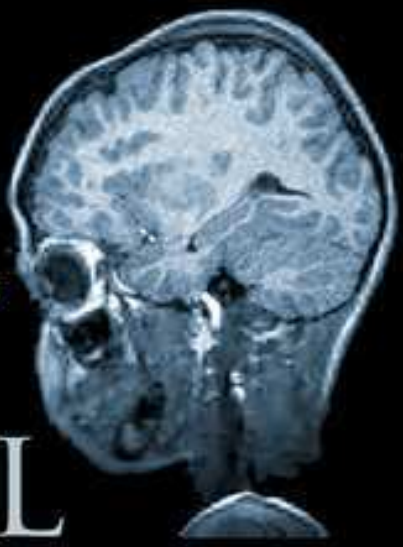
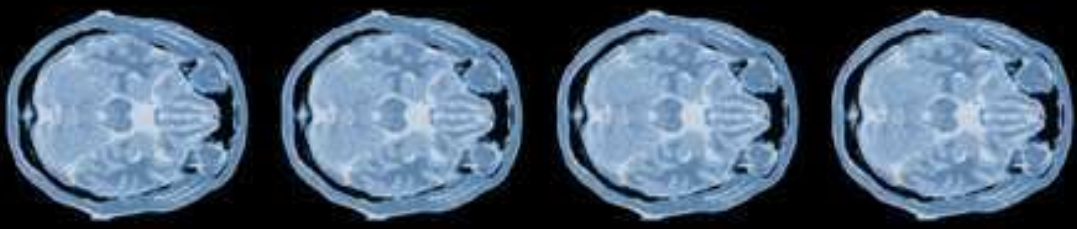
A 5-Volume Set

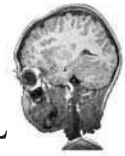


MEDICAL IMAGING SYSTEMS TECHNOLOGY

Methods in General Anatomy



Cornelius T Leondes
editor



Medical Imaging Systems Technology

Methods in General Anatomy

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

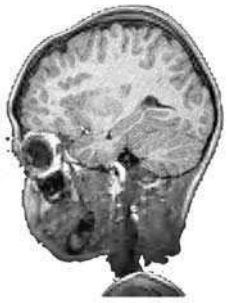
Analysis and Computational Methods
ISBN 981-256-993-6

Modalities
ISBN 981-256-992-8

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

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

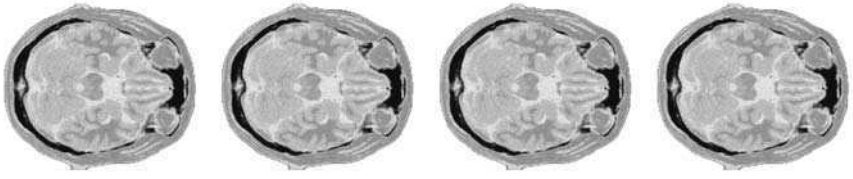
Methods in Cardiovascular and Brain Systems
ISBN 981-256-989-8



Medical Imaging Systems Technology

Methods in General Anatomy

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
Methods in General Anatomy

ISBN 981-256-364-4 (Set)
ISBN 981-256-991-X

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 — methods in general anatomy, 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 multivolume 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 — Methods in General Anatomy" includes essential subjects like:

(a) Medical imaging analysis of the three dimensional (3D) architecture of trabecular bone: Techniques and their applications
(b) Medical image-based preformed titanium membranes for bone reconstruction
(c) Techniques for tracheal segmentation in medical imaging
(d) Knowledge-based system for contouring the spinal cord in computed tomography images
(e) From global to local approaches for non-rigid registration
(f) Automated image segmentation: Issues and applications
(g) Techniques in image guided surgery based on integrated rate sensing, segmentation and registration framework methods
(h) Image registration and fusion for interventional MRI-guided treatment of prostate cancer
(i) Detection and segmentation of Drusen deposits on retina images

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

CHAPTER 1

MEDICAL IMAGING ANALYSIS OF THE THREE DIMENSIONAL (3D) ARCHITECTURE OF TRABECULAR BONE: TECHNIQUES AND THEIR APPLICATIONS

AGOSTINO P. ACCARDO*
Department of Electronics (DEEI), University of Trieste
Via Valerio 10, I-34127 Trieste, Italy
accardo@deei.units.it

IGOR STROLKA†
Department of Radiology, University of Trieste
Cattinara Hospital, I-34149 Trieste, Italy
umerstro@savba.sk

RENATO TOFFANIN
Department BBCM, University of Trieste
Via Giorgieri 1, I-34127 Trieste, Italy
PROTOS Research Institute, P. O. Box 972, I-34100 Trieste, Italy
toffanin@units.it

FRANCO VITTUR
Department BBCM, University of Trieste
Via Giorgieri 1, I-34127 Trieste, Italy
vittur@bbcm.units.it

The purpose of this chapter is to provide a perspective on the current techniques in the imaging analysis of the three-dimensional architecture of trabecular bone and their relevance to the diagnosis of osteoporosis. The emphasis lies on the analysis of images obtained by high resolution X-ray-based CT and MRI techniques. The description of these acquisition techniques is followed by a presentation of the most common image processing methods. Different approaches (morphological, topological, fractal, etc.) used to derive the main architectural features of trabecular bone are illustrated and discussed.

Keywords: Image analysis; MRI; X-ray tomography; trabecular bone architecture; structural parameters and models.

1. Introduction

Mechanical support, maintenance of calcium homeostasis and hematopoiesis in the bone marrow are the main functions of the skeleton. Strength and low weight of the

*Corresponding author.
†Permanent address: Institute of Measurement Science, Slovak Academy of Sciences, Dubravska Cesta 9, SK-84104, Bratislava, Slovak Republic.

skeleton are given by a sophisticated spatial distribution of two main kinds of bone: (a) cortical or compact bone, and (b) trabecular or cancellous bone. The cortical bone, which contributes about 80% to the skeleton weight, is found principally in the shafts of long bones. It mainly consists of a number of irregularly spaced, frequently overlapping, cylindrical units, termed Haversian systems.[1] The trabecular bone forms about 20% of the skeleton and is found principally at the ends of long bones, in vertebral bodies and in flat bones. This kind of bone is composed of a meshwork of trabeculae within which are intercommunicating spaces containing the bone marrow. Typically, the thickness of the trabecular elements is around 100 μm in humans. As a consequence of a significantly larger surface area, the trabecular bone is more metabolically active compared to the cortical bone. Therefore, the skeleton osteoporotic status can be more easily evaluated through the investigation of trabecular bone.

In 1994, WHO defined osteoporosis as *a systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture.*[2] The disease is usually developed without any symptoms for several decades while the bone is losing its sophisticated structure until an osteoporotic fracture occurs as a result of minimal injury. Most often the osteoporotic fractures of the proximal femur (hip), distal radius and vertebral bodies arise. These fractures have significant consequences to the quality of life of the elderly population. It was estimated that in developed countries about 5–6% of the population suffers from osteoporosis. Up to 1 in 2 women and 1 in 3 men will sustain an osteoporotic fracture during their lifetime. In the USA, the total cost for treating all types of osteoporotic fractures was estimated at $14 billion in 1999.

The bone osteoporotic status is routinely estimated by the bone mineral density (BMD) parameter based on X-ray measurements (e.g. DXA or QCT measurements). However, a considerable overlap in density measurements between patients with and those without fractures exists. The variations of density explain around 60% of the bone's mechanical competence.[3,4] Therefore, it is necessary to find new reliable methods for the evaluation of the bone osteoporotic status. It was suggested and experimentally well proven that changes in the trabecular bone structure affect the bone strength independently of the bone mineral density.[5] In order to evaluate the bone structure quality, it is possible either to measure some physical parameter, which is affected by the bone structure (e.g. the T_2^* relaxation constant of the MR signal, the ultrasound attenuation or velocity, etc.) or directly depict the three-dimensional structure of the trabecular bone using X-ray or MR tomographic imaging and analyze its properties.

Imaging of the trabecular bone structure and the analysis of the acquired images are currently limited by the fact that the thickness of a trabecular element is comparable to the highest resolution achievable *in vivo* by both X-ray and MR imaging methods. This results in the partial volume blurring. Other artifacts, which further hamper the bone analysis are characteristic of a given imaging modality.

2. Image Generation

There are three main approaches to generate 3D trabecular bone images: *serial sectioning techniques*, *X-ray Computed Tomography* (*CT*) and *Magnetic Resonance Imaging* (*MRI*). Serial sectioning techniques provide high-resolution high-contrast images but this approach is limited to bone biopsy specimens (is invasive), and prevents further analysis of the specimen (is destructive). The micro-computed tomography (μCT) images may have a resolution similar to the images obtained by the serial sectioning approach, whereas being completely nondestructive. For small animals studies, *in vivo* trabecular bone imaging may be carried out as well. The *in vivo* imaging of human skeleton peripheral sites may be performed using dedicated peripheral CT scanners (pQCT) or using clinical whole body MRI scanners using dedicated coils and imaging protocols.

Because of the different physical principles governing the image generation, each imaging modality introduces specific artifacts into the processed images. Therefore, the processing of acquired images and the evaluation of quantitative parameters, which describe the bone structure, may have different requirements for each imaging method.

2.1. *Histomorphometry*

The standard procedure to assess structural features of trabecular bone is based on 2D sections of bone biopsies. Three-dimensional morphological parameters are derived from the 2D images using stereological methods.[6] These methods are prone to error because of the inherently three-dimensional and highly anisotropic nature of the trabecular bone structure.[5] Moreover, there is no relation between the connectivity in a random 2D section and the connectivity in 3D space.[7] The methods, therefore, do not provide a reliable means to characterize the 3D bone topology.

Early 3D reconstructions were based on tedious and time consuming serial sectioning techniques, which did not allow routine applications. The introduction of new computerized methods[8,9] have significantly improved the situation. These methods, however, require substantial preparation of the samples. The throughput of automated serial sectioning techniques may be as high as 600 sections per hour.[9] With the implementation of the computer numerically controlled (CNC) milling machine[10] thin slices are serially removed from embedded bone specimens and acquired by a CCD camera. This imaging technique generates high resolution, three-dimensional models of trabecular bone, in which image resolution is independent of specimen size.

