along versus transverse to the local fiber axis, and the degree of anisotropy is quite different in the local extracellular versus intracellular spaces ("unequal anisotropy"). One immediate problem is that cardiac muscle conductivity is not scalar. In particular, resistance is lowest along the fiber axis — but excitation also occurs transverse to the fiber axis in a higher resistance direction. Cleavage planes also exist in cardiac tissue meaning that there is a preferred direction associated with the degree of tightest transverse coupling of the fibers. Thus σ is really orthotropic in cardiac tissue. Significant conductivity variations exist in other tissues, for example the gastric wall has layers of transverse and circular muscle fibers.

The Inverse Problem is to compute $\nabla \cdot \mathbf{j}_s$ (or some feasible "equivalent" source representation) given some knowledge of ϕ (e.g. on the body surface).

Forward and inverse problems are fundamentally different for the following reasons.[53] Differential operators such as $L[\cdot]$ in (31), are "unbounded" — that is, there exist functions of unit norm that are mapped by $L[\cdot]$ to functions of arbitrarily large norm. For example, consider that the operator d^2/dx^2 over functions on interval $[-1, 1]$ with homogeneous boundary conditions sends the function $\sin n\pi x$ to the function $n^2\pi^2 \sin n\pi x$, whose norm is higher by the factor $n^2\pi^2$, for arbitrarily large n. As is the case for the operator d^2/dx^2, the sequence of eigenvalues of $L[\cdot]$ in (31) tends to infinity. This means that the sequence of eigenvalues of its *inverse* will tend to zero. To solve the forward problem (i.e. the problem of computing ϕ given the source $\nabla \cdot \mathbf{j}_s$), one in fact applies the "inverse" of the differential operator $L[\cdot]$. The application of this inverse to any bounded data function will be a stable procedure — because the inverse operator's amplification of the data function component in any eigensubspace will be bounded, and the amplitude of such components will tend to zero for the higher-order subspaces (since the eigenvalues of $L[\cdot]$ tend to infinity, the eigenvalues of its "inverse" will tend to zero). If the function we begin with is noise-corrupted, the noise components in the higher-order subspaces will in fact be progressively attenuated when the inverse operator is applied.

In other words, the forward problem is that of integrating a partial differential equation, and integration is a stable procedure. Thus the forward problem is well-posed. It also follows that the inverse problem is "ill-posed". That is, amplification of components in the higher order eigenspaces (typically noise dominated) will receive unbounded amplification. This is analogous to the instability of differentiation of a noise corrupted function.

4.1. *The forward problem*

To solve (31), one could look for a function $G(x, y)$ such that

$$L[G(x, y)] = \delta(x - y), \tag{32}$$

subject to the condition that for y on the boundary, $\partial G(x, y)/\partial \mathbf{n} = \mathbf{0}$. In that case, it is clear that

$$\phi(y) = \int_X G(x, y)\nabla \cdot \mathbf{j}_s(x)\, dX, \tag{33}$$

(i.e. just insert (33) into (31) and use (32) — and also note that the homogeneous boundary conditions will automatically be satisfied). In other words, one simply finds the potential that would be induced by a point source at an arbitrary location, and then one can solve the problem for an arbitrary source distribution via superposition. But it is evident that such superposition will not work unless that boundary conditions are homogeneous.

The way to take inhomogeneous boundary conditions into account is as follows: Solve the homogeneous equation $L[\phi] = 0$ with the specified inhomogeneous boundary conditions, and add this solution to the above integral (33). This gives a solution satisfying the inhomogeneous equation with inhomogeneous boundary conditions. Interestingly, the latter can be treated as a second inhomogeneous equation with homogeneous boundary conditions, where the source is actually the boundary conditions themselves. Typically, the Green's function for this latter case is highly related to the Green's function computed for the former inhomogeneous problem.[78] This sort of thinking is important for treating boundary element formulations of the inverse electrocardiography problem in a medium of inhomogeneous conductivity.

Thus, we need only compute the Green's function to solve the forward problem. If the body was an infinite homogeneous isotropic medium, it is easy to show that $G(x, y) = 1/\|x - y\|^2$ (as in the preceding Physics section). Because of the inhomogeneity of the medium, and irregular geometry, we must use a general numerical method. Choices are:

- finite difference,
- finite element,
- boundary element.

4.1.1. *The finite difference approach*

The ingredients are as follows:

- Place a grid on the volume
- Replace the partial derivatives by their finite difference counterparts, e.g. in two dimensions

  $$\frac{\partial^2 \phi}{\partial x^2} \to \frac{\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j}}{(\Delta x)^2},$$

  etc.

- One evidently replaces the continuous operator equation with a matrix equation,

  $$L[\phi] = f \to \mathbf{K}\vec{\phi} = \vec{f}$$

  with $\vec{\phi} = (\phi_k)$, and where we have made the three dimensional array (e.g. $\{\phi_{i,j}\}$ with $i = 1, \ldots, q$) into a vector by some appropriate means (e.g. $\phi_{i,j} \to \phi_k$ with $\phi_{i,j} = \phi_{jq+i}$).

In the bioelectric problem, the grid points can be imagined to be connected to their neighbors by intervening resistances, and Kirchoff's current law at each point results in simultaneous equations to be solved.

A regular grid can lead to geometry imprecisions in regions where ϕ is expected to vary rapidly in space (the finite element approach is potentially more accurate than the finite difference approach for a given mesh density). Other disadvantages of finite differences compared with finite elements are

- it cannot take advantage of the weak formulation, for which one less is derivative required (see below),
- for irregular grids a symmetric operator will not in general lead to a symmetric matrix.

However, the finite difference method is a viable choice in forward electroencephalography and forward electrocardiography.[81,143] A regular grid is not an absolute requirement (but abandoning it requires care, as a naive algorithm might not then converge).

4.1.2. *The finite element approach*

The first step is to produce a general form for the unknown function (potential) in terms of "finite elements".

- An appropriate tesselation of the domain is made (e.g. bricks or tetrahedrons).
- These can be made to conform to the geometry well, and can be given greater density where more rapid variation of the function is expected.

Now let (ϕ_i) be the vector whose components are the (unknown) potential at each node i. We define *shape functions*, $\{N_i(x)\}$, such that $N_i(x)$ is unity at $x = i$th node, and zero in any finite element which does not have i as a node (x, of course, is a variable over 3-space). Also, $N_i(x)$ is continuous. Such requirements (and the possibility of higher derivative continuity of the solution) are nicely fulfilled by the Lagrange polynomials (the $n - 1$ degree polynomial going through n specified points), or splines (cubic Hermite polynomials, for example). Then we set our discretized potential to be

$$\hat{\phi} = \sum_i \phi_i N_i(x). \tag{34}$$

Note that $\hat{\phi}$ will be continuous, and capable of approximating any function.

So now we need an appropriate equation on which to apply the discretization. To formulate it we use the residual

$$R \equiv L[\hat{\phi}] - f \neq 0. \tag{35}$$

For the Galerkin method (one of several possible numerical schemes), we insist that for all i,

$$\int_V R N_i dV = 0, \tag{36}$$

where V is the volume of interest. Thus, the requirement that the residual be orthogonal to the space spanned by the shape (or basis) functions determines the system. Note that

$$L[\hat{\phi}] = \sum \phi_j L[N_j(x)],$$

so substituting this into (35), it is seen that (36) gives

$$\sum_j \phi_j \int_V L[N_j(x)] N_i(x) dV = \int_V f N_i(x) dV. \tag{37}$$

Now, the integral inside the sum on the left-hand-side above can be integrated by parts — which in this case represents the divergence theorem (recall $L[\cdot]$ is defined by (31)). For each term we get a volume integral involving $\nabla N_i \cdot \nabla N_j$ and a surface integral bounding the support of N_i and N_j. All these surface integrals will cancel each other (due to common surface orientation differences) except for those with elements on the boundary of V itself. Ultimately (37) leads to an equation of the form

$$K(\phi_i) = (f_i).$$

Now, we have a format where the boundary conditions are explicitly inserted *and* the weak formulation (i.e. the integration by parts yields the "weak" formulation — the necessity of solution smoothness is weakened, and the solution need not be as differentiable as the differential equation would imply).

The above formulation already contains the (natural) boundary conditions. To deal with these, it is only necessary to use first order (i.e. linear) shape functions. However, it might be desired to also satisfy "essential" boundary conditions — e.g. that the potential and current density be continuous throughout the domain. One can use splines or Lagrange polynomials of higher order so as to satisfy these.[39,117]

The book-keeping involving global and local coordinate systems is extensive, and one ends up with multiply indexed expressions like

$$\hat{\phi} = \sum N_a(\xi, \eta, \zeta) \phi^h(x_a^e)$$

and a stiffness matrix for each finite element as

$$K^e = \left(\int_{V_e} \sigma_{ij} \frac{\partial N_b}{\partial x_i} \frac{\partial N_a}{\partial x_j} dV \right)_{ij}, \tag{38}$$

which can then be assembled into the global stiffness matrix.

4.1.3. *The boundary element approach*

As derived in the Physics section, in a medium of scalar conductivity where regions of differing uniform conductivity are separated by distinct interfaces, the relevant integral equations are over these boundary surfaces. Discretization for numerical analysis then involves tesselation of these surfaces along the lines of the volume tesselations of the finite element method. This defines the "boundary element method". Recent implementations of the boundary element method with piecewise homogeneous conductivity are found in Refs. 39 and 118.

4.1.4. *A note regarding Green's functions*

The Green's functions relevant to (31) in a finite body of heterogeneous conductivity are never computed explicitly. Rather, either simpler (free-space) Green's functions are used (referrable to (24) and (25) in a boundary element context) or the governing equations are solved without specific recourse to any Green's functions at all (e.g. as in a finite difference or finite element formulation). The resultant transfer matrices embody the Green's functions for the appropriate problem, but only in and indirect and implicit manner. Thus, for the problem of determining the relationship between body surface potentials and epicardial potentials (for example), V is the body volume external to the epicardial surface, and the finite element derived matrices resulting from (38) can be arranged in a block matrix system like

$$\begin{pmatrix} K_{TT} & K_{TV} & K_{TE} \\ K_{VT} & K_{VV} & K_{VE} \\ K_{ET} & K_{EV} & K_{EE} \end{pmatrix} \begin{pmatrix} \Phi_T \\ \Phi_V \\ \Phi_E \end{pmatrix} = \begin{pmatrix} J_E \\ 0 \\ 0 \end{pmatrix},$$

where the components of J_E are the normal current densities on the epicardial surface, the components of Φ_V are the potentials inside the thoracic volume, the components of Φ_T are the potentials on the surface of the thorax, and the components of Φ_E are potentials on the epicardial surface.[76] We can use the last two equations implied by the above expression to ignore the unknown J_E, eliminate Φ_V, and thus derive the transfer matrix between Φ_T and Φ_E (analogous to (25)). This, of course, applies to the inverse EEG problem as well.

4.2. *The inverse problem*

Computation of potentials given the sources can be formally reduced to that of Green's function computation. It is a "well-posed" problem, in that one asks for an integration (i.e. solving a PDE) — a smoothing procedure.

On the other hand, the source computation problem is that of solving an integral equation, where the kernel is the Green's function. Inverting the (well-behaved) integral operator (sort of a differentiation) is *not* well posed — i.e. it tends to be an unstable procedure with potentially great noise amplification (i.e. if one has a square integrable Green's function). Treatment therefore requires care. Integral equations that can be stably solved have (in some sense) an intrinsically singular nature. For those without this singular nature, the process of solving the integral equation is akin to the unstable act of differentiation. Thus, the act of taking a derivative is equivalent to solving a particular integral equation, e.g. if $z(t)$ is the nth order derivative of $u(t)$, then it is the solution of the equation

$$u(t) = \int_0^1 \frac{H(t-\tau)}{(n-1)!}(t-\tau)^{n-1} z(\tau) d\tau,$$

easily verified via integration by parts, where $H(s)$ is the Heaviside function (zero for $s < 0$ and unity for $s > 0$).

For an inhomogeneous linear differential equation with homogeneous boundary conditions and Green's function $f(x, y)$, the relationship of source to field is

$$h(x) = \int_Y f(x, y) g(y) dY,$$

where $h(x)$ is field, and $g(y)$ is source (i.e. after discretization this is $h = Fg$, with h, g vectors and F the transfer matrix). Here, $f(x, y)$ is square-integrable, which means that we can write its "singular value expansion"[61] as

$$f(x, y) = \sum_n f_n f_{L,n}(x) f_{R,n}(y), \tag{39}$$

where $\{f_{L,n}(x)\}$ and $\{f_{R,n}(y)\}$ are the eigenfunctions of the operators $\int_X \int_Y f(x, y) f(x', y)(\cdot)\, dY\, dX'$ and $\int_Y \int_X f(x, y) f(x, y')(\cdot)\, dX\, dY'$ and $\{f_n\}$ is the set of eigenvalues. Thus,

$$\begin{aligned} h(x) &= \int_Y \left[\sum_n f_n f_{L,n}(x) f_{R,n}(y) \right] g(y) dy \\ &= \sum_n f_n \langle g(y), f_{R,n}(y) \rangle f_{L,n}(x), \end{aligned}$$

where "$\langle \cdot, \cdot \rangle$ is inner product. But simply expanding $h(x)$ as a Fourier series in $\{f_{L,n}(x)\}$, we get

$$h(x) = \sum_n \langle h(x), f_{L,n}(x) \rangle f_{L,n}(x).$$

The above two equations then imply that

$$g(y) = \sum_n \frac{\langle h(x), f_{L,n}(x) \rangle}{f_n} f_{R,n}(y). \tag{40}$$

Since f_n tends to zero as n gets large (this must be the case for a square-integrable kernel), if $h(x)$ is noise contaminated (even very slightly), we will generally not have

$$\langle h(x), f_{L,n}(x) \rangle = o(f_n).$$

Hence, the sum in (40) diverges (unbounded noise amplification).

Since we have a decay in singular values of $f(x, y)$ toward zero, the obvious thing to do is truncate the series on the right-hand side of (39) and insert the resulting estimate into (40). But how does one decide on where or how to truncate? Regularization is the process by which this problem is (hopefully) optimally dealt with.

4.2.1. *Linear estimation and regularization*

In reference to $h = Fg + \nu$, we are required to estimate g given h and F. It might first occur to us that a useful estimate for g would be the one which maximizes $p(h|g)$ — i.e. the choice that maximizes the conditional probability (density) that the measured h would occur given a particular candidate for g. This is known as the "maximum likelihood estimate". Assuming F has an inverse F^{-1}, and the noise is zero mean white Gaussian, this is given by $g_{ml} = F^{-1}h$ — because the likeliest noise vector is the zero vector (in this case, the maximum likelihood estimate coincides with the least-squares solution). If F is ill-conditioned, this is in general a very poor estimate for g, and is very unstable to noise variations.

But one could alternatively take the estimate for g to be the choice which maximizes $p(g|h)$ — i.e. the choice that has the greatest chance of being correct, given the observed noisy data h. This is referred to as the "maximum *a posteriori*" estimate, g_{map}. The relationship between the two conditional probability densities above is given by a version of Bayes Theorem: $p(g|h) = [p(h|g)p(g)]/p(h)$.

g_{map} has the great advantage of being stable to noise variations — but the disadvantage is that its calculation requires one to first supply nontrivial information concerning statistical properties of g (in fact, the entire "prior" probability density $p(g)$). However, if $p(g)$ is available and reliable, the methodology is referred to as "Bayesian". If one supplies some of the statistical properties as estimated from the given data h itself, the methodology is referred to as "empirical" Bayesian. The general approach is also referred to as "Statistical Regularization". Evidently, in discrete settings, the maximum likelihood estimate is equivalent to the Bayesian approach in a minimum information setting where every realization of g is considered equally likely to occur.

All of this suggests that once noise is introduced it is useful to carefully consider statistical notions. Specifically, each component of noise vector ν is a "realization" (outcome) of a "random variable". For our purposes, a random variable is an entity that associates a probability density value with every real number. From this, we can compute the probability that some realization of the random variable (outcome of a measurement) will yield a value falling in some given interval. Accordingly, the expectation $\mathcal{E}[\cdot]$ of some expression involving a random variable is the integral of the expression over all possible values of the random variable weighted by the probability density associated with each value. A "Gaussian" random variable α has a Gaussian probability density (the familiar bell-shaped curve), and is fully characterized by its particular expectation $\mathcal{E}[\alpha]$ and variance $\mathcal{E}[(\alpha - \mathcal{E}[\alpha])^2]$. Similarly, a zero mean Gaussian random vector w is a column vector of zero mean jointly Gaussian random variables. A zero mean random vector is further characterized by its autocovariance matrix $\mathcal{E}[ww']$, where superscript "$'$" denotes transpose. The autocovariance matrix describes the dependence between all different pairs of components of w. Similarly, the cross-covariance matrix of zero mean random vectors v, w is given by $\mathcal{E}[vw']$, and describes the mutual dependence of v and w.

Just as ν is a random vector, so too can g be considered to be an (unknown) realization of a random vector (the basic assumption of the Bayesian approach). In approaching $h = Fg + \nu$, a good objective is to find g_{opt} such that $\mathcal{E}[\|g - g_{opt}\|^2]$ is minimum (this being the "minimum-mean-square-error" estimate). If g and ν are realizations of independent zero mean Gaussian random vectors, g_{opt} is obtained via the Wiener filter. Under these conditions, the maximum *a posteriori* estimate g_{map} is equivalent to g_{opt}. This linear estimation procedure develops as follows.

A linear estimate of g is given by application of an "estimation matrix" M_{est} to data h, i.e. $g_{est} = M_{est}h$. Ideally, we desire the solution estimate

$$g_{opt} = M_{opt}h, \tag{41}$$

such that $\mathcal{E}[\|g - M_{opt}h\|^2]$ is minimum. Thus, it is sufficient to calculate M_{opt}. The way to proceed follows from the "Orthogonality Principle", which asserts that g_{opt} minimizes the mean-square-error when

$$\mathcal{E}[(g - g_{opt})h^t] = \mathbf{0}, \tag{42}$$

i.e. when the cross-covariance matrix of the "error of the estimate" $(g - g_{est})$ and the "data vector" h is the zero matrix (so that the error and the data have no dependence). Intuitively, the Orthogonality Principle assures that every bit of useful information is extracted from the data h in making the solution estimate g_{opt}. Substitution of (29) and (41) into (42) immediately gives

$$\mathcal{E}[gg^tF^t + g\nu^t] - \mathcal{E}[(M_{opt}Fg + M_{opt}\nu)(g^tF^t + \nu^t)] = \mathbf{0}. \tag{43}$$

Thus, assuming that g and ν are independent (i.e. $\mathcal{E}[g\nu^t] = \mathbf{0}$), (43) can be written as

$$C_gF^t - M_{opt}(FC_gF^t + C_\nu) = \mathbf{0},$$

where $C_g \equiv \mathcal{E}[gg^t]$ and $C_\nu \equiv \mathcal{E}[\nu\nu^t]$ are the autocovariance matrices of signal g and noise ν. Hence,

$$M_{opt} = C_gF^t(FC_gF^t + C_\nu)^{-1}. \tag{44}$$

Thus, g_{opt} is provided by (41), where M_{opt} is given by (44). Of course, it is assumed that we know the autocovariance matrices of signal and noise. Evidently, the Bayesian methods potentially allow (or require) introduction of a larger class of *a priori* information.

If $C_\nu = \sigma_\nu^2 I$ (i.e. if the noise is white), (44) becomes

$$M_{opt} = (F^t F + \sigma_\nu^2 C_g^{-1})^{-1} F^t, \tag{45}$$

since $(F^t F + \sigma_n^2 C_g^{-1}) C_g F^t = F^t (F C_g F^t + \sigma_\nu^2 I)$.

An alternative applicable formulation is provided by the maximum likelihood method in concert with a deterministic constraint (such as that the signal power is equal to some *a priori* value E). Maximum likelihood then corresponds to minimizing $\|Fg - h\|^2$ subject to $\|g\|^2 = E$ (assuming Gaussian noise). This is ultimately the technique of Lagrange multipliers, and leads to a linear estimation matrix of the form (45), with $\sigma_\nu^2 C_g^{-1}$ replaced by λI (where λ is the Lagrange multiplier).

A third alternative is Tikhonov regularization.[133] In this method, one chooses the estimate that minimizes

$$\|Fg - h\|^2 + \lambda R[g], \tag{46}$$

where $R[\cdot]$ is a non-negative functional, and λ is a "regularization parameter". If $R[g]$ is of the form $\|Lg\|^2$, where L is a matrix and $\|\cdot\|$ is Frobenius norm, application of the variational principle (setting the functional derivative of (46) to zero) leads again to (45), with $\sigma_\nu^2 C_g^{-1}$ replaced by λL. When L is the identity, this procedure is known as zero-order Tikhonov regularization (and first or second order if L is the gradient or Laplacian operator, respectively).

The question then becomes one of selecting a good choice of λ. There are several methods for accomplishing this. One popular choice results from constructing an "L-curve" (a plot of $\|Lg\|^2$ versus residual $\|h - Fg\|^2$ for different values of λ). The plots typically have the form of the letter "L", and the solution estimate (and regularization parameter) associated with the corner is chosen. In fact, for Gaussian noise, the solution which minimizes the error is found at the L-curve corner.[61]

For the choice of $L = I$, a similar sort of thing results from setting to zero the terms in (40) corresponding to the smallest singular values — with "small" as determined by the L-curve. This is known as Truncated SVD regularization (TSVD).

As stated earlier, a square-integrable (i.e. well-behaved) function (or image) must be such that its higher order Fourier coefficients tend to zero. In the presence of white noise, whose Fourier coefficients therefore do not tend to zero, it is then clear that higher order Fourier coefficients of data h are hopelessly noise-corrupted. The Wiener filter and Tikhonov regularization achieve stable results by removing any attempt at meaningful reconstruction of the high resolution components.

There is only one way out of the dilemma of resolution loss. If physiological constraints exist which effectively reduce the dimension of the solution space to be commensurate with the number of useful data Fourier coefficients, one can anticipate that it will be possible to preserve spatial resolution. For example, for the inverse electroencephalography problem (where one wishes to image the brain sources of the scalp electrical potentials), it might be true that only a single focus is responsible for inciting an epileptic seizure, and that this focus can be modeled as a single current source dipole located at some unknown location in the brain. In that case, one is searching for an entity with six degrees of freedom (reflecting its location, orientation, and magnitude). High spatial resolution could conceivably be possible assuming there are six or more data Fourier coefficients (with respect to the transfer matrix SVD-derived coordinate system) that are not dominated by noise.

At first, such an obvious constraint does not appear to be applicable to the heart, since the heart is not faithfully modeled as a small number of dipoles. However, a deeper look at the geometry reveals that such constraints do in fact apply for the "critical points" of ventricular activation — from which an activation map can be fashioned, in principle.[47]

4.2.2. *Incorporation of time in the regularization procedure*

The source imaging problem is actually posed as

$$h(x,t) = \int_Y f(x,y)g(y,t)dY,$$

where $h(x,t)$ is the time-varying data (e.g. body surface potentials), $f(x,y)$ is the Green's function, and $g(y,t)$ is the unknown time-varying source. As indicated above, one can expect to preserve resolution if one has the powerful geometrical constraint that the source is a point. In fact, using the "MUltiple SIgnal Classification" (MUSIC) algorithm,[124] it is possible to resolve several point sources whose activity overlap in time, so long as their time series are linearly independent of each other. The simple underlying idea is that if there are only a few point sources, then there will only be a few components of the singular value expansion of the spatiotemporal data $h(x,t)$ that are not due to noise. The noise components can presumably be recognized by their "uniformly" small singular values (if the noise power is sufficiently small, and the noise components are independent and identically distributed Gaussian random variables uncorrelated to the signal). In that case, a "signal space" spanned by the non-noise data spatial singular vectors can be constructed from the spatiotemporal data matrix. Any candidate for a point source which (after being operated on by F) does not map to a vector close to being in this signal space, can thereby be rejected as a possible source. In effect, there is a point source at y' (i.e. a multiple of $\delta(y-y')$) "if and only if" $f(x,y')$ is in the space spanned by the spatial singular functions of $h(x,t)$. This approach evidently relies in a fundamental way on the geometrical assumption that the sources are "nearly" zero dimensional. Such a supposition is quite relevant to the brain.

On the other hand, the point source constraint (other than in the context of the Critical Point Theorem[47]) does not pertain to the cardiac situation. Thus, based on the last subsection, it might seem like what we want to do is find the optimal Tikhonov solution for each distinct time point (independently of all other times), and call this our regularized solution for $g(y,t)$ (e.g. this is the supposition of the current approach found in Ramanathan *et al.* (2004)[120]). However, this is potentially a suboptimal thing to do because

- members of the time series $\{g(y,t)\}$ are typically correlated,
- one often notices "noise" in the reconstructed time series using that approach.

Thus, several methods to effect an intrinsically spatiotemporal regularization have been proposed.[11,18,132] Each of those approaches have *ad hoc* features. Instead, if one adopts a Bayesian perspective, and insists that the utilized prior (i) be invariant to any transformation which does not impact the known specifications of the setting, and (ii) is consistent with the data statistics, one arrives at a very specific spatiotemporal regularization procedure. It can be shown that the resulting method (in the setting where there is no prior knowledge of any specific temporal constraints) is as follows: Write

$$\begin{aligned} h(x,t) &= \sum_i h_i h_{L,i}(x) h_{R,i}(t) \\ &= \int_Y f(x,y) g(y,t) dY, \end{aligned} \tag{47}$$

where the sum is the singular value expansion of $h(x,t)$. Multiplying both sides of (47) by $h_{R,j}(t)$, and integrating over data acquisition interval T, we get integral equations for Fourier components of $g(y,t)$ with respect to $\{h_{R,j}(t)\}$, i.e.

$$h_j h_{L,j}(x) = \int_Y f(x,y)\gamma_j(y) dY, \tag{48}$$

$j = 1, 2, \ldots,$ with

$$g(y,t) = \sum_j h_{R,j}(t)\gamma_j(y). \tag{49}$$

It can be shown that the optimal solution estimate for $\gamma_j(y)$, obtained from treatment of the isolated equation (48), inserted for $\gamma_j(y)$ on the right-hand-side of (49), for $j = 1, 2, \ldots,$ gives the optimal solution estimate for $g(y,t)$ — assuming the unavailability of valid prior temporal constraints.[51]

A comprehensive treatment of (47) follows from the recognition that it is a "Partial Inverse Problem", i.e. an inverse problem where the defining operator $(\int_Y f(x,y)(\cdot)\, dY)$ does not address all of the variables of the unknown $g(y,t)$. This leads to the inevitability of symmetries in the prior, and a clear means for incorporation of data statistics in the estimation procedure.[52]

SOURCE CATEGORIES

5. The Brain

5.1. *The nature of the sources*

The main building blocks of the brain are the neurons. As is the case with the myocardial cells in the heart, and the smooth- and skeletal muscle cells, electrical current sources are generated by the cell membrane in a non-steady state. Neurons are excitable cells as well, and can generate action potentials. In fact, these action potentials are the carriers of information within the brain and between the brain and the rest of the body. However, it is not the action potentials that are the generators of the electric potentials and magnetic fields that can be measured on and over the scalp surface as the Electro- and Magneto Encephalogram (EEG and MEG).

The brain contains about 10^{10} neurons. They are found in various aggregates in both cortex and subcortical areas and in the cerebellum. The neuron consists of a cell-body called *soma* with tree-like extensions called *dendrites* that receive information and a single long trunk-like extension called *axon* that outputs information through action potentials. Connections between different neurons are through *synapses* that provide a one-way electro-chemical connection between the output axon of one neuron and the input dendrites of another.

If an action potential arrives at a synapse (through a process of diffusion of certain chemical agents called *neuro-transmitters*), the dendritic membrane can become depolarized, i.e. the resting membrane potential becomes less negative and more excitable, or hyperpolarized — more negative and less excitable. These inferred changes on the transmembrane potential are called Excitatory Post Synaptical Potentials (EPSP's) and Inhibitory Post Synaptical Potentials (IPSP's) respectively. IPSP's and EPSP's are relatively long lasting (10–200 ms). They can be considered as sub-threshold responses from a system similar to that described by the Hodgkin–Huxley equations (2).

The dendrites can receive input from the axons of various other neurons, each of which can generate either an EPSP or an IPSP. If enough EPSP's are present simultaneously (with respect to the relative long duration of these processes) they add up, and a fast (1–2 ms) action potential is generated that propagates from the soma to the output axon of the neuron.

An individual EPSP can be modeled by a current sink at the location of the synapse. The rest of the membrane acts as a (distributed) current source (for the IPSP source and sink are interchanged). Such a configuration has a non vanishing dipole moment. The currents involved with each individual EPSP or IPSP in the dendritric tree are of course extremely small. However, in neocortex, neurons called *pyramidal cells* are arranged in an ordered way with dendritic tree, soma and axon trunk aligned and perpendicular to the cortical surface (Fig. 5).[17] In addition to that, neighboring cells receive the same input. As a result, given the relative long duration, the semi-synchronous currents associated with individual EPSP's (IPSP's) may add up to produce a net current distribution that from a distance can be characterized by a current dipole moment, oriented perpendicular to the cortical surface. On the other hand, the characteristic *macroscopic* geometry of the neocortex with *gyri* and *sulci* (Fig. 5) can cause cancellation of simultaneous activity on both sides of a sulcus.

For the short and fast propagating action potentials, the duration is too short to allow for any effective addition. Also, the current distribution of an individual

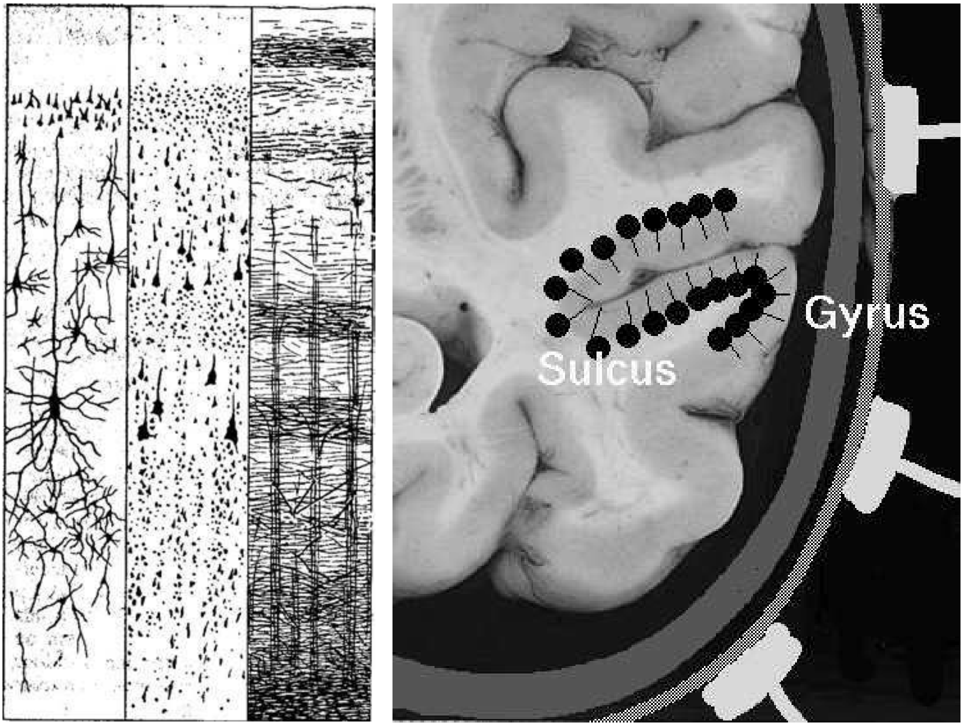


Fig. 5. Left: Ordered structure of the microscopic columns of pyramidal cells in neocortex shown in three different stainings, after Brodmann (1909).[17] Right: Macroscopic *Gyri* and *sulci* of neocortex, with oriented equivalent dipoles indicated.

propagating action potential is quadrupolar rather than dipolar, and thus decreases rapidly with distance. So, the EEG and MEG are determined by synchronous post-synaptical activity in the ordered structure of the columns of pyramidal cells in neocortex that can be modeled as a current dipole. Interestingly, recent research has shown that post-synaptical activity is also the major source of the signals measured in BOLD-fMRI.[83]

5.1.1. *Spontaneous and evoked activity*

The ensemble of neurons in the brain constitutes a bioelectromagnetic source that has a reportoire the size of which vastly exceeds that of the heart and other bioelectromagnetic sources of the body. As already pointed out by Helmholtz, reconstruction of any full 3D current distribution that is associated with this activity from remote measurements by means of EEG or MEG is impossible, since the associated inverse problem is not unique.[60] In order to arrive at any sensible approach to bioelectromagnetic imaging of the brain, severe restrictions on both the nature of the sources and of the data that are to be imaged are necessary.

The patterns most suited for bioelectromagnetic source imaging are those that can be assumed to be originating from a small number of areas of limited extent. But in addition to those patterns arising from localized activity (to be called "signal"), there will always be EEG and MEG components that cannot be attributed to a specific well localized site. This latter activity, which cannot be characterized as artifact or random noise, can be correlated and can have a frequency range from DC to 100 Hz. This is referred to as *background activity.* For spontaneous activity, if the patterns of interest have a high enough amplitude e.g. *an epileptic spike*, or have a narrow frequency content that allows for bandpass filtering such as some *brain rhythms*, episodes containing signal can be easily identified.

However, another class of signals consists of those that can be elicited by external stimuli — so called Evoked or Event Related Potentials, EP's and ERP's (or fields, in the case of MEG). In those situations, a trigger is available that allows for the averaging of data over an interval following the trigger. Assuming that background and (evoked) signal are uncorrelated, the averaging of the interval over N triggers will result in a reduction of the background level by a factor of $\sqrt{N}$.

5.2. *The nature of the volume conductor*

The EEG is measured on and over the scalp surface at a distance from the generators which for many cortical regions is not larger than 2 cm (for MEG the distance may be somewhat larger because of the distance of the measurement coils from the skin, and the location of the tangential sources, to which it is sensitive, that are located deeper in the sulci of the brain). In spite of this proximity, when compared to the inverse problem of electrocardiography, the inverse EEG/MEG problem is by no means less ill-posed. This is because of the low conductivity of the skull as compared to that of the skin and the brain itself. This configuration acts as a spatial low-pass filter: Spatial detail in the potential distribution at the cortical surface appears severely smeared on the scalp.

Since this low skull conductivity is such a determining factor in the EEG inverse problem, knowledge of its actual value and of the geometry and thickness of the skull, is of paramount importance. The fact that in a practical application this knowledge generally is not available at an individual level poses perhaps a bigger problem to the EEG inverse problem than the ill-posedness as such, since in the case of dipole localization it can lead to systematic errors of the order of centimeters.[68] Also, if the BEM is used for solving the forward problem, special approaches (the "Isolated Problem Approach" or IPA) are required in order to avoid numerical instabilities that arise from having compartments of significantly differing conductivity.[58,90] Literature values of the skull/scalp conductivity ratio show values ranging from 1/80 to 1/10, often related to the way in which conductivity is measured.[65] Also, skull thickness is difficult to estimate from MRI, whereas CT in individual cases is often unacceptable for both patients and volunteers, because of the radiation issues involved. Approaches in which an effective skull-scalp conductivity ratio is estimated from impedance tomography measurements, using the EEG electrodes and the same volume conductor geometry as for the inverse problem, may be the most promising.[45]

It is also this mostly unknown skull conductivity and thickness that constitutes the great advantage of MEG over EEG. Equation (27) can be interpreted as stating that $\mathbf{B}$ is to first order determined by the impressed currents J_s with a correction based on the volume currents in each compartment. It is clear that most volume currents flow in the compartment enclosed by the skull, and that current flowing in the skull and in the scalp (the latter of course ultimately causing a measureable EEG) are negligible. As a result only the geometry of the inner surface of the skull is relevant to the MEG, and this can be extracted fairly reliable from individual 3DT1 MR.[35]

5.3. *Methods for the inverse problem*

In the following a scalar function B rather than the vector $\mathbf{B}$ will be used when possible. This reflects the fact that in the vast majority of MEG systems used the recording coils are arranged as *gradiometers* that measure a gradient (either planar or axial) of the magnetic flux.

5.3.1. *Dipole localization*

Equations for the electrical potential and magnetic field generated by current dipoles in bounded, partially homogeneous volume conductors of arbitrary shape are given by Eqs. (23), (16) and (28), (19). The Boundary Element Method that can be used to solve these equations can be described as the computation of the linear operator L^0 that transforms infinite medium potentials $\Phi_\infty(\mathbf{r})$ (or fields $\mathbf{B}_\infty$) to the potentials (fields) $\Phi(\mathbf{r})$ that are actually measured: $\Phi = L^0\Phi_\infty$. Operator L^0 contains all conductivity and geometrical factors defining the volume conductor. For a current dipole $\mathbf{p}$ at location $\mathbf{r}'$, an analytical expression for $\Phi_\infty(\mathbf{r})$ that follows from (18) is available:

$$\Phi_\infty(\mathbf{r}) = \frac{1}{4\pi\sigma_s}\mathbf{p}\cdot\frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}.$$

Similarly for the magnetic field:

$$\mathbf{B}_\infty(\mathbf{r}) = \frac{\mu}{4\pi}\mathbf{p}\times\frac{(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}.$$

If a single localized source can be assumed, the bioelectromagnetic imaging problem reduces to the nonlinear but well-posed problem of finding the location $\mathbf{r}'$ and strength $\mathbf{p}$ of a current dipole given the measurements $\Phi(\mathbf{r})$ or $\mathbf{B}(\mathbf{r})$. Since L^0 needs to be computed only once for a given volume conductor, and given the analytical expressions available, this can be efficiently done using e.g. a Quasi-Newton, Marquard, or Simplex algorithm.[43]

The quasi-static nature of the bioelectromagnetic source imaging problem demands that $\Phi(\mathbf{r}, t)$ directly reflects dipolar strength $\mathbf{p}(t)$ at location $\mathbf{r}'(t)$. For a truly localized source, $\mathbf{r}'$ does not change over time and the cortical columns at $\mathbf{r}'$ define dipole orientation, $\mathbf{p}(t) = \hat{\mathbf{p}}\, s(t)$, with $s(t)$ scalar dipole strength. This model is often referred to as the *spatio-temporal* or *fixed* dipole model. The *rotating* dipole model refers to the situation where no constraint on orientation is assumed. If the position of the dipole is allowed to change with time as well, the term *moving* dipole model is used.

If a small known number of localized sources is assumed, the nonlinear problem is still well-posed — but solving it becomes tedious. The algorithms mentioned above easily get locked up in local minima, especially when the noise level (i.e. the background activity) is high. Methods like simulated annealing may help here, but in practice these do not converge easily. If available, constraints, e.g. demanding symmetry of dipole location and (vector) strength, can sometimes increase stability.

Figure 6 shows moving dipole solutions (crosses) and a two-spatio-temporal dipole solution (arrows) for EEG (left panels) and simultaneously measured MEG data (right panels) of an epileptic spike. The cortical area shown is around the left

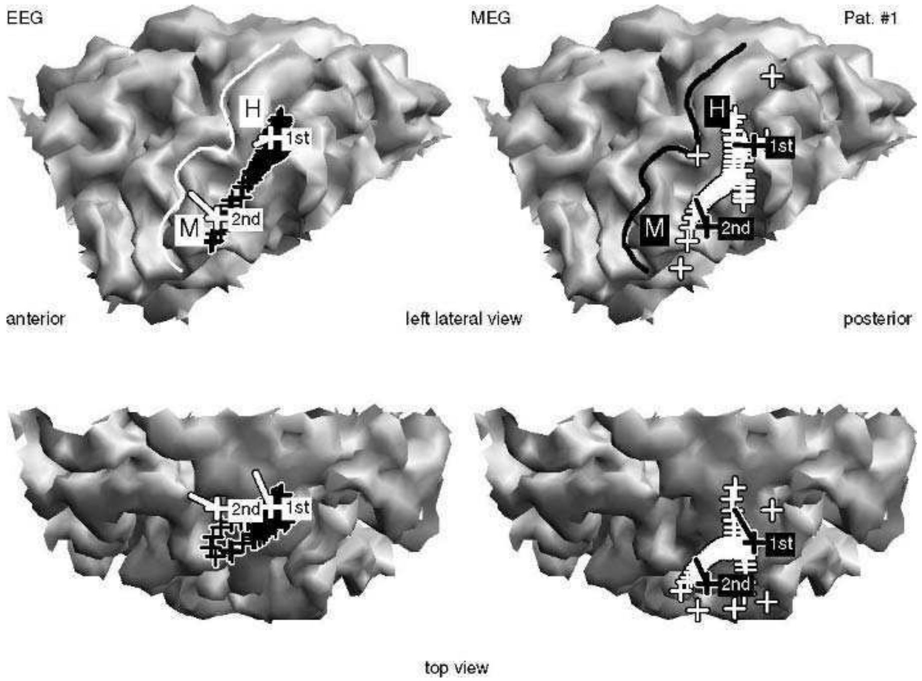


Fig. 6. Moving 1-dipole (plusses) and spatio-temporal 2-dipole (poles) solutions for EEG (left panels) and simultaneous MEG (right panels) for epileptic spike data. Left lateral view for top panels, superior view for lower panels. Reproduced with permission from Huiskamp *et al.* (2004),[69] ©American Clinical Neurophysiology Society.

central sulcus (drawn line). “1st” and “2nd” indicate the sequence of activation of the spatio-temporal dipoles. “H” and “M” indicate estimates of the hand and foot area as known from functional neuro-anatomy.[69]

One could argue that source localization based on the single or (limited) multiple dipole model is not obviously interpreted as “source imaging”, especially when results are presented in the graphical form of an arrow symbol placed in a stylized representation of a head and brain. Raw data methods have been described in which MEG or EEG is pre-processed by means of a bandpass filter and single dipoles are fitted to each time sample passing a certain predefined amplitude threshold.[29] If the residual error of the fit does not exceed a preset value, a counter for the voxel to which the fitted dipole position can be assigned is incremented. By thresholding the total number of counts thus gathered for the whole segment of data under consideration a functional image comparable to e.g. one obtained using the Statistical Parametrical Mapping (SPM) technique in fMRI is obtained.[33]

5.3.2. *The lead field*

Most other techniques in EEG and MEG source imaging use the concept of the lead field or transfer function. If the forward problem is solved for a dipole of arbitrary position and strength (i.e. $\Phi = L^0\Phi_\infty$ from the previous section) a formal function $L_{x_k}(\mathbf{r}, \mathbf{r}')$ linking orthogonal current dipoles of unit strength $p_{x_{k,k=1,2,3}}$ at $\mathbf{r}'$ to surface data Φ or B at $\mathbf{r}$ can be defined as:

$$L^E_{x_k}(\mathbf{r}, \mathbf{r}') \equiv \Phi(\mathbf{r}; p_{x_k}(\mathbf{r}')) = L^{0^E}\Phi_\infty(\mathbf{r}; p_{x_k}(\mathbf{r}')), \tag{50}$$

$$L^M_{x_k}(\mathbf{r}, \mathbf{r}') \equiv B(\mathbf{r}; p_{x_k}(\mathbf{r}')) = \mathbf{L}^{0^M} B_\infty(\mathbf{r}; p_{x_k}(\mathbf{r}')). \tag{51}$$

If the geometry of the cortical surface (i.e. the layer containing the ordered pyramidal cells) is known from MRI, unit dipoles $\hat{p}$ normal to that surface can be defined instead of the more general 3 orthonormal components. In that case the lead field links a *scalar* dipole strength at $\mathbf{r}' \epsilon S_{\text{cortex}}$ to potential Φ or Magnetic flux B at $\mathbf{r}$ on (or just above) the scalp. Since the layer of pyramidal cells is relatively thin (3 mm), an oriented double layer instead of oriented point dipoles can be assumed, leading to a description similar to that of the electrical generator in the heart (noted in Sec. 5). This has the advantage that $L(\mathbf{r}, \mathbf{r}')$ does not tend to infinity as $\mathbf{r}$ approaches $\mathbf{r}'$, since a double layer is described by the solid angle as in (22).

5.3.3. *Lead field scanning methods*

If a single focal source is assumed, the nonlinear dipole fitting method can be replaced by a scanning of the lead field computed for a pre-defined number of locations in the full volume of the brain or constrained to the cortical surface. If the resolution of the computed lead field is chosen high enough, in this way it can be guaranteed that a global minimum residual error solution is found. The residual error, or the explained variance, at each location in the brain, in combination with an estimate of the noise (or background) level of the data can be used to construct and threshold a functional image of brain activity.[36] If more than one focal source is assumed, the method mentioned above will fail because of the computational burden posed by the implied combinatorial problem. A very elegant method with the attractive acronym MUSIC (MUltiple SIgnal Classification) that originates from radar technique was first introduced to the EEG/MEG field by Mosher *et al.* (1992),[96] and circumvents this problem in the case of multiple dipolar sources with *temporally independent activity.*

Consider k normalized dipolar sources $\hat{p}_i$ at fixed, distinct locations $\mathbf{r}'_i$ having a fixed orientation normal to the cortical surface, having source strengths represented by sampled time series $s_i(t_j)$ which are independent during a time interval $[t_1, t_n]$. These sources can be represented by a $k \times n$ matrix S. Measurements in m sensors over the same time interval $[t_1, t_n]$ can be defined as a $m \times n$ matrix M, which is related to the sources through the matrix equation $M = LS + N$ where L is a matrix composed of the lead fields of each of the sources $\hat{p}_i$ and N is a $m \times n$ matrix representing additive white noise. If $m > k$, and the noise time series are uncorrelated with the signal (and the SNR > 1), a Singular Value Decomposition (SVD)[44] of $M, M = U\Sigma V^t$ can be partitioned as

$$M = U_s \Sigma_s V_s^t + U_n \Sigma_n V_n^t,$$

where the matrix U_s of rank k defines the spatial (in terms of sensors) eigenvectors of *signal space* and U_n (rank $m - k$) represents *noise space.* The diagonal matrix Σ_s contains k positive singular values $\sigma_{s_1}, \ldots, \sigma_{s_k}$, so that $\sigma_{s_{k+1}}, \ldots, \sigma_{s_m} = 0$. The first k singular values of $\Sigma_n, \sigma_{n_1}, \ldots, \sigma_{n_k}$, will therefore be zero, while for stationary white Gaussian noise, $\sigma_{n_{k+1}}, \ldots, \sigma_{n_m}$ are all equal, positive and smaller than σ_{s_k}. The matrices V_s and V_n represent temporal signal and noise space respectively.

It is important to realize that the subspace spanned by U_s, span (U_s) is equal to the subspace span (LS). If at a particular location $\mathbf{r}'$ on the cortical surface a source $\mathbf{p}$ is independently active, the vector $\mathbf{L}(\mathbf{r}')$ defined by its lead field in the sensors at $\mathbf{r}_{i,\, i=1,M}$ is in this subspace, and orthogonal to noise space. Conversely, if at a particular location there is *no* activity, the lead field for that location has a non-zero projection in noise space. The metric

$$Q \equiv \frac{\mathbf{L}^t(\mathbf{r}')U_s \cdot U_s^t \mathbf{L}(\mathbf{r}')}{\mathbf{L}^t(\mathbf{r}') \cdot \mathbf{L}(\mathbf{r}')} \tag{52}$$

is used in MUSIC to quantify this projection. $Q = 1$ if the lead field for $\mathbf{r}'$ is totally contained in signal space, $Q = 0$ if it is orthogonal to it. By examining $(1 - Q)^{-1}$ in all possible source locations ("lead field scanning"), a plot is obtained which peaks at the locations of multiple, independently active sources *in a single scan.* Derivations of a similar metric for situations where source orientations are unknown or allowed to vary are given in Ref. 96.

Practical problems arise with the application of MUSIC in situations where assumptions are violated or where the noise level is high. If highly correlated sources exist, the metric Q will not show a pronounced maximum or show one at an erroneous location. Also, the determination of the rank k that separates signal space from noise space is not straightforward, and may depend on assumptions of the character of noise.[80] In general, if signal and noise are truly uncorrelated and the SNR is high, over-estimation of the number of independent signal components k will not produce erroneous results — though it will reduce resolution.[67] However, if k is under-estimated, false peaks will appear. In a more recent adaptation of the classical MUSIC algorithm, unavoidably named RAP-MUSIC, some of the drawbacks of the original approach have been overcome.[97]

An example of a rank two MUSIC solution for high resolution EEG data of a Rolandic epilepsy spike is shown in Fig. 7. The area shown is the left central sulcus, with the white spot fMRI BOLD activation during tongue movement measured from the same patient.[146]

A scanning approach that is similar to MUSIC, but that computes for each scanning location an optimal spatial filter from the lead field, has recently gained more attention in the MEG field. There it is referred to as Synthetic Aperture Magnetometry (SAM),[7] in the EEG field has been called linearly constrained minimum variance spatial filtering, or LCMV beamforming.[141]

5.3.4. *Linear inverse methods*

If no assumptions about the existence of a small number of focal sources can be made, the source imaging problem takes the general form of the linear inverse problem (as discussed in Sec. 3).

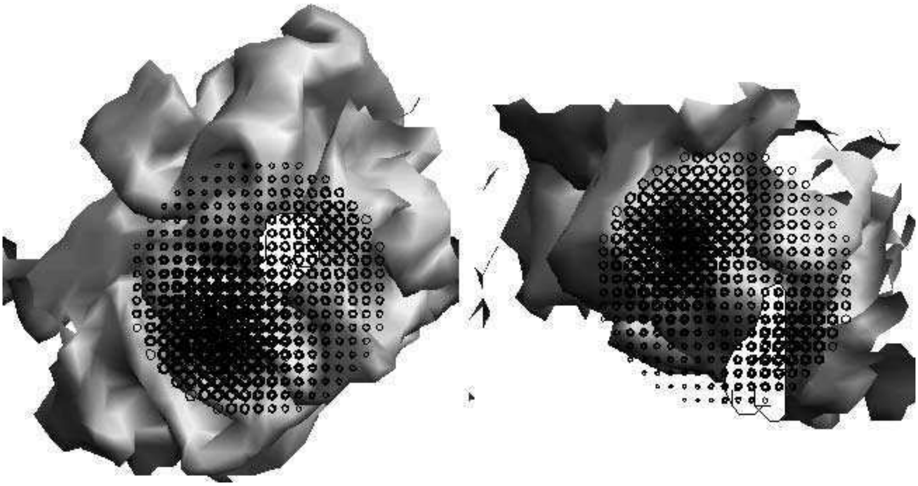


Fig. 7. Rank 2 MUSIC solution for EEG Rolandic epilepsy spike data showing two distinct extrema. Size of black circles is proportional to value of MUSIC metric. Left lateral view of central sulcus area for left panels, superior view for right panels. White spot indicates fMRI BOLD activation during tongue movement of same patient.

$$\Phi(\mathbf{r},t) = \sum_{x_k} \int_X L^E_{x_k}(\mathbf{r},\mathbf{r}')S_{x_k}(\mathbf{r}',t)dX + N^E(\mathbf{r},t), \tag{53}$$

$$B(\mathbf{r},t) = \sum_{x_k} \int_X L^M_{x_k}(\mathbf{r},\mathbf{r}')S_{x_k}(\mathbf{r}',t)dX + N^M(\mathbf{r},t), \tag{54}$$

where potentials $\Phi(\mathbf{r},t)$ or Magnetic flux $B(\mathbf{r},t)$ are given, $L^E_{x_k}(\mathbf{r},\mathbf{r}')$ and $(L^M_{x_k}(\mathbf{r},\mathbf{r}'))$ are available through the solution of the forward problem, $N^E(\mathbf{r},t)$ and $N^M(\mathbf{r},t)$ are unknown noise and $S_{x_k}(\mathbf{r}',t)$ is the source strength to be estimated.

In the most general case the sum is over all three orthonormal dipole components x_1, x_2, x_3 and over the brain Volume X. This problem is ill-posed, but in addition to that solutions are not unique. Nevertheless, standard zero-order Tikhonov ("the minimum norm solution") is traditionally used in order to solve this problem.[57] Implicitly, second order Tihonov was utilized in a paper by Pascual-Marquis *et al.* (1994).[111] Figure 8 depicts an application of this method, referred to in the EEG modeling field as "LORETA".[148]

Minimum l^1-norm solutions have also been proposed.[37] These proceed from minimization of (46), where $R = \|g\|$. For the magnetic case, an additional problem arises from the fact that pure radial sources produce near-zero magnetic fields. This can be circumvented by truncating the SVD of the lead field matrix (defined by $L^M_{x_k}(\mathbf{r},\mathbf{r'}_i)$ for each position i in the brain) to rank 2. In practice methods using Tikhonov regularization in this type of 3D inverse problem favor superficial solutions to deeper ones. Normalization methods for the columns of the lead field matrix have been used,[111] a more formal way of handling this was presented by Grave *et al.* (1997)[46] using the concept of resolution kernels. Other methods mentioned in Sec. 3 have been applied to the problem defined by (53), e.g. Bayesian approaches[5,12] and Wiener estimation.[126]

If the cortical surface geometry is available, the sum over the three orthonormal dipole components in (53) reduces to the single unit dipole normal to this surface (or a double layer element), and the integral is a surface integral. Such an approach

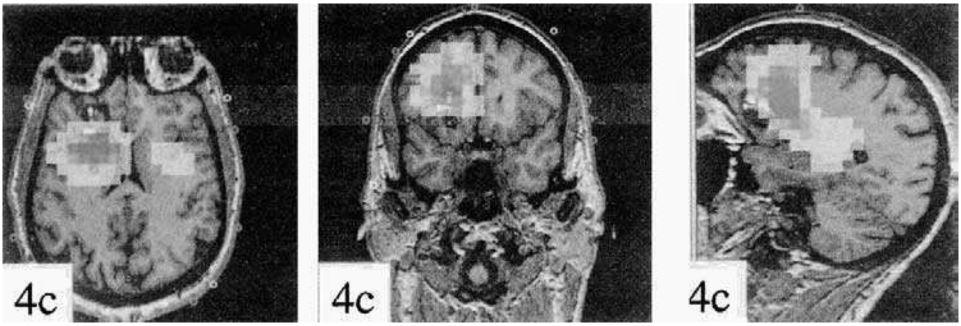


Fig. 8. LORETA solutions for ictal epileptic data, overlay on axial, coronal and saggital 3DT1 MRI. Reproduced with permission from Worell *et al.* (2000),[148] ©Human Sciences Press, Inc.

was first introduced by Dale and Serreno (1993).[27] Although this introduces additional constraints to the problem, the specific geometry of the sulci, with nearby sources of opposing orientation, adds additional ill-posedness. For the magnetic field, to make things worse, the gyri sources are oriented mainly radially, producing a negligible MEG. Baillet *et al.* (1999) proposed combined MEG-EEG modeling and measurement using limited regions of interest.[6]

5.3.5. *Cortical potential imaging*

This method, in which no assumptions about sources are made has its basis in Eq. (25). In the EEG field (a similar method for MEG would not make sense) this is often referred to as "spatial deblurring of scalp potentials".[82] It aims at estimating potentials that could be measured on the brain surface (in fact on a surface just beneath the skull surface) from scalp potentials. For this, only knowledge of the geometry and conductivity of scalp and skull are required. This imaging problem has a unique solution, but is ill-posed as well.

Since for most parts of the head the geometry resembles that of concentric spheres, if has been suggested that reasonable approximation of brain-surface potentials Φ_B is given by[104]

$$\Phi_B(\mathbf{r}', t) \approx \Phi(\mathbf{r}, t) - \frac{\sigma_{sc}}{\sigma_{sk}} d_{sk} d_{sc} \nabla_S^2 \Phi(\mathbf{r}, t), \tag{55}$$

where Φ is scalp potential (EEG), σ_{sc} and σ_{sc} are scalp and skull conductivity, d_{sc} and d_{sk} scalp and skull thickness and ∇_S^2 is the surface Laplacian. The validity of (55) is based on the fact that (homogeneous) skull conductivity is much lower than that of cortex and d_{sk} and d_{sc} are much smaller than the size of the brain, and that as a result currents through the skull are mainly radial. In practice these assumptions will be less valid, especially in the case of superficial tangential sources, and the surface Laplacian has to be approximated from a limited number of measurement sites by numerical techniques. Edlinger *et al.* presented an analytical solution of (25) for a concentric spherical brain-skull-scalp model, which is more accurate when inter-electrode distances are less than 2.5 cm and is not dependent on assumptions about trans-skull currents and the need to construct a surface Laplacian operator.[32]

For realistically shaped head models Gevins *et al.* used the FEM for the solution of (25).[41] A BEM approach, adapted from the method introduced in by Barr *et al.* (1977)[8] for the corresponding ECG problem, was presented by He *et al.* (1999).[59]

The full array of potentially applicable inverse methodology has not yet been applied to this imaging problem. For the case of superficial sources involved in the generation of evoked potentials, it is expected that progress can be made towards non-invasive potential imaging (e.g. approximating Cortical SEP (CSEP) by imaged surface SEP's). However, for spontaneous signals such as epileptic spikes, the application makes less sense. In epilepsy patients with implanted electrodes, inversion strategies applied to recorded Electro Cortico Grams (ECoG's) require an assumed source model or other details that the imaging method cannot itself provide.

5.4. *Validation*

Validation of results obtained by the various methods for EEG-MEG source imaging have been mainly done on basis of data available from epilepsy patients that have been treated surgically, or that of candidates for such a procedure. In the former case, ECoG's are measured on the exposed brain in order to tailor the resection, or cortical evoked responses are measured and cortical stimulation is performed in order to localize functional areas. In the case of surgery candidates, often electrodes are implanted subdurally or intracortically and the patient is monitored for seizures during some period. Cortical evoked response and stimulation may be performed as well.

However, few of the data available from such studies is used for a quantitative validation. Reports from Gevins *et al.*[41,42] present examples of a qualitative comparison between computed cortical potential distributions and those measured on the brain in the case of median nerve evoked responses. Ossenblok *et al.* (2003),[109] present potential distributions measured in grids placed over the central sulcus and compare these to dipoles localized on basis of MEG evoked by median nerve stimulation. Similar comparisons of localization of evoked responses for MEG, EEG and ECoG were presented by Sutherling *et al.* already in 1988.[130]

For the *spontaneous* inter-ictal spikes and ictal discharges that occur in epilepsy patients, simultaneous measurements of EEG or MEG and ECoG have shown that activity that is truely localized intracranially (i.e. that is observed around a single electrode) is not detectable with scalp EEG or MEG.[91,106] If an activity with a dipolar character is observed in EEG or MEG, it will be accompanied by widespread cortical activity over several cm^2. For any quantitative analysis and validation a model for such activity that goes beyond the single dipole is required.

6. The Heart

6.1. *Cardiac muscle structure and function*

Cardiac muscle cells or myocytes are roughly cylindrical with a length in human ventricular tissue of 80 μM to 100 μM and a diameter of 10 μM to 20 μM, and are bounded by the cell membrane or sarcolemma.

There are two sets of filaments present in muscle cells: Thin filaments composed of a globular protein called actin, and thick filaments formed as an aggregate of a much larger protein called myosin. The thick and thin filaments interdigitate to form a sarcomere between the thin filament tethering points at Z lines. The sarcomere is the basic contractile unit.

A large number of sarcomeres are present in a single cell. Cells are joined end-to-end with other cells through intercalated disks and also branch and interconnect with neighboring, nearly parallel cellular strands. Electrical connection between adjacent cells is through gap junctions which are located in the intercalated disks. Additionally, there are several important internal structures. Due to the almost constant energy requirements of repeated cellular contraction in the cardiac cell there are a large number of mitochondria which provide oxygen to the cell. A network of Sarcoplasmic Reticulum (SR) provides a large region of uptake pumps which remove Ca^{+2} from the cellular matrix and stores it in the Junctional SR (JSR) awaiting release for enabling contraction. Deep invaginations of the sarcolemma into the fiber are known as the transverse tubular system or T-tubules. This system is used primarily to conduct the action potential down into the cell, and additionally to transport components from the interstitial fluid surrounding the cell deep into the cell, and is of particular importance in the excitation-contraction coupling.

An averaged myocyte direction can be defined at any point, which is known as the local fiber orientation. Measurements of fiber orientation at a large number of sites spread throughout the ventricular myocardium have been made.[102] There also exists a comprehensive organization of extracellular connective tissue, including a substantial hierarchy of collagen structures which constrain the movement of the muscle fibers.

Mechanical contraction of the heart is caused by the electrical activation of the myocardial cells. The heart is electrically self-contained, having the ability to initiate its own beat with a regular period, and will continue to beat after being removed from the body. Cells capable of initiating electrical activity are called pacemaker cells, and exist in several places throughout the heart. Only those pacemaker cells with the fastest rate of pacemaker discharge control the electrical activity of the entire heart. The region of tissue with the shortest period of spontaneous electrical activity is the SinoAtrial (SA) node, which is located on the atrial wall near the junction of the superior vena cava and the right atrium, and consists of a group of pacemaker cells. Action potentials are normally generated here at the rate of 60 to 100 per minute. From the SA node, the action potential is propagated from cell to cell through firstly the right atrium, followed closely by the left atrium at a conduction velocity of approximately 1 meter/second until it reaches the AtrioVentricular (AV) node. The AV node consists of similar pacemaker-type cells as are found in the SA node, but because they beat spontaneously at a slower rate (approximately 40 to 55 beats per minute) they are governed by the propagation from the SA node. In the event that the SA node is removed or destroyed, or that conduction is slowed through the atria, the cells in the AV node will take over as pacemaker for the heart.

Conduction through the AV node is at a much slower rate (around 0.05 meter/second) giving time for the atria to contract and pump blood into the ventricles before the action potential conducts through the ventricles and causes them to contract. The AV node is normally the only electrical connection between the atria and the ventricles. From the AV node, the electrical propagation enters the bundle of His which is the upper portion of the ventricular conduction system and runs down the right side of the septum. This common bundle divides after a short distance into right and left bundle branches. The right branch continues down the right septal wall, and the left perforates the septum and splits into two further main branches on the left septal wall. All of these branches continue to subdivide into a complex network of fibers called the Purkinje fiber network, spreading across the endocardial surface of both ventricles and into the subendocardial region of the ventricular myocardium. Due to the arrangement of connecting fibers the septum is activated first. The papillary muscles are also activated early so that they can prevent the AV valves from inverting during systole. Due to the faster conduction of approximately 2 meter/second through the bundle and Purkinje fibers, the entire endocardium is excited almost simultaneously. The apical regions contract first and the basal regions are usually the last regions to be excited. Excitation spreads outwards through the ventricular wall at a rate of about 0.3 to 0.4 meter/second, and the first epicardial region to be excited is the thinnest portion of the right ventricular wall.

The bioelectric sources which arise during the heart's excitation process produce a flow of electric current in the surrounding tissues. It is therefore possible to detect, with a pair of electrodes external to the heart, time-varying potential differences known as electrocardiograms.

The first recording of a human ECG was by Waller in 1887.[144] The electrical activity of the exposed heart was already known (and galvanometers had been invented) but Waller decided to investigate the possibility of recording potentials from the limbs of animals and from man. He dipped his right hand and left foot into a couple of basins of salt solution which were connected to two poles of an electrometer and "at once had the pleasure of seeing the mercury column pulsate with the pulsation of the heart".

Waller demonstrated this feat at St Mary's Laboratory in 1887 to an audience which included Professor Willem Einthoven. Einthoven went on to refine the process, developing the string galvanometer, and in 1924 was awarded the Nobel prize for Physiology or Medicine for the discovery of electrocardiogram mechanism.

6.2. *Derivation of the basic imaging equations for various source imaging models*

As noted earlier, the heart supports propagation of "action potentials". These result from transmembrane currents subsequent to time-varying membrane conductance to various ions in the presence of previously established transmembrane ionic concentration gradients. While highly sophisticated models of such events exist and continue to be refined (mostly taking inspiration from the Hodgkin–Huxley model of the neuron), for our purposes a simplified version of the so-called bidomain model[63] is sufficient. Each bidomain "point" contains a bit of intracellular space and a bit of extracellular space. Continuum notions, quasi-static assumptions, and Ohm's Law imply that the current density $\mathbf{j}$ at any of these bidomain points must satisfy

$$\mathbf{j} = -G_i \nabla \phi_i - G_e \nabla \phi_e,$$

where G_i, G_e are intracellular and extracellular conductivity tensors and ϕ_i, ϕ_e are the intracellular and extracellular potential. Any excess current (non-zero divergence) appearing in the extracellular component of a bidomain point must come from the intracellular component of the bidomain (via the cell membrane), and vice versa. Hence $\nabla \cdot \mathbf{j} = 0$, and so we have

$$\nabla \cdot [G_e \nabla \phi_e] = -\nabla \cdot [G_i \nabla \phi_i]. \tag{56}$$

The divergences on both sides of the above equation express the location's role as a *source* of extracellular current (net transmembrane current). Transmembrane currents capable of influencing body surface potentials only occur in excitable tissue (i.e. sensory-neuro-muscular tissue) — the predominant one being the heart. Thus, in the body volume excluding the heart muscle, both sides of the above equation are zero. In particular, the above becomes Laplace's equation,

$$\nabla \cdot [G_e \nabla \phi_e] = 0, \tag{57}$$

for the volume external to the heart. Note that the time-independence of the operators in the above equations follows from the quasi-electrostatic assumption.

- One "continuous volume" where (57) is valid is the body region external to the heart (pericardium). The boundary conditions are the zero normal component of current at the body surface (Neumann) and the epicardial potentials (Dirichlet). This leads to the *epicardial imaging approach.*
- Another continuous volume where (57) holds is within a blood-filled chamber of the heart. The boundary conditions would then be the zero component of current normal to a multi-electrode catheter tip (floated into the chamber via a peripheral vein or artery) and the endocardial potentials. This leads to the *endocardial imaging approach.*

In either case, Green's second identity supplies a linear equation for (measured) body surface potentials or catheter tip potentials in terms of the unknown epicardial or endocardial potentials. Thus, the relationship can be expressed by

$$f = K\phi, \tag{58}$$

where (after discretization) f is a vector whose components are the measured body surface or catheter tip potentials at various locations, ϕ is a vector of epicardial or endocardial potentials at various locations, and K is the Green's-function-derived transfer matrix (computation of K constitutes the "Forward Problem"[55]). Estimation of ϕ via "inversion" of (58) supplies an image of the epicardial (or endocardial) potentials at any point in time. Reconstruction of the epicardial potentials are by far the most investigated imaging formulations of the inverse electrocardiography problem.[88]

However, it is also possible to base an imaging formalism on reconstruction of transmembrane potential $\phi_m = \phi_i - \phi_e$. Substituting ϕ_m into (56), we obtain

$$-\nabla \cdot [(G_e + G_i)\nabla\phi_e] = \nabla \cdot [G_i \nabla\phi_m]. \tag{59}$$

We still have the homogeneous Neumann boundary condition (zero normal components of current density on the body surface). Applying a Green's function formalism (as described in the Mathematical Methods section), we get

$$\phi(y,t) = \int_V \psi(x,y)\nabla \cdot [G_i(x)\nabla\phi_m(x,t)]\, dV_x, \tag{60}$$

where $\psi(x,y)$ is the Green's function of (59), V is the volume of heart muscle, and $\phi(y,t)$ is the (extracellular) potential at y (e.g. on the body surface) at time t. Integrating by parts twice, one obtains[149]

$$\phi(y) = -\int_S [G_i \nabla\psi(x,y)] \cdot \mathbf{n}_x \phi_m(x) dS_x + \int_V \nabla \cdot [G_i \nabla\psi(x,y)]\phi_m(x) dV_x, \tag{61}$$

where S is the boundary of V, and $\mathbf{n}_x$ is the outward unit normal on S at x. If tensors G_i and G_e are related to each other by a scalar (equal anisotropy), then the volume integral on the above right vanishes (by definition of the Green's function). The resulting equation then relates body surface potentials to transmembrane potential on the heart surface (note that S includes both epicardial *and* endocardial surfaces). However, as noted earlier, the equal anisotropy assumption is not accurate.[24,116]

An important distinction between the above source formulations is that computation of the Green's function for the transmembrane potential (59) requires knowledge of the anisotropic conductivity of the heart — whereas in the epicardial potential formulation (57) only knowledge of tissue conductivity external to the heart is required (and in the endocardial potential formulation one requires only a value for blood conductivity). Though standard conductivity values from the literature are typically used, it is noted that strategies exist for patient-specific tissue conductivity tensor imaging via MRI.[138,139]

6.3. *Epicardial potential imaging*

The concept of epicardial potential imaging arose in a formal way with the work of Martin and Pilkington[89] in the early 1970s. They immediately recognized the severely ill-posed nature of the problem and discussed appropriate statistically based regularization strategies. Barr *et al.* introduced a numerically favorable boundary element scheme,[8] which was evaluated in chronically instrumented dogs.[9] Over the years, there has been a great deal of work on refinements of regularization methodology, beginning with Colli-Franzone and coworkers.[25] The advent of CT and MRI obviated many difficulties regarding reliable identification of the anatomy for solution of the forward problem, though quantification of conductivity variations remains extremely problematic. While the impact of conductivity imprecisions has been contraversial, there is in any case reason to believe that MRI could provide very useful information relevant to relative tissue conductivity, including anisotropy.[138]

By far the most sustained effort in the realm of invasive investigations appropriate to verification of non-invasive imaging strategies, has been conducted by Taccardi (presently at the University of Utah) and coworkers, extending back to the 1960s. The Utah group, sometimes in collaboration with the group at Case Western led by Rudy, have produced a long series of investigations of non-invasive epicardial potential imaging (which they term Electrocardiographic Imaging), invasively verified via the Utah torso tank system. The experimental setup employs a heart suspended in a torso shaped electrolytic tank, perfused by an anesthetized dog external to the tank.[110] Electrodes are present on the outer margin of the tank, and also in proximity to the epicardium (either with electrodes in proximity to the heart via projections from the tank surface, or via an epicardial electrode sock). Such work has addressed inverse reconstruction of epicardial potentials and activation in the setting of sinus rhythm, pacing, and arrhythmias,[21] and has also been used to a limited degree to assess the impact of torso inhomogeneities.[119] Data recorded from infarcting dogs has been used in a simulation study to demonstrate the potential promise of the epicardial potential imaging formulation.[20] This approach has also been applied to demonstrate the feasibility of reconstructing repolarization properties of interest (e.g. increased dispersion of repolarization[38]). Another report[114] (extending earlier work[87]) identified local changes in inversely computed epicardial electrograms in patients whose data was accessed during coronary catheterization, preceeding and following angioplasty balloon catheter inflation. In the eighteen study patients, the predicted region of ischemia following balloon inflation correlated with the expected region of perfusion deficit based on the vessel occluded.

Recently, the Case Western group has also presented non-invasively reconstructed human epicardial potential images in human hearts in a variety of normal and pathological conditions,[120] although only qualitative validation of the results could be made (from the known sequence of atrial depolarization in normal subjects, and from beats artificially paced from a known location). These results conflict with earlier results of a different group using invasive data from patients undergoing arrhythmia surgery, which had suggested that the usual epicardial potential regularization methodology was able to usefully image epicardial potential during the QRS interval only in its initial portions.[127]

There remain many issues regarding the efficacy and accuracy of epicardial potential imaging (or "electrocardiographic imaging"). Most importantly, *in vivo* validation is required — something very difficult to obtain. In addition, the impact of present imprecise knowledge of organ conductivities on the quality of the inversely computed images has still not been definitively settled. Another issue involves accuracy of activation isochrones based on identification of the inversely computed intrinsic deflection.[112] Particularly as regards the latter, there remain questions regarding optimal regularization techniques. For example, it is evident from numerical simulations that for a wide range of signal-to-noise ratio conditions there will be significant noise in the electrogram reconstructions unless time is jointly considered in the reconstructions[52] — clearly impacting isochrone computation. Finally, a measurable clinical impact of epicardial potential imaging has not yet been verified (though a potential role clearly could exist). However, all of the above questions and concerns could be settled in the not distant future.

6.4. *Endocardial potential imaging*

Interventional cardiologists employ transvenous catheter procedures to treat arrhythmogenic foci and aberrant conduction pathways. Such treatment first requires mapping the endocardial potential. This initial invasive imaging is cumbersome, tedious, and lengthy. Typically, a roving probing transvenous electrode catheter is brought into contact with many endocardial locations over the course of many heartbeats, and a depiction of an endocardial activation map is thereby inferred (an improved technology along these lines is described in Gepstein *et al.* (1996)[40]). The number of sites accessed is limited, and there is no accounting for beat-to-beat activation variability when reconstructing the maps from the many beats. Recently introduced expandable basket electrode arrays[125] have their own problems related to limited numbers of electrodes, the need to contact (and perhaps irritate) the endocardium, and the possibility of difficulties in collapsing the basket at the end of the acquisition.