The advent of X-ray computed tomography as well as of MRI allows to carry out the study also *in vivo* although a high resolution like that present in histomorphometric images is still available only *in vitro*. Compared to these imaging modalities, the main disadvantage of the serial sectioning method is the specimen destruction in the process of imaging, which prevents the use of the sample for further measurements, e.g. mechanical testing.

2.2. *X-ray computed tomography*

The contrast in X-ray images of the trabecular bone structure originates from the large difference in the attenuation coefficients between the mineralized bone matrix and the bone marrow. Apart from the 60–80% greater physical density of the mineralized bone compared to soft tissues, this difference can be explained by the presence of higher atomic number elements, P and Ca, which being only 10% the number of atoms of the mineralized bone matrix, absorb together more than 90% of X-ray photons in the range of 20–40 keV.[11] In the X-ray images, mineralized bone appears bright while spaces containing the bone marrow are dark since they have a low attenuation coefficient.

Conventional radiography (Fig. 1) provides 2D projection images of the bone structure with spatial resolutions up to 40 μm at peripheral skeletal sites.[12] However, the relationship between the complicated three-dimensional trabecular bone architecture and the associated two-dimensional radiographic pattern is not well understood at present. Results of a study, where plain radiographs were simulated using 3D μCT images,[13] suggest that the information on the average architecture fabric, i.e. mean intercept length, is contained in 2D images.

3D images with in-plane resolution up to 400 μm and slice thickness about 1 mm can be obtained using clinical whole body CT scanners (Fig. 2).[12] However, this

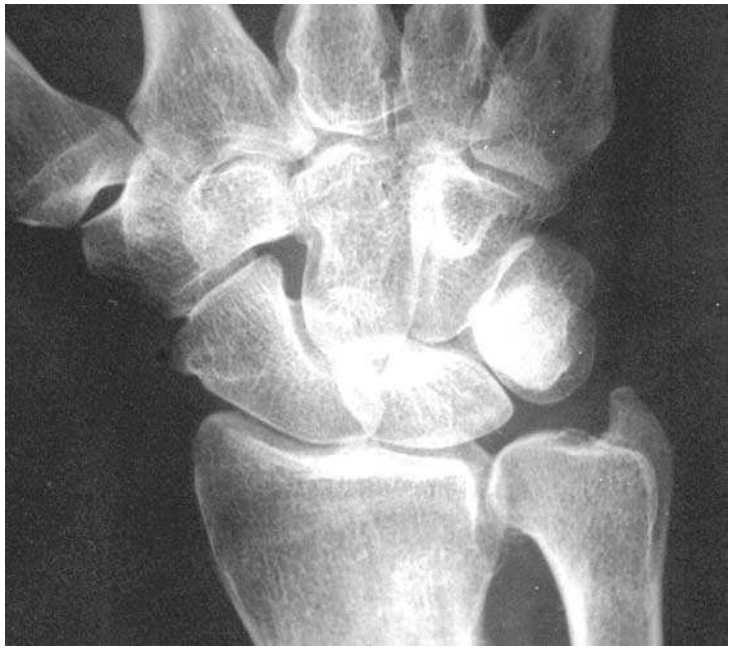


Fig. 1. Radiograph of the right hand of a healthy volunteer where the trabeculae are not well identifiable (Courtesy G. Guglielmi — IRCCS *Casa Sollievo della Sofferenza*, S. Giovanni Rotondo, Italy).

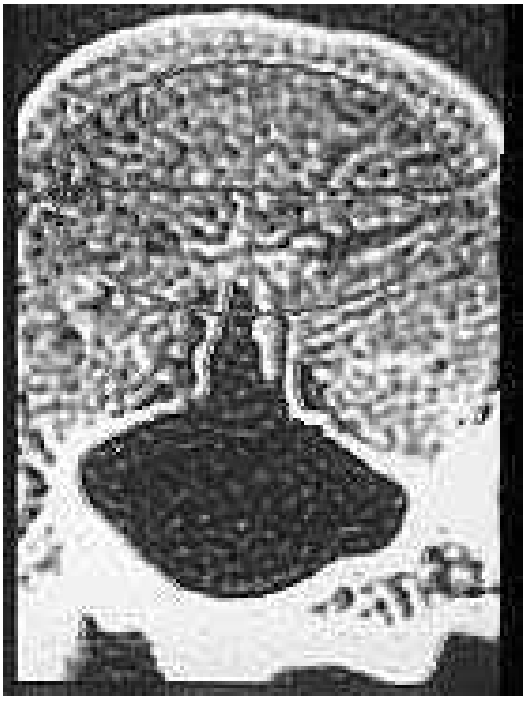


Fig. 2. CT image of a human vertebral body where the trabeculae are well identifiable. (Courtesy G. Guglielmi — IRCCS *Casa Sollievo della Sofferenza*, S. Giovanni Rotondo, Italy).

resolution is not sufficient for the accurate depiction of individual trabeculae. Special purpose small gantry CT scanners were built to acquire high-resolution images (compared to the usual CT scan) of the trabecular bone structure in peripheral sites of the skeleton (distal radius).[14,15]

The experimental low-dose 3D-QCT system, developed in Zurich by Rüegsegger and coworkers,[16,17] provides images with approximately isotropic resolution of about 250 μm (reconstructed 3D images had an isotropic voxel size of 170 μm). This system makes use of a 0.2×10.0 mm line-focus X-ray tube and a 10×16 channels 2D detector array allowing the simultaneous assessment of 10 tomographic images in each measurement cycle. Acquisition of 60 high-resolution slices can be performed in 10 minutes. Even though at this resolution excessive partial volume blurring of trabecular elements occurs, sophisticated image processing methods[16] allow the bone structure extraction, visualization and characterization. Furthermore, the system showed an excellent reproducibility.[17] Recent advances in the X-ray detector technology allowed to increase the *in vivo* resolution of pQCT. A system providing images having 100 μm isotropic resolution is commercially available now.[18] However, the maximum tolerable radiation dose limits the further increase of the resolution.

Much higher resolutions can be reached using μCT scanners dedicated to *in vitro* measurements of small bone biopsy samples or to *in vivo* studies of small animals. Most of these systems use a fixed X-ray source and a detector array whereas the specimen is rotated. Among the technical factors affecting the resolution of the CT scanners are the X-ray source focal spot size, the detector array resolution and the specimen position with respect to the source and detector.[19] The position of the sample governs its geometrical magnification being projected in the detector plane. Placing the specimen closer to the X-ray source improves the effective resolution. However, with this increase of the geometric magnification, the finite size of the detector array limits the field-of-view of the system and the blurring due to the X-ray focal spot size increases as well.

Microfocus X-ray tubes with stationary targets are employed in some μCT systems. X-ray sources with focal spot sizes as small as 5 μm are commercially available. Since the maximum power that can be applied to a stationary target of the microfocus X-ray tube is proportional approximately to the focal spot diameter (because of the heat removal rate from the target),[19] the decrease of the focal spot size reduces the geometrical blurring of the image at the expense of a lower beam intensity, which in turn means longer exposure times for a given signal-to-noise ratio.

A linear array of solid-state or gas ionization X-ray detectors are used in clinical scanners. The in-plane spatial resolution is to a large extent determined by the effective detector width, whereas the slice thickness is determined by the thickness of the beam or by the detector height. In this case, fan beam reconstruction algorithms are applied to obtain 2D tomographic images. In radiography as well as in μCT devices this kind of detector is often replaced by a thin scintillating (or phosphor) layer, optically coupled to a 2D CCD chip. In μCT the 3D images are reconstructed directly applying a cone beam reconstruction algorithm.[20] An example is the experimental μCT scanner developed at the Physical Department of the University of Bologna.[21] Here, the scintillating Gd_2O_2STb layer is deposited on the 40×20 mm entrance window of a 2:1 glass fiberoptic taper. The small end of the taper is coupled to a Peltier cooled CCD with 1024×512 useful pixels. Experimentally accessed Nyquist cut off of the detection system is at about 16 lp/mm.[21] This system was employed to visualize the structure of cylindrical bone samples with a diameter of approximately 7 mm.[22]

A common problem affecting these systems concerns the noise influence on image quality. In μCT imaging the finite number of X-ray photons emitted by the X-ray source and the small element sizes of the detector lead to fluctuations of measured intensity values, which follow the Poisson statistics. In this case, the standard deviation σ, about the mean number of X-ray quanta, N_0, falling on a detector element, is equal to $\sqrt{N_0}$. The noise distribution in the reconstructed image is often considered to be close to Gaussian.[23] Even in the case when the noise in individual projections is white, there is more noise at high spatial frequencies than at low frequencies in the reconstructed image. In fact, the power spectrum depends on the

reconstruction algorithm actually applied. Analysis of the noise power spectrum properties in CT images has been described by Kijewski and Judy.[24]

Another problem bound to the use of conventional X-ray tubes, which produce a polychromatic X-ray radiation, is the beam hardening. Beam low-frequencies absorption occurs while the beam travels through the measured object because of the energy dependence of X-ray radiation absorption coefficients (the beam becomes "harder"). As a consequence, the measured CT values tend to be higher near the edges of the object. Aluminum or copper filter plates[11] can be employed to reduce the spectral width of the X-ray beam selectively, removing the low energies. However, this results in great loss of the X-ray beam intensity. Software methods for the reduction of the beam hardening artifacts were developed too,[25,26] but the application of these methods may cause some loss of the information in the reconstructed images. Because of these limitations, the highest resolution of μCT systems with microfocus X-ray sources is limited to approximately 10 μm.

Employing synchrotron X-ray source instead of conventional X-ray tubes allows the acquisition of images with significantly higher resolutions (Fig. 3). Synchrotron radiation offers the possibility to select X-rays with a small energy bandwith, thus no beam hardening occurs. Because of the high photon flux of the X-ray beam, high signal-to-noise can be retained even when the image resolution is increased. Moreover, the X-ray beam is parallel as opposed to the cone-beam generated by conventional X-ray sources, thus reducing some image distortions. In trabecular bone images with a resolution of around 3–5 μm features like osteocyte lacunae on the bone surface and trabeculae microcracks were seen.[27]

2.3. *Magnetic resonance imaging*

Recent efforts have been directed toward quantifying trabecular bone architecture non-invasively. In particular, MR imaging has emerged as a potential method for determining both bone density and structure. In trabecular bone images, acquired by MRI techniques, the structures corresponding to the mineralized bone matrix appear dark, whereas regions with large brightness values correspond to the intertrabecular spaces filled with bone marrow. This contrast originates from the significant difference between the transverse (spin-spin) relaxation constants (T_2) of protons present in the bone tissue and those present in the bone marrow, which consists mainly of water and lipids (fat). Since these molecules are highly mobile in the bone marrow, their T_2 relaxation times are around 30–50 ms at 1.5 Tesla.[28] On the other hand, the bone tissue contains about 15% of water molecules (interstitial or bound to collagen or to hydroxyapatite) having a T_2 of about 250 μs.[29] This proton pool, characterized by very low T_2 values, cannot be detected when liquid-phase proton MRI sequences are used. For this reason, a negative image of the bone structure is generally produced by MRI.