These problems can be potentially addressed by the use of a transvenous catheter whose tip is studded with multiple electrodes, and which is placed somewhere in the midst of a cardiac chamber (without contacting the endocardium). Once the catheter location relative to the endocardium is registered, it becomes theoretically possible to inversely compute the endocardial potentials from a single heartbeat — indeed, to follow dynamic isopotential maps within a single beat, as well as beat-to-beat changes in activation maps. For this inverse problem, the volume is bounded by the endocardial surface and the multielectrode probe surface. Laplace's equation holds in this volume, and the boundary conditions are the (unknown) endocardial potentials, and the zero normal current density at the multi-electrode probe surface. As indicated earlier, a linear relationship is derived between the endocardial potentials and the catheter electrode potentials (Fig. 9).

Notwithstanding the inconvenience of the required cardiac catheterization, there are two very significant advantages of this formulation over the technique of imaging the epicardial potentials from the body surface. First, the electrodes are relatively close to all portions of the surface to be imaged (e.g. as opposed to the distance between body surface electrodes and the posterior wall of the heart). Second, the relevant volume is composed only of blood in the lumen of the cardiac chamber. Therefore, the modeling required for estimation of the transfer matrix is vastly less, and the uncertainties in the values of key components of the model (i.e. tissue conductivities) are markedly diminished (the blood has uniform isotropic conductivity).

The initial proposal and work on a multielectrode non-contact array, placed in a cardiac chamber for purposes of accessing endocardial potentials, was due to Taccardi *et al.*[131] In the past few years there has been much significant work

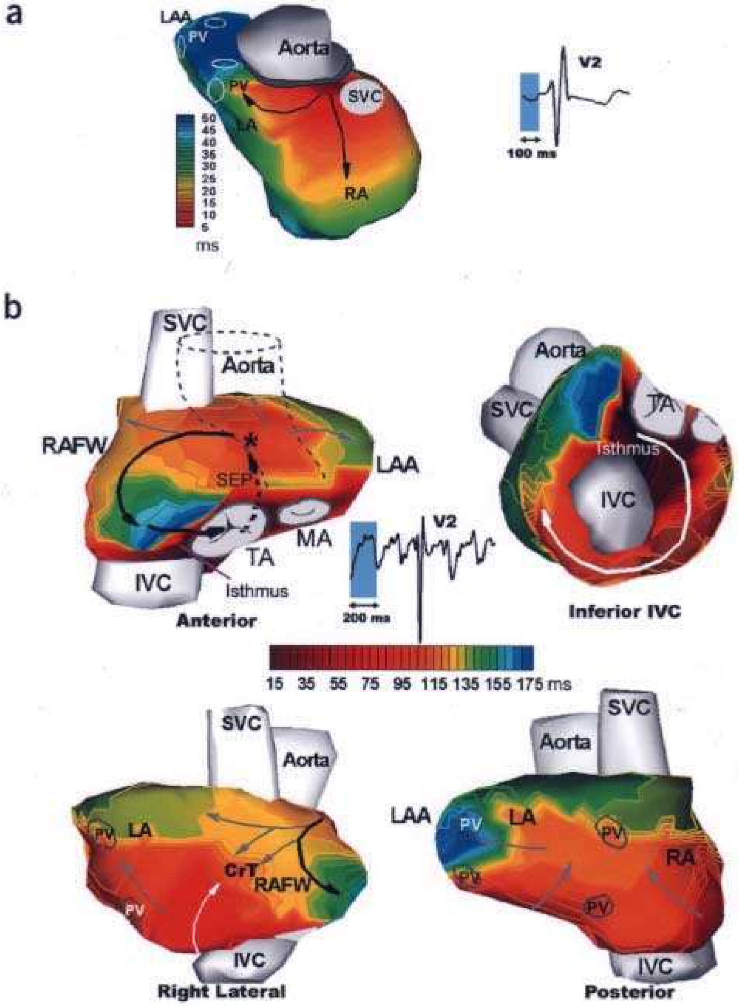

a
LAA
PV
Aorta
PV
SVC
LA
RA
50
45
40
35
30
25
20
15
10
5
ms
V2
100 ms
b
SVC
Aorta
RAFW
*
SEP
LAA
MA
TA
IVC
Isthmus
Anterior
V2
200 ms
Aorta
SVC
TA
Isthmus
IVC
Inferior IVC
15 35 55 75 95 115 135 155 175 ms
SVC
Aorta
PV
LA
CrT
RAFW
PV
IVC
Right Lateral
SVC
Aorta
LAA
PV
LA
PV
RA
PV
PV
IVC
Posterior

Fig. 9. Atrial activation. (a) A normally conducted sinus originated beat is imaged, along with the V2 ECG tracing. (b) Atria in four projections imaged during atrial flutter, with the V2 ECG tracing. From Ramanathan *et al.*, *Nature Medicine*, 10:422–428 (2004),[120] used by permission.

reported on successors to this idea. For example, in experiments on dogs, Khoury *et al.*[79] used a 128 electrode catheter, inserted via a purse string suture in the left ventricular apex, and showed that faithful renditions of endocardial activation, both with paced and spontaneous beats, was possible by solving the inverse problem. Ischemic zones were also well defined. A spiral catheter design has also been investigated.[74]

An impressive series of experiments has been performed with a competing system, developed by Endocardial Solutions, Inc. In addition to a 64-electrode 7.5 ml inflatable balloon catheter, a second transvacular catheter is passed and dragged along the endocardium. As it is dragged, a several kHz signal is passed between it and the electrode catheter, localizing its position with respect to the electrode catheter. In this way, a rendition of the endocardium with respect to the electrode catheter is produced. Following construction of a "virtual endocardium" via a convex hull algorithm applied to the above anatomical data, the inverse problem is then solved, generating several thousand "virtual electrograms" on the virtual endocardium.

One study of endocardial potential imaging describes the classification of atrial fibrillation in humans in terms of numbers of independent reentrant wavefronts identified.[123] Another report[129] describes the successful ablation of fifteen instances of ventricular tachycardial guided by this catheter system. A further report[113] describes the utility of the system in directing catheter ablative therapy in subjects with atrial arrhythmias refractory to pharmacologic therapy.

Also notable is the most recent work with the spiral catheter design[75] where the images obtained were verified by 92 intramural needle electrodes placed into the myocardium of an isolated dog heart. Invasively verified reconstructions were obtained in normal and infarcting heart, with reportedly faithful construction of a variety of electrophysiologic characteristics.

6.5. *Transmembrane potential imaging*

Epicardial and endocardial potential imaging addresses the need to reconstruct something that is currently accessed invasively, and is thus of evident interest. However, such potentials are not themselves a clinical endpoint. Ultimately, clinicians are interested in the action potential — or at least, features of the action potential. The most important features of the action potential are activation time (time of arrival of phase zero at every location, the aggregate of which globally describe conduction disturbances), phase zero amplitude (reflecting ischemia), and action potential duration (reflecting refractory periods, potentially associated with propensity for re-entrant arrhythmias).

The marker for activation in an electrogram (a tracing of epicardial or endocardial potential at a given cardiac site) is the "intrinsic deflection" — defined as its steepest downward deflection. Recall that the source derives from the gradient of transmembrane potential (e.g. (60)), and that during cardiac activation this is usually appreciably non-zero only at the locus of points undergoing action potential phase 0. This locus is approximately a surface (the interface between depolarized and nondepolarized muscle). Electrically, this behaves approximately as a propagating surface of dipole moment density (a double layer). There is a discontinuity of potential as the double layer is crossed. Ideally, as an extracellular location is passed over by the activation wavefront, there will then be a sharp downward deflection in the extracellular (electrogram) potential — the intrinsic deflection. However, the reality is that it is not infrequent that there is more than one reasonable candidate for the intrinsic deflection within a given location's electrogram. Furthermore, the intrinsic deflection is often rather lengthy, so the selection of a single activation time within the intrinsic deflection is to some extent arbitrary.[71,112] The activation time is presumably the inflection point of the deflection (which itself is poorly defined in the noisy setting). To a large extent, these problems are inherent in the source formulation: The epicardial (or endocardial) potential at a location actually reflects contributions from electrical activity at all surrounding locations, when in fact we desire to resolve results of the membrane function at a single location — i.e. the local action potential.

One can begin with consideration of (60), which expresses the linear relationship between cardiac muscle current sources (or sinks) expressed as $\nabla \cdot [G_i(x)\nabla\phi_m]$ and accessible potentials. However, the null space of this operator is non-trivial: In a heart of uniform conductivity (or equal anisotropy), it includes every arrangement of ϕ_m such that the set of points where $\nabla\phi_m \neq 0$ effectively constitutes a closed surface. One must be ready to accept a similarly annoying null space for the setting of anisotropy as well. Therefore, to date, all forays in this direction have concentrated on (61), where one ultimately plans on "ignoring" the second integral on the right-hand-side, so as to compute ϕ_m solely on the surface bounding the heart muscle S. Ignoring of the second integral, of course, can only be definitively justified by successful verification of estimate accuracy. There also remains the need to know the anisotropic conductivity of the heart muscle for computation of the Green's function in the first integral on the right-hand side, presumably via strategies connected with MRI.[138]

The numerically sophisticated forays in this direction were due to Cuppen and van Osteroom.[26] Ultimately, these relied on modeling the local transmembrane potential during the activation interval as a step function, thus

$$\phi_m(x,t) = a + b(x)H(t-\tau(x)), \tag{62}$$

where $H(t)$ is the Heaviside function (zero for $t < 0$, unity for $t > 0$) and $\tau(x)$ is the time of phase 0. Substituting in (60), one evidently obtains a nonlinear equation for $\tau(x)$, which was treated by a quasi-Newton routine[66] with initial seed obtained from the linear equation

$$\int_{QRS} \phi(y,t)dt = -\int_S A(x,y)\tau(x)dS_x, \tag{63}$$

(obtained by substituting (62) into (61) and then integrating by parts (while ignoring the volume integral and assuming $G_i(x)$ is a constant scalar).

The problem of reconstruction of $\tau(x)$ is well posed if its relative extrema are known,[48] in that case being merely a question of interpolation. Relevant to this, the Critical Point Theorem[47,49] states that x' is a relative extremum of $\tau(x)$ as a function over S (in particular, a source or sink of surface activation) if and only if $\phi(x',y)$ is in the space spanned by the eigenfunctions of the operator $\int_Y \int_X [G_i\nabla\psi(x,y)] \cdot \mathbf{n}_x [G_i\nabla\psi(x,\hat{y})] \cdot \mathbf{n}_x \, dX(\cdot)d\hat{Y}$. An algorithm based on this theorem is presented in Huiskamp and Greensite (1997).[67] Oostendorp *et al.* at the

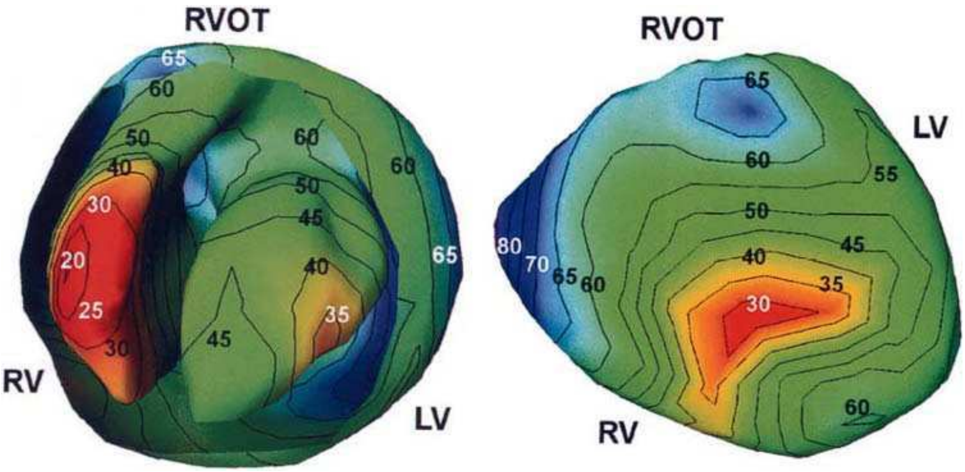


Fig. 10. Non-invasive ventricular surface activation maps of a human subject, produced by Tilg *et al.*, *IEEE Trans. Biomed. Eng.*, (2002).[136] The image on the left has the anterior cardiac wall cut away, so that endocardial activation is visualized. The image on the right is a view of the anterior epicardium. Used by permission.

University of Nijmegen/University of Helsinki have produced work evaluating this approach, both *in vitro*[107] and *in vivo*[108] (validation in hearts removed at the time of cardiac transplantation). The group at Graz and Innsbruck have produced much work on transmembrane potential imaging, some of which utilizes the Critical Point Theorem[94,95,134–136,142]. Figure 10 is an example of activation imaging obtained by a nonlinear algorithm using the Critical Point Theorem to provide an initial seed.

On the other hand, a recent study[92] proposes convex optimization treatment of (61), where the side constraint is that $\phi_m(x,t)$ is monotonic (the cardiac cycle is split into two parts: During the QRS interval $\phi_m(x,t)$ is monotonic increasing, and outside this interval it is monotonic decreasing).

A rather different approach was proposed by Ohyu,[105] where a covariance matrix for $b(x)$ and a covariance matrix for $\tau(x)$ were used to define a maximum *a posteriori* expression solved by simulated annealing.

7. Skeletal Muscle

7.1. *Skeletal muscle structure and function*

Contraction of skeletal muscles causes them to shorten and, hence, move the bones that they are attached to. Each muscle has one or more origins and insertions. The origin point of the muscle is the point of attachment (of the muscle to bone) that is less movable during contraction of the muscle. The insertion point is the point of attachment that moves the most during a contraction. The idea of origins and insertions provides a macroscopic view of muscle structure and attachment. In reality, muscles are compartmentalized — fibers are grouped into different bundles that may run in slightly different directions to other bundles. Muscle fibers can be further separated into myofibrils that are approximately 1 μm in diameter.

Skeletal muscles are activated by the central and peripheral nervous systems. The process by which fine motor control can be achieved by the nervous system is a very complex one. The Motor Unit (MU) is the basic functional unit in the muscle that the nervous system has control over. Each motor unit consists of a single motor neuron, its axon and the group of fibers that it innervates or supplies. The number of fibers in a motor unit is highly variable — this can range from 3 to 2000. The fibers in a MU are not situated side-by-side but spread over a certain volume known as the Motor Unit Territory. MU territory diameters can vary from 2 to 10 mm.

Each fiber in a motor unit is innervated by one of the several branches of the axon belonging to that motor unit. Each branch of the axon is attached to a muscle fiber at the Neuromuscular Junction or endplate, which is generally situated in the middle of the fiber. When a neuron controlling a motor unit fires, all the fibers in that motor unit contract to produce twitches. In order to grade the contractions of the whole muscle, therefore, the number of motor units activated is varied and synchronization of motor unit activation is also varied. The central nervous system can increase the force produced by a muscle by increasing the number of motor units that are already active or by increasing the rate of excitation of each motor unit.

The motor unit action potential is built up from individual fiber activity and is dependent on the distributions of fibers within the motor unit, the variation of the innervation of each fiber within the unit, the motor unit recruitment and firing behavior, and the interpulse interval. A surface electrode placed on or over a skeletal muscle records the summation of the various motor unit action potentials, and is termed a surface EMG (electromygram) recording.

7.2. *Skeletal muscle inverse problems*

The inverse problem of electromyography generally involves the estimation of certain parameters of the bioelectric sources that generate the surface EMG signals. EMGs also provide valuable information in analysis of biomechanical models. The focus, in the latter case, is not on the determination of bioelectric source strength or position but rather in processing the SEMG signals to provide a measure of muscle activation, timing of activation, muscle force or some other similar parameter(s) that can be easily used as input to other models or as an indication of muscle pathology.

Given the earlier description of skeletal muscle physiology, the task of identifying the underlying source is extremely difficult. Not only are the distribution of fibers within a motor unit unknown, but also unknown is the distribution of the end plates of each fiber within that unit. This variation in innervation within fibers of the same motor unit means that there can be a spread of one to two milliseconds in the activation of a given motor unit. Also different motor units can be activated to produce the same response. However, typically what one is interested in when performing EMG recordings is not so much the individual fiber activation *per se*, but the result of the activation, i.e. the force produced. For this reason, most EMG inverse analysis attempt to reconstruct muscle forces (or estimate fatigue) from the derived signals. This is mostly done empirically i.e. inverse dynamics, rather than source localization.

8. Gastrointestinal Smooth Muscle

The muscular layers in the walls of the GastroIntestinal (GI) system produce peristaltic waves through interplay between the enteric nervous system, the Interstitial Cells of Cajal (ICC) and smooth muscle cells. These smooth muscle cells are unitary (or visceral) cells that act like a syncytium in which the electrical activity is conducted from fiber to fiber.[15] The muscular system is characterized by spontaneous activity with regular cellular depolarizations that are known as the slow wave or electrical control activity (ECA).[137] The ICC are critical for initiating this slow wave activity and are believed to regulate the pacing of the GI ECA.[70] As with all muscle, electrical action potentials initiate each muscular contraction. However, not every electrical impulse in the GI tract gives rise to a contraction, and hence these impulses are not considered to be a true action potential. Muscular contraction occurs only when the Electrical Response Activity (ERA) or spike activity is triggered by the depolarization of the cell membrane above threshold. Thus, the temporal and spatial coordination of the muscular activity of the gut is based upon the ECA.

The ECA exhibits a frequency that is location dependent with pacemakers present in both the stomach and duodenum.[147,122] It is the slow wave membrane depolarization generated by the ICC that determines the maximum frequency and the propagation characteristics of the intestinal rhythmic contractions. The ICC associated with Auerbach's plexus (a group of ganglion cells between the circular and longitudinal muscle layers of the muscularis externa in the digestive tract) spontaneously depolarize three times a minute to initiate the ECA in the stomach.[100] In the duodenum, the frequency of the slow wave activity is about 12 cycles per minute (cpm) and this decreases over the length of the small intestine to about 8 cpm in the terminal ileum. Transection of the small intestine allows local pacemakers to take control. Thus, the ECA frequency of the intestine distal to the transection is less than when previously controlled by the more proximal pacemaker, especially if the transection is made in the proximal frequency plateau.[30]

The electrical activity of the stomach can be recorded with cutaneous electrodes giving rise to what is termed the ElectroGastroGram (EGG). This was first recorded by Alvarez in 1922[3] from "a little old woman whose abdominal wall was so thin that her gastric peristalsis was easily visible". It was independently discovered again in 1957[28] but it was not until 1975 that the gastric origin of the EGG was conclusively demonstrated.[19] In the last 25 years much has been learnt about the electrical activity of the stomach, and research into the relationship between gastric electrical activity and the EGG has recently substantially increased. This research has also resulted in significant advances in the study of intestinal ECA, but the cutaneous recording of intestinal potentials continues to be problematic.

Analysis of electrical recordings from gastrointestinal activity have mostly been limited to an analysis of the various frequency components within the signals.

Recordings of the magnetic field associated with the stomach and small intestine have been made, e.g. Bradshaw *et al.* (1997).[16] These data are beginning to be used in inverse studies. Some simple inverse calculations have been presented in Irimia and Bradshaw (2004)[72] in which the body was modeled as an infinite plane. Anatomically based models were used in Cheng *et al.* (2004),[23] in which qualitatively normal gastric activity was constructed from 19 channels of magnetic recordings.

From an inverse point of view, this field is still in its infancy and much work has to be done. However, there is the possibility, given the complete lack of spatial information obtainable from the EGG (particularly compared with the ECG) that some useful information can be obtained via the use of inverse analysis. For instance, the EGG can fail to detect even the most simplest of rhythm disorders, for instance retrograde propagation, or asynchronous activation of the fundus and corpus occurring at the same frequency. Such activity can be potentially identified with a model-based inverse analysis.

Acknowledgments

Andrew Pullan currently holds a James Cook Research Fellowship and is grateful to the Royal Society of New Zealand for this funding. The application of source imaging ideas to the gastrointestinal field is supported by NIH grant RO1 DK64775. Andrew Pullan also acknowledges Dr. Buist, Dr. Cheng and Ms. Raghu for their respective inputs.

References

1. R. H. Adrian, W. K. Chandler and A. L. Hodgkin, Voltage clamp experiments in striated muscle fibers, *Journal of Physiology* **208** (1970) 607–644.
2. R. H. Adrian and L. D. Peachey, *Reconstruction of the Action Potential of Frog Sartorius Muscle*, *J. Physiol.* **235** (1973) 103–131.
3. W. C. Alvarez, The electrogastrogram and what it shows, *J. Am. Med. Assoc.* **78** (1922) 1116–1119.
4. R. R. Aliev, W. O. Richards and J. P. Wikswo, A simple nonlinear model of electrical activity in the intestine, *J. Theor. Biol.* **204** (2000) 21–28.
5. S. Baillet and L. Garnero, A Bayesian approach to introducing anatomo-functional priors in the EEG/MEG inverse problem, *IEEE Trans. Biomed. Eng.* **BME-44** (1997) 374–385.

6. S. Baillet, L. Garnero, G. Marin and J. Hugonin, Combined MEG and EEG source imaging by minimization of mutual information, *IEEE Trans. Biomed. Eng.* **BME-46** (1999) 522–534.
7. G. Barnes and A. Hillebrand, Statistical flattening of MEG beamformer images, *Neuroimage* **18** (2003) 1–12.
8. R. Barr, M. Ramsey and M. Spach, Relating epicardial to body surface potentials by means of transfer coefficients based on geometry measurements, *IEEE Trans. Biomed. Eng.* **BME-24** (1977) 1–11.
9. R. C. Barr and M. S. Spach, Inverse calculation of QRS-T epicardial potentials from body surface potential distributions for normal and ectopic beats in the intact dog, *Circ. Res.* **42** (1978) 661–675.
10. G. W. Beeler and H. Reuter, Reconstruction of the action potential of ventricular myocardial fibre, *Journal of Physiology* **268** (1977) 177–210.
11. K. L. Berrier, D. C. Sorensen and D. S. Khoury, Solving the inverse problem of electrocardiography using a Duncan and Horn formulation of the Kalman filter, *IEEE Trans. Biomed. Eng.* **51** (2004) 507–515.
12. C. Bertrand, Y. Hamada and H. Kado, MRI prior computation and parallel tempering algorithm: A probabilistic resolution of the MEG/EEG inverse problem, *Brain Topogr.* **14** (2001) 57–68.
13. see www.biosemi.com/products.htm.
14. see www.biosemi.com/faq/without_past.htm.
15. E. Bozler, Physiological evidence for the syncytial character of smooth muscle, *Science* **86** (1937) 476–478.
16. A. L. Bradshaw, A. H. Allos, J. P. Wikswo and W. O. Richards, Correlation and comparison of magnetic and electric detection of small intestinal electrical activity, *Am. J. Physiol.* **272** (1997) 1159–1167.
17. K. Brodmann, *Vergleichende Lokalisationslehre der Grosshirnrinde* (Barth, Leipzig, 1909).
18. D. H. Brooks, G. F. Ahmad, R. S. MacLeod and G. M. Maratos, Inverse electrocardiography by simultaneous imposition of multiple constraints, *IEEE Trans. Biomed. Eng.* **46** (1999) 3–18.
19. B. H. Brown, R. H. Smallwood, H. L. Duthie and C. J. Stoddard, Intestinal smooth muscle electrical potentials recorded from surface electrodes, *Med. Biol. Eng.* **13** (1975) 97–103.
20. J. E. Burnes, B. Taccardi, R. S. MacLeod and Y. Rudy, Non-invasive ECG imaging of electrophysiologically abnormal substrates in infarcted hearts, a model study, *Circulation* **101** (2000) 533–540.
21. J. E. Burnes, B. Taccardi, P. R. Ershler and Y. Rudy, Non-invasive ECG imaging of substrate and intramural ventricular tachycardia in infarcted hearts, *J. Am. Col. Cardiol.* **39** (2001) 2071–2078.
22. see www.cardiomag.com.
23. L. K. Cheng, M. L. Buist, J. A. Sims, R. L. Palmer, L. A. Bradshaw, W. O. Richards and A. J. Pullan, Spatial localization of gastric electrical activity using non-invasive magnetic field recordings, Proceedings of Digestive Disease Week, New Orleans, May 16–21, (1994).
24. L. Clerc, Directional differences of impulse spread in trabecular muscle from mammalian heart, *J. Physiol.* (London) **255** (1976) 335–346.
25. P. Colli-Franzone, L. Guerri, S. Tentoni, C. Viganotti, S. Baruffi, S. Spaggiari and B. Taccardi, A mathematical procedure for solving the inverse potential problem of electrocardiography, Analysis of the time-space accuracy from in vitro experimental data, *Math. Biosci.* **77** (1985) 353–396.

26. J. Cuppen and A. van Oosterom, Model studies with inversely calculated isochrones of ventricular depolarization, *IEEE Trans. Biomed. Eng.* **BME-31** (1984) 652–659.
27. A. Dale and M. Sereno, Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach, *J. Cogn. Neurosci.* **5** (1993) 162–176.
28. R. C. Davis, L. Garafolo and F. P. Gault, An exploration of abdominal potential, *J. Comp. Physiol. Psychol.* **52** (1957) 519–523.
29. J. De Munck, A. De Jongh and B. Van Dijk, The localization of spontaneous brain activity: An efficient way to analyze large data sets, *IEEE Trans. Biomed. Eng.* **BME-48** (2001) 1221–1228.
30. N. E. Diamant and A. Bortoff, Nature of the intestinal slow wave frequency gradient, *Am. J. Physiol.* **216** (1969) 301–307.
31. D. DiFrancesco and D. Noble, A model of cardiac electrical activity incorporating ionic pumps and concentration changes, *Phil. Trans. R. Soc. (Lond)* **B307** (1985) 353–398.
32. G. Edlingner, P. Wach and G. Pfurtscheller, On the realization of an analytic high-resolution EEG, *IEEE Trans. Biomed. Eng.* **BME-45** (1998) 736–745.
33. K. Friston, Statistical parametric maps in functional imaging: A general linear approach, *Human Brain Mapping* **2** (1995) 189–210.
34. R. A. Fitzhugh, Impulses and physiological states in theoretical models of nerve membrane, *Biophys. J.* **1** (1961) 445–466.
35. M. Fuchs, M. Wagner and J. Kastner, Boundary element method volume conductor models for EEG source reconstruction, *Clin. Neurophysiol.* **112** (2001) 1400–1407.
36. M. Fuchs, M. Wagner and J. Kastner, Confidence limits of dipole source reconstruction results, *Clin. Neurophysiol.* **115** (2004) 1442–1451.
37. M. Fuchs, M. Wagner, T. Kohler and H. Wischmann, Linear and nonlinear current density reconstructions, *J. Clin. Neurophysiol.* **16** (1999) 267–295.
38. R. N. Ghanem, J. E. Burnes, A. L. Waldo and Y. Rudy, Imaging dispersion of myocardial repolarization, II, *Circulation* **104** (2001) 1306–1312.
39. N. G. Gencer and I. O. Tanzer, Forward problem solution of electromagnetic source imaging using a new BEM formulation with high-order elements, *Phys. Med. Biol.* **44** (1999) 2275–2287.
40. L. Gepstein, G. Hayam and S. A. Ben-Haim, A novel method for nonfluoroscopic catheter-based electroanatomical mapping of the heart: In vitro and in vivo accuracy results, *Circulation* **95** (1996) 1611–1622.
41. A. Gevins, J. Le, P. Brickett, J. Desmond and B. Reutter, Seeing through the skull: Advanced EEGs use MRIs to accurately measure cortical activity from the scalp, *Brain Topogr.* **3** (1991) 53–64.
42. A. Gevins, M. Smith, L. McEvoy, H. Leong and J. Le, Electroencephalographic imaging of higher brain function, *Phil. Trans. R. Soc., Lond. B* **354** (1999) 1125–1134.
43. P. Gill, W. Murray and M. Wright, *Practical Optimization* (Academic Press, New York, 1981).
44. G. Golub and C. Van Loan, *Matrix Computations* (North Oxford Academic, Oxford, 1983).
45. S. Gonçalves, J. de Munck, J. Verbunt, F. Bijma, R. Heethaar and F. Lopes da Silva, In vivo measurement of the brain and skull resistivities using an EIT-based method and realistic models for the head, *IEEE Trans. Biomed. Eng.* **BME-50** (2003) 754–767.
46. R. Grave de Peralta Menendez, O. Hauk, S. Gonzalez Andino, H. Vogt and C. Michel, Linear inverse solutions with optimal resolution kernels applied to electromagnetic tomography, *Human Brain Mapping* **5** (1997) 454–476.

47. F. Greensite, The mathematical basis for imaging cardiac electrical function, *Critical Reviews in Biomedical Engineering* **22** (1994) 347–399.
48. F. Greensite, Well-posed formulation of the inverse problem of electrocardiography, *Ann. Biomed. Eng.* **22** (1994) 172–183.
49. F. Greensite, Remote reconstruction of confined wavefront propagation, *Inverse Problems* **11** (1995) 361–370.
50. F. Greensite and G. Huiskamp, An improved method for estimating epicardial potentials from the body surface, *IEEE Trans. Biomed. Eng.* **BME-45** (1998) 1–7.
51. F. Greensite, The temporal prior in bioelectromagnetic source imaging problems, *IEEE Trans. Biomed. Eng.* **50** (2003) 1152–1159.
52. F. Greensite, Dynamic medical imaging as a partial inverse problem, *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2004).
53. C. Groetsch, *Inverse Problems in the Mathematical Sciences* (Viewig, Braunschweig, 1993).
54. R. Gulrajani, *Bioelectricity and Biomagnetism* (John Wiley and Sons, New York, 1998).
55. R. Gulrajani, The forward problem of electrocardiography: Theoretical underpinnings and applications, in: *Modeling and Imaging of Bioelectric Activity*, Bin He (ed.), Kluwer, New York, 2004), pp. 281–319.
56. M. Hämäläinen, R. Hari, J. Ilmoniemi, J. Knuutila and O. Lounasmaa, MEG: Theory, instrumentation and applications to non-invasive studies of the working human brain, *Rev. Mod. Phys.* **65** (1993) 413–497.
57. M. Hämäläinen and R. Ilmoniemi, Interpreting magnetic fields in the brain: Minimium norm estimates, *Med. Biol. Eng. Comp.* **32** (1994) 35–42.
58. M. Hämäläinen and J. Sarvas, Realistic conductivity geometry model of the human head for the interpretation of neuromagnetic data, *IEEE Trans. Biomed. Eng.* **BME-36** (1989) 165–171.
59. B. He, Y. Wang and D. Wu, Estimating cortical potentials from scalp EEG's in a realistically shaped inhomogeneous head model by means of the boundary element method, *IEEE Trans. Biomed. Eng.* **BME-46** (1999) 1264–1268.
60. H. Helmholtz, Ueber einige Gezetze der Verteilung elektrischer Ströme in körperliche Leitern mit Anwendung auf die thierisch-elektrischen Versuche, *Pogg. Ann. Physik und Chemie* **89** (1853) 211–233; 353–377.
61. P. C. Hansen, Numerical tools for analysis and solution of Fredholm integral equations of the first kind, *Inverse Problems* **8** (1992) 849–872.
62. K. Henneberg and F. A. Roberge, Simulation of propagation along an isolated skeletal muscle fiber in an isotropic volume conductor, *Annals of Biomedical Engineering* **25** (1997) 15–28.
63. C. Henriquez, Simulating the electrical behavior of cardiac tissue using the bidomain model, *Crit. Rev. Biomed. Eng.* **21** (1993) 1–77.
64. A. L. Hodgkin and A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, *Journal of Physiology* **117** (1952) 500–544.
65. R. Hoekema, G. Wieneke, F. Leijten, C. van Veelen, P. van Rijen, G. Huiskamp, J. Ansems and A. van Huffelen, Measurement of the conductivity of skull, temporarily removed during epilepsy surgery, *Brain Topogr.* **16** (2003) 29–38.
66. G. J. M. Huiskamp and A. van Oosterom, The depolarization sequence of the human heart surface computed from measured body surface potentials, *IEEE Trans. Biomed. End.* **35** (1988) 1047–1058.

67. G. Huiskamp and F. Greensite, A new method for myocardial activation imaging, *IEEE Trans. Biomed. Eng.* **BME-44** (1997) 433–446.
68. G. Huiskamp, M. Vroeijenstijn, R. van Dijk, G. Wieneke and A. van Huffelen, The need for correct realistic geometry in the inverse EEG problem, *IEEE Trans. Biomed. Eng.* **46** (1999) 1281–1287.
69. G. Huiskamp, W. van der Meij, A. van Huffelen and O. van Nieuwenhuizen, High resolution spatio-temporal EEG-MEG analysis of rolandic spikes, *J. Clin. Neurophysiol.* (in press, 2004).
70. J. D. Huizinga, Physiology and pathophysiology of the interstitial cell of Cajal: From bench to bedside II. Gastric motility: Lessons from mutant mice on slow waves and innervation, *Am. J. Physiol.* **281** (2001) G1129–G1134.
71. R. E. Ideker, W. M. Smith, S. M. Blanchard, S. L. Reiser, E. V. Simpson, R. D. Wolf and N. D. Danieley, The assumptions of isochronal cardiac mapping, *PACE* **12** (1989) 456–478.
72. A. Irimia and L. A. Bradshaw, Theoretical and computational methods for the non-invasive detection of gastic electrical source coupling, *Physical Review E: Statistical Nonlinear and Soft Matter Physics* **69** (2004) 051920.
73. J. Jackson, *Classical Electrodynamics* (John Wiley and Sons, New York, 1975).
74. P. Jia, B. Punske, B. Taccardi and Y. Rudy, Electrophysiologic endocardial mapping from a non-contact non-expandable catheter, *J. Cardiovasc. Electrophysiol.* **11** (2000) 1238–1251.
75. P. Jia, B. Punske, B. Taccardi and Y. Rudy, Endocardial mapping of electrophysiologically abnormal substrates and cardiac arrhythmias using a non-contact non-expandable catheter, *J. Cardiovasc. Electrophys.* **13** (2002) 888–895.
76. C. R. Johnson, Adaptive finite element and local regularization methods for the inverse ECG problem, in: *Computational Inverse Problems in Electrocardiography* (WIT Press, Southampton, 2001), pp. 51–88.
77. A. Kadish, J. Hauck, B. Pederson, G. Beatty and C. Gornick, Mapping of atrial activation with a non-contact, multielectrode catheter in dogs, *Circulation* **99** (1999) 1906–1913.
78. J. P. Keener, *Principles of Applied Mathematics* (Addison-Wesley Publishing Co., Redwood City, CA, 1988).
79. D. S. Khoury, K. L. Berrier, S. M. Badruddin and W. A. Zoghbi, Three-dimensional electrophysiological imaging of the intact canine left ventricle using a non-contact multielectrode cavitary probe: Study of sinus, paced and spontaneous premature beats, *Circulation* **97** (1998) 399–409.
80. T. Knösche, E. Berends, H. Jagers and M. Peters, Determining the number of independent sources of the EEG: A simulation study on information criteria, *Brain Topogr.* **11** (1998) 111–124.
81. P. Laarne, P. Kauppinen, J. Hyttinen, J. Malmivuo and H. Eskola, Effects of tissue resistivities on electroencephalogram sensitivity distribution, *Med. Biol. Eng. Comput.* **37** (1999) 555–559.
82. J. Le and A. Gevins, Methods to reduce blur distortions from EEG's using a realistic head model, *IEEE Trans. Biomed. Eng.* **BME-40** (1993) 517–527.
83. N. Logothetis, The underpinnings of the BOLD functional magnetic resonance imaging signal, *J. Neurosci.* **23** (2003) 3963–3971.
84. C. H. Luo and Y. Rudy, A model of the ventricular cardiac action potential, depolarisation, repolarisation, and their interaction, *Circ. Res.* **68** (1991) 1501–1526.
85. C. H. Luo and Y. Rudy, A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes, *Circ. Res.* **74** (1994) 1071–1096.

86. R. L. Lux, C. R. Smith, R. F. Wyatt and J. A. Abildskow, Limited lead selection for estimation of body surface potential maps in electrocardiography, *IEEE Trans. Biomed. Eng.* **25** (1978) 270–276.
87. R. S. MacLeod, M. Gardner, R. M. Miller and B. M. Horacek, Application of an electrocardiographic inverse solution to localize ischemia during coronary angioplasty, *J. Cardiovasc. Electrophys.* **6** (1995) 2–18.
88. R. S. MacLeod and D. H. Brooks, Recent progress in inverse problems of electrocardiography, *IEEE Eng. Med. Biol.* **17** (1998) 73–83.
89. R. O. Martin and T. Pilkington, Statistically constrained inverse electrocardiography, *IEEE Trans. Biomed. Eng.* **BME-22** (1975) 487–492.
90. J. Meijs, O. Weier, M. Peters and A. van Oosterom, On the numerical accuracy of the boundary element method, *IEEE Trans. Biomed. Eng.* **BME-36** (1989) 1038–1049.
91. I. Merlet and J. Gotman, Dipole modeling of scalp electroencephalogram epileptic discharges: Correlation with intracerebral fields, *Clin. Neurophysiol.* **112** (2001) 414–430.
92. B. Messnarz, B. Tilg, R. Modre, G. Fischer and F. Hanser, A new spatiotemporal regularization approach for reconstruction of cardiac transmembrane potential patterns, *IEEE Trans. Biomed. Eng.* **51** (2004) 273–281.
93. R. N. Miftakhov, G. R. Abdusheva and J. Christensen, Numerical simulation of motility patterns of the small bowel. 1. Formulation of a mathematical model, *J. Theor. Biol.* **197** (1999) 89–112.
94. R. Modre, B. Tilg, G. Fischer and P. Wach, An iterative algorithm for myocardial activation time imaging, *Computer Methods and Programs in Biomedicine* **64** (2001) 1–7.
95. R. Modre, B. Tilg and G. Fischer, Stability of activation time imaging from single beat data under clinical conditions, *Biomedizinishe Technik* **46** (2001) 213–215.
96. J. Mosher, P. Lewis and R. Leahy, Multiple dipole modeling and localization from spatio-temporal MEG data, *IEEE Trans. Biomed. Eng.* **BME-39** (1992) 541–557.
97. J. Mosher and R. Leahy, Source localization using recursively applied and projected (RAP) MUSIC, *IEEE Trans. Signal Processing* **47** (1999) 332–340.
98. J. Nagumo, S. Animoto and S. Yoshizawa, An active pulse transmission line simulating nerve axon, *Proc. Inst. Radio Engineers* **50** (1962) 2061–2070.
99. M. P. Nash, C. P. Bradley, A. Kardos, A. J. Pullan and D. J. Paterson, An experimental model to correlate simultaneous body surface and epicardial electropotential recordings in vivo, *Chaos, Solitons and Fractals* **13** (2002) 1735–1742.
100. T. S. Nelsen and J. C. Becker, Simulation of the electrical and mechanical gradient of the small intestine, *Am. J. Physiol.* **214** (1968) 749–757.
101. see www.neuromag.com.
102. P. M. F. Nielsen, I. J. Le Grice, B. H. Smaill and P. J. Hunter, Mathematical model of geometry and fibrous structure of the heart, *Am. J. Physiol. 260(Heart Circ. Physiol. 29 1991)*, pp. H1365–H1378.
103. D. Noble, A. Varghese, P. Kohl and P. Noble, Improved guinea-pig ventricular cell model incorporating a diadic space, IKr and IKs, and length- and tension-dependent processes, *Can. J. Cardiol.* **14** (1998) 123–134.
104. P. Nunez and F. Westdorp, The surface Laplacian, high resolution EEG and controversies, *Brain Topogr.* **6** (1994) 221–226.
105. S. Ohyu, Y. Okamoto and S. Kuriki, Use of the ventricular propagated excitation model in the magnetocardiographic inverse problem for reconstruction of electrophysiological properties, *IEEE Trans. Biomed. Eng.* **49** (2002) 509–519.

106. M. Oishi, H. Otsubo, S. Kameyama, N. Morota, H. Masuda, M. Kitayama and R. Tanaka, Epileptic spikes: Magnetoencephalography versus simultaneous electrocorticography, *Epilepsia* **43** (2002) 1390–1395.
107. T. Oostendorp, R. MacLeod and A. van Oosterom, Non-invasive determination of the activation sequence of the heart: Validation with invasive data, *Proc. 19th Annual Int. Conf. IEEE EMBS*, CD-ROM (1997).
108. T. Oostendorp and K. Pesola, Non-invasive determination of the activation time sequence of the heart: Validation by comparison with invasive human data, in: *Computers in Cardiology*, New York: IEEE **25** (1998) 313–316.
109. P. Ossenblok, F. Leijten, J. de Munck, G. Huiskamp, F. Barkhof and P. Boon, Magnetic source imaging contributes to the presurgical identification of sensorimotor cortex in patients with frontal lobe epilepsy, *Clin. Neurophysiol.* **114** (2003) 221–232.
110. H. Oster, B. Taccardi, R. Lux, P. Ershler and Y. Rudy, Non-invasive electrocardiographic imaging, *Circulation* **96** (1997) 1012–1024.
111. R. Pascual-Marqui, C. Michel and D. Lehmann, Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain, *Int. J. Psychophysiol.* **18** (1994) 49–65.
112. T. Paul, J. P. Moak, C. Morris and A. Garson, Epicardial mapping: How to measure local activation, *PACE* **12** (1990) 285–292.
113. T. Paul, B. Windhagen-Mahnert, T. Kriebel, H. Bertram, R. Kaulitz, T. Korte, M. Niehaus and J. Tebbenjohanns, Atrial re-entrant tachycardia after surgery for congenital heart disease endocardial mapping and radiofrequency catheter ablation using a novel, non-contact mapping system, *Circulation* **103** (2001) 2266–2271.
114. C. J. Penney, J. C. Clements and B. M. Horacek, Non-invasive imaging of epicardial electrograms during controlled myocardial ischemia, *Computers in Cardiology 2000* **27** (2000) 103–106.
115. R. Plonsey, *Bioelectric Phenomena* (New York, McGraw-Hill, 1969).
116. M. Peters, J. Stinstra and I. Leveles, The electrical conductivity of living tissue: A parameter in the bioelectric inverse problem, in: *Modeling and Imaging of Bioelectric Activity*, Bin He (ed.), (Kluwer, New York, 2004), pp. 281–319.
117. A. Pullan, A high-order coupled finite element/boundary element torso model, *IEEE Trans. Biomed. Eng.* **43** (1996) 292–298.
118. A. J. Pullan, L. K. Cheng, M. P. Nash, C. P. Bradley and D. J. Paterson, Non-invasive electrical imaging of the heart — theory and model development, *Annals of Biomedical Engineering* **29** (2001) 817–836.
119. C. Ramanathan and Y. Rudy, Electrocardiographic Imaging: II. Effect of torso inhomogeneities on non-invasive reconstruction of epicardial potentials, electrograms and isochrones, *J. Cardiovasc. Electrophysiol.* **12** (2001) 242–252.
120. C. Ramanathan, R. Ghanem, P. Jia, K. Ryu and Y. Rudy, Non-invasive electrocardiographic imaging for cardiac electrophysiology and arrhythmia, *Nature Medicine* **10** (2004) 422–428.
121. P. Rosenfalck, Intra- and extracellular potential fields of active nerve and muscle fibers: A physico-mathematical analysis of different models, *Acta Physiologica Scandinavica Supplementum* **321** (1969) 1–168.
122. S. Sarna, K. Bowes and E. Daniel, Gastric pacemakers, *Gastro.* **70** (1976) 226–231.
123. R. J. Schilling, A. H. Kadish, N. S. Peters, J. Goldberger and D. Wyn Davies, Endocardial mapping of atrial fibrillation in the human right atrium using a non-contact catheter, *European Heart Journal* **21** (2000) 550–564.
124. R. O. Schmidt, Multiple emitter location and signal parameter estimation, *IEEE Transactions on Antennas and Propagation* **AP-34** (1986) 276–280.

125. C. Schmitt, B. Zrenner, M. Schneider, M. Karch, G. Ndrepepa, I. Deisenhofer, S. Weyerbrock, J. Schreieck and A. Schoemig, Clinical experience with a novel multielectrode basket catheter in right atrial tachycardias, *Circulation* **99** (1999) 2414–2422.
126. K. Sekihara, Generalized wiener estimation of three-dimensional current distribution from biomagnetic measurements, *IEEE Trans. Biomed. Eng.* **BME-43** (1996) 281–291.
127. A. V. Shahidi, P. Savard and R. Nadeau, Forward and inverse problems of electrocardiography: Modeling and recovery of epicardial potentials in humans, *IEEE Trans. Biomed. Eng.* **BME-41** (1994) 249–256.
128. A. S. Groenewegen, H. Spekhorst, N. M. van Hemel, J. H. Kingma, R. N. W. Hauer, J. M. T. de Bakker, C. A. Grimbergen, M. J. Janse and A. J. Dunning, Localization of the site of origin of postinfarction ventricular tachycardia by endocardial pace mapping. Body surface mapping compared with the 12-lead electrocardiogram, *Circulation* **88** (1993) 2290–2306.
129. S. A. Strickberger, B. P. Knight, G. F. Michaud, F. Pelosi and F. Morady, Mapping and ablation of ventricular tachycardia guided by virtual electrograms using a noncontact, computerized mapping system, *J. Am. Col. Cardiol.* **35** (2000) 414–421.
130. W. Sutherling, P. Crandall, T. Darcey, D. Becker, M. Levesque and D. Barth, The magnetic and electric fields agree with intracranial localizations of somatosensory cortex, *Neurology* **38** (1988) 1705–1714.
131. B. Taccardi, G. Arisi, E. Macchi, S. Baruffi and S. Spaggiari, A new intracavitary probe for detecting the site of origin of ectopic ventricular beats during one cardiac cycle, *Circulation* **75** (1987) 272–281.
132. R. D. Throne, L. G. Olson and J. R. Windle, A new method for incorporating weighted temporal and spatial smoothing in the inverse problem of electrocardiography, *IEEE Trans. Biomed. Eng.* **49** (2002) 1054–1059.
133. A. Tikhonov, *Solutions of Ill-Posed Problems* (John Wiley and Sons, New York, 1977).
134. B. Tilg, P. Wach, A. S. Groenwegen, G. Fischer, R. Modre, F. Roithinger, M. Mlynash, G. Reddyuu, T. Roberts, M. Lesh and P. Steiner, Closed-chest validation of source imaging from human ECG and MCG mapping data, *Proceedings of the 21st Annual International Conference of the IEEE EMBS, October 1999/First Joint BMES/EMBS Conference*, Atlanta, Georgia (1999).
135. B. Tilg, G. Fischer and R. Modre, Feasibility of activation time imaging within the human atria and ventricles in the catheter laboratory, *Biomedizinishe Technik* **46** (2001) 213–215.
136. B. Tilg, G. Fischer, R. Modre, F. Hanser, B. Messnarz, M. Schocke, C. Kremser, T. Berger, F. Hintringer and F. X. Roithinger, Model-based imaging of cardiac electrical excitation in humans, *IEEE Trans. Med. Imag.* **21** (2002) 1031–1039.
137. T. Tomita, Electrical properties of mammalian smooth muscle, in E. Bulbring, A. Brading, A. Jones and T. Tomita (eds.), *Smooth Muscle* (Baltimore, Williams and Wilkins, 1970), pp. 197–243.
138. D. S. Tuch, V. J. Wedeen, A. M. Dale and J. W. Belliveau, Conductivity maps of white matter fiber tracts using magnetic resonance diffusion tensor imaging, *Proc. Third Int. Conf. on Fundamental Mapping of the Human Brain*, *Neuroimage* **5** (1997) s44.
139. S. Ueno and N. Iriguchi, Impedance magnetic resonance imaging: A method for imaging of impedance distribution based on magnetic resonance imaging, *J. Appl. Phys.* **83** (1998) 6450–6452.

140. see www.unemap.com.
141. B. van Veen, W. van Drongelen, M. Yuchtman and A. Suzuki, Localization of brain electrical activity via linearly constrained minimum variance spatial filtering, *IEEE Trans. Biomed. Eng.* **BME-44** (1997) 867–880.
142. P. Wach, R. Modre, B. Tilg and G. Fischer, An iterative linearized optimization technique for nonlinear ill-posed problems applied to cardiac activation time imaging, *COMPEL* **20** (2001) 676–688.
143. S. Walker and D. Kilpatrick, Forward and inverse electrocardiographic calculations using resistor network models of the human torso, *Circ. Res.* **67** (1987) 504–513.
144. A. D. Waller, A demonstration on man of electromotive changes accompanying the heart's beat, *J. Physiol.* **8** (1887) 22934.
145. W. Wallinga, S. L. Meijer, M. J. Alberink, M. Vliek, E. D. Wienk and D. L. Ypey, Modelling action potentials and membrane currents of mammalian skeletal muscle fibers in coherence with potassium concentration changes in the T-tubular system, *European Biophysics Journal* **28** (1999) 317–329.
146. W. van der Meij, G. Huiskamp, G. Rutten, G. Wieneke, A. van Huffelen and O. van Nieuwenhuizen, The existence of two sources in rolandic epilepsy: Confirmation with high resolution EEG, MEG and fMRI, *Brain Topogr.* **13** (2001) 275–282.
147. N. Weisbrodt, Motility of the small intestine, in L. Johnson (ed.), *Physiology of the Gastrointestinal Tract* (Raven Press, New York, 1987), pp. 631–659.
148. G. Worrell, T. Lagerlund, F. Sharbrough, B. Brinkmann and N. Busacker, Localization of the epileptic focus by low-resolution electromagnetic tomography in patients with a lesion demonstrated by MRI, *Brain Topogr.* **12** (2000) 273–282.
149. Y. Yamashita and D. Geselowitz, Source-field relationships for cardiac generators on the heart surface based on their transfer coefficients, *IEEE Trans. Biomed. Eng.* **BME-32** (1985) 964–970.

This page intentionally left blank

CHAPTER 7

ELECTRICAL IMPEDANCE TOMOGRAPHY FOR IMAGING AND LESION ESTIMATION

JIN KEUN SEO
Department of Mathematics, Yonsei University
Seoul 120-749, Korea
seoj@yonsei.ac.kr

OHIN KWON
Department of Mathematics, Konkuk University
Seoul 143-701, Korea
oikwon@konkuk.ac.kr

EUNG JE WOO
Department of Biomedical Engineering, Kyung Hee University
Kyungki 449-701, Korea
ejwoo@khu.ac.kr

This chapter describes electrical impedance imaging techniques providing information on electrical properties of biological tissues. Sections on Electrical Impedance Tomography (EIT) deal with cross-sectional image reconstructions of conductivity and/or permittivity distributions from boundary measurements of current-voltage data. After mathematically defining the imaging problem in EIT, image reconstruction algorithms are discussed. Describing measurement techniques, examples of EIT images are presented. Noting that EIT suffers from the ill-posed nature of the corresponding inverse problem, Magnetic Resonance Electrical Impedance Tomography (MREIT) has been proposed to provide images with better spatial resolution and accuracy. MREIT utilizes internal information on the induced magnetic field in addition to the boundary measurements. Mathematical theory, algorithms, and experimental results of current MREIT research are described. As another way of handling the ill-posedness in EIT, lesion estimation techniques have been suggested. Assuming that a lesion has a different conductivity and/or permittivity compared with normal tissues, the lesion estimation including its location and size estimate could be a well-posed problem. Introducing mathematical models, lesion estimation algorithms, and measurement techniques, some of experimental results are explained. At the end of the chapter, potential applications of EIT and MREIT are discussed. Breast cancer detection is taken as the primary application area of the lesion estimation techniques.

Keywords: Electrical impedance tomography; static imaging; MREIT; lesion estimation.

1. Introduction

Numerous experimental findings have shown that different biological tissues in the human body have different electrical properties of conductivity and permittivity at the frequency range of tens of Hz to several MHz.[1] These properties change with ion concentrations in extra and intracellular fluids, cellular structure and density, molecular composition, membrane capacitance, and so on. Therefore, they manifest structural, functional, and pathological conditions of the tissue providing valuable diagnostic information.

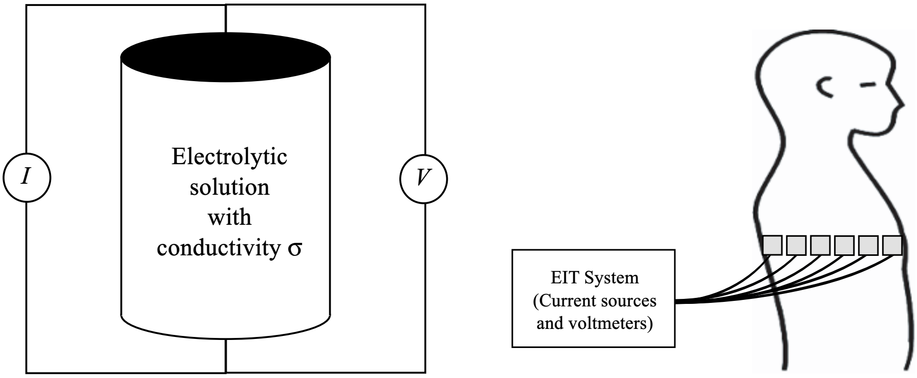
In most studies of electrical bioimpedance, we either inject current or apply voltage using electrodes attached on a subject. Measuring the induced voltage or current, electrical properties of the subject are evaluated. An extracted tissue sample could also be investigated for *in vitro* experimental studies. Electrical Impedance Spectroscopy (EIS) is a technique to characterize a biological tissue from its bioimpedance measurements at multiple frequencies. In Electrical Impedance Tomography (EIT), we try to reconstruct cross-sectional images of conductivity and/or permittivity distributions inside the subject.

For example, consider the problem of finding the conductivity of a homogeneous electrolytic solution inside a plastic cylindrical container shown in Fig. 1(a). Attaching electrodes on the top and bottom of the container, we inject current I and measure the induced voltage V between them. Then, the conductivity σ can be found from the Ohm's law:

$$R = \frac{V}{I} = \frac{L}{\sigma S}, \tag{1}$$

where R is the measured resistance, L the length, and S the cross-sectional area of the cylinder. Here, we neglected the characteristics of the electrode-electrolyte interface.

When the subject is a human body as shown in Fig. 1(b), it is not cylindrical and its conductivity distribution is inhomogeneous. Furthermore, there are limited areas where we can attach electrodes. In EIT, therefore, finding the inhomogeneous conductivity and/or permittivity distribution in a form of cross-sectional image becomes a complicated inverse problem also known as the inverse conductivity


I
V
Electrolytic solution with conductivity σ
EIT System (Current sources and voltmeters)

(a) (b)

Fig. 1. Problems of finding (a) a single conductivity value of a homogeneous electrolytic solution and (b) an inhomogeneous conductivity distribution inside the human body.

problem. In most cases, surface electrodes as many as 8 to 256 are attached in a two or three-dimensional configuration. Injecting patterns of currents through all or chosen pairs of electrodes, induced boundary voltages on all or selected electrodes are measured. The measured boundary current-voltage data set is utilized to reconstruct cross-sectional images of the conductivity and/or permittivity distribution inside the subject.[2–6] It is known that most biological tissues are anisotropic in terms of their electrical properties. Throughout this chapter, however, we only consider isotropic cases since there is no solid technique yet to deal with the anisotropy in EIT.

When we inject current into a subject, the internal current pathway or the current density distribution is affected nonlinearly by the global structure of the electrical properties of the subject. Any change in the conductivity of an internal region alters the current pathway and its effect is conveyed to the corresponding change in the boundary voltage. However, these boundary measurements are very insensitive to a local change away from measuring points. For this reason, EIT suffers from the ill-posed characteristics of the corresponding inverse problem. This makes it difficult to reconstruct accurate conductivity and/or permittivity images with a high spatial resolution under realistic environments where modeling and measurement errors are unavoidable.

In practice, we can attach only a limited number of electrodes and this means that we are limited by a fixed amount of information from boundary measurements. Using a larger number of electrodes requires a more complicated instrument and cumbersome electrode attachment procedure and these are prone to increase the total amount of measurement errors in practice. With these kinds of technical restrictions, it is desirable for EIT to find clinical applications where its portability and high temporal resolution to monitor changes in electrical properties are significant merits. This kind of impedance imaging has been called the dynamic or difference imaging in EIT. In the following sections on EIT, we will describe its mathematical formulation, image reconstruction algorithms, measurement techniques, and examples of EIT images.

For the static or absolute imaging, Magnetic Resonance Electrical Impedance Tomography (MREIT) has been lately proposed to overcome the technical limitations of EIT.[7–12] When we inject current into a subject, it produces distributions of voltage, current density, and also magnetic flux density inside the subject. The basic idea of MREIT is to utilize the information on magnetic as well as electric field induced by an injection current. While EIT is limited by the boundary measurements of current-voltage data, MREIT utilizes the internal magnetic flux density data obtained using a Magnetic Resonance Imaging (MRI) scanner. Recent progress in MREIT providing conductivity images with an improved spatial resolution and accuracy is presented including the mathematical theory, algorithms, and experimental outcomes.

While MREIT is capable of static impedance imaging with better performance, it requires an MRI scanner. As another way of handling the ill-posedness in EIT, lesion estimation techniques are suggested.[13–15] They utilize boundary measurements of current-voltage data as in EIT. However, the goal is not the cross-sectional imaging but the extraction of core information on any lesions with different electrical properties. Introducing theory and algorithms for feature extractions, experimental results and potential applications are discussed.

At the end of the chapter, we will introduce most promising application areas of impedance imaging. Dynamic imaging using EIT techniques are mainly for functional imaging and monitoring of physiological events. Providing static images of conductivity distributions, MREIT could find important contributions in the areas of neuronal source localization and mapping. There are numerous methods of applying electromagnetic energy to the human body mostly for therapeutic purposes. Conductivity information from MREIT will be valuable for the optimization and evaluation of these therapeutic treatments. Feature extraction of lesions may find applications in breast cancer detection, non-destructive testing such as bubble detection in two-phase flow, and others.

2. Electrical Impedance Tomography (EIT)

After many years of worldwide research investments on EIT, several review papers describe numerous aspects of the EIT technique.[2–6] We can classify previous works into two categories. One is a kind of works with theoretical meanings that enhance our understanding of the problem. All other works with practical applicability in mind belong to the other kind. In this second category, we may often have to sacrifice the mathematical rigorousness for practical implementations.

We will first introduce mathematical theories and image reconstruction algorithms in EIT. After discussing the ill-posed nature of the EIT problem together with experimental limitations, typical examples of EIT images are presented. The dynamic EIT imaging rather than static imaging is emphasized in the following sections on EIT.

2.1. *Mathematical formulation of EIT*

2.1.1. *Forward problem: Voltage due to an injection current for a given conductivity*

Let $\Omega \subset \mathbb{R}^3$ be an electrically conducting subject with its boundary $\partial\Omega$ as shown in Fig. 2. Surface electrodes $\mathcal{E}_j$ for $j = 1, \ldots, E$ are attached on the boundary $\partial\Omega$. Assume that we inject current I at a fixed angular frequency ω through a pair of chosen electrodes. Denoting a position vector in $\mathbb{R}^3$ as $\mathbf{r}$, the time harmonic electric field $\mathbf{E}$ and magnetic flux density $\mathbf{B}$ due to the injection current satisfy the time harmonic Maxwell equations:

$$\nabla \times \left(\frac{1}{\mu}\mathbf{B}(\mathbf{r})\right) = (\sigma(\mathbf{r}) + i\omega\epsilon(\mathbf{r}))\mathbf{E}(\mathbf{r}) \quad \text{and} \quad \nabla \times \mathbf{E}(\mathbf{r}) = -i\omega\mathbf{B}(\mathbf{r}),$$

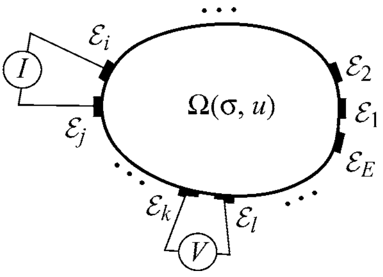


Fig. 2. Electrically conducting subject Ω with a conductivity σ and voltage u distribution. Surface electrodes $\mathcal{E}_j, j = 1, \ldots, E$ are attached on the boundary $\partial\Omega$. Here, we assume that current is injected between the pair of electrodes $\mathcal{E}_i$ and $\mathcal{E}_j$ and voltage is measured between $\mathcal{E}_k$ and $\mathcal{E}_l$.

where σ is the conductivity, ϵ permittivity, and μ magnetic permeability. We may express a complex conductivity τ at $\mathbf{r} \in \Omega$ as

$$\tau(\mathbf{r}) = \sigma(\mathbf{r}) + i\omega\epsilon(\mathbf{r}). \tag{2}$$

Since $\nabla \cdot \nabla \times \left(\frac{1}{\mu}\mathbf{B}(\mathbf{r})\right) = 0$, $\mathbf{E}$ satisfies $\nabla \cdot (\tau(\mathbf{r})\mathbf{E}(\mathbf{r})) = 0$. At a relatively low frequency range, we may assume $\nabla \times \mathbf{E}(\mathbf{r}) = \mathbf{0}$. Then, we can define a voltage distribution u in Ω satisfying $-\nabla u(\mathbf{r}) = \mathbf{E}(\mathbf{r})$. Now, we can formulate the following boundary value problem with the Neumann boundary condition:

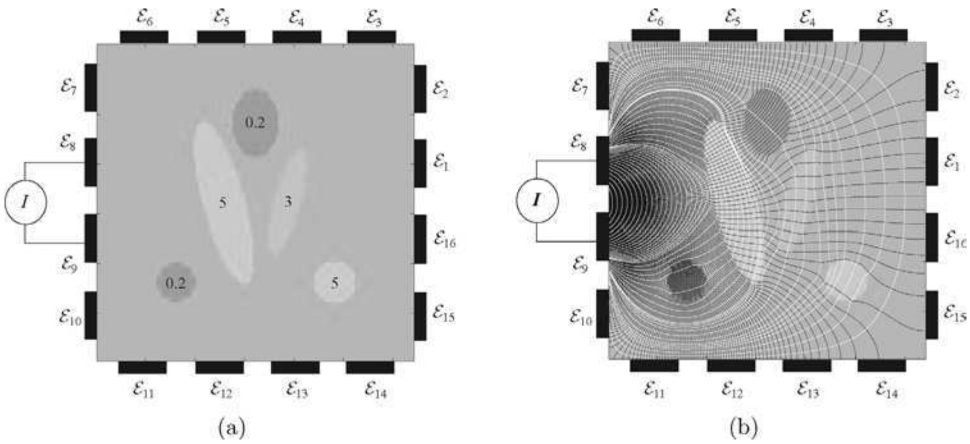
$$\begin{cases} \nabla \cdot [\tau(\mathbf{r})\nabla u(\mathbf{r})] = 0 \text{ in } \Omega \\ -\tau\nabla u \cdot \mathbf{n} = g \text{ on } \partial\Omega \end{cases}, \tag{3}$$

where $\mathbf{n}$ is the outward unit normal vector on $\partial\Omega$ and g the magnitude of the current density on $\partial\Omega$ due to the injection current I. On a current injection electrode $\mathcal{E}_j$, we have $\int_{\mathcal{E}_j} g\, ds = \pm I$ where ds is the surface element and the sign depends on the direction of the injection current. The Neumann data g is zero on the regions of the boundary not contacting with current injection electrodes. Setting a reference voltage $u(\mathbf{r}_0) = 0$ for $\mathbf{r}_0 \in \partial\Omega$, we can obtain a unique solution u of (3) from σ and g. The current density $\mathbf{J}$ in Ω due to the injection current is given by

$$\mathbf{J}(\mathbf{r}) = -\tau(\mathbf{r})\nabla u(\mathbf{r}) = \tau(\mathbf{r})\mathbf{E}(\mathbf{r}) \quad \text{in } \Omega. \tag{4}$$

If we apply currents with sinusoidal waveform at a low frequency range, the imaginary part of τ, $\omega\epsilon$, is relatively small. Throughout this chapter, we will assume $\tau = \sigma$ for simplicity. However, note that the permittivity ϵ is also frequently imaged in EIT in addition to the conductivity σ. By EIT technique, we cannot determine an anisotropic conductivity distribution of a subject, since the boundary information of current-voltage data are not sufficient to recover the anisotropic conductivity distribution. This means that many different anisotropic conductivity distributions may have the same boundary measurements in EIT.[16] For this reason, we will assume that σ is isotropic.

Using a simple numerical example, we now illustrate how the voltage and current density distributions are formed from (3) and (4). Figure 3(a) shows



(a) (b)

Fig. 3. (a) An example of an electrically conducting subject with a given conductivity distribution. Numbers inside ellipsoids are conductivity values in S/m. (b) Voltage and current density distribution induced by the injection current. Black and white lines are equipotential and current density streamlines, respectively.

a two-dimensional model of an electrically conducting subject with a given conductivity distribution σ. With 16 electrodes on its boundary, the injection current is applied between the chosen pair of electrodes. Figure 3(b) shows the computed voltage and current density distribution using the Finite Element Method (FEM).[17,18]

In EIT, we measure the boundary voltage f, the restriction of u to the boundary $\partial\Omega$, to reconstruct an image of τ. If the subject is homogeneous, a single measurement of the boundary current-voltage pair (g, f) will be enough to determine the constant τ. For the general case where the subject is inhomogeneous, the reconstruction of the target image τ requires applying several independent currents g_j, $j = 1, \ldots, N$ and measuring the corresponding boundary voltages f_j. As an example, Figs. 4(a) and (b) show the Neumann data g and voltage f, respectively, on the boundary of the model in Fig. 3.

With E electrodes attached to the subject as in Fig. 2, there could be different methods to inject several independent currents.[2,3] Since all currents entering the subject must exit from it to form a closed circuit including one or multiple current sources, any injection current must satisfy $\int_{\partial\Omega} g_j ds = 0$ and the number N of the independent injection currents should be $N \leq (E - 1)$. For each injection current g_j, we can measure $(E-1)$ boundary voltages since we use one electrode as the reference electrode. However, the reciprocity theorem tells that the boundary voltage between a pair of electrodes $\mathcal{E}_k$ and $\mathcal{E}_l$ due to an injection current between another pair $\mathcal{E}_i$ and $\mathcal{E}_j$ is the same as the case where the two pairs exchange their roles. Removing the redundant data, the number M of the independent voltage measurements satisfies $M \leq \frac{(E-1)(E-1)+(E-1)}{2} = \frac{E(E-1)}{2}$. This indicates that the

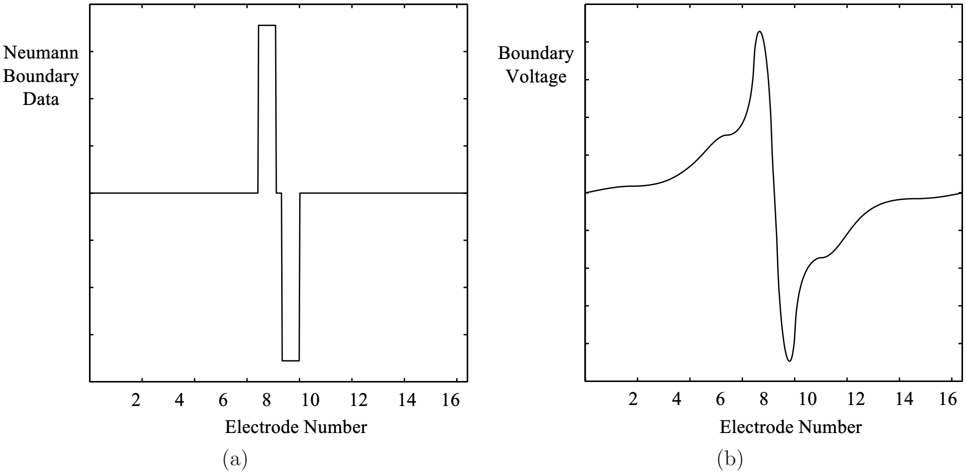


Fig. 4. (a) Neumann boundary data of the model in Fig. 3. We assumed that the current density underneath each current injection electrode is uniform. (b) Boundary voltage on 16 electrodes of the model. Here, we neglected the effects of electrode contact impedances.

number of electrodes is the primary limiting factor determining the total amount of information available from the boundary measurements.

Now, the inverse problem is to reconstruct τ from the boundary current-voltage pairs (g_j, f_j), $j = 1, \ldots, N$. To search τ, it is necessary to interpret how the change in conductivity distribution affect the current-to-voltage relation on the boundary. Also, we should take inevitable measurement noise and modeling errors into account.

2.1.2. *Inverse problem: Conductivity from Neumann-to-Dirichlet (NtD) map*

There has been a significant progress in theoretical study of the inverse problem in EIT. To provide a quick survey on these results, we denote the solution u of the Neumann boundary value problem in (3) by $u[\sigma, g]$ because it is determined uniquely by the conductivity distribution σ and the Neumann data g. To explain the inverse problem, we define the map Λ_σ: $g \to f$ by $\Lambda_\sigma[g] = u[\sigma, g]|_{\partial\Omega}$ where $u|_{\partial\Omega}$ is the restriction of u to $\partial\Omega$. The map Λ_σ is called the Neumann-to-Dirichlet (NtD) map. The reconstruction of σ requires us to invert the following map:

$$\sigma \to \{(g_j, f_j)\}_{j=1}^{N}, \quad \text{with } f_j = \Lambda_\sigma[g_j]$$

for a given sequence of injection currents $\{g_j\}_{j=1}^{N}$.

The NtD map Λ_σ is closely related with the Neumann function restricted on $\partial\Omega$. The Neumann function $\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')$ is the solution of the following Neumann problem: for each $\mathbf{r}$,

$$\begin{cases} \nabla_{\mathbf{r}'} \cdot (\sigma(\mathbf{r}')\nabla_{\mathbf{r}'}\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')) = \delta(\mathbf{r} - \mathbf{r}') \quad \text{for all } \mathbf{r}' \in \Omega \\ \sigma(\mathbf{r}')\nabla_{\mathbf{r}'}\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}') \cdot \mathbf{n}(\mathbf{r}') = 0 \quad \text{for all } \mathbf{r}' \in \partial\Omega \end{cases},$$

where δ is the Dirac delta function. With the use of the Neumann function $\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')$, we can represent $u[\sigma, g](\mathbf{r})$ in terms of the singular integral:

$$\begin{aligned} u[\sigma, g](\mathbf{r}) &= \int_\Omega \delta(\mathbf{r} - \mathbf{r}') u[\sigma, g](\mathbf{r}') d\mathbf{r}' \\ &= \int_\Omega \nabla \cdot (\sigma(\mathbf{r}') \nabla \mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')) u[\sigma, g](\mathbf{r}') d\mathbf{r}' \\ &= -\int_\Omega \sigma(\mathbf{r}') \nabla \mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}') \cdot \nabla u[\sigma, g](\mathbf{r}') d\mathbf{r}' \\ &= \int_{\partial\Omega} \mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}') g(\mathbf{r}') ds_{\mathbf{r}'}. \end{aligned}$$

Since $\Lambda_\sigma[g]$ is the restriction of $u[\sigma, g]$ to the boundary $\partial\Omega$, it can be represented as

$$\Lambda_\sigma[g](\mathbf{r}) = \int_{\partial\Omega} \mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}') g(\mathbf{r}') ds_{\mathbf{r}'}, \quad \mathbf{r} \in \partial\Omega. \tag{5}$$

Therefore, the kernel $\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')$ with $\mathbf{r}, \mathbf{r}' \in \partial\Omega$ can be viewed as a kind of an expression of the NtD map Λ_σ. Note that the map Λ_σ is sensitive to a change of the geometry of the surface $\partial\Omega$ since $\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')$ is singular at $\mathbf{r} = \mathbf{r}'$.

It has been proved that the knowledge of Λ_σ is sufficient to uniquely determine σ provided that σ satisfies some minor regularity condition. For the uniqueness results, please see the previous works.[19–27] Nachman provided a direct method of reconstructing a two dimensional conductivity from the NtD map Λ_σ.[26]

To reconstruct σ, it would be ideal if we could measure the full NtD map Λ_σ. However, in practice, it is not possible to get the complete knowledge of Λ_σ due to a limited number of electrodes and the difficulty in capturing the geometry of $\partial\Omega$. Furthermore, it is well known that this map is highly nonlinear and insensitive to perturbations of an internal σ. Hence, it is generally well accepted that there is no stable way to reconstruct σ with a high spatial resolution.

2.2. *EIT image reconstruction techniques*

2.2.1. *Dynamic imaging using sensitivity matrix method*

Dynamic imaging in EIT produces images of any difference in conductivity distributions between two states. It could be a temporal change or frequency-dependent change in a multi-frequency EIT system. The dynamic EIT imaging was first proposed by Barber and Brown[28] using the backprojection method following numerous variations. In this section, however, we describe the dynamic imaging by introducing the sensitivity matrix method since it is more widely used.[4,29,30]

In (5), we expressed the voltage $u[\sigma, g]$ with the aid of Neumann kernel $\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')$ and it is rewritten as

$$u[\sigma, g](\mathbf{r}) = \int_{\partial\Omega} \mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}') g(\mathbf{r}') ds_{\mathbf{r}'}, \quad \mathbf{r} \in \partial\Omega. \tag{6}$$

Providing that we have a full knowledge of the NtD map, we can compute the Neumann kernel $\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')$ for $\mathbf{r}, \mathbf{r}' \in \partial\Omega$ and, therefore, the inverse problem becomes determining σ from the Neumann kernel $\mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}')$. However, any given electrode configuration with a finite number of electrodes provides only a partial knowledge of the NtD map with unavoidable measurement noise in it. Also, for a numerical implementation, an appropriate discrete version of the inverse problem should be considered. In such a discrete version, we must pay special attention to the effects of inaccurate and incomplete data set due to the ill-posed nature (the prime cause of inaccuracy in EIT imaging) of the problem.

Let E be the number of electrodes and $\mathcal{E}_j$ with $j = 1, \ldots, E$ indicates each surface electrode. We denote by $\chi_{\mathcal{E}}$ the indicator function of $\mathcal{E}$; $\chi_{\mathcal{E}} = 1$ on $\mathcal{E}$ and 0 otherwise. As described before, we inject $(E-1)$ number of linearly independent currents I_j for $j = 1, \ldots, (E-1)$. Let g_j denote the corresponding Neumann data due to the current I_j. In order to derive a discrete version of (6), we multiply both sides of (6) by $\chi_{\mathcal{E}_k}$, $k = 1, \ldots, E$ and integrate them over the boundary $\partial\Omega$:

$$\begin{bmatrix} \int_{\partial\Omega} u[\sigma, g_j](\mathbf{r}) \chi_{\mathcal{E}_1}(\mathbf{r}) ds \\ \vdots \\ \int_{\partial\Omega} u[\sigma, g_j](\mathbf{r}) \chi_{\mathcal{E}_E}(\mathbf{r}) ds \end{bmatrix} = \int_{\partial\Omega} \begin{bmatrix} \int_{\partial\Omega} \mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}') \chi_{\mathcal{E}_1}(\mathbf{r}) ds \\ \vdots \\ \int_{\partial\Omega} \mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}') \chi_{\mathcal{E}_E}(\mathbf{r}) ds \end{bmatrix} \quad g_j(\mathbf{r}') ds \tag{7}$$

for each g_j. The Neumann data g_j is supported on $\bigcup_{k=1}^{E} \mathcal{E}_k$ and we assume that it is uniform underneath each electrode. Then, g_j can be approximated as

$$g_j(\mathbf{r}) \approx \sum_{k=1}^{E} \left(\frac{1}{a_k} \int_{\mathcal{E}_k} g_j \, ds \right) \chi_{\mathcal{E}_k}(\mathbf{r}), \quad \mathbf{r} \in \partial\Omega,$$

where a_k is the surface area of $\mathcal{E}_k$. Here, we used the simple uniform current density electrode model. One may consider the complete electrode model[31,32] that results in the same sensitivity matrix.[17]

Substituting the above approximation into (7) yields

$$\mathbf{u}_j^*[\sigma] = \mathbb{N}^*[\sigma] \mathbf{g}_j^*, \tag{8}$$

where

$$\mathbf{u}_j^*[\sigma] = \begin{bmatrix} \frac{1}{a_1} \int_{\mathcal{E}_1} u[\sigma, g_j](\mathbf{r}) ds \\ \vdots \\ \frac{1}{a_E} \int_{\mathcal{E}_E} u[\sigma, g_j](\mathbf{r}) ds \end{bmatrix}, \quad \mathbf{g}_j^* = \begin{bmatrix} \frac{1}{a_1} \int_{\mathcal{E}_1} g_j ds \\ \vdots \\ \frac{1}{a_E} \int_{\mathcal{E}_E} g_j ds \end{bmatrix},$$

and $\mathbb{N}^*[\sigma]$ is the $E \times E$ matrix with

$$(i, j)\text{th-component of } \mathbb{N}^*[\sigma] = \frac{1}{a_i} \int_{\mathcal{E}_i} \int_{\mathcal{E}_j} \mathcal{N}_\sigma(\mathbf{r}, \mathbf{r}') ds_{\mathbf{r}} \, ds_{\mathbf{r}'}.$$

Due to the reciprocity relation between current and voltage, the matrix $\mathbb{N}^*[\sigma]$ is symmetric.

Since we use one electrode, say $\mathcal{E}_E$, as a reference electrode, the Eth component of $\mathbf{u}_j^*$ does not have any information. Moreover, since $\int_{\partial\Omega} g_j\, ds = 0$, the Eth component of $\mathbf{g}_j$ is determined by the remaining $(E-1)$ components of $\mathbf{g}_j^*$. This means that the system (8) having the $E \times E$ matrix is essentially the same as the following reduced system having the $(E-1)\times(E-1)$ matrix:

$$\mathbf{u}_j[\sigma] = \mathbb{N}[\sigma]\mathbf{g}_j, \tag{9}$$

where $\mathbb{N}[\sigma]$ is the $(E-1)\times(E-1)$ sub-matrix of $\mathbb{N}^*[\sigma]$ eliminating its Eth row and Eth column and $\mathbf{u}_j$ and $\mathbf{g}_j$ are the corresponding sub-vectors of $\mathbf{u}_j^*$ and $\mathbf{g}_j^*$, respectively.

The $(E-1)\times(E-1)$ matrix $\mathbb{N}[\sigma]$ having $\frac{E(E-1)}{2}$ number of unknown components can be determined from the relationship between $(E-1)$ number of linearly independent currents g_j and the corresponding measured voltages f_j. The inverse problem is to determine σ from $\mathbb{N}[\sigma]$ containing $\frac{E(E-1)}{2}$ data. It should be noticed that the resolution of a reconstructed conductivity image σ is restricted by the number of data, $\frac{E(E-1)}{2}$. Therefore, the reconstruction with a limited knowledge leads us to approximate σ as a piece-wise constant function:

$$\sigma = \sum_{j=1}^{P} \sigma_j \chi_{\Delta_j}, \quad P \le \frac{E(E-1)}{2},$$

where σ_j is a constant and Δ_j is a voxel, a partition of the domain Ω. With this approximation, σ can be viewed as a vector, $\sigma = (\sigma_1, \ldots, \sigma_P)$.

Let the vector $\sigma \in \mathbb{R}^P$ be a small perturbation of $\sigma_0 \in \mathbb{R}^P$ with a perturbing term $\delta\sigma \in \mathbb{R}^P$:

$$\sigma = \sigma_0 + \delta\sigma.$$

We can view σ_0 as the background conductivity of a subject and $\delta\sigma$ as a change of the conductivity due to a physiological event. The mathematical problem in dynamic impedance imaging is to visualize $\delta\sigma$.

From (9), we have

$$\mathbf{u}_j[\sigma] - \mathbf{u}_j[\sigma_0] = (\mathbb{N}[\sigma] - \mathbb{N}[\sigma_0])\mathbf{g}_j \approx \delta\sigma \cdot \nabla_\sigma \left(\mathbb{N}[\sigma_0]\mathbf{g}_j\right), \quad j = 1, \ldots, E-1. \tag{10}$$

Unification of the above approximations with all injection currents leads to

$$\delta\mathbf{u}[\sigma] = \mathbb{S}[\sigma_0]\delta\sigma,$$

where

$$\delta\mathbf{u}[\sigma] = \begin{bmatrix} \mathbf{u}_1[\sigma] - \mathbf{u}_1[\sigma_0] \\ \vdots \\ \mathbf{u}_{E-1}[\sigma] - \mathbf{u}_{E-1}[\sigma_0] \end{bmatrix} \in \mathbb{R}^{(E-1)^2}$$

and $\mathbb{S}[\sigma_0]$ is a $(E-1)^2 \times P$ matrix given by

$$\mathbb{S}[\sigma_0] = \begin{bmatrix} \vdots \\ \boxed{\begin{matrix} ((E-1)(j+1))\text{th row of } \mathbb{S}[\sigma_0] \\ \vdots \\ ((E-1)(j+1))\text{th row of } \mathbb{S}[\sigma_0] \end{matrix}} \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \vdots \\ \boxed{\begin{matrix} \uparrow & & \uparrow \\ \frac{\partial}{\partial\sigma_1}\left(\mathbb{N}[\sigma_0]\mathbf{g}_j\right) & \cdots & \frac{\partial}{\partial\sigma_m}\left(\mathbb{N}[\sigma_0]\mathbf{g}_j\right) \\ \downarrow & & \downarrow \end{matrix}} \\ \vdots \end{bmatrix}.$$

The box in the above equation corresponds to the formula (10) for a certain j, and the matrix $\mathbb{S}[\sigma_0]$ is the simple arrangement of the contents within the box for $j = 1, \ldots, E-1$. This matrix $\mathbb{S}[\sigma_0]$ is called the sensitivity matrix.

We now explain this sensitivity matrix in a variational point of view. If we know the background conductivity σ_0, we can compute $u[\sigma_0, g_j]$ by solving the direct problem:

$$\nabla \cdot \big(\sigma_0 \nabla u[\sigma_0, g_j]\big) = 0 \quad \text{in } \Omega, \qquad -\sigma_0 \frac{\partial u[\sigma_0, g_j]}{\partial \mathbf{n}} = g_j \quad \text{on } \partial\Omega.$$

For simplicity, we write $u[\sigma_0, g_k] = u_k^0$ and $u[\sigma, g_k] = u_k = u_k^0 + \delta u_k$. We try to reconstruct the conductivity change $\delta\sigma$ from the voltage difference $(u_k - u_k^0)$ on the electrodes. Using the properties of u_k and u_k^0 and the divergence theorem, we have the following identity:

$$\begin{aligned} \int_\Omega \delta\sigma \nabla u_j^0 \nabla u_k^0 \, d\mathbf{r} &= \int_\Omega [\sigma \nabla u_j \nabla u_k^0 - \sigma_0 \nabla u_j \nabla u_k^0] d\mathbf{r} - \int_\Omega \delta\sigma \nabla \delta u_j \nabla u_k^0 \, d\mathbf{r} \\ &= \int_{\partial\Omega} g_j [u_k^0 - u_k] ds - \int_\Omega \delta\sigma \nabla \delta u_j \nabla u_k^0 \, d\mathbf{r}. \end{aligned}$$

Since the last term in the above identity can be regarded as negligibly small, we obtain

$$\int_\Omega \delta\sigma \nabla u_j^0 \nabla u_k^0 \, d\mathbf{r} = \int_{\partial\Omega} g_j [u_k^0 - u_k] ds.$$

Since we assume that $\sigma = \sum_{j=1}^{P} \sigma_j \chi_{\Delta_j}$ is piecewise constant, the above identity can be expressed as the following matrix form:

$$\mathbb{A}[\sigma_0]\delta\sigma = \mathbf{b}, \tag{11}$$

where

$$\mathbf{b} = \begin{bmatrix} b[1,1] \\ \vdots \\ b[1,E-1] \\ b[2,1] \\ \vdots \\ \vdots \\ b[E-1,E-1] \end{bmatrix} \in \mathbb{R}^{(E-1)^2} \quad \text{with } b[j,k] = \int_{\partial\Omega} g_j [u_k^0 - u_k] ds,$$

and $\mathbb{A}[\sigma_0]$ is the $(E-1)^2 \times P$ matrix given by

$$\mathbb{A}[\sigma_0] = \begin{bmatrix} \vdots \\ \boxed{\begin{matrix} ((E-1)(j+1))\text{th row} \\ \vdots \\ ((E-1)(j+1))\text{th row} \end{matrix}} \\ \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \vdots \\ \boxed{\begin{matrix} \int_{\Delta_1} \nabla u_j^0 \nabla u_1^0 d\mathbf{r} & \cdots & \int_{\Delta_P} \nabla u_j^0 \nabla u_1^0 d\mathbf{r} \\ \vdots & \cdots & \vdots \\ \int_{\Delta_1} \nabla u_j^0 \nabla u_{E-1}^0 d\mathbf{r} & \cdots & \int_{\Delta_P} \nabla u_j^0 \nabla u_{E-1}^0 d\mathbf{r} \end{matrix}} \\ \vdots \end{bmatrix}.$$

The biggest advantage of the dynamic imaging comes from the fact that the right hand side of (11) contains the difference in measured boundary voltage, $u_k^0 - u_k$. Since any systematic errors common to both u_k^0 and u_k are cancelled out, the dynamic imaging could be more robust against different kinds of noise. However, in order to solve the system (11) for $\delta\sigma$, we need to compute u_k^0 and the background conductivity σ_0 is usually unknown in advance. So, in most dynamic imaging methods, a homogeneous conductivity distribution is often used as σ_0. Alternatively, we may utilize as much *a priori* information as possible to setup a better background conductivity distribution σ_0.

Now, one may wonder if we use an iterative scheme to reconstruct a static image of σ starting from a homogeneous initial guess σ_0. Actually, this is the basic idea of the static imaging in EIT as described in the next section. If successful, the static imaging may provide a good background conductivity image for a subsequent dynamic imaging. However, considering the ill-posedness of the EIT problem implying that the sensitivity matrix is severely ill-conditioned, any iterative approach is very difficult to be successful. For this reason, lately MREIT techniques have been suggested for the static conductivity imaging and will be discussed later. Static images from MREIT may improve the quality of dynamic images in EIT when both techniques are properly combined.

2.2.2. *Static imaging using data fitting method*

Most static image reconstruction algorithms in EIT can be viewed as a data fitting method as described in Fig. 5. We first construct a computer model of the subject satisfying (3). Since we do not know the true σ of the subject, we assume an initial conductivity distribution σ_k with $k = 0$ for the model. When we inject several currents into both the subject and the model, the corresponding measured and computed boundary voltages are different since $\sigma_k \neq \sigma$ in general. An image reconstruction algorithm iteratively updates σ_k until it minimizes the difference between measured and computed boundary voltages.

To illustrate this idea, we define the following minimization problem:

$$\hat{\sigma} = \arg\,\min_{\sigma_k}\left[\frac{1}{2}\sum_m \|\mathbf{f}^m - \mathbf{f}_c^m(\sigma_k)\|_2^2\right], \tag{12}$$

where "arg min" is an operator which gives an energy functional minimizer, $\mathbf{f}^m$ is a vector of measured boundary voltages with the mth injection current, and $\mathbf{f}_c^m(\sigma_k)$ is a corresponding vector of computed voltages. For the solution of (12), we may use the Newton-Raphson method.[33] Since the Jacobian matrix of this minimization problem is ill-conditioned, we often use some Tikhonov type regularization:

$$\hat{\sigma} = \arg\,\min_{\sigma^k}\left[\frac{1}{2}\sum_m \|\mathbf{f}^m - \mathbf{f}_c^m(\sigma_k)\|_2^2 + \lambda\eta(\sigma_k)\right], \tag{13}$$

where λ is a small regularization parameter and $\eta(\sigma)$ is a function measuring regularity of σ. The computation of the Jacobian matrix is basically the same as the

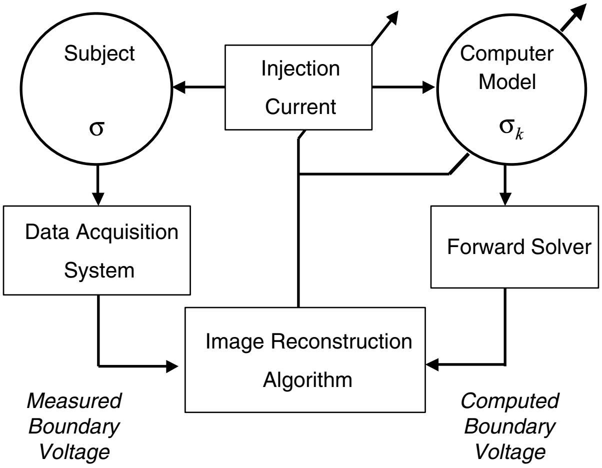

Subject
σ
Injection Current
Computer Model
σ_k
Data Acquisition System
Forward Solver
Image Reconstruction Algorithm
Measured Boundary Voltage
Computed Boundary Voltage

Fig. 5. Static image reconstruction as a data fitting method.

computation of the sensitivity matrix at σ_k instead of σ_0 as described in the previous section.

This kind of method was first introduced in EIT by Yorkey *et al.*[33] following numerous variations and improvements.[34–39] These include utilization of *a priori* information, various forms of regularization, adaptive mesh refinement, and so on. Even though this approach is widely adopted for static imaging by many researchers, it requires a large amount of computation time for producing static images even with a low spatial resolution and poor accuracy. Though it seems to be generally regarded as a classical technique in static impedance imaging, new ideas are still desired to get images with better quality.

2.3. *Ill-posedness in EIT*

EIT reveals technical difficulties in producing high-resolution conductivity images. This stems from the inherent insensitivity problem that perturbations of an internal conductivity distribution deliver relatively small changes of the measured boundary voltage data in a highly nonlinear way. In this section, we investigate this ill-posedness of EIT in detail.

The value of the voltage at a point inside the subject can be expressed as a weighted average of its neighboring voltages where the weights are determined by the conductivity distribution. In this weighted averaging way, the information on the conductivity distribution is conveyed to the boundary voltage as illustrated in Fig. 3(b). Therefore, the current-voltage data on the boundary is entangled with the global structure of the conductivity distribution in a highly nonlinear way.

Only for simplicity, we assume that the domain Ω of the subject is a square region in $\mathbb{R}^2$. We divide Ω uniformly into $K \times K$ sub-squares $\Omega_{i,j}$ with the center point (x_i, y_j) where $i, j = 0, \ldots, K-1$. Our goal is to determine $K \times K$ conductivity values under the assumption that the conductivity σ is constant on each sub-square $\Omega_{i,j}$, say $\sigma_{i,j}$. Let

$$\Sigma = \{\sigma | \sigma|_{\Omega_{i,j}} = \text{constant } \forall\, i, j = 0, \ldots, K-1\}.$$

For a given $\sigma \in \Sigma$, the solution u_σ of the direct problem in (3) can be approximated by a vector $\mathbf{u} = (u_0, u_1, \ldots, u_{K^2-1})$ such that each interior voltage $u_k, k = i + jK$ is determined by the weighted average (depending on the conductivity σ) of the four neighboring voltages. To be precise, the conductivity equation

$$\nabla \cdot (\sigma \nabla u(\mathbf{r})) = 0, \quad \mathbf{r} \in \Omega$$

can be written as the following discretized form of

$$u_k = \frac{1}{a_{k,k}}[a_{k,k_N} u_{k_N} + a_{k,k_S} u_{k_S} + a_{k,k_E} u_{k_E} + a_{k,k_W} u_{k_W}] \tag{14}$$

with

$$a_{k,k} = -\sum_d a_{k,k_d} \quad \text{and} \quad a_{k,k_d} = \frac{\sigma_k \sigma_{k_d}}{\sigma_k + \sigma_{k_d}} \quad \text{for } d = N, S, E, W, \tag{15}$$

where k_N, k_S, k_E, and k_W denote north, south, east and west neighboring points of the kth point, respectively. The discretized conductivity Eq. (14) with the Neumann boundary condition can be rewritten as the linear system $A_\sigma \mathbf{u} = \mathbf{g}$ where $\mathbf{g}$ is the injection current vector associated with a Neumann boundary data g. Let $\mathbf{f}$ denote a small-size sub-vector of $\mathbf{u}$ restricted to $\partial\Omega$, which corresponds to the boundary voltage on $\partial\Omega$ where sensing electrodes are attached. Then the inverse problem is to determine the conductivity σ, which is equivalent to obtain A_σ, from several measurements of current-to-voltage pairs $(\mathbf{g}^m, \mathbf{f}^m)$, $m = 1, \ldots, N$.

The fundamental shortcoming of EIT for providing high-resolution images is due to the fact that reconstructing A_σ from $(\mathbf{g}^m, \mathbf{f}^m)$, $m = 1, \ldots, N$ is exponentially difficult as the matrix size of A_σ increases. Let us explain this difficulty more precisely. According to (14), the value of the voltage at each pixel inside the region can be expressed as the weighted average of its neighboring voltages where weights are determined by the conductivity distribution. Therefore the measured data $\mathbf{f}$ is entangled in the global structure of the conductivity distribution in a highly nonlinear way and any internal conductivity value $\sigma_{i,j}$ influences little to boundary measurements if $\sigma_{i,j}$ is away from measuring points.

This ill-posed nature and severe nonlinearity make it very difficult for EIT to provide good static conductivity images. Increasing the matrix size for better spatial resolution makes the problem more ill-conditioned. Hence, having a larger number of electrodes beyond a certain limit may result in poorer images since the deteriorated ill-conditioning problem may take over the benefit of additional information from the increased number of electrodes. Furthermore, we must consider inevitable errors caused by modeling and measurement system. For these reasons, the static EIT imaging is still far from clinical applications though the dynamic imaging has been tried in numerous clinical application areas.[3,4,40]

2.4. *Measurement techniques in EIT and images*

There exist several EIT systems already developed and also being developed with different design concepts and technical details in their implementations. We may classify recent EIT systems into two types. The first may be characterized as one current source and multiple voltmeters. In this case, current is sequentially injected between a chosen pair of electrodes and there always exists only one active current source. The second type uses multiple current sources and multiple voltmeters. Here, we inject a pattern of current through multiple electrodes using multiple active current sources to maximize the distinguishability.[41] The sum of all currents from multiple current sources must be zero. In most current EIT systems belonging to both types, voltages from all electrodes are simultaneously measured using multiple voltmeters. Typical examples of the first and second type are Mk3.5 from Sheffield[42] and ACT-3 from RPI,[43] respectively. Boone *et al.*[3] summarized numerous techniques in the development of EIT system and it is still controversial to decide which type is superior. Figure 6 shows a typical example of EIT system.

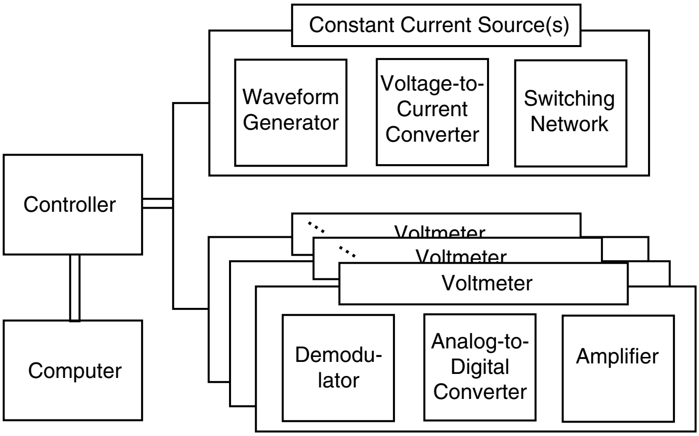


Fig. 6. Block diagram of a typical EIT system.

When the injection current $i_{ij}(t)$ between ith and jth electrode is

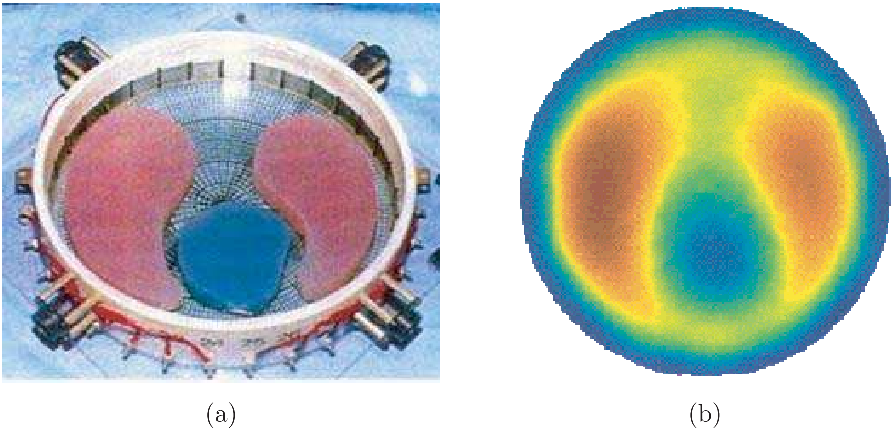
$$i_{ij}(t) = I \sin \omega t,$$

the measured voltage between kth and lth electrode is

$$u_{kl}(t) = IZ \sin(\omega t + \theta),$$

where $\mathbf{Z} = Ze^{i\theta}$ is the trans-impedance. Extracting the trans-impedance from the measured voltage requires the phase-sensitive demodulation. Important factors affecting the performance of an EIT system include the amplitude stability and source impedance of the current source. For the voltmeter, dynamic range, common-mode rejection ratio, and noise level are of primary concerns. When the EIT system include many switches and cables, special care must be given to minimize stray capacitances and cross-talks. The capability of providing the information on the boundary shape and size of the subject and electrode positions would be necessary especially for static imaging.

At the frequency of tens of kHz, the range of trans-impedance is from a few mΩ to tens of Ω depending on the subject, number of electrodes, and their configuration. Assuming the magnitude of the injection current is 1 mA, for example, the voltage is in the range of a few μV to tens of mV. Allowing the noise of 1% of the smallest voltage, we are dealing with a noise level of much less than 1 μV requiring the state-of-the-art electronic instrumentation technology. Modern EIT system usually acquires a complete set of data within 10 ms enabling the temporal resolution of more than 16 frames/s using a fast dynamic image reconstruction algorithm. This means that it is possible to produce real-time images of most physiological time-varying events. In designing a new EIT system, we should consider the factors such as the complexity of the system, technology available at present, and future direction of EIT system development including miniaturization and wireless interconnection.



(a) (b)

Fig. 7. (a) Two dimensional conductivity phantom and (b) reconstructed static conductivity image using ACT-3 EIT system from the RPI group. See text for details and also http://www.rpi.edu/newelj/eit.html.

Figure 7(a) shows a two-dimensional conductivity phantom with 32 electrodes constructed by the RPI group.[5] The phantom is circular with 30 cm diameter and saline-filled. On its inside surface, there are 32 stainless steel electrodes with the size of $2.54 \times 2.54\,\text{cm}^2$. Two lung-shaped structures and a heart-shaped structure made of agar are immersed in saline. Figure 7(b) is a static conductivity image of the phantom reconstructed from a set of boundary data.

Metherall described numerous *in-vivo* three-dimensional EIT images using Sheffield Mk3.5 EIT system.[4] Figure 8 shows one set of those images showing the impedance changes during ventilation. Figure 8(a) are 8 images at 8 different cross-sectional planes of a human thorax with both lungs at residual volume. The images are difference images with respect to a set of reference data at total lung capacity. Figure 8(b) and (c) are corresponding images at functional residual capacity and peak tidal volume, respectively.

Tidswell *et al.* applied the EIT technique to image functional activity in the brain.[44] Figure 9 shows a time series of EIT images during visual stimulation. The goggles were used to produce a bright flash which stimulated the infant's vision. Each column represents averaged images over 4 seconds at six slices of the baby's head. From the EIT images, we can see that visual stimulation produced impedance changes at the front and the back of the head. The area on the back of the head corresponds to the position of the visual cortex and the impedance decrease there is probably due to the increased blood volume in response to the visual stimulus. The larger change at the front of the head was interpreted as artifacts due to blinking or muscle movement during the bright visual stimulus. More information on EIT techniques and numerous applications can be found at http://www.eit.org.uk/index.html.

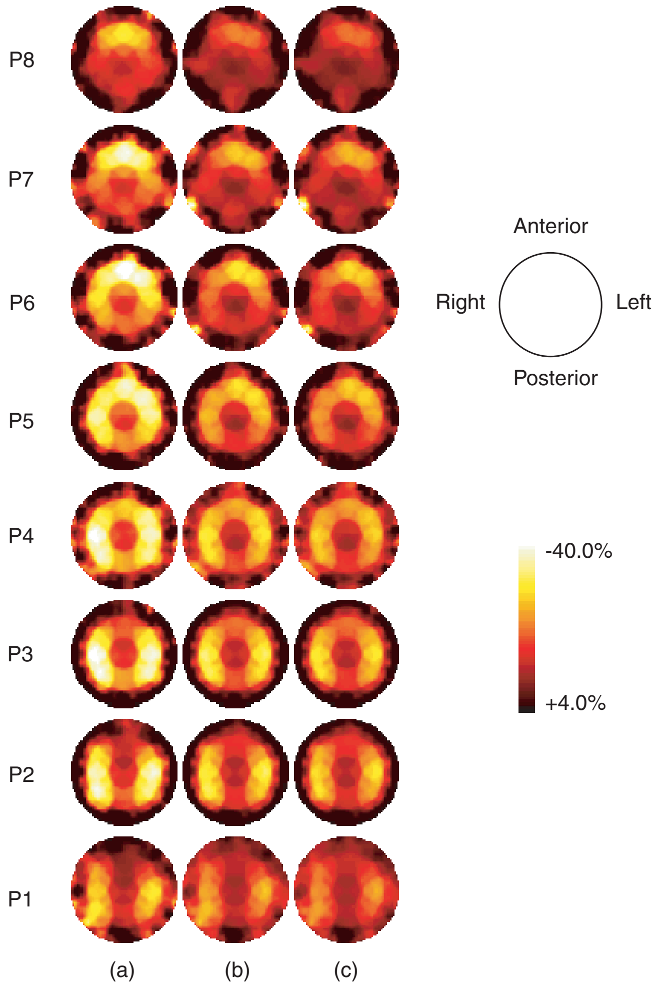


Fig. 8. Three-dimensional *in-vivo* EIT images of a human thorax using Sheffield Mk3.5 EIT system. See text for details and also http://www.shef.ac.uk/uni/academic/I-M/mpce/rsch/funimg.html.

3. Magnetic Resonance Electrical Impedance Tomography (MREIT)

Injected current in an electrically conducting subject produces a magnetic field as well as an electric field. In EIT, the information on the electric field in a form of boundary current-voltage data is used to reconstruct conductivity images. Noting that the magnetic field inside the subject can also be measured by a non-contact method using an MRI scanner, we may transform the ill-posed problem into a well-posed one utilizing this additional information. This new idea initiated the research area called Magnetic Resonance Electrical Impedance Tomography (MREIT).

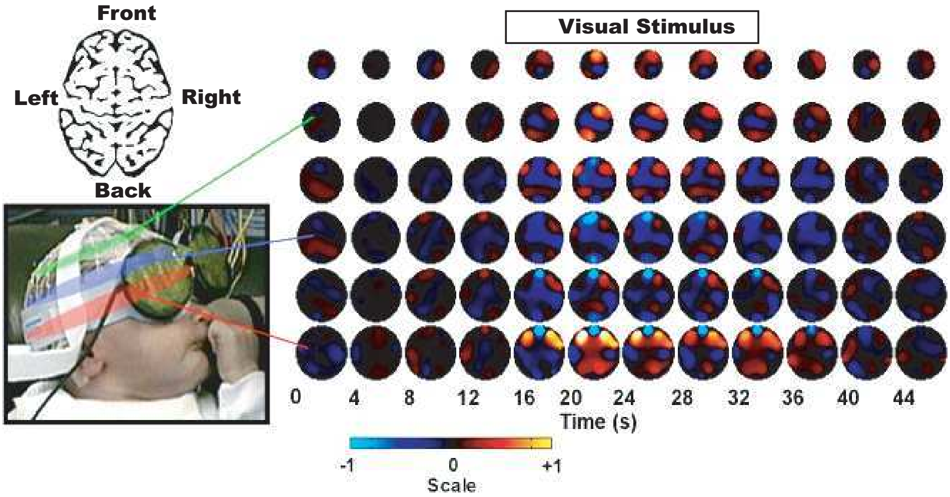


Fig. 9. Time series EIT images of a neonatal brain during visual stimulation from the UCL group. See text for details and also http://madeira.physiol.ucl.ac.uk/midx-group/.

Since late 1980s, measurements of the internal magnetic flux density due to an injection current have been studied in Magnetic Resonance Current Density Imaging (MRCDI) to visualize the internal current density distribution.[45–47] This requires an MRI scanner as a tool to capture internal magnetic flux density images. Once we obtain the magnetic flux density $\mathbf{B} = (B_x, B_y, B_z)$ due to an injection current I, we can produce an image of the corresponding internal current density distribution $\mathbf{J}$ from the Ampere's law $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$ where μ_0 is the magnetic permeability of the free space.

The basic concept of MREIT was proposed by combining EIT and MRCDI techniques.[7–12] In MREIT, we measure the induced magnetic flux density $\mathbf{B}$ inside a subject due to an injection current I using an MRI scanner. Then, we may compute the internal current density $\mathbf{J}$ as is done in MRCDI. From $\mathbf{B}$ and/or $\mathbf{J}$, we can perceive the internal current pathways due to the conductivity distribution to be imaged.

However, if we try to utilize $\mathbf{J} = \nabla \times \mathbf{B}/\mu_0$ by measuring all three components of $\mathbf{B}$, there occurs a serious technical problem. Since any currently available MRI scanner measures only one component of $\mathbf{B}$ that is parallel to the direction of the main magnetic field of the MRI scanner, measuring all three orthogonal components of $\mathbf{B} = (B_x, B_y, B_z)$ requires subject rotations. These subject rotations are impractical and also cause other problems such as misalignments of pixels. Therefore, it is highly desirable to reconstruct conductivity images from only B_z instead of $\mathbf{B}$ where z is the direction of the main magnetic field of the MRI scanner. For this reason, most recent MREIT techniques focus on analyzing the information embedded in the measured B_z data to extract any constructive relations between B_z and the conductivity distribution to be imaged.

This section addresses the image reconstruction problem in MREIT as a well-posed inverse problem taking advantage of the information on internal magnetic flux density distributions. Assuming that the magnetic flux density $\mathbf{B} = (B_x, B_y, B_z)$ or only B_z is available, a mathematical formulation for the MREIT problem is presented to explain the fundamental concept. Following the description on image reconstruction algorithms and measurement methods with some of experimental results, practical limitations in terms of the measurement noise and the amount of injection current are discussed. At the end of this section, future research directions are summarized.

3.1. *Mathematical formulation of MREIT*

3.1.1. *Forward problem: Magnetic flux density due to an injection current for a given conductivity*

As shown in Fig. 10, we assume an electrically conducting domain Ω with its boundary $\partial\Omega$ and a conductivity distribution σ. We choose a pair of electrodes attached on $\partial\Omega$, for example, $\mathcal{E}_i$ and $\mathcal{E}_j$ to inject current I. Lead wires carrying the injection current I are denoted as $\mathcal{L}_i$ and $\mathcal{L}_j$. Then, the voltage u in Ω satisfies the Neumann boundary value problem in (3). Knowing the voltage distribution u, the current density $\mathbf{J}$ is given by (4). Denoting the Neumann boundary data due to the injection current as g, those equations are rewritten as

$$\begin{cases} \nabla \cdot [\sigma(\mathbf{r})\nabla u(\mathbf{r})] = 0 & \text{in } \Omega \\ -\sigma\nabla u \cdot \mathbf{n} = g & \text{on } \partial\Omega \end{cases}, \tag{16}$$

and

$$\mathbf{J}(\mathbf{r}) = -\sigma(\mathbf{r})\nabla u(\mathbf{r}) = \sigma(\mathbf{r})\mathbf{E}(\mathbf{r}) \quad \text{in } \Omega. \tag{17}$$

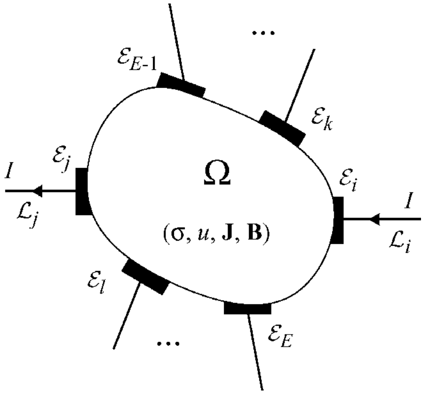


Fig. 10. Electrically conducting subject Ω with a conductivity σ and voltage u distribution. Surface electrodes $\mathcal{E}_j, j = 1, \ldots, E$ are attached on the boundary $\partial\Omega$. Here, we assume that current is injected between the diagonal pair of electrodes $\mathcal{E}_i$ and $\mathcal{E}_j$.

We now consider the magnetic field produced by the injection current. The induced magnetic flux density $\mathbf{B}$ in Ω can be decomposed into three parts as

$$\mathbf{B}(\mathbf{r}) = \mathbf{B}_{\Omega}(\mathbf{r}) + \mathbf{B}_{\mathcal{E}}(\mathbf{r}) + \mathbf{B}_{\mathcal{L}}(\mathbf{r}) \quad \text{in } \Omega, \tag{18}$$

where $\mathbf{B}_{\Omega}$, $\mathbf{B}_{\mathcal{E}}$ and $\mathbf{B}_{\mathcal{L}}$ are magnetic flux densities due to $\mathbf{J}$ in Ω, $\mathbf{J}$ in $\mathcal{E} = \mathcal{E}_i \cup \mathcal{E}_j$ and I in $\mathcal{L} = \mathcal{L}_i \cup \mathcal{L}_j$, respectively. From the Biot-Savart law,

$$\mathbf{B}_{\Omega}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_{\Omega} \mathbf{J}(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3}\, dv', \tag{19}$$

$$\mathbf{B}_{\mathcal{E}}(\mathbf{r}) = \frac{\mu_0}{4\pi} \int_{\mathcal{E}} \mathbf{J}(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3}\, dv', \tag{20}$$

and

$$\mathbf{B}_{\mathcal{L}}(\mathbf{r}) = \frac{\mu_0 I}{4\pi} \int_{\mathcal{L}} \mathbf{a}(\mathbf{r}') \times \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3}\, dl', \tag{21}$$

where $\mathbf{a}(\mathbf{r}')$ is the unit vector in the direction of the current flow at $\mathbf{r}' \in \mathcal{L}$. From the Ampere's law, the current density $\mathbf{J}$ is also given by

$$\mathbf{J}(\mathbf{r}) = \frac{1}{\mu_0} \nabla \times \mathbf{B}(\mathbf{r}) \quad \text{in } \Omega. \tag{22}$$

We must have

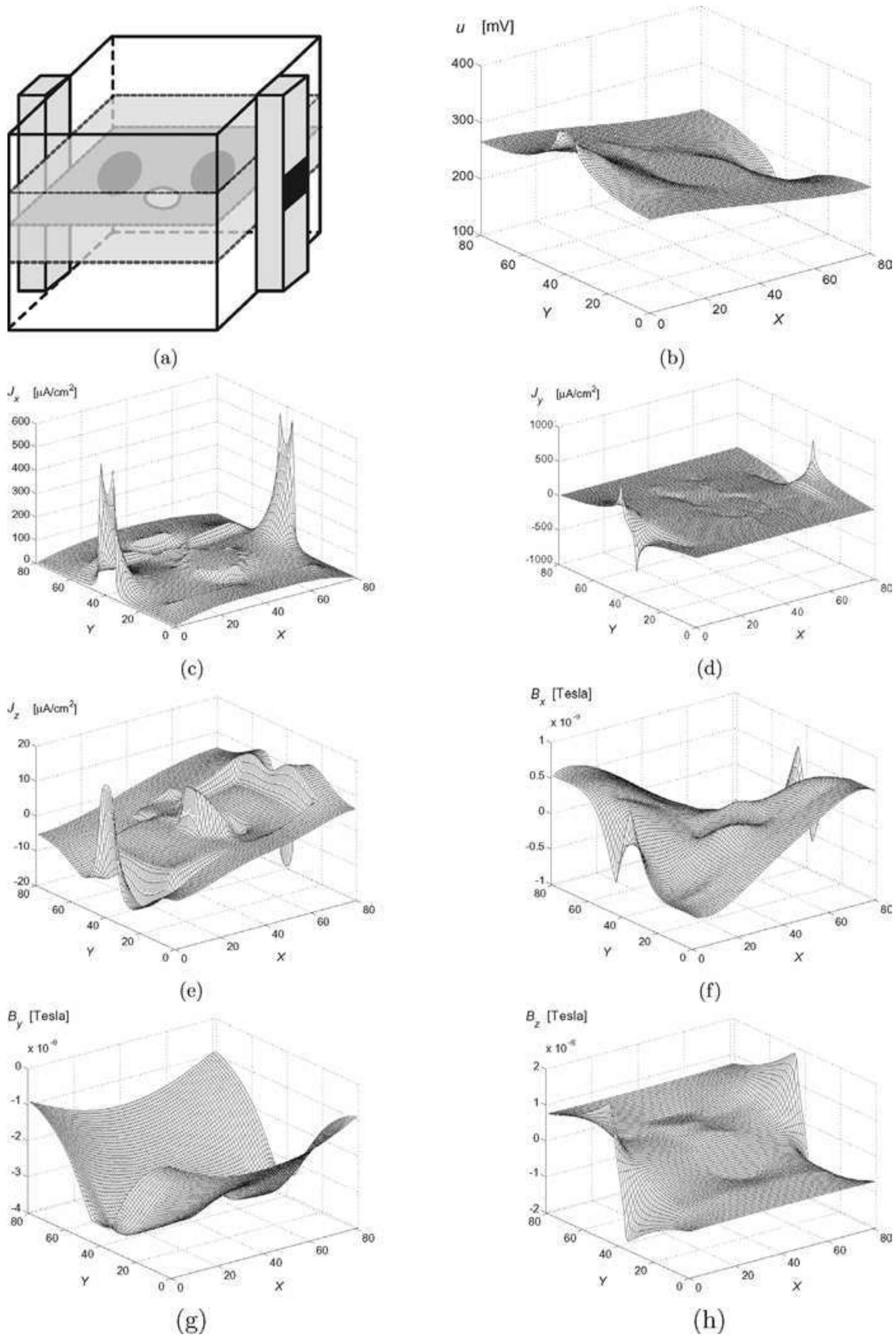
$$\frac{1}{\mu_0} \nabla \times \mathbf{B}(\mathbf{r}) = -\sigma(\mathbf{r}) \nabla u(\mathbf{r}) \quad \text{and} \quad \nabla \cdot \mathbf{J}(\mathbf{r}) = 0 \quad \text{in } \Omega. \tag{23}$$

Numerical techniques in solving (16)–(22) are described by Lee *et al.* with several examples.[18] Figure 11 shows a typical example.

3.1.2. *Inverse problem: Conductivity from Neumann-to-B_z (NtB$_z$) map*

Now, the problem of interest is to reconstruct an image of σ in Ω from measured magnetic flux density and boundary voltage. For the uniqueness of a reconstructed conductivity image, it has been shown that we need to inject at least two currents using more than three electrodes and measure the corresponding magnetic flux densities.[48,49] In addition, at least one boundary voltage measurement is needed to recover the absolute values of the conductivity distribution. Since the measurement of $\mathbf{B} = (B_x, B_y, B_z)$ requires the impractical subject rotations, in this section, we assume that we measure only B_z without rotating the subject.

The description of the inverse problem in MREIT is based on the following setup. We place a subject Ω inside an MRI scanner and attach surface electrodes. The conductivity distribution σ of the subject is assumed to be isotropic with $0 < \sigma < \infty$. When the number of electrodes is E, we can sequentially select one of $N \leq \frac{E(E-1)}{2}$ different pairs of electrodes to inject currents into the subject. Let the injection current between the jth pair of electrodes be I_j for $j = 1, \ldots, N$ with $N \geq 2$. The current I_j produces a current density $\mathbf{J}_j = (J_x^j, J_y^j, J_z^j)$ inside the subject. The presence of the internal current density $\mathbf{J}_j$ and the current I_j



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 11. Typical numerical results of a simple thorax model of $50 \times 50 \times 50\,\mathrm{mm}^3$. Current of $1\,\mathrm{mA}$ is injected between the pair of electrodes on the surface of the model in (a). Computed results of (b) u, (c) J_x, (d) J_y, (e) J_z, (f) B_x, (g) B_y, and (h) B_z.

in external lead wires generates a magnetic flux density $\mathbf{B}_j = (B_x^j, B_y^j, B_z^j)$ and $\mathbf{J}_j = \nabla \times \mathbf{B}_j/\mu_0$ holds inside the electrically conducting subject. We now assume that we have measured B_z^j for $j = 1, \ldots, N$.

Let u_j be the voltage due to the injection current I_j for $j = 1, \ldots, N$. Since σ is approximately independent of injection currents, each u_j is a solution of the following Neumann boundary value problem:

$$\begin{cases} \nabla \cdot (\sigma(\mathbf{r})\nabla u_j(\mathbf{r})) = 0 \quad \text{in } \Omega \\ -\sigma\nabla u_j \cdot \ \mathbf{n} = g_j \quad \text{on } \partial\Omega \end{cases}, \tag{24}$$

where g_j is the normal component of current density on the boundary of the subject for the injection current I_j. If σ, I_j, and electrode configuration are given, we can solve (24) for u_j using a numerical method such as FEM.[17,18]

Now, we introduce a map relating B_z^j with the Neumann data g^j:

$$\Lambda_\sigma[g^j](\mathbf{r}) = B_z^j(\mathbf{r}), \quad \mathbf{r} \in \Omega.$$

We will call this map Λ_σ by the Neumann-to-B_z map (NtB_z-map). According to the Biot–Savart law with a given g^j, $\Lambda_\sigma[g^j]$ is expressed as

$$\Lambda_\sigma[g^j](\mathbf{r}) = \frac{\mu_0}{4\pi}\int_\Omega \frac{\sigma(\mathbf{r}')[(x-x')\frac{\partial u^j}{\partial y}(\mathbf{r}') - (y-y')\frac{\partial u^j}{\partial x}(\mathbf{r}')]}{|\mathbf{r}-\mathbf{r}'|^3}\, d\mathbf{r}', \tag{25}$$

where u^j is the solution of (24). The inverse problem in MREIT is to reconstruct σ from several NtB_z data, $\Lambda_\sigma[g^j]$, $j = 1, \ldots, N$. In order for MREIT to be more practical, N should not be a large number.

3.2. *MREIT image reconstruction techniques*

When $\mathbf{B} = (B_x, B_y, B_z)$ is available, we may use $\mathbf{J}$ from (22) to reconstruct conductivity images using image reconstruction algorithms such as the J-substitution algorithm,[10,50,51] current constrained voltage scaled reconstruction (CCVSR) algorithm,[52] and equipotential line methods.[49,53,54] However, since these methods require the impractical subject rotation procedure, we describe algorithms utilizing only B_z data.

3.2.1. *Harmonic B_z algorithm*

Based on the relation of $\nabla^2\mathbf{B} = -\mu_0\nabla u \times \nabla\sigma$ observed by Scott *et al.*,[46] Seo *et al.* derived the following expression that holds for each position in Ω.[55]

$$\frac{1}{\mu_0}\nabla^2 B_z^j = \left(\frac{\partial\sigma}{\partial x}, \frac{\partial\sigma}{\partial y}\right) \cdot \left(\frac{\partial u_j}{\partial y}, -\frac{\partial u_j}{\partial x}\right) = \frac{\partial\sigma}{\partial x}\frac{\partial u_j}{\partial y} - \frac{\partial\sigma}{\partial y}\frac{\partial u_j}{\partial x}, \quad j = 1, \ldots, N. \tag{26}$$

Note that the magnetic flux density due to the injection current I_j along external lead wires becomes irrelevant by using $\nabla^2 B_z^j$. Using a matrix form, (26) becomes

$$\mathbf{Us} = \mathbf{b}, \tag{27}$$

where $\mathbf{U} = \begin{bmatrix} \frac{\partial u_1}{\partial y} & -\frac{\partial u_1}{\partial x} \\ \vdots & \vdots \\ \frac{\partial u_N}{\partial y} & -\frac{\partial u_N}{\partial x} \end{bmatrix}$, $\mathbf{s} = \begin{bmatrix} \frac{\partial \sigma}{\partial x} \\ \frac{\partial \sigma}{\partial y} \end{bmatrix}$, and $\mathbf{b} = \frac{1}{\mu_0} \begin{bmatrix} \nabla^2 B_z^1 \\ \vdots \\ \nabla^2 B_z^N \end{bmatrix}$.

For the case where two injection currents are used ($N = 2$), we can obtain $\mathbf{s}$ provided that two voltages u_1 and u_2 corresponding to two injection currents I_1 and I_2 satisfy

$$-\frac{\partial u_1}{\partial y}\frac{\partial u_2}{\partial x} + \frac{\partial u_1}{\partial x}\frac{\partial u_2}{\partial y} \neq 0. \tag{28}$$

We can argue that (28) holds for almost all positions within the subject since two current densities $\mathbf{J}_1$ and $\mathbf{J}_2$ due to appropriately chosen I_1 and I_2 will not have the same direction.[48,49,56]

We use N injection currents to better handle measurement noise in B_z and improve the condition number of $\mathbf{U}^T\mathbf{U}$ where $\mathbf{U}^T$ is the transpose of $\mathbf{U}$. Using the weighted regularized least square method suggested by Oh *et al.*,[11] we can get $\mathbf{s}$ as

$$\mathbf{s} = (\tilde{\mathbf{U}}^T\tilde{\mathbf{U}} + \lambda\mathbf{I})^{-1}\ \tilde{\mathbf{U}}^T\tilde{\mathbf{b}}, \tag{29}$$

where λ is a positive regularization parameter, $\mathbf{I}$ is the 2×2 identity matrix, $\tilde{\mathbf{U}} = \mathbf{W}\mathbf{U}$, $\tilde{\mathbf{b}} = \mathbf{W}\mathbf{b}$ and $\mathbf{W} = \mathrm{diag}(w_1, \ldots, w_N)$ is an $N \times N$ diagonal weight matrix. Oh *et al.* discussed different ways of determining the value of λ and the weight w_j. Computing (29) for each position or pixel, we obtain a distribution of $\mathbf{s} = \left[\frac{\partial \sigma}{\partial x}\ \frac{\partial \sigma}{\partial y}\right]^T$ inside the subject.

We now tentatively assume that the imaging slice $\mathcal{S}$ is lying in the plane $\{z = 0\}$ and the conductivity value at a fixed position $\mathbf{r}_0 = (x_0, y_0, 0)$ on its boundary $\partial\mathcal{S}$ is 1. For a moment, we denote $\mathbf{r} = (x, y)$, $\mathbf{r}' = (x', y')$ and $\sigma(x, y, 0) = \sigma(\mathbf{r})$. In order to compute σ from $\nabla\sigma = \left(\frac{\partial \sigma}{\partial x}, \frac{\partial \sigma}{\partial y}\right)$, Oh *et al.* suggested a layer potential technique in two dimension.[11] Then,

$$\begin{aligned} \sigma(\mathbf{r}) &= \int_{\mathcal{S}} \nabla^2\Phi(\mathbf{r} - \mathbf{r}')\sigma(\mathbf{r}')d\mathbf{r}' \\ &= -\int_{\mathcal{S}} \nabla_{\mathbf{r}'}\Phi(\mathbf{r} - \mathbf{r}') \cdot \nabla\sigma(\mathbf{r}')d\mathbf{r}' + \int_{\partial\mathcal{S}} \mathbf{n}_{\mathbf{r}'} \cdot \nabla_{\mathbf{r}'}\Phi(\mathbf{r} - \mathbf{r}')\ \sigma(\mathbf{r}')dl_{\mathbf{r}'}, \end{aligned} \tag{30}$$

where $\Phi(\mathbf{r} - \mathbf{r}') = \frac{1}{2\pi}\log|\mathbf{r} - \mathbf{r}'|$ and $\nabla_{\mathbf{r}'}\Phi\left(\mathbf{r} - \mathbf{r}'\right) = -\frac{1}{2\pi}\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^2}$. It is well known[57] that for $\mathbf{r} \in \partial\mathcal{S}$

$$\lim_{t \to +0} \int_{\partial\mathcal{S}} \mathbf{n}_{\mathbf{r}'} \cdot \nabla_{\mathbf{r}'}\Phi(\mathbf{r} - t\mathbf{n}_{\mathbf{r}} - \mathbf{r}')\sigma(\mathbf{r}')dl_{\mathbf{r}'} = \frac{\sigma(\mathbf{r})}{2} + \int_{\partial\mathcal{S}} \mathbf{n}_{\mathbf{r}'} \cdot \nabla_{\mathbf{r}'}\Phi(\mathbf{r} - \mathbf{r}')\sigma(\mathbf{r}')dl_{\mathbf{r}'}.$$

Hence, as $\mathbf{r} \in \mathcal{S}$ approaches the boundary $\partial\mathcal{S}$ in (30), we have

$$\frac{\sigma_{\partial\mathcal{S}}(\mathbf{r})}{2} + \frac{1}{2\pi}\int_{\partial\mathcal{S}} \frac{(\mathbf{r} - \mathbf{r}') \cdot \mathbf{n}_{\mathbf{r}'}}{|\mathbf{r} - \mathbf{r}'|^2}\sigma_{\partial\mathcal{S}}(\mathbf{r}')dl'_{\mathbf{r}} = \frac{1}{2\pi}\int_{\mathcal{S}} \frac{(\mathbf{r} - \mathbf{r}') \cdot \nabla\sigma(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^2}d\mathbf{r}', \tag{31}$$

where $\sigma_{\partial\mathcal{S}}$ denotes the conductivity restricted at the boundary $\partial\mathcal{S}$. It is also well known that the solvability of the integral equation (31) for $\sigma_{\partial\mathcal{S}}$ is guaranteed for a given right side of (31).[57] Since $\nabla\sigma$ is known in $\mathcal{S}$, so does the right side of (31).

This enables us to obtain the value $\sigma_{\partial\mathcal{S}}$ by solving the integral equation (31). Now, we can compute the conductivity σ in $\mathcal{S}$ by substituting the boundary conductivity $\sigma_{\partial\mathcal{S}}$ into (30) as

$$\sigma(\mathbf{r}) = -\int_{\mathcal{S}} \nabla_{\mathbf{r}'}\Phi(\mathbf{r}-\mathbf{r}')\cdot\nabla\sigma(\mathbf{r}')d\mathbf{r}' + \int_{\partial\mathcal{S}} \mathbf{n}_{\mathbf{r}'}\cdot\nabla_{\mathbf{r}'}\Phi(\mathbf{r}-\mathbf{r}')\,\sigma_{\partial\mathcal{S}}(\mathbf{r}')dl_{\mathbf{r}'}. \tag{32}$$

The process of solving (29) for each pixel and (31) and (32) for each imaging slice can be repeated for all imaging slices of interest within the subject as long as the measured data B_z are available for the slices. Furthermore, we can apply the method described in this section to any imaging slice of axial, coronal and sagittal direction.

As expressed in (24), voltages u_j depend on the unknown true conductivity σ and, therefore, we do not know the matrix $\mathbf{U}$ corresponding to σ. This requires us to use the iterative algorithm described below. For $j = 1, \ldots, N$, we sequentially inject current I_j through a chosen pair of electrodes and measure the z-component of the induced magnetic flux density B_z^j. For each injection current I_j, we also measure boundary voltages $u_j\mid_{\partial\mathcal{S}}$ on electrodes not injecting the current I_j. Then, the $\nabla^2 B_z$ algorithm is as follows.

(i) Let $n = 0$ and assume an initial conductivity distribution σ_0.

(ii) Compute u_j^n by solving the following Neumann boundary value problems for $j = 1, \ldots, N$:

$$\begin{cases} \nabla\cdot\left(\sigma_n\nabla u_j^n\right) = 0 & \text{in } \Omega \\ -\sigma_n\nabla u_j^n\cdot\mathbf{n} = g_j & \text{on } \partial\Omega. \end{cases} \tag{33}$$

(iii) Compute σ_{n+1} using (29), (31) and (32). Scale σ_{n+1} using the measured boundary voltages $u_j\mid_{\partial\mathcal{S}}$ and the corresponding computed ones $u_j^n\mid_{\partial\mathcal{S}}$.

(iv) If $\|\sigma_{n+1}-\sigma_n\|_2 < \epsilon$, go to Step 5. Here, ϵ is a given tolerance. Otherwise, set $n \leftarrow (n+1)$ and go to Step 2.

(v) If needed, compute current density images as $\mathbf{J}_j \leftarrow -\sigma_{n+1}\nabla u_j^M$ where u_j^M is a solution of the boundary value problem in (24) with σ_{n+1} replacing σ.

3.2.2. *Gradient B_z decomposition algorithm*

After the introduction of the harmonic B_z algorithm, there has been an effort to improve its performance especially in terms of the way we numerically differentiate the measured noisy B_z data. Based on a novel analysis utilizing the Helmhortz decomposition, Park *et al.* suggested the gradient B_z decomposition algorithm.[58]

In order to explain the algorithm, let us assume $\Omega = D\times[-\delta,\delta] = \{\mathbf{r} = (x,y,z)|(x,y)\in D,\ -\delta < z < \delta\}$ is an electrically conducting subject where D is a two dimensional smooth simply connected domain. Let u be the solution of the Neumann boundary value problem (16) or (24) with the Neumann data g. We parameterize ∂D as $\partial D := \{(x(t),y(t))\colon 0\le t\le 1\}$ and define $\tilde{g}(x(t),y(t),z) := \int_0^t g((x(t),y(t),z))\sqrt{|x'(t)|^2+|y'(t)|^2}\,dt$ for $(x,y,z)\in\partial D\times(-\delta,\delta)$.

The gradient B_z decomposition algorithm is based on the following key identity:

$$\sigma = \frac{\left|-\left(\frac{\partial H}{\partial y} + \Lambda_x[u]\right)\frac{\partial u}{\partial x} + \left(\frac{\partial H}{\partial x} + \Lambda_y[u]\right)\frac{\partial u}{\partial y}\right|}{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2} \quad \text{in } \Omega, \tag{34}$$

where

$$\Lambda_x[u] := \frac{\partial \psi}{\partial y} - \frac{\partial W_z}{\partial x} + \frac{\partial W_x}{\partial z} \quad \text{and} \quad \Lambda_y[u] := \frac{\partial \psi}{\partial x} + \frac{\partial W_z}{\partial y} - \frac{\partial W_y}{\partial z} \quad \text{in } \Omega,$$

and

$$H = \phi + \frac{1}{\mu_0} B_z, \quad W(\mathbf{r}) := \int_{\Omega_\delta} \frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} \frac{\partial(\sigma \nabla u(\mathbf{r}'))}{\partial z} \, d\mathbf{r}'.$$

Here, ϕ and ψ are solutions of the following equations:

$$\begin{cases} \nabla^2 \phi = 0 & \text{in } \Omega \\ \phi = \tilde{g} - \frac{1}{\mu_0} B_z & \text{on } \partial\Omega_{side} \\ \frac{\partial \phi}{\partial z} = -\frac{1}{\mu_0}\frac{\partial B_z}{\partial z} & \text{on } \partial\Omega_{tb} \end{cases} \quad \text{and} \quad \begin{cases} \nabla^2 \psi = 0 & \text{in } \Omega \\ \nabla \psi \cdot \tau = \nabla \times W \cdot \tau & \text{on } \partial\Omega_{side}\,, \\ \frac{\partial \psi}{\partial z} = -\nabla \times W \cdot \epsilon_z & \text{on } \partial\Omega_{tb} \end{cases}$$

where $\epsilon_z = (0,0,1)$, $\partial\Omega_{side} = \partial D \times (-\delta, \delta)$, Ω_{tb} is the top and bottom surfaces of Ω, and $\tau := (-\nu_y, \nu_x, 0)$ is the tangent vector on the lateral boundary $\partial D \times (-\delta, \delta)$.

Since the term u in (34) is a highly nonlinear function of σ, the identity (34) can be viewed as an implicit reconstruction formula for σ. It should be noticed that we cannot identify σ with a single g using (34). Hence, we may use an iterative reconstruction scheme with multiple Neumann data g_j, $j = 1, \ldots, N$ to find σ. Let u_j^m be the solution of (24) with $\sigma = \sigma_m$ and g_j. Then, the reconstructed σ is the limit of a sequence σ^m that is obtained by the following formula:

$$\sigma_{m+1} = \frac{\sum_{i=1}^N \left|-\left(\frac{\partial H_i}{\partial y} + \Lambda_x[u_i^m]\right)\frac{\partial u_i^m}{\partial x} + \left(\frac{\partial H_i}{\partial x} + \Lambda_y[u_i^m]\right)\frac{\partial u_i^m}{\partial y}\right|}{\sum_{i=1}^N \left[\left(\frac{\partial u_i^m}{\partial x}\right)^2 + \left(\frac{\partial u_i^m}{\partial y}\right)^2\right]}.$$

3.2.3. *Other algorithms*

There are other algorithms such as the variational gradient B_z algorithm.[59] Lately, a new algorithm that can handle anisotropic conductivity distributions has also been suggested.[60] All of these algorithms are quite successful in numerical simulations and experimental studies using a relatively large amount of injection current. Therefore, the algorithm development in MREIT should be focused on how to handle random and systematic noise in the measured B_z data. It should include efficient denoising techniques utilizing the fundamental properties of the induced magnetic flux density. Since conventional MR images providing excellent structural information are always available in MREIT, we may incorporate this *a priori* information in MREIT conductivity image reconstructions.

3.3. *Measurement techniques in MREIT and images*

Let z be the coordinate that is parallel to the direction of the main magnetic field $\mathbf{B}_0$ of an MRI scanner. Using a constant current source and a pair of surface electrodes, we sequentially inject two current pulses of $I^{\pm}$ and $I^{\mp}$ synchronized with the standard spin echo pulse sequence shown in Fig. 12. The application of the injection current during MR imaging induces a magnetic flux density $\mathbf{B} = (B_x, B_y, B_z)$. Since the magnetic flux density $\mathbf{B}$ produces inhomogeneity of the main magnetic field changing $\mathbf{B}_0$ to $(\mathbf{B}_0 + \mathbf{B})$, it causes phase changes that are proportional to the z-component of $\mathbf{B}$, that is B_z. Then, the corresponding MRI signals are

$$S^{I^{\pm}}(m,n) = \iint_{-\infty}^{\infty} M(x,y)e^{j\delta(x,y)}e^{j\gamma B_z(x,y)T_c}e^{j(xm\Delta k_x + yn\Delta k_y)}dx\,dy, \tag{35}$$

and

$$S^{I^{\mp}}(m,n) = \iint_{-\infty}^{\infty} M(x,y)e^{j\delta(x,y)}e^{-j\gamma B_z(x,y)T_c}e^{j(xm\Delta k_x + yn\Delta k_y)}dx\,dy. \tag{36}$$

Here, M is the transverse magnetization, δ is any systematic phase error, $\gamma = 26.75 \times 10^7$ rad/Tesla $\cdot$ s is the gyromagnetic ratio of the hydrogen, T_c is the duration of current pulses.

Two-dimensional discrete Fourier transformations of $S^{I^{\pm}}(m,n)$ and $S^{I^{\mp}}(m,n)$ result in two complex images of $\mathcal{M}_c^{\pm}(x,y)$ and $\mathcal{M}_c^{\mp}(x,y)$, respectively. Dividing the two complex images, we get

$$\mathrm{Arg}\left(\frac{\mathcal{M}_c^{\pm}(x,y)}{\mathcal{M}_c^{\mp}(x,y)}\right) = \mathrm{Arg}\big(e^{j2\gamma B_z(x,y)T_c}\big) = \tilde{\Phi}_z(x,y),$$

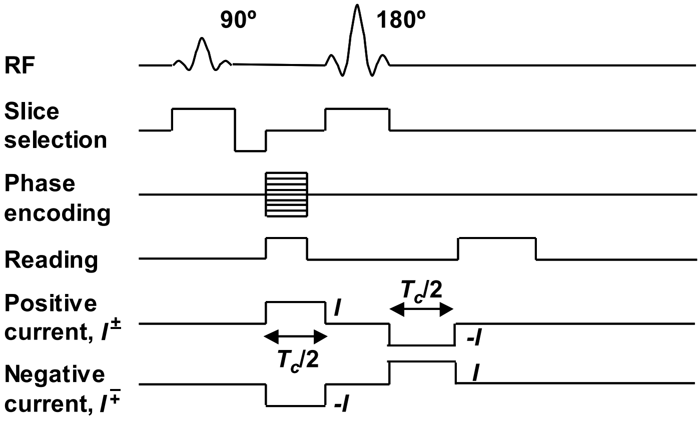


Fig. 12. An example of spin echo pulse sequence for MREIT.

where $\mathrm{Arg}(\omega)$ is the principal value of the argument of the complex number ω. Since $\tilde{\Phi}_z$ is wrapped in $-\pi < \tilde{\Phi}_z \leq \pi$, we must unwrap $\tilde{\Phi}_z$ to obtain Φ_z.[61] Finally, we get

$$B_z(x,y) = \frac{1}{2\gamma T_c}\Phi_z(x,y). \tag{37}$$

Scott *et al.* analyzed the Gaussian random noise in measured B_z.[47] Denoting the standard deviation of the Gaussian random noise as s_B, it can be estimated as

$$s_B = \frac{1}{2\gamma T_c \Upsilon_M}, \tag{38}$$

where Υ_M is the Signal-to-Noise Ratio (SNR) of the corresponding MR magnitude image $M(x,y)$ in (35) or (36). The noise standard deviation is inversely proportional to the size of each pixel since Υ_M in (38) is proportional to the size. With $T_c = 50\,\mathrm{ms}$, we obtain $s_B = 1.43 \times 10^{-9}$ and 5.68×10^{-9} Tesla when $\Upsilon_M = 50$ and 25, respectively.

In MREIT, we are interested in the SNR of measured magnetic flux density images and it is mainly determined by the noise standard deviation s_B in (38), amount of injection current, size of the subject, and electrode configuration. To reduce s_B, we must increase the SNR of the MR magnitude image, Υ_M in (38). This can be done by increasing the voxel size, number of averaging, strength of the main magnetic field, and so on. In doing so, it is inevitable to sacrifice the spatial and/or temporal resolution to some extent.

Regarding the amount of injection currents, it should be lower than the level that can stimulate muscle or nerve tissues. Although the amount depends on several factors such as the size and shape of electrodes, anatomical structure, and type of tissues, it is desirable to conform to the safety guideline. According to the guideline, the current should be limited below 0.1 mA at the frequency range of below 1 kHz. The safety limit increases as frequency goes up and a current up to 5 mA is allowed at 50 kHz and beyond.

From the Biot–Savart law in (19), we can see that the magnitude of magnetic flux density at one point is strongly dependent on the current density near the point. The current density distribution inside the subject could be quite inhomogeneous and very low current density could appear at some local regions depending on the dimension of the subject and electrode configuration. If we use small electrodes compared with the subject size, the current density at the vicinity of the electrodes will be much higher than that in the far region. To alleviate the spatial dependency of the SNR, it may be desirable to use electrodes with an appropriate size.

Since MREIT is still at its early stage of development, most of the published MREIT conductivity images are from numerical simulations and preliminary phantom experiments.[11,12,55,58–60] Lately, Woo *et al.* used a 3 Tesla MREIT system to produce conductivity images of a biological tissue phantom shown in Fig. 13.[62] They injected current pulses with 48 mA amplitude and 10 ms width. Magnetic flux density images are shown in Fig. 14. Figures 15(a) and (b) are the MR magnitude image at the middle imaging slice and the corresponding reconstructed conductivity

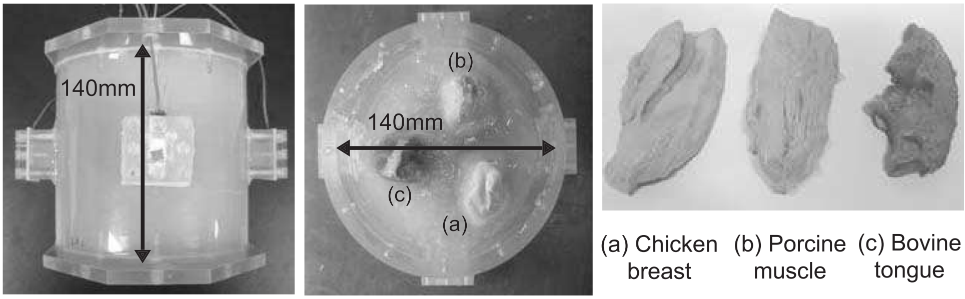
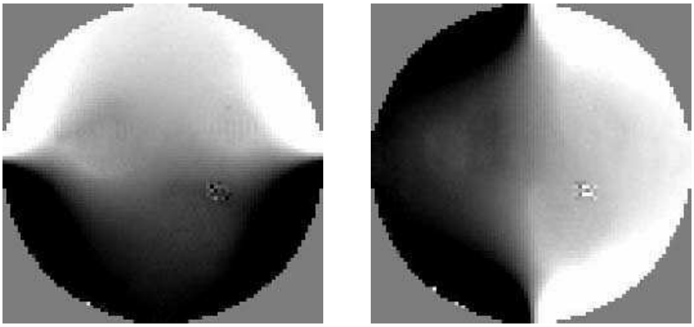
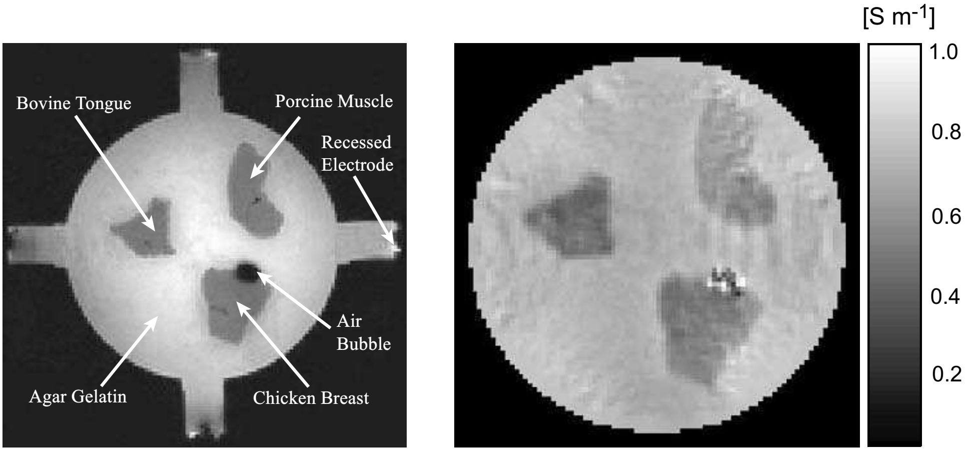

140mm
(b)
140mm
(c)
(a)
(a) Chicken breast
(b) Porcine muscle
(c) Bovine tongue

Fig. 13. Biological tissue phantom.



(a) (b)

Fig. 14. Magnetic flux density images of the tissue phantom in Fig. 13 at the middle imaging slice for (a) horizontal and (b) vertical injection current.


Bovine Tongue
Porcine Muscle
Recessed Electrode
Air Bubble
Agar Gelatin
Chicken Breast
[S m-1]
1.0
0.8
0.6
0.4
0.2

(a) (b)

Fig. 15. (a) MR magnitude image of the tissue phantom in Fig. 13 at the middle imaging slice and (b) reconstructed conductivity image at the same slice.

image, respectively. The pixel size is $1.56 \times 1.56\,\mathrm{mm}^2$ and there are 90×90 pixels in the reconstructed conductivity image. They found that the reconstructed conductivity values of the image in Fig. 15(b) are very close to the measured ones using an impedance analyzer after the experiment. This result demonstrates the feasibility of the MREIT technique in producing conductivity images of different biological tissues with a high spatial resolution and accuracy when we use a sufficient amount of injection current.

Though there are still several technical problems to be solved including the reduction of the amount of injection current, MREIT has the potential to provide cross-sectional conductivity images with better accuracy and spatial resolution. Reconstructed static conductivity images will allow us to obtain internal current density images for any arbitrary injection currents and electrode configurations.

4. Lesion Estimation using EIT

This section handles the problem of estimating or detecting lesions or anomalies inside an electrically conducting subject using boundary measurements of current-voltage data as in EIT. We assume that there exists a high contrast between conductivity or permittivity values of a lesion and the surrounding medium. Here, the major difficulty basically comes from the followings. First, its reconstruction map from the current-voltage data to the geometry of an anomaly is highly nonlinear. Second, the sensitivity of the current-voltage data to the inhomogeneity due to the anomaly is very low. Therefore, as already discussed in the previous sections on EIT, the cross-sectional conductivity and/or permittivity imaging of the subject may not be able to provide enough spatial resolution needed to localize the anomaly. Without appropriately managing this difficulty, any static or absolute EIT image would be suspicious in terms of its accuracy. Hence, this section focuses on the feature extraction of anomalies inside the subject instead of its cross-sectional imaging.

Based on the understanding of fundamental shortcomings in static EIT imaging, Kwon *et al.* looked at the problem in a different way and showed that the estimation of locations and sizes of anomalies is a well-posed problem.[13,14,63] Suppose anomalies $D_1, \ldots, D_M$ occupy a region $D = D_1 \cup \cdots \cup D_M$ inside a background medium Ω. Since the conductivity σ changes abruptly across the interface ∂D, a clear contrast exists between the anomalies and the surrounding medium. To distinguish them, we denote

$$\sigma = \begin{cases} \sigma_0 & \text{in } \Omega \backslash \bar{D} \\ \sigma_j & \text{in each } D_j \end{cases}, \tag{39}$$

for $j = 1, \ldots, M$. Along the interface ∂D, the tangential component of the electric field is continuous while the normal component changes abruptly. If u is the voltage in (3), it satisfies the transmission conditions of

$$\begin{aligned} \sigma_j\, \mathbf{n}(\mathbf{r}) \cdot \nabla u^{int}(\mathbf{r}) &= \sigma_0\, \mathbf{n}(\mathbf{r}) \cdot \nabla u^{int}(\mathbf{r}) \quad \text{for each } \mathbf{r} \in \partial D_j, \\ u^{ext}(\mathbf{r}) &= u^{int}(\mathbf{r}) \quad \text{for } \mathbf{r} \in \partial D, \end{aligned} \tag{40}$$

where $u^{int} := u|_D$ and $u^{ext} := u|_{\Omega \setminus \bar{D}}$ are voltages inside and outside of D, respectively. The inverse problem is to recover anomalies D from the relationship between current $g = -\sigma \nabla u \cdot \mathbf{n}|_{\partial\Omega}$ and boundary voltage $u|_{\partial\Omega}$. The goal of this problem is to develop an algorithm for extracting a quantitative core information of D with a few measured data in such a way that the core information of D is reasonably stable against measurement errors. This section considers non-iterative anomaly estimation algorithms for searching location of anomaly and estimating its size.