As a consequence of these different chemical composition, bone and bone marrow exhibit a difference in magnetic susceptibility (about 2.5–3 ppm). Also this

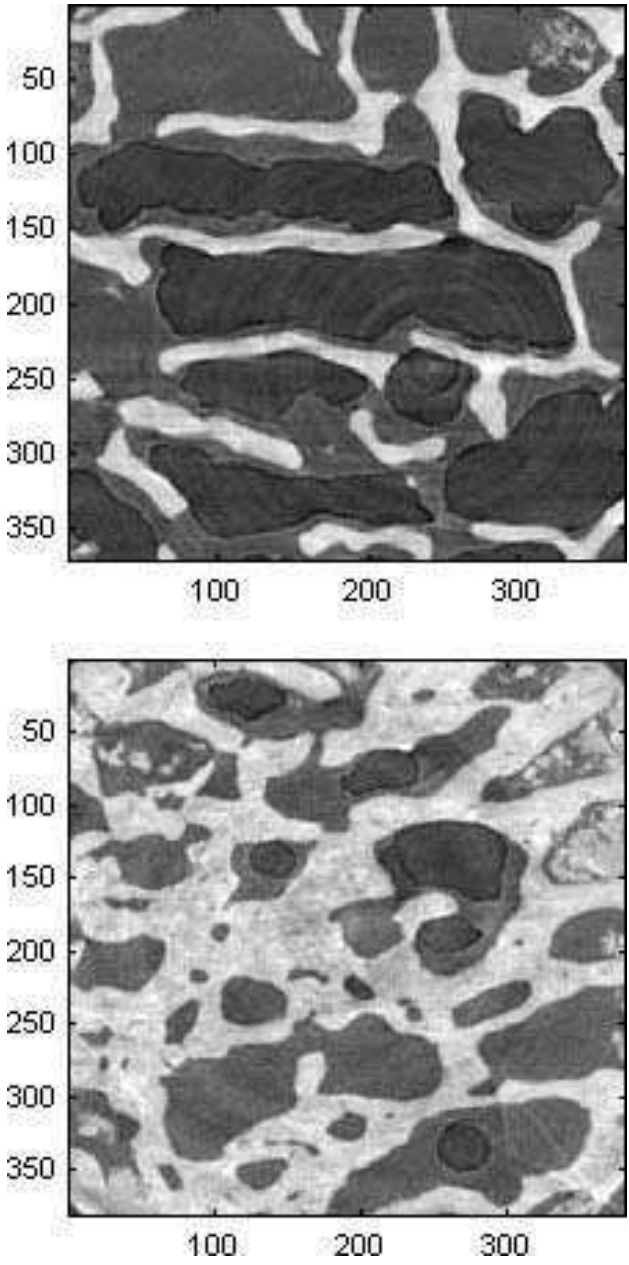


Fig. 3. μCT images of bovine femur explants with an in-plane resolution of 14 μm. (Courtesy D. Dreossi, ELETTRA Synchrotron Light Source — Trieste, Italy).

difference is relevant for the analysis of the trabecular bone.[30,31] In fact, a field perturbation is induced near the boundary of two materials having different susceptibility. The effective transverse relaxation time (T_2^*) also provides a valuable physical mechanism for the quantitative assessment of the bone structure.[31,32] The field distribution is less homogeneous in samples with a large spatial density of trabeculae compared to those which are less dense. This field inhomogeneity may introduce severe artifacts in trabecular bone images like an apparent amplification of trabeculae dimensions if the images are acquired using gradient echo imaging sequences.[33,34] In this case, the magnetic field spread, induced near the boundary of the bone and marrow, causes a dephasing of the transverse magnetization and thus a severe signal attenuation from the unity volume of the bone marrow compared to the same volume of the marrow positioned in regions with low field variation. This artifact can be reduced using the gradient echo imaging sequence with short echo times, which requires the use of a space encoding gradient system with high amplitudes and slew rates, and/or partial echo acquisition schemes. Since the application of the 180° refocusing pulse restores the loss of the phase coherence caused by the field inhomogeneity, spin echo sequences are not sensitive to this kind of artifact (Fig. 4). At high fields, the susceptibility artifacts are more pronounced, requiring wider acquisition bandwidths.

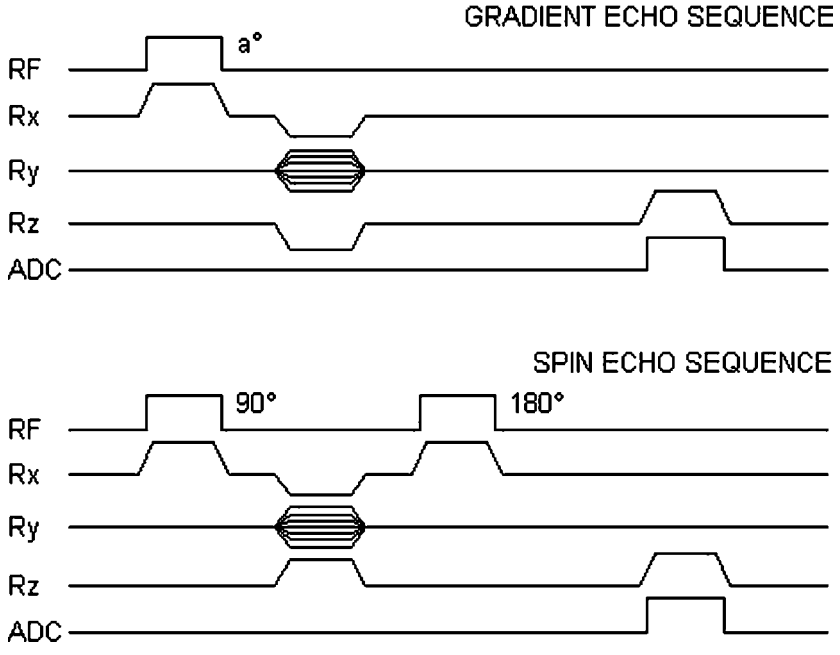


Fig. 4. Timing diagrams of the standard 2D gradient-echo and spin-echo sequences. RF = Radio Frequency transmitter; Rx = slice selection gradient x; Ry = phase encoding gradient y; Rz = frequency encoding gradient z, also called readout gradient; ADC = Analog to Digital Conversion and sampling; a° = flip angle between 0° and 90°.

Another kind of distortion of the trabecular bone images occurs as a direct consequence of the application of controlled magnetic field gradients, with intensity linearly varying with some spatial coordinate, necessary for the spatial encoding of the MR signal. The field distribution in the sample, affected by field inhomogeneity produced by the susceptibility difference, is further influenced by these gradients. In the case of spin warp imaging, this results in the apparent shift of spins in the read gradient direction with consequent image distortions[35] and loss of the resolution. These artifacts can be minimized by the application of stronger read gradients. However, the increase of the acquisition bandwidth (which is proportional to the gradient strength) reduces the signal-to-noise ratio.

The resolution of the acquired trabecular bone images is fundamentally limited by the acceptable signal-to-noise ratio (SNR), since the signal available from each volume element decreases as the voxel size is reduced. The noise in MR experiments is typically dominated by the receiver coil thermal noise and by the noise arising from *r.f.* field energy losses in the sample (the most significant loss originates from magnetically-induced eddy currents). In the case of very small bone biopsy samples, the coil noise dominates over the sample noise and SNR is approximately proportional to the square of the resonance frequency.[35,36] For larger samples, the sample noise becomes more important, and the SNR increases linearly with the resonance frequency.[5]

Dedicated coils are applied in order to obtain images with high SNR (e.g. the quadrature birdcage coil for microimaging of the wrist[37]). In fact, the coil sensitivity increases when the coil diameter is reduced.[5,35,38] Therefore, peripheral sites of the skeleton, which can be positioned into small coils, have to be selected for high-resolution trabecular bone MRI. The SNR increases with the square root of the number of accumulations. In the case of *in vivo* studies, however, the imaging time cannot exceed 10–20 minutes, therefore, typically no more than two accumulations can be acquired.

The dominant noise is uncorrelated (white) noise with Gaussian distribution. The real and imaginary components of the noise signals are not mutually correlated. In the case of the Cartesian k-space sampling, commonly applied in MRI sequences, the image is obtained as a 3D Fourier transform of the MR signal. Since the Fourier transform is an orthogonal linear operation, the noise signal present in both the real and the imaginary components of the image is still independent, spatially invariant and Gaussian distributed.[36] It is common, however, to work on magnitude images, which are obtained as the absolute value of the complex image. The computation of the absolute value is a nonlinear operation and the noise in the resulting images is characterized by a Rician distribution.[36,38] The conditional probability that the measured intensity is I_{meas}, given that the actual intensity is I_{act}, is given by

$$P_{meas|act}\left(I_{meas}|I_{act}\right) = \frac{I_{meas}}{\sigma^2}\exp\left[-\frac{I_{act}^2 + I_{meas}^2}{2\sigma^2}\right] I_0\left(\frac{I_{act}I_{meas}}{\sigma^2}\right). \qquad (1)$$

Here I_0 is the modified zeroth-order Bessel function of the first kind and σ is the standard deviation of the noise in the acquired signal.

The point spread function in MR images acquired with Cartesian k-space sampling is $sinc(\pi x/\Delta x)$, with Δx representing the voxel size. Since the width of this function is comparable to the voxel size, it is often assumed for simplicity, that the point spread is given by a box function equal to 1 in the voxel's volume and 0 otherwise and the partial volume blurring model is applied in image processing methods.

It has to be noted, however, that for non-Cartesian acquisition methods (e.g. Radial MRI[39]) the noise properties and the point spread function will be different. The same is true for recently proposed parallel acquisition methods with coil sensitivity encoding (e.g. SENSE), which allow a considerable reduction in scan time by the use of multiple receiver coils in parallel.[40] In this case, the noise signals in the receiver channels are correlated, and the reconstruction matrix is not unitary. Consequently, the noise in the resulting image is correlated and the noise level is spatially variable.[40] The noise may become spatially variant also when methods for the $r.f.$ field inhomogeneity correction are employed.

High field (typically 7 Tesla or more) MR microscopes are used to obtain 3D high-resolution images of small bone samples (Fig. 5). From these bone samples,

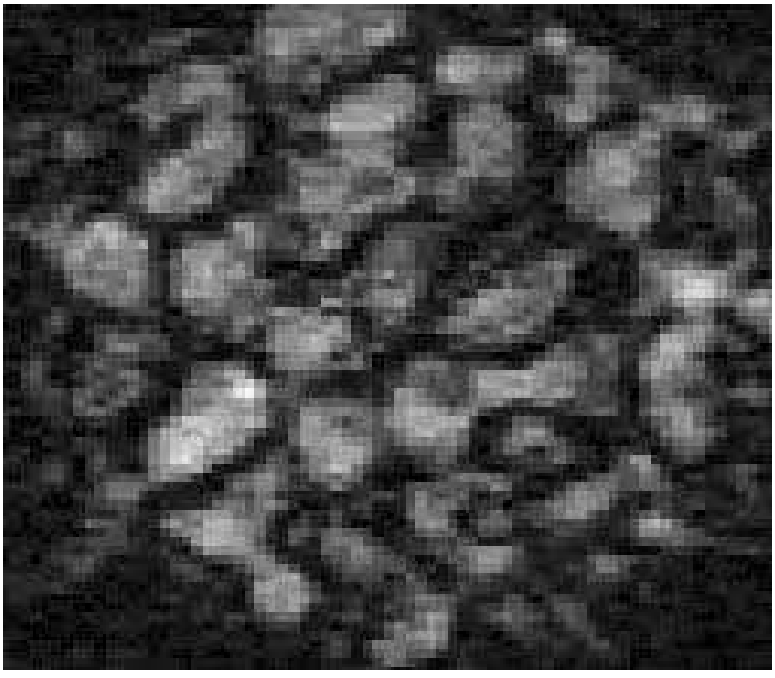


Fig. 5. Spin-echo μMR image of a porcine humer explant obtained at 7.05 Tesla (in-plane resolution of 42 μm).

usually the bone marrow is removed. One reason for this is the transition of the hemoglobin in hematopoietic bone marrow from the diamagnetic oxy to the paramagnetic deoxy state after the contact with air. Therefore, the magnetic properties of the sample would be changed in comparison to the intact bone. Another rationale for the marrow removal is the possible presence of air cavities in the sample.[32] The abrupt change of the magnetic susceptibility at the air–marrow boundary would affect the image appearance significantly. Therefore, the defatted bone structure is generally refilled with doped water. Images with pixel volumes ranging from $92 \times 92 \times 92\,\mu\text{m}^3$ [41] to $47 \times 47 \times 47\,\mu\text{m}^3$ [38] or $42 \times 42 \times 84\,\mu\text{m}^3$ [39] were obtained in various studies. Because of high magnetic fields, the spin echo imaging sequences were preferred in order to avoid the susceptibility induced trabecular thickness amplification.

Both spin echo and gradient echo imaging sequences are applied in trabecular bone structure studies in various sites of the skeleton *in vivo* using 1.5T clinical scanners (Fig. 6). Studies of the calcaneus with nominal resolutions of

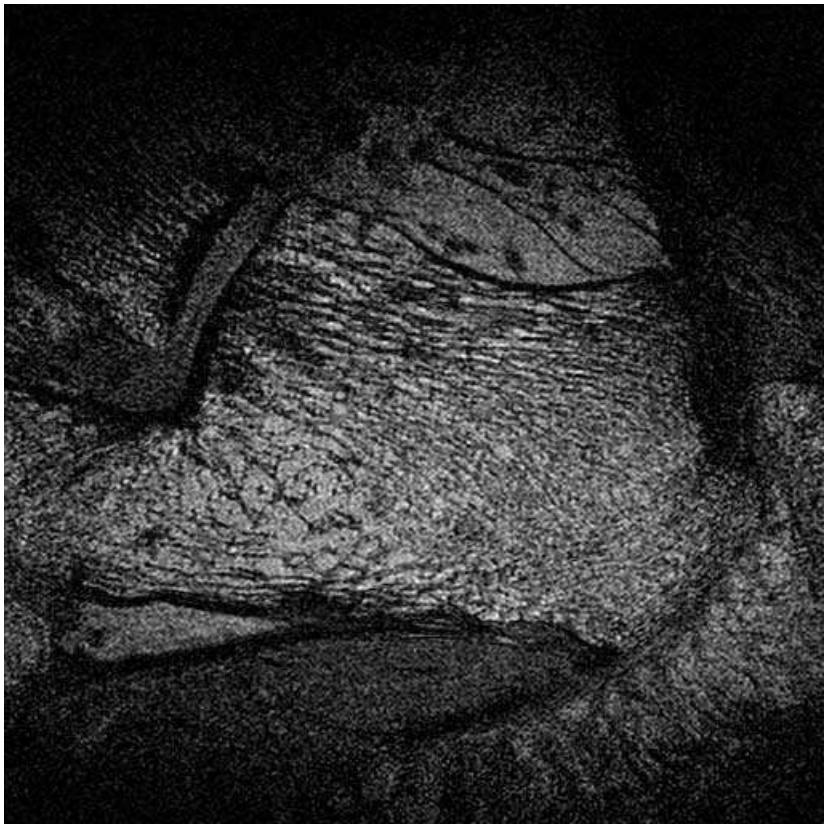


Fig. 6. High resolution gradient-echo MR image of the right calcaneus of a young healthy man obtained at 1.5 Tesla using a two-element experimental microcoil (in-plane resolution $156 \times 156 \times 500\,\mu\text{m}^3$).

$172 \times 172 \times 700\,\mu\text{m}^3$ [42] or of $195 \times 195 \times 1000\,\mu\text{m}^3$ [43] and of the distal radius with resolution of $156 \times 156 \times 500\,\mu\text{m}^3$ [44] were performed. In all of these cases, the gradient-recalled imaging sequences were applied. It is argued, that gradient echo sequences can acquire the 3D images of the trabecular bone structure in a considerably less time than conventional spin echo sequences and, with proper parameters settings, the image distortions can be reduced.[34] However, a modified spin-echo sequence was proposed (i.e. Fast 3D Large-Angle Spin-Echo — 3D FLASE), which reduces the imaging time considerably.[33] In this case, a large angle excitation pulse is applied, which inverts the longitudinal magnetization. The subsequent 180° refocusing pulse restores the inverted longitudinal magnetization. Using this imaging sequence, distal radius images with resolution $135 \times 135 \times 350\,\mu\text{m}^3$ and with SNR ~ 8 were acquired. A limitation of this imaging technique is the higher *r.f.* power requirement compared to the gradient-recalled steady state acquisition protocols.

For *in vivo* measurements, the quality of the acquired images and the precision of the bone architecture analysis can be significantly reduced by the presence of motion artifacts due to the long time necessary for the acquisition. Several research groups apply dedicated fixation tools for the immobilization of the studied site (e.g. foot holder),[42,43] which also provide a reproducible positioning for successive scans of the same subject. Another efficient solution of this problem is the use of navigator echoes.[37] In this case, an additional set of 1D projections of the measured object in both x- and y-axes (in-plane) in each phase-encoding step is acquired and these projections are used to correct translational motion. The motion along the z-axis direction is not compensated, taking into consideration the greater slice thickness compared to the in-plane pixel size. The shift between the reference and the current navigator projections is estimated using the least-square approach.

3. Image Processing

Most of the methods which provide the quantitative characterization of the trabecular bone structure, require a preliminary image processing involving the *binary segmentation* of the acquired image, i.e. assigning the voxels to the mineralized bone structure or to the intertrabeculae spaces. There are, however, techniques able to derive architectural parameters from bone volume fraction (BVF) maps, since the image intensity values are proportional to the bone amount in the corresponding voxels in the volume. Random and structured noise, partial volume effect (blurring) and spatial sensitivity variations of the imaging system may make both the voxel classification and the BVF map estimation a very demanding task.

3.1. *Binary segmentation*

The most straightforward way to the image voxel classification is by the use of a simple threshold.[20] Voxels with intensity values above the threshold are assigned to one material phase (e.g. bone in the case of X-ray CT imaging or marrow in the

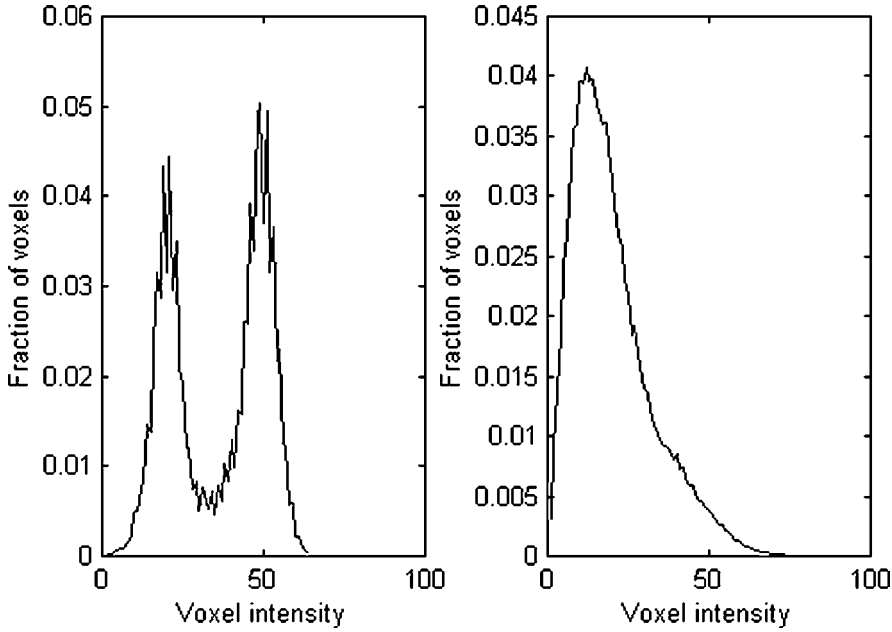


Fig. 7. Image intensity histograms presenting one (right) or two (left) peaks. The images were obtained by μCT and MR imaging systems, respectively.

case of MRI), while those with lower intensities are assigned to the other material phase. In the case of high-resolution (e.g. $14 \times 14 \times 14\,\mu\text{m}^3$) trabecular bone images with high signal-to-noise ratio, acquired using the μCT scanners with synchrotron radiation, the image intensity histogram is clearly bimodal, with two relatively narrow peaks corresponding to the bone and to the intertrabeculae spaces (Fig. 7). The threshold may be estimated as a histogram minimum lying between these two peaks.[45] Another simple thresholding scheme was applied to bimodal trabecular bone images with a resolution of $47 \times 47 \times 47\,\mu\text{m}^3$ acquired by μMRI.[46] In this case, the bone sample was immersed in water. Because of the presence of Rician noise, the water peak is approximately Gaussian, while the bone peak follows the Rayleigh distribution. The water peak maximum was estimated as the mean intensity value of the water in intertrabeculae spaces and the threshold was selected to be 50% of this value (assuming zero signal formed by protons bound in the bone structure).