4.1. *Global measurement: Ring of electrodes encircling a subject*

To provide a feasible representation formula for estimating an anomaly, we begin with considering the simplest case where $\sigma_0, \sigma_1, \ldots, \sigma_M$ are constants. We assume that the domains $\{D_j\}_{j=1}^M$ are small relative to Ω, separated apart from each other, and away from the boundary. As in Figs. 2 and 3, we place surface electrodes $\mathcal{E}_j$ for $j = 1, \ldots, E$ on the boundary $\partial\Omega$. Let u be the solution of the Neumann boundary value problem in (3) and f be its boundary voltage.

4.1.1. *Mathematical formulation*

In order to extract the information of the collection of anomalies D, it is desirable to express u in terms of D. In the section, we derive a representation formula of u involving D.[13,14,64] Let Φ be the fundamental solution of the Laplace equation:

$$\Phi(\mathbf{r}, \mathbf{r}') = -\frac{1}{4\pi|\mathbf{r} - \mathbf{r}'|} = \frac{-1}{4\pi\sqrt{(x-x')^2 + (y-y')^2 + (z-z')^2}}. \tag{41}$$

Carefully using the transmission condition in (40), we have the following identities:

$$\begin{aligned} u(\mathbf{r}) &= \sum_{j=1}^{M} \left(\frac{\sigma_j}{\sigma_0} - 1\right) \int_{D_j} \nabla u(\mathbf{r}) \cdot \nabla\Phi(\mathbf{r}, \mathbf{r}') d\mathbf{r}' + H(\mathbf{r}; g, f) \quad \text{for } \mathbf{r} \in \Omega, \\ 0 &= \sum_{j=1}^{M} \left(\frac{\sigma_j}{\sigma_0} - 1\right) \int_{D_j} \nabla u(\mathbf{r}) \cdot \nabla\Phi(\mathbf{r}, \mathbf{r}') d\mathbf{r}' + H(\mathbf{r}; g, f) \quad \text{for } \mathbf{r} \in \mathbb{R}^3 \setminus \bar{\Omega}. \end{aligned} \tag{42}$$

Here, $H(\mathbf{r}; g, f)$ is a harmonic function that is computed directly from the data g and f:

$$H(\mathbf{r}; g, f) := \frac{1}{\sigma_0} \int_{\partial\Omega} \Phi(\mathbf{r}, \mathbf{r}') g(\mathbf{r}') ds_{\mathbf{r}'} + \int_{\partial\Omega} \mathbf{n}(\mathbf{r}') \cdot \nabla_{\mathbf{r}'} \Phi(\mathbf{r}, \mathbf{r}') f(\mathbf{r}') ds_{\mathbf{r}'} \tag{43}$$

for $\mathbf{r} \in \mathbb{R}^3 \setminus \partial\Omega$. The derivation of the representation formula (42) is given by Kwon *et al.*[14] In order to extract the core information of D, we need to find a direct interplay between the unknown D and known $H(\mathbf{r}; g, f)$ from the formula (42).

Since (42) involves the unknown u that depends on D in a highly nonlinear way, it is desirable to approximate u in terms of $H(\mathbf{r}; g, f)$ and D.

We express the integral term involving D_j in (42) as a single layer potential with the weight φ_j:

$$\left(\frac{\sigma_j}{\sigma_0} - 1\right) \int_{D_j} \nabla_y \Phi(\mathbf{r} - \mathbf{r}') \cdot \nabla u(\mathbf{r}') d\mathbf{r}' = \int_{\partial D_j} \Phi(\mathbf{r} - \mathbf{r}') \varphi_j(\mathbf{r}') ds_{\mathbf{r}'}, \tag{44}$$

where φ_j with $j = 1, \ldots, M$ is the normal component of ∇u on the interface ∂D multiplied by the constant $(\frac{\sigma_j}{\sigma_0} - 1)$:

$$\varphi_j = \left(\frac{\sigma_j}{\sigma_0} - 1\right) \nabla u \cdot \mathbf{n}|_{\partial D_j}.$$

The main advantage of introducing φ_j is that we can represent φ_j as a function depending only on D_j and the known function $H(\mathbf{r}; g, f)$. To be precise, φ_j is the solution of the following integral identity:

$$\frac{\sigma_0 + \sigma_j}{2(\sigma_j - \sigma_0)} \varphi_j(\mathbf{r}) - \mathcal{K}^*_{D_j} \varphi_j(\mathbf{r}) = \mathbf{n}(\mathbf{r}) \cdot \nabla H_j(\mathbf{r}), \quad \text{for } \mathbf{r} \in \partial D_j, \tag{45}$$

where

$$\begin{aligned} \mathcal{K}^*_{D_j} \varphi_j(\mathbf{r}) &= \frac{1}{4\pi} \int_{\partial D_j} \frac{(\mathbf{r} - \mathbf{r}') \cdot \mathbf{n}(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|^3} \varphi_j(\mathbf{r}') ds_{\mathbf{r}'}, \quad \text{for } \mathbf{r} \in \partial D_j, \\ H_j(\mathbf{r}) &:= H(\mathbf{r}; g, f) + \sum_{k \neq j} \int_{\partial D_j} \Phi(\mathbf{r} - \mathbf{r}') \varphi_k(\mathbf{r}') ds_{\mathbf{r}'}, \quad \text{for } \mathbf{r} \in \mathbb{R}^3. \end{aligned}$$

This enables us to provide the following approximations:

$$\begin{aligned} u(\mathbf{r}) &\approx H(\mathbf{r}; g, f) + \sum_{j=1}^{m} \frac{1}{\lambda_j} \int_{D_j} \nabla_{\mathbf{r}'} \Phi(\mathbf{r} - \mathbf{r}') \cdot \nabla H_j(\mathbf{r}') d\mathbf{r}' \quad \text{for } \mathbf{r} \in \Omega, \\ H(\mathbf{r}; g, f) &\approx -\sum_{j=1}^{m} \frac{1}{\lambda_j} \int_{D_j} \nabla_{\mathbf{r}'} \Phi(\mathbf{r} - \mathbf{r}') \cdot \nabla H_j(\mathbf{r}') d\mathbf{r}' \quad \text{for } \mathbf{r} \in \mathbb{R}^3 \setminus \bar{\Omega}, \end{aligned} \tag{46}$$

where $\lambda_j = \frac{\sigma_0 + \sigma_j}{2(\sigma_j - \sigma_0)}$. The detailed explanation of (46) is given by Ammari *et al.*[63] In the next section, we use the above identities to extract features of D.

4.1.2. *Location search method*

The location of an anomaly can be determined by the pattern of a simple weighted combination of injection current and boundary voltage.[14] We assume that the subject contains a single anomaly $D = D_1$ that is small compared with the subject and separated away from the boundary $\partial\Omega$. We also assume that σ_0 and σ_1 are constants. According to (46), D and $H(\mathbf{r}, g, f)$ satisfy the following approximate identity:

$$\mathbf{H}(\mathbf{r}; g, f) \approx -\frac{1}{\lambda_1}\int_D \nabla_{\mathbf{r}'}\Phi(\mathbf{r}-\mathbf{r}')\cdot\nabla H(\mathbf{r}'; g, f)d\mathbf{r}', \quad \mathbf{r}\in\mathbb{R}^3\backslash\bar{\Omega}. \tag{47}$$

The location search algorithm is based on simple aspects of the function $H(\mathbf{r}; g, f)$ outside the domain Ω which can be computed directly from the data g and f.

For the injection current pattern, we choose $g(\mathbf{r}) = \vec{a}\cdot\mathbf{n}(\mathbf{r})$ for some fixed constant vector $\vec{a}$. Now, we are ready to explain the location search method using only the boundary current-voltage data.

(1) *Take two observation regions Σ_1 and Σ_2 contained in $\mathbb{R}^3\backslash\Omega$ given by*

$$\Sigma_1 := \textit{a line parallel to } \vec{a},$$
$$\Sigma_2 := \textit{a plane (or a line if } n=2\textit{) normal to } \vec{a}.$$

(2) *Find two points $P_i \in \Sigma_i (i = 1, 2)$ so that*

$$H(P_1; g, f) = 0,$$

and

$$H(P_2; g, f) = \begin{cases}\min_{\mathbf{r}\in\Sigma_2} H(\mathbf{r}; g, f) & \textit{if } (\sigma_j-\sigma_0) > 0,\\ \max_{\mathbf{r}\in\Sigma_2} H(\mathbf{r}; g, f) & \textit{if } (\sigma_j-\sigma_0) < 0.\end{cases}$$

(3) *Draw the corresponding plane $\Pi_1(P_1)$ and the line $\Pi_2(P_2)$ given by*

$$\Pi_1(P_1) := \{\mathbf{r}\,|\,\vec{a}\cdot(\mathbf{r}-P_1) = 0\},$$
$$\Pi_2(P_2) := \{\mathbf{r}\,|\,(\mathbf{r}-P_2)\textit{ is parallel to }\vec{a}\}.$$

(4) *Find the intersecting point P of the plane $\Pi_1(P_1)$ and the line $\Pi_2(P_2)$, then this point P can be viewed as the location of the anomaly D.*

In order to provide more insight on the above location search method, we let u_0 be the voltage in the absence of the anomaly D. With the same injection current $g = \vec{a}\cdot\mathbf{n}$, the voltage u_0 satisfies $\nabla u_0|_\Omega = \vec{a}/\sigma_0$. If we denote by f_0 the boundary voltage of u_0, it follows from (42) that

$$0 \approx H(\mathbf{r}; g, f_0) \quad \text{for } \mathbf{r}\in\mathbb{R}^3\backslash\bar{\Omega}. \tag{48}$$

Subtracting (48) from (47) gives

$$H(\mathbf{r}; g, f_0) - H(\mathbf{r}; g, f) \approx \frac{1}{\lambda_1}\int_D \nabla_{\mathbf{r}'}\Phi(\mathbf{r}-\mathbf{r}')\cdot\nabla H(\mathbf{r}'; g, f)d\mathbf{r}', \quad \text{for } \mathbf{r}\in\mathbb{R}^3\backslash\bar{\Omega}.$$

Since $H(\mathbf{r}; g, f_0) = 0$ for $r\in\mathbb{R}^3\backslash\bar{\Omega}$,

$$H(\mathbf{r}; g, f) \approx \frac{-1}{\lambda_1}\int_D \nabla_{\mathbf{r}'}\Phi(\mathbf{r}-\mathbf{r}')\cdot\nabla H(\mathbf{r}'; g, f)d\mathbf{r}', \quad \text{for } \mathbf{r}\in\mathbb{R}^3\backslash\bar{\Omega}. \tag{49}$$

Due to the special injection current $g = \vec{a}\cdot\mathbf{n}$, $\nabla H(\mathbf{r}; g, f_0) = \vec{a}/\sigma_0$. Using the assumption that the anomaly D is relatively small and situated away from $\partial\Omega$, we obtain

$$\nabla H(\mathbf{r}; g, f) \approx \nabla H(\mathbf{r}; g, f_0) = \vec{a}/\sigma_0, \quad \text{for } \mathbf{r}\in D.$$

Hence, (49) is reduced to

$$H(\mathbf{r}; g, f) \approx \frac{1}{4\pi\sigma_0\lambda_1} \int_D \frac{(\mathbf{r}-\mathbf{r}')\cdot\vec{a}}{|\mathbf{r}-\mathbf{r}'|^3}\, d\mathbf{r}', \quad \text{for } \mathbf{r} \in \mathbb{R}^3 \setminus \bar{\Omega}. \tag{50}$$

Examining the integrand of (50), we can see that the sign of $H(\cdot; g, f)$ is determined by $\vec{a}$, D, and the sign of λ_1. Indeed, the identity (50) leads to the crucial observation that for $P_1 \in \mathbb{R}^n \setminus \bar{\Omega}$ with $H(P_1; g, f) = 0$, the plane or the line $\Pi_1(P_1) := \{\mathbf{r} \,|\, \vec{a} \cdot (\mathbf{r} - P_1) = 0\}$ divides the domain D.

4.1.3. *Total size estimation*

The total size estimation of anomalies $D = \cup_{j=1}^M D_j$ also uses the projection current $g = \vec{a} \cdot \mathbf{n}$, where $\vec{a}$ is a unit constant vector. In this section, we assume that the conductivity values of all anomalies are the same constant σ_1 so that $\sigma(\mathbf{r}) = \sigma_0$ in the background region $\Omega \setminus \bar{D}$ and $\sigma = \sigma_1$ in the anomalies D. We may assume Ω contains the origin. Define the scaled domain $\Omega_t = \{t\mathbf{r}\colon \mathbf{r} \in \Omega\}$ for a scaling factor $t > 0$. Let v_r be the solution of the problem

$$\begin{cases} \nabla \cdot ((\sigma_0 \chi_{\Omega\setminus\Omega_t} + \sigma_1 \chi_{\Omega_t})\nabla v_t) = 0 & \text{in } \Omega \\ \sigma_0\, \mathbf{n} \cdot \nabla v_t = g \quad \text{on } \partial\Omega, & \displaystyle\int_{\partial\Omega} v_t = 0 \end{cases},$$

where $\chi_{\mathcal{T}}$ is the indicator function of the domain $\mathcal{T}$.

The total size of D is very close to the size of the domain Ω_{t_0} where t_0 with $0 < t_0 < 1$ is determined uniquely from

$$\int_{\partial\Omega} v_{t_0} g\, ds = \int_{\partial\Omega} u g\, ds. \tag{51}$$

Various numerical experiments indicate that the algorithm gives a nearly exact estimate for arbitrary multiple anomalies even though some restrictions on anomalies are necessary in its rigorous proof.[15]

We now show why t_0 is uniquely determined in the interval $(0, 1)$. Let $\eta(t) := \int_{\partial\Omega} v_t g\, ds$ as a function of t defined in the interval $(0, 1)$. If $t_1 < t_2$, it follows from integration by parts that

$$\begin{aligned} \eta(t_1) - \eta(t_2) &= \int_{\partial\Omega} (v_{t_1} - v_{t_2}) g\, ds \\ &= \int_\Omega (\sigma_0 \chi_{\Omega\setminus\Omega_{t_1}} + \sigma_1 \chi_{\Omega_{t_1}}) |\nabla(v_{t_1} - v_{t_2})|^2\, d\mathbf{r} \\ &\quad + (\sigma_1 - \sigma_0) \int_{\Omega_{t_2}\setminus\Omega_{t_1}} |\nabla v_{t_2}|^2\, d\mathbf{r}, \\ \eta(t_1) - \eta(t_2) &= -\int_\Omega (\sigma_0 \chi_{\Omega\setminus\Omega_{t_2}} + \sigma_1 \chi_{\Omega_{t_2}}) |\nabla(v_{t_1} - v_{t_2})|^2\, d\mathbf{r} \\ &\quad + (\sigma_1 - \sigma_0) \int_{\Omega_{t_2}\setminus\Omega_{t_1}} |\nabla v_{t_1}|^2\, d\mathbf{r}. \end{aligned}$$

These identities give a monotonicity of $\eta(t)$:

$$\eta(t_1) < \eta(t_2) \quad \text{if } (\sigma_1 - \sigma_0) < 0 \quad \text{and} \quad \eta(t_1) > \eta(t_2) \quad \text{if } (\sigma_1 - \sigma_0) > 0.$$

Since $D \subset \Omega$, a similar monotonicity argument leads to the following inequalities:

$$\begin{aligned}
\eta(0) < \int_{\partial\Omega} ug\, ds < \eta(1) \quad \text{if } (\sigma_1 - \sigma_0) < 0,\\
\eta(0) > \int_{\partial\Omega} ug\, ds > \eta(1) \quad \text{if } (\sigma_1 - \sigma_0) > 0.
\end{aligned}$$

Since $\eta(t)$ is continuous, there exists a unique t_0 so that $\eta(t_0) = \int_{\partial\Omega} ug\, ds$.

Next, we try to provide an explanation on the background idea of the following size estimation:

$$\text{the volume of } \Omega_{t_0} \approx \text{the total volume of } \cup_{j=1}^{M} D_j. \tag{52}$$

We begin with the following identities which can be obtained easily from integration by parts:

$$\begin{aligned}
\int_{\partial\Omega} (u - v_t) g\, d\sigma &= \int_{\Omega} \sigma \nabla (u - v_t)|^2\, d\mathbf{r} + (\sigma_1 - \sigma_0) \int_{\Omega_t} |\nabla v_t|^2\, d\mathbf{r}\\
&\quad - (\sigma_1 - \sigma_0) \int_{D} |\nabla v_t|^2\, d\mathbf{r},\\
\int_{\partial\Omega} (u - v_t) g\, d\sigma &= -\int_{\Omega} \sigma_t |\nabla (u - v_t)|^2\, d\mathbf{r} + (\sigma_1 - \sigma_0) \int_{\Omega_t} |\nabla u|^2\, d\mathbf{r}\\
&\quad - (\sigma_1 - \sigma_0) \int_{D} |\nabla u|^2\, d\mathbf{r}
\end{aligned}$$

where $\sigma = \sigma_0 \chi_{\Omega \setminus \bar{D}} + \sigma_1 \chi_D$ and $\sigma_t = \sigma_0 \chi_{\Omega \setminus \bar{\Omega}_t} + \sigma_1 \chi_{\Omega_t}$. By adding the above two identities, we obtain

$$\begin{aligned}
2\int_{\partial\Omega} (u - v_t) g\, d\sigma &= (\sigma_1 - \sigma_0) \left[\int_{D} |\nabla (u - v_t)|^2\, d\mathbf{r} + \int_{\Omega_t} |\nabla v_t|^2 + |\nabla u|^2\, d\mathbf{r} \right]\\
&\quad - (\sigma_1 - \sigma_0) \left[\int_{\Omega_t} |\nabla (u - v_t)|^2\, d\mathbf{r} + \int_{D} |\nabla v_t|^2 + |\nabla u|^2\, d\mathbf{r} \right]\\
&= 2(\sigma_1 - \sigma_0) \left[\int_{\Omega_t} \nabla u \cdot \nabla v_t\, d\mathbf{r} - \int_{D} \nabla u \cdot \nabla v_t\, d\mathbf{r} \right].
\end{aligned}$$

According to the choice of t_0,

$$\int_{\Omega_{t_0}} \nabla u \cdot \nabla v_{t_0}\, d\mathbf{r} = \int_{D} \nabla u \cdot \nabla v_{t_0}\, d\mathbf{r}.$$

The above identity is possible when the volume of Ω_{t_0} is close to the total volume of $D = \cup_{j=1}^{M} D_j$.

4.1.4. *Experimental settings and results*

In order to test the feasibility of the location search and size estimation methods, Kwon *et al.* carried out phantom experiments.[15] They used a circular phantom with 290 mm diameter as a container and it was filled with NaCl solution of 0.69 S/m conductivity. Anomalies with different conductivity values, shapes, and sizes were placed inside the phantom. Equally spaced 32 electrodes were attached on the interior surface of the phantom. Using a 32-channel EIT system with a measurement error of about 1.73%, they applied the algorithms described in the previous sections to the measured boundary current-voltage data. In this section, we describe one example of applying the algorithms.

The circular phantom can be regarded as a unit disk $\Omega := B_1(0,0)$ by normalizing the length scale. In order to demonstrate how the location search and size estimation algorithm work, Kwon *et al.*[15] place four insulators $D = \cup_{j=1}^4 D_j$ into the phantom as shown in Fig. 16:

$$\begin{aligned} D_1 &= B_{0.1138}(0.5172, 0.5172), & D_2 &= B_{0.1759}(-0.5172, 0.5172), \\ D_3 &= B_{0.1828}(-0.5172, -0.5172), & D_4 &= B_{0.2448}(0.1724, -0.1724). \end{aligned}$$

We inject a projection current $g = \vec{a} \cdot \mathbf{n}$ with $\vec{a} = (0,1)$ and measure the boundary voltage f.

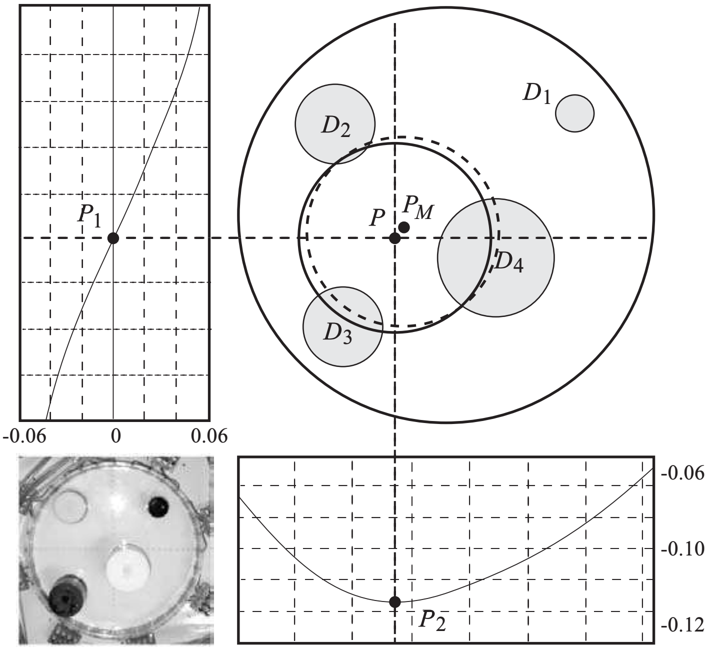


Fig. 16. Illustration of the location and size estimation process. Four anomalies are all insulators and the conductivity of the saline is 0.69 S/m. Lower-left: Configuration of anomalies. Upper-left: H-plot on Σ_1. Lower-right: H-plot on Σ_2. Upper-right: Estimation of the location and size.

For the location search described in Sec. 4.1.2, we choose two observation lines,

$$\Sigma_1 := \{(-1.5, s)|s \in \mathbb{R}\} \quad \text{and} \quad \Sigma_2 := \{(s, -1.5)|s \in \mathbb{R}\}.$$

We evaluate the two dimensional version of $H(\mathbf{r}; f, g)$ defined in (43) with Φ replaced by $\Phi(\mathbf{r}) = \frac{1}{2\pi} \log \sqrt{x^2 + y^2}$. In Fig. 16, the upper-left plot is the graph of $H(\mathbf{r}; f, g)$ on Σ_1 and the lower-right plot is the graph of $H(\mathbf{r}; f, g)$ on Σ_2. We find the zero point of $H(\mathbf{r}; f, g)$ on Σ_1 and the maximum point of $|H(\mathbf{r}; f, g)|$ on Σ_2 denoted by P_1 and P_2, respectively, in Fig. 16. The intersecting point was calculated as $P(-0.1620, -0.0980)$ which is close to the center of mass $P_M(-0.1184, -0.0358)$ of the four anomalies. For the case of a single anomaly or a cluster of multiple anomalies, the intersecting point furnishes a meaningful location information.

For the size estimation, we use (51) and (52) to compute the total size of $\cup_{j=1}^{4} D_j$. The estimated total size was 0.4537 compared with the true total size of 0.4311. In Fig. 16, the corresponding disk with the size of 0.4537 centered at $P(-0.1620, -0.0980)$ is drawn with a solid line and the corresponding disk with the true size centered at P_M is drawn with a dotted line. The relative error of the estimated size was about 5.24%.

4.2. *Local measurement: Planar array of electrodes placed on a portion of a subject*

In this section, we describe lesion estimation techniques using a local measurement from a planar array of electrodes placed on a portion of an electrically conducting subject. Since the breast cancer detection problem is a typical and most important example of this setting, we explain the method in the context of the breast cancer detection problem.

Lately, a commercial system called TransScan TS2000 (TransScan Medical, Ltd., Migdal Ha'Emek, Israel) has been released for adjunctive clinical uses with X-ray mammography in the diagnosis of breast cancer.[65,66] Interestingly, TS2000 system is similar to the frontal plane impedance camera that initiated EIT research early in 1978.[67] In TS2000 system, a metallic cylindrical electrode is held by a patient in her hand. A scan probe with a planar array (16 × 16 or 8 × 8) of electrodes is placed on the breast. A constant voltage of 1 to 2.5 V with frequencies spanning from 100 Hz to 100 kHz is applied to the hand-held electrode and all the electrodes in the scan probe are kept at the ground potential. From each electrode in the planar array, the amplitude and phase shift of the exit current are measured. The measured trans-admittance data from all electrodes of the planar array are displayed as images of conductivity and permittivity. Then, any white spots on these images are interpreted as showing the possibility of cancerous lesions under them. Figure 17 shows the configuration of TS2000 system and the primary goal is to decrease equivocal findings and thereby reducing unnecessary biopsies.

However, the diagnostic information from the currently available TS2000 system lacks of a sophisticated reconstruction method of finding lesions. Definitely,

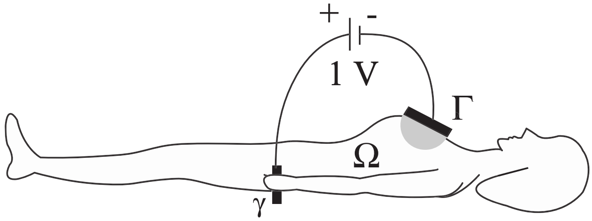


Fig. 17. TS2000 configuration for breast cancer detection.

systematic studies are essential to achieve a higher performance in the breast cancer detection, and it is necessary to derive agreements between experimental results and mathematical theory that provides the relationship between lesions and a trans-admittance map acquired by the scanning probe on the breast.

In order to develop a more sophisticated analysis and lesion estimation algorithm, an explicit representation of the relation between the measured trans-admittance data and the core information of the lesion is required. In this section, we describe a mathematical model based on TS2000 system and develops a lesion estimation algorithm in a systematic way. The spatial distribution of the trans-admittance data from all electrodes of the planar scan probe will be exploited to provide the core information of the lesion.

For the development of a practical estimation algorithm, we must take account the following limitations:

- We only measure the data in a small surface Γ instead of the whole surface $\partial\Omega$.
- Since $\partial\Omega$ differs for each subject, the algorithm should not depend much on the global geometry of $\partial\Omega$.
- Electrical safety regulation limits the amount of total current flowing through the human subject and therefore the range of the applied voltage is also limited.

The challenge of this problem is to develop a proper analysis to provide a quantitative information of a tumor in the breast region from a few measured data on the breast. This must be done in such a way that a reconstruction formula for the tumor is reasonably stable to any change of the conductivity distribution outside breast region.

4.2.1. *Mathematical formulation*

Let Ω be a bounded domain in $\mathbb{R}^3$ with a smooth boundary $\partial\Omega$. We denote by γ the contact surface of the reference electrode and by Γ the probe plane, that is, contact plane of the scan probe on the breast as shown in Fig. 17. If a constant voltage of 1 volt with frequency ω is applied to the hand-held electrode, the resulting voltage $u(\mathbf{r})$ at position $\mathbf{r} = (x, y, z)$ in Ω satisfies the following mixed boundary value problem:

$$\begin{cases} \nabla \cdot ((\sigma + i\omega\epsilon)\nabla u(\mathbf{r})) = 0, & \mathbf{r} \in \Omega \\ u(\mathbf{r}) = 0, & \mathbf{r} \in \Gamma \\ u(\mathbf{r}) = 1, & \mathbf{r} \in \gamma \\ (\sigma + i\omega\epsilon)\nabla u(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0, & \mathbf{r} \in \partial\Omega \setminus \overline{(\Gamma \cup \gamma)}. \end{cases} \tag{53}$$

Both σ and ϵ depend on the position $\mathbf{r}$ and frequency ω.

Through the probe plane Γ, we measure

$$(\sigma + i\omega\epsilon)\nabla u(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) := -g(\mathbf{r}), \quad \mathbf{r} \in \Gamma. \tag{54}$$

The goal is to detect a tumor within a region underneath the probe plane Γ. Now we suppose that there is a tumorous lesion D underneath the probe plane Γ so that the complex conductivity $(\sigma + i\omega\epsilon)$ changes abruptly across the interface ∂D as shown in Fig. 18. A clear contrast exists between the lesion and surrounding normal tissues. To distinguish them, we denote

$$\sigma + i\omega\epsilon = \begin{cases} \sigma_1 + i\omega\epsilon_1 & \text{in } \Omega \setminus \bar{D} \\ \sigma_2 + i\omega\epsilon_2 & \text{in } D. \end{cases} \tag{55}$$

We fix the voltage $f = 0$ on Γ as in TS2000 system. Keeping $f = 0$ on Γ has an advantage because it forces the level surface of the voltage in the breast region to be approximately parallel to the probe plane Γ. Its electric field $-\nabla u$ will be in the direction perpendicular to the level surface, and so more currents will flow along D of which the conductivity σ_2 is much higher than the surrounding.

We denote $\tau_1 = \sigma_1 + i\omega\epsilon_1$ and $\tau_2 = \sigma_2 + i\omega\epsilon_2$. For simplicity, we let z be the label of the axis normal to Γ and let Ω be contained in the lower half space

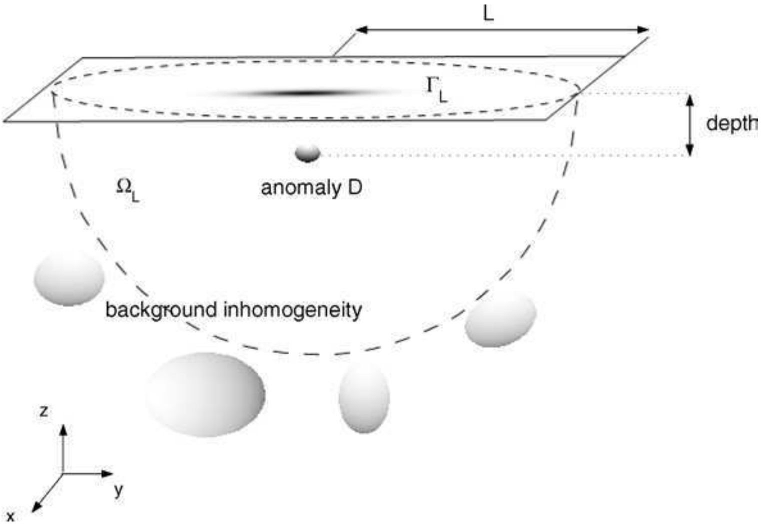


Fig. 18. A lesion to be detected in an inhomogeneous background.

$\Omega = \mathbb{R}^3_- := \{\mathbf{r} = \ \mid z < 0\}$. Let the center of Γ be the origin $\mathbf{0} = (0,0,0)$ and let $B_\rho = \{\mathbf{r} \in \mathbb{R}^3 \colon |\mathbf{r}| < \rho\}$, the ball with radius ρ. Assume that anomaly D is included in the breast region $\Omega_\rho = \Omega \cap B_\rho$. If u_0 is the voltage in the absence of D and $g_0 = -\tau_1 \frac{\partial u_0}{\partial z}|_\Gamma$, then we have the following approximation:

$$\frac{1}{2}[g(\mathbf{r}) - g_0(\mathbf{r})] \approx \frac{\partial}{\partial z} \int_D (\tau_2 - \tau_1) \nabla \Phi(\mathbf{r}, \mathbf{r}') \cdot \nabla u(\mathbf{r}') d\mathbf{r}', \quad \mathbf{r} \in \Gamma_\rho, \tag{56}$$

where $\Gamma_\rho = \Gamma \cap B_\rho$. Under the assumption that τ_1 and τ_2 are almost constant in $\Omega \backslash \bar{D}$ and D, respectively, we can get

$$\frac{1}{2}[g(\mathbf{r}) - g_0(\mathbf{r})] \approx \frac{3\alpha(\tau_1 - \tau_2)}{2\tau_1 + \tau_2} |D| \frac{2\xi_3^2 - (x - \xi_1)^2 - (y - \xi_2)^2}{4\pi |\mathbf{r} - \xi|^5}, \quad \mathbf{r} \in \Gamma_\rho, \tag{57}$$

where ξ is the gravitational center of the anomaly D and $\alpha = \frac{1}{|\Gamma_\rho|} \int_{\Gamma_\rho} g \, ds$ denotes the average of g over Γ_ρ.

In practice, we often do not have *a priori* knowledge of the background τ_1, and so we cannot compute the data g_0. In this case, we may use more than two different frequencies provided that the frequency dependencies of conductivity and permittivity values for background and anomaly are different. Let $\tilde{\omega}$ be a frequency such that the corresponding $\tilde{\tau}_2$ is quite different from τ_2, while $\tilde{\tau}_1$ is close to τ_1. Let $\tilde{u}$ be the solution of (53) at the frequency $\tilde{\omega}$ and let $\tilde{g} = -\tilde{\tau}_1 \frac{\partial \tilde{u}}{\partial z}|_\Gamma$. Suppose now that the difference $\tau_2 - \tilde{\tau}_2$ in cancerous region D is much larger than the difference $\tau_1 - \tilde{\tau}_1$ in the normal region. Then, the expression corresponding to (57) is

$$\frac{1}{2}[\tilde{g}(\mathbf{r}) - g(\mathbf{r})] \approx \frac{9\alpha\tau_1(\tau_2 - \tilde{\tau}_2)}{(2\tau_1 + \tau_2)(2\tau_1 + \tilde{\tau}_2)} |D| \frac{2\xi_3^2 - (x_1 - \xi_1)^2 - (x_2 - \xi_2)^2}{4\pi |\mathbf{r} - \xi|^5}, \quad \mathbf{r} \in \Gamma_\rho. \tag{58}$$

4.2.2. *Lesion estimation algorithm*

The following anomaly estimation procedure for breast cancer detection is based on the formula (57).[68] Similar algorithm can be derived from (58).

- *Transversal position.* We let (x_0, y_0) be the point at which the absolute value $|g(x,y) - g_0(x,y)|$ has the greatest quantity:

$$|g(x_0, y_0) - g_0(x_0, y_0)| = \max_{(x,y) \in \Gamma} |g(x,y) - g_0(x,y)|. \tag{59}$$

- *Depth.* Let (a,b) be any chosen point in Γ near (x_0, y_0) and let d be the distance between (x_0, y_0) and (a,b), that is, $d = \sqrt{(x_0 - a)^2 + (y_0 - b)^2}$. The depth z_0 is determined by the identity:

$$\frac{|g(a,b) - g_0(a,b)|}{|g(x_0, y_0) - g_0(x_0, y_0)|} = \frac{|2 - \frac{d^2}{z_0^2}|}{2\left(\frac{d^2}{z_0^2} + 1\right)^{5/2}}. \tag{60}$$

- *Size.* If we know τ_1 and τ_2, the size $|D|$ is determined by

$$|D| = \frac{\pi |2\tau_1 + \tau_2| |z_0|^3}{|\tau_1 - \tau_2|} \frac{|g(x_0, y_0) - g_0(x_0, y_0)|}{3\bar{g}_0}. \tag{61}$$

4.2.3. *Experimental settings and results*

The method described in the previous section requires applying a voltage between the reference electrode at patient's hand and the planar array of electrodes kept at the ground potential as shown in Fig. 17. The measured data are exit currents through each grounded electrode in the array. This means that we measure a map of trans-admittance using a multi-channel ammeters. Figure 19 shows a typical configuration of a trans-admittance scanner. It is very similar to the typical EIT system shown in Fig. 6 and shares a lot of technical details. Kim *et al.* described the development of a trans-admittance scanner for breast cancer detection including a hand-held electrode, scan probe with planar array of electrodes, constant sinusoidal voltage source, multi-channel ammeters, controller, and computer.[69] Figure 20 shows trans-admittance maps measured from a saline phantom. For the maps in Figs. 20(a) and (b), a small anomaly was located at 5 and 10 mm depth, respectively, from the surface where the scan probe was placed. The size of the scan probe was $33 \times 33\,\text{mm}^2$ and it included 8×8 array of electrodes. We can see that the distributions of the trans-admittance maps provide enough information to distinguish the two different anomalies.

Seo *et al.* evaluated the performance of the estimation algorithm described in the previous section using various anomaly configurations.[68] One of them was a ball-shaped lesion D with its radius $r^* = 0.05$. It was located at $(0, 0, z^*)$ and the depth z^* was set to be one of four different values of -0.2, -0.3, -0.4, and -0.5. Assuming that $\tau_1 = 1$ and $\tau_2 = 5$, Fig. 21 shows four different images of $|g - g^0|$ corresponding to four different depths of D. The brightness of the image indicates the magnitude of $|g - g^0|$. Applying the algorithm, they provided the results summarized in Table 1. From other tests to estimate anomalies inside an

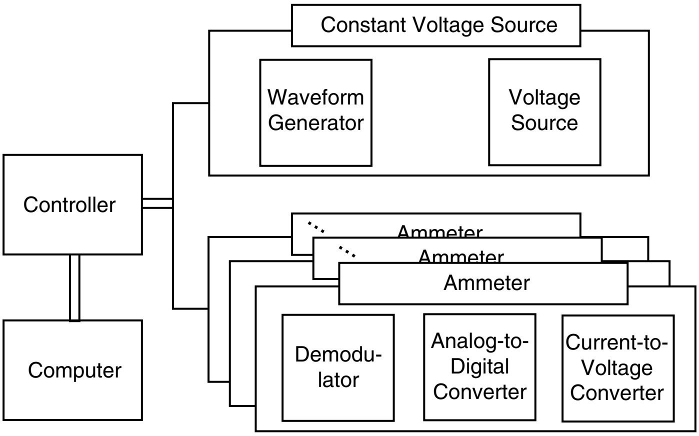
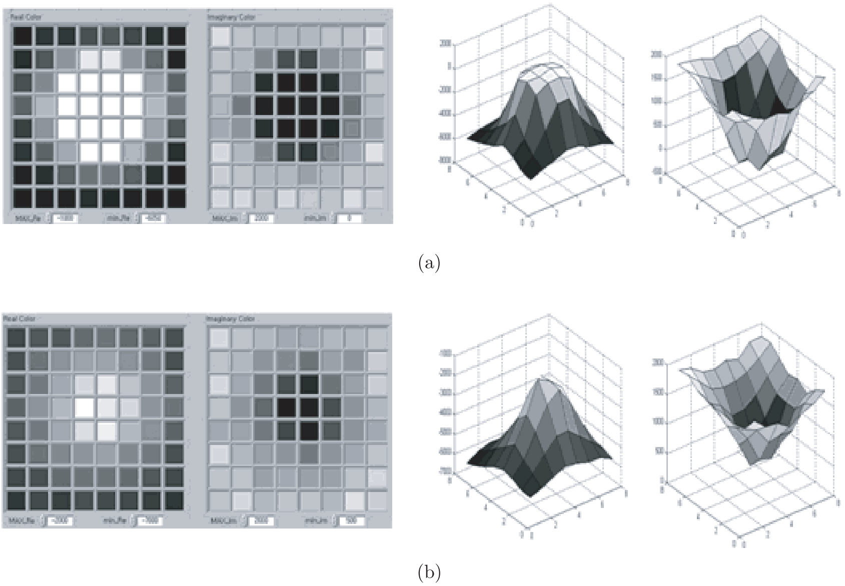


Fig. 19. Block diagram of a typical trans-admittance scanner.



(a)

(b)

Fig. 20. Trans-admittance maps from a saline phantom with a small anomaly at (a) 5 and (b) 10 mm depth from the surface. The maps were measured using a scan probe of 33 × 33 mm^2 with 8×8 array of electrodes. For the maps in (a) and (b), the left is the real part and the right is the imaginary part of the trans-admittance map. Three-dimensional plots are just different views of the corresponding maps.

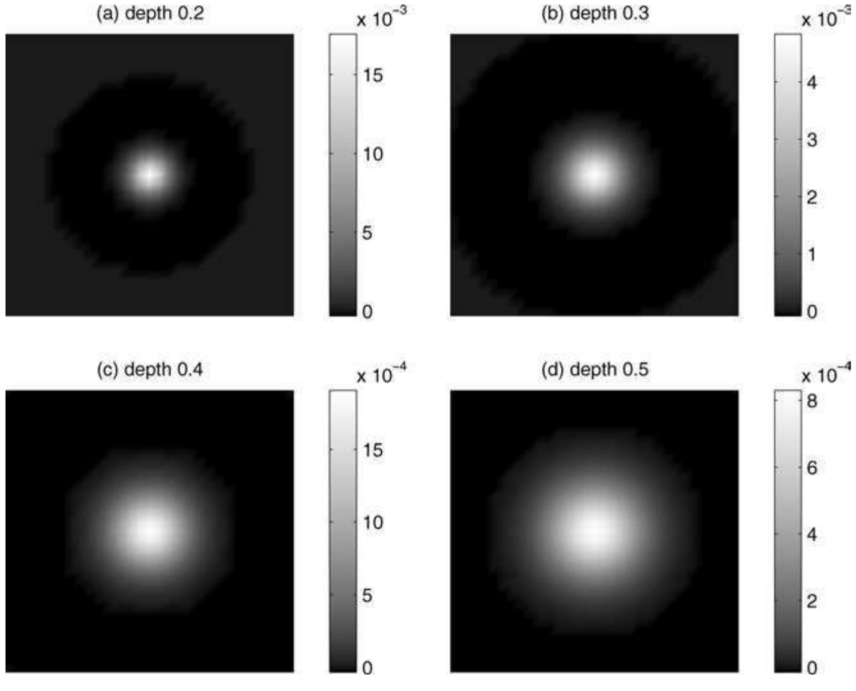


Fig. 21. Magnitude images of $|g - g^0|$ with a ball-shaped lesion D at different depths of -0.2, -0.3, -0.4, and -0.5. The radius of D is 0.05 for all cases.

Table 1. Estimated depth z and radius r of the lesion D at $(0, 0, z^*)$ with $r^* = 0.05$.

z^*	r^*	z	r
−0.2	0.05	−0.18976	0.05838
−0.3	0.05	−0.28973	0.05762
−0.4	0.05	−0.38969	0.05665
−0.5	0.05	−0.48966	0.05501

inhomogeneous background, they found that both location and size estimate are within an acceptable range of accuracy assuming that the measurement noise is below 1%.

5. Applications of EIT, MREIT, and Lesion Estimation

EIT can be used to visualize physiological activities in the human body such as respiration, cardiac circulation, brain function, stomach emptying, fracture healing, bladder filling, and others.[2,40] One of the most tried clinical applications of EIT is the imaging of the thorax including lungs and cardiovascular system. There are numerous experimental studies in this area as summarized by Metherall.[4] These include imaging of respiratory related impedance changes, cardio-synchronous changes, and also perfusion related changes. Especially, EIT has been suggested to diagnosis pulmonary embolism and other lung diseases.

Applying EIT techniques to image brain activities has been tried by Tidswell *et al.*[30,44] Reginal cerebral blood flow and blood volume changes during brain activity alter the local impedance of the corresponding cortical area. Even though current EIT images have a relatively poor spatial resolution, the high temporal resolution and portability could be a big advantage in neuroimaging area especially for the imaging of epilepsy, migraine, and stroke.

In this chapter, we took the breast cancer detection as an example for the lesion estimation technique. There are, however, different approaches using EIT imaging methods for the detection of breast cancers. Kerner *et al.* placed a circular array of electrodes around the breast and produced cross-sectional conductivity images.[70] There are also several investigations for the usefulness of planar electrode arrays instead of conventional circular electrode arrays in breast EIT imaging.[71–73] Cherepenin *et al.* used a planar array of 256 electrodes placed on the breast to reconstruct so called electrical impedance mammograms by sequentially injecting currents through chosen electrodes and measuring voltage data on other electrodes in the array.[74,75]

Clinical applications of MREIT have not been tried yet since it is still in its early stage of development. Once we could reconstruct cross-sectional conductivity and current density images with improved spatial resolution and accuracy, MREIT will find numerous clinical applications. These include all clinical application areas of EIT where static or absolute values of conductivity and current density are needed. However, the temporal resolution of MREIT is expected to be much worse than that of EIT and MREIT lacks the portability. Therefore, MREIT will never replace EIT in application areas where monitoring of fast physiological events is requested. For this kind of applications, MREIT may provide static conductivity images to be utilized as *a priori* information in subsequent dynamic EIT image reconstructions.

Bodurka *et al.* tried a direct mapping of neural activity by measuring very weak transient magnetic field changes using a 3.0 Tesla MRI scanner.[76] Providing static images of conductivity and current density distributions, MREIT could find important contributions in the areas of neuronal source localization and mapping. There are numerous methods of applying electromagnetic energy to the human body mostly for therapeutic purposes. Conductivity information from MREIT will be valuable for the optimization of these therapeutic treatments and current density images could be used to visualize how therapeutic electric currents are actually distributed within the subject. Based on the temperature dependency of tissue conductivity, MREIT could also be used for internal temperature mappings.

The feature extraction or lesion estimation techniques have numerous applications including breast cancer detection, corrosion detection, crack detection, electric field sensing, and bubble detection. These methods may also be utilized as a parametric dynamic imaging technique where temporal and/or spatial changes in locations and sizes of lesions are of primary concern. These may include the monitoring of impedance related physiological events, bubble detection in two-phase flow, and others in medicine and nondestructive testing.

6. Conclusions and Outlook

EIT has been an active research area since early 1980s. Struggling to overcome the ill-posed nature of the inverse problem in EIT image reconstructions, numerous techniques have been suggested. Even with these efforts, further improvements in image quality are needed for more successful clinical applications. Three-dimensional dynamic EIT imaging with a wireless miniaturized EIT system is believed to make the next breakthrough in EIT technology. Obtaining and utilizing the accurate data for the shape and size of a subject with electrode positions, static EIT imaging could be improved. However, since the ill-posedness in EIT still remains anyway, we should not expect EIT to compete with other medical imaging modalities such as MRI and X-ray CT in terms of spatial resolution. The significance of EIT should be emphasized based on the fact that it provides the information on electrical properties of biological tissues. Since this kind of information is not available from any other imaging modality, EIT should keep finding unique application areas especially in dynamic functional imaging. Based on the frequency dependent characteristics of tissue conductivity and permittivity, multi-frequency three-dimensional EIT imaging is also quite promising.

The latest research outcomes in MREIT show the definite feasibility of the technique for high-resolution static conductivity and current density imaging. With many possible clinical and biological applications in mind, future research direction in MREIT should follow the way to reduce the amount of the required injection current. Anisotropic conductivity imaging is believed to be pursuable in MREIT even though it is not feasible in EIT. MREIT should always include the current density imaging to provide more information from the same measured data. The performance of the MRI system itself has been greatly enhanced to make 3 Tesla systems available to clinical settings. The progress in MREIT techniques will follow this trend.

The lesion estimation technique can be viewed as a parametric EIT imaging. Providing the core information on lesions inside a subject, it can effectively avoid the ill-posedness in conventional cross-sectional EIT imaging. Considering the fact that there are numerous medical and also industrial applications, further developments in its theory, algorithms, and measurement techniques are desirable. Currently, the potential applicability of the technique to breast cancer detection is the strongest motivation for future research efforts.

Acknowledgments

This work was supported by grant R11-2002-103 from Korea Science and Engineering Foundation.

References

1. S. Grimnes and O. G. Martinsen, *Bioimpedance and Bioelectricity Basics* (Academic Press, London, UK, 2000).
2. J. G. Webster (ed.), *Electrical Impedance Tomography* (Adam Hilger, Bristol, UK, 1990).
3. K. Boone, D. Barber and B. Brown, *J. Med. Eng. Tech.* **21** (1997) 201.
4. P. Metherall, *Three Dimensional Electrical Impedance Tomography of the Human Thorax* (PhD Thesis, University of Sheffield, Dept. of Med. Phys and Clin. Eng., 1998).
5. M. Cheney, D. Isaacson and J. C. Newell, *SIAM Rev.* **41** (1999) 85.
6. G. J. Saulnier, R. S. Blue, J. C. Newell, D. Isaacson and P. M. Edic, *IEEE Sig. Proc. Mag.* **18** (2001) 31.
7. N. Zhang, *Electrical Impedance Tomography Based on Current Density Imaging* (MS Thesis, Dept. of Elec. Eng., Univ. of Toronto, Toronto, Canada, 1992).
8. E. J. Woo, S. Y. Lee and C. W. Mun, *SPIE* **2299** (1994) 377.
9. Y. Z. Ider and O. Birgul, *Elektrik* **6** (1998) 215.
10. O. Kwon, E. J. Woo, J. R. Yoon and J. K. Seo, *IEEE Trans. Biomed. Eng.* **48** (2002) 160.
11. S. H. Oh, B. I. Lee, E. J. Woo, S. Y. Lee, M. H. Cho, O. Kwon and J. K. Seo, *Phys. Med. Biol.* **48** (2003) 3101.
12. S. H. Oh, B. I. Lee, E. J. Woo, S. Y. Lee, M. H. Cho, O. Kwon and J. K. Seo, *Mag. Reson. Med.* **51** (2004) 1292.
13. O. Kwon and J. K. Seo, *Inv. Prob.* **17** (2001) 59.
14. O. Kwon, J. K. Seo and J. R. Yoon, *Comm. Pure Appl. Math.* **LV** (2002) 1.
15. O. Kwon, J. R. Yoon, J. K. Seo, E. J. Woo and Y. G. Cho, *IEEE Trans. Biomed. Eng.* **50** (2003) 89.

16. A. Greenleaf, M. Lassas and G. Uhlmann, *Physiol. Meas.* **24** (2003) 413.
17. N. Polydorides and W. R. B. Lionheart, *Meas. Sci. Technol.* **13** (2002) 1871.
18. B. I. Lee, S. H. Oh, E. J. Woo, S. Y. Lee, M. H. Cho, O. Kwon, J. K. Seo, J. Y. Lee and W. S. Baek, *Phys. Med. Biol.* **48** (2003) 1971.
19. R. Kohn and M. Vogelius, *Comm. Pure Appl. Math.* **37** (1984) 113.
20. J. Sylvester and G. Uhlmann, *Comm. Pure Appl. Math.* **39** (1986) 91.
21. J. Sylvester and G. Uhlmann, *Ann. Math.* **125** (1987) 153.
22. A. Nachman, *Ann. Math.* **128** (1988) 531.
23. V. Isakov, *Comm. Pure Appl. Math.* **41** (1988) 856.
24. A. Friedman and V. Isakov, *Indiana Univ. Math. J.* **38** (1989) 553.
25. G. Alessandrini, V. Isakov and J. Powell, *Trans. Amer. Math. Soc.* **347** (1995) 3031.
26. A. Nachman, *Ann. Math.* **142** (1996) 71.
27. G. Uhlmann, in *Harmonic Analysis and Partial Differential Equations*, M. Christ, C. E. Kenig and C. Sadosky (eds.) (Univ. Chicago Press, Chicago, IL, 1999), p. 295.
28. D. C. Barber and B. H. Brown, *J. Phys. E: Sci. Inst.* **17** (1984) 723.
29. P. Metherall, D. C. Barber, R. H. Smallwood and B. H. Brown, *Nature* **380** (1996) 509.
30. A. T. Tidswell, A. Gibson, R. H. Bayford and D. S. Holder, *Neuroimage* **13** (2001) 283.
31. K. Cheng, D. Isaacson, J. C. Newell and D. G. Gisser, *IEEE Trans. Biomed. Eng.* **36** (1989) 918.
32. P. J. Vauhkonen, M. Vauhkonen, T. Savolainen and J. P. Kaipio, *IEEE Trans. Biomed. Eng.* **46** (1999) 1150.
33. T. J. Yorkey, J. G. Webster and W. J. Tompkins, *IEEE Trans. Biomed. Eng.* **34** (1987) 843.
34. M. Cheney, D. Isaacson, J. C. Newell, S. Simske and J. Goble, *Int. J. Imag. Sys. and Tech.* **2** (1990) 66.
35. E. J. Woo, P. Hua, J. G. Webster and W. J. Tompkins, *IEEE Trans. Med. Imag.* **12** (1993) 137.
36. C. Cohen-Bacrie and R. Guardo, *IEEE Trans. Med. Imag.* **16** (1997) 562.
37. P. Edic, D. Isaacson, G. Saulnier, H. Jain and J. C. Newell, *IEEE Trans. Biomed. Eng.* **45** (1998) 899.
38. M. Vauhkonen, D. Vadasz, P. A. Karjalainen, E. Somersalo and J. P. Kaipio, *IEEE Trans. Med. Imag.* **17** (1998) 285.
39. A. L. Hyaric and M. K. Pidcock, *IEEE Trans. Biomed. Eng.* **48** (2001) 230.
40. D. Holder (ed.), *Clinical and Physiological Applications of Electrical Impedance Tomography* (University College London Press, London, UK, 1993).
41. D. Isaacson, *IEEE Trans. Med. Imag.* **5** (1986) 91.
42. A. J. Wilson, P. Milnes, A. R. Waterworth, R. H. Smallwood and B. H. Brown, *Physiol. Meas.* **22** (2001) 49.
43. R. D. Cook, G. J. Saulnier, D. G. Gisser, J. C. Goble, J. C. Newell and D. Isaacson, *IEEE Trans. Biomed. Eng.* **41** (1994) 713.
44. A. T. Tidswell, A. Gibson, R. H. Bayford and D. S. Holder, *Physiol. Meas.* **22** (2001) 177.
45. M. L. G. Joy, G. C. Scott and R. M. Henkelman, *Magn. Reson. Imag.* **7** (1989) 89.
46. G. C. Scott, M. L. G. Joy, R. L. Armstrong and R. M. Henkelman, *IEEE Trans. Med. Imag.* **10** (1991) 362.
47. G. C. Scott, M. L. G. Joy, R. L. Armstrong and R. M. Henkelman, *J. Magn. Reson.* **97** (1992) 235.
48. Y. J. Kim, O. Kwon, J. K. Seo and E. J. Woo, *Inv. Prob.* **19** (2003) 1213.

49. S. Onart, Y. Z. Ider and W. R. B. Lionheart, *Physiol. Meas.* **24** (2003) 591.
50. H. S. Khang, B. I. Lee, S. H. Oh, E. J. Woo, S. Y. Lee, M. H. Cho, O. Kwon, J. R. Yoon and J. K. Seo, *IEEE Trans. Med. Imag.* **21** (2002) 695.
51. B. I. Lee, S. H. Oh, E. J. Woo, S. Y. Lee, M. H. Cho, O. Kwon, J. K. Seo and W. S. Baek, *Physiol. Meas.* **24** (2003) 579.
52. B. M. Eyuboglu, O. Birgul and Y. Z. Ider, *Phys. Med. Biol.* **48** (2003) 653.
53. O. Kwon, J. Y. Lee and J. R. Yoon, *Inv. Prob.* **18** (2002) 1089.
54. J. Y. Lee, *Inv. Prob.* **20** (2004) 847.
55. J. K. Seo, J. R. Yoon, E. J. Woo and O. Kwon, *IEEE Trans. Biomed. Eng.* **50** (2003) 1121.
56. S. W. Kim, O. Kwon, J. K. Seo and J. R. Yoon, *SIAM J. Math. Anal.* **34** (2002) 511.
57. G. Folland, *Introduction to Partial Differential Equations* (Princeton University Press, Princeton, New Jersey, 1976).
58. C. Park, O. Kwon, E. J. Woo and J. K. Seo, *IEEE Trans. Med. Imag.* **23** (2004) 388.
59. C. Park, E. J. Park, E. J. Woo, O. Kwon and J. K. Seo, *Physiol. Meas.* **25** (2004) 257.
60. J. K. Seo, H. C. Pyo, C. J. Park, O. Kwon and E. J. Woo, *Phys. Med. Biol.* **49** (in press 2004).
61. D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms and Software* (Wiley Interscience, New York, 1998).
62. E. J. Woo, S. Y. Lee, J. K. Seo, O. Kwon, S. H. Oh and B. I. Lee, *26th Ann. Int. Conf. IEEE EMBS* (in press 2004).
63. H. Ammari and J. K. Seo, *Adv. Appl. Math.* **30** (2003) 679.
64. H. Kang and J. K. Seo, *Inv. Prob.* **15** (1999) 851.
65. M. Assenheimer, O. Laver-Moskovitz, D. Malonek, D. Manor, U. Nahliel, R. Nitzan and A. Saad, *Physiol. Meas.* **22** (2001) 1.
66. B. Scholz, *IEEE Trans. Med. Imag.* **21** (2002) 588.
67. R. P. Henderson and J. G. Webster, *IEEE Trans. Biomed. Eng.* **25** (1978) 250.
68. J. K. Seo, O. Kwon, H. Ammari and E. J. Woo, *IEEE Trans. Biomed. Eng.* **51** (in press 2004).
69. K. S. Kim, S. M. Baek, J. S. Lee, T. I. Oh, E. J. Woo, O. Kown and J. K. Seo, *Proc. XII ICEBI and V EIT* (2004), p. 679.
70. T. E. Kerner, K. D. Paulsen, A. Hartov, S. K. Soho and S. P. Poplack, *IEEE Trans. Med. Imag.* **21** (2002) 638.
71. J. L. Larson-Wiseman, *Early Breast Cancer Detection Utilizing Clustered Electrode Arrays in Impedance Imaging* (PhD Thesis, Rensselaer Polytechnic Institute, Troy, New York, 1998).
72. J. L. Mueller, D. Isaacson and J. C. Newell, *IEEE Trans. Biomed. Eng.* **46** (1999) 1379.
73. T. Kao, J. C. Newell, G. J. Saulinier and D. Isaacson, *Physiol. Meas.* **24** (2003) 403.
74. V. Cherepenin, A. Karpov, A. Korjenevsky, V. Kornienko, A. Mazaletskaya, D. Mazourov and D. Meister, *Physiol. Meas.* **22** (2001) 9.
75. V. A. Cherepenin, A. Y. Karpov, A. V. Korjenevsky, V. N. Kornienko, Y. S. Kultiasov, M. B. Ochapkin, O. V. Trochanova and J. D. Meister, *IEEE Trans. Med. Imag.* **21** (2002) 662.
76. J. Bodurka and P. A. Bandettini, *Magn. Reson. Med.* **47** (2002) 1052.

This page intentionally left blank

CHAPTER 8

SINGLE-SHOT MAGNETIC RESONANCE IMAGING (MRI) TECHNIQUES AND THEIR APPLICATIONS

YIHONG YANG*, HONG GU, THOMAS J. ROSS, WANG ZHAN and SHAOLIN YANG
Neuroimaging Research Branch, National Institute on Drug Abuse
5500 Nathan Shock Drive, Building C Room 383, Baltimore, MD, USA
**yihongyang@intra.nida.nih.gov*

Single-shot fast MRI acquisition techniques are described in this chapter, with an emphasis on echo-planar imaging and spiral imaging. Advantages and potential shortcomings of these imaging schemes are discussed. Applications of the fast acquisition techniques in advanced MRI techniques such as functional neuroimaging, perfusion imaging with arterial spin labeling, diffusion tensor imaging and perfusion imaging with injection of susceptibility contrast agents are demonstrated.

Keywords: Magnetic resonance imaging; image acquisition; distortion; BOLD.

1. Introduction

Fast MRI acquisition techniques[1–4] have been of great interest in a number of application areas including functional neuroimaging (fMRI),[5] Diffusion Tensor Imaging (DTI),[6] and perfusion imaging based on Arterial Spin-Labeling (ASL)[7] or Dynamic Susceptibility Contrast (DSC).[8] Many of the fast imaging techniques allow data acquisition of a slice in a single excitation (in 20–100 ms). The capability of scanning a whole brain volume in a few seconds provides excellent opportunities for observing dynamic changes of physiological variations (e.g. blood volume, flow and oxygenation) associated with brain activity. Single-shot acquisition is generally less sensitive to motion artifacts. This feature is critical for diffusion tensor imaging and arterial spin-labeling perfusion imaging, in which either the signal is dependent on microscopic movements of molecules (for DTI) or the inherit image contrast is very subtle (for ASL). Furthermore, the efficient data acquisition of single-shot fast imaging leads to higher Signal-to-Noise Ratio (SNR) per unit time than conventional imaging techniques, which is essential for many applications.

In this chapter, we will first introduce a number of single-shot image acquisition techniques, with an emphasis on two popular techniques: Echo-Planar Imaging (EPI) and spiral imaging. The advantages and potential pitfalls of these techniques will be discussed. Then, we will demonstrate applications of fast image acquisition in several advanced MRI techniques, including fMRI, DTI, ASL and DSC perfusion imaging. Through this chapter, readers are expected to learn the basic concepts of the single-shot image acquisition schemes and their potential applications.

2. Single-Shot Image Acquisition Techniques

2.1. *Data acquisition in spatial-frequency space (k-space)*

For a 2D object $I(x, y)$, the signal detected in MRI $S(t)$, neglecting relaxation, can be described as a planar integral of the object function multiplied by a phase factor Φ

$$S(t) = \iint I(x, y)e^{-i2\pi\Phi}\,dx\,dy. \tag{1}$$

The phase factor is modulated by magnetic field gradients, G_x and G_y, in order to encode spatial information into frequency and phase. Based on the angular frequency of nuclear precession (Larmor frequency), Φ is determined by

$$\begin{aligned}\Phi &= (\gamma/2\pi)\left[x\int_0^t G_x(\tau)d\tau + y\int_0^t G_y(\tau)d\tau\right],\\ &= k_x(t)x + k_y(t)y\end{aligned} \tag{2}$$

where γ is the gyromagnetic ratio, and $k_x(t)$ and $k_y(t)$ are defined as

$$\begin{aligned}k_x(t) &= (\gamma/2\pi)\int_0^t G_x(\tau)d\tau,\\ k_y(t) &= (\gamma/2\pi)\int_0^t G_y(\tau)d\tau.\end{aligned} \tag{3}$$

Therefore, the MRI signal (Eq. (1)) can be expressed as

$$S(k_x, k_y) = \iint I(x, y)e^{-i2\pi[k_x(t)x + k_y(t)y]}dx\,dy. \tag{4}$$

This equation states that the MRI signal $S(k_x,\ k_y)$ is simply a 2D Fourier transform of the object function $I(x, y)$. The Fourier transform space is usually called spatial-frequency space or k-space,[9] where k represents the spatial-frequency variable. Typically, k_x and k_y are in the units of cycle/cm. Figure 1 illustrates the Fourier transform relationship between an object and the corresponding MRI signal in k-space.

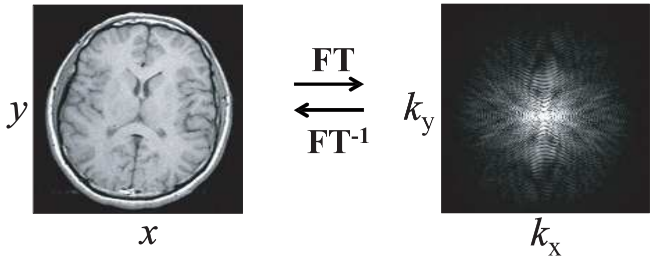


Fig. 1. Illustration of the Fourier transform relationship between an object and the corresponding MRI signal in the k-space. A brain image is shown on the left, and its Fourier transform on the right.

Given the description above, the process of image acquisition in MRI is to collect data in k-space that can be reconstructed to an image as an estimate of the original object. In traditional MRI techniques, such as T_1-weighted spin-echo imaging, only a single line in the k-space is collected per excitation, leading to a lengthy acquisition time. Imaging time can be reduced significantly by techniques acquiring multi-lines per excitation or using partial excitation with reduced repetition time for each line. Single-shot imaging techniques collect the entire data in k-space after a single excitation. These techniques, with the advantages of high temporal resolution, high signal-to-noise ratio, and insensitivity to motion artifacts, have been widely used in research and clinical applications.

Various types of single-shot techniques have been proposed. These techniques can be categorized into two major groups based on their traversal patterns in k-space during data acquisition. The first group collects data in rectilinear patterns, and echo-planar imaging[1,2] is a representative of this group. Image reconstruction is convenient and rapid for these techniques since a two-dimensional (2D) Fast Fourier Transform (FFT) algorithm can be used directly. The second group acquires data in non-rectilinear trajectories, such as spirals,[3,4] rosettes,[10] Lissajous,[11] and radial patterns.[12] These techniques usually have faster acquisition speed due to their efficient coverage of k-space. However, image reconstruction is usually not straightforward, because regridding of the data is often required before using the FFT algorithm.

2.2. *Echo-planar imaging*

Echo-Planar Imaging (EPI) is the first single-shot imaging technique and was proposed by Mansfield,[1] shortly after the inception of MRI. EPI is characterized by the acquisition of planar data in k-space per excitation with the data collected on a rectilinear trajectory. Figure 2 shows a typical EPI pulse sequence (time series of radio-frequency pulse and magnetic gradients). A 2D slice is first selected by a slice-selective Radio-Frequency (RF) pulse together with a z-gradient (G_z). In-plane gradients (G_x and G_y) are then used to encode spatial information into frequency

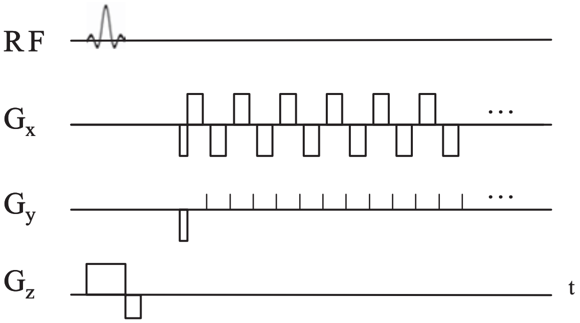


Fig. 2. An EPI pulse sequence for collecting a set of 2D data in a single excitation.

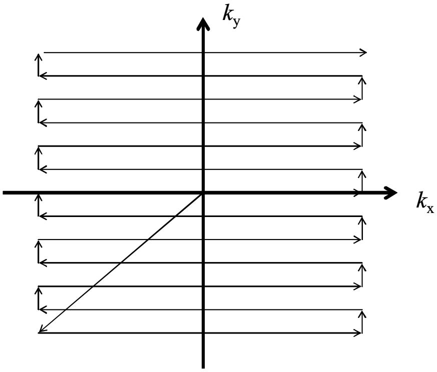


Fig. 3. k-space trajectory of the EPI pulse sequence shown in Fig. 2.

and phase of the MRI signal. Since k_x and k_y are proportional to the integral of G_x and G_y along the time, respectively, as indicated in Eq. (3), an appropriate design of the in-plane gradients will allow acquisition of data on Cartesian grid.

As shown in Fig. 3, after initial movement to the lower left corner of k-space, the trajectory moves along the k_x direction driven by the positive G_x, and blips upwards in the k_y direction due to G_y. Then, the trajectory travels back along the k_x direction by the negative G_x, and moves up by G_y. These steps are repeated until the entire k-space is scanned. G_x is called the frequency-encoding gradient because it makes the oscillation frequency of the MRI signal linearly dependent on the spatial location. G_y, on the other hand, is called the phase-encoding gradient, due to the fact that signals from different y-positions in the object accumulate different phase angles. The pulse sequence in Fig. 2 assumes ideal switching of the gradients. In practice, due to the limited gradient slew rate (dB/dt), the rising and falling time of the gradients have to be taken into account. Usually, trapezoidal gradient waveforms, instead of rectangular waveforms, are used in G_x, and triangular "blips" are used in G_y.

2.3. *Spiral imaging*

Spiral imaging usually acquires data along an Archimedian spiral trajectory in k-space.[3,4] The trajectory typically starts at the center of k-space and spirals out radially with certain constraints, such as at a constant angular velocity or a constant linear velocity. Advantages of spiral methods include superior performance in the presence of motion and flow, and less demanding gradient strength and slew rate. Spiral imaging has been used in a variety of applications, including cardiac imaging, flow imaging, abdominal imaging, and functional neuroimaging.

An Archimedian spiral in k-space is described by

$$k(t) = A\theta(t)e^{i\theta(t)}, \tag{5}$$

where $k(t) = k_x(t) + ik_y(t)$ is the complex location in k-space, A is a constant that is determined by the Field-Of-View (FOV) of the image and the number of spiral interleaves N ($N = 1$ for single-shot), with $A = N$ / FOV. The required gradient waveform in complex form is given by[13,14]

$$G = \frac{1}{\gamma}\frac{dk}{dt} = \frac{A}{\lambda}\dot{\theta}(1 + i\theta)e^{i\theta}, \tag{6}$$

where $\dot{\theta} \equiv d\theta/dt$. The gradient slew rate is then given by

$$S = \frac{dG}{dt} = \frac{A}{\lambda}\{\ddot{\theta} - \theta\dot{\theta}^2 + i[\theta\ddot{\theta} + 2\dot{\theta}^2]\}e^{i\theta}, \tag{7}$$

where $\ddot{\theta} \equiv d^2\theta/d^2t$. Considering hardware constraints for gradient amplifiers, Eqs. (6) and (7) can be rearranged as[14]

$$\ddot{\theta} = \frac{f(\theta, \dot{\theta}) - \theta\dot{\theta}^2}{1 + \theta^2}, \tag{8}$$

where

$$f(\theta, \dot{\theta}) = \begin{cases} [\alpha^2(1 + \theta^2) - \dot{\theta}^4(2 + \theta^2)^2]^{1/2}, & \text{if } |G| < G_{max}, \\ \quad 0 \quad \text{otherwise} \end{cases} \tag{9}$$

and $\alpha = \gamma S_{max}/A$, and G_{max} and S_{max} are the maximum available gradient strength and slew rate, respectively.

Equations (8) and (9) can be solved by numerical methods such as a 4th-order Runge–Kutta algorithm. Alternatively, simple analytic algorithms with appropriate approximations can be found to speed up the spiral waveform design for real-time applications. An example of spiral trajectories generated with slew-rate constraints is illustrated in Fig. 4.

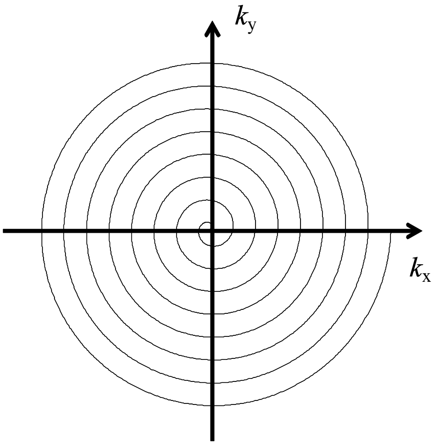


Fig. 4. k-space trajectory of a spiral pulse sequence.

Reconstruction of a spiral image includes correction of non-uniform sampling density in k-space, resampling of the data onto an orthogonal equidistant grid, Fourier transform of the k-space data to image space, and correction of the apodizing effects of the regridding.[15]

3. Advantages and Potential Artifacts

Single-shot MRI techniques can collect an MR image in about 20–100 ms. Such a rapid acquisition has a number of advantages over conventional MR imaging that requires multiple RF excitations. First, the efficient data acquisition schemes in single-shot imaging techniques provide rapid imaging speed (or high temporal resolution) for dynamic imaging applications. Secondly, single-shot techniques can achieve higher Signal-to-Noise Ratio (SNR) per unit time than multi-shot methods due to their capability to collect a larger number of images within a fixed time duration. For comparable voxel size, theoretical analyses and empirical measurements have demonstrated a nearly five-fold SNR improvement of EPI over FLASH techniques.[2] Thirdly, single-shot MRI is less sensitive to patient motion, and thus is helpful for reducing motion-related artifacts. With these advantages, single-shot techniques have been widely employed to study dynamic and functional processes in biological systems.

There are several potential artifacts in single-shot imaging techniques. In principle, the reconstructed MR image should reflect the original object authentically, irrespective of the k-space sampling trajectory. In practice, however, each technique, with its specific k-space traversal pattern, has its own traits and artifact behavior. In this section, the properties and basic problems of EPI and spiral imaging are discussed.

3.1. *Chemical shift*

In MRI, the spatial location of protons in the object can be accurately encoded by external gradients, with the assumption that the local magnetic field to which the protons are exposed is solely determined by the externally superimposed gradients. However, this is not true for protons with different resonant frequencies, such as fat and water. The resonant frequency of a nucleus is affected by its local molecular structure, and this frequency shift is called chemical shift. Without taking the chemical shift effect into consideration, signals from fat and water at the same position will be mapped to different image locations, resulting in chemical shift artifacts as shown in Fig. 5.