The beam hardening effect and the noise present in trabecular bone images obtained using the μCT systems with microfocus X-ray sources, require a more careful selection of the threshold value, since no significant isolated peaks may appear in the image histogram. Rossi *et al.* suggested a threshold estimation based on the second derivative of the intensity histogram.[21] Three thresholds were found as zeros of the derivative function, the first corresponding to the background, the second to the marrow peak while the third one was the marrow-bone transition threshold.

Estimation of the threshold value may become a nontrivial task, especially in the case of images acquired *in vivo*. In this case, because of the noise fluctuations as well as of the limited spatial resolution, the intensity histogram of images obtained *in vivo* does not consist of two separate peaks representing the bone and the bone marrow. Such a kind of histogram is called monomodal (Fig. 7).

Empirical approaches were applied to the threshold selection in the case of *in vivo* MRI images of the calcaneus and the distal radius. Ouyang *et al.*[43] described a procedure applied to calcaneus images with resolution $195 \times 195 \times 1000\,\mu\text{m}^3$ where a low intensity value was estimated as the average brightness of three different regions, manually selected in the processed image. These regions correspond to air, tendon and cortical bone. The high intensity value was estimated in selected regions of the bone marrow and subcutaneous fat. The resulting threshold value was then set to lie between the estimated low and high values, employing an empirical constant governing the threshold selection.[43] This "internal calibration" procedure served to optimize the threshold value to the scan-to-scan signal and noise variations due to MR scanner adjustments.

The idea of "internal calibration" was preserved in another segmentation method, which was applied to distal radius images with resolution $156 \times 156 \times 700\,\mu\text{m}^3$.[44] The noise variation was suppressed using a 2D median filter with a kernel of 3×3 voxels. The gray scale of the acquired images was reversed to ease the trabecular network and the cortical rim visualization. The expected mineralized bone brightness was specified as a mean intensity value in the cortical rim region, whereas the marrow brightness level was empirically estimated from the trabecular bone region intensity histogram, assuming that the most frequently occurring pixels in this region consist of a mixture of trabecular bone and bone marrow. Therefore, this value was taken as a lower intensity value at which the histogram reached half of its peak value. The apparent bone volume fraction was estimated assuming the expected intensities of bone and marrow phases and from the computed mean intensity in the trabecular bone region. The threshold was then selected as a value, at which the fraction of the number of bone voxels after segmentation to the total number of voxels in the Region of Interest (ROI) was equal to the apparent bone volume fraction. This standardized method of threshold selection ensured consistency of the image processing across all the subjects studied.

Since the classification of a voxel, when threshold segmentation schemes are applied, is based only on its intensity value, this kind of a segmentation algorithm is sensitive to the presence of noise.[45] Various kinds of denoising algorithms may be applied to suppress the noise variation. Among these, 2D or 3D linear or median digital filters are the most simple, both causing a significant blurring of the image. These noise suppression methods may not be suitable in the case, when the voxel size exceeds the individual trabeculae dimensions (typically in the case of *in vivo* imaging). More sophisticated image processing methods, e.g. adaptive digital filters or nonlinear diffusion filters, may be more effective, especially when taking

into consideration the spatial anisotropy of the trabecular network as well as the anisotropy of the image resolution.

Another difficulty bound to the thresholding segmentation approaches is the sensitivity to intensity inhomogeneities, which can occur in MR images.[45] This may lead to an apparent thickening of the trabeculae or to a loss of thinner trabeculae within the same image.[44] For *in vitro* studies, a calibration measurement using a homogenous specimen may be performed to estimate the sensitivity map.[39] However, the coil placement may differ significantly between individual experiments performed *in vivo*. In this case, it is necessary to determine the sensitivity maps of the receiver coils with the definitive set-up in addition to actual imaging protocol[40] or to estimate a sensitivity map by analyzing the local intensity histogram in various regions of the image.

Besides the thresholding approach, various other segmentation methods are applied to trabecular bone images. To process images acquired using a peripheral CT system having an isotropic voxel size of 170 μm, an in-plane resolution of 250 μm and a slice thickness of 480 μm, an edge detection segmentation method was proposed.[16] By this approach, the image is first filtered using a Gaussian filter to suppress the noise variation. Afterwards, the discrete image is approximated locally (within a $5 \times 5 \times 5$ neighborhood of a voxel) using a least-square continuous polynomial fit approximation of the discrete CT image. The used 3D cubic polynomial fit function is defined by 20 coefficients. An orthogonal set of polynomials in the 3D space was found to perform a computationally efficient least square approximation of the discrete image function. Using this polynomial fit approximation, a voxel belongs to the step edge (describing a borderline between two regions with different intensities) only if there is a zero crossing of the second directional derivative within the immediate volume of the voxel. The fit is analyzed following the direction of the gradient computed in the origin of the fit coordinate system. Similarly, a voxel belongs to the roof edge (ridge or valley) only if there is a zero crossing of the first directional derivative following the direction of maximal second derivative (maximal curvature). The roof edges approximate the topology of the object, whereas the step edges are useful to derive the morphometric characteristics of the bone structure.

A 2D watershed algorithm was applied to process images of human lumbar vertebras.[47] The images were acquired using an experimental high-resolution X-ray CT system employing a stabilized X-ray source (90 kVp) and a linear CCD detector with 1024 elements (element size $0.225 \times 0.5\,\text{mm}^2$). The voxel size was $108 \times 108 \times 300\,\mu\text{m}^3$ and the experimentally estimated resolution was about 150 μm within the slice. The intensities of a given pixel and of the pixels situated in its neighborhood were compared to decide whether the given pixel lies on a slope or represents a local maximum. The gray-scale levels of the pixels located on a slope were decreased, whereas the intensities of local maxima and minima were preserved. The image was updated iteratively until there was no change and a gray-scale watershed image was obtained in this way. The next step is the crest lines detection. A set of four masks

selected four pairs of pixels in the neighborhood of the classified pixel positioned symmetrically with respect to the central classified pixel. If the intensities of both neighboring pixels, for some mask, were inferior compared to a chosen percentage of the classified pixel gray level, the pixel was considered to be on a crest line. The result is a binary "quasi-skeleton". The last step performed is the skeleton expansion in order to include the total area of trabecular bone. This thickening procedure was performed toward the two sides of the bone component in a direction perpendicular to the skeleton. An additional contrast criterion was included to ensure adaptability to variation in gray levels in the original image. Two parameters governing the segmentation process have to be adjusted by a radiologist expert, comparing the original and segmented images. One of them is the percentage of the image function applied in the crest lines detection, the other one governs the contrast criterion in the thickening procedure. Once adjusted, the processing of the acquired images becomes entirely automatic. Comparison of this segmentation technique with a conventional contrast enhancement segmentation method (a modification of the thresholding approach) showed a higher precision of the watershed transformation based approach.[47]

3.2. *Subvoxel methods*

The image processing methods described above assign voxels to the mineralized bone structure or to the intertrabeculae spaces via binary segmentation. Clearly, this approach leads to a significant information loss when the spatial resolution of the imaging system is comparable to the dimension of individual trabeculae. Linear interpolation is a common method applied to increase the apparent resolution of digital images. In this case, however, the interpolated values remain between the intensity values of the original image, which contradicts the expectation that the proportion of voxels having the voxel bone volume fraction values near zero or one is higher in images with increased resolution.[38] Alternative methods were proposed, which make use of prior information about the properties of the trabecular network and allow the increment of resolution of the resulting images.

Wu *et al.* proposed a Bayesian approach to subvoxel tissue classification in MR images of trabecular bone.[48] In this method, the volume of each voxel was divided into M subvoxels, which could be assigned to bone or to bone marrow. For the noise-free image model, it was assumed that the intensity of voxel having the bone volume fraction ν is

$$I(\nu) = \nu \cdot I_{bone} + (1 - \nu) \cdot I_{marrow} \tag{2}$$

where $I_{bone} = 0$ is the intensity of voxels containing only the mineralized bone whereas a constant nonzero intensity I_{marrow} was assigned to voxels including solely the bone marrow. A uniform distribution of the voxel bone volume fraction was expected for boundary voxels. The relation between the intensity distribution in the image with and without the presence of noise is given by Eq. (1). The parameters

of the distribution $p(I)$ (proportion of voxels containing only the bone, or marrow, respectively, and the marrow voxels intensity) as well as the parameter σ of the noise distribution were estimated by a least square fitting of the distribution $p(I_{meas})$ to the histogram of the processed image. This parameters estimation may be reliable, however, only in the case when an image having a clearly bimodal histogram is processed. A more accurate method will be described briefly later in this paragraph.

Each voxel could contain from 0 to M subvoxels, therefore the interval [0, 1] of voxel bone volume fraction was divided into $M + 1$ equal length subintervals, each corresponding to a given number of bone subvoxels in the voxel volume. Knowing the relation between the voxel volume fraction and gray scale values I and I_{meas}, the conditional probability $p(I_{meas}|m)$, where m is the number of subvoxels assigned to the mineralized bone structure, was estimated. This is equal to the likelihood that a given voxel with measured gray-scale level I_{meas} contains m bone subvoxels.