The chemical shift between fat and water protons is about 3.5 parts per million (ppm). This corresponds to a 447 Hz difference of resonant frequency at 3 Tesla. The chemical shift artifact is inversely proportional to the receiver bandwidth, i.e. the range of frequencies used to encode the MR signal. In EPI sequences, the receiver bandwidth along frequency-encoding direction is usually

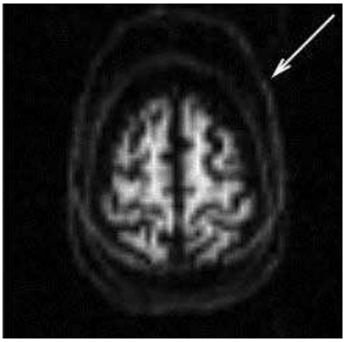


Fig. 5. Chemical shift artifacts caused by different resonant frequencies of water and fat.

very high, typically 250 kHz, amounting to 3906 Hz/pixel for 64 sampling points. In this case, the fat signal will be shifted 0.11 pixels from its correct location. However, the receiver bandwidth along phase encoding direction is narrow, usually 40 Hz/pixel if 64 lines are acquired. With such a low receiver bandwidth, the fat and water component in the same location will be shifted from each other by about 11 voxels. Therefore chemical shift artifacts are typically seen along the phase encoding direction in EPI images. To suppress this artifact, a simple method is to saturate the signal from the unwanted component and only image the component of interest, which is usually water in neuroimaging. Alternatively one can either use spatial-spectral selective RF excitation pulses to excite only water spins,[16] or apply the Dixon technique to decompose the water and fat images.[17,18]

3.2. *Geometric distortion*

Apart from the chemical shift effect, the heterogeneous nature of EPI sequence makes it vulnerable to various off-resonance effects along the phase-encoding direction. Unlike the fixed spatial shift in chemical-shift artifacts, the displacement artifacts caused by field inhomogeneity and susceptibility vary across the image, leading to highly nonlinear image deformations.[19] In brain imaging, macroscopic susceptibility-induced local field inhomogeneity usually exists around petrous bone and the sinuses, resulting in image distortions in the orbito-frontal and temporal lobes.

In a slice-selective imaging experiment, the induced MR signal $S(t)$, obtained from an object $I(x, y)$, can be described as:

$$S(t) = \iint I(x, y) e^{-i2\pi[k_x(t)x + k_y(t)y]} e^{-t/T_2^*} \, dx \, dy, \tag{10}$$

where T_2^* is the effective transverse relaxation time. During the EPI scans, $S(t)$ is discretely sampled at times $t = t_0 + nT_{sy} + mT_{sx}$ $(0 \leq m \leq N-1, 0 \leq n \leq N-1)$, where t_0 is the time interval between the RF excitation pulse and the beginning of the EPI encoding gradient, T_{sx} is the interval between two sampling points along the readout direction, also referred to as dwell time, T_{sy} is the inter-echo time between two adjacent lines, and the image matrix size is $N \times N$. In the absence of field inhomogeneity, the phase evolution of spins within a voxel, or the k-space position (k_x, k_y) at time t, is determined by the readout gradient G_x and the phase-encoding blip gradient ΔG_y, i.e. $k_x = m\gamma G_x T_{sx} = m\Delta k_x$ and $k_y = n\gamma \Delta G_y \tau = n\Delta k_y$, where τ is the duration during which the blip gradient is applied. Assuming a local magnetic field offset ΔB_{sus} at location (x, y), the signal in Eq. (10) becomes

$$\begin{aligned} S_{mn} &= \iint I(x,y) e^{i(m\Delta k_x x + n\Delta k_y y)} e^{i\gamma \Delta B_{sus}(x,y)(mT_{sx}+nT_{sy})} e^{-t/T_2^*}\, dx\, dy \\ &= \iint I(x,y) e^{im\Delta k_x \left(x + \frac{\gamma \Delta B_{sus}(x,y) T_{sx}}{\Delta k_x}\right)} e^{in\Delta k_y \left(y + \frac{\gamma \Delta B_{sus}(x,y) T_{sy}}{\Delta k_y}\right)} e^{-t/T_2^*}\, dx\, dy. \end{aligned} \tag{11}$$

It can be seen from Eq. (11) that the extra phase caused by ΔB_{sus} will erroneously map the spins' actual position (x, y) to a pair of new spatial coordinates $\left(x + \frac{\gamma \Delta B_{sus}(x,y) T_{sx}}{\Delta k_x}, y + \frac{\gamma \Delta B_{sus}(x,y) T_{sy}}{\Delta k_y}\right)$. Usually Δk_x equals to Δk_y in most imaging experiments. Therefore under same field inhomogeneity, the ratio of pixel shift along phase-encoding direction over shift along readout direction is determined by T_{sy}/T_{sx}. Since the phase-encoding dwell time T_{sy} is much longer than the readout dwell time T_{sx}, the image distortion consists of pixel shifts primarily along the phase-encoding direction. Taking typical parameters of $T_{sx} = 4\,\mu\text{s}$, $T_{sy} = 0.4\,\text{ms}$, and a 64×64 imaging matrix over a $24 \times 24\,\text{cm}^2$ FOV, a field inhomogeneity of $\frac{\gamma}{2\pi}\Delta B_{sus}(x,y) = 16\,\text{Hz}$ will lead to a pixel shift of 0.004 pixels along the readout direction, and a shift of 0.4 pixels along the phase-encoding direction, with a ratio determined by $T_{sy}/T_{sx} = 100$. In areas of the brain with stronger field inhomogeneity, the pixel shift can be much larger.

In functional brain mapping, it is desirable to superimpose the brain activation maps onto the high-resolution structural images that are acquired using a conventional non-EPI sequence and have negligible geometric distortion. The distortion in EPI images makes it difficult to accurately coregister the functional and structural data sets. To correct the geometric distortion, a commonly used method is to separately acquire a set of complex gradient-echo images[20–22] with different echo times (TE_1 and TE_2), through which the information on field inhomogeneity across the object can be obtained:

$$\Delta B_{sus}(x,y) = \frac{\phi_{TE1}(x,y) - \phi_{TE2}(x,y)}{\gamma\, \Delta\text{TE}}, \tag{12}$$

where ϕ is the phase information calculated from the gradient-echo image data sets. This field map is then used to correct the pixel shift along the phase-encoding direction. An example of distortion correction based on field maps measurement is shown in Fig. 6. It can be observed that the corrected EPI image in Fig. 6(c) shows

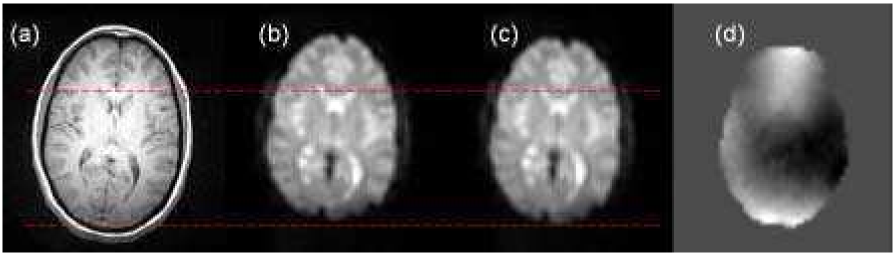


Fig. 6. Geometric distortion and correction. (a) The 256 × 256 anatomical image. (b) The uncorrected 64 × 64 EPI image. (c) The corrected 64 × 64 EPI image. (d) The field map calculated from separately acquired complex image data sets.

better correspondence to the high-resolution anatomical image than the uncorrected image in Fig. 6(b), as indicated by the two dashed horizontal lines.

Besides the geometric correction algorithms based on a field map, there are other correction approaches, e.g. the point spread function method[23] and the multiple reference scan method,[24,25] proposed to circumvent the technical difficulty of phase unwrapping in the field mapping methods.

3.3. *Signal loss*

The presence of field inhomogeneity, regardless of the source, alters the intended k-space trajectory defined by the spatial-encoding gradients and will induce not only geometric distortion, but also signal dropout in the reconstructed images, especially at high magnetic fields.[26] In the neighborhood of bone-tissue and air-tissue boundary where macroscopic susceptibility gradients exist, spins located at different positions within a voxel will experience a different magnetic field, and thus precess at different frequencies and become out of phase over time.

The effects of such dephasing on image quality depend on several factors, such as the particular sequence, data acquisition speed, echo time, voxel size, slice orientation, and phase-encoding direction in the case of EPI.[27] Slower data sampling rate and longer echo time will give spins more opportunity for intravoxel dephasing and cause larger signal loss. Larger voxel size means larger variations in the spins' precession frequency across the voxel, leading to more signal cancellation from the spins and more signal loss. Since slice thickness is usually larger than the in-plane resolution, signal dropout along the through-plane direction (assuming z-direction) is more severe, and much effort has been made to compensate for dephasing along this direction.

The most straightforward method to alleviate the through-plane signal loss along slice select direction is to reduce the slice thickness,[28] which, however, will reduce SNR and sacrifice spatial coverage or increase the total acquisition time. The tailored Radio Frequency (RF) pulse technique is more complicated, which designs RF pulses with specific phase profiles to compensate for the intravoxel

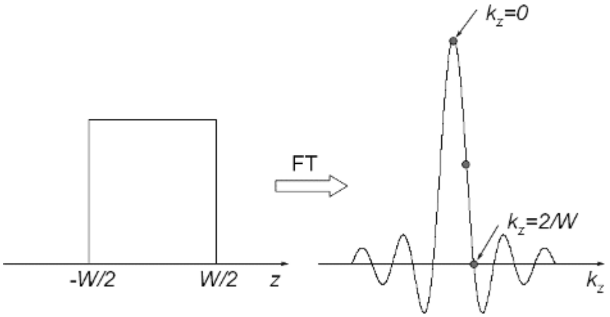


Fig. 7. Signal profile in the presence of a susceptibility gradient along slice-selective direction.

dephasing.[29–31] This technique is useful for recovering the signal loss in the susceptibility affected areas, but in general has the drawback of reducing SNR elsewhere. Another method is called the z-shim technique.[26,32–37] Assume a gradient G_{susz} is present along the through-plane direction (z-direction). The signal magnitude at the time of acquisition t will be a sinc function of G_{susz}, as shown in Fig. 7, where $k_z = (\gamma/2\pi)G_{susz}t$. In the ideal situation ($G_{susz} = 0$), except for the imaging gradients applied along frequency- and phase-encoding direction, there should not be any gradients influencing the spins' phase evolution. Therefore, spins at the same position (x, y) but with different z coordinates will precess in synchronization after the excitation and contribute maximum signal at the time of acquisition. This is the case of $k_z = 0$ in Fig. 7. However, if a susceptibility gradient G_{susz} is present, the sampling point will be shifted away from $k_z = 0$ position and signal loss will occur after large enough dephasing has accrued. If an external z gradient with comparable strength is applied to counteract the susceptibility gradients G_{susz}, intravoxel dephasing will decrease and signal can thus be recovered. This is the fundamental principle of z-shim methods.

To account for varying susceptibility gradients in different areas, a z-shim encoding method typically acquires multiple images of the same location with different amplitudes of z refocus gradient, and then combines these images to form a composite image. Figure 8(b) illustrates a set of composite images combined from 16 z-shim encoding images. Compared with conventional EPI images in Fig. 8(a), the z-shim composite images demonstrate satisfactory signal recovery in the ventral prefrontal and lateral temporal lobes, as indicated by the arrows. The limitation of this technique is the reduced temporal resolution due to multiple acquisitions of the same imaging plane. If it takes 2 sec to cover the whole brain using a conventional EPI sequence, z-shim EPI with 16 encodings will require 32 sec with other parameters kept the same. To shorten the effective repetition time, first z-shim techniques with reduced encodings,[32,35] and then single-shot z-shim techniques[36,37] were proposed. Figure 8(c) shows the composite images combined from two z-shim encodings images, whose compensation results demonstrate significant improvement over conventional EPI images.

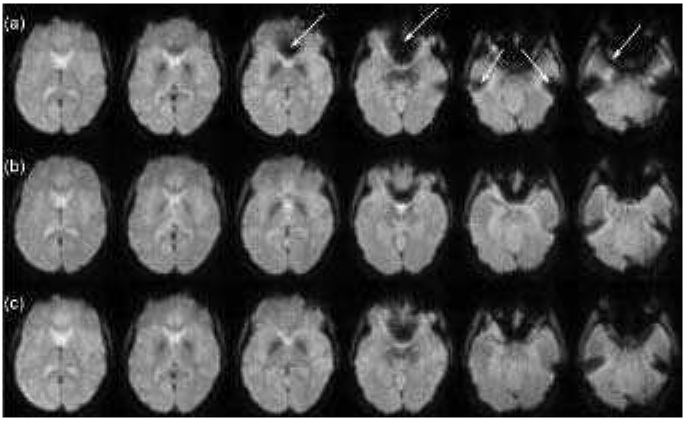


Fig. 8. (a) Conventional single-shot EPI Images; (b) 16 z-shim encoded EPI composite images; (c) 2 z-shim encoded EPI composite images.

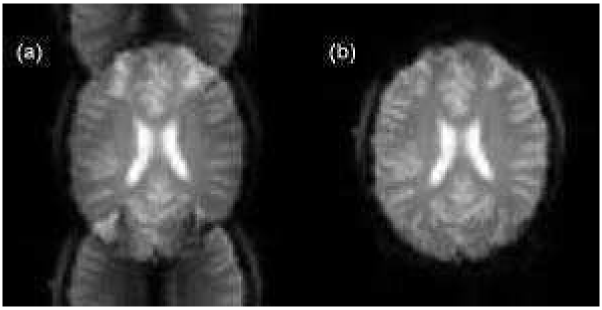


Fig. 9. (a) Demonstration of Nyquist ghost on brain image; (b) Improved image after phase correction.

3.4. *Nyquist ghosting*

Ideally, an echo will form at the refocusing point of each readout gradient waveform. However, due to inaccurate timing of the readout gradients, eddy current effects from rapid gradient switching, and off-resonance effects, the actual echo position could be delayed or advanced relative to the k-space center. In conventional imaging, since each line of k-space is traversed in the same direction, the echo shift in k-space leads to a constant phase shift in the reconstructed images, which will not appear in the magnitude images and thus causes no problems. But the back and forth k-space trajectory used in EPI is different. Before reconstruction, every other line collected in k-space must be flipped, which could produce different echo offsets between the odd and even lines, leading to a “ghost” image shifted by half of the field-of-view in the phase-encoding direction after Fourier transform as shown in Fig. 9(a). This artifact is usually referred to as a “Nyquist ghost” or “$N/2$ ghost”.

Hardware improvement in gradients and filter performance can suppress Nyquist ghost to some extent, but other post-processing methods are still needed to further reduce ghosting intensity. Several methods have been proposed to solve this problem.[38–40] The commonly used technique is to acquire a reference scan with phase-encoding gradients turned off. Such a reference scan can be either a very short separate scan or incorporated into the beginning of the imaging sequence without affecting its single-shot nature. The 0th- and 1st-order phase information about the echo offset difference between the odd and even lines can be extracted from the reference scan and then the odd/even lines can be corrected by multiplying the exponential of the conjugate phase difference in image space. Figure 9(b) illustrates the removal of ghosting artifact using this reference scan technique.

3.5. *Spatial resolution degradation*

Image blurring is another important issue in fast imaging techniques. The blurring process can be characterized by a Point Spread Function (PSF), whose Full Width at Half Maximum (FWHM) is often used to assess the spatial resolution of the imaging system. The spatial resolution of an MR imaging system is determined not only by the sampling duration, but also by the intrinsic spin-spin relaxation T_2 effect and the intravoxel inhomogeneity T_2' effect.

Assuming T_2 and T_2' are spatially invariant, the point spread function in a conventional gradient-echo sequence is given by[41]

$$PSF(x) = T_{xacq}\frac{\sin(\gamma\bar{G}_x x T_{xacq}/2)}{\gamma\bar{G}_x x T_{xacq}/2} \otimes \frac{\gamma\bar{G}_x T_2^*}{1 + i\gamma\bar{G}_x T_2^* x}, \tag{13}$$

where $\otimes$ denotes convolution, $\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'}$, T_{xacq} is the echo sampling duration, $\bar{G}_x = \frac{1}{T_{xacq}}\int_0^{T_{xacq}} G_x(t)dt$ is the mean of the trapezoidal gradient waveform $G_x(t)$ during T_{xacq}. The first term in Eq. (13) is the sampling point spread function caused by the finite acquisition time, whose full width at half maximum is $\mathrm{FWHM}_{xacq} = 1.2\pi/(\gamma\bar{G}_x T_{xacq})$; the second term, with a full width at half maximum of $\mathrm{FWHM}_{T_2^*} = 2/(\gamma\bar{G}_x T_2^*)$, reflects the T_2^* decay effect.[42] The impact of these two terms in degrading spatial resolution is determined by their relative FWHM:

$$\frac{\mathrm{FWHM}_{xacq}}{\mathrm{FWHM}_{T_2^*}} = \frac{1.2\pi T_2^*}{2T_{xacq}} \approx \frac{1.9T_2^*}{T_{xacq}}. \tag{14}$$

If the FWHM of one term is significantly larger than the other, it will dominate the resolving power. In conventional imaging, a typical receiver bandwidth of 32 kHz with 256 sampling points requires 8 ms to acquire an echo, not considering ramp sampling. Given a typical T_2^* relaxation time of 27 ms at 3T in brain tissue, FWHM_{xacq} is significantly greater than $\mathrm{FWHM}_{T_2^*}$. Therefore the effects of T_2^* in degrading spatial resolution are negligible in conventional imaging.

The receiver bandwidth along the readout direction in EPI is even higher (typically 250 kHz) than conventional imaging, hence the pixel broadening caused by

T_2^* decay is minimal in this direction. After ignoring the contribution of T_2^* effect to the blurring of the readout direction, a similar expression as Eq. (14) can be derived for the phase-encoding direction:

$$\frac{\mathrm{FWHM}_{yacq}}{\mathrm{FWHM}_{T_2^*}} \approx \frac{1.9T_2^*}{T_{yacq}}. \tag{15}$$

Assuming an inter-echo time T_{sy} of 0.4 ms, the total acquisition time T_{yacq} for 64 views will be 25.6 ms. It can be seen from Eq. (15) that the FWHM of the sampling point spread function still dominates but to a lesser degree, so there will be small spatial resolution degradation induced by T_2^* decay along the phase-encoding direction.

3.6. *Off-resonance artifacts in spiral imaging*

Unlike EPI's distinct behaviors between readout and phase-encoding directions, spiral trajectories do not have a preferable encoding orientation and hence, instead of prominent artifacts along a particular direction, the off-resonance effects (e.g. chemical shift, field inhomogeneities, susceptibilities) will give rise to radial blurring in spiral imaging.[43]

Numerous off-resonance correction methods for spiral imaging, either extended from correspondent EPI solutions or specifically developed for spiral itself, have been proposed. To avoid chemical shift artifacts, the fat signal suppression techniques used in EPI, such as spatial-spectral selective RF pulses, fat signal presaturation, and modified Dixon techniques,[44] can all be employed to spiral imaging as well. The removal of blurring artifacts introduced by magnetic field inhomogeneity and susceptibility can be achieved through image reconstruction with or without field map acquisitions.[45–48] As to the signal loss along slice-selective direction, all of the compensation techniques used in EPI can be applied in spiral imaging. An interesting method unique to spiral acquisition to alleviate in-plane signal loss is the reverse spiral scanning technique,[49–51] which has been demonstrated to have improved SNR and reduced susceptibility artifacts around interfaces between different materials.

4. Applications in Functional MRI

Functional MRI (fMRI) is the non-invasive *in vivo* measuring of neuronal activity using MRI techniques. Although there are emerging methods to directly measure changes in the MRI signal caused by the current of neurons firing,[52] the vast majority of fMRI experiments measure changes in hemodynamic parameters as an indirect measure of neuronal activity. The earliest of these experiments was reported by Bellivieu *et al.*[53] in 1991 and utilized an exogenous contrast agent using a now seldom-used technique called dynamic susceptibility contrast MRI (DSC-MRI) described in further detail in Sec. 7. It is possible to measure Cerebral Blood Flow

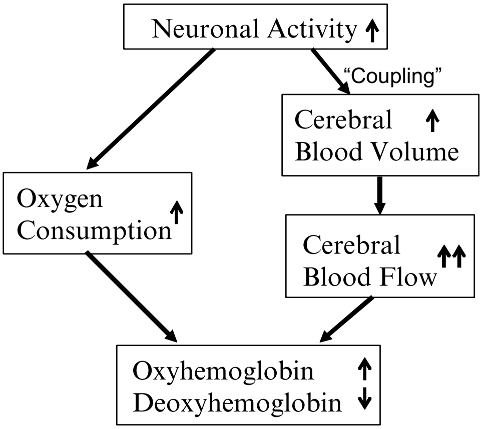


Fig. 10. Physiologic changes that give rise to the BOLD signal.

(CBF) directly using single-shot techniques, described in the next section, but the majority of fMRI experiments utilize single-shot techniques optimized for blood oxygen level dependant (BOLD) contrast.

4.1. *The BOLD effect*

The BOLD phenomenon was first described by Ogawa *et al.*[54] It arises from the fact that deoxyhemoglobin is paramagnetic whereas oxyhemoglobin is diamagnetic. An increase in the concentration of deoxyhemoglobin [deoxy-Hb] will cause a local perturbation in the magnetic field. Perturbations in the magnetic field, both macroscopic and microscopic, cause an increase in dephasing and thus a decrease in T_2^*. Physiologically, there are three causes to a change in [deoxy-Hb]: A change in the metabolic rate of oxygen extraction, CMRO_2, a change in regional cerebral blood flow, rCBF, or a change in regional cerebral blood volume, rCBV. The chain of events leading to the BOLD signal is shown in Fig. 10. Note that there is a mismatch between the change in rCBF and CMRO_2, leading to the seemingly paradoxical increase in MRI signal with increased neuronal activity. This overperfusion is sometimes described as "Watering the whole garden for the sake of one flower", and results in a *decrease* in [deoxy-Hb] with increased neuronal activity.

4.2. *The relationship between BOLD and neuronal activity*

It is natural to ask exactly how changes in BOLD relate to the underlying neuronal activity. The mechanism by which changes in neuronal activity cause changes in hemodynamic parameters, termed "coupling", is unknown, although there are several candidate mediators including nitric oxide, adenosine, changes in pH, K+ and neurogenic mechanisms.[55,56] In a non-human primate model using simultaneous BOLD and direct recording of electrical activity, Logothetis *et al.*[57] demonstrated that changes in the BOLD signal were best predicted by changes in local field potentials, indicating that BOLD most strongly reflects the input and intracortical processing of an area.

4.3. *The hemodynamic response*

The temporal dynamics of the three physiologic parameters leading to a change in [deoxy-Hb] are complicated and still poorly understood. Various models have been put forth, the most famous being the balloon model.[58,59] An empirical description is shown in Fig. 11 and is as follows: Assuming a brief burst of neuronal activity there is a delay of about 2 seconds. Following this, some experimenters see a dip in the MRI signal, attributed to an increase in [dexoy-Hb] due to an increase in oxygen extraction.[60] There is even some evidence that activation maps based upon this initial dip have greater spatial specificity to the sites of neuronal activation,[61] but the size and irreproducibility of this effect make it unsuitable for most studies. Following this dip, there is a broad increase in the MRI signal, peaking at about 4 seconds after onset and having a full width at half max of around 4.5 seconds. This is attributed to the increase in perfusion that overcompensates for the increase in oxygen extraction and removes the deoxy-Hb. The vast majority of BOLD imaging experiments make maps based upon this signal increase. Finally, there is a broad so-called "post undershoot" attributed to an increase in regional CBV due to the vasodilation and thus a greater amount of deoxy-Hb. The positive-going portion of the hemodynamic response has been modeled with several functions, the most common being a gamma-variate function.[62]

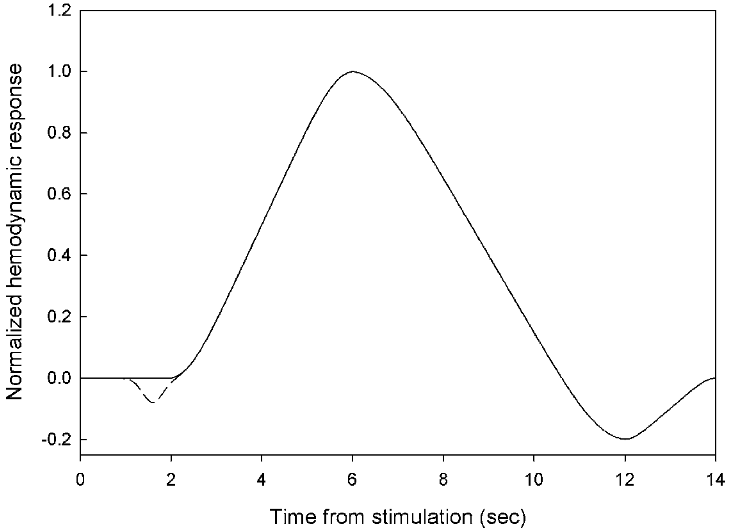


Fig. 11. Model of the hemodynamic response function, which is the BOLD response to a brief neuronal activity. The dashed portion indicates the controversial "initial dip".

There is considerable heterogeneity, both inter- and intrasubject, in the onset time and shape of the hemodynamic response.[63] Critically, the variations in the onset are most likely due to variations in local vasculature and not asynchronies in neuronal firing. By and large, with a few notable exceptions[64,65] this makes BOLD unsuitable for determining the chronology of neuronal events. To determine the temporal order, there has been considerable advancement in the simultaneous measuring of BOLD and EEG,[66] the former providing the spatial specificity and the latter providing temporal information.

4.4. *Experimental design using BOLD*

The hemodynamic response shown in Fig. 11 is for a brief burst of neuronal activity. If one assumes the activity is infinitely brief, and infinitely large, then the hemodynamic response function shown takes on the role of an impulse response function and basic signal processing theory states that the response to any activation will be the convolution of that activation with the hemodynamic response function. For this to be true, two conditions must be met: Linearity (the response to the sum of two events is equal to the sum of the responses to each event) and time invariance (a given event will always generate the same response, no matter when it happens). Unfortunately, either/both of these are violated in certain circumstances. If two neuronal events are too close together (about $< 2\,\text{sec}$) the response to the second is reduced.[67] Further, there is evidence that the size of the responses varies during an experiment.[68] Despite these caveats, most experiments are designed and analyzed with the assumption that the convolution is valid.

The duration of the hemodynamic response gives rise to two domains of experimental design. The first is when the duration of neuronal activity is long compared to the hemodynamic response. These designs include the earliest fMRI experiments and are commonly referred to as "block-designs". These experiments consist of blocks when the subject is continuously performing some task alternating with blocks when the subject is performing a control task and/or blocks when the subject is resting. Analysis can be as simple as subtracting the average of the task images from the average of the control/rest images. The other regime of experimental design is when the duration of neuronal activity is short compared to the hemodynamic response such that each event is expected to produce a single hemodynamic response — commonly referred to as "event-related" designs. In initial event-related experiments the events were separated far enough in time that they could be averaged in a time-locked fashion. Currently, events are overlapped with random[69] or pseudorandom[70] inter-stimulus intervals for efficiency. Event-related designs have the advantage that individual events can be discarded, for example incorrect responses, the stimuli can be less predictable, and the hemodynamic response can be determined; block designs are typically more sensitive.[71] An ideal experiment combines the two.[71]

Analysis of fMRI data is at its most fundamental level the analysis of time course data. Single-shot MRI techniques permit the acquisition of whole brain with a temporal resolution around 2 seconds and with a spatial resolution of around $3 \times 3 \times\ 3\,\text{mm}$. Typically a design matrix is formed by convolving the expected neuronal response (based upon the experimental design) with a set of basis functions (for example, a gamma-variate function and its derivative to help account for variations in onset) and the data in each voxel is modeled under a general linear model.[72] A second level of analysis is then performed, also under a general linear model across subjects/treatments. Statistical parameter maps are then formed from this second level analysis. More advanced multivariate approaches, which analyze both the spatial and temporal data simultaneously are starting to be more common. These include partial least squares,[73] structural equation modeling/path analysis[74] and data driven approaches that look for patterns in the data without any model such as independent component analysis,[75] canonical correlation[76] and fuzzy clustering.[77]

4.5. *Application of BOLD imaging*

In the just over one decade of its existence, fMRI using single-shot techniques has undergone explosive growth. There are books and journals dedicated to its application; see, for example Refs. 78 and 79. It would be impossible to review such a large literature here — instead we will present some examples from various applications.

The earliest fMRI experiments were simple sensory or motor tasks.[80–82] These tasks consist of activating the primary visual system via flashing lights or changing visual patterns, or activating the motor system via timed hand movements. Over a range of frequencies, graded neuronal activation will occur by varying the frequency in either of these paradigms.[83,84] This, their simplicity, and the robustness of their activation have made sensory/motor paradigms the *de facto* standard for the testing of new pulse sequences, analysis techniques and other basic advances in fMRI. Fast imaging techniques have permitted retnotopic[85] and tonotopic[86] mapping of human cortex. Figure 12 gives an example of a simple motor task and a typical time course for these types of experiments.

Arguably the largest effort in fMRI has gone toward the understanding of human cognition, here broadly defined to include learning and memory and attention. Whereas much was already known about sensory/motor systems through the preclinical and human lesion literatures, with neuroimaging performing a mostly confirmatory role, the non-invasive nature of fMRI allows the study of what makes humans unique, *in vivo,* and in healthy individuals. Some of the work here has utilized standard neuropsychological tests, or slight variations thereof, to probe various cognitive systems. Excellent examples of this are the Stroop task,[87] the continuous performance task[88] and the Wisconsin card sorting task.[89] In addition, a plethora of new tasks have been created to probe systems/constructs such as the storage and retrieval of long-term memory,[90] attention,[91] decision-making,[92] language,[93] and executive functions such as working memory,[94] set switching,[95] inhibition.[96]

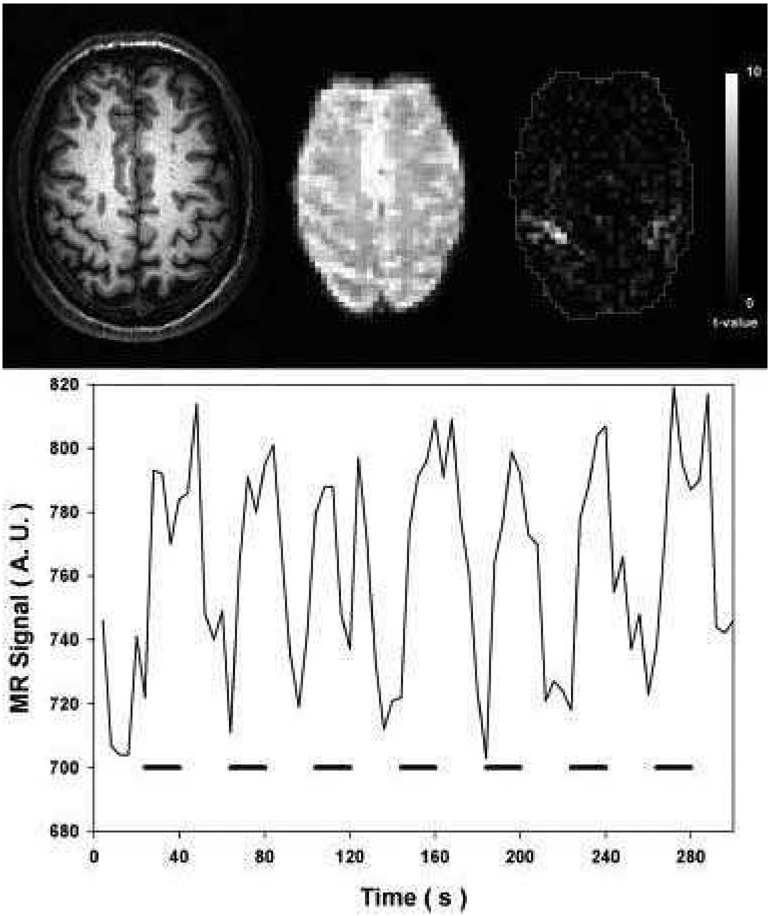


Fig. 12. Example of fMRI activation, in this case bilateral finger-tapping in a block design. The image on the left is the high-resolution T_1 anatomical image, in the middle a single-shot EPI image of the same slice, on the right a map of t-values indicating the correspondence between the fMRI data and the experimental design. Note the highest activation is in primary motor cortex. The bottom figure is the time-course of a typical activated voxel, with the black bars indicating when the subject was tapping their fingers.

Functional neuroimaging has also made contributions to the affective literature. It has been long known from the preclinical literature that the amygdala is involved in fear conditioning, confirmed in humans using neuroimaging.[97] But there is evidence that it may be arousal, irrespective of the valence of the stimulus, that activates this key brain region.[98] Fast imaging techniques have also permitted the investigation of drive states such as hunger[99] and thirst,[100] satiation,[101] sexual arousal,[102] and drug craving.[103]

By combining the fast imaging of single-shot techniques with pharmacological challenges an offshoot neuroimaging technique has been created, abbreviated phMRI. Although it lacks the ability to look at specific transmitter/receptor systems compared to PET, phMRI has the advantage of high temporal and spatial resolution. The high temporal resolution permits the characterization of the response based upon known pharmacokinetic parameters.[104] Using these methods, the neuronal sites of action of cocaine[105] and nicotine[106] have been mapped. A difficulty with these methods is that the BOLD signal has a $1/f$ noise characteristic,[107] and thus is subject to slow drifts. It can be difficult to separate these drifts from signal induced by the pharmacological manipulations. A second approach is to measure how a drug affects a second process. Two examples of this are nicotine's effect on attentional processes[108] and the effect of remifentanil (an analgesic) on painful thermal stimuli.[109]

While the previous examples mostly point to the utility of fMRI in the investigation of normal neural functioning, it is the application to clinical medicine and disease states that will ultimately be neuroimaging's greatest benefit. Examples here include the investigation of the neural underpinnings and consequences of mental disorders such as attention deficit/hyperactivity disorder,[110] depression,[111] obsessive-compulsive disorder[112] and schizophrenia.[89] There is evidence that BOLD imaging may someday be used in the diagnosis of Alzheimer's disease.[113] Finally, BOLD imaging is being used as a preoperative tool to aid in the planning of neurosurgery. Two examples here are using BOLD imaging to locate language areas prior to surgery to help relieve epileptic seizures[114] and to locate motor regions prior to tumor extraction.[115]

5. Applications in Perfusion Imaging with Arterial Spin Labeling

Cerebral Blood Flow (CBF) is the amount of arterial blood delivered to a local volume of brain tissue per unit time. It is usually quantified in milliliters of blood per gram of tissue per second (ml/g/sec). Cerebral perfusion determines the effectiveness of the blood circulation to provide oxygen and nutrients to the brain tissue, and to remove waste products from the tissue. Measurements of perfusion in brain and other organs provide information about tissue viability and function, and are therefore of fundamental significance in medical research and clinical diagnostics. It is of particular interest that local perfusion changes in the brain reflect regional neuronal activity and metabolism, and thus CBF can be used as an index for mapping functional neuroanatomy.[116]

Perfusion MR imaging with Arterial Spin-Labeling (ASL) is a non-invasive technique developed in the past decade that can be used to map blood flow in the brain, heart, kidney and other organs.[7] Single-shot fast imaging techniques have been successfully used for ASL perfusion measurements, due to their high SNR, high temporal resolution and insensitivity to motion artifacts.

5.1. *Principles of ASL perfusion imaging*

Arterial spin-labeling perfusion imaging utilizes magnetically labeled water in arteries as an endogenous tracer to obtain quantitative blood flow information in the

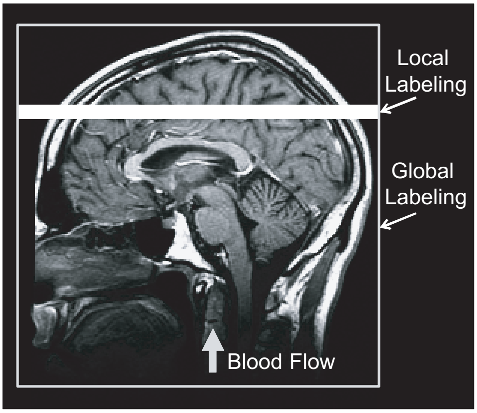


Fig. 13. Illustration of spin labeling in FAIR perfusion imaging. The image acquired with local labeling (in the imaging slice) is compared to the one with global labeling (entire brain) to generate perfusion contrast.

tissue.[7] The labeling is usually accomplished by inverting or saturating the magnetization of the inflowing arterial blood with respect to the tissue of interest. Several strategies for spin labeling have been proposed,[117–121] but here we will focus on discussing one of the most popular techniques called Flow-sensitive Alternating Inversion Recovery (FAIR).[118] In FAIR imaging, as illustrated in Fig. 13, perfusion contrast is obtained by comparing images with local and global spin labeling. Typically, a time series of inversion recovery images are acquired, with alternative slice-selective inversion (local labeling) or a non-selective inversion (global labeling).[118] The difference image between the local and global labeling is equivalent to labeling only the spins outside the imaging slices, and thus reflects the signal change associated with perfusion.

To quantify blood flow, a perfusion model describing the blood/tissue water exchange kinetics and the magnetization characteristics is necessary. A perfusion model was proposed incorporating arterial transit and trailing times for improved quantification of CBF.[121] In the model, the amplitude of the difference signal at time t obtained from the locally and globally labeled images, $\Delta M(t)$, is described as[121]

$$\frac{\Delta M(t)}{M_0} = -\,(2Q/\lambda)\int_0^t \alpha_{cap}(\xi)e^{-R_1(t-\xi)}d\xi, \tag{16}$$

where M_0 is the amplitude of the fully relaxed signal, Q is the cerebral blood flow, λ is the partition coefficient of brain water, R_1 is the longitudinal relaxation rate of brain water, and $\alpha_{cap}(\xi)$ is the extent of arterial labeling at the capillary exchange site in the brain. The labeling is due to the difference in preparation for spins outside the imaging slices. We assume that

$$\alpha_{cap}(t) = h(t)e^{-R_{1a}t}, \tag{17}$$

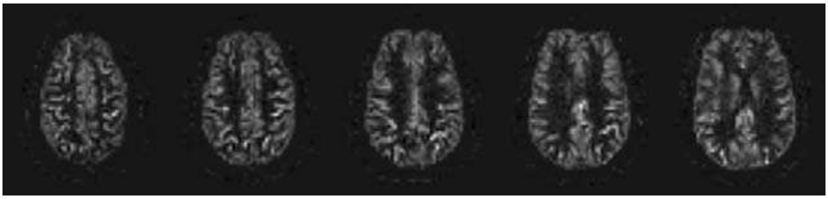


Fig. 14. Difference images between local and global labeling.

where $h(t)$ is defined as

$$h(t) = \begin{cases} 0, & \text{for } 0 \leq t < \tau_a \\ 1, & \text{for } \tau_a \leq t < \tau_d, \\ 0, & \text{for } t \geq \tau_d \end{cases} \tag{18}$$

where τ_a and τ_d are arterial transit and trailing times respectively, and R_{1a} is the longitudinal relaxation rate of arterial blood. Inserting Eq. (17) into Eq. (16) gives

$$\frac{\Delta M(t)}{M_0} = 0, \quad \text{for } 0 < t \leq \tau_a, \tag{19}$$

$$\frac{\Delta M(t)}{M_0} = -2Q/(\lambda\,\Delta R)e^{-R_{1a}t}[1 - e^{-\Delta R(t-\tau_a)}], \quad \text{for } \tau_a \leq t < \tau_d, \tag{20}$$

$$\frac{\Delta M(t)}{M_0} = -2Q/(\lambda\,\Delta R)e^{-R_{1a}\tau_d}[1 - e^{-\Delta R(\tau_d-\tau_a)}]e^{-R_1(t-\tau_d)}, \quad \text{for } t \geq \tau_d, \tag{21}$$

where $\Delta R = R_{1a} - R_1$. The τ_a and τ_d can be measured from an ASL experiment with a series of inversion times, and R_1 can be measured using inversion recovery experiments. Tissue perfusion (Q) can then be calculated from Eqs. (19)–(21) if τ_a, τ_d, R_1 and R_{1a} are known. An example of the difference images ΔM on a normal subject is shown in Fig. 14.

5.2. *Brain activation measurement using ASL imaging*

In the past decade, ASL perfusion imaging techniques have been successfully used for cerebral activation studies.[117–121] The mechanism for detecting brain activation using ASL techniques is that increased brain activity is accompanied by local changes of cerebral blood flow, as demonstrated by PET studies.[116] In perfusion-based brain activation studies, as in BOLD functional imaging described previously, a time series of ASL images is collected during which the subject is asked to perform specific tasks. Brain activation areas can be detected by comparing images acquired at different states (e.g. performing task versus rest). Such studies have been carried out with visual and sensorimotor activation paradigms,[117–121] as well as cognitive paradigms.[122] The increase in CBF during brain activation varied from 30% to 90% depending on the activation paradigms, and was generally consistent with the blood flow change observed in PET studies.[123] Compared to BOLD imaging, which is relatively more venous-weighted, ASL perfusion techniques target signals closer to the capillaries and tissue, and are therefore more closely related to neuronal activity. Another advantage of ASL techniques, comparing to BOLD, is their capability to obtain a quantitative and physiologically more meaningful index for brain activation, which may provide for more precise interpretation of the data.

Whereas ASL perfusion techniques are attractive for functional brain imaging studies, there are a few disadvantages for the techniques that one should be aware of. *First*, sensitivity of the ASL techniques is inherently low, due to the small perfusion-related signal difference between control and label ($< 1\%$) and the requirement of image subtraction to obtain blood flow information. In general, perfusion Contrast-to-Noise Ratio (CNR) is lower than BOLD CNR in brain activation studies, although the CNR difference of the two techniques depends on a number of factors including perfusion/BOLD methodologies, field strength, and image acquisition parameters. *Second*, temporal resolution is usually low in most of ASL techniques, because the perfusion image is obtained from a subtraction of two raw images, resulting in a temporal resolution half of that of the raw images. In addition, inversion time (TI) is required for labeled (inverted) spins to recover, leading to a typical temporal resolution of 4–8 s for ASL imaging. This poor temporal resolution makes an event-related functional study[124] difficult to perform. Recently, several pulsed ASL techniques with improved temporal resolution have been suggested,[125–127] and significant improvements of temporal resolution (0.5–1.0 s) have been achieved for ASL perfusion imaging. *Third*, ASL techniques usually cover less brain volume than BOLD imaging. Typically, ASL imaging is limited to acquisition of a few slices per labeling, due to potential artifacts and technical difficulties. Single-shot fast imaging techniques are particularly important for developing multi-slice or 3-dimensional perfusion techniques, covering a larger brain volume. For example, the utilization of efficient image acquisition of fast spiral scanning allows 10 or more slices to be acquired per labeling,[121] greatly facilitating brain functional studies with multiple activation areas.

Simultaneous measurement of ASL perfusion and BOLD signals in a single scan session has been proposed in recent years.[128–131] In these methods, ASL signal is obtained by labeling the in-flowing spins in arteries, whereas BOLD contrast is achieved by effective transverse relaxation (T_2^*) weighting. From a time series of images acquired with alternating control and labeling, ASL and BOLD signals can be obtained by subtracting or adding the control and labeled images, respectively, in the same data set. The capability of simultaneous detection of ASL and BOLD signals has a number of advantages, including efficient acquisition of two functional images (improvement of CNR per unit time) and minimization of temporal and spatial variations between the two types of images. These advantages may be particularly important for functional brain studies involving transient and dynamic signal changes such as those with drug-induced brain activation.[105,106] Simultaneous image acquisition is also useful for more accurate determination of Cerebral Metabolic Rate of Oxygen consumption ($\mathrm{CMRO_2}$) from a combination of perfusion and BOLD techniques.[134–136]

6. Applications in Diffusion Magnetic Resonance Imaging

"Diffusion" generally refers to the random thermal motion of small particles, e.g. water molecules, in a given medium. The strength of a diffusion process can be quantified by the Diffusion Coefficient (DC) as the mean squared diffusion displacements of an ensemble of molecules within a unit duration of time. The value of the DC characterizes the molecular diffusive mobility that is associated with the temperature and the compartment restriction of the medium. For ideal free diffusion, the DC has a uniform value in all direction, which is called *isotropic* diffusion. However, a diffusion process in biological tissue, e.g. brain white matter, could be *anisotropic* because the diffusion process may experience different restrictions in different directions due to the orientation-dependent structure of the tissue. Quantitative assessment of diffusion thus provides an indirect measurement of the complex structure of biological tissue on a cellular scale that could be well beyond the imaging resolution itself.

Diffusion MRI (dMRI) is a technique to non-invasively measure the diffusion process by incorporating diffusion-sensitive gradients into an MRI sequence. The concept of diffusion measurement by NMR dates back to year before the birth of MRI, but vulnerability to noise and motion artifacts of the techniques had hindered its wide application.[137] It was the single-shot fast imaging techniques, especially EPI, that practically enabled dMRI to acquire diffusion images of a living biological organization with clinically acceptable quality and acquisition time. Recent methodological developments focused on investigating anisotropic properties of biological tissue based on dMRI acquired at a number of directions. Diffusion Tensor Imaging (DTI) and High Angular Resolution Diffusion (HARD) analysis models have been developed to identify the orientation and more complex structures of the white matter.[138,139] These methods have shown promising ability to virtually delineate the anatomical connectivity of a neural system *in vivo*.[140] Due to the increasing demands for rapid and robust acquisition, single-shot fast imaging techniques are playing more important roles in dMRI applications.

6.1. *Principles of diffusion MRI*

For a diffusion process in three-dimensional (3D) space, the dependence of the net molecular displacements of an ensemble of spins on the diffusion coefficient is described by the Einstein relationship[141]

$$\left\langle \mathbf{r}^2 \right\rangle = 6D\tau_D, \tag{22}$$

where $\langle \mathbf{r}^2 \rangle$ is the mean squared displacement, τ_D is the diffusion time, and D is the diffusion coefficient that is generally linearly related to the temperature of the medium.[142] The conditional probability distribution of the net displacement is given by a 3D Gaussian function

$$P(\mathbf{r}, \tau_D) = \frac{1}{\sqrt{(4\pi D\tau_D)^3}} \exp\left(\frac{-\mathbf{r}^2}{4D\tau_D}\right). \tag{23}$$

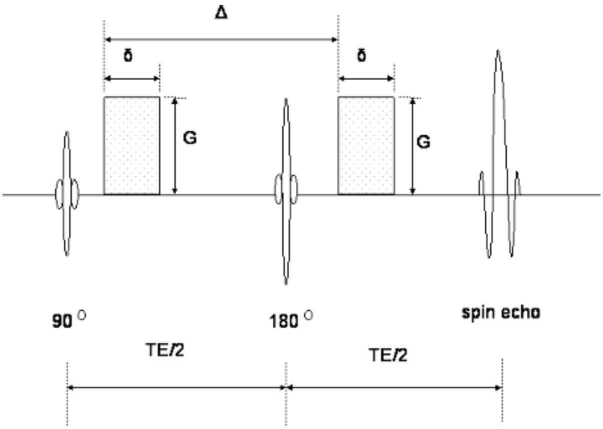


Fig. 15. Schematic diagram of the Stejskal–Tanner pulsed gradient experiment.

Essential principles of diffusion MRI can be illustrated by the Stejskal and Tanner pulsed gradient experiment,[143] which is designed to measure the spin echo attenuation caused by the diffusive motion of the spins during the presence of a pair of diffusion-sensitive gradients.

As illustrated in Fig. 15, identical magnetic field gradients (shaded areas) with the same duration δ and magnitude G are placed on both sides of the 180-degree RF pulse in a typical spin-echo MR experiment. Any spin motion that occurs during a pulsed gradient generates a phase shift that is proportional to the product of δ and the displacement in the direction of the applied gradient. Therefore, static spins would produce a pair of phase shifts with the same magnitude but opposite signs during the two separated gradients, generating no net phase change at the echo time. In contrast, diffusive spin motion with random displacements over the time would produce a random net phase shift with the same Probability Distribution Function (PDF) as the diffusive displacements. Due to the ensemble effects of a large number of spins, an attenuation of the spin-echo magnitude would occur because of the phase dispersion (dephase) caused by the pair of gradients. Based on the Gaussian PDF of the displacements as in Eq. (23), the spin-echo attenuation in the Stejskal–Tanner experiment can be explicitly written as[143]

$$\ln\left(\frac{S(G,\Delta,\delta)}{S(0)}\right) = -\gamma^2 G^2 \Delta^2 \left(\Delta - \frac{\delta}{3}\right) D, \tag{24}$$

where $S(G,\Delta,\delta)$ is the spin-echo magnitude obtained from the experiment with the 2 diffusion-weighting gradients separated by interval Δ, whereas $S(0)$ denotes the echo magnitude in a reference experiment without the diffusion weighting. For convenience, we generally use the so-called "b factor" to characterize the strength of the diffusion-weighting gradients, such that

$$b = \gamma^2 G^2 \Delta^2 \left(\Delta - \frac{\delta}{3}\right). \tag{25}$$

Thus, the echo attenuation $S(b)/S(0)$ has a negative exponential relationship with the b factor and the diffusion coefficient, i.e.

$$S(b) = S(0)\exp(-bD). \tag{26}$$

Given the b factor and the diffusion and non-diffusion weighted spin echo signal intensities measured in the experiments, the diffusion coefficient D of the spin ensemble can then be estimated by

$$D = -\ln(S(b)/S(0))/b. \tag{27}$$

It should be noted that regular imaging gradients of an MRI experiment also produce a small diffusion-weighting effect. In general, the b factor of an arbitrary MRI gradient sequence is defined by

$$b = \int_0^{TE} \mathbf{k}(t') \cdot \mathbf{k}(t')dt', \tag{28}$$

where $\mathbf{k}(t)$ represents the vector of the image acquisition trajectory in k-space. In general, the b factor of a regular spin echo EPI sequence for clinical brain imaging would be less than 50 (s/mm^2), whereas a typical diffusion-weighting EPI sequence would have a b factor of 1000 (s/mm^2) or more. Therefore, the diffusion-weighting effect of regular imaging gradients is negligible and will not be discussed further.

Diffusion MRI can be implemented by a modified single-shot fast imaging sequence incorporating the diffusion-weighting gradients, so that the Stejskal–Tanner experiment can be performed voxel-wise. However, water diffusion in biological tissue with complex microscopic structures may only be approximated by the free diffusion model as described above. Because free diffusion is an imperfect model of real diffusive processes, diffusion strength measured in biological systems is generally termed "apparent diffusion coefficient" or ADC.

A typical value of the ADC measured in room-temperature free water is $(2.30 \pm 0.02) \times 10^{-3}$ (mm^2/s). The ADC measured in the Grey Matter (GM) of a human brain is about $(0.76 \pm 0.03) \times 10^{-3}$ (mm^2/s). Due to its anisotropic structure, White Matter (WM) exhibits different diffusion strengths in directions orthogonal and parallel to the fiber orientation, with ADC values of $(0.45 \pm 0.03) \times 10^{-3}$ and $(0.95 \pm 0.03) \times 10^{-3}$ (mm^2/s) respectively. The decreased ADC value of water in tissue with respect to that of free water provides an indication of the confined environment within biological systems. These ADC differences are generally owing to restricted/hindered diffusion, or multiple compartments in the tissue.[144]

6.2. *Diffusion tensor imaging*

The quantitative characterization of the anisotropy of the diffusion systems has been one of the primary objectives of diffusion MRI. In early studies, the diffusion-sensitive gradients were usually applied separately along the x-, y- and z-axis, to acquire the Diffusion-Weighted Images (DWI) along the three directions, respectively. The diffusion anisotropy can be *roughly* inferred from the intensity differences among these DWI images. However, an obvious disadvantage is that the result is rotation-variable, i.e. it depends on the position/orientation of the subject with respect to diffusion gradients.

To overcome the difficulties of regular DWI, the Diffusion Tensor Imaging (DTI) technique has been established to offer a rotation invariant model for diffusion anisotropy.[145] In the formulation of DTI, the diffusion coefficient is no longer characterized by a scalar parameter D, but rather by a 3×3 tensor $\mathbf{D}$:

$$\mathbf{D} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}, \tag{29}$$

where $D_{ij}(i,j = x,y,z)$ denotes the cross-correlation of the diffusion coefficient between the i and j axis, and thus $\mathbf{D}$ is always symmetric, i.e. $D_{ij} = D_{ji}$. Similarly, another 3×3 matrix $\mathbf{b}$ is employed in DTI with each element b_{ij} representing the b-factor corresponding to the element D_{ij} in $\mathbf{D}$. Thus, the diffusion-weighted echo attenuation described in Eq. (26) can be expressed in the DTI formulation as

$$S(b) = S(0)\exp\left(-\sum_{i=x,y,z}\sum_{j=x,y,z} b_{ij}D_{ij}\right). \tag{30}$$

Considering the symmetry of $\mathbf{D}$ and $\mathbf{b}$, we have

$$\ln(S(b)/S(0)) = -b_{xx}D_{xx} - b_{yy}D_{yy} - b_{zz}D_{zz} - 2b_{xy}D_{xy} - 2b_{xz}D_{xz} - 2b_{yz}D_{yz}. \tag{31}$$

To determine the six independent elements in $\mathbf{D}$, we need to perform the diffusion-weighted imaging experiment at least six times with the $\mathbf{b}$ matrices independent from each other. An additional experiment is also needed to provide a non-diffusion weighted reference image. Therefore, a total of at least seven MRI images are required to fully determine the diffusion tensor $\mathbf{D}$ based on the linear relationship stated in Eq. (31). A typical set of the gradients for DTI are given by following six vectors with unit magnitude:

$$\begin{aligned} &\left[1/\sqrt{2}, 0, 1/\sqrt{2}\right]; \quad \left[0, 1/\sqrt{2}, 1/\sqrt{2}\right]; \quad \left[1/\sqrt{2}, 1/\sqrt{2}, 0\right]; \\ &\left[-1/\sqrt{2}, 0, 1/\sqrt{2}\right]; \quad \left[0, 1/\sqrt{2}, -1/\sqrt{2}\right]; \quad \left[-1/\sqrt{2}, 1/\sqrt{2}, 0\right], \end{aligned} \tag{32}$$

where the three elements in each vector correspond to the magnitudes of the gradient components applied on the x-, y-, and z-axis, respectively. These six directions are linearly independent from each other. When more than six independent diffusion-encoding directions are used, the diffusion tensor can still be fully estimated by solving a set of over-determined equations. In fact, several studies have indicated that using more directions for diffusion encoding generally helps to improve the accuracy and/or the efficiency of the DTI technique, if these directions are appropriately optimized.[146]

6.3. *Quantification and visualization of DTI*

An essential procedure in DTI is to quantify and visualize useful diffusion information. Many analysis methods are based on the eigen-analysis theorem:[147] A real, symmetric and positive-defined matrix (tensor) can be decomposed into the product of three matrices, such that

$$\begin{pmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_2 & \mathbf{V}_3 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_2 & \mathbf{V}_3 \end{pmatrix}^{\mathrm{T}}, \tag{33}$$

where the three diagonal elements λ_1, λ_2 and λ_3 in the middle matrix on the right side are defined as the three eigenvalues of the diffusion tensor. The first and the third matrixes are orthogonal and transposes of each other, and the three vectors $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ are called the eigenvectors of the diffusion tensor associated with the eigenvalues λ_1, λ_2 and λ_3, respectively.

The eigen-analysis can be understood as a geometric interpretation for the diffusion tensor: Each tensor corresponds to a 3D ellipsoid with three axes defined by the eigenvalues and the eigenvectors in terms of their lengths and orientations, respectively. Assume $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$; the primary eigenvector $\mathbf{V}_1$ is associated with the primary eigenvalue λ_1 and corresponds to the longest axis of the diffusion ellipsoid, which reflects the dominant direction of the diffusion process.

Several eigen-system based invariant *indexes* have been widely used to visualize the diffusion tensor map in various DTI applications. These include[146] the Mean Diffusivity (MD) for visualizing the average diffusion strength in all directions:

$$\mathrm{MD} = \langle \mathbf{D} \rangle = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3}. \tag{34}$$

The relative anisotropy (RA) for visualizing the degree of anisotropy of the diffusion tensor:

$$\mathrm{RA} = \sqrt{\frac{(\lambda_1 - \langle \mathbf{D} \rangle)^2 + (\lambda_2 - \langle \mathbf{D} \rangle)^2 + (\lambda_3 - \langle \mathbf{D} \rangle)^2}{3\langle \mathbf{D} \rangle}} \tag{35}$$

and the fractional anisotropy (FA) for visualizing the *normalized* $(0 \leq \mathrm{FA} < 1)$ degree of the diffusion anisotropy:

$$\mathrm{FA} = \sqrt{\frac{3\big((\lambda_1 - \langle \mathbf{D} \rangle)^2 + (\lambda_2 - \langle \mathbf{D} \rangle)^2 + (\lambda_3 - \langle \mathbf{D} \rangle)^2\big)}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}. \tag{36}$$

Figure 16 illustrates maps of MD, RA, and FA calculated from the diffusion tensor map acquired from a human brain. In general, the MD map shows high intensities in ventricles and GM, where the diffusion is relatively isotropic with higher strength. Both the RA and FA maps highlight the WM tracts in the brain, where the diffusion is highly anisotropic. The primary eigenvector map is also useful for indicating the orientation of the fibrous tissues. Figure 17 illustrates the primary eigenvector map of a partial brain by using a line-field technique. The primary eigenvector

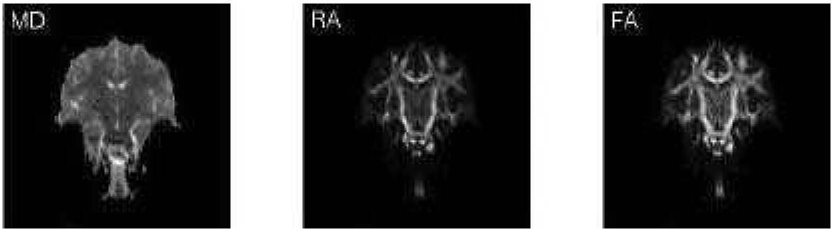


Fig. 16. The Mean Diffusivity (MD), Relative Anisotropy (RA), and Fractional Anisotropy (FA) maps of a same slice on a human brain.

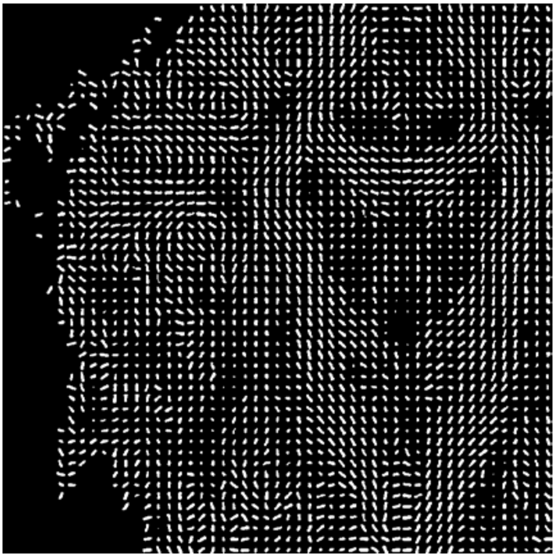


Fig. 17. A primary eigen-vector map calculated from the diffusion tensor data of a part of a coronal slice of human brain.

is represented by a line segment in each voxel. Directionally Encoded Color (DEC) is another method to visualize the $\mathbf{V}_1$ orientation by color encoding.[149] For example, a color mixed by red, green and blue (r, g, b) components in color space can be used to indicate the orientation of vector (r, g, b) in 3D space. Recently, a new technique called "fiber-tracking" was developed to virtually reconstruct the connectivity of the white matter tracts in the brain based on the primary eigenvector map.[150]

Despite the successes of DTI in a wide variety of applications, there are still challenges in dealing with some complex diffusion patterns. A major challenge is caused by fiber crossings, i.e. the multiple fiber compartments share a single voxel. In this case, the major eigenvector of the tensor can be substantially biased from the real fiber orientation. Unfortunately, the DTI formulation cannot offer a complete solution to this fiber-crossing problem, mainly because of the limitation of the tensor model itself. This is because the diffusion tensor is only a second order approximation (in terms of mean square fitting) to the real 3D diffusion process.[151]

Diffusion models beyond the tensor, such as High Angular Resolution Diffusion (HARD) and q-Space Imaging (QSI) are essential to overcome these difficulties.[152]

7. Application in Perfusion Imaging with Susceptibility Contrast Agents

MR perfusion imaging using exogenous contrast agents was developed at the end of 1980s and the early of 1990s.[153–160] In the so-called Dynamic Susceptibility Contrast Magnetic Resonance Imaging (DSC-MRI) techniques, a bolus of paramagnetic contrast agent is administrated and rapid MR imaging techniques are employed to collect the MRI signal change during its first pass through the tissue.[155] The paramagnetic contrast agents, on the one hand, shorten the longitudinal relaxation time T_1 due to the short-reaching dipolar coupling between the unpaired electrons of the paramagnetic contrast agents and the proton nuclear spins. On the other hand, the paramagnetic contrast agents shorten the tissue T_2 as well. This effect relies on the compartmentalization of the contrast agent, which creates a susceptibility gradient between the intravascular and extravascular space. The effect of the susceptibility gradient reaches far beyond the vessels into the surrounding tissue and thus leads to greater signal attenuation.[154] The understanding of the relationship between the MR signal change and the contrast agent is one of the keys to DSC-MRI perfusion imaging.

7.1. *MR signal change and contrast agent*

7.1.1. *Susceptibility*

When a sample is placed in the magnetic field B_0, the actually established magnetic field within the sample may be different from B_0. This phenomenon is explained in terms of "susceptibility" as follows

$$B = B_0 + M = B_0(1 + \chi), \tag{37}$$

where M is the resultant magnetization in the sample and χ is the sample magnetic susceptibility constant. The induced magnetization may align with or opposite to the applied magnetic field, respectively, depending on the number of unpaired or paired electrons in the sample. Gadolinium chelates (such as Gd-DTPA) are the commonly used contrast agents in DSC-MRI techniques, possessing sufficient unpaired electron spins that tend to align with the applied magnetic field, and therefore, are typical paramagnetic contrast agents.[154]

7.1.2. *Change in T_1 in blood and tissue with contrast agents*

The delivery of contrast agents into tissue can reduce the T_1 of the tissue. In general, a linear relationship between the increase in the relaxation rate R_1 (i.e. $1/T_1$) and the blood concentration C of contrast agent exists. Assuming the original relaxation rate of the tissue is R_1^0, the relaxation rate after the injection of a contrast agent, R_1, can be expressed as

$$R_1 \equiv \frac{1}{T_1} = R_1^0 + r_1 C, \tag{38}$$

where r_1 is called the longitudinal relaxivity (T_1 relaxivity). When a T_1-weighted pulse sequence is used, the MR signal will increase after the injection of a contrast agent. The linear relationship between the longitudinal relaxation rate and the concentration of the contrast agent is the basis of the T_1-weighted perfusion imaging techniques.[156,161]

7.1.3. *Change in T_2 or T_2^* in blood and tissue with contrast agent*

In the blood, the Red Blood Cells (RBC) and plasma possess different magnetic susceptibility constants, χ_{RBC} and χ_{plasma}. Therefore, a susceptibility gradient exists in the blood. The contrast agent is usually distributed in the plasma and will also contribute to the susceptibility difference. The transverse relaxation of blood is predominantly due to diffusion of water protons through field gradients arising from the susceptibility difference between the red blood cell and plasma, $\Delta\chi_{RBC/plasma}$, expressed as

$$\Delta\chi_{RBC/plasma} = \chi_{RBC} - (\chi_{plasma} + \chi_{ca}), \tag{39}$$

where χ_{ca} is the volume susceptibility constant of the contrast agent.[156] Since water exchange between intravascular and extravascular space is relatively slow, the change of tissue T_2 and T_2^* in the extravascular space is mainly not due to the water exchange, but to the dephasing of the extravascular spins in the spatially varying field caused by the magnetic susceptibility gradient between the blood and tissue.[156] At a large dose of a contrast agent, gradient-echo based techniques are equally sensitive to both macro- and microvasculature and can be used to measure the total vascular volume. A similar linear relationship between the change of the relaxation rate ΔR_2^* and the blood concentration of contrast agent C is given by

$$R_2^* \equiv \frac{1}{T_2^*} = R_2^{*0} + r_2^* C, \tag{40}$$

where r_2^* is called the transverse relaxivity. This relationship is the basis of perfusion imaging using DSC-MRI techniques. In contrast to gradient-echo based techniques, spin-echo based techniques are mainly sensitive to the microvasculature.[156,162]

7.2. *DSC-MRI perfusion imaging techniques*

7.2.1. *Theoretic basis*

The theoretical model used in the DSC-MRI perfusion imaging is established in the context of tracer kinetics of nondiffusible tracers.[163–166] In the model, several key quantities are defined as follows[155,167]:

$C_{VOI}(t)$ is the tissue concentration of the tracer in the Volume Of Interest (VOI) at time t after a bolus of tracer has been injected into the tissue.

$C_a(t)$ is the blood concentration of the tracer in the feeding artery at time t, and is generally called *Arterial Input Function* (AIF).
$h(t)$ is the probability density function of transit time t of the tracer through the VOI following an ideal instantaneous unit bolus injection. This quantity expresses the distribution of transit time of the tracer through the voxel and is determined only by the vascular structure. It is generally called *transport function*.
$R(t)$ is the fraction of the tracer still present in the VOI at time t following an ideal instantaneous unit bolus injection. It is generally called the *residue function*, and $R(t) = 1 - \int_0^t h(\tau)d\tau$, where the integral term represents the fraction of the total tracer that has left the VOI, and $R(0) = 1$, i.e. all of the tracer is present within the VOI at time $t = 0$ immediately after the ideal instantaneous unit bolus injection.

We firstly establish the inter-relationship between these quantities, from which we then derive Cerebral Blood Flow (CBF) and other physiologic parameters related to CBF, Cerebral Blood Volume (CBV) and the Mean Transit Time (MTT) of blood through a volume of tissue.[155,167] Considering a tracer is delivered to the VOI with an arterial input function $C_a(t)$ and the VOI possess the transport function $h(t)$, the concentration of tracer in the venous blood $C_V(t)$ follows the convolution relationship

$$C_V(t) = C_a(t) \otimes h(t) = \int_0^t C_a(\tau)h(t-\tau)d\tau. \tag{41}$$

Assuming a constant arterial blood flow F_{VOI} feeding the VOI, the amount of tracer entering per unit volume of the VOI is a time dependent product, $F_{VOI}C_a(t)$. Given that the microvasculature of the VOI possesses a residue function $R(t)$, the amount of tracer remained in the VOI, i.e. time dependent tissue concentration of tracer, $C_{VOI}(t)$, follows the convolution relationship:

$$C_{VOI}(t) = (F_{VOI}C_a(t)) \otimes R(t) = F_{VOI}\int_0^t C_a(\tau)R(t-\tau)d\tau, \tag{42}$$

As commonly used in the literature,[155,156] tissue density ρ needs to be added to the above expression. In addition, considering the difference in hematocrit between capillaries and large vessels and the fact that tracers are only distributed in intravascular and extracellular space (i.e. plasma space), another factor $k_H = (1 - H_{art})/(1 - H_{cap})$ also needs to be added to the above expression. So the final expression is as follows

$$C_{VOI}(t) = \frac{\rho}{k_H}(F_{VOI}C_a(t)) \otimes R(t) = \frac{\rho}{k_H}F_{VOI}\int_0^t C_a(\tau)R(t-\tau)d\tau. \tag{43}$$

Equation (43) is the central equation in the model to estimate the blood flow with nondiffusible tracer.[167] Similarly, we can derive cerebral blood volume in the VOI as follows[155]

$$\text{CBV} = \frac{k_H}{\rho}\frac{\int_0^\infty C_{VOI}(t)dt}{\int_0^\infty C_a(t)dt}. \tag{44}$$

According to the central volume theorem,[163,164] MTT can be calculated from CBV and CBF

$$\mathrm{MTT} = \frac{\mathrm{CBV}}{\mathrm{CBF}}. \tag{45}$$

7.2.2. *Perfusion calculation from DSC MR signal*

The equations given above indicate that the concentration of contrast agent is the fundamental quantity in the calculation of perfusion parameters. However, MR scanners cannot measure the concentration directly. So, an important step in MR perfusion imaging is to convert the DSC MR signal changes during the pass of contrast agent through tissue to the concentration of contrast agent in blood. As described earlier, a linear relationship between the change in relaxation rate, R_2 or R_2^*, and the blood concentration of contrast agent is the basis of this conversion. If the T_1-weighting effect during the pass of contrast agent is negligible, the MR signal, $S(t)$, after the injection of contrast agent follows a simple single exponential relation[154,155,167]

$$S(t) = S_0 \exp(-\mathrm{TE}\,\Delta R_2^*(t))\,, \tag{46}$$

where S_0 represents the baseline MR signal before the injection of contrast agent. Based on Eqs. (40) and (46), the concentration of contrast agent can be obtained from the following relation[154,155,167]

$$C(t) = \kappa_{VOI}\,\Delta R_2^* = -\frac{\kappa_{VOI}}{\mathrm{TE}} \ln\left(\frac{S(t)}{S_0}\right), \tag{47}$$

where TE is the echo time of the pulse sequence, and κ_{VOI} is a proportionality constant depending on the tissue, the type of contrast agent, the field strength, and the pulse sequence parameters. Similarly, the concentration of contrast agent in arterial blood, i.e. the AIF, can also be obtained using Eq. (47) with a different proportionality constant, κ_{art}, due to different physical environments in the arterial blood. After the tissue concentration-time curve and AIF are available, CBF, CBV and MTT can be obtained using Eqs. (43), (44) and (45), respectively. Equation (44) only involves integration of the tissue concentration-time curve and the AIF. Calculation of Eq. (45) is also straightforward.

To obtain CBF, one needs to deconvolve the tissue concentration-time curve $C_{VOI}(t)$ with the arterial input function $C_a(t)$ in Eq. (43). A successful deconvolution is critical to obtain the blood flow, F_{VOI}, accurately. A comprehensive treatment to this issue has been given in Refs. 167 and 168. There are two main categories of approaches to deconvolve Eq. (43): Model dependent techniques and model independent techniques. In the model dependent techniques, a specific analytical expression is assumed for the residue function, $R(t)$. Because the residue function is determined only by the tissue microvasculature, this empirical expression implicitly assumes a specific tissue microvasculature, which, however, may not have sufficient precision in the actual application. A single decreasing exponential is commonly used as the residue function

$$R(t) = \exp(-t/\mathrm{MTT}), \tag{48}$$

where MTT is a parameter representing the mean transit time under this model. This expression models the vascular bed as a single, well-mixed compartment. Incorporating Eq. (48) into Eq. (43), the two parameters F_{VOI} and MTT can be estimated by nonlinear least squares curve fitting, and CBV can then be obtained with Eq. (45). This is the general processing procedure of the model dependent deconvolution techniques.

In the model independent approaches, the residue function, $R(t)$, along with the blood flow F_{VOI}, is estimated from a non-parametric deconvolution procedure without any prior assumption. The model independent approaches are divided into two subcategories, the *transform approach* and the *algebraic approach*. The transform approach uses a Fourier transform or some other transform, combined with the associated convolution theorem, to deconvolve Eq. (43). Generally, however, this approach is very sensitive to noise and filtering is needed to suppress high frequency noise in the data.[167,169]

In the algebraic approach, Eq. (43) is reformed as a matrix equation by discretization of the convolution integral[167]:

$$\begin{aligned} C_{VOI}(t_j) &= \frac{\rho}{k_H} F_{VOI} \int_0^{t_j} C_a(\tau) R(t-\tau) d\tau \\ &\approx \frac{\rho}{k_H} F_{VOI} \left(\sum_{i=0}^{t_j} C_a(t_i) R(t_j - t_i) \right) \Delta t, \end{aligned} \tag{49}$$

where Δt is the sampling interval of the AIF and the tissue concentration-time curve. Equation (49) can be written as matrix notation:

$$\frac{\rho}{k_H} \Delta t \begin{pmatrix} C_a(t_1) & 0 & \cdots & 0 \\ C_a(t_2) & C_a(t_1) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ C_a(t_1) & C_a(t_1) & \cdots & C_a(t_1) \end{pmatrix} \left(F_{VOI} \begin{pmatrix} R(t_1) \\ R(t_2) \\ \vdots \\ R(t_N) \end{pmatrix} \right) = \begin{pmatrix} C_{VOI}(t_1) \\ C_{VOI}(t_2) \\ \vdots \\ C_{VOI}(t_N) \end{pmatrix}, \tag{50}$$

or

$$\boldsymbol{A} \cdot \boldsymbol{b} = \boldsymbol{c}, \tag{51}$$

where $\boldsymbol{b}$ represents the vector of elements $F_{VOI}\, R(t_i)$ and $\boldsymbol{c}$ represents the vector of tissue concentrations. This is generally an inverse problem and the goal is to obtain $\boldsymbol{b}$, from which the estimation of the blood flow and the residue function is straightforward. Two approaches are commonly used to solve this inverse problem, one using regularization[167,170] and the other is based on Singular Value Decomposition (SVD).[167,171,172]

Some other issues in perfusion quantification include elimination of the recirculation of contrast agent and the measurement of the AIF. The first one is due to the theoretical model assuming only the first pass of contrast agent through the tissue. The most commonly used method is to fit the concentration-time curve to a gamma-variate function. A detailed investigation on gamma-variate function fitting to the concentration-time curves is made in Ref. 173. The measurement of the AIF is an important factor to the perfusion measurement using DSC-MRI techniques. The effect of the delay and dispersion of the AIF on the estimation of CBF has been assessed in numerical simulation in Ref. 174. It was shown that an approximately 40% underestimation of CBF was introduced by a delay of 1 to 2 seconds in the AIF.[175] The effect of a delay in the AIF can be eliminated using a circular deconvolution.[172] However, the effect of dispersion of the AIF on the quantification of CBF cannot be ignored because the theoretical model cannot differentiate the dispersion of the AIF and the dispersion of contrast agent within the VOI that the indicator dilution theory is based upon.[175] Several factors, even conflicting with each other, can influence the measurement of AIF. To reduce the dispersion of the AIF, the site for the AIF measurement should be as close to the tissue of interest as possible. But, the partial volume effect will appear more easily in this case. It will bring about an underestimation of the AIF, and correspondingly, an overestimation of the CBF. To minimize the effect of partial volume, the sampling site for the AIF should be in a large artery. However, this is opposite to the effort of reducing the dispersion of the AIF due to a long transit from the artery to the tissue. Some related issues include which artery should be chosen as the sampling site for the AIF, and whether a common AIF is used for all the VOIs, or different AIFs for different VOIs, etc. On the whole, it seems that there is no consensus in the literature to the issues about the AIF measurement,[175] and further studies are required to provide better solution to these problems.