The localization of bone subvoxels within the voxel volume is executed via a Gibbs prior distribution model (which is related to the Markov random field theory). The Gibbs probability of the segmented image is the exponential function of a so-called total energy of the segmented image configuration multiplied by a factor -1.[49] Image configurations having a low total energy are highly probable, whereas the probability of images having high energy is low. The total energy of the image is given by the sum of the so-called potentials of local subvoxel configurations. In the prior image model it was assumed that the potential takes lowest values (meaning a high probability of the local configuration) when neighboring voxels are assigned to the same classes (i.e. all of them are assigned to bone or to marrow, respectively). In this way, the assumptions that a bone subvoxel should be connected to other bone subvoxels and that the intertrabeculae spaces are interconnected, are incorporated into the image segmentation process. The amplitudes of potential for different voxel configurations govern the segmentation method behavior. With the increasing of these amplitudes, the prior model becomes rigid possibly leading to an overmuch smooth segmentation (and to a possible loss of smaller details of the bone structure). Therefore, the value of this parameter has to be optimized.

The product of the image likelihood and prior probability is proportional to the posterior probability of the resulting image configuration. Finding a global maximum of this probability is computationally demanding and, therefore, a "good" local maximum solution was found using a block Iterative Conditional Mode (block ICM) algorithm. A performance evaluation of the proposed segmentation method was carried out using synthetic phantom images. The method was compared with the maximum likelihood threshold segmentation. The fraction of the misclassified volume was significantly higher in the case of the threshold segmentation method compared to the subvoxel approach. It is argued by the authors, however, that the benefits of the subvoxel classification will vanish when the noise variance increases.[48]

3.3. *BVF map estimation*

Hwang and Wehrli[46] proposed a method for the bone volume fraction (BVF) estimation in each voxel of trabecular bone MR images acquired *in vivo*, which showed a considerable accuracy. The histogram of the BVF in a ROI (equivalent to the noiseless intensity histogram) is estimated via iterative deconvolution of the processed noisy image intensity histogram and a bone volume fraction map is computed. The process of iterative deconvolution is straightforward in the case of Gaussian noise. The predicted histogram in the presence of noise is expressed as a linear convolution of the noise distribution and the noiseless intensity histogram estimate in a given iteration. The difference (error) between the measured and predicted intensity histograms is added to the noiseless histogram estimate to obtain a new approximation. These steps are repeated iteratively until a good agreement between the predicted and measured histograms is reached. The problem becomes more complicated in the case of MR images, characterized by Rician noise distribution and spatial sensitivity inhomogeneities of the imaging system. In this case, the local bone marrow intensity is estimated as the most commonly occurring intensity greater than the mean intensity in a selected region surrounding a given location.

Taking into consideration the Rician noise distribution, the maximum likelihood estimate of the voxel intensity in the image without the noise was computed and the intensity shading correction was applied afterwards. The voxels were sorted in the order of decreasing likelihood of containing bone (order of increasing image intensity). Additional criteria (e.g. a local connectivity measure, the intensity extremum — minimum or maximum — in the neighborhood of the voxel) were used to order voxels having the same intensity value. The voxel BVF values were assigned to each voxel according to the most recent noiseless histogram estimate. The highest voxel BVF in the noiseless histogram was assigned to the set of voxels having the highest likelihood of containing bone. Number of these voxels was selected according to this histogram. Similarly, the following set of voxels in the ordered list was assigned to the second highest value of the BVF in the noiseless intensity histogram, etc. In this way the noiseless image estimate was obtained. The intensity shading was applied to this image and the intensity histogram was computed. Using the Rician histogram broadening (Eq. (1)) the predicted histogram of the image in the presence of noise was computed and was compared to the histogram of the acquired image. Their difference, however, must not be added to the noiseless intensity histogram directly, as in the case of deconvolution in the presence of Gaussian noise. This is due to the asymmetry of the Rician distribution, meaning that the maximum contribution from some intensity value in the noisefree histogram may be shifted to some other value in the histogram of the noisy image. Introducing this weighting of contributions, the difference of the acquired and the predicted histograms is converted to the error without the presence of noise, which is, however, still distorted by intensity shading. The latter is corrected by spatially mapping of this shaded error distribution to the distribution of the noiseless image.

The error distribution in the noiseless image is added to the noiseless image intensity histogram to obtain the new noiseless image histogram estimate. These steps are iteratively repeated.

Comparison of the method with a simple thresholding approach showed that threshold segmentation depending on the threshold selection either falsely disconnected trabeculae, or unrealistically inflated the bone volume fraction, whereas the BVF maps obtained using the described method preserved well the connectivity of the trabecular network. The error of the estimated BVF, obtained from computational phantom images, was less than 1% processing, assuming the signal-to-noise ratio higher than 8 in the processed images.

An extension of this algorithm, which increases the apparent resolution of the voxel BVF map was proposed later.[38] The method employs some empirical assumptions, which arise from the prior knowledge about the trabecular bone network properties. Firstly, it is assumed that smaller voxels are more likely to have BVF closer to zero or one, compared to voxels having large dimensions. Secondly, because the bone network is an interconnected structure, bone is expected to be in close proximity to more bone. The volume of each voxel was divided into eight subvoxels (resolution was doubled in each direction), and the total BVF of the voxel was allocated to these subvoxels. The weight of a given subvoxel in this BVF redistribution was computed as a sum of BVF of all voxels adjacent to this subvoxel divided by a sum of BVF of all voxels surrounding given voxel. The voxel's BVF is assigned to subvoxels in the order of decreasing weights. The BVF assigned to the first subvoxel is given as its weight multiplied by the given voxel's BVF. In the case that result is greater than one, the BVF of the voxel is set to one. The same process is repeated with other subvoxels; their weights, however, are multiplied by the residual BVF, which is the part of the BVF that was not yet assigned to some subvoxel. In this way the bone mass in each voxel volume is strictly conserved.

Comparison of the voxel size in processed images ($137 \times 137 \times 350\,\mu\text{m}^3$) and typical trabeculae thicknesses and dimensions of intertrabeculae spaces results in the third assumption: that the occurrence of multiple trabeculae within a single voxel is highly unlikely. Subvoxels are therefore classified as boundary or central voxel using the adjacency and connectivity criteria. To be classified as a boundary voxel, the voxel has to have in its 6- or 4-neighborhood (in the 3D or 2D case, respectively) at least one voxel not containing the bone. This voxel also must not be critical to the trabecular network connectivity. The local connectivity measure is based on the number of disconnected regions in the neighborhood of the voxel, which appear by temporarily assuming the central subvoxel contains marrow only. Boundary voxel weights are set to zero and the bone is distributed among subvoxels. Weights of these subvoxels are set now according to adjacent subvoxels located outside the voxel. The algorithm of the bone allocation was modified in such a way that boundary subvoxels are not assigned any bone unless the BVF of central subvoxels is equal to one.

To compare the accuracy of this method for subvoxel image processing with trilinear interpolation, computational MR image phantoms were created on the base of high-resolution ($22 \times 22 \times 22\,\mu m^3$) μCT trabecular bone images modeling the partial volume effect of MR images acquired *in vivo*. If no noise was added to the phantom image, using the trilinear interpolation, only 26–50% of voxels were assigned to the correct BVF, compared to 85–92% of subvoxels using the proposed method. In the case of SNR ~ 10, using the trilinear interpolation this proportion decreased to 3–5% of voxels, whereas 82–90% of voxels were assigned to the correct BVF using the new subvoxel processing method.

4. Structural Parameters and Models

To determine the bone mechanical properties (e.g. bone strength, bone stiffness, Young's modulus) methods such as compression and bend testings are used. Nevertheless, two important problems arise with all mechanical studies: first, the mechanical properties have almost never been described completely because they vary with the loading orientation due to the anisotropy of trabecular bone; second, the results of tests are only rarely accurate because of the errors and problems in the measurement (see Odgaard[7] for a review). On the other hand, following the Wolff's observations,[50] the mechanical properties of trabecular bone are influenced by its architecture. Thus, analyzing the bone architecture it is possible to study the bone quality.

Other than density (BMD in particular), able to well predict most of the elastic properties and strength,[51,52] trabecular thickness and other parameters like connectivity and anisotropy (i.e. orientation of trabeculae) have been proposed for a complete description of the trabecular bone architecture. Different approaches (morphological, topological, fractal, etc.), characterized by indices extracted from 2D or 3D images, allow the assessment of these features. For some methods, high correlations ($r > 0.84$[53,54]; $r > 0.76$[55]; $r > 0.97$[56]) between the parameters evaluated from 3D high resolution images (like μCT) and those estimated from the corresponding conventional histological sections were found. Others methods to calculate in a non destructive way the bone mechanical properties, based on finite element analysis of bone structure, have also been proposed. Below a description of the most used methods and indices follows.

4.1. *Morphological parameters*

A set of morphometric indices such as *relative bone volume or bone volume fraction* (generally denoted as BV/TV, being BV the Bone Volume and TV the Total Volume), *surface density* (BS/BV, being BS the Bone Surface Area), *trabecular thickness* (Tb.Th), *trabecular separation* (Tb.Sp) and *trabecular number* (Tb.N) is used to characterize the trabecular bone structure describing its morphological properties. Originally, these quantities characterizing the 3D bone structure were

calculated in a ROI of 2D images of thin sections of bone biopsies using stereological methods[57,58] and assuming a constant structure model. The bone volume fraction was set equal to the 2D area fraction of the mineralized bone and the bone perimeter was measured to estimate the bone surface density. The surface density was obtained as $(\textit{2D perimeter})/(\textit{tissue area}) \times k$. For isotropic structures the constant k is $4/\pi$, however, in the case of iliac trabecular bone this value is set to 1.2.[58] Furthermore, it was assumed that the trabecular bone structure consists mainly of interconnecting plates. Applying this, the so-called *parallel plate model*, the trabecular thickness is given by $\text{Tb.Th} = 2 * \text{BV/BS}$, trabecular separation by $\text{Tb.Sp} = 2 * (\text{TV} - \text{BV})/\text{BS}$, and the trabecular number by $\text{Tb.N} = 0.5 * \text{BS/TV}$ or equivalently by $\text{Tb.N} = 1/(\text{Tb.Sp} + \text{Tb.Th})$. The Tb.Th represents the mean trabecular plate thickness, Tb.Sp is a measure of the mean trabecular plate separation (measure the thickness of the marrow cavities) and the Tb.N is the mean number of plates traversed by a line of unit length perpendicular to the plates.

The Marching cubes method, proposed by Lorensen and Cline,[59] permits to triangulate the surface of the mineralized bone phase in order to obtain BS in 3D. In this case, BV can be obtained by approximating with tetrahedrons the volume limited by the triangulated surface used for BS calculation.[60]

For the analysis of 3D trabecular bone images acquired *in vitro* (resolution $92 \times 92 \times 92\,\mu\text{m}^3$) using an MR microscope, Hipp *et al.*[41] subsequently proposed a 3D version of the directed secant algorithm. The three dimensional binarized image was scanned with an array of uniformly distributed parallel test lines spaced $200\,\mu$m apart and the number of test lines intersections with the bone-marrow interface was determined for 128 randomly oriented rotations of the test line grid. The Mean Intercept Length (MIL), i.e. the total line length divided by the number of intersections, for each direction was defined. Apart from the information about the bone anisotropy, which will be discussed later, morphological parameters were estimated assuming the parallel plate model, using this analysis of the binary image. In this case BV/TV is equal to the ratio of the number of voxels that fall within bone to the total number of voxels in the volume of interest (VOI) and the Tb.N is the number of test line intersections with bone-marrow interface per unit test line length. Remaining parameters can be computed indirectly using the parallel plate model assumption.

The resolution of the processed images may significantly affect the morphological parameters estimates. Kothary *et al.*[61] studied images by modeling the resolution of trabecular bone images obtained *in vivo*. Models of MR images having an in-plane resolution of $100\,\mu$m and a slice thickness varying from 100 to $1000\,\mu$m were derived from trabecular bone images acquired using the serial milling system and optical serial reconstruction with resolution $20 \times 20 \times 20\,\mu\text{m}^3$. An image with resolution $240 \times 240 \times 240\,\mu\text{m}^3$ was generated as well to mimic the resolution of *in vivo* trabecular bone CT images. Images were binarized using a threshold segmentation algorithm and the directed secant algorithm was applied to derive the traditional morphometric measures using the parallel-plate model. This study showed that the

increasing of the slice thickness induced a rapid increase of the trabecular thickness, whereas both trabecular number and trabecular spacing diminished somewhat with slice thickness. The error of the estimated parameters was dependent on the imaging plane orientation as well. It was also found that the morphological parameters could be estimated with higher accuracy from images with resolution $100 \times 100 \times 500\,\mu\text{m}^3$ compared to images with resolution $240 \times 240 \times 240\,\mu\text{m}^3$. Because of the dependence on the image resolution, the parameter estimates based on limited resolution images were denoted as apparent (app BV/TV, app Tb.Th, app Tb.N, app Tb.Sp).[44]

These morphological parameters, called also standard stereology measures, were used to quantify the properties of the trabecular bone structure in several studies. The MRI studies of Ouyang *et al.*[43] and Majumdar *et al.*[44] showed a correlation of these parameters with age, osteoporotic status and bone mineral density. In the MRI investigation of Boutry *et al.* on a cohort of osteoporotic men[42] the app BV/TV and app Tb.Sp were proposed as possible predictors of bone fracture risk.

In the work of Hildebrand and Rüegsegger, it was pointed out that due to aging and disease, trabecular bone may change its structure type from the plate-like to the rod-like structure continuously.[62] Figure 8 shows typical examples of plate and rod-like structures in different human skeletal sites. When the bone structure does not correspond to the assumed model, unpredictable errors in estimated parameters may be induced. Verification of the model assumption is a nontrivial task.[7] Thus a series of direct model-independent methods have been proposed.[63–65] For the analysis of high-resolution μCT images of trabecular bone (voxel size $14 \times 14 \times 14\,\mu\text{m}^3$, spatial resolution $28\,\mu$m), Hildebrand and Rüegsegger[63] estimate the bone surface and bone volume using the marching cubes algorithm.[59] From these measures, the values BV/TV and BS/BV follow directly. They define the local thickness in some point p of a structure as the diameter of the largest sphere, containing the point p and which is completely inside the structure.[63] Trabecular thickness (Tb.Th*) is determined as the arithmetic mean value of the local thickness taken over all points in the bone structure. Similarly, measures of the marrow cavities thickness give the estimate of the trabecular separation (Tb.Sp*). Trabecular number is found as the inverse of the mean distance between the midaxes[64] of the observed structure. From the thickness distribution, other interesting parameters, e.g. the thickness variation, may be estimated as well.

Within the framework of a BIOMED I project of the European Union, Hildebrand *et al.*[54] compared model-dependent and their direct estimates of morphological parameters. In their 3D μCT study on a total of 260 human bone biopsies taken from five different skeletal sites (femoral head, vertebral bodies $L2$ and $L4$, iliac crest and calcaneus) the authors showed that trabecular thickness was systematically underestimated with the plate-model assumption (by 17% in the femoral head having the plate-like structure, and by 30% in the rod-like arranged lumbar spine). Moreover, the images depicted in this work showed clearly that the bone structure in the samples of the iliac crest and calcaneus cannot be associated to a single model. However, the accuracy of the direct method depends on the resolution

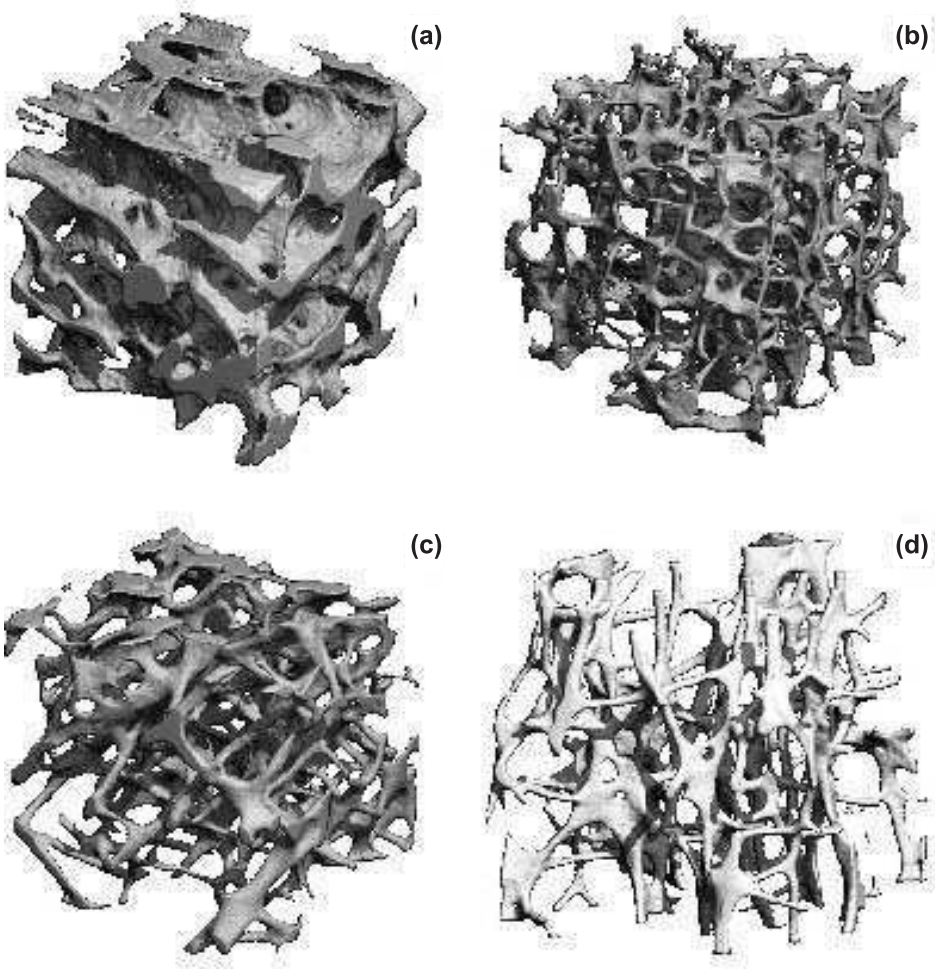


Fig. 8. Examples of plate-like and rod-like trabecular bone structures presented by samples (about $4 \times 4 \times 4\,\mathrm{mm}^3$) of the iliac crest (a, c) and the lumbar spine $L2$ (b, d) in male (a) and female (b, c, d) subjects. (Courtesy A. Laib, ETH Zurich, Switzerland).

of the image acquisition system used and underestimation of the actual thickness may be caused by the presence of noise in the acquired images.[63]

A thickness independent structure extraction was introduced in the work of Laib *et al.*[66] to process trabecular bone images acquired *in vivo* using a special purpose low-dose 3D-QCT system. The nominal resolution of the reconstructed images was $165\,\mu$m (isotropic) whereas the actual resolution was about $240\,\mu$m. Average mineral density (D_{trab}) within the trabecular region was measured. Assuming a constant and known mineral density of fully mineralized bone ($1.2\,\mathrm{g\,HA/cm}^3$), the BV/TV_{deriv} could be derived. 3D ridges (the center points of the trabeculae) were detected in the acquired images and the mean spacing of the ridges Tb.N* was

estimated using the distance transformation. Knowing these two values, the Tb.Th and Tb.Sp parameters could be found similarly as in the directed secant method using the parallel plate model.

Other model independent 3D parameters were used by Borah *et al.*[65]: *Direct Trabecular Thickness* (Dir_Tb.Th, mm), directly measured in the three-dimensional dataset; *Marrow Star Volume* (Ma.St.V, mm^3) and *Percent Bone in Load Direction* (%BoneLD). The Marrow star volume measures the "voids" within the trabecular structure, which increase with progressive bone loss. It is derived from line lengths in random directions measured from random points in the marrow to trabecular bone. The average cubed length (L^3) is used to calculate the marrow star volume as $4\pi/3 * (L^3)$. It is a sensitive descriptor to quantify bone loss either through trabecular thinning or loss of entire trabeculae. The Percent Bone in Load Direction is a sensitive measure of architectural changes related to the bone's alignment in the load direction. The measurement of %BoneLD can provide a quantitative estimate of amount of vertical and horizontal trabeculae and may be a useful indicator of preferential loss of horizontal struts in specimens of osteoporotic vertebrae.