7.3. *Clinical application of the DSC-MRI perfusion imaging techniques*

Conventional MRI is insensitive to the detection of acute stroke. With conventional MRI, therefore, the cerebral ischemia is seldom studied during the hyperacute phase (first few hours) until the affected area was already visible on T_2-weighted images. As early assessment of the affected area become critical with the advent of thrombolytic therapy, perfusion imaging with DSC-MRI techniques begins to be extensively used in hyperacute diagnosis, identification of tissue at risk, and prediction of neurological outcome.[155]

A combined analysis using DSC-MRI, Diffusion-Weighted Imaging (DWI), and Magnetic Resonance Angiography (MRA) could be used to identify patients who might benefit from various treatment modalities and those who can avoid the potential risks associated with thrombolysis and neuroprotective agents. DSC-MRI and DWI could improve patient selection, guide stroke therapy, and, therefore, improve the evaluation of new therapeutic strategies.[155]

Due to abnormal blood flow and vascular patterns, brain tumors are also common targets of DSC-MRI studies. Most of these studies use CBV, and a few studies use the distribution of CBF and other parameters. A good correlation between CBV and tumor grade has been demonstrated. Cerebral blood volume mapping has also been used to differentiate recurrent tumor from radiation necrosis, and to monitor the changes associated with radiotherapy in a group of patients with astrocytomas. Tumor heterogeneity, not observed in conventional MRI, has been shown on the CBV maps and the DSC-MRI technique has shown high sensitivity to small regional CBV variations in the tumor.[155]

References

1. P. Mansfield, *J. Phys. C: Solid State Phys.* **L55** (1977).
2. M. S. Cohen and R. M. Weisskoff, *Magn. Reson. Imag.* **1** (1991).
3. A. B. Ahn, J. H. Kim and Z. H. Cho, *IEEE Trans. Med. Imag.* **2** (1986).
4. C. Meyer, B. Hu, D. Nishimura and A. Macovski, *Magn. Reson. Med.* **202** (1992).
5. S. Ogawa, T. M. Lee, A. S. Nayak and P. Glynn, *Magn. Reson. Imag.* **68** (1990).
6. P. J. Basser, J. Mattiello and D. Le Bihan, *J. Magn. Reson.* **247** (1994).
7. J. A. Detre, J. S. Leigh, D. S. Williams and A. P. Koretsky, *Magn. Reson. Imag.* **37** (1992).
8. B. R. Rosen, J. W. Belliveau and D. Chien, *Magn. Reson. Q.* **263** (1989).
9. D. Twieg, *Med. Phys.* **610** (1983).
10. D. C. Noll, *IEEE Trans. Med. Imag.* **372** (1997).
11. H. Feng, H. Gu, D. Silversweig, E. Stern and Y. Yang, *IEEE Trans. Med. Imag.* **925** (2003).
12. G. E. Sarty, *Magn. Reson. Med.* **445** (2004).
13. G. Glover, *Magn. Reson. Med.* **412** (1999).
14. K. F. King, T. K. F. Foo and C. Crawford, *Magn. Reson. Med.* **156** (1995).
15. J. I. Jackson, C. H. Meyer, D. G. Nishimura and A. Macovski, *IEEE Trans. Med. Imag.* **193** (1987).
16. C. H. Meyer, J. M. Pauly, A. Macovski and D. G. Nishimura, *Magn. Reson. Med.* **287** (1990).
17. W. T. Dixon, *Radiology* **189** (1984).
18. G. H. Glover and E. Schneider, *Magn. Reson. Med.* **371** (1991).
19. J. Hennig, *Eur. Radiol.* **1020** (1999).
20. P. Jezzard and R. S. Balaban, *Magn. Reson. Med.* **65** (1995).
21. P. J. Reber, E. C. Wong, R. B. Buxton and L. R. Frank, *Magn. Reson. Med.* **328** (1998).
22. P. Munger, G. R. Crelier, T. M. Peters and G. B. Pike, *IEEE Trans. Med. Imag.* **681** (2000).
23. H. Zeng and R. T. Constable, *Magn. Reson. Med.* **137** (2002).
24. X. Wan, G. T. Gullberg, D. L. Parker and G. L. Zeng, *Magn. Reson. Med.* **932** (1997).
25. N. K. Chen and A. M. Wyrwicz, *Magn. Reson. Med.* **1206** (1999).
26. J. Frahm, K.-D. Merboldt and W. Hanicke, *Magn. Reson. Med.* **474** (1988).
27. J. G. Ojemann, E. Akbudak, A. Z. Snyder, R. C. McKinstry, M. E. Raichle and T. E. Conturo, *Neuroimage* **156** (1997).

28. R. Young, I. J. Cox, D. J. Bryant and G. M. Bydder, *Magn. Reson. Imag.* **585** (1988).
29. Z. H. Cho and Y. M. Ro, *Magn. Reson. Med.* **193** (1992).
30. N. K. Chen and A. M. Wyrwicz, *Magn. Reson. Med.* **807** (1999).
31. V. A. Stenger, F. E. Boada and D. C. Noll, *Magn. Reson. Med.* **525** (2000).
32. R. T. Constable, *J. Magn. Reson. Imag.* **746** (1995).
33. Q. X. Yang, B. J. Dardzinski, S. Li, P. J. Eslinger and M. B. Smith, *Magn. Reson. Med.* **331** (1997).
34. G. H. Glover, *Magn. Reson. Med.* **290** (1999).
35. D. Cordes, P. A. Turski and J. A. Sorenson, *Magn. Reson. Imag.* **1055** (2000).
36. A. W. Song, *Magn. Reson. Med.* **407** (2001).
37. H. Gu, H. Feng, W. Zhan, S. Xu, D. A. Silbersweig, E. Stern and Y. Yang, *Neuroimage* **1358** (2002).
38. H. Bruder, H. Fisher, H.-E. Reinfelder and F. Schmitt, *Magn. Reson. Med.* **311** (1992).
39. X. Hu and T. H. Le, *Magn. Reson. Med.* **166** (1996).
40. M. H. Buonocore and L. Gao, *Magn. Reson. Med.* **89** (1997).
41. F. Farzaneh, S. J. Riederer and N. J. Pelc, *Magn. Reson. Med.* **123** (1990).
42. M. T. Vlaardingerbroek and J. A. den Boer, in *Magnetic Resonance Imaging Theory and Practice* (Springer, New York, 1996), p. 128.
43. E. Yudilevich and H. Stark, *IEEE Trans. Med. Imag.* **337** (1987).
44. H. Moriguchi, J. S. Lewin and J. L. Duerk, *Magn. Reson. Med.* **915** (2003).
45. L. C. Man, J. M. Pauly and A. Macovski, *Magn. Reson. Med.* **785** (1997).
46. T. B. Harshbarger and D. B. Twieg, *IEEE Trans. Med. Imag.* **196** (1999).
47. D. C. Noll, J. M. Pauly, C. H. Meyer, D. G. Nishimura and A. Macovski, *Magn. Reson. Med.* **319** (1992).
48. S. Nayak, C.-M. Tsai, C. H. Meyer and D. G. Nishimura, *Magn. Reson. Med.* **521** (2001).
49. P. Bornert, B. Aldefeld and H. Eggers, *Magn. Reson. Med.* **479** (2000).
50. G. H. Glover and C. S. Law, *Magn. Reson. Med.* **515** (2001).
51. Y. Yang, H. Gu, W. Zhan, S. Xu, D. A. Silbersweig and E. Stern, *Magn. Reson. Med.* **278** (2002).
52. J. Xiong, P. T. Fox and J. H. Gao, *Hum. Brain Mapp.* **41** (2003).
53. J. W. Belliveau, D. N. Kennedy, Jr., R. C. McKinstry, B. R. Buchbinder, R. M. Weisskoff, M. S. Cohen, J. M. Vevea, T. J. Brady and B. R. Rosen, *Science* **716** (1991).
54. S. Ogawa, T. M. Lee, A. R. Kay and D. W. Tank, *Proc. Natl. Acad. Sci. USA* **9868** (1990).
55. P. J. Magistretti in *Primer on Cerebrovascular Diseases* (Academic Press, San Diego, 1997), p. 70.
56. R. J. Gerrits, E. A. Stein and A. S. Greene, *Brain Res.* **108** (2002).
57. N. K. Logothetis, J. Pauls, M. Augath, T. Trinath and A. Oeltermann, *Nature* **150** (2001).
58. R. B. Buxton and L. R. Frank, *J. Cereb. Blood Flow Metab.* **64** (1997).
59. R. B. Buxton, E. C. Wong and L. R. Frank, *Magn. Reson. Med.* **855** (1998).
60. X. Hu, T. H. Le and K. Ugurbil, *Magn. Reson. Med.* **877** (1997).
61. T. Q. Duong, D. S. Kim, K. Ugurbil and S. G. Kim, *Magn. Reson. Med.* **231** (2000).
62. M. S. Cohen, *Neuroimage* **93** (1997).
63. G. K. Aguirre, E. Zarahn and M. D'Esposito, *Neuroimage* **360** (1998).

64. R. S. Menon, D. C. Luknowsky and J. S. Gati, *Proc. Natl. Acad. Sci. USA* **10902** (1998).
65. S. Ogawa, T. M. Lee, R. Stepnoski, W. Chen, X. H. Zhu and K. Ugurbil, *Proc. Natl. Acad. Sci. USA* **11026** (2000).
66. G. Bonmassar, K. Anami, J. Ives and J. W. Belliveau, *Neuroreport* **1893** (1999).
67. C. Janz, S. P. Heinrich, J. Kornmayer, M. Bach and J. Hennig, *Magn. Reson. Med.* **482** (2001).
68. J. R. Duann, T. P. Jung, W. J. Kuo, T. C. Yeh, S. Makeig, J. C. Hsieh and T. J. Sejnowski, *Neuroimage* **823** (2002).
69. M. A. Burock, R. L. Buckner, M. G. Woldorff, B. R. Rosen and A. M. Dale, *Neuroreport* **3735** (1998).
70. G. T. Buracas and G. M. Boynton, *Neuroimage* **801** (2002).
71. T. T. Liu, L. R. Frank, E. C. Wong and R. B. Buxton, *Neuroimage* **759** (2001).
72. K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith and R. S. J. Frackowiak, *Hum. Brain. Mapp.* **189** (1995).
73. R. McIntosh, F. L. Bookstein, J. V. Haxby and C. L. Grady, *Neuroimage* **143** (1996).
74. R. McIntosh, C. L. Grady, L. G. Ungerleider, J. V. Haxby, S. I. Rapoport and B. Horwitz, *J. Neurosci.* **655** (1994).
75. M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell and T. J. Sejnowski, *Hum. Brain Mapp.* **160** (1998).
76. Friman, M. Borga, P. Lundberg and H. Knutsson, *Neuroimage* **454** (2002).
77. R. Baumgartner, G. Scarth, C. Teichtmeister, R. Somorjai and E. Moser, *J. Magn. Reson. Imag.* **1094** (1997).
78. T. W. Moonen and P. A. Bandettini *Functional MRI* (Springer-Verlag, Berlin, 1999).
79. S. A. Huettel, A. W. Song and G. McCarths *Functional Magnetic Resonance Imaging* (Sinauer, Massachusetts, 2004).
80. P. A. Bandettini, E. C. Wong, R. S. Hinks, R. S. Tikofsky and J. S. Hyde, *Magn. Reson. Med.* **390** (1992).
81. K. K. Kwong, J. W. Belliveau, D. A. Chesler, I. E. Goldberg, R. M. Weisskoff, B. P. Poncelet, D. N. Kennedy, B. E. Hoppel, M. S. Cohen, R. Turner, H. Cheng, T. J. Brady and B. R. Rosen, *Proc. Natl. Acad. Sci. USA* **5675** (1992).
82. S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. G. Kim, H. Merkle and K. Ugurbil, *Proc. Natl. Acad. Sci. USA* **5951** (1992).
83. N. Sadato, V. Ibanez, G. Campbell, M. P. Deiber, D. Le Bihan and M. Hallett, *J. Cereb. Blood Flow Metab.* **670** (1997).
84. G. Thomas and R. S. Menon, *Magn. Reson. Med.* **203** (1998).
85. A. DeYoe, P. Bandettini, J. Neitz, D. Miller and P. Winans, *J. Neurosci. Methods* **171** (1994).
86. C. M. Wessinger, M. H. Buonocore, C. L. Kussmaul and G. R. Mangun, *Hum. Brain Mapp.* **18** (1997).
87. M. T. Banich, M. P. Milham, R. A. Atchley, N. J. Cohen, A. Webb, T. Wszalek, A. F. Kramer, Z. Liang, V. Barad, D. Gullett, C. Shah and C. Brown, *Brain Res. Cogn. Brain Res.* **1** (2000).
88. H. P. Volz, C. Gaser, H. J. Mentzel, F. Hager, W. A. Kaiser and H. Sauer, *Eur. Arch. Psychiatry Clin. Neurosci.* **161** (1998).
89. H. P. Volz, C. Gaser, F. Hager, R. Rzanny, H. J. Mentzel, I. Kreitschmann-Andermahr, W. A. Kaiser and H. Sauer, *Psychiatry Res.* **145** (1997).

90. D. Wagner, D. L. Schacter, M. Rotte, W. Koutstaal, A. Maril, A. M. Dale, B. R. Rosen and R. L. Buckner, *Science* **1188** (1998).
91. K. J. Friston, *Cereb. Cortex* **768** (1997).
92. R. D. Rogers, N. Ramnani, C. Mackay, J. L. Wilson, P. Jezzard, C. S. Carter and S. M. Smith, *Biol. Psychiatry* **594** (2004).
93. J. R. Binder, S. M. Rao, T. A. Hammeke, F. Z. Yetkin, A. Jesmanowicz, P. A. Bandettini, E. C. Wong, L. D. Estkowski, M. D. Goldstein and V. M. Haughton, *Ann. Neurol.* **662** (1994).
94. M. D'Esposito, J. A. Detre, D. C. Alsop, R. K. Shin, S. Atlas and M. Grossman, *Nature* **279** (1995).
95. H. Garavan, T. J. Ross, S. J. Li and E. A. Stein, *Cerebral Cortex* **585** (2000).
96. H. Garavan, T. J. Ross and E. A. Stein, *Proc. Natl. Acad. Sci. USA* **8301** (1999).
97. D. T. Cheng, D. C. Knight, C. N. Smith, E. A. Stein and F. J. Helmstetter, *Behav. Neurosci.* **3** (2003).
98. H. Garavan, J. C. Pendergrass, T. J. Ross, E. A. Stein and R. C. Risinger, *Neuroreport* **2779** (2001).
99. K. S. LaBar, D. R. Gitelman, T. B. Parrish, Y. H. Kim, A. C. Nobre and M. M. Mesulam, *Behav. Neurosci.* **493** (2001).
100. E. de Araujo, M. L. Kringelbach, E. T. Rolls and F. McGlone, *J. Neurophysiol.* **1865** (2003).
101. Y. Liu, J. H. Gao, H. L. Liu and P. T. Fox, *Nature* **1058** (2000).
102. H. K. Kang, J. J. Seo, H. J. Kim, S. B. Ryu and G. W. Jeong, *Urology* **1189** (2001).
103. H. Garavan, J. Pankiewicz, A. Bloom, J. K. Cho, L. Sperry, T. J. Ross, B. J. Salmeron, R. Risinger, D. Kelley and E. A. Stein, *Am. J. Psychiatry* **1789** (2000).
104. S. Bloom, R. G. Hoffmann, S. A. Fuller, J. Pankiewicz, H. H. Harsch and E. A. Stein, *Hum. Brain Mapp.* **235** (1999).
105. H. C. Breiter, R. L. Gollub, R. M. Weisskoff, D. N. Kennedy, N. Makris, J. D. Berke, J. M. Goodman, H. L. Kantor, D. R. Gastfriend, J. P. Riorden, R. T. Mathew, B. R. Rosen and S. E. Hyman, *Neuron* **591** (1997).
106. E. A. Stein, J. Pankiewicz, H. H. Harsch, J. K. Cho, S. A. Fuller, R. G. Hoffmann, M. Hawkins, S. M. Rao, P. A. Bandettini and A. S. Bloom, *Am. J. Psychiatry* **1009** (1998).
107. E. Zarahn, G. K. Aguirre and M. D'Esposito, *Neuroimage* **179** (1997).
108. N. S. Lawrence, T. J. Ross and E. A. Stein, *Neuron* **539** (2002).
109. R. G. Wise, R. Rogers, D. Painter, S. Bantick, A. Ploghaus, P. Williams, G. Rapeport and I. Tracey, *Neuroimage* **999** (2002).
110. J. A. Frazier, S. L. Rauch, L. J. Seidman, P. J. Whalen, M. A. Jenike, B. R. Rosen and J. Biederman, *Biol. Psychiatry* **1542** (1999).
111. M. Beauregard, J. M. Leroux, S. Bergman, Y. Arzoumanian, G. Beaudoin, P. Bourgouin and E. Stip, *Neuroreport* **3253** (1998).
112. C. M. Adler, P. McDonough-Ryan, K. W. Sax, S. K. Holland, S. Arndt and S. M. Strakowski, *J. Psychiatr. Res.* **317** (2000).
113. S. J. Li, Z. Li, G. Wu, M. J. Zhang, M. Franczak and P. G. Antuono, *Radiology* **253** (2002).
114. E. Desmond, J. M. Sum, A. D. Wagner, J. B. Demb, P. K. Shear, G. H. Glover, J. D. Gabrieli and M. J. Morrell, *Brain* **1411** (1995).
115. F. Nitschke, U. H. Melchert, C. Hahn, V. Otto, H. Arnold, H. D. Herrmann, G. Nowak, M. Westphal and K. Wessel, *Acta Neurochir. (Wien)* **1223** (1998).

116. M. E. Raichle, *Handbook of Physiology — The Nervous System* (Vol. 5). *American Physiological Society* (1987), p. 643.
117. R. R. Edelman, B. Siewert, D. G. Darby, V. Thangaraj, A. C. Nobre, M. M. Mesulam and S. Warrash, *Radiology* **513** (1994).
118. S. G. Kim, *Magn. Reson. Med.* **293** (1995).
119. E. C. Wong, R. B. Buxton and L. R. Frank, *NMR Biomed* **237** (1997).
120. F. Q. Ye, J. Pekar, P. Jezzard, J. H. Duyn, J. A. Frank and A. C. McLaughlin, *Magn. Reson. Med.* **219** (1996).
121. Y. Yang, J. A. Frank, L. Hou, F. Q. Ye, A. C. McLaughlin and J. H. Duyn, *Magn. Reson. Med.* **825** (1998).
122. F. Q. Ye, A. M. Smith, V. S. Mattay, U. E. Ruttimann, J. A. Frank, D. R. Weinberger and A. C. McLaughlin, *Neuroimage* **44** (1998).
123. M. A. Mintun, P. T. Fox and M. E. Raichle, *J. Cereb. Blood. Flow. Metab.* **96** (1989).
124. R. L. Buckner, P. A. Bandettini, K. M. O'Craven, R. L. Savoy, S. E. Petersen, M. E. Raichle and B. R. Rosen, *Proc. Natl. Acad. Sci. USA* **14878** (1996).
125. H. L. Liu and J. H. Gao, *Magn. Reson. Med.* **1011** (1999).
126. Y. Yang, W. Engelien, H. Pang, S. Xu, D. A. Silbersweig and E. Stern, *Neuroimage* **287** (2000).
127. E. C. Wong, W. M. Luh and T. T. Liu, *Magn. Reson. Med.* **511** (2000).
128. A. C. Silva, S. P. Lee, G. Yang, C. Iadecola and S. G. Kim, *J. Cerebr. Blood. Flow. Met.* **871** (1999).
129. W. M. Luh, E. C. Wong, P. A. Bandettini, B. D. Ward and J. S. Hyde, *Magn. Reson. Med.* **137** (2000).
130. C. Schwarzbauer, *NMR Biomed.* **37** (2000).
131. Y. Yang, *Concepts Magn. Reson.* **347** (2002).
132. H. L. Kantor, D. R. Gastfriend, J. P. Riorden, R. T. Mathew and B. P. Rosen, *Neuron* **591** (1997).
133. S. M. Rao, P. A. Bandettini and A. S. Bloom, *Am. J. Psychiatry* **1009** (1998).
134. S. G. Kim and K. Ugurbil, *Magn. Reson. Med.* **59** (1997).
135. T. L. Davis, K. K. Kwong, R. M. Weisskoff and B. R. Rosen, *Proc. Natl. Acad. Sci. USA* **1834** (1998).
136. R. D. Hoge, J. Atkinson, B. Gill, G. R. Crelier, S. Marrett and G. B. Pike, *Proc. Natl. Acad. Sci. USA* **9403** (1999).
137. D. Le Bihan, *Magn. Reson. Quart* **7** (1991) 1.
138. L. R. Frank, *Magn. Reson. Med.* **45** (2001) 935.
139. W. Zhan, H. Gu, S. Xu, D. A. Silbersweig, E. Stern and Y. Yang, *Magn. Reson. Med.* **49** (2003) 1077.
140. R. Xue, P. C. M. Van Zijl, B. J. Crain, M. Solaiyappan and S. Mori, *Magn. Reson. Med.* **42** (1999) 1123.
141. A. Einstein, *Investigations on the Theory of Brownian Motion* (New York, Dover, 1926).
142. D. Le Bihan, *Temperature imaging by NMR diffusion* **181** (New York, Raven, 1995).
143. E. O. Stejskal and J. E. Tanner, *J. Chem. Phys.* **42** (1965) 288.
144. D. L. Thomas, M. F. Lythgoe, G. S. Pell, F. Calamante and R. J. Ordidge, *Phys. Med. Biol.* **45** (2000) R97.
145. P. J. Basser, J. Mattiello and D. Le Bihan, *J. Magn. Reson.* **103** (1994) 247.
146. D. Le Bihan, J. F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko and H. Chabriat, *J. Magn. Reson. Imag.* **13** (2001) 534.
147. G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins University Press 1996).

148. P. J. Basser and C. Pierpaoli, *J. Magn. Reson. Ser. B* **111** (1996) 209.
149. S. Pajevic and C. Piepaoli, *Magn. Reson. Med.* **42** (1999) 526.
150. C. Poupon, C. A. Clark, V. Frouin, J. Regis, I. Bloch, D. Le Bihan and J. Mangin, *Neuroimage* **184** (2000).
151. P. J. Basser, S. Pajevic, C. Pierpaoli, C. Duda and A. Aldroubi, *Magn. Reson. Med.* **44** (2000) 625.
152. C. Lin, V. J. Wedeen, J. Chen, C. Yao and W. I. Tseng, *Neuroimage* **19** (2003) 482.
153. D. Le Bihan and R. Turner*, Magn. Reson. Med.* **171** (1992).
154. M. K. Stehling, R. Brüning and B. R. Rosen, in *Echo-Planar Imaging Theory, Technique and Application.* (eds.) F. Schmitt, M. K. Stehling and R. Turner (Springer, Berlin, 1998) pp. 419–464.
155. F. Calamante, D. L. Thomas, G. S. Pell, J. Wiersma and R. Turner, *J. Cereb. Blood Flow Metab.* **701** (1999).
156. E. L. Barbier, L. Lamalle and M. Decorps, *J. Magn. Reson. Imag.* **496** (2001).
157. B. R. Rosen, J. W. Belliveau, J. M. Vevea and T. J. Brady, *Magn. Reson. Med.* **249** (1990).
158. A. Villringer, B. R. Rosen, J. W. Belliveau, J. L. Ackerman, R. B. Lauffer, R. B. Buxton, Y. S. Chao, V. J. Wedeen and T. J. Brady, *Magn. Reson. Med.* **164** (1988).
159. C. R. Fisel, J. L. Ackerman, R. B. Buxton, L. Garrido, J. W. Belliveau, B. R. Rosen and T. J. Brady, *Magn. Reson. Med.* **336** (1991).
160. B. R. Rosen, J. W. Belliveau, H. J. Aronen, D. Kennedy, B. R. Buchbinder, A. Fischman, M. Gruber, J. Glas, R. M. Weisskopf, M. S. Cohen, F. H. Hochberg and T. J. Brady, *Magn. Reson. Med.* **293** (1991).
161. E. M. Haacke, R. W. Brown, M. R. Thompson and R. Venkatesan, *Magnetic Resonance Imaging — Physical Principles and Sequence Design* (John Wiley & Sons, Inc, New York, 1999) pp. 367–370.
162. J. L. Boxerman, L. M. Hamberg, B. R. Rosen and R. M. Weisskoff, *Magn. Reson. Med.* **555** (1995).
163. G. N. Steward, *J. Physiol* (London) **1** (1894).
164. P. Meier and K. L. Zierler, *J. Appl. Physiol.* **731** (1954).
165. K. L. Zierler, *Circ. Res.* **393** (1962).
166. L. Axel, *Radiology* **676** (1980).
167. L. Østergaard, R. M. Weisskoff, D. A. Chesler, C. Gyldensted and B. R. Rosen, *Magn. Reson. Med.* **715** (1996).
168. L. Østergaard, A. G. Sorensen, K. K. Kwong, R. M. Weisskoff, C. Gyldensted and B. R. Rosen, *Magn. Reson. Med.* **726** (1996).
169. K. A. Rempp, G. Brix, F. Wenz, C. R. Becker, F. Guckel and W. J. Lorenz, *Radiology* **637** (1994).
170. F. Calamante, D. G. Gadian and A. Connelly, *Magn. Reson. Med.* **1237** (2003).
171. H. Liu, Y. Pu, Y. Liu, L. Nickerson, T. Andrews, P. T. Fox and J. Gao, *Magn. Reson. Med.* **167** (1999).
172. O. Wu, L. Østergaard, R. M. Weisskoff, T. Benner, B. R. Rosen and A. G. Sorensen, *Magn. Reson. Imag.* **164** (2003).
173. T. Benner, S. Heiland, G. Erb, M. Forsting and K. Sartor, *Magn. Reson. Imag.* **307** (1997).
174. F. Calamante, D. G. Gadian and A. Connelly, *Magn. Reson. Imag.* **466** (2000).
175. F. Calamante, D. G. Gadian and A. Connelly, *Stroke* **1146** (2002).

CHAPTER 9

SHAPE BASED INTERPOLATION METHODS FOR MEDICAL IMAGES AND THEIR APPLICATION

TONG-YEE LEE* and CHAO-HUNG LIN
Computer Graphics Group/Visual System Lab
Department of Computer Science and Information Engineering
National Cheng Kung University, Taiwan, ROC
**tonylee@mail.ncku.edu.tw*

In the past, a variety of interpolation approaches have been proposed for medical images. The shape-based interpolation is a well-known and most commonly used method that can be implemented efficiently and achieve reasonable results. In this chapter, the morphology-based interpolation is first introduced and it can solve drawbacks in the shape-based interpolation method. Next, a powerful feature-guided image interpolation is presented. This method automatically finds feature-line segments and integrates image-warping technique to interpolate shapes. In comparison with the shape-based and morphology-based methods, the feature-guided method can manage more general image cases and generate better shape interpolation.

1. Introduction

Clinicians exploit computer graphics tools to enable them to visualize, manipulate, and quantitate the 3D internal structures of patients. Major sources of data in these medical applications are gathered from 2D medical-imaging devices such as CT, MRI and PET. A 3D image, formed by stacking a contiguous series of 2D images, can be used to visualize complex structures in 3D. However, generally, the number of image slices generated from these instruments is not adequate enough to produce high-quality 3D images. Therefore, the 3D reconstruction must be accomplished by appropriate interpolation methods to fill gaps between available image slices.

A variety of approaches have been proposed to reconstruct 3D objects. Among these works, the simplest method is to linearly interpolate the gray values in the slices to fill in the gray values in the missing slices.[1–3] With this scheme, an artifact always arises when the location of a boundary between two uniform regions shifts considerably between two adjacent slices.

Raya *et al.*[4] and Herman *et al.*[5] proposed shape-based methods by encoding the segmented image with distance codes. This approach interpolates the distance instead of the gray values and therefore maintains better geometric changes than the gray-value interpolation. Because the shape-based method can be implemented efficiently and achieve reasonable interpolation results, it has become a widely used method. However, it cannot deal effectively with objects with holes, large offsets, or heavy invagination. To overcome these drawbacks, Guo *et al.*[6] and Lee *et al.*[7]

develop a morphology-based interpolation method. This method first overlaps the two slices to obtain a Morphologically Difference Image (MDI) and then applies a sequence of dilation and erosion operations on non-overlapping regions to achieve interpolation. This method successfully resolves the problems in objects with holes and large offsets.

In this chapter, the shape-based interpolation and the morphology-based interpolation are introduced first. A feature-guided image interpolation scheme[8] is then presented. It is an effective and improved shape-based interpolation method used for interpolating image slices in medical applications. This approach integrates feature line-segments to guide the shape-based method for better shape interpolation. An automatic method for finding these line segments is given. This approach can manage translation, rotation and scaling situations when the slices have similar shapes. It can also interpolate intermediate shapes when the successive slices do not have similar shapes.

2. Shape-Based Interpolation

Shape-based interpolation converts segmented binary images into distance maps. The value of distance maps is the *Euclidean shortest distance* between the pixel and the boundary of the organ (positive values for inside the organ and negative values for outside). Therefore, the distance of a pixel P is defined by (1)

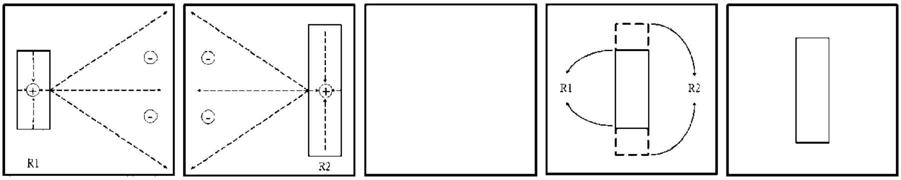
$$dist\_map(P) = \begin{cases} 0 & \text{if } P \in \partial X \\ +dis(P, \partial X) & \text{if } P \in X \\ -dis(P, \partial X) & \text{if } P \notin X \end{cases}, \tag{1}$$

where

X represents the object;
∂X represents the contour of the object X;
$dist(P, \partial X)$ represents the shortest distance from P to ∂X.

The intermediated slices can be interpolated by thresholding the distance maps at zero. Unfortunately, if there is no *prior* alignment between two input images, the shape-based interpolation cannot perform well, For example, in Figs. 1(a) and (b), there are two contours denoted as R_1 and R_2 on two binary images. Without an appropriate alignment, the shape-based scheme creates a bad interpolation in Fig. 1(c), where there is no contour in this interpolated image. If an appropriate alignment is performed (i.e. match the centroids of two objects *prior* to distance transformation (as shown in Fig. 1(d)), a better interpolation (as shown in Fig. 1(e)) can be obtained.

In Refs. 4 and 5 a variety of distance transformations for digital images are discussed. In general, most approaches concentrate on using difference convolution masks such as city block or chamfer template to efficiently approximate the *Euclidean shortest distance*. Considering the example of Fig. 2. There are two objects X_0 and X_{n+1}. The object X_{n+1} fully enclosed the object X_0. In this case,



(a) (b) (c) (d) (e)

Fig. 1. Shape-based interpolation and object centralization.

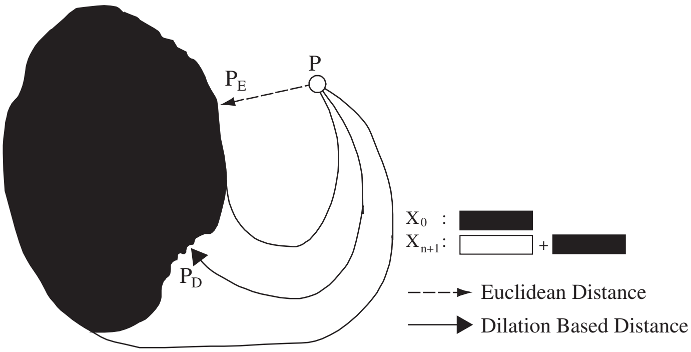


Fig. 2. Dilation based distance versus Euclidean distance.

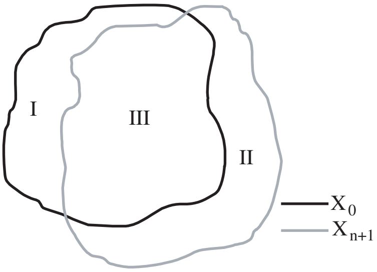


Fig. 3. Morphology difference between X_0 and X_{n+1} after object centralization.

a pixel P belonging to X_{n+1} will correspond to P_E on X_0 using *Euclidean shortest distance* transformation. This result seems awkward, since a better result would have the intermediate pixel moving from P to P_D.

To overcome these drawbacks, Guo *et al.*[6] and Lee *et al.*[7] present the morphology-based interpolation method. In this method, only the non-overlapping regions are interpolated using a sequence of dilation and erosion operations. First, the two corresponding objects X_0 and X_{n+1} are aligned using *object centralization*. After this alignment, there are three kinds of possible portions: Regions I, II, and III, respectively (as shown in Fig. 3). Regions I and II represent the morphological difference between the two objects X_0 and X_{n+1}. Then, apply a *dilation* operator to both regions I and II, respectively. The purpose of this step is to obtain the dilation-based distance from the pixel P (i.e. on region I or II) to the boundary of region III. Next, apply an *erosion* operator to interpolate the results.

```
Algorithm Dilation Based Distance Coding
Begin
    /* a. initialize */
    $index \leftarrow 0$;
    Initialize all elements of Array $A_0$ and $A_{n+1}$ to be $-1$;
    Set elements $A_0(x,y)$ to be 0, if the pixel $(x,y) \in X_0$;
    Set elements $A_{n+1}(x,y)$ to be 0, if the pixel $(x,y) \in X_{n+1}$;
    For each pixel $(x,y)$ of $\partial X_0$
        Insert a new node $M(x,y,0,n+1,0)$ into the active dilation contour list $L_D$;
        For each pixel $(x,y)$ of $\partial X_{n+1}$
            Insert a new node $M(x,y,n+1,0,0)$ into the active dilation contour list $L_D$
    /* b. entering dilation process */
    While $L_D$ is not empty
        Retrieve the first node of the list $L_D$ and denoted as $N(x,y,a,b,d)$
        /* c. the same dilation layer */
        If $N \cdot d = index$ then
        /* d. check if 4-neighbors could be updated as the next dilation layer*/
            For each four-neighbor $P(x',y')$ of $(N.x, N.y)$
                /* e. check if inside another object */
                If $P \in X_b$ then
                    If $A_a(P) \geq 0$ and $A_a(P) \leq index + 1$ then
                        No update on $A_a(P)$ and $L_D$;
                    Else
                        Insert point P into the tail of the list $L_D$
                        with $(x', y', a, b, index + 1)$;
                        $A_a(P) \leftarrow index + 1$;
        /* f. next dilation layer */
        Else
            $index \leftarrow index + 1$;
            Insert node N back to the first position of the list $L_D$;
End
```

The above pseudo-code is the dilation-based distance transformation algorithm.[7] In this algorithm, two arrays A_0 and A_{n+1} are used to store the distance codes for regions II and I, respectively. Initially [part (a) in algorithm], each pixel of the distance maps is set to be -1 (i.e. outside X_0 or X_{n+1}) or 0 (i.e. inside X_0 or X_{n+1}). Then, all pixels of both ∂X_0 and ∂X_{n+1} are inserted into a list called active dilation list L_D. The distance code of these contour pixels is all zero. Assume a smaller object X_0 is fully enclosed by a larger object X_{n+1}. Start from each node at active linked list L_D to perform the dilation [part (b)]. In other words, start the dilation from layer zero, since the distance code of all nodes at L_D is zero (i.e. contour pixels at both ∂X_0 and ∂X_{n+1}). In the course of dilation, there will be many dilation layers created in a monotonically increasing order (distance code value) and each layer has an equal distance code [part (c)]. Furthermore, once a layer is totally completed, another layer is then started [part (f)]. While performing dilation, a node will insert a new node belonging to the next layer into the tail of L_D. Therefore, this newly added node will not start dilation until those nodes (with smaller distance codes) in front of it finish dilation. The whole dilation will be repeated until there is no pixel with −1 within the morphology difference region [part (e)].

3. Feature-Guided Shape-Based Image Interpolation

In this section, a feature-guided interpolation scheme is presented. Figure 4 shows a brief overview of this interpolation scheme. This scheme can be divided into the following steps. For any given two consecutive segmented images (binary images), extract the contours for the objects of interest. Compute the object matching and create positive and negative object (i.e. hole) pairs. For each object pair (positive or negative), generate the corresponding distance maps. The feature line-segments are found automatically and warping is used to interpolate the intermediate distance maps. Next, by threshold the distance map at zero to obtain the interpolated contours. If creating multiple contours is required, such as branching or holes, some of the above procedures must be processed several times (see the dashed line in Fig. 4). Finally a contour-blending task is required to combine all interpolated contours together to obtain the correct results. In the following subsections, these main steps for this interpolation scheme are presented.

3.1. *Pseudo-object generation*

In the image slices, the holes will be treated as negative objects and the organs will be treated as positive objects. For any negative object pair, there must exist a corresponding positive object pair. Sometimes one hole is inside one object, but there is no corresponding hole inside the other object. In such a situation, a pseudo-object must be created. Similarly, if a positive object is on a slice and there is no corresponding positive object on the other slice, a pseudo-positive object will

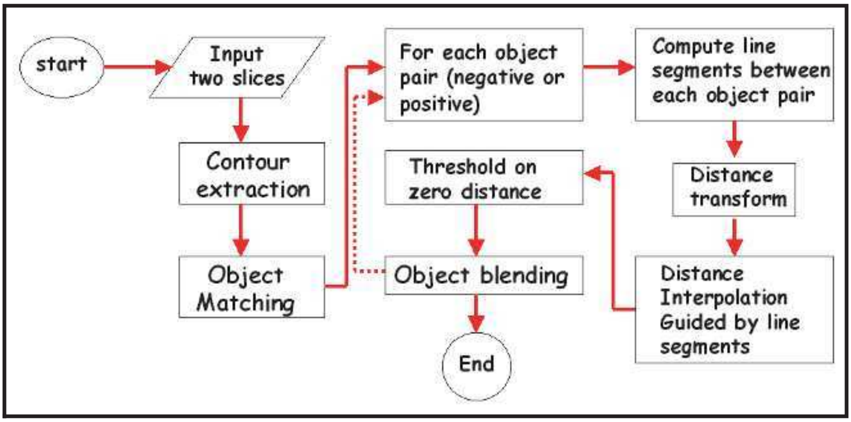

start
Input two slices
Contour extraction
Object Matching
For each object pair (negative or positive)
Compute line segments between each object pair
Distance transform
Distance Interpolation Guided by line segments
Threshold on zero distance
Object blending
End

Fig. 4. The flowchart of the feature-guided shape-based interpolation algorithm.

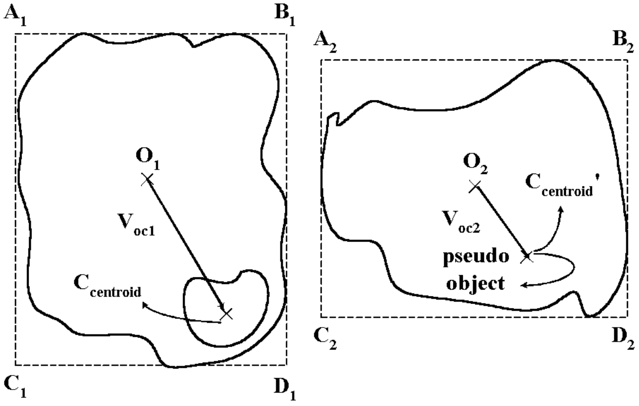


Fig. 5. Pseudo-object generation.

be created. To generate a pseudo-positive object is very straightforward. Assume there is a positive object and its center is at $C_{centroid}$. Then, on the other slice, a corresponding pseudo-positive object is created at $C_{centroid}$ with a size of one pixel. For a pseudo-negative object, there is some extra work described as follows. In Fig. 5, there are two corresponding objects and their centroids are O_1 and O_2. These two objects are bounded by boxes $A_1B_1C_1D_1$ and $A_2B_2C_2D_2$. Each box is further divided into four subregions (A_iO_i, B_iO_i, C_iO_i, D_iO_i) based on O_i. Assume there is a hole on the first and its centroid is $C_{centroid}$. In Fig. 5(a), suppose $C_{centroid}$ is located in D_1O_1 (denoted as region$_1$). Then, a pseudo-object will be created at D_2O_2 (denoted as region$_2$) in Fig. 5(b), too. The size of this pseudo-object is one pixel and its centroid $C'_{centroid}$ can be computed in (2) and (3).

$$\begin{aligned}\overrightarrow{V_{oc2}\cdot x} &= \overrightarrow{V_{oc1}\cdot x}\cdot\frac{length\ of\ region_2}{length\ of\ region_1}\\ \overrightarrow{V_{oc2}\cdot y} &= \overrightarrow{V_{oc1}\cdot y}\cdot\frac{width\ of\ region_2}{width\ of\ region_1}\end{aligned}, \tag{2}$$

$$C'_{centroid} = O_2 + \overrightarrow{V_{oc2}}\,. \tag{3}$$

In (4) and (5), V_{ocl} is a vector from O_1 to $C_{centroid}$. If, instead, one tries to compute $C'_{centroid}$ directly by $O_2 + V_{ocl}$, it could incorrectly create a pseudo-object outside the object (as shown in Fig. 6).

3.2. *Object matching*

Multiple objects may exist on two input slices. In this situation, the matching problem must be solved. In here, a simple method is provided. First, for each potential object pair (a, b), a score of matching can be evaluated by (4). Then, if this score is higher than a threshold, this object pair is matched. In (4), two factors are taken into account: (i) Object pair (a, b) is overlapped or not (i.e. returns 1 or 0), and (ii) the distance between two object centers is within a range or not.

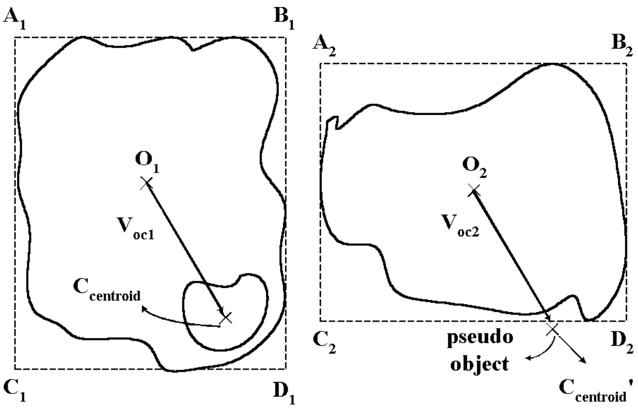


Fig. 6. Example of incorrect pseudo-object generation.

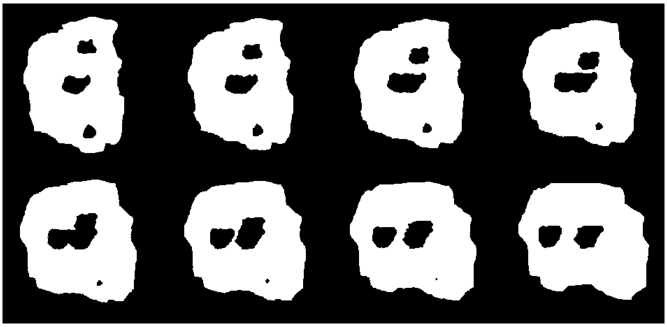


Fig. 7. An example of interpolation using the matching policy.

Both w_1 and w_2 are user-specified weights to compute (4). For a given object pair, the $dist_{threshold}$ can be selected as the sum of width and length of the larger object. Figure 7 shows an interpolation result using this matching policy.

$$score\_match = overlap(a,b) \times w_1 + (dist_{threshold} - dist(a,b)) \times w_2. \tag{4}$$

3.3. *Shape-based interpolation using warping*

The object contour pairs (C_s, C_{s+1}) are extracted from two matched objects X_0 and X_{n+1}. Two distance maps (D_s, D_{s+1}) are computed using any distance transformation algorithm. After the corresponding feature line-segments are computed (will be described later) for (C_s, C_{s+1}), a warping technique[9] is used to fill in the distance information for the intermediate distance map D_t. To create D_t, the warping is divided into two main steps. First, two deformed distance maps (D'_s, D'_{s+1}) from (D_s, D_{s+1}) are computed according to the control line-segments. The two deformed (D'_s, D'_{s+1}) maps are then linearly blended to generate D_t. Each control line segment is directed. For each corresponding pair of line segments $\overline{P_sQ_s}$ and $\overline{P_{s+1}Q_{s+1}}$ on $(D_s,\, D_{s+1})$, the warping computes an intermediate line segment $\overline{P_tQ_t}$, where $0 \le t \le 1$, $P_t = P_s * (1-t) + t * P_{s+1}$ and $Q_t = Q_s * (1-t) + t * Q_{s+1}$. The deformed D'_s can then be computed as follows. For each pixel coordinate P of D'_s and its corresponding image pixel P' on D_s, the warping computes P' using:

$$u = \frac{(P - P_t) \bullet (Q_t - P_t)}{\|Q_t - P_t\|^2}, \tag{5}$$

$$v = \frac{(P - P_t) \bullet \mathit{Perpendicular}(Q_t - P_t)}{\|Q_t - P_t\|}, \tag{6}$$

$$P' = P_s + u \bullet (Q_s - P_s) + \frac{v \bullet \mathit{Perpendicular}(Q_s - P_s)}{\|Q_s - P_s\|}, \tag{7}$$

where the value u is the position along the line normalized by the distance $\overline{P_tQ_t}$ and v is the distance from the line and procedure *Perpendicular*() returning a vector that is perpendicular to the input vector. The idea is very simple in the above equations. Both directed lines $\overline{P_tQ_t}$ and $\overline{P_sQ_s}$ define their local coordinate systems. For the corresponding P and P', their local coordinates are (u, v) in these two local systems defined by $\overline{P_tQ_t}$ and $\overline{P_sQ_s}$. The warping method uses multiple line segments. Assume that $P'_1, P'_2, \ldots, P'_n$ have been computed due to n corresponding line segments. The combined point P' for P can be calculated as follows:

$$P' = \frac{\sum_{i=1}^{n} w_i * P'_i}{\sum_{i=1}^{n} w_i} \quad \text{and} \quad w_i = (a + \mathit{distance})^{-b}, \tag{8}$$

where the distance is the distance from point P to each line segment on D'_s. The parameter a is a small number to avoid w_i being a zero value and the parameter b is used to control the rate of degradation influence per each line segment. In this manner, the warping function calculates P' for P. We then let the distance value for P on D'_s be the distance for P' on D_s. Using procedures similar to those above we can also compute the deformed D'_{s+1}. Now, we have computed two deformed $(D'_s,\, D'_{s+1})$ maps for the intermediate D_t. Next, both D'_s and D'_{s+1} are blended in a linear manner to calculate D_t using:

$$D_t(P) = D'_s(P) * (1 - t) + D'_{s+1} * t, \tag{9}$$

where $0 \le t \le 1$ and P is a pixel coordinate. The above procedures are called a feature-guided shape-based interpolation method.

3.4. *Automatically computing control line segments*

In this section an approach to compute corresponding line segments for a given contour pair $(C_s,\, C_{s+1})$ will be presented. This approach consists of three main tasks: (i) Finding the principle axis of each contour, (ii) simplifying the input contours and (iii) contour matching. These three tasks will be presented in the following sections.

3.4.1. *Principle axis alignment*

Given two input contours, align their principle axes must be aligned before finding their corresponding line segments. A 2D contour C consists of n points and any two consecutive points p_i and p_{i+1} of these n points can form a line segment. The coordinate of p_i is denoted as $(p_i \cdot x, p_i \cdot y)$. The principle axis of a contour C can be computed from its Σ, covariance matrix, where Σ is defined by:

$$\Sigma = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\{(p_i \cdot x - m \cdot x)(p_i \cdot x - m \cdot x)\} & \frac{1}{n}\sum_{i=1}^{n}\{(p_i \cdot x - m \cdot x)(p_i \cdot y - m \cdot y)\} \\ \frac{1}{n}\sum_{i=1}^{n}\{(p_i \cdot y - m \cdot y)(p_i \cdot x - m \cdot x)\} & \frac{1}{n}\sum_{i=1}^{n}\{(p_i \cdot y - m \cdot y)(p_i \cdot y - m \cdot y)\} \end{bmatrix}. \tag{10}$$

In the above equation, $(m \cdot x, m \cdot y)$ represents the average of n points. The corresponding eigenvector of maximum eigenvalue λ of Σ defines the principle axis of the contour C. The principle axes for the two input contours (C_s and C_{s+1}) can be computed and their axes are denoted V_s and V_{s+1}, respectively. The included angle θ between V_s and V_{s+1} is then determined. Using this included angle θ, each p_i of C_s is then rotated to achieve alignment using:

$$p_i' = R(\theta) * p_i, \tag{11}$$

where $R(\theta)$ is the rotation matrix.

3.4.2. *Contour simplification*

In the feature-guided shape-based interpolation scheme, line segments are used to control the shape-based interpolation. Ideally, the line segments will not be too many and should be features of the given contours. For this purpose, the optimal polygonal approximation of digitized curves[10] is performed. In Ref. 10, the error function, *error* (i, j) is used to estimate the local approximation and is equal to the square Euclidean distance from contour each point from (x_i, y_i) to (x_j, y_j) to its orthogonal projection onto the line $y = ax + b$ defined by (x_i, y_i) and (x_j, y_j) using:

$$error(i, j) = \sum_{t=i}^{j} \frac{(y_{P_t} - ax_{P_t} - b)^2}{a^2 + 1}, \tag{12}$$

where P_t is a point from (x_i, y_i) to (x_j, y_j). Figure 8 shows an example using this method. After simplification, only the feature points of a given contour are left. In the next subsection, we will describe a matching method to determine the correspondence among these feature points.

3.4.3. *Contour matching*

Let $C_s(u)$, $u \in [0, 1]$ and $C_{s+1}(v)$, $v \in [0, 1]$ be two parametric curves for the source and input contours. To establish the correspondence between the two curves, two

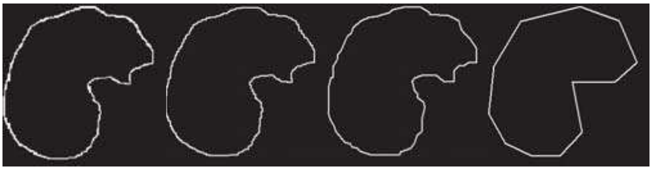


Fig. 8. Contour simplification: (a) An input contour consists of 999 points. (b), (c) and (d) are simplified contours of (a) consisting of 250, 60 and 12 points, respectively.

matching criteria are considered: Intensity and geometry similarities. The image correlation is used to evaluate the intensity similarity. Assume that two contours are originally extracted from two given gray-level image slices I_s and I_{s+1}. For two contour points p_s^i and p_{s+1}^j on I_s and I_{s+1}, their image correlation can be computed using:

$$C(p_s^i, p_{s+1}^j) = \frac{\sigma_{s,s+1}^2}{(\sigma_s^2 \sigma_{s+1}^2)^{1/2}}, \tag{13}$$

where σ_s^2 and σ_{s+1}^2 are the variances in intensity value for two m^*n blocks centered on p_s^i and p_{s+1}^j, respectively. σ_s^2 and σ_{s+1}^2 are computed using:

$$\sigma_k^2 = \sum_{j=1}^{n}\sum_{i=1}^{m}\{I_k(i,j) - \mu_k\}^2/(\mathrm{mn}) \quad \text{for } k = s, \text{ or } s+1, \tag{14}$$

and $\sigma_{s,s+1}^2$ is covariance of I_s and I_{s+1}, and can be computed using:

$$\sigma_{s,s+1}^2 = \sum_{j=1}^{n}\sum_{i=1}^{m}[\{I_s(i,j) - \mu_s\}\{I_{s+1}(i,j) - \mu_{s+1}\}]/(\mathrm{mn}). \tag{15}$$

For a continuous parametric curve $C(x)$, we can compute its unit tangent vector $T(x)$ using the following:

$$T(x) = \frac{\frac{dC(x)}{dx}}{\left\|\frac{dC(x)}{dx}\right\|} = \frac{C'(x)}{\|C'(x)\|}. \tag{16}$$

The geometric similarity of $C_s(p_s^i)$ and $C_{s+1}(p_{s+1}^j)$ is evaluated using the inner product of $T_s(p_s^i)$ and $T_{s+1}(p_{s+1}^j)$. This inner product is denoted using:

$$\langle T_s(p_s^i), T_{s+1}(p_{s+1}^j)\rangle. \tag{17}$$

The basic idea is that when two curves are matched, each correspondence pair (p_s^i, p_{s+1}^j) has an equal tangent vector (i.e. inner product is equal to 1). Therefore, when all of the equal tangent vectors between two curves are found, the best match occurs when the sum of Eq. (17) for all correspondences is the maximum. After the contour simplification discussed in Sec. 3.4.2, assume that both $C_s(u)$ and $C_{s+1}(v)$ consist of m points and each point is denoted as $C_s(p_s^i)$ and $C_{s+1}(p_{s+1}^j)$, where

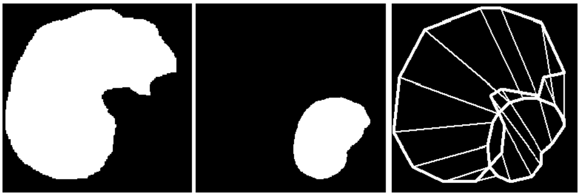


Fig. 9. Contour matching: (a) and (b) are the input contours on I_s and I_{s+1}. Each corresponding point pair is shown using a line connecting two corresponding points on two contours in (c).

$1 \leq i,\, j \leq m$. To find the best matched points between $C_s(u)$ and $C_{s+1}(v)$, the solution is computed using the following:

$$\max_{j(i)} \sum_{i=1}^{m} \{w_1 * \langle T_s(p_s^i), T_{s+1}(p_{s+1}^{j(i)})\rangle + w_2 * C(p_s^i, p_{s+1}^{j(i)})\}. \tag{18}$$

To evaluate Eq. (18), $C_{s+1}(p_{s+1}^j)$ is re-parameterized using $C_{s+1}(p_{s+1}^{j(i)})$ to find the solutions. Where w_1 and w_2 are the weights for the intensity and geometry similarities. In addition, the re-parameterization must be subject to $j(1) = 1$, $j(m) = m$ and $j(i) \leq j(i+1)$. In here, the dynamic programming can be adopted to solve this optimization problem. First, a cost function is defined as (19)

$$\begin{aligned} Cost(i,j) &= \min(Cost(i-1,j-1), Cost(i-1,j), Cost(i,j-1)) \\ &\quad + w_1 * \langle T_s(p_s^i), T_{s+1}(p_{s+1}^j)\rangle + w_2 * C(p_s^i, p_{s+1}^j). \end{aligned} \tag{19}$$

The meaning of $Cost(i,j)$ is interpreted as the optimal cost for matching the first i samples of $C_s(u)$ with the first j samples of $C_{s+1}(v)$. The time complexity of $Cost(i,j)$ is $O(ij)$ using the dynamic programming approach. Figure 9 shows an example of matching two curves using the proposed method. In this example, each corresponding point pair is shown using a line connecting two corresponding points. After the matching task, these matched points will define corresponding points between the two contours and two consecutive feature-points from a feature line-segment on each contour.

3.5. *Solving interpolation problems*

The example shown in Fig. 10 is a brief overview of the feature-guided shape-based interpolation scheme. In this scheme, the distance maps D_s (Fig. 10(a)) and D_{s+1} (Fig. 10(b)) for two input contours C_s and C_{s+1} on image slices I_s and I_{s+1} are created first. The feature line-segments are then computed from the algorithms presented in Sec. 3.4 to generate two warping distance maps D'_s (Fig. 10(c)) and D'_{s+1} (Fig. 10(d)). Next used Eq. (9) to linearly interpolate any number of intermediate distance maps.

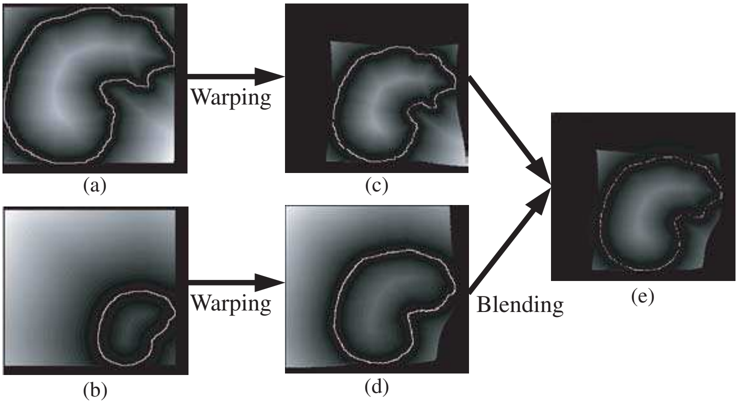


Fig. 10. The warping procedures to interpolate contours.

As mentioned earlier, this interpolation scheme separately interpolates positive object pairs and negative object pairs. Therefore, these objects must be combined using:

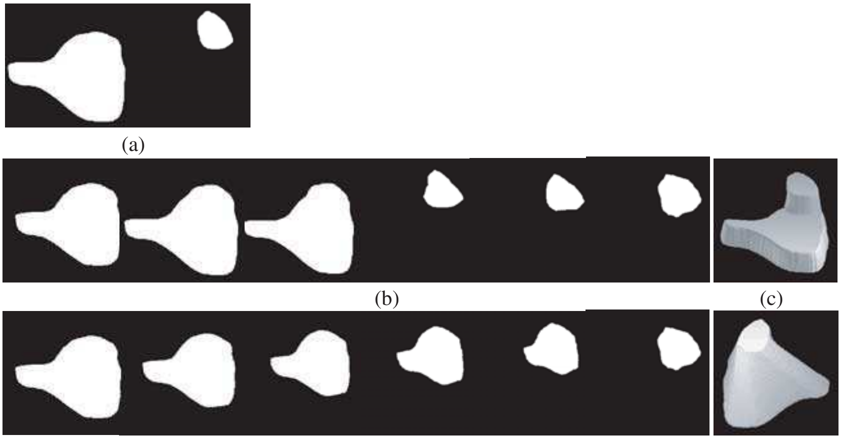
$$\begin{cases} X_R^+ = \bigcup X_i & \text{if } X_i \in \textit{positive object} \\ X_R^- = \bigcup X_i & \text{if } X_i \in \textit{negative object}. \\ X_R = X_R^+ - X_R^- \end{cases} \tag{20}$$

In (20), objects X_R^+ and X_R^- are blended results for the positive and negative objects, respectively. This equation defines the *blending order*: Blend all positive and negative objects separately and then subtract X_R^- from X_R^+.

4. Experimental Results and Discussion

Figure 11 shows a comparison between the interpolation result generated by the shape-based interpolation method and that generated by the feature-guided shape-based interpolation scheme. In this case, the shape-based interpolation method does not perform well due to the matched objects with large offsets. In the morphology-based interpolation, the object centralization is used to have one object enclosed by another before interpolation.[6,7] However, the conventional centralization (i.e. aligning the centroids of the two objects) sometimes failed when the objects are concave. In this issue, the feature-guided interpolation method can manage this problem well. This method interpolates the slices with a warping function described in Sec. 3.3. Therefore, this method has the ability to manage the matched object with large difference.

In the following examples, all corresponding point pairs are shown using lines connecting the corresponding points among the contours. Figure 12 shows an example of the object slices with holes. In this example two contour pairs are created: A positive pair for the outer contours and a negative pair for the inner

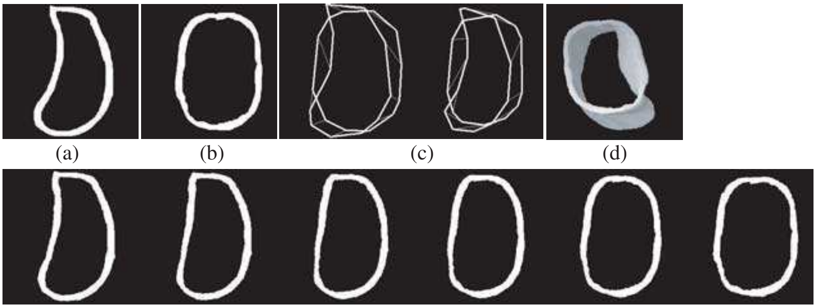


(a)

(b) (c)

(d) (e)

Fig. 11. (a) is the original frames; (b) is the interpolation result generated by the shape-based interpolation method; (d) is the interpolation result generated by the feature-guided shape-based interpolation scheme; (c) and (e) are the rendered results of the (b) and (d), respectively.



(a) (b) (c) (d)

(e)

Fig. 12. Objects with a hole: (a), (b) are the original frames, (c) left: Positive pair and (c) right: Negative pair; (d) the rendered result of interpolated volume and (e) shows a sequence of interpolated object.

hole-contours. The feature-guided interpolation scheme will generate two interpolated contours: One for a positive pair and one for a negative pair. The two contours are then blended to generate a desired contour. Figure 13 shows a branching case. In this example, there are three matched pairs. The feature-guided interpolation scheme will independently interpolate these three positive object pairs. The final contour can then be reconstructed from the union of these three interpolated contours.

Figure 14 shows the interpolation result using the feature-guided method to real 3D molar data. The number of the original slices of the molar volume is 16 as shown in Fig. 14(a). Figure 14(b) shows two views of the reconstructed results.

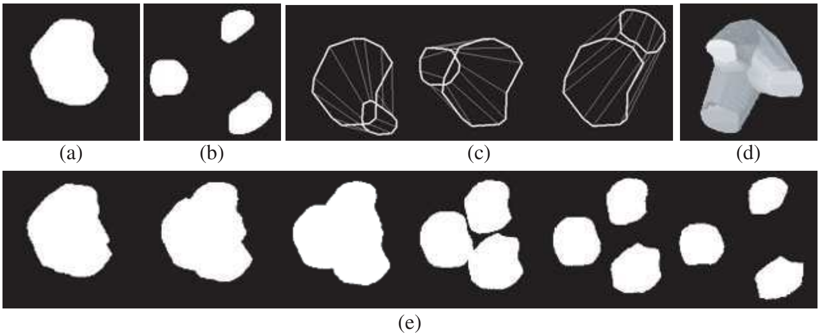


Fig. 13. Branching case: (a) and (b) original slices, (c) matched object pairs, (d) the rendered result of interpolated volume and (e): a sequence of interpolation.

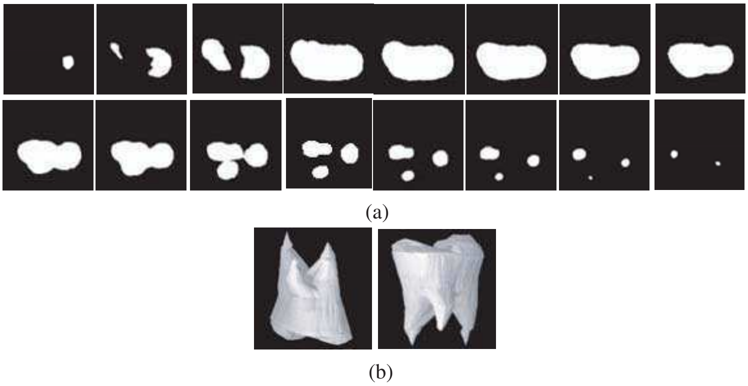


Fig. 14. Molar volume reconstruction. Original input images were scanned from V. Chatzis and I. Pitas.[11]

Acknowledgments

The authors would like to thank the National Science Council, ROC for supporting this work under Grant Nos. NSC-93-2213-E-006-026, NSC-93-2213-E-006-060 and NSC 94-2213-E-006-005. The authors also would like to thank V. Chatzis and I. Pitas,[11] since they scanned their test input images to conduct experiments.

References

1. C. C. Liang, C. T. Chen and W. C. Lin, Intensity interpolation for reconstructing 3D medical images from serial cross-sections, in *Proc. IEEE Eng. Med. Bio. Soc. 10th Int. Conf.*, New Orleans, LA **CH2566-8** (1988) 1389–1390.
2. G. J. Grevera and J. K. Udupa, Shape-based interpolation of multidimensional grey-level images, *IEEE Trans. Med. Imag.* **15** (1996) 881–892.
3. K. Chuang, C. Chen, L. Yuan and C. Yeh, Shape-based grey-level image interpolation, *Phys. Med. Biol.* **44** (1999) 1565–1577.

4. S. P. Raya and J. K. Udupa, Shape-based interpolation of multi-dimensional objects, *IEEE Trans. on Medical Imaging* **9**, 1 (1990) 32–42.
5. G. T. Herman, J. Zheng and C. A. Bucholtz, Shaped-based interpolation, *IEEE Computer Graphics & Applications* (1992) pp. 69–79.
6. J.-F. Guo, Y.-L. Cai and Y.-P. Wang, Morphology-based interpolation for 3D medical image reconstruction, *Computerized Medical Imaging and Graphics* **19**, 3 (1995) 267–279.
7. T.-Y. Lee and W.-H. Wang, Morphology-based three-dimensional interpolation, *IEEE Transactions on Medical Imaging* **19**, 7 (2000) 711–721.
8. T.-Y. Lee and C.-H. Lin, Feature-guided shape-based image interpolation, *IEEE Transactions on Medical Imaging* **21**, 12 (2002) 1479–1489.
9. T. Berier and S. Neely, Feature-based image metamorphosis, *In ACM SIGGRAPH Conference Proceedings* (1992) pp. 35–42.
10. M. Salotti, An efficient algorithm for the optimal polygonal approximation of digitized curves, *Pattern Recognition Letters* **22** (2001) 215–221.
11. V. Chatzis and I. Pitas, Interpolation of 3D binary images based on morphological skeletonization, *IEEE Transactions on Medical Imaging* **19** (2000) 699–710.

This page intentionally left blank

CHAPTER 10

NON-PARAMETRIC PIXEL APPEARANCE PROBABILITY MODEL USING GRID QUANTIZATION FOR LOCAL IMAGE INFORMATION REPRESENTATION

MINGZHOU SONG
Department of Computer Science, New Mexico State University
P.O. Box 30001, MSC CS, Las Cruces, NM 88003, USA
joemsong@cs.nmsu.edu

ROBERT M. HARALICK
Doctoral Program in Computer Science, Graduate Center, City University of New York
365 Fifth Ave., New York, NY 10016, USA
haralick@gc.cuny.edu

We describe a non-parametric pixel appearance probability model to represent local image information. It allows an optimal image analysis framework that integrates low- and high-level stages to substantially improve overall accuracy of object reconstruction. In this framework, feature detection would be an overall consequence rather than an intermediate result. The pixel appearance probability model is a probability density function obtained by grid quantization. A grid is found by a genetic algorithm and a local refinement algorithm. The density values in each cell of the grid are computed by smoothing neighboring cells. We apply the pixel appearance probability model to represent features of echocardiographic images. We illustrate the substantially improved performance on left ventricle surface reconstruction due to the proposed model.

Keywords: Non-parametric pixel appearance probability model; grid quantization; ultrasound imaging.

1. Introduction

The ultimate goal of medical image analysis is to acquire quantitative representations of objects that are of medical concern from observed images. In echocardiography, one objective is to create a three-dimensional (3D) Left Ventricle (LV) surface model, including the EPIcardium (EPI), the outer surface of the LV, and the ENDOcardium (ENDO), the inner surface of the LV. A standard two-stage approach comprises feature detection and object reconstruction. Feature detection classifies each pixel into categories such as edges and regions. Established techniques based on pixel intensities and their derivatives, known as low-level operators, are able to reliably perform feature detection on images of good quality. Detected features are fed into an algorithm for object reconstruction, known as high-level methods. Such a two-stage framework is computationally efficient and is feasible when image quality is high. However, noisy imagery is particularly common in ultrasound imaging. Under uncertainty due to low signal-to-noise ratio, low-level feature detection by inspecting only the local neighborhood of each pixel is inaccurate. To overcome the unreliable feature detection, we advocate an image analysis framework that integrates low- and high-level stages to substantially improve overall accuracy of object reconstruction.

Rather than using detected features to represent information in images, we will adopt a much richer representation of the low-level information using a grid quantization technique. Statistical and computational considerations lead to this choice. This representation is non-parametric and carries much less biases than a parametric one such as an appearance-based model using a multivariate normal distribution. Grid quantization alleviates the online computation and storage burden of standard non-parametric ones. A grid can be trained either in conjunction with a high-level model or directly from low-level groundtruth.

The grid quantization technique is used to obtain the pixel appearance probability model in three steps. In the first step, a global grid is found by a genetic algorithm; in the second step, the grid is refined by a fast local algorithm; in the last step, probability density values are obtained for each cell in the grid.

Training of the pixel appearance probability model usually involves much more than a single run of grid quantization, when pixel classification is not available or unreliable. We provide a generalized EM algorithm to handle such situations when only images and object models are available with no edge information.

We applied the pixel appearance probability model in reconstructing the 3D left ventricle surface model from 2D images. We reduced the reconstruction error by about 2.6 mm when compared with a standard two-stage approach.

In this chapter, we will review related work to pixel appearance probability models in Sec. 2. We describe an integrated framework in Sec. 3 and explain the role of a pixel appearance probability model in the integrated framework. In Sec. 4, we deal with the technicalities of grid quantization. We introduce a pixel appearance probability model for echocardiographic images in Sec. 5 and a pixel class prediction probability model in Sec. 6. A method to train the two models is given in Sec. 7, when low-level edge information is not available. We illustrate the performance of 3D left ventricle surface reconstruction in Sec. 8. In Sec. 9, we summarize this chapter.

2. Related Work

Chakraborty *et al.* integrate gradient and region information when performing pixel classification.[1] The region information is modeled by Markov random fields without shape statistics applied. Cootes *et al.* model image pixels by the statistical active appearance models,[2,3] which are equivalent to parametric multivariate normal probability models. They combine the active appearance models and the active shape models to find the best 2D contour from images. Pixel classification decisions are still made but might change during the model fitting iterations. This iterative classification scheme allows shape statistics to guide local feature detection, though it may not necessarily be optimal in the Bayesian sense. As far as we know in the literature, only Mignotte and Meunier have explored the idea of modeling shapes from images without an explicit feature detection stage.[4] Their treatment is more or less intuitive and lacks for a systematic account. By contrast, we will use the Bayesian framework to integrate the low-level image information and high-level prior shape knowledge through a pixel class prediction mechanism.

To represent image feature vectors, parametric statistical models are widely used in general and Gaussian distributions in particular. However, in a noisy imaging environment, Gaussian models or other simple parametric models are not effective because of their large modeling biases. Standard non-parametric kernel methods are seldom employed owing to their low efficiency when dealing with millions of feature vectors, which is not an especially large sample size for pixel based applications. An alternative to kernel methods is quantization, which is a function that maps a larger set to a smaller one. Two steps, discretization and representation, are performed during quantization. The discretization step determines the size and shape of the cells by partitioning the larger set into smaller subsets. The representation step assigns summary statistics to each cell.

For discretization, equal bin width histograms or their simple extension in multidimensions have often been adopted, because they are computationally efficient. However, they are not statistically effective for two reasons. First, the cells are blindly allocated in advance, not adapting to the data. Second, the normalized frequency may be zero for many cells and there is no guarantee for the consistency of the density estimates. The statistically equivalent blocks approach is an early tree-structured partitioning scheme.[5] The CART algorithm is a tree-structured classifier.[6] Grid-based partition schemes are studied, e.g. multivariate ASH,[7] STING algorithm,[8] OptiGrid algorithm,[9] STUCCO algorithm[10] and Adaptive Grids.[11] All these multivariate discretization approaches are sub-optimal algorithms; most of them are greedy. For tree-based algorithms, it is evident since a tree is constructed level by level using a certain greedy method and does not possess a global optimal measure. For grid-based algorithms, a grid can be acquired randomly[12,13]; the grid lines can be equally spaced[7]; a grid is improved by merging adjacent intervals by hypothesis testing[10]; the adaptive grids technique merges dense cells.[11]

Splines or local polynomials have been used to delineate a probability density function over each cell, but they are not computationally efficient in a multidimensional space. On the other hand, using the empirical density directly leads to empty cells. Existence of empty cells is particular prominent in a high dimensional space, due to the increased sparsity of data. WARPing[14] and averaging shifted histograms[15] smooth cell density estimates. Otherwise there are relatively few methods. We will present a closely related smoothing method for grid quantization.

3. A Framework to Integrate Low- and High-Level Stages

Although the pixel appearance probability model to be discussed can be used separately from high-level object reconstruction, it is best understood under a framework that integrates the low- and high-level stages. We use a Bayesian framework to formulate the overall object reconstruction problem:

$$\max_{\Theta} p(\Theta|Z) \quad \text{(MAP Rule)},$$

where Θ is the object model, Z is the feature vector of a pixel, $p(\Theta|Z)$ is the posterior probability of Θ given Z. This formulation is known as the maximum *à posteriori* (MAP) rule. By Bayes' Theorem, the MAP rule is equivalent to

$$\max_{\Theta} p(\Theta|Z) = \max_{\Theta} \frac{p(\Theta)p(Z|\Theta)}{p(Z)} \propto \max_{\Theta} p(\Theta)p(Z|\Theta), \tag{1}$$

where $p(\Theta)$ is the prior probability of object model Θ and $p(Z|\Theta)$ is the conditional probability of observed feature vector Z given an object model Θ. We also call $p(Z|\Theta)$ the object appearance model, capturing the overall imaging system behavior.

The prior probability $p(\Theta)$ is application-specific. For a surface object model, it can be the prior probability characterizing smoothness, or the shape of the objects, or several user input points. $p(Z)$ is the probability of observing a particular feature vector Z, the knowledge of which is only necessary when the exact posterior probability is desired. Computing $p(Z|\Theta)$ directly is difficult because of the many degrees of freedom of the feature vector Z.

Feature detection precisely avoids finding the functional form of $p(Z|\Theta)$. We use Y to denote the class label of a pixel, marking each pixel to be visible as either on or off the object. If we can detect the class labels for each pixel, we can search for an object model Θ^* that fits the class labels best, instead of fitting to the original images. These two stages, i.e. feature detection and model fitting, form a standard approach of object reconstruction, summarized as Algorithm 1. $P(Y|Z)$ is the posterior of a class label Y given the feature vector Z. $p(Z|Y)$ is the likelihood of the class label Y for the feature vector Z. $p(\Theta|Y)$ is the posterior of the object model Θ given the class label Y. $P(Y|\Theta)$ is the likelihood of the object model Θ for the class label Y. Estimation of $p(Z|\Theta)$ is unnecessary in this framework. Instead, we estimate $p(Z|Y)$ and $P(Y|\Theta)$, as well as apply suitable priors $P(Y)$ and $p(\Theta)$. However, if we have to detect class labels on fuzzy images, the two-stage framework does not yield an optimal Θ^* because the detected class label Y^* from the first stage may be unreliable.

```
**Algorithm 1** Two-stage object reconstruction.
Stage 1. Feature detection. Find $Y^*$ that solves
    $\max_{Y} P(Y|Z) \propto \max_{Y} P(Y)p(Z|Y)$;
Stage 2. Model fitting. Find $\Theta^*$ that solves
    $\max_{\Theta} p(\Theta|Y^*) \propto \max_{\Theta} p(\Theta)P(Y^*|\Theta)$.
```

In the integrated approach to be proposed, we will not assign a class label to each pixel, but will profile each pixel by probabilities of having different class labels. We still need the class label Y to avoid direct off-line estimation and online computation of $p(Z|\Theta)$. Being not directly observable, a class label serves as a hidden bridge between the feature vector Z and the object model Θ. Under the assumption that an object model Θ_1 inferred from both the feature vector Z and the class label Y has the same probability distribution as the object model Θ_2 inferred from only the class label Y, we can obtain the following theorem.[16]

Theorem 1. (Integrated object inference) The posterior probability of an object model Θ given the observed feature vector Z can be written as

$$p(\Theta|Z) = \frac{p(\Theta)}{p(Z)} \sum_{y=1}^{K} P(Y = y|\Theta)p(Z|Y = y), \tag{2}$$

with the assumption $p(\Theta|Z, Y) = p(\Theta|Y)$. K is the number of pixel classes.

The integrated object inference theorem leads to the *integrated approach of object model optimization*

$$\Theta^* = \operatorname*{argmax}_{\Theta} p(\Theta) \sum_{y} P(Y = y|\Theta)p(Z|Y = y), \tag{3}$$

where Θ^* is an optimal object model. The interpretation behind the integrated approach is as follows. Every image pixel is assigned a likelihood profile of being different classes $p(Z|Y)$. Another class probability $P(Y|\Theta)$ profile is predicted from an object model. When the likelihood profile $P(Y|Z)$ and the predicted class probability profile $P(Y|\Theta)$ match well, the object model that generates the predicted class probability profile is a good explanation of the images. If $p(Z)$ can be calculated, we can get the posterior $p(\Theta^*|Z)$ which indicates the goodness of the solution Θ^*. When $P(Y|Z)$ has a single narrow peak, the maximization of $p(\Theta|Z)$ can be approximated by the two-stage approach. However, the two-stage approximation generally is not optimal in the sense of the Bayesian framework and may lead to less consistent results. We call $P(Y|\Theta)$ the *Pixel Class Prediction* (PicPre) probability model, meaning that the class label Y can be predicted probabilistically from a given object model Θ. We call $p(Z|Y)$ the *Pixel Appearance* (PixApp) probability model, because $p(Z|Y)$ depicts probabilistically the appearance of a pixel with the class label Y.

4. Grid Quantization

Grid quantization we describe is a non-parametric statistical pattern recognition technique. It partitions the space into hyper-rectangular cells shown in Fig. 1 and estimates the probability density for each cell. Non-parametric methods do not necessarily guarantee a minimum variance estimate. However, it does not have the potentially large modeling biases inherent in parametric models that do not fit reality. Non-parametric models usually require a larger sample size than parametric models, since the degrees of freedom of non-parametric models are usually much greater than parametric models. However, a larger sample size almost always implies more CPU cycles and memory capacity. Best control variables in a non-parametric model are typically determined during the time consuming off-line training. The online CPU cycles and the memory requirement are evidently high for standard non-parametric models, which might be unacceptable in some applications. For example, a kernel method often estimates the density value for a single point from the sample directly, typically involving the entire data with the implication that time and space are at least in proportion to the sample size.

Grid quantization alleviates the online computation and storage burden of standard non-parametric techniques. Grid quantization finds the most effective non-parametric representation of the data, for given computation resources, in terms of CPU cycles, memory requirement and the targeted performance. It requires an intensive off-line training in order to determine the best representation. Grid quantization possesses the scalability to trade more resources for performance. In contrast, it is prohibitive to scale most other non-parametric models. A grid quantization locates the most important regions in the feature space which are then finely quantized, while unimportant regions are coarsely quantized. What determines

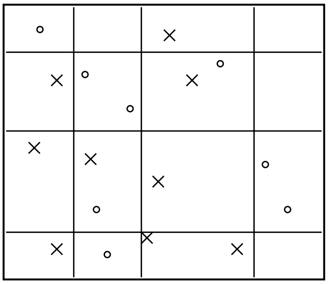


Fig. 1. The grid partition pattern.

importance relies on the pattern recognition task. We will use a convex combination of average log likelihood, entropy, classification accuracy. Kernel methods treat equally everywhere in the space and for unimportant regions, there exists the potential of wasting resources. A concern might arise for the quantization effect, but to accomplish acceptable results does not necessarily entail infinite resolution quantization. Since a grid is constructed by adapting to the data, the quantization effect can be minimized for a particular task.

We choose grid quantization pattern because it is a reasonable tradeoff between computational and statistical efficiencies. Other options are less attractive. For an equal spacing grid, retrieval efficiency is linear in the dimension and does not depend on the number of cells. However, an equal spacing grid may not be statistically efficient. Variable spacing grid lines can dramatically improve the statistical efficiency while having low computational complexity. A tree structured quantization pattern can have very high statistical efficiency and low computational complexity, but to optimize a tree for a global criterion is a hard problem. Almost all tree structured quantizers use greedy approaches. An irregular quantization pattern can further improve statistical efficiency, but the computational complexity could be much higher — more expensive to use than a kernel method. Another consideration to choose a grid pattern is that smoothing can be carried out efficiently. The purpose of smoothing is to improve the generalization ability of a quantizer. Smoothing of a grid quantizer is analogous to pruning a tree structured quantizer.

A quantizer is constructed off-line on training data by optimizing a certain performance measure. Training is two fold. The first is to achieve as good performance as possible on the training data. The second is to obtain a consistent density estimate of each cell by smoothing. Our smoothing strategy is a close approximation to the k_N spacing approach. The latter subdivides the space into cells such that each cell contains k_N data points. It has been shown that the k_N density estimate is both L_1[17] and L_2[15,18] consistent. A quantizer is used online by table-lookup. A cell is located for a given data vector and then the density value of that cell is returned.

4.1. *Notations, definitions and problem statement*

Let the random vector $X \in \mathbb{R}^D$ represent an individual data or pattern. $X(d)$ is the dth dimension random variable of X. We call the sequence $\mathcal{X}_N = \{x_1, x_2, \ldots, x_N\}$ a data set or a sample of size N. This set contains i.i.d. data vectors from x_1 to x_N. Let K be the total number of classes and $\{1, 2, \ldots, K\}$ be the class label set. Let random variable $Y \in \{1, 2, \ldots, K\}$ be the class assignment of X. We call the sequence $\mathcal{Y}_N = \{y_1, y_2, \ldots, y_N\}$ the class assignment set of $\mathcal{X}_N$, where x_1 has class label y_1, x_2 has class label y_2 and so on. We also consider a more general case where the class assignment is not exclusive, but instead weighted. Let random vector $W \in [0,1]^K$ be the weighted class assignment vector. $W(y)$ is the weight for class y.

We also require $\sum_{y=1}^{K} W(y) = 1$. We call the sequence $\mathcal{W}_N = \{w_1, w_2, \ldots, w_N\}$ the weighted class assignments of the data set $\mathcal{X}_N$.

Definition 2. A *quantizer* Q is a function that maps $\mathbb{R}^D$ to $\mathcal{I}$. $\mathcal{I}$, called quantization index set, is a finite set of integers or integer vectors.

Definition 3. A *quantization cell* of Q is a set $\mathcal{C} \subset \mathbb{R}^D$ such that for any $x_1, x_2 \in \mathcal{C}$, $Q(x_1) = Q(x_2)$.

For each cell, we assign an index $q \in \mathcal{I}$ to it. We use the function notation $q = Q(x)$ to denote that x is quantized to q or x belongs to cell q. We use $N(y)$ to denote the total number of class y data in $\mathcal{X}_N$, that is

$$N(y) = \sum_{n=1}^{N} w_n(y).$$

Let N_q be the total number of data in cell q. Let $N_q(y)$ be the total number of data of class y in cell q, that is

$$N_q(y) = \sum_{n \in \{n | q = Q(x_n)\}} w_n(y).$$

Let L be the number of cells. The index to the first cell is $q = 1$, and the last cell $q = L$.

Let $p_{X|Y}(x|Y = y)$ be the class y conditional probability density function (p.d.f.). We use $\hat{p}_{X|Y}(x|Y = y)$ to denote the estimated p.d.f. In most cases, we shall drop the subscript $X|Y$ from $p_{X|Y}(x|Y = y)$ and $\hat{p}_{X|Y}(x|Y = y)$ to simplify the notation. The overall grid quantization problem is to estimate the p.d.f.s

$$p_{X|Y}(x|Y = 1),\ p_{X|Y}(x|Y = 2),\ \ldots\ ,\ p_{X|Y}(x|Y = K),$$

such that a certain quantizer performance measure $T(\ )$, which we will define later, is maximized for the training data $\mathcal{X}_N$, $\mathcal{Y}_N$ or $\mathcal{W}_N$. It is equivalent to solving

$$\max_{Q} T(\mathcal{X}_N,\ \mathcal{Y}_N) \quad \text{or} \quad T(\mathcal{X}_N,\ \mathcal{W}_N). \tag{4}$$

4.2. *The performance measure of a quantizer*

The quantizer performance measure includes three components: log likelihood, entropy and correct classification probability. We explain each of them as follows.

4.2.1. *Log likelihood*

Kullback-Leibler divergence from $\hat{p}(x)$ to $p(x)$ is

$$D(p||\hat{p}) = \int p(x) \log \frac{p(x)}{\hat{p}(x)}\, dx = \mathbf{E}[\log p(X)] - \mathbf{E}[\log \hat{p}(X)],$$

which, being non-negative (zero if and only if $\hat{p}(x) = p(x)$), should be minimized. As $p(x)$ is fixed, minimizing $D_{KL}(p\|\hat{p})$ is equivalent to maximizing $\mathbf{E}[\log \hat{p}(X)]$. Let $p(q|y)$ be the density of cell q. Then $\mathbf{E}[\log \hat{p}(X|Y)]$ can be estimated by $\frac{1}{N(y)} \log \prod_{q=1}^{L} (p(q|y))^{N_q(y)}$. The *overall average log likelihood* of a quantizer Q is

$$J(Q) = \frac{1}{N} \sum_{y=1}^{K} N(y) \mathbf{E}[\log \hat{p}(X|y)] = \frac{1}{N} \sum_{y=1}^{K} \log \prod_{q=1}^{L} (p(q|y))^{N_q(y)}.$$

When the class number ratio $N(1) : N(2) : \cdots : N(K)$ is representative for the data population, the overall average log likelihood is preferred, with the log likelihood of popular classes being emphasized.

The *mean class average log likelihood* is

$$J(Q) = \frac{1}{K} \sum_{y=1}^{K} \mathbf{E}[\log \hat{p}(X|y)] = \frac{1}{K} \sum_{y=1}^{K} \frac{1}{N(y)} \log \prod_{q=1}^{L} (p(q|y))^{N_q(y)}.$$

When the class number count $N(y)$ is randomly decided or every class is considered to have equal importance, the mean class average log likelihood is preferred, with every class contributing equally to the log likelihood of the quantizer.

4.2.2. *Correct classification probability*

Let $P(y)$ be the prior probability of class Y. Within cell q, the Bayes' rule is equivalent to

$$y_q^* = \operatorname*{argmax}_{y} P(y) \frac{N_q(y)}{N(y)}.$$

Let $N_c(q)$ be the number of correct decisions in cell q, i.e. $N_c(q) = N_q(y_q^*)$. We define the correct classification probability in two situations. The *overall correct classification probability* is

$$P_c(Q) = \frac{\sum_{q=1}^{L} N_c(q)}{N}. \tag{5}$$

The *mean class correct classification probability* is

$$P_c(Q) = \frac{1}{K} \sum_{y=1}^{K} \sum_{q=1}^{L} I(y = y_q^*) \frac{N_c(q)}{N(y)}. \tag{6}$$

In the above equation, I is indicator function. The choice of (5) or (6) should follow the same considerations for the choice of average log likelihood.

4.2.3. *Entropy*

Similar to the case of average log likelihood, we give two options: Overall entropy and mean class entropy. Again, the choice of either should also follow the considerations for the choice in average log likelihood and correct classification probability. We define the *overall entropy* by

$$H(Q) = \frac{N_q}{N} \log \frac{N}{N_q}. \tag{7}$$

We define the *mean class entropy* by

$$H(Q) = \frac{1}{K} \sum_{y=1}^{K} \sum_{q=1}^{L} \frac{N_q(y)}{N(y)} \log \frac{N(y)}{N_q(y)}. \tag{8}$$

Entropy has been used as a class impurity measure. But we use entropy as a measure of the consistence or generalization ability of a quantizer.

4.2.4. *The performance measure function*

We define the quantizer performance measure function, by linearly combining average log likelihood, entropy and the log of correct classification probability, as follows

$$T(Q) = W_J J(Q) + W_H H(Q) + W_c \log P_c(Q), \tag{9}$$

where W_J, W_H and W_c are given non-negative weights for average log likelihood $J(Q)$, entropy $H(Q)$ and log of correct classification probability $\log P_c(Q)$, respectively.

4.3. ***Preprocessing***

The performance of classifiers does not monotonically increase with dimensions when sample size is fixed, because the required sample size would have grown exponentially to achieve a similar performance. Dimension reduction may be necessary. Popular techniques including principal component analysis, projection pursuit and independent component analysis. In addition to dimension reduction, for data in a high dimension, we may also want to view the data in a new coordinate system such that the most interesting dimensions come first, and use more quantization levels. Let $z_1, z_2, \ldots, z_N \in \mathbb{R}^M$ denote the feature vectors, each corresponding to a pixel. Normalization may also be necessary to make the dimension reduction sensible. We use matrix B to represent the normalization and matrix W to represent dimensional reduction and coordinate change. A feature vector z is projected to $x \in \mathbb{R}^D$ by $x = W^T Bz$.

4.4. ***Relative quantization levels***

The choice of a proper total number of quantization cells L depends on the sample size and the underlying distribution. It is also confined by available storage resource. When the sample size is large enough, it would be mostly determined by the available storage resource, as a small quantization cell is always preferred for asymptotic optimality, which means at least a large L. Once L is fixed, quantization levels $L_1, \ldots, L_D$ for each dimension are to be assigned. The information content of a random variable can be measured by its entropy. We assign the number of quantization levels for each dimension based on the marginal entropy in that dimension. We use the scale invariant portion of the continuous histogram entropy, or the discrete histogram entropy, to guide the assignment of quantization levels. Let $H_d(X)$ be the marginal histogram entropy for the dth dimension of X. The bit-allocation rule is defined by

$$\frac{H_1(X)}{\log L_1} = \frac{H_2(X)}{\log L_2} = \cdots = \frac{H_D(X)}{\log L_D}, \tag{10}$$

$$\log L_1 + \log L_2 + \cdots + \log L_D = \log L. \tag{11}$$

Solving the above equations for $L_d,\ d = 1, \ldots, D$, we get

$$L_d = L^{\frac{H_d(X)}{\sum_{m=1}^{D} H_m(X)}}, \quad d = 1, 2, \ldots, D. \tag{12}$$

4.5. *Obtaining a grid quantizer*

Having defined the quantizer performance measure function and applied certain pre-processing of the data, we will obtain a grid quantizer by three algorithms in turn. The first algorithm attempts to find a good grid in a global sense using a genetic algorithm. The second algorithm refines a grid locally by adjusting the grid lines one by one. The third algorithm obtains a smoothed density estimate for each grid cell.

4.5.1. *Finding a globally good grid*

We cast the grid optimization problem under a genetic algorithms model[19] in the following way. Each individual has a single chromosome, which is a grid. A gene is the decision boundaries in a particular dimension of the grid. A nucleotide is a single decision boundary. How well an individual adapts to the environment is measured by a fitness function. The choice of a fitness function depends on how selection is carried out. We use fitness proportionate selection, the chance of selecting individual being in proportion to its fitness. Therefore, we would normally require the fitness function be non-negative in order to directly relate it with probabilities.

Definition 4. The *fitness function* of a grid G is

$$\varphi(G) = \exp(T(G)) = [\exp(J(G))]^{W_J} [\exp(H(G))]^{W_H} [P_c(G)]^{W_c}. \tag{13}$$

$\varphi(G)$ is non-negative since it is a real exponential function of $T(G)$. $\exp(J(G))$ corresponds to the geometric average of likelihood. $\exp(H(G))$ is the information content of the grid expressed in terms of the size of a code book. $P_c(G)$ is exactly the correct classification probability. $\varphi(G)$ is a weighted geometric combination of the above components.

Algorithm 2 (Optimize-Grid-Genetic) outlines the main steps of the grid optimization genetic algorithm. It starts with an initial population of N_p random grids. The population evolves in the selection-reproduction-selection cycle as described in the **for** loop from line 2 to 16. In each cycle, or generation, N_p children are reproduced in the **for** loop from line 7 to 15. Every execution of the loop produces two children C_1 and C_2 by parents G_1 and G_2. G_1 and G_2 are randomly selected (line 8 and 9) from the population and the chance of their being selected is in proportion to their fitness function values. Selection is illustrated in Fig. 2. A cross-over site d_r is randomly decided for the parent chromosomes or grids G_1 and G_2. The cross-over occurs with a probability of P_r. A cross-over example of a grid is shown in Fig. 3. Once the cross-over is finished, two children C_1 and C_2 are produced (line 11). The

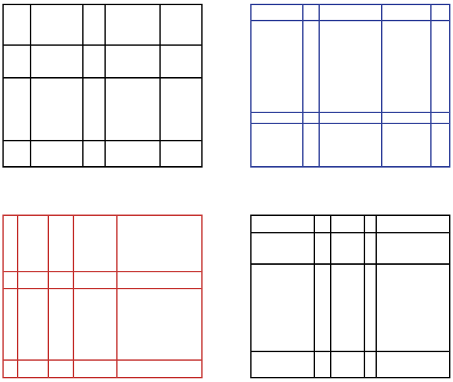


Fig. 2. Grid selection. In a population of four grids, two (lower left and upper right) are randomly selected as the parents by a chance in proportion to their fitness.

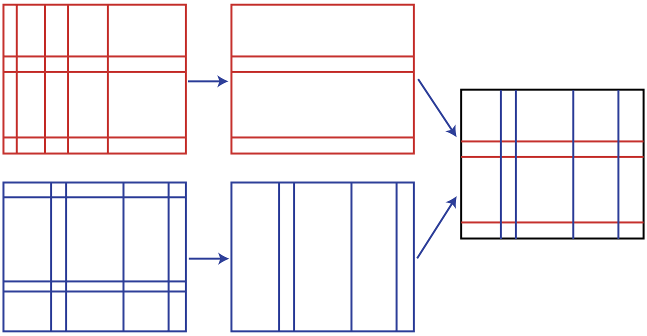


Fig. 3. Grid crossover. The quantization of the vertical axis in the top grid and the quantization of the horizontal axis of the bottom grid are crossed over to form a next generation grid on the right.

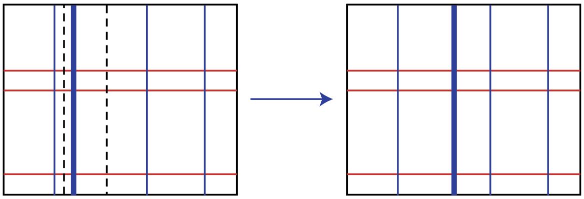


Fig. 4. Grid mutation. The thick vertical grid line could be mutated to the dashed lines on its sides. In this example it was mutated to the one on the right.

Algorithm 2 Optimize-Grid-Genetic($\mathcal{X}_N$, $\mathcal{Y}_N$, N_p, N_g, P_r, P_u)

```
 1: $\mathcal{P}^0 \leftarrow$ a population of $N_P$ random grids;
 2: for $j \leftarrow 0$ to $N_g - 1$ do
 3:    if $\varphi(G^*) < \max_{G \in \mathcal{P}^j} \varphi(G)$ then
 4:       $G^* \leftarrow \operatorname{argmax}_{G \in \mathcal{P}^j} \varphi(G)$;
 5:    end if
 6:    $\mathcal{P}^{j+1} \leftarrow \phi$;
 7:    for $i \leftarrow 0$ to $N_p - 1$ with increment of 2 do
 8:       Select a grid $G_1$ from $\mathcal{P}^j$ with a probability proportional to its fitness value;
 9:       Select a grid $G_2$ from $\mathcal{P}^j$ with a probability proportional to its fitness value;
10:       Randomly decide a dimension $d_r$ as the cross-over site;
11:       Exchange the decision boundaries of dimensions 1 to $d_r$ between
          $C_1$ and $C_2$ with probability $P_r$ or do not exchange;
12:       Mutation: Adjust each decision boundary of $C_1$ with probability $P_u$;
13:       Mutation: Adjust each decision boundary of $C_2$ with probability $P_u$;
14:       $\mathcal{P}^{j+1} \leftarrow \mathcal{P}^{j+1} \cup \{C_1, C_2\}$;
15:    end for
16: end for
17: if $\varphi(G^*) < \max_{G \in \mathcal{P}^{N_g}} \varphi(G)$ then
18:    $G^* \leftarrow \operatorname{argmax}_{G \in \mathcal{P}^{N_g}} \varphi(G)$;
19: end if
20: return $G^*$;
```

decision boundaries of each child grid are randomly changed by mutation (line 12 to 13), illustrated in Fig. 4. Then both children are added to the next generation (line 14). The best grid is kept as G^*, and is returned after a certain number of generations have evolved.

4.5.2. *Local refinement of a grid*

When the current best solution is close to the global optimal, genetic algorithms will be less efficient. Here we design a more efficient grid refinement algorithm by adjusting the decision boundaries one by one. An adjusted boundary is accepted only when the performance measure increases. The idea is explained in Fig. 5. By

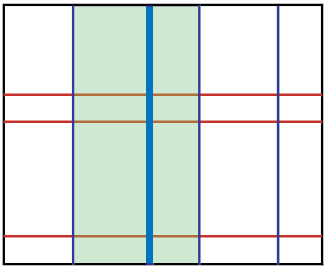


Fig. 5. Grid refinement. We adjust the thick grid line within the shaded area until a local maximum is reached.

definitions of $J(G), H(G)$ and $P_c(G)$, they are additive, i.e.

$$J(G) = \sum_{q=1}^{L} J(q), \quad H(G) = \sum_{q=1}^{L} H(q), \quad P_c(G) = \sum_{q=1}^{L} P_c(q).$$

Therefore, when one decision boundary of a grid is moved, we need only recalculate the change of the additive measures on those affected cells and data.

Algorithm 3 describes the Refine-Grid algorithm. The input includes a grid G, data sets $\mathcal{X}_N$ and $\mathcal{Y}_N$, number of refinements N_R and the convergence parameter δ. J, H, P_c and T record the performance measures of current grid. The refinement is done dimension by dimension in the **for** loop between line 6 and 20. The decision boundaries of each dimension are refined one by one in the **for** loop from line 7 to 19. A sub-grid G_s is formed by all the cells that is affected by a current decision boundary. The data that fall in the sub-grid are kept in sets $\mathcal{X}_{N'}$ and $\mathcal{Y}_{N'}$. Other data will not affect the refinement of the current decision boundary. The additive measures J^s, H^s and P_c^s are calculated for the sub-grid G_s. Line 12 finds the optimal decision boundary that maximizes the overall measure $T(G_s)$ and assigns the optimal boundary to the optimal sub-grid G_s^*. The changes in the additive measures of the sub-grid are calculated and the additive measures of the grid G are updated by the changes (line 13 to 16). The optimal decision boundary replaces the original one in G (line 17). The overall measure T is also updated. The refinement repeats until convergence is achieved as measured by $\Delta T/T$.

4.5.3. *Density estimation for each quantization cell*

The consistency of a quantizer is determined by the bias and variance of the density estimates. To reduce the bias, the sample size N must be large, and N_q/N of each cell must be small. To diminish the variance, the sample size N_q within a cell must be large. Overfitting manifests the large variance in cell density estimates, as a result of cell emptiness. Smoothing can reduce the variance of the density estimates. When smoothing is overdone, a quantizer gives very consistent result on both the training sample or an unseen sample, but carries large biases.

Algorithm 3 Refine-Grid(G, $\mathcal{X}_N$, $\mathcal{Y}_N$, N_R, δ)

```
$J \leftarrow J(G)$, $H \leftarrow H(G)$, $P_c \leftarrow P_c(G)$;
$T \leftarrow W_J J + W_H H + W_c \log P_c$;
$j \leftarrow 0$;
repeat
  $T^- \leftarrow T$;
  for $d \leftarrow 1$ to $D$ do
    for $q \leftarrow 1$ to $L_d - 1$ do
      Form a sub-grid $G_s$ by the cells sharing the decision boundaries $G[d, q]$;
      $G_s^* \leftarrow G_s$;
      $\mathcal{X}_{N'}, \mathcal{Y}_{N'} \leftarrow \{(x_n, y_n) | x_n(d) \in (G_s[d, 0], G_s[d, 2]]\}$;
      $J^s \leftarrow J(G_s)$, $H^s \leftarrow H(G_s)$, $P_c^s \leftarrow P_c(G_s)$, all on $\mathcal{X}_{N'}, \mathcal{Y}_{N'}$;
      $G_s^*[d, 1] \leftarrow \operatorname*{argmax}_{G_s[d,1]} T(G_s)$ on $\mathcal{X}_{N'}, \mathcal{Y}_{N'}$;
      $\Delta J \leftarrow J(G_s^*) - J^s$, $\Delta H \leftarrow H(G_s^*) - H^s$, $\Delta P_c \leftarrow P_c(G_s^*) - P_c^s$, all on $\mathcal{X}_{N'}, \mathcal{Y}_{N'}$;
      $J \leftarrow J + \Delta J$;
      $H \leftarrow H + \Delta H$;
      $P_c \leftarrow P_c + \Delta P_c$;
      $G[d, q] \leftarrow G_s^*[d, 1]$;
      $T \leftarrow W_J J + W_H H + W_c \log P_c$;
    end for
  end for
  $j \leftarrow j + 1$;
  $\Delta T \leftarrow T - T^-$;
until $j = N_R$ or $|\Delta T / T| < \delta$;
```

We offer a smoothing algorithm to assign a density estimate to a quantization cell. The algorithm is an approximation to the k nearest neighbor density estimate. Let $V(q)$ be the volume of cell q. Let $V_k(q)$ be the volume of a minimum neighborhood of cell q that contains at least k points. Let k_q be the actual number of points in the neighborhood. A smoothed probability density estimate of cell q is

$$p(q) = \frac{k_q}{V_k(q) \sum_r \frac{k_r}{V_k(r)} V(r)}. \tag{14}$$

Searching for the exact kth nearest neighbor is computationally expensive. We introduce an approximate, but no less than k, nearest neighbor smoothing algorithm after the following definitions.

Definition 5. We call cell a and b *neighbor cells* if they share at least a partial decision boundary.

Definition 6. The *radius* 0 *neighborhood of cell* q is a set that contains exactly the cell itself. We use $\mathcal{N}(q, 0)$ to denote the radius 0 neighborhood of cell q.

Definition 7. The *radius* R ($R \in \mathbb{Z}^+$) *neighborhood of cell* q, $\mathcal{N}(q, R)$, is the union of the radius $R - 1$ neighborhood $\mathcal{N}(q, R - 1)$, and the set of all the neighbor cells to those in $\mathcal{N}(q, R - 1)$.

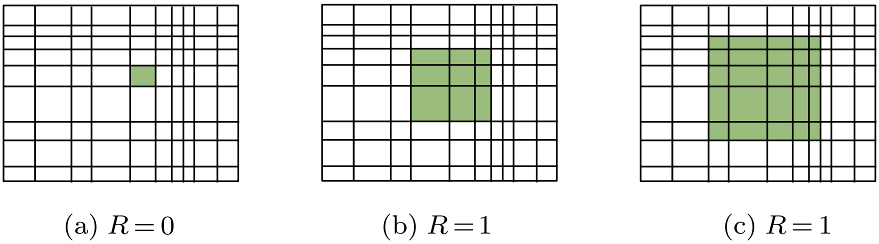

(a) $R=0$ (b) $R=1$ (c) $R=1$

Fig. 6. Radius R neighborhood of a cell.

Figure 6 shows the neighborhoods of a cell with different radii. In Fig. 6(a), the cell of interest is the cell in gray. The cell itself is also its radius 0 neighborhood. Figures 6(b) and (c) draw its radius 1 and 2 neighborhoods.

Definition 8. The *k neighborhood of a cell* is the smallest radius R neighborhood of the cell that contains at least k points.

Algorithm 4 Radius-Smoothing is based on the k neighborhood concept. The algorithm searches for a minimum radius R neighborhood with at least k data points for a current cell. The density of the k neighborhood is assigned to the cell as its density estimate. M is the total mass on the density support. M can be considered an adjusted data count by smoothing and is closely related to N. For cells with less than k data points, the initial guess of R is the radius of the k neighborhood of a previously processed neighbor cell. To make this initial guess more realistic, we shall go through the cells in an order that a pair of cells visited consecutively are neighbor cells. k_q is the actual number of data points in current radius R neighborhood. We either increase R until there are at least k data points in the neighborhood, or decrease R until the $R-1$ neighborhood contains less than k data points.

The extent of smoothing is usually controlled by k. We use cross-validation to determine an optimal control parameter k^* for smoothing, such that it maximizes the average quantizer performance.

5. A Pixel Appearance Probability Model for Echocardiographic Images

The appearance of a pixel is defined by its local information, such as intensity, contrast, directional derivatives, gradient, and etc. It is a result of the imaging process of a point on the object with some structural type corresponding to the pixel location. However, it may not be strictly a function of structural types. As we have discussed earlier, the PixApp probability model is used to capture such appearance uncertainty. We present in this section a PixApp probability model for echocardiographic image pixels.

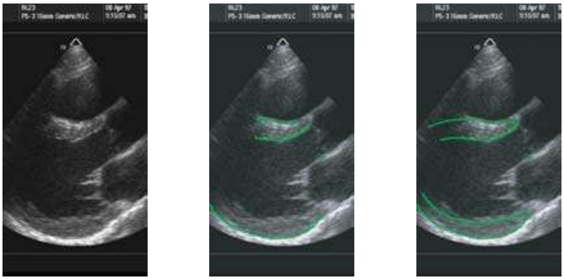
Algorithm 4 Radius-Smoothing(Q, k)

```
$M \leftarrow 0$, $R \leftarrow -1$;
for each cell $q$ do
  if $N(q) = k$ then
    $R \leftarrow 0$, $k_q \leftarrow N(q)$;
  else
    if $R < 0$ then
      $R \leftarrow 0$, $k_q \leftarrow N(q)$;
    else
      $\mathcal{A} \leftarrow \mathcal{N}(q^-, R) \cap \mathcal{N}(q, R)$;
      $k_q \leftarrow k_q - |\mathcal{N}(q^-, R) - \mathcal{A}| + |\mathcal{N}(q, R) - \mathcal{A}|$;
    end if
    while $k_q > k$ and $R > 0$ do
      $k_q \leftarrow k_q - |\mathcal{N}(q, R) - \mathcal{N}(q, R-1)|$, $R \leftarrow R - 1$;
    end while
    while $k_q < k$ and $R < R_{\max}$ do
      $k_q \leftarrow k_q + |\mathcal{N}(q, R+1) - \mathcal{N}(q, R)|$, $R \leftarrow R + 1$;
    end while
  end if
  $\rho(q) \leftarrow \frac{k_q}{V(\mathcal{N}(q,R))}$, $M \leftarrow M + \rho(q)V(q)$, $q^- \leftarrow q$
end for
for each cell $q$ do
  $p(q) \leftarrow \frac{\rho(q)}{M}$;
end for
```

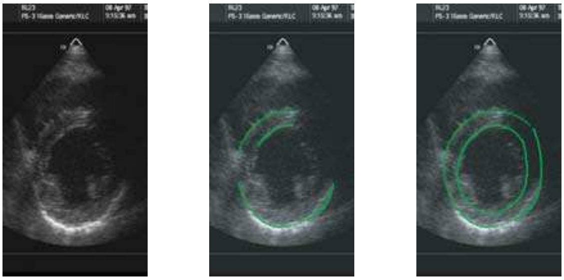
Figures 7(a), (d), (g) show ultrasound images of the left ventricle and Figs. 7(b), (e), (h) show the same images with the visible left ventricle boundary overlaid. Figures 7(c), (f), (i) overlay the complete contour of the underlying left ventricle surface on the original images. It is quite evident that pixels on the underlying surface contour do not have uniform appearance everywhere: Some pixels are bright with high contrast, while others do not differ too much from the background. The background pixels also have variable appearance.

During ultrasound imaging, signals arriving at an interface between media with different acoustic impedance produce strong echo when the angle of incidence is near perpendicular; signals arriving at an interface at near tangential angles produce very weak echo. Thus, the image intensity and its spatial variation are important. The derivatives carry spatial variation information. We fit a cubic facet model[20] to the pixel intensities in a local window centered at each pixel. A facet is a smooth surface patch in 3D. We analytically derive all first- and second-order derivatives for each pixel by fitting a facet centered at that pixel. Each pixel feature vector contains

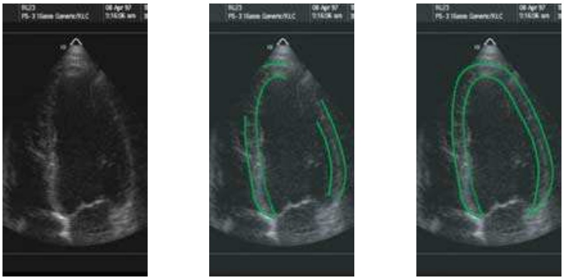
(i) the spatially and temporally smoothed pixel intensity value, capturing absolute strength of echo signals. A 3D median filter, with a window of 5 pixels × 5 pixels × 5 frames, is used to obtain the smoothed pixel values;



(a) Parasternal long axis view (b) With visible contours (c) With complete contours



(d) Short axis view (e) With visible contours (f) With complete contours

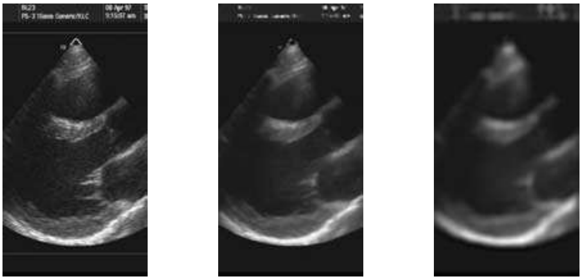


(g) Apical 4 chambar view (h) With visible contours (i) With complete contours

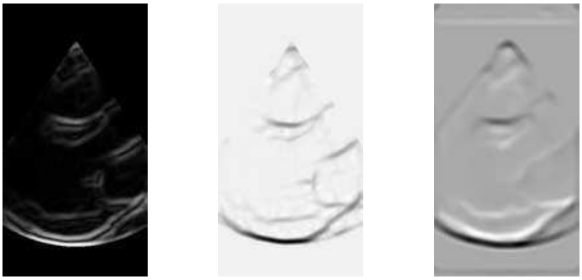
Fig. 7. Ultrasound images of the left ventricle and surface model contours.

(ii) the third-order facet model approximation of the pixel intensity value, giving the absolute strength of echo signals but on a larger scale. A window of 21 pixels × 21 pixels is used to compute each facet;

(iii) the directional derivative along the gradient direction, providing a local measure of edginess. A window of 21 pixels × 21 pixels is used to compute each facet;



(a) The original (b) Feature 1 map (c) Feature 2 map



(d) Feature 3 map (e) Feature 4 map (f) Feature 5 map

Fig. 8. An original parasternal long axis view image and its five feature maps.

(iv) the minimum second directional derivative, among second derivatives along all directions, indicating relative strength of echo signals. A window of 21 pixels × 21 pixels is used to compute each facet;

(v) the directional derivative from a point inside the LV, to help distinguish ENDO and EPI surfaces. The inner point is derived from user input points. A 3D median filter with a window of 17 pixels × 17 pixels × 17 frames is first applied to get the smoothed pixel values. Then a window of 21 pixels × 21 pixels is used to compute each facet.

Examples of the feature vector maps are shown in Fig. 8.

The PixApp probability model describes the pixel appearance uncertainty under noise for a given structural type. Let classes $Y = 1$ and $Y = 2$ correspond to EPI and ENDO pixels, respectively. An additional class $Y = 3$ labels the background. Thus, the PixApp probability model for echocardiographic images includes three p.d.f.s: $p(Z|1), p(Z|2)$ and $p(Z|3)$. We use grid quantization to represent them. Our technique requires a larger sample size in order to obtain a better representation for the pixel appearance than a multivariate normal model. For a multivariate sample of size on the order of millions, not uncommon for image pixels, we are reasonably grounded. The PixApp probability model can be estimated using the technique described in Sec. 4 directly if the pixel class information Y is available for each pixel. Section 7 introduces a strategy to estimate PixApp probability models without explicit knowledge of pixel classes.

6. A Pixel Prediction Probability Model for Ultrasound Imaging

When pixel classification is not available for observed training images, the PixApp probability model cannot be trained directly. If high-level object models are given for the observed training images, which is common in practice, a pixel class can be predicted probabilistically through the PicPre probability model using an object model. In this section, we describe a PicPre probability model for ultrasound imaging. In Sec. 7, we describe how this model is trained simultaneously with the PixApp probability model.

Pixel class prediction associates every pixel on an imaging plane with some physical properties of a given object model and its environment. Pixel class prediction and classification are fundamentally different. Pixel classification assigns class label information to each pixel based on observed images, not from an object model. We denote the deterministic prediction from an object model to pixel class by $Y|\Theta$, by which each pixel has an exclusive class assignment. We represent the probabilistic prediction by the conditional probability $P(Y|\Theta)$, which provides a soft pixel class prediction, allowing a more precise relationship to be captured.

Deterministic prediction methods can be considered special cases of the probabilistic prediction methods to be described. To predict the output of a physical system, we need to model both systematic and random behaviors of the system. Well understood systematic behaviors are often described by functional models. Less studied and more complex physical processes, often represented by probabilistic models, account for the random behaviors, as well as random noises from the environment. The overall probabilistic prediction is shown in the diagram in Fig. 9. A surface model Θ is used to produce simulated images via a physical simulation. By the distance transform, the distance $d(y)$ to and the intensity $I(y)$ of the closest surface y pixel are calculated. Details of $d(y)$ and $I(y)$ will be explained shortly. Then the pixel class probability profile $P(Y|\Theta)$ is obtained by probability modeling.

When a 2D image is scanned from a 3D object in ultrasound imaging, two phenomena occur:

(i) A 3D point on the object is transformed to a 2D pixel on the image by plane intersection. The plane is defined by ultrasound beams emitted in a single B-scan.

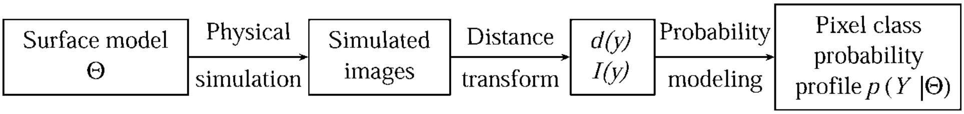


Fig. 9. Probabilistic pixel class prediction.

(ii) the physical properties of the 3D point yield a 2D pixel intensity, determined mostly by the reflective properties of the object.

Simulation generates images from an object model by functional modeling of a real imaging system. We have implemented an ultrasound imaging simulation system to synthesize echocardiographic images from 3D surface models of the left ventricle.[21,22] The object model of LV includes two geometric surface models, one for EPI and the other for ENDO. The simulation software is capable of performing backscattering, attenuation, and reflection, implemented by a ray-tracing algorithm. We only do reflections in this study, since our purpose is to predict the systematic image dropout rather than the stochastic behavior of the speckle noise. The dropout is mostly due to weak reflection at interfaces. The randomness is accounted for in the PixPre and PixApp probability models.

The distance from a pixel p to its closest class-y neighbor pixel q on the simulated image is denoted by $d(y)$. The intensity of the neighbor pixel q is denoted by $I(y)$. $d(y)$ and $I(y)$ of every pixel on a simulated image can be efficiently found by the distance transform.[20,23]

During imaging, a point on the surface may be transformed to a pixel to look more like the background; a point not on the surface may be transformed to a pixel as if on the surface. The PicPre probability model allows such variations than simply saying that a pixel coming from a point on surface y must have label y. In addition, we have further considerations in the PicPre probability model for the following observations. An off-surface point closer to an on-surface point may appear as a pixel with similar location and intensity with the type of the pixel from the on-surface point. An off-surface point that has a stronger on-surface neighbor point is more likely to appear as a pixel that is similar to the type of a pixel from the on-surface neighbor point. To satisfy the above considerations, we design the following parametric PicPre probability model $P(Y|\Theta)$:

$$\frac{P(Y=y|\Theta)}{f(I(y),d(y)|\lambda_y)} = \frac{P(Y=K|\Theta)}{\beta}, \quad y = 1,2,\ldots,K-1, \tag{15}$$

with the constraints

$$\sum_{y=1}^{K} P(Y=y|\Theta) = 1 \quad \text{and} \quad P(Y=y|\Theta) \geq 0, \quad y = 1,\ldots,K. \tag{16}$$

In Eq. (15), $f(I,d|\lambda) : [0,\infty) \times [0,\infty) \to (0,\infty)$ is a decay function decreasing with d but increasing with I; $\lambda_1, \lambda_2, \ldots, \lambda_{K-1}$ are non-negative decay rates of different classes; and β is a non-negative parameter which corresponds to the strength of a pixel being the background. Solving Eqs. (15) and (16), we get

$$P(Y=y|\Theta) = \begin{cases} \dfrac{f(I(y),d(y)|\lambda_y)}{\beta + \sum_{k=1}^{K-1} f(I(k),d(k)|\lambda_k)}, & y = 1,2,\ldots,K-1 \\ \dfrac{\beta}{\beta + \sum_{k=1}^{K-1} f(I(k),d(k)|\lambda_k)}, & y = K \end{cases}. \tag{17}$$

In the above model, the probability of a pixel being class y is in proportion to some monotonically decreasing function of its distance to the nearest pixel coming directly from surface y; the probability of a pixel being on background is in proportion to some function of the smallest distances for this pixel to all other types of non-background pixels. Meanwhile, a pixel is more likely to be from surface y if its neighbor pixel coming directly from surface y has a larger intensity.

In our study, we design the intensity exponential decay model with $f(I, d|\lambda) = Ie^{-\lambda d}$, that is

$$P(Y = y|\Theta) = \begin{cases} \dfrac{I(y)e^{-\lambda_y d(y)}}{\beta + \sum_{k=1}^{K-1} I(k)e^{-\lambda_k d(k)}}, & y = 1, 2, \ldots, K-1 \\ \dfrac{\beta}{\beta + \sum_{k=1}^{K-1} I(k)e^{-\lambda_k d(k)}}, & y = K \end{cases}. \tag{18}$$

We will illustrate this PicPre probability model by an example in next section.

7. Training PixApp and PicPre Probability Models without Low-Level Edge Groundtruth

In some applications, an example of object models Θ and their images are given, but the class labels Y are not available. In other situations, the class labels Y are too inaccurate to use. To achieve the overall optimality, we need to consider simultaneous estimation of the PixApp and PicPre probability models. Our strategy will allow each pixel to participate in a different manner on a scale of 0 to 1 in the PixApp probability models for different classes. We solve the problem of joint estimation of the PixApp and the PicPre probability models by a generalized EM algorithm. In the off-line training phase, the aim is to make an accurate and consistent estimation of $p(Z|\Theta)$. We use the Kullback-Leibler divergence as the criterion for the density estimation, equivalent to the expected log likelihood. A caveat is that the target of estimation is not Θ, but the conditional p.d.f. $p(Z|\Theta)$.

It is necessary to inspect how the models are estimated in the two-stage approach, i.e. the feature detection and model fitting approach. In the feature detection stage, the PixApp probability model $p(Z|Y)$ is used. Estimation of $p(Z|Y)$ requires the knowledge of class labels. However, the class labels are not observed data and they are typically produced by human experts. A class label has to be unique for each pixel. In the model fitting stage, the PicPre probability model $P(Y|\Theta)$ is used. Using the class labels and known surface models, $P(Y|\Theta)$ can be estimated. A problem with these two estimations is that the uncertainty of class label Y as described by $P(Y|\Theta)$ is not taken into account in the estimation of $p(Z|Y)$. The isolation can seriously degrade the performance seriously when uncertainty of class labels for given surface models is prominent.

In the integrated approach, the conditional probability $p(Z|\Theta)$ is given in terms of a summation over Y by

$$p(Z|\Theta) = \sum_y P(Y = y|\Theta)p(Z|Y = y).$$

In this form, we still need to estimate the PicPre probability model $P(Y|\Theta)$ and the PixApp probability model $p(Z|Y)$, but we do not have to make a decision on the class label Y of each pixel, because every possibility of Y is considered. Since we have decided that $P(Y|\Theta)$ is a parametric model and $p(Z|Y)$ is a non-parametric model, the overall model $p(Z|\Theta)$ is a hybrid model. On one hand, maximum likelihood estimation for $p(Z|\Theta)$ requires joint estimation of the PixApp and the PicPre probability models. On the other hand, joint estimation of a hybrid model poses a computational challenge.

Although it is typically a solution to parametric density estimation with missing or hidden variables, the Expectation Maximization (EM) algorithm suitably performs maximum likelihood estimation on p.d.f.s that can be written as an integral or a sum. The missing or hidden variable is precisely the integral or summation variable. Whether the targeted p.d.f. is parametric, non-parametric or hybrid will affect neither the applicability nor the convergence of the EM algorithm. In the integrated model, the goal is to maximize

$$\mathbf{E}[\log p(Z|\Theta)] \tag{19}$$

over all possible p.d.f.s $p(Z|\Theta)$ (not over Θ in our case.) When $p(Z|\Theta)$ is written in the integrated form, Y is the missing or hidden variable. Instead of maximizing Eq. (19), the EM algorithm maximizes an approximation of

$$\mathbf{E}[\log p(Y, Z|\Theta)] \tag{20}$$

over $p(Y, Z|\Theta)$ in its iterations. We shall link $p(Y, Z|\Theta)$ to the PixApp and the PicPre probability models soon. Through the EM algorithm, the maximization of $\mathbf{E}[\log p(Y, Z|\Theta)]$ is substantially computationally easier than that of $\mathbf{E}[\log p(Z|\Theta)]$. We initialize $p(Y, Z|\Theta)$ by a guess $p_0(Y, Z|\Theta)$. It is then followed by iterations of expectation steps (E-steps) and maximization steps (M-steps). In the E-step of iteration m, we first compute the conditional probability $\pi_m(Y|Z, \Theta)$, using $p_m(Y, Z|\Theta)$. Then we find the expectation $\phi_m(p(Y, Z|\Theta)) = \mathbf{E}_{\pi_m}[\log p(Y, Z|\Theta)]$, using the conditional probability $\pi_m(Y|Z, \Theta)$. In the M-step of iteration m, we find a solution that maximizes the expectation $\phi_m(p(Y, Z|\Theta))$, assigning the optimal solution to $p_{m+1}(Y, Z|\Theta)$. The E-step and the M-step alternate until convergence is achieved. When the M-step returns a sub-optimal solution that does not decrease $\phi_m(p(Y, Z|\Theta))$, the algorithm is known as a generalized EM algorithm. Both the original and the generalized EM algorithms increase the targeted expected log likelihood $\mathbf{E}[\log p(Z|\Theta)]$ monotonically as a function of the iteration number.[24]

Now we associate $p(Y, Z|\Theta)$ with the PixApp probability model $p(Z|Y)$ and the PicPre probability model $P(Y|\Theta)$. By the conditional independence assumption $p(\Theta|Z, Y) = p(\Theta|Y)$, we have

$$p(Y, Z|\Theta) = P(Y|\Theta)p(Z|Y, \Theta) = P(Y|\Theta)p(Z|Y).$$

The above equation implies that $p(Y, Z|\Theta)$ is exactly the product of the prediction probability $P(Y|\Theta)$ and the appearance probability $p(Z|Y)$. Hence the M-step can be written as

$$\begin{aligned} &\max_{p(Y,Z|\Theta)} \mathbf{E}_{\pi_m}[\log p(Y, Z|\Theta)] \\ &\quad = \max_{p(Y,Z|\Theta)} \mathbf{E}_{\pi_m}[\log P(Y|\Theta)] + \mathbf{E}_{\pi_m}[\log p(Z|Y)] \\ &\quad = \max_{P(Y|\Theta)} \mathbf{E}_{\pi_m}[\log P(Y|\Theta)] + \max_{p(Z|Y)} \mathbf{E}_{\pi_m}[\log p(Z|Y)]. \end{aligned} \tag{21}$$

Thus, the M-step is broken into two separate optimization problems. One is the parametric estimation of the PicPre probability model, and the other is the non-parametric estimation of the PixApp probability model. Replacing $p(Y, Z|\Theta)$ by $P(Y|\Theta)p(Z|Y)$, we present Algorithm 5, Estimate-Integrated-Model.

Algorithm 5 Estimate-Integrated-Model

```
Initialization:
        $P_0(Y|\Theta)$ and $p_0(Z|Y)$
Iteration:
(1) E-step.
        $\pi_m(Y|Z, \Theta) = \frac{P_m(Y|\Theta)p_m(Z|Y)}{\sum_k P_m(Y = k|\Theta)p_m(Z|Y = k)}$    (22)
        $\phi_m(P(Y|\Theta)) = \mathbf{E}_{\pi_m}[\log P(Y|\Theta)]$    (23)
        $\psi_m(p(Z|Y)) = \mathbf{E}_{\pi_m}[\log p(Z|Y)]$    (24)
(2) M-step.
        $P_{m+1}(Y|\Theta) = \operatorname{argmax}_{P(Y|\Theta)} \phi_m(P(Y|\Theta))$    (25)
        $p_{m+1}(Z|Y) = \operatorname{argmax}_{p(Z|Y)} \psi_m(p(Z|Y))$    (26)
```

The Estimate-Integrated-Model algorithm differs in Eq. (22) from the two-stage estimation solution. In the two-stage approach, every pixel is assigned a unique class label y, equivalent to setting $\pi_m(Y|Z, \Theta) = \delta(Y - y)$. Here, $\pi_m(Y|Z, \Theta)$ signifies the probability profile of class labels for the images and the surface model given. In addition, Estimate-Integrated-Model iterates over the two steps, while the two-stage approach does the two steps only once.

Here is a summary of the overall training strategy. Training images and the corresponding ground-truth surface models are input data to the estimation. Training images are pre-processed to remove noise and obtain feature vectors. Imaging simulation produces simulated images using ground-truth surface models. Before the first iteration, initial guesses for PixApp and PicPre probability models are made. With the PixApp probability model, feature vectors are optimally quantized to calculate the PixApp probability of each pixel for each class. K maps of PixApp probabilities are generated per image. Meanwhile, the PicPre probability profile of each pixel is calculated on the simulated images using the PicPre probability model. A total of K PicPre probability maps are created for each image. With the PixApp and PicPre probability maps, we can calculate a class probability profile for each pixel given both the observed images and the ground-truth surface models. The class profile of each pixel, as a weight vector, participates in both finding a grid quantization for PixApp probability model and the parameter estimation of the PicPre probability model. With the weight vectors and the feature vectors, we obtain an updated PixApp probability model. With the weight vectors and the simulated images, we obtain an updated PicPre probability model. Then we start the next iteration with the newly updated models, until the overall log likelihood $\mathbf{E}_{\pi_m}[\log p(Y, Z|\Theta)]$ converges.

7.1. *PixApp probability model estimation*

One of the two expectations to be maximized in the M-step is $\mathbf{E}_{\pi}[\log p(Z|Y)]$. The expectation is with respect to both Y and Z. However the unknown conditional probability $P(Y|Z, \Theta)$ is replaced by an approximation $\pi(Y|Z, \Theta)$. Hence $\mathbf{E}_{\pi}[\log p(Z|Y)]$ can be written as

$$\mathbf{E}_{\pi}[\log p(Z|Y)] = \int p(z|\Theta) \sum_{k=1}^{K} \pi(Y = k|z, \Theta) \log p(z|Y = k) dz.$$

Taking the sample average log likelihood as the expected value, we obtain

$$\mathcal{L}_1 = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \pi(y_n = k|z_n, \Theta) \log p(z_n|y_n = k).$$

Since $\pi(Y|Z, \Theta)$ is given in the E-step, we simplify the notation by letting

$$\pi_n^k = \pi(y_n = k|z_n, \Theta), \tag{27}$$

which can be thought of as a normalized class weight. Then $\mathcal{L}_1$ can be written as

$$\mathcal{L}_1 = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \pi_n^k \log p(z_n|y_n = k),$$

which is the weighted log likelihood of the sample. Thus, maximization of $\mathbf{E}_{\pi}[\log p(Z|Y)]$ is approximated by that of $\mathcal{L}_1$. As Z is usually a continuous vector, $p(Z|Y)$ is a p.d.f. conditioned on the discrete variable Y. Assumptions on $p(Z|Y)$ reflects how well the imaging process is understood. When there is complex and less studied noise during the imaging process, we would like to make as few assumptions as possible. In the worst case where the least information is available about the noise, we use the grid quantization technique described in Sec. 4 to describe the density function $p(Z|Y)$.

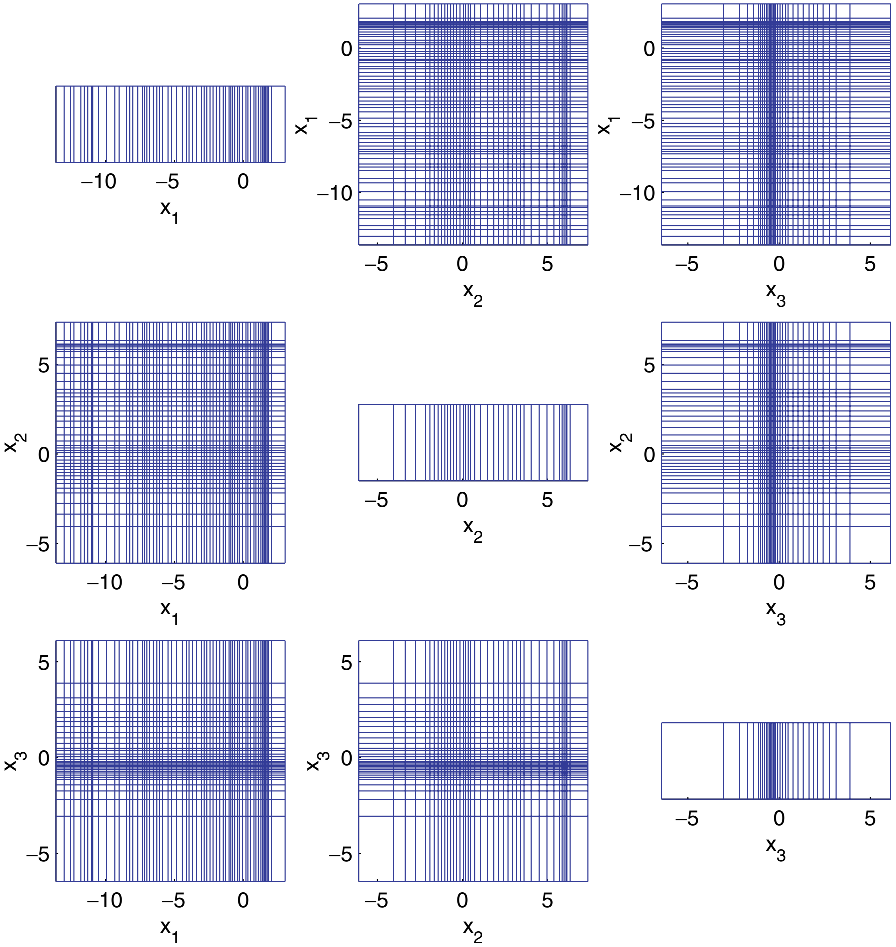
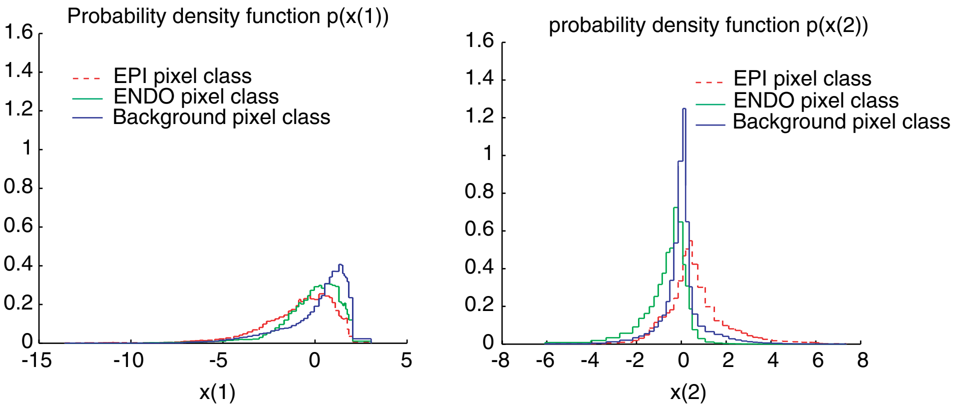


Fig. 10. The 3D quantization grid.

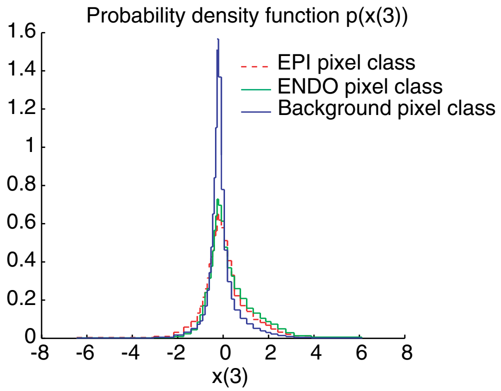
For the original pixel feature vector Z of 5 dimensions, we reduce it to a 3 dimensional vector X. We perform grid quantization on X instead of Z. The optimal 3D quantization grid we obtained for the three classes are displayed in Fig. 10, where 1D and 2D combinations of the grid are drawn. We show the 1D and 2D marginal densities of the 3D PixApp probability densities $p(X|Y)$ in Figs. 11 and 12, respectively. Figure 13 shows the estimated PixApp probability maps for the three classes of a given image.

7.2. *PicPre probability model estimation*

The other expectation to be maximized in the M-step is $\mathbf{E}_\pi[\log P(Y|\Theta)]$. The expectation is on both Z and Y, where Z is implicitly expressed in $\pi(Y|Z,\Theta)$.



(a) $x(1)$ (b) $x(2)$



(c) $x\,(3)$

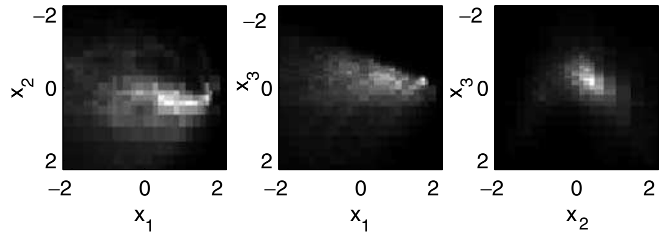
Fig. 11. One-dimensional marginal densities of estimated PixApp probability model.

$\mathbf{E}_\pi[\log P(Y|\Theta)]$ is an estimate of $\mathbf{E}[\log P(Y|\Theta)]$ with $P(Y|Z,\Theta)$ replaced by $\pi(Y|Z,\Theta)$. Therefore we have

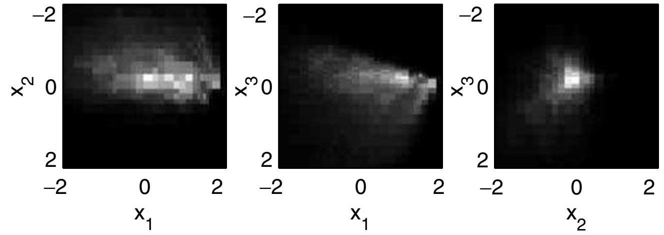
$$\mathbf{E}_\pi[\log P(Y|\Theta)] = \int p(z|\Theta) \sum_{k=1}^{K} \pi(Y=k|z,\Theta) \log P(Y=k|\Theta)\, dz.$$

We can further obtain an estimate of $\mathbf{E}_\pi[\log P(Y|\Theta)]$ by the average log likelihood of the sample, that is

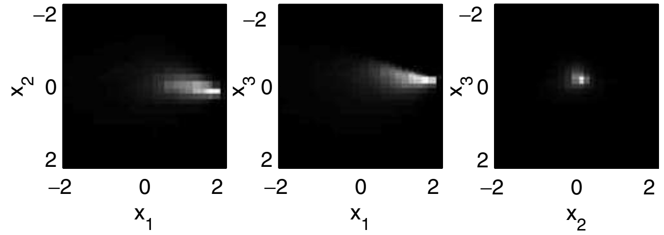
$$\mathcal{L}_2 = \sum_{n=1}^{N} \sum_{k=1}^{K} \pi(y_n = k|z_n,\Theta) \log P(y_n = k|\Theta).$$



(a) Class 1(EPI pixel)



(b) Class 2 (ENDO pixel)



(c) Class 3 (Background pixel)

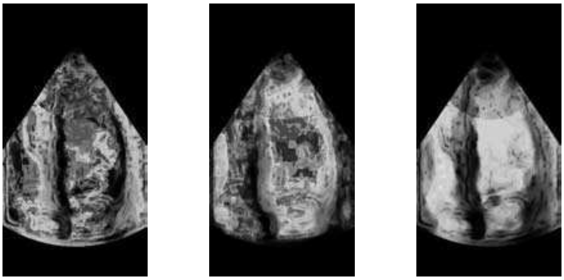
Fig. 12. Two-dimensional marginal densities of estimated PixApp probability model.

The average factor $1/N$ is not shown because it does not affect the maximization. Therefore,

$$\mathcal{L}_2 = \sum_{n=1}^{N} \sum_{k=1}^{K} \pi_n^k \log P(y_n = k|\Theta),$$

which is the weighted log likelihood of the PicPre label assignments. Hence the maximization of $\mathbf{E}_\pi[\log P(Y|\Theta)]$ reduces to that of the weighted log likelihood $\mathcal{L}_2$. As we have defined $P(Y|\Theta)$ by a continuous parametric model previously, $\mathcal{L}_2$ is a function of the PicPre model parameters $\lambda_1, \lambda_2, \ldots, \lambda_{K-1}, \beta$. Since the parameters are all non-negative, we can re-parameterize them by

$$\begin{aligned} \lambda_k &= \tau_k^2, \quad k = 1, \ldots, K-1 \\ \beta &= \alpha^2. \end{aligned}$$



(a) Class 1 Epicardium PixApp ($p(Z|1)$) map

(b) Class 2 Endocardium PixApp ($p(Z|2)$) map

(c) Class 3 Background PixApp ($p(Z|3)$) map

Fig. 13. PixApp probability maps of apical four chamber view.

We denote the parameter vector by

$$u = [\tau_1, \tau_2, \ldots, \tau_{K-1}, \alpha],$$

and the likelihood $\mathcal{L}_2$ by $\mathcal{L}_2(u)$. We adopt a quasi-Newton method that has been widely applied in many unconstrained optimization problems, called the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. It updates the Hessian matrix with a rank two difference matrix during every iteration and guarantees the approximated Hessian matrix is positive definite for minimization problems.[25] The major steps include finding a Newton search direction, the line search and the Hessian update. Since we are to maximize $\mathcal{L}_2(u)$, the objective function of the minimization is $-\mathcal{L}_2(u)$. Figure 14 shows the estimated intensity exponential decay PicPre probability model. Figure 15 illustrates the pixel class prediction process. We obtain a simulated image shown in Fig. 15(a) through ultrasound imaging simulation. Then we compute the distance transform of the epicardium and endocardium contours, shown as Figs. 15(b) and (c). Figures 15(d) and (e) are the intensity maps of the closest on-surface pixels. We apply the estimated PicPre probability model on the distance and intensity maps and display the PicPre probability maps in Figs. 15(f) to (h).

8. Surface Reconstruction for Left Ventricle Using the Estimated Pixel Appearance Probability Model

In our experiment, we used a total of 45 *in vivo* clinical studies. There are 16 normal studies and 29 diseased studies. There are six condition groups among the 45 studies. Forty-four sets of image sequences were acquired from ATL ultrasound machines; one set of image sequences was acquired from an HP ultrasound machine. These image sequences were for other studies by three operators over a period of two years, so that they incorporate some amount of operator and system setting

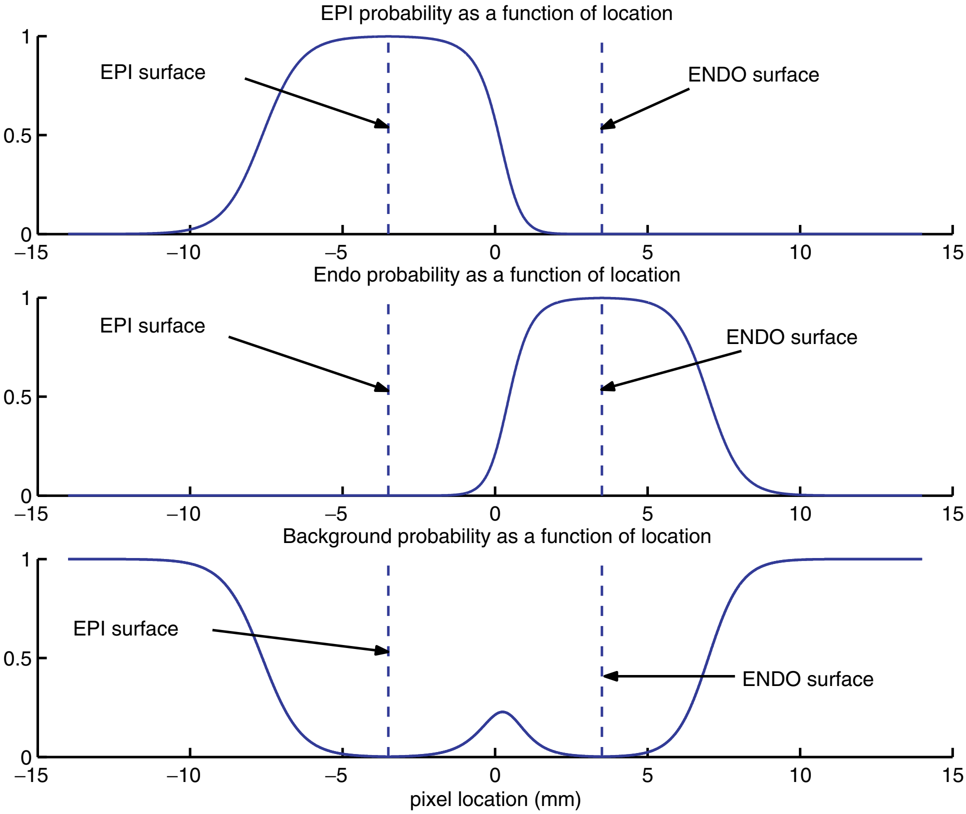
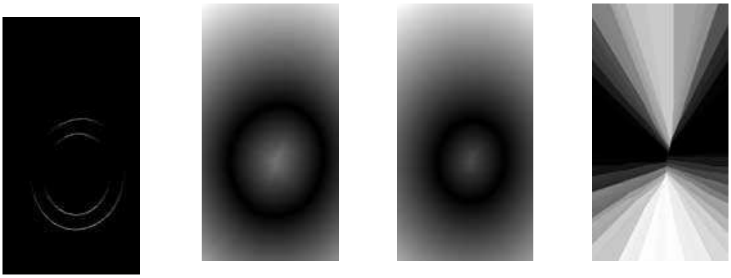


Fig. 14. The estimated PicPre probability model.

variability. The frame rate was 30 per second. The horizontal and vertical resolutions of the images were, respectively, 0.37–0.46 mm and 0.37–0.41 mm per pixel. For each study, we selected subsequences of images from four or five different views, including three or four long-axis views and one short-axis view. Each view was further cut into an upper sector and a lower division, divided by an arc passing an inner point of the LV and centered at the transducer location. We selected 20 studies with good image quality as the training set. The remaining 25 studies formed the test set. We performed the experiment at *end diastole*. We measured the projection distance between the optimized and the ground-truth surface models. The projection distance from surface A to surface B is defined as the mean vertex projection distance from all the vertices of surface A to surface B. The projection distance between surface A and B is the average of the projection distances from A to B and from B to A.