Another interesting approach to measure the morphology of the trabecular bone structure was proposed by Hwang *et al.*[67] In this method, the classification of voxels into bone and bone marrow (causing a significant information loss specially when the voxel size is comparable to trabeculae dimensions) is avoided and the morphological parameters are derived from the voxel bone volume fraction map. The BVF of a voxel is regarded as the probability that a single randomly chosen point in the voxel is in bone. Probability of two points in two different voxels being both in bone is equal to the product of BVF's of the two voxels. Averaging this probability through all voxel pairs with a given mutual offset n in a given direction results in the *autocorrelation function* (ACF) estimate for given n in that direction. Because the method was applied to distal radius trabecular bone specimens, which are approximately transversely isotropic, the transversal (average through all directions perpendicular to the forearm) and longitudinal ACFs were estimated. The distance to the second peak of the ACF is the average spacing of the trabeculae in a given direction. It is discussed, however, that the distance to the first minimum of ACF is more suitable to characterize the bone properties, since this quantity is estimated with higher precision due to the finite image size and to the interpolations required to compute the transverse ACF in different directions. Based on ACF estimates, parameters such as *transverse contiguity* and *tubularity* (which are not morphological measures) were introduced as well. The *transverse contiguity* is defined as the ratio of the probability that a point from a voxel and a point from one of the voxel's eight nearest neighbor voxels are both in bone divided by the probability that two points from the same voxel are both in bone. Analogously, the *tubularity* is equal to the probability that two points from adjacent voxels in the longitudinal direction are both in bone, divided by the probability that two points from the same voxel are both in bone.

Recently, Saha *et al.*[68,69] proposed a model independent method to derive the morphometric parameters of the trabecular bone structure in images acquired *in vivo* using μCT and μMRI. By analogy with the previously described method,[67] the image binarization is avoided and the 3D BVF map image is analyzed, and similarly to the direct structure thickness estimation proposed by Hildebrand and Rüegsegger,[63] the local object thickness is given as a distance from the object's medial axis to the nearest point not belonging to the object. In this case, however, the *fuzzy distance* between these two points is measured. When we consider a path from one point to another, the length of the path is defined as the integral of local BVF values along the path, which is, in other words, the amount of the bone material traversed by the path. There are infinitely many paths from one point to another, and the fuzzy distance is the infimum of their lengths. When computing the object thickness, the skeletonization of the object was performed to obtain the medial axis of the structure. Then the fuzzy distances to the object borders (voxels having BVF = 0) along this axis are computed. In this method, Saha and coworkers[68,69] introduce a resolution dependent error correction as well, considering the digital skeleton representation in a discrete grid. Analysis of the thickness resolution dependence showed that resampling the 3D μCT images from 22 μm to 176 μm (isotropic voxel size) caused only approximately 7% maximum variation of the derived thickness when the resolution dependent error correction was applied.[69] The direct structure thickness estimates from binarized images are close to the values estimated using the fuzzy distance at the 22 μm resolution. However, the method requiring the image binarization fails completely at voxel sizes greater than mean trabecular bone thickness. Another important property of the method is the relatively high noise robustness. For a SNR > 5, the noise causes approximately a 3% error of the estimated thickness, while at lower SNR values, the thickness becomes significantly underestimated.

4.2. *Anisotropy*

Another parameter often used to understand the biomechanical properties of bone and also to optimize sampling strategies for bone histomorphometry is the geometric anisotropy degree (DA) of a structure. The anisotropy corresponds to the preferential orientation(s) of trabeculae. It is constituted under the influence of strengths applied to bone and permits to establish resistance to these strengths in a given preferential direction. A structure is isotropic if it has no preferred orientation or, more rigorously, if the perpendicular to any element of surface has an equal probability of lying within any element of solid angle. The trabecular bone, from a single inspection, is obviously anisotropic and DA represents an important architectural index.

DA can be quantified by different methods. In particular, in 2D, Singh *et al.*[70] developed a semi-quantitative index based on the existence of several arches of trabeculae in the femoral neck. Whitehouse[57] expressed the difference from isotropy

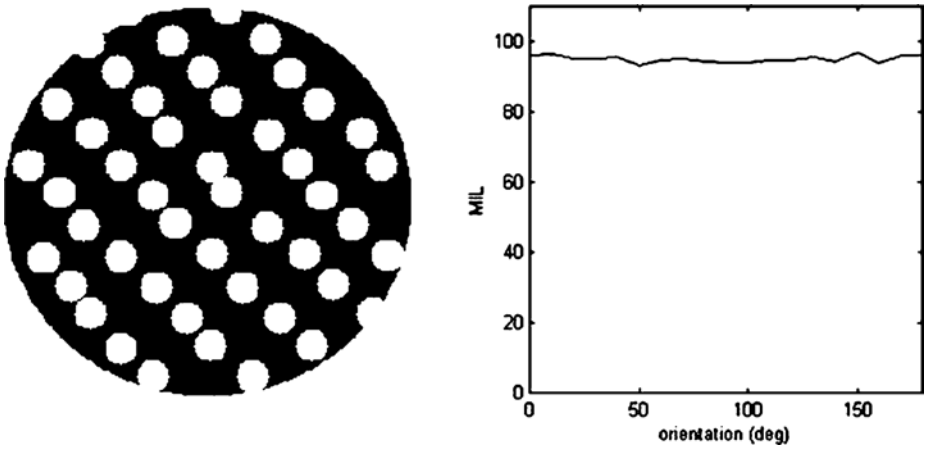


Fig. 9. Example of an anisotropic structure (left) in which the MIL result is isotropic. In this case, the MIL (measured in an arbitrary length measurement) is unable to identify the correct orientation.

as a polar diagram of the mean intercept length (MIL) between two bone/marrow interfaces and the angle of orientation. Such a diagram consists of an elliptical plot in which the ratio of the major and minor axis (eccentricity) indicates the degree of anisotropy and the angle between the axes indicates the overall orientation relative to an arbitrarily chosen reference direction. As perfect isotropy is approached, the ellipse becomes a circle. Whitehouse[57] defined the MIL both in bone (MILb) and marrow (MILm), linked by the relation: $\mathrm{MILm} = \mathrm{BVF} * 2 * \mathrm{MIL}$. Since the MIL results depend entirely on the interface between bone and marrow, surprising effects could happen, however. In fact, obviously anisotropic structures sometimes may appear isotropic when examined with the MIL method (Fig. 9).

The theory can be extended to three dimensions by using a second-rank tensor, of which the quadratic form (or fabric) describes an ellipsoid.[71] Cowin[72] developed a formulation of the dependency of mechanical anisotropy upon fabric, which was based on the assumption that fabric tensor main orientations correspond to mechanical main orientation and that the material properties of trabecular bone are isotropic. In the 3D approach, a MIL tensor can be calculated by fitting the MIL values to an ellipsoid. Since the eigenvectors of the tensor give information about the direction of the axes of the ellipsoid, and the eigenvalues express the radii of the ellipsoid, the latter can be used to define the degree of anisotropy, which denotes the ratio between the maximal and minimal radii of the MIL.

Because the MIL is unable to detect some forms of architectural anisotropy, volume-based measures were introduced with the volume orientation (VO) method.[73] The results of using the volume-based measures may also be expressed as positive definite, second-rank tensors, and consequently they are competitors for the position of providing the best fabric tensor for Cowin's structure-mechanics

relations. The volume-based methods shift the interest of architectural anisotropy from interface to volume. Three parameters, used as a description of the typical distribution of trabecular bone volume around a typical point within a trabecula, were proposed: (1) a *local volume orientation*, that is the orientation of the longest intercept through each bone point belonging to a randomly translated point grid placed on the 3D structure, (2) the *star length distribution* (*SLD*) and (3) the *star volume distribution* (*SVD*) which represent the distributions either of the intercept lengths or of their cube, respectively, through each bone point (of the previous grid) for several 3D orientations. As compared with other anisotropy measures, preliminary tests suggest that with the SLD, a more accurate description of the mechanical properties of porous structures may be obtained. However, due to possible secondary orientations that become apparent with the SLD, a fabric tensor must be of rank higher than two in order to properly describe an orthogonal structure mathematically.[74]

Another way to estimate anisotropy was developed by Gomberg *et al.*[75] which proposed the digital topology-based orientation analysis (DTA-O). In this case, the input is constituted by a 3D BVF image map of the trabecular network, from which the voxels belonging to plates, identified by means of DTA,[76] are extracted and the local normal surfaces are determined by fitting a plane through the local neighborhood BVF map. Modeling regional distributions of these vectors by means of a Gaussian model allows assessment of anisotropy measures, such as mean and variance of the orientation distribution.

4.3. *Rod-plate classification*

Some additional methods were proposed to identify the correct model or to classify trabeculae as either plate-like or rod-like.[62,77,78] An estimation of the plate-rod characteristic of the structure can be achieved using the *structure model index* (SMI), calculated by a differential analysis of a triangulated surface of a structure.[62] SMI is defined as

$$\mathrm{SMI} = 6 * (\mathrm{BV} * (d\mathrm{BS}/dr)/\mathrm{BS}^2) \tag{3}$$

where $d\mathrm{BS}/dr$ is the surface area derivative with respect to a linear measure r, corresponding to the half thickness or the radius assumed constant over the entire structure. This derivative is estimated by a simulated thickening of the structure by translating the triangulated surface by a small extent in its normal direction and dividing the associated change of surface area by the length of the extent. For a structure with both plates and rods of equal thickness, the SMI value is between 0 and 3, depending on the volume ratio between rods to plates, being largest for an element of circular cross section (i.e. a "rod") and smallest for a plate. The method has been used successfully for the characterization of structures derived from μCT images[54] but it is not suited for analysis of *in vivo* images.[5]

Borah *et al.*[77] proposed the *Percent Plate* (% Plate) parameter in order to delineate rods from plates. The measure of plates and rods is obtained by comparing

the direct (Dir_Tb.Th) and model-plate (Tb.Th) derived parameters of thickness. If the entire bone is rod-like, it should have a thickness measure equal to the thickness derived from the rod model. Likewise, if the bone is all plate-like, it would have a thickness value equal to the plate model. However, the direct trabecular thickness measure usually falls between these two extremes; this is due to the fact that some of the bone structure is plate-like and some rod-like. The %Plate can be calculated as:

$$\%\text{Plate} = (2 - (\text{Dir