In our study of 3D left ventricle surface reconstruction from 2D echocardiographic images, we obtained much better results using the pixel appearance probability model with the integrated approach than the two-stage approach. In the

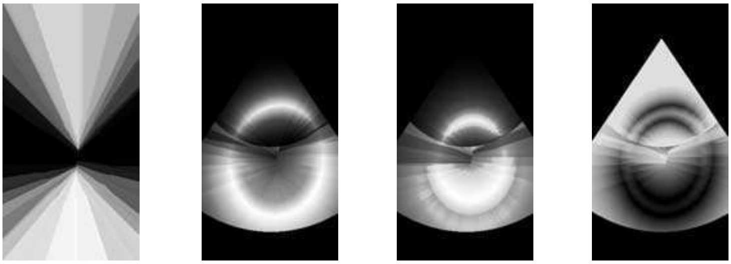


(a) The simulated image

(b) $d(1)$ distance map

(c) $d(2)$ distance map

(d) $I(1)$ intensity map



(e) $I(2)$ intensity map

(f) $P(1|\Theta)$ map

(g) $P(2|\Theta)$ map

(h) $P(3|\Theta)$ map

Fig. 15. A simulated image and its intensity, distance and PicPre probability maps.

two stage approach, we used Canny edge detector to find boundaries and then reconstruct 3D surface model from the detect edges. We were only able to obtain meaningful results for the studies with the best quality images. The distance errors between the manually delineated surfaces and the reconstructed ones range from 3.1 mm to 6.6 mm for normal cases, which were far from the requirement for practical clinical use. For the integrated approach, we were able to handle images with modest image quality. On all the normal studies, we achieved distance errors from 1.1 mm/1.7 mm (endocardium/epicardium) to 3.1 mm/4.0 mm, the average being 1.9 mm/2.4 mm.[16]

9. Conclusions

In this chapter, we have presented the pixel appearance probability model for representation of local pixel information, and illustrated its usage in echocardiographic image analysis. It is vastly different from standard feature detection based low-level image processing techniques, in that it preserves much richer information from the original images. The model is obtained by a grid quantization technique, which is a statistically effective and computationally efficient approach to estimating probability density functions. The pixel appearance probability model can be used in its own right for purposes such as pixel classification. We have argued that this model can be used much more effectively in an integrated object reconstruction framework as opposed to the traditional two-stage approach. In our study, the adoption of the pixel appearance probability model has reduced the max construction error of the left ventricle groundtruth surface by about 2.6 mm.

Acknowledgments

The authors thank Dr. Florence H. Sheehan and Dr. Richard K. Johnson for discussions and suggestions related to the work reported in this chapter. The authors also thank the Cardiovascular Research and Training Center at the University of Washington for providing echocardiographic image sequences and left ventricle surface models reconstructed from manually delineated features.

References

1. A. Chakraborty, L. H. Staib and J. S. Duncan, Deformable boundary finding in medical images by integrating gradient and region information, *IEEE Transactions on Medical Imaging* **15**, 6 (1996) 859–870.
2. T. F. Cootes, A. Hill, C. J. Taylor and J. Haslam, The use of active shape models for locating structures in medical images, *Image and Vision Computing* **23**, 6 (1994) 355–366.
3. T. F. Cootes, G. J. Edwards and C. J. Taylor, Active appearance models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 6 (2001) 681–685.
4. M. Mignotte and J. Meunier, Deformable template and distribution mixture-based data modeling for the encodcarial contour tracking in an echographic sequence, In *Proceedings of 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1999), pp. 225–230.
5. M. P. Gessaman, A consistent non-parametric multivariate density estimator based on statistically equivalent blocks, *The Annals of Mathematical Statistics* **41** (1970) 1344–1346.
6. L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, *Classification and Regression Trees* (Statistics/Probability Series. Wadsworth & Brooks/Cole, Pacific Grove, California, 1984).
7. D. W. Scott and G. Whittaker, Multivariate applications of the ASH in regression, *Communications in Statistics A: Theory and Methods* **25** (1996) 2521–2530.
8. W. Wang, J. Yang and R. R. Muntz, STING: A statistical information grid approach to spatial data mining, In *Proceedings of the 23rd VLDB Conference* (1997), pp. 186–195.
9. A. Hinneburg and D. A. Keim, Optimal grid-clustering: Towards breaking the curse of dimensionality in high-dimensional clustering, In M. P. Atkinson, M. E. Orlowska, P. Valduriez, S. B. Zdonik and M. L. Brodie (eds.), *Proceedings of 25th International Conference on Very Large Data Bases*, pp. 506–517, Edinburgh, Scotland, UK, September 1999. Morgan Kaufmann.
10. S. D. Bay, Multivariate discretization for set mining, *Knowledge and Information Systems* **3**, 4 (2001) 491–512.

11. H. Nagesh, S. Goil and A. Choudhary, Adaptive grids for clustering massive data sets, In V. Kumar and R. Grossman (eds.), *Proceedings of the First SIAM International Conference on Data Mining, April 5–7, 2001, Chicago, IL USA*, pp. 506–517 (Society for Industrial & Applied Mathematics, 2001).
12. L. B. Hearne and E. J. Wegman, Maximum entropy density estimation using random tessellations, In *Computing Science and Statistics* **24** (1992) 483–487.
13. L. B. Hearne and E. J. Wegman, Fast multidimensional density estimation based on random-width bins, In *Computing Science and Statistics* **26** (1994) 150–155.
14. W. Härdle and D. W. Scott, Smoothing by weighted averaging of rounded points, *Computational Statistics* **7** (1992) 97–128.
15. D. W. Scott, *Multivariate Density Estimation — Theory, Practice and Visualization* (Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1992).
16. M. Song, R. M. Haralick, F. H. Sheehan and R. K. Johnson, Integrated surface model optimization for freehand 3D echocardiography, *IEEE Transactions on Medical Imaging* **21**, 9 (2002) 1077–1090.
17. G. Lugosi and A. B. Noble, Consistency of data-driven histogram methods for density estimation and classification, *Annals of Statistics* **24** (1996) 687–706.
18. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, *Computer Science and Scientific Computing* (Academic Press, 2nd edition, 1990).
19. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, 1989).
20. R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Vol. I (Addison-Wesley, 1992).
21. M. Song, Ultrasound imaging simulation and echocardiographic image synthesis, Masters thesis, Department of Electrical Engineering, University of Washington, Seattle, Washington (June 1999).
22. M. Song, R. M. Haralick and F. H. Sheehan, Ultrasound imaging simulation and echocardiographic image synthesis, In *International Conference on Image Processing 2000*, Vancouver, Canada (September 2000).
23. M. J. Seaidoun, *A Fast Exact Euclidean Distance Transform with Application to Computer Vision and Digital Image Processing*, PhD dissertation, Northeastern University (1993).
24. A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B* **39**, 1 (1977) 1–38.
25. J. V. Burke, MATH 516 numerical optimization lecture notes. Department of Mathematics, University of Washington (1999).

This page intentionally left blank

CHAPTER 11

MEASUREMENT OF CAROTID ARTERY STENOSIS FROM MAGNETIC RESONANCE ANGIOGRAPHY

PETER J. YIM* and J. KEVIN DEMARCO
Department of Radiology, UMDNJ-Robert Wood Johnson Medical School
Medical Education Building 404, New Brunwick, NJ 08903, USA
**yimpj@umdnj.edu*

The Degree of Stenosis (DS) is a primary consideration in determining the course of treatment for atherosclerotic disease of the Internal Carotid Artery (ICA). This chapter describes recent developments in Magnetic Resonance Angiography (MRA) for measurement of DS of the ICA. The primary focus is on computational image analysis methodology for the quantification of DS. Background on MRA acquisition strategies, physical models of the MRA acquisition and clinical experiences with MRA are provided that form a basis for the design of image interpretation methodology. Measurement of DS is viewed largely as a segmentation problem. The most promising algorithms for segmentation of the carotid artery from MRA are Geometric Deformable Models (GDM's). Several GDM's have been developed that incorporate a tubular topology. These GDM's are based on the tensor-product-b-spline and on level-sets surface representations. Delineation of the vessel centerline plays an important role in the tubular GDM and can be performed by algorithms based on various models of the image-intensity structure including a differential-geometry model, a moment-of-inertial model, a shortest-path model and a skeletonization model. A non-tubular GDM, the Isosurface Deformable Model, is also discussed. Consideration is given to both conceptual aspects of the algorithms and their performance.

Keywords: Carotid artery stenosis; Magnetic Resonance Angiography (MRA); geometric deformable model.

1. Introduction

Atherosclerotic disease of the internal carotid artery (ICA) is a leading cause of stroke.[1] The formation of atherosclerotic plaque typically occurs in the vicinity of the carotid bifurcation. As the plaque grows, the ICA progressively narrows and becomes prone to causing stroke. Stroke occurs when a thrombus (blood clot), that is formed at the plaque, breaks off, flows downstream and embolizes a cerebral artery. The risk of stroke has been shown to be related to the degree of narrowing of the ICA which is defined as the percent diameter reduction. In the North American Symptomatic Carotid Endarterectomy Trialists (NASCET) study, stroke was seen in 22% of symptomatic patients (with a prior history of transient ischemic attack or stroke) with moderate stenosis of the ICA (30–70%) in a five-year follow-up period.[2] In contrast, stroke was seen in 26% of symptomatic patients with severe ICA stenosis (70–99%) in a *two-year* follow-up period.[3] Comparison of the risks of

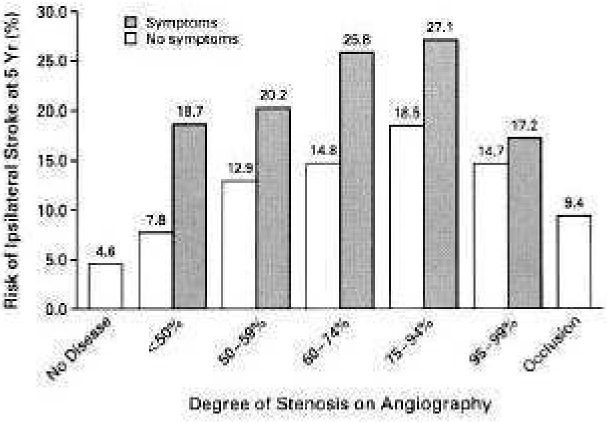


Fig. 1. Relationship between five-year risk of ipsilateral stroke and the degree of stenosis of the internal carotid artery. Reproduced with permission from Inzitari *et al.*, *N. Engl. J. Med.*, Volume 342(23), 2000.

stroke with DS are shown in Fig. 1. Surgical removal of ICA plaque was found to be beneficial, in spite of perioperative risks associated with endarterectomy, only for symptomatic patients with severe ICA stenosis.[3] Similar results have been obtained by the European Carotid Surgery Trial (ECST).[4]

DS of the ICA in the NASCET and ECST was obtained from Digital Subtraction Angiography (DSA) which unambiguously depicts ICA stenosis. However, DSA requires intra-arterial injection of contrast media that is expensive and causes stroke in 1–4% of cases.[5] Thus, alternatives to DSA have emerged for non-invasive evaluation of carotid artery stenosis including Magnetic Resonance Angiography (MRA), Computed Tomographic Angiography (CTA) and Doppler Ultrasound (DU). This chapter describes methodology for measurement of carotid artery stenosis from MRA. The primary focus of the chapter is on computational methodology for quantification of carotid artery stenosis. However, the problem is manifold in nature. Perspectives are provided on MRA acquisition strategies and on clinical studies of the accuracy of MRA.

2. Contrast-Enhanced Magnetic Resonance Angiography

Contrast-enhanced Magnetic Resonance Angiography (MRA) has emerged as a promising non-invasive alternative to Digital Subtraction Angiography (DSA) and MRA obtained by this mechanism is the primary focus of this chapter. In this technique a bolus of Gd Chelate contrast is injected intravenously and the angiogram is acquired during the arterial phase of the circulation of the contrast. This technique was first introduced by Prince *et al.*[6] The use of contrast media for obtaining MRA has been found to be particularly advantageous for detection and evaluation of arterial stenoses. Another contrast mechanism for MRA is Time-Of-Flight (TOF) in which the lumen-to-background contrast is produced by the presence of blood flow.

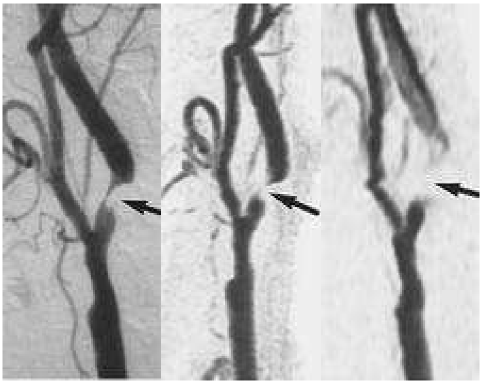


Fig. 2. Comparison of MIP rendering of contrast-enhanced MRA (center) and time-of-flight MRA (right) against gold-standard DSA (left). The location of the stenosis is indicated by the arrows. Reproduced with permission from J. Huston *et al.*, *Radiology*, Volume 218, 2001.

In 2D TOF MRA, for example, the magnetization of stationary tissue within a slice is saturated and produces mimimal image intensity whereas blood flowing into the plane of the slice produces a relatively high image intensity. A significant limitation of this acquisition method is that regions of vessels associated with atherosclerotic plaque formation often have slow or reversed flow, such as in the internal carotid artery immediately distal to the carotid bifurcation and on the side opposite to the flow division.[7] In these locations, a normal vessel may have a stenotic appearance.[8] Visualization of high-grade stenoses may be superior for contrast-enhanced MRA as compared to TOF MRA as shown in Fig. 2 although contrary findings have also been reported.[9] Limitations in TOF MRA can be potentially overcome by more sophisticated interpretation. Ahn *et al.* have shown that the artifactual appearance of stenosis in TOF MRA can potentially be discriminated from true stenosis by consideration of other image features.[10,11] DeMarco *et al.* found that the flow void that may occur at carotid artery stenoses with "classic" 2D TOF MRA is predictive of severe stenosis.[12] Other motivations for the use of contrast for obtaining MRA are to reduce the long image-acquisition times and motion artifacts associated with TOF MRA.

Typically, contrast-enhanced MRA is obtained using a 3D gradient-echo pulse sequence that allows for high-speed image acquisition. 3D MRA can be obtained by either Fourier imaging or projection imaging.[13] At this time, MRA is primarily performed by Fourier imaging. In Fourier imaging, the image is acquired in the spatial-frequency domain and reconstructed by the Fourier Transform. Since contrast-enhanced MRA is acquired under dynamic conditions in which the concentration of the contrast media in the arteries is varying with time, the order in which the spatial frequencies are acquired has a significant impact on the image quality. Elliptical centric acquisitions, in which the low-spatial frequencies are acquired during the peak of the arterial enhancement, have been found to produce good image contrast while suppressing the appearance of veins.[14,15] This strategy is particularly important for imaging of the carotid artery where there is a very short time interval (about 8 seconds) between the arrival of the contrast media in the carotid artery and the return of the contrast through the jugular vein.[16] In comparison, the time for acquisition of the complete image may exceed 20 seconds for a high-resolution scan. In a commercially available Elliptical centric protocol, for example, image acquisition takes approximately 30 seconds using a resolution (number of sample points) of $Nx = 256, Ny = 183$ and $Nz = 36$ and a repetition rate (TR) of 5.6 msec (total imaging time is $Ny \times Nz \times TR$).[17] The elliptical centric acquisition methodology requires that the start of the image acquisition is exactly synchronized with the arrival of the bolus of the contrast media. Methods for synchronization include the use of either a test bolus or so-called fluoroscopic or real-time triggering.[18–20]

An alternative approach to the acquisition of contrast-enhanced MRA is to acquire the images in a repetitive manner that does not require synchronization with the arrival of the contrast bolus. One method for rapid and repetitive acquisition of MRA is the so-called Time-Resolved Imaging of Contrast Kinetics (TRICKS).[21] In this methodology, a temporal frame-rate of as high as 2 seconds can be obtained. The high frame rate is largely achieved by acquiring only a selective set of the spatial frequencies during each time frame with more frequent sampling of the low-spatial frequencies that are known to contribute the most to the image contrast.

Currently, contrast-enhanced MRA is performed with contrast media which is only effective for arterial imaging during its first pass through the arteries. For these contrast media the concentration rapidly declines after the first pass due to uptake within tissue and clearance by the kidneys. However, a new class of contrast media, known as blood-pool agents, maintain a high concentration in the blood over long periods of time and allow for imaging of the vasculature under steady-state conditions.[22] One limitation of this approach is that veins have equal intensity when imaged under steady-state conditions and significantly complicate the image interpretation.

3. Radiological Studies

A number of radiological studies have been carried out to validate contrast-enhanced MRA. The most recent of these studies have involved the use of either a test bolus or fluoroscopic (real-time) imaging for determining the arrival of the contrast bolus. Butz *et al.* compared detection of severe carotid stenosis ($>70\%$ diameter reduction) using contrast-enhanced MRA with detection using Digital Subtraction Angiography (DSA) in 50 consecutive patients.[23] The MRA was acquired at high resolution (0.88 mm slice thickness, 160×512 matrix) with elliptical centric k-space ordering and the acquisition was triggered fluoroscopically. They found a sensitivity of 95.6% and a specificity of 90.4%.

U-King-Im *et al.* evaluated contrast-enhanced MRA for 167 consecutive patients against DSA.[24] Patients were selected for the MRA based on Doppler ultrasound examination. Images were acquired at high resolution (0.8 mm slice thickness, 256 × 256 matrix) using elliptical centric k-space ordering. The acquisition was triggered based on the timing of a test bolus. They found a sensitivity of 93.0% and a specificity of 80.6%. However, they attribute significant portion of the disagreement between MRA and DSA to variability in interpretation of both the MRA and DSA. The inter-observer agreement was assessed for the three readers and the kappa statistic ranged from 0.85 to 0.88 for DSA and 0.86 to 0.88 for MRA. Young *et al.* determined the limits-of-agreement (Bland-Altman) of DSA interpretation to be ±23%.[25] Similar results were obtained by Cosottini *et al.* in a study of 92 patients.[26] They found a sensitivity of 97% and specificity of 82% for detection of severe stenosis of the internal carotid artery.

Huston *et al.* compared the effectiveness of Elliptical centric contrast-enhanced MRA versus DSA for 98 carotid arteries from 50 patients.[27] MRA was evaluated with respect to the ability to detect stenoses of 70% or greater diameter reduction. They found that the estimate of the severity of the stenosis depends on the method of interpretation of the MRA. When the Maximum Intensity Projection (MIP) was used for interpretation, the sensitivity and specificity was 93.3% and 85.1% respectively. When the interpretation was carried out based on obliquely reformatted images, the sensitivity and specificity were 83.3% and 97.0%. They also found that the accuracy of interpretation could not be improved by the use of a multiple regression model that incorporated interpretation based on both the MIP rendering and the obliquely reformatted images. Comparison of measurements from DSA and MRA using MIP and source images is shown in Fig. 3.

Hathout *et al.* evaluated contrast-enhanced MRA for 22 patients who underwent MRA and DSA.[28] They found that at the 95% confidence level, using MRA to predict the stenosis measured by DSA there is at least a variability of ±13.6%. This inaccuracy could be attributed, in part, to variability in interpretation of DSA.

Morasch *et al.* evaluated contrast-enhanced MRA for 29 patients who underwent both DSA and endarterectomy.[29] The stenosis of the internal carotid artery was

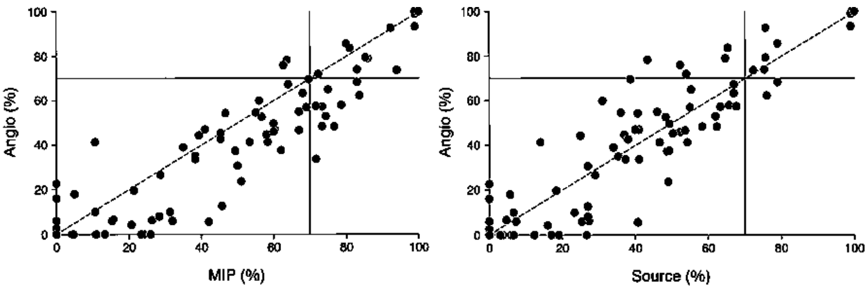


Fig. 3. Relationship between DSA and contrast-enhanced MRA measurements of the degree of stenosis using the MIP (left) and the source images (right). Reproduced with permission from J. Huston *et al.*, *Radiology* Volume 218, 2001.

measured in terms of the reduction in the cross-sectional area of the vessel from a reformatted image in a plane orthogonal to the vessel axis. The area reduction was compared with both the percent stenosis from DSA and the area reduction as measured from the endarterectomy specimen that was resected *en bloc.* A strong correlation was obtained between the minimum cross-sectional area obtained with contrast-enhanced MRA and imaging of the surgical specimen ($R = 0.9$) and there was no statistically significant difference between the minimum cross-sectional area measured with MRA and from imaging of the surgical specimen ($p = 0.61$).

A more basic issue relates to the value of precision in the measurement of DS of the ICA; it is not entirely clear what degree of precision is necessary for MRA of the ICA. On the one hand, the cutoff level for deciding whether a patient should undergo medical or surgical treatment is itself imprecise with reports of various cutoffs from 50% to 70%.[3,30] However, Rothwell *et al.* show that different approaches should be taken for patients with 50–69% stenosis compared with patients with stenoses of 70% or greater.[31]

4. Physical Studies

Fundamental limitations of contrast-enhanced MRA have been identified in phantom studies and with mathematical modeling. These studies have focused on the effects of the time-varying concentration of the contrast media during the acquisition and on the effects of highly heterogeneous blood velocity. The time-varying concentration of the contrast media has the effect of selectively enhancing certain spatial frequencies within the MRA image.[32] Mis-timing of the acquisition can degrade the image by introducing blurring and a ringing artifact.[33,34] Fain *et al.* show that even for optimal timing of the image acquisition, the elliptical centric k-space acquisition has the same effect as a low-pass filter since the low-frequency components of the image are acquired at the peak of the concentration of the contrast media.[35] The spatial filtering can be described in terms of the Full-Width-at-Half-Max (FWHM) of the point spread function (psf). Based on this analysis, the maximal isotropic image resolution that can be obtained with a repetition time (T_r) of 6.5 msec and a total acquisition time of 50 seconds is about 1.0 mm.

Extreme heterogeneity of the blood velocity causes a loss of MR signal due to the "dephasing" of the magnetic spins within a given voxel.[36] This effect is observed both at the neck of the stenosis and immediately distal to the stenosis.[37,38] The spin dephasing at the neck of the stenosis can be attributed to laminar flow with high shear rates while the spin dephasing distal to the stenosis can be attributed to turbulent or complex flow. Mitzuzaki *et al.* found, in a vessel model, that significant spin dephasing can occur even with 50% stenosis under physiologically realistic flow conditions. However, the effect was minimal for a high concentration of the contrast media (2 mmol/l) and when using a short echo time (1.4 msec). Townsend *et al.* found that spin dephasing can cause the vessel stenosis to be severely overestimated, in a flow-through model of the carotid bifurcation with approximately an 80%

stenosis.[39] This effect may have been exaggerated due to the relatively large section thickness employed. That study also showed that for this flow-through model, a TOF MRA acquisition provided a significantly more accurate visualization of the stenosis than did the contrast-enhanced MRA.

5. Computational Algorithms for Measurement of Stenosis

Quantification of DS from the MRA image is an important aspect of imaging of the carotid artery. A variety of computational algorithms have been proposed for quantifying DS. These algorithms will be reviewed in detail below.

5.1. *Delineation of the vessel centerline*

One computational approach to measurement of the DS is to first delineate the vessel centerline. The vessel centerline is potentially useful since it can serve as geometrical reference for detection of the boundaries of the vessel. Several approaches have emerged for delineation of the vessel centerline from MRA or other 3D angiography.

One approach to delineation of the vessel centerline is based on a differential geometry model. This model was first proposed by Aylward *et al.*[40,41] In this model, the vessel is considered to have a tubular shape that appears in the image such that the image becomes progressively more intense towards the center of any given vessel in MRA. This assumption of the vascular intensities is valid for smaller vessels in MRA due to the partial volume effect and can be imposed on larger vessels by blurring the image, for example, by convolution with a Gaussian kernel. The central axis of such tubular shapes can be modeled as ridges in the hyper-surface representing the image. Such ridges can be defined in a rigorous manner in terms of the eigenvalues of the Hessian matrix.[42] For an image, f, the Hessian is defined as follows:

$$H = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial x \partial z} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} & \dfrac{\partial^2 f}{\partial y \partial z} \\ \dfrac{\partial^2 f}{\partial z \partial x} & \dfrac{\partial^2 f}{\partial z \partial y} & \dfrac{\partial^2 f}{\partial z^2} \end{bmatrix}. \tag{1}$$

Since f is a discrete function, the Hessian is formed using Gaussian derivatives. The Gaussian derivatives are defined by convolution:

$$\frac{\partial^2 I}{\partial a \partial b} \equiv K_{ab} * f, \tag{2}$$

where K_{ab} is the discrete kernel sampled from the second partial derivatives of a normalized Gaussian, $G(x, y, z, \sigma)$:

$$G(x, y, z, \sigma) = \frac{1}{\sqrt{(2\pi\sigma^2)^3}} e^{-(x^2+y^2+z^2)/(2\sigma^2)}. \tag{3}$$

The scale factor, σ, dictates the size of the vessel for which the centerline can be accurately located using this model. In this model, for points on the central axis the two smallest eigenvalues, λ_1 and λ_2 are less than zero. The eigenvectors, $\hat{v}_1$ and $\hat{v}_2$ corresponding to the two smallest eigenvalues define a plane that is orthogonal to the vessel axis. In this plane, the image intensity is a local maxima and thus, the projection of the gradient of the image onto either of $\hat{v}_1$ or $\hat{v}_2$ has a small magnitude.

$$\hat{v}_1 \cdot \nabla f \cong 0 \quad \text{and} \quad \hat{v}_2 \cdot \nabla f \cong 0. \tag{4}$$

Under this assumption of the intensity structure in the vascular image, the vessel directionality at the center of the vessel is an eigenvector of the Hessian matrix that corresponds to the largest eigenvalue. Aylward *et al.*[40,41] develop a vessel-tracking algorithm based on this centerline model.

Alternatively, the vessel centerline can be enhanced based on the differential geometry model.[43] The filter is based on the behavior of the eigenvalues of the Hessian in response to various structures. Given an ordering of the eigenvalues, in this case:

$$|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|, \tag{5}$$

line-like structures, for example, produce two eigenvalues that are negative (assuming that the vessel is bright) and have a large magnitude:

$$\lambda_1 \approx 0 \quad \text{and} \quad \lambda_2, \lambda_3 \ll 0. \tag{6}$$

Alternatively, a bright plate-like structure, for example, has two small eigenvalues and one large negative eigenvalue:

$$\lambda_1, \lambda_2 \approx 0 \quad \text{and} \quad \lambda_3 \ll 0. \tag{7}$$

A disciminant function, $V(x, y, z, \sigma)$, can thus be constructed from the eigenvalues of the Hessian that is maximal for line-like structures. As indicated, the function is dependent on the scale, σ, of the Gaussian kernel used for obtaining the Hessian. The scale-dependence is removed by finding the maximal value of the discriminant function across all scales:

$$V_{\max}(x, y, z) = \max_{\sigma_{\min} \leq \sigma \leq \sigma_{\max}} V(x, y, z, \sigma). \tag{8}$$

The results for application of this filter to MRA of the carotid are shown in Fig. 4.

This vessel-enhancement function was adapted by van Bemmel *et al.* for delineation of the vessel centerline.[44] In the proposed method, the centerline is a minimum-cost path between proximal and distal points along the vessel, identified by the user, where the cost of a path, C_{path}, is sum of the reciprocal of the intensities of the enhanced image intensity at each point, $\boldsymbol{p}$, along the path.

$$C_{path} = \sum_{p \in path} \frac{1}{V_{\max}(\boldsymbol{p})}. \tag{9}$$

Such a cost function can be minimized using Dijkstra's shortest-path algorithm.[45]

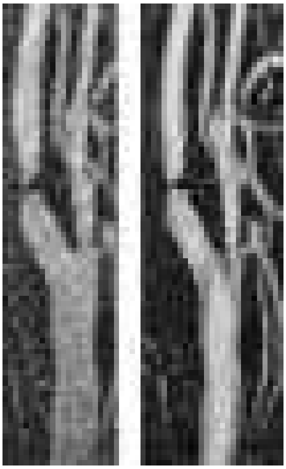


Fig. 4. MIP rendering of contrast-enhanced MRA before (left) and after (right) application of centerline filter derived from eigenvalues of Hessian (Eq. (8)). Reproduced with permission from Frangi *et al.*, *Magn. Reson. Med.* 45(2), 2001.

Vessel directionality can also be determined based on a moment-of-inertia model developed by Hernández-Hoyos *et al.*[46] In this model, a physical moment of inertia is calculated for a given region of the image where the image intensity is considered to be a physical density. The inertia of the given region, or cell, in the image is described by the physical inertia matrix:

$$\boldsymbol{I} = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix}. \tag{10}$$

The diagonal elements of the matrix are moments of inertia and the non-diagonal elements are products of inertia. For example:

$$I_{xx} = \sum_{\boldsymbol{x}_j \in S(\boldsymbol{x}_j, r)} f(\boldsymbol{x}_j)(y_j^2 + z_j^2), \tag{11}$$

and

$$I_{xy} = -\sum_{\boldsymbol{x}_j \in S(\boldsymbol{x}_j, r)} f(\boldsymbol{x}_j) x_j y_j. \tag{12}$$

$S(\boldsymbol{x}_j, r)$ is the set of points in a spherical region centered at $\boldsymbol{x}_j$ with a radius of r. $f(\boldsymbol{x}_j)$ is the image intensity and is analogous to physical density.

Similar analysis of the eigenvalues, as was applied to the Hessian, can be used to obtain the vessel directionality. In this case, the eigenvector associated with the smallest eigenvalue of this matrix corresponds to the vessel directionality, provided that the given region is centered on the vessel. The vessel centerline can then be extracted by tracking along the vessel from a point at the center of a vessel identified by the user. The vessel centerline is tracked by prediction-estimation process. The

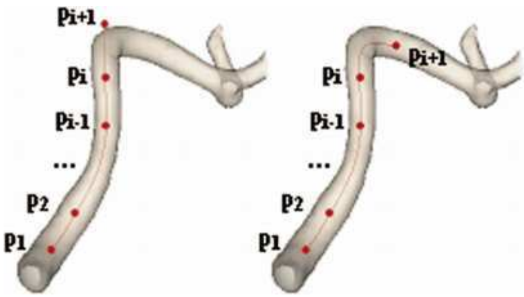


Fig. 5. Two-step process for vessel tracking based on prediction (left) and correction (right). Reproduced with permission from Hernandez-Hoyos *et al.*, *Radiographics*, 22(2), 2002.

tracking proceeds in an iterative manner. In the first stage of each iteration, the extension of the centerline is predicted based on the moment-of-inertial analysis. In the next stage, an estimate of the vessel center is then derived from the predicted vessel center. The estimate of the vessel center takes into account the center-of-mass at the predicted vessel center and the smoothness constraints on the centerline. The concept of this algorithm is illustrated in Fig. 5.

Another approach to delineation of 3D vessel centerlines is that of skeletonization. The concept of skeletonization is that central axes of objects in images can be extracted by progressively stripping away the outer layers of pixels until only the centerlines remain. The concept of skeletonization applies most directly to binary images. Methods for obtaining curvilinear skeletons from 3D binary images include those of Lee *et al.*,[47] Ge *et al.*,[48] and Siddiqi *et al.*[49] Skeletonization can also be applied to gray-scale images. In this case, the order in which pixels or voxel are removed from the image is determined by the image intensity of the pixel or voxel subject to constraints for preserving the topology of the object. Thus, the resulting centerline tends to fall along the ridge with the maximal image intensity. Methods for skeletonization of 2D gray-scale images include those of Arcelli and Ramella[50] and Salari and Siy.[51]

An algorithm for skeletonization of the 3D gray-scale image has been developed by Yim *et al.*[52,53] This algorithm is based on Ordered Region Growing (ORG), a process for obtaining optimal paths between pairs of points in the image. The optimization criteria is the following: Any segment of an ORG path, P_{ORG}, has a minimum intensity greater than or equal to that of any alternative path, P_{alt}, between the same two points.[54]

$$\min\left\{\underset{p_{ORG}\in P_{ORG}}{Y} f(p_{ORG})\right\} \geq \min\left\{\underset{p_{alt}\in P_{alt}}{Y} f(p_{alt})\right\}, \tag{13}$$

where $f(x)$ is the image intensity. ORG is initialized by placement of a point, s_0, at a proximal location on a vessel of interest. An iterative growth process then takes place; new points are incorporated into the growth region, R_n, that

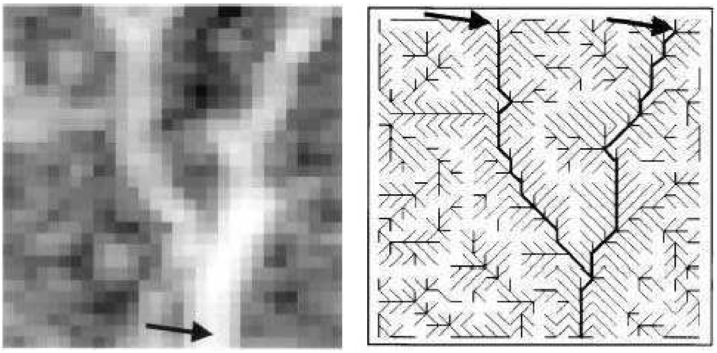


Fig. 6. Gray-scale skeletonization of vascular structures based on Ordered Region Growing algorithm that produces acyclic graph (right) from the image (left). Vessel axes are extracted by identification of proximal and distal points along vessel (arrows).

are adjacent to the point of maximum intensity, s_n, on the boundary, B_n, of the growth region, G_n.

$$\begin{aligned} s_n &= \arg\max_{x \in B_n}\{f(x)\}, \\ G_n &= \text{Neighbors}\{s_n\} \backslash R_n. \end{aligned} \tag{14}$$

The parent-child relations developed by ORG are represented by a directed acyclic graph (DAG) as shown in Fig. 6. A path from any point in the image to the point of initialization of the ORG is found by simply tracing the upstream path through the DAG. This algorithm for delineation of paths in the gray-scale image has been shown to be analogous to a thinning process.

5.2. *Vessel surface reconstruction*

Methods for reconstruction of the carotid artery surface have primarily employed the Geometric Deformable Model (GDM) approach. GDM methods segment objects in an image through a deformation process in which an initial approximation to the object shape is deformed to accurately represent the object shape. The deformation is driven by the image intensity structure and by constraints on the surface smoothness. Typically, the deformation process also preserves the topology such that, for example, holes or bifurcations of the object are neither created nor destroyed.

The GDM was introduced by Kass *et al.* in the context of segmentation of objects in 2D images.[55] Active contours, as the algorithm was called, produces an optimal segmentation of an image by minimization of an "energy" functional. The energy functional is defined with respect to properties of a parametric curve $\boldsymbol{C}\colon [0,1] \to \mathbf{R}^2$.

$$E(\boldsymbol{C}) = \int_0^1 E_{image}(\boldsymbol{C}(u)) + E_{internal}(\boldsymbol{C}, u)du. \tag{15}$$

The image energy reflects the strength of the edges in the image and is defined as:

$$E_{image}(x, y) = -\left\|\nabla(G_\sigma * f(x, y))\right\|, \tag{16}$$

where G_σ represents the Gaussian function with a space constant of σ and f represents the image-intensity function. The internal energy reflects the smoothness of the curve and is defined as:

$$E_{internal}(\boldsymbol{C}, u) = \frac{\alpha}{2}\left\|\frac{d\boldsymbol{C}}{du}\right\|^2 + \frac{\beta}{2}\left\|\frac{d^2\boldsymbol{C}}{du^2}\right\|^2. \tag{17}$$

In practice, the curve can be represented as a set of n discrete nodes, $\boldsymbol{C}_{discrete}$: $\{0, 1, \ldots, n\} \to \mathbf{R}^2$. For discrete curves that are parametrized by the node index, finite differences correspond to the first and second derivatives of the curve that are used to construct the internal energy:

$$\frac{d\boldsymbol{C}}{du} \to \frac{\boldsymbol{C}_{i+1} - \boldsymbol{C}_{i-1}}{2}, \tag{18}$$

$$\frac{d^2\boldsymbol{C}}{du^2} \to \boldsymbol{C}_{i+1} - 2\boldsymbol{C}_i + \boldsymbol{C}_{i-1}. \tag{19}$$

The active contour solution is then obtained in the following manner. First, an initial curve is obtained manually, that is an approximation of the object boundary. That curve then deforms in a manner so as to successively lower the energy functional. Mechanical analogies can be used to produce the desired deformation, where each node is considered to be a particle under the influence of elastic forces, bending forces and forces derived from the image.

A method for reconstruction of the carotid artery surface has been proposed by Frangi *et al.* based on a Tensor-Product B-Spline (TPBS) surface model.[56] The surface model is formed from B-spline basis functions that are parametrized by the circumferential and longitudinal position:

$$W(v, u) = \sum_{j=0}^{q}\sum_{k=0}^{r} N_{jl}(u) N_{km}(v) \boldsymbol{P}_{jk}, \tag{20}$$

where $\boldsymbol{P}_{jk}$ are the control points, $N_{jl}(u)$ is the jth B-spline periodic basis function of order $l, N_{km}(v)$ is the kth B-spline non-periodic basis function of order m. u and v are parameters describing the circumferential and longitudinal positions, respectively. The parametrization is derived from the central axis of the vessel.

The energy functional for this surface model is analogous to that of Active Contours:

$$E^W = E^W_{external} + \bar{\gamma}^W_s \cdot \bar{E}^W_{stretching} + \bar{\gamma}^W_b \cdot \bar{E}^W_{bending}, \tag{21}$$

where the vectors $\bar{\gamma}^W_s$ and $\bar{\gamma}^W_b$ determine the relative weights of the stretching and bending energies. Since, in this case, the surface model is continuous in nature, the internal energy of the surface can be defined in an analytical manner:

$$\bar{E}^{W}_{stretching} = \frac{1}{S}\int_0^1 \int_0^{2\pi} \begin{pmatrix} \|W_v\|^2 \\ \|W_u\|^2 \end{pmatrix} \|W_v \times W_u\| \, dv \, du, \tag{22}$$

$$\bar{E}^{W}_{bending} = \frac{1}{S}\int_0^1 \int_0^{2\pi} \begin{pmatrix} \|W_{vv}\|^2 \\ 2\,\|W_{vu}\|^2 \\ \|W_{uu}\|^2 \end{pmatrix} \|W_v \times W_u\| \, dv \, du, \tag{23}$$

where S represents the surface area of the model and subscript notation is used to indicate first and second partial derivatives of W.

The B-spline GDM was evaluated in flow-though vascular phantoms representing stenotic carotid arteries and for MRA of patients with carotid artery disease.[57] The B-spline GDM tended to over-estimate the lumenal diameter for all image acquisition methods including 2D Time-Of-Flight (TOF) MRA, 3D TOF MRA, 3D Phase-Contrast (PC) angiography and for contrast-enhanced MRA. In particular, the overall average error in the diameter measurement (our calculation from published data) was $+0.28 \pm 0.44$ mm for contrast-enhanced MRA from images with an in-plane resolution of 1.0 mm. The B-spline GDM was also evaluated for contrast MRA of 19 carotid arteries in comparison with digital subtraction angiography. The B-spline GDM measurement of the DS was significantly more correlated with the DSA than either of the observers (our calculation of the statistical significance based on the published correlation coefficients).

GDM's for the carotid artery have also been developed based on the level-sets method for shape representation. In the level-set approach, a surface is defined in an implicit manner as a zero-level set:

$$\Gamma = \{\xi \in \mathbf{R}^3 \mid \varphi(\xi) = 0\}, \tag{24}$$

where φ is the so-called embedding function. Optimization of the surface can, again, be obtained by a deformation process. In this case, deformation of the surface is produced by an evolution of the embedding function that is governed by:

$$\frac{\partial \varphi(\bar{x}, t)}{\partial t} + \boldsymbol{F}\, \|\nabla \varphi(\bar{x}, t)\| = 0, \tag{25}$$

where $\boldsymbol{F}$ is a force derived from smoothing constraints and the local image intensity structure. Evolution of the embedding function of this form produces an evolution of the level-set surface where the velocity of the surface is proportional to the force in the direction normal to the surface, $\hat{N}$, that is applied to it:

$$\frac{\partial \Gamma(t)}{\partial t} = \boldsymbol{F} \cdot \hat{N}. \tag{26}$$

The process of deforming or evolving the surface can be produced by an equivalent process of evolution of the embedding function. As with other GDM's, the level-set evolution produces a surface that is both smooth and aligned with image edges. Lorigo *et al.* proposed a level-sets GDM for vascular surface reconstruction in which smoothing is only applied based only on the minor principal curvature of the surface, presumably corresponding to curvature of the vessel axis. It is appropriate to discount the major principal curvature when performing the smoothing, since the major principal curvature is typically the circumferential curvature and should be allowed to assume high values to allow for narrow vessels.[58] Van Bemmel *et al.* propose a level-sets GDM in which the embedding function is initialized based on the vessel centerline determined by the method described in the previous section.[59] The vessel edge is defined using the Full-Width-at-Half-Max (FWHM) criteria, with the local maxima derived from the vessel centerline. This GDM also only considers the minor principal curvature for smoothing of the surface.

The level-sets GDM of van Bemmel *et al.* was evaluated for measurement of the DS in 15 carotid arteries from contrast-enhanced MRA. In this study, the DS was defined as the percent reduction in the cross-sectional area in the stenotic region relative to the normal region distal to the stenosis. Measurements of the DS by the GDM were found to be consistent with measurements made by three physicians with expertise in interpretation of MRA. The difference between the GDM measurement of the stenosis and the mean of the measurements made by the experts was $0.8 \pm 8.7\%$. The results from four of the cases in this study are shown in Fig. 7.

A physics-based GDM has been developed for vascular surface reconstruction by Yim *et al.* referred to as an Isosurface Deformable Model (IDM).[60] In this approach to surface reconstruction, the surface deformation is governed by mechanical interactions within the surface mesh and between the surface mesh and the image. The behavior of this GDM can also be described in terms of the minimization of an energy functional as follows:

$$E = E_{image} + \alpha\, E_{stretching} + \beta\, E_{bending}, \tag{27}$$

where α and β are the relative weight constants of the stretching and bending energies respectively and:

$$E_{image} = -\sum_{v \in V} \left\| (\bar{K} * f)|_v \cdot \hat{\boldsymbol{n}}_v \right\|, \tag{28}$$

$$E_{stretching} = \sum_{v \in V} \left\| \hat{\boldsymbol{n}}_v \times \left(\boldsymbol{x}_v - \boldsymbol{x}_{ave}\left(N^{vertices}(v) \right) \right) \right\|, \tag{29}$$

$$E_{bending} = \sum_{t \in T} \left\| \hat{\boldsymbol{n}}_v \times \hat{\boldsymbol{n}}_{ave}\left(N^{triangles}(t) \right) \right\|, \tag{30}$$

where:
V is the set of vertices in the surface mesh
T is the set of triangles in the surface mesh
$\bar{K}$ is the convolution kernel that represents the gradient of the Gaussian
$\boldsymbol{x}_v$ is the position of the vertex
$\hat{\boldsymbol{n}}_v$ is the unit vector in the direction normal to the surface at a given vertex
f is the image intensity function
$\boldsymbol{I}$ is the identity matrix
$N^{vertices}$ is the set of adjacent vertices

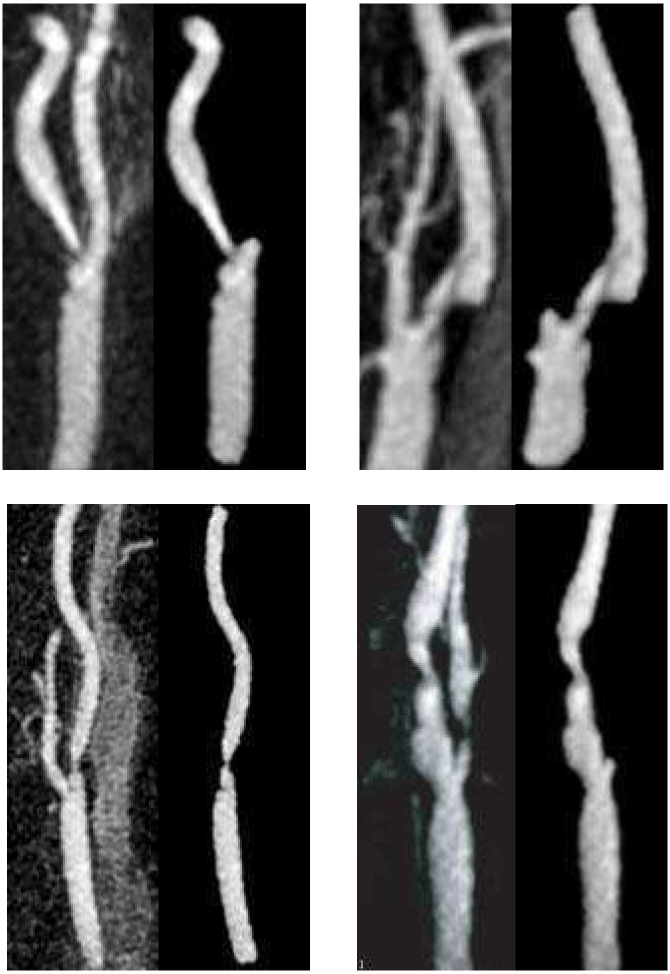


Fig. 7. Segmentation of contrast-enhanced MRA using the level-sets algorithm of van Bemmel *et al.* Each vessel is shown in MIP rendering before (left part of pair) and after (right part of pair) level-sets segmentation. Reproduced with permission from van Bemmel *et al.*, *Magn. Reson. Med.* 51(4), 2004.

$N^{triangles}$ is the set of adjacent triangles
$\boldsymbol{x}_{ave}$ is the mean position of a set of vertices
$\hat{\boldsymbol{n}}_{ave}$ is the unit vector in the direction of the mean of a set of vectors.

The surface mesh, as the name this model suggests, is based on an isosurface constructed using the Marching Cubes algorithm.[61] The IDM has been evaluated for the measurement of the DS in 10 carotid arteries from contrast-enhanced MRA.[62] In this study, the DS was quantified from the surface reconstruction by applying the distance transform to the reconstructed surface and then determining the ORG centerline of the vessel from the distance transform. The value of the distance

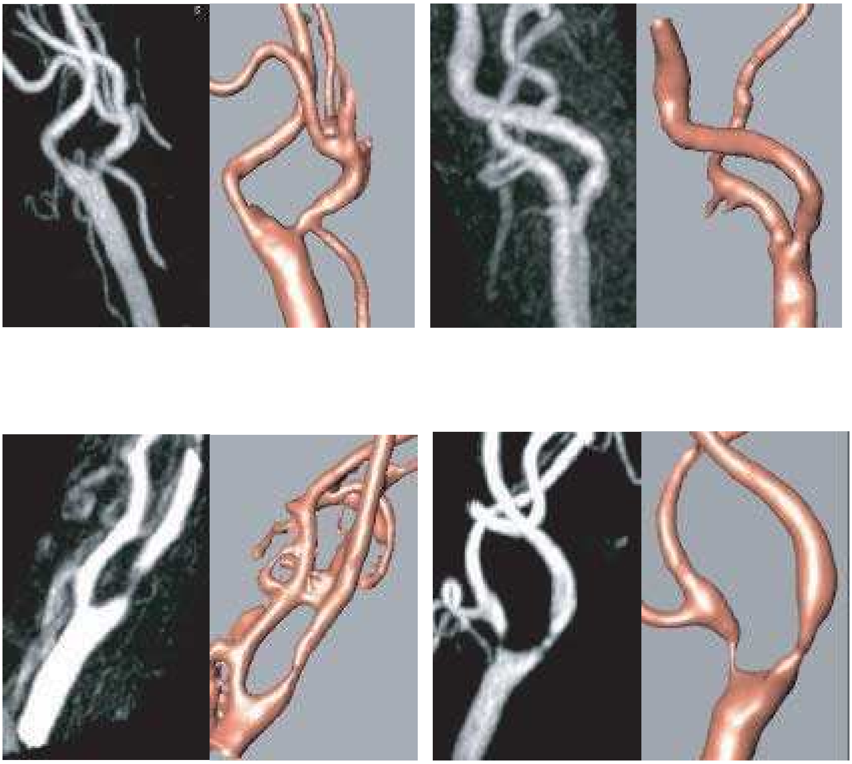


Fig. 8. Segmentation of contrast-enhanced MRA using the IDM. Each MRA is shown in MIP rendering (left part of pair) and shaded-surface display based on IDM segmentation (right part of pair).

transform along the centerline is then considered to be the vessel radius at that point. Measurement of the DS by the IDM was consistent with measurement of the DS by two observers ($r = 0.8836$). The IDM also significantly reduced inter-observer variability in the measurement of the DS ($p = 0.006$). The results of IDM segmentation are shown for four cases in Fig. 8.

6. Summary

Evaluation of carotid artery stenosis with contrast-enhanced MRA has been found to be reliable and accurate in a number of clinical studies. However, MRA is still very much an evolving technology. Significant prospective improvements in quantification of carotid artery stenosis using computational methods have begun to emerge, as have been discussed in detail in the preceding sections. These computational methodologies offer the promise of reducing or removing subjectivity in the measurement of carotid artery stenosis as well as reducing the level of expertise necessary for interpretation of carotid MRA. These methodologies may also facilitate the adaptation of alternative criteria for characterization of arterial stenoses such as percent reduction in lumenal *area*, that are more appropriate for characterization of lumenal shape from 3D imagery.

The computational algorithms, discussed in the preceding sections, may already be sufficient for clinical use in quantification of carotid artery stenosis, although certainly improvements in computational speed and the degree of automation are desirable. However, the accuracy of the measurement of carotid stenosis using computational technology remains the primary consideration for clinical acceptance. Thus, more thorough demonstration of such accuracy must be addressed in clinical studies.

References

1. G. W. Petty, R. D. Brown, J. P. Whisnant, J. D. Sicks, W. M. O'Fallon and D. O. Wiebers, Ischemic stroke subtypes; a population-based study of functional outcome, survival, and recurrence, *Stroke* **31** (2000) 1062–1068.
2. H. J. M. Barnett, D. W. Taylor and M. Eliasziw, Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis, *N. Engl. J. Med.* **339** (1998) 1415–1425.
3. NASCET, Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenoses, *New England Journal of Medicine* **325**, 7 (1991) 445–453.
4. MRC European Carotid Surgery Trial: Interim results for symptomatic patients with severe (70–99%) or with mild (0-29%) carotid stenosis. European Carotid Surgery Trialists' Collaborative Group, *Lancet* **337**, 8752 (1991) 1235–1243.
5. K. E. Garth, B. A. Carroll, F. G. Sommer and D. A Oppenheimer, Duplex ultrasound scanning of the carotid arteries with velocity spectrum analysis, *Radiology* **147**, 3 (1983) 823–827.
6. M. R. Prince, E. K. Yucel, J. A. Kaufman, D. C. Harrison and S. C. Geller, Dynamic gadolinium-enhanced three-dimensional abdominal MR arteriography, *J. Magn. Reson. Imaging* **3**, 6 (1993) 877–881.
7. B. K. Bharadvaj, R. F. Mabon and D. P. Giddens, Steady flow in a model of the human carotid bifurcation. Part I — flow visualization, *J. Biomech.* **15**, 5 (1982) 349–362.
8. M. R. Patel, R. A. Klufas, D. Kim, R. R. Edelman and K. C. Kent, MR angiography of the carotid bifurcation: Artifacts and limitations, *AJR Am. J. Roentgenol.* **162**, 6 (1994) 1431–1437.
9. T. C. Townsend, D. Saloner, X. M. Pan and J. H. Rapp, Contrast material-enhanced MRA overestimates severity of carotid stenosis, compared with 3D time-of-flight MRA, *J. Vasc. Surg.* **38**, 1 (2003) 36–40.
10. K. J. Ahn, W. J. You, J. H. Lee, B. J. Kang, Y. J. Kim, B. S. Kim and S. T. Hahn, Re-circulation artefact at the carotid bulb can be differentiated from true stenosis, *Br. J. Radiol.* **77**, 919 (2004) 551–556.
11. P. J. Nederkoorn, Y. van der Graaf, B. C. Eikelboom, A. van der Lugt, L. W. Bartels and W. P. Mali, Time-of-flight MR angiography of carotid artery stenosis: Does a flow void represent severe stenosis? *AJNR Am. J. Neuroradiol.* **23**, 10 (2002) 1779–1784.
12. J. K. DeMarco, J. Huston III and M. A. Bernstein, Evaluation of classic 2D time-of-flight MR angiography in the depiction of severe carotid stenosis, *AJR Am. J. Roentgenol.* **183**, 3 (2004) 787–793.

13. J. Du, T. J. Carroll, E. Brodsky, A. Lu, T. M. Grist, C. A. Mistretta and W. F. Block, Contrast-enhanced peripheral magnetic resonance angiography using time-resolved vastly undersampled isotropic projection reconstruction, *J. Magn. Reson. Imaging* **20**, 5 (2004) 894–900.
14. A. H. Wilman and S. J. Riederer, Performance of an elliptical centric view order for signal enhancement and motion artifact suppression in breath-hold three-dimensional gradient echo imaging, *Magn. Reson. Med.* **38**, 5 (1997) 793–802.
15. H. M. Lee and Y. Wang, Dynamic k-space filling for bolus chase 3D MR digital subtraction angiography, *Magn. Reson. Med.* **40**, 1 (1998) 99–104.
16. J. Huston, III, S. B. Fain, S. J. Riederer, A. H. Wilman, M. A. Bernstein and R. F. Busse, Carotid arteries: Maximizing arterial to venous contrast in fluoroscopically triggered contrast enhanced MR angiography with elliptic centric view ordering, *Radiology* **211** (1999) 265–273.
17. J. K. De Marco, S. Schonfeld, I. Keller and M. A. Bernstein, Contrast-enhanced carotid MR angiography with commercially available triggering mechanisms and elliptic centric phase encoding, *AJR Am. J. Roentgenol.* **176**, 1 (2001) 221–227.
18. J. K. Kim, R. I. Farb and G. A. Wright, Test bolus examination in the carotid artery at dynamic gadolinium-enhanced MR angiography, *Radiology* **206**, 1 (1998) 283–289.
19. A. H. Wilman, S. J. Riederer, J. Huston, III, J. T. Wald and J. P. Debbins, Arterial phase carotid and vertebral artery imaging in 3D contrast-enhanced MR angiography by combining fluoroscopic triggering with an elliptical centric acquisition order, *Magn. Reson. Med.* **40**, 1 (1998) 24–35.
20. T. K. Foo, V. B. Ho and P. L. Choyke, Contrast-enhanced carotid MR angiography, Imaging principles and physics, *Neuroimaging Clin. N. Am.* **9**, 2 (1999) 263–284.
21. F. R. Korosec, R. Frayne, T. M. Grist and C. A. Mistretta, Time-resolved contrast-enhanced 3D MR angiography, *Magn. Reson. Med.* **36**, 3 (1996) 345–351.
22. T. M. Grist, F. R. Korosec, D. C. Peters, S. Witte, R. C. Walovitch, R. P. Dolan, W. E. Bridson, E. K. Yucel and C. A. Mistretta, Steady-state and dynamic MR angiography with MS-325: Initial experience in humans, *Radiology* **207**, 2 (1998) 539–544.
23. B. Butz, U. Dorenbeck, I. Borisch, N. Zorger, M. Lenhart, S. Feuerbach and J. Link, High-resolution contrast-enhanced magnetic resonance angiography of the carotid arteries using fluoroscopic monitoring of contrast arrival: Diagnostic accuracy and interobserver variability, *Acta Radiol.* **45**, 2 (2004) 164–170.
24. J. M. U-King-Im, R. A. Trivedi, M. J. Graves, N. J. Higgins, J. J. Cross, B. D. Tom, W. Hollingworth, H. Eales, E. A. Warburton, P. J. Kirkpatrick, N. M. Antoun and J. H. Gillard, Contrast-enhanced MR angiography for carotid disease: Diagnostic and potential clinical impact, *Neurology* **62**, 8 (2004) 1282–1290.
25. G. R. Young, P. R. Humphrey, T. E. Nixon and E. T. Smith, Variability in measurement of extracranial internal carotid artery stenosis as displayed by both digital subtraction and magnetic resonance angiography: An assessment of three caliper techniques and visual impression of stenosis, *Stroke* **27** (1996) 467–473.
26. M. Cosottini, A. Pingitore, M. Puglioli, M. C. Michelassi, G. Lupi, A. Abbruzzese, R. Calabrese, M. Lombardi, G. Parenti and C. Bartolozzi, Contrast-enhanced three-dimensional magnetic resonance angiography of atherosclerotic internal carotid stenosis as the non-invasive imaging modality in revascularization decision making, *Stroke* **34**, 3 (2003) 660–664.
27. J. Huston, III, S. B. Fain, J. T. Wald, P. H. Luetmer, C. H. Rydberg, D. J. Covarrubias, S. J. Riederer, M. A. Bernstein, R. D. Brown, F. B. Meyer, T. C. Bower and

C. D. Schleck, Carotid artery: Elliptic centric contrast-enhanced MR angiography compared with conventional angiography, *Radiology* **218**, 1 (2001) 138–143.

28. G. M. Hathout, M. J. Duh and S. M. El-Saden, Accuracy of contrast-enhanced MR angiography in predicting angiographic stenosis of the internal carotid artery: Linear regression analysis, *AJNR Am. J. Neuroradiol.* **24**, 9 (2003) 1747–1756.
29. M. D. Morasch, A. N. Gurjala, E. Washington, A. C. Chiou, O. P. Simonetti, J. P. Finn and J. S. Yao, Cross-sectional magnetic resonance angiography is accurate in predicting degree of carotid stenosis, *Ann. Vasc. Surg.* **16**, 3 (2002) 266–272.
30. J. D. Vijungco and W. H. Pearce, Carotid artery revascularization, *Top Stroke Rehabil.* **10**, 3 (2003) 46–60.
31. P. M. Rothwell, M. Eliasziw, S. A. Gutnikov, C. P. Warlow and H. J. Barnett, Carotid endarterectomy Trialists Collaboration, Endarterectomy for symptomatic carotid stenosis in relation to clinical subgroups and timing of surgery, *Lancet* **20** (2004) 915–924.
32. K. Ito, J. Kato, S. Okada and T. Kumazaki, K-space filter effect in three-dimensional contrast MR angiography, *Acta Radiol.* **38**, 1 (1997) 173–175.
33. J. Svensson, J. S. Petersson, F. Stahlberg, E. M. Larsson, P. Leander and L. E. Olsson, Image artifacts due to a time-varying contrast medium concentration in 3D contrast-enhanced MRA, *J. Magn. Reson. Imaging* **10**, 6 (1999) 919–928.
34. J. H. Maki, M. R. Prince, F. J. Londy and T. L. Chenevert, The effects of time varying intravascular signal intensity and k-space acquisition order on three-dimensional MR angiography image quality, *J. Magn. Reson. Imaging* **6**, 4 (1996) 642–651.
35. S. B. Fain, S. J. Riederer, M. A. Bernstein and J. Huston, III, Theoretical limits of spatial resolution in elliptical-centric contrast-enhanced 3D-MRA, *Magn. Reson. Med.* **42**, 6 (1999) 1106–1116.
36. J. C. Gatenby, T. R. McCauley and J. C. Gore, Mechanisms of signal loss in magnetic resonance imaging of stenoses, *Med. Phys.* **20**, 4 (1993) 1049–1057.
37. K. Mitsuzaki, Y. Yamashita, M. Onomichi, T. Tsuchigame and M. Takahashi, Delineation of simulated vascular stenosis with Gd-DTPA-enhanced 3D gradient echo MR angiography: An experimental study, *J. Comput. Assist. Tomogr.* **24**, 1 (2000) 77–82.
38. B. R. Mustert, D. M. Williams and M. R. Prince, In vitro model of arterial stenosis: Correlation of MR signal dephasing and trans-stenotic pressure gradients, *Magn. Reson. Imaging* **16**, 3 (1998) 301–310.
39. T. C. Townsend, D. Saloner, X. M. Pan and J. H. Rapp, Contrast material-enhanced MRA overestimates severity of carotid stenosis, compared with 3D time-of-flight MRA, *J. Vasc. Surg.* **38**, 1 (2003) 36–40.
40. S. R. Aylward, E. Bullitt, S. M. Pizer and D. Eberly, Intensity ridge and widths for tubular object segmentation and registration, in *Proc. IEEE Workshop Mathematical Methods in Biomedical Image Analysis* (1996) 131–138.
41. S. R. Aylward and E. Bullitt, Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction, *IEEE Trans. Med. Imaging* **21**, 2 (2002) 61–75.
42. D. Eberly, R. Gardner, B. Morse, S. Pizer and C. Scharlach, Ridges for image analysis, *Journal of Mathematical Imaging and Vision* **4** (1994) 353–373.
43. A. F. Frangi, W. J. Niessen, K. L. Vincken and M. A. Viergever, Multiscale vessel enhancement, *Itering.* In W. M. Wells, A. Colchester and S. Delp (eds), *Medical Image Computing and Computer-Assisted Intervention*, Volume 1496, *Lecture Notes in Computer Science* (1998) 130–137.

44. C. M. van Bemmel, M. A. Viergever and W. J. Niessen, Semiautomatic segmentation and stenosis quantification of 3D contrast-enhanced MR angiograms of the internal carotid artery, *Magn. Reson. Med.* **54**, 4 (2004) 753–760.
45. E. W. Dijkstra, A note on two problems in connection with graphs, *Numerische Mathematic* **1** (1959) 269–271.
46. M. Hernández-Hoyos, M. Orkisz, J. P. Roux and P. Douek, Inertia-based vessel axis extraction and stenosis quantification in 3D MRA images, in *CARS'99 — Computer Assisted Radiology and Surgery*, H. U. Lemke, M. W. Vannier, K. Inamura and A. G. Farman (eds) (Amsterdam: Elsevier-Verlag, 1999), pp. 189–193.
47. T. C. Lee, R. L. Kashyap and C. N. Chu, Building skeleton models via 3D medial surface/axis thinning algorithms, *CVGIP: Graph Models Image Proc.* **56** (1994) 462–478.
48. Y. Ge, D. R. Stelts, J. Wang and D. J. Vining, Computing the centerline of a colon: A robust and efficient method based on 3D skeletons, *J. Comput. Assist. Tomogr.* **23**, 5 (1999) 786–794.
49. K. Siddiqi, S. Bouix, S. Tannenbaum and S. W. Zucker, The Hamilton-Jacobi Skeleton, *International Conference on Computer Vision* **2** (1999) 828–834.
50. C. Arcelli and G. Ramella, Finding grey-skeletons by iterated pixel removal, *Image and Vision Computing* **13**, 3 (1995) 159–167.
51. E. Salari and P. Siy, The ridge-seeking method for obtaining the skeleton of digital images, *IEEE Transactions on Systems, Man and Cybernetics* **14**, 3 (1984) 524–528.
52. P. J. Yim, P. L. Choyke and R. M. Summers, Gray-scale skeletonization of small vessels in magnetic resonance angiography, *IEEE Trans. Med. Imaging* **19**, 6 (2000) 568–576.
53. P. L. Choyke, P. Yim, H. Marcos, V. B. Ho, R. Mullick and R. M. Summers, Hepatic MR angiography: A multiobserver comparison of visualization methods, *AJR Am. J. Roentgenol.* **176**, 2 (2001) 465–470.
54. P. J. Yim, D. Kim and P. L. Choyke, The watershed and skeletonization of angiography, *Proceedings of SPIE* **5032** (2003) 1685–1698.
55. M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active contour models, *International Journal of Computer Vision* **1**, 4 (1987) 321–331.
56. A. F. Frangi, W. J. Niessen, R. M. Hoogeveen, T. van Walsum and M. A. Viergever, Model-based quantitation of 3D magnetic resonance angiographic images, *IEEE Trans. Med. Imaging* **18**, 10 (1999) 946–956.
57. A. F. Frangi, W. J. Niessen, P. J. Nederkoorn, J. Bakker, W. P. Mali and M. A. Viergever, Quantitative analysis of vascular morphology from 3D MR angiograms: In vitro and in vivo results, *Magn. Reson. Med.* **45**, 2 (2001) 311–322.
58. M. L. Lorigo, O. D. Faugeras, W. E. Grimson, R. Keriven, R. Kikinis, A. Nabavi and C. F. Westin, CURVES: Curve evolution for vessel segmentation, *Med. Image Anal.* **5**, 3 (2001) 195–206.
59. C. M. van Bemmel, M. A. Viergever and W. J. R. Niessen, Semiautomatic segmentation and stenosis quantification of 3D contrast-enhanced MR angiograms of the internal carotid artery, *Magn. Reson. Med.* **51**, 4 (2004) 753–760.
60. P. J. Yim, G. Boudewijn, C. Vasbinder, V. B. Ho and P. L. Choyke, Isosurfaces as deformable models for magnetic resonance angiography, *IEEE Trans. Med. Imaging* **22**, 7 (2003) 875–881.
61. W. E. Lorensen and H. E. Cline, Marching cubes: A high resolution 3D surface construction algorithm, *Computer Graphics* **21**, 4 (1987) 163–168.
62. P. J. Yim, M. Mishra and J. K. DeMarco, Precision measurement of carotid artery stenosis using the isosurface deformable model and skeletonization, *Proceedings of the ISMRM 2004*, Kyoto, Japan, May 15–21 2004.

Index

adaptive multiscale products thresholding, 90
adenosine triphosphate, 136
airway segmentation, 62
Ampere's law, 211
angiography, 37
anomaly, 222
Archimedian spiral, 244
arterial input function, 272
arterial spin-labeling (ASL), 241

b factor, 264
bioimpedance, 194
Biot–Savard law, 144, 213
block-designs, 256
blood oxygen level dependant, 254
boundary value problem, 197
BPF algorithms, 19
breast cancer, 229
breast cancer detection, 230

Canny edge detector, 327
carboplatin, 111, 121, 125, 127
cardiac imaging, 8
cerebral blood flow, 253, 254
cerebral blood volume, 254
chemical shift, 246
chemotherapy, 106–108
computed tomography (CT), 105, 106, 110, 112–114, 120, 128
computer-aided diagnosis, 22
conductivity, 194
conductivity image, 221
cone-angle induced aliasing, 26
conebeam CT, 16
contrast to noise ratio (CNR), 97, 262
contrast-enhanced magnetic resonance angiography (MRA), 332
cortical potential imaging, 169
CT angiography, 7
CT dose, 28
CT fluoroscopy, 10
CT perfusion imaging, 10
current density, 197, 211
current source, 219

data fitting method, 205
desired region of interest (DROI), 97
diffusion coefficient (DC), 263
diffusion tensor imaging (DTI), 241
diffusion-weighted images (DWI), 265
digital subtraction angiography (DSA), 332, 334
dipole localization, 163
discrete decomposition algorithms, 86
dose reduction strategies, 30
drug delivery system, 106, 108, 114
dyadic wavelet transform (DWT), 85
dynamic imaging, 200
dynamic susceptibility contrast (DSC), 241

echo-planar imaging (EPI), 241
EIT images, 209
EIT system, 207
electric field, 196, 231
electrical impedance tomography (EIT), 196
electrode model, 201
electrodes, 194, 219, 229
end diastole, 326
endocardial potential imaging, 176, 178
ENDOcardium (ENDO), 297
epicardial potential imaging, 174, 175
EPIcardium (EPI), 297
event-related, 256
excitable tissue, 134
expectation maximization (EM) algorithm, 319

FDK-style methods, 18
field inhomogeneity, 249
finite element method (FEM), 198
fitness function, 307

Fitzhugh–Nagumo model, 138
flow imaging, 54
forward problem, 148–150, 196, 212
Fourier transform, 242
fractional anisotropy (FA), 267
frequency-encoding, 244
frontal plane impedance camera, 229
full width at half maximum (FWHM), 252
functional MRI, 253
functional neuroimaging (fMRI), 241
fundamental solution, 223

Gadolinium, 269
gastrointestinal smooth muscle, 182
geometric deformable model (GDM), 341
geometric distortion, 247
gradient B_z decomposition algorithm, 217
Green's functions, 153
grid quantization, 302
groundtruth, 298
gyromagnetic ratio, 242

hard thresholds, 89
harmonic B_z algorithm, 215
heart, 170
helical CT, 1
Helmhortz decomposition, 217
hemodynamic response, 255
Hessian matrix, 337, 338
high angular resolution diffusion (HARD), 269
Hodgkin–Huxley equations, 135
hologram, 44
holographic reconstruction, 43

ill-posedness, 206
image reconstruction, 12
impulse response function, 256
integrated approach, 301
integrated approach of object model optimization, 301
internal carotid artery (ICA), 331
interstitial cells of Cajal (ICCs), 138
intratumoral chemotherapy, 108
inverse problem, 148, 149, 153, 163, 199, 213
in vitro/in vivo correlation, 114
in vitro/in vivo validation, 106, 111, 112
iohexol, 111–119
isosurface deformable model (IDM), 344

k-edge, 114
k neighborhood, 312
k-space, 242

Larmor frequency, 242
layer potential technique, 216
lead field scanning, 165
least square method, 216
lesion estimation, 222
lesion estimation algorithm, 232
level-set, 343
Lipschitz regularity, 87
liquid crystal-spatial light modulator (LC-SLM), 51
location search, 224
lung cancer screening, 11

magnetic field gradients, 242
magnetic flux density, 196, 211
magnetic flux density images, 221
magnetic resonance current density imaging (MRCDI), 211
magnetic resonance diffractive imaging, 37
magnetic resonance electrical impedance tomography (MREIT), 210
magnetic resonance imaging (MRI), 81, 105, 109, 110, 129, 165, 174, 241
marching cubes algorithm, 345
Markov random fields, 298
Maxwell's equations, 139
mean class average log likelihood, 305
mean class correct classification probability, 305
mean diffusivity (MD), 267
mean transit time (MTT), 271
mean-to-standard-deviation-ratio (MSR), 97
metabolic rate of oxygen extraction, 254
minimization problem, 205
moment-of-inertia, 339
MRI scanner, 211
multi-phase organ studies, 9
multiplanar reformation (MPR), 67
multiscale edge detector, 84
multiscale products, 88
multiscale products threshold, 94
multislice computed tomography, 1
multislice helical pitch, 6

neighbor cells, 311
Nernst equation, 134
Neumann boundary condition, 197
Neumann boundary data, 212
Neumann boundary value problem, 212
Neumann function, 199, 200
Neumann-to-B_z map (NtB_z-map), 215
Neumann-to-Dirichlet (NtD) map, 199
neuronal activity, 254
noise standard deviation, 220
non-parametric, 297
nonlinearity, 207
nuclear magnetic resonance (NMR), 37
Nyquist ghosting, 251

object appearance model, 300
object reconstruction, 300
off-resonance, 253
ordered region growing (ORG), 340
orthogonal wavelet transform (OWT), 95
over-complete wavelet expansion (OWE), 97
overall average log likelihood, 305
overall correct classification probability, 305

perfusion imaging, 241
pharmacokinetic, 105, 108, 109
phase-encoding, 244
pi-line, 16
PicPre probability model, 317
PixApp probability model, 315
pixel appearance probability model, 297, 302
pixel class prediction, 316
pixel class prediction (PicPre) probability model, 301
Poisson equation, 140, 148
positron emission tomography (PET), 105, 109, 110
preferred pitch, 24
probability distribution function (PDF), 264
projection distance, 326
pulse sequence, 219

q-space imaging (QSI), 269
quadratic field gradient, 39
quantization, 298
quantization cell, 304
quantization levels, 307
quantizer, 304
quantizer performance measure, 304
quantizer performance measure function, 306
quasi-static approximation, 140

radiofrequency ablation (RFA), 106–108, 118
radius 0 neighborhood, 311
radius R ($R \in \mathbb{Z}^+$) neighborhood, 311
radius-smoothing, 312
radon-transform based methods, 18
rebinning algorithms, 18
reciprocity theorem, 198
reference electrode, 230
regularization, 205
relative anisotropy (RA), 267
residue function, 272
respiratory motion artifact, 116, 117, 128

S-CT data acquisition, 62
sampling and aliasing, 23
scan probe, 229
sensitivity matrix, 201
shape-based interpolation, 282, 287
signal-to-noise ratio (SNR), 220
single layer potential, 224
single photon emission computed tomography (SPECT), 105, 109
singular value decomposition (SVD), 273
site-specific drug release, 105–107
size estimation, 226
skeletal muscle, 180, 181
skeletonization, 340
soft thresholds, 89
solid-state detectors, 5
speckle noise, 317
spin-echo, 243
spiral imaging, 241
static imaging, 205
statistical active appearance models, 299
surface rendering, 21
susceptibility, 249
synthetic aperture magnetometry (SAM), 167

temporal resolution, 208
tensor-product B-spline (TPBS), 342
tissue phantom, 221

trans-admittance, 229
trans-admittance map, 234
trans-admittance scanner, 233
trans-impedance, 208
transmembrane potential imaging, 178

ultrasound imaging, 313
ultrasound imaging simulation, 317
undesired region of interest (UROI), 97

variational gradient B_z algorithm, 218
vessel centerline, 337–339
virtual colonoscopy, 11
volume rendering, 21

wavelet-based thresholding, 89
well-posed problem, 222

z-shim, 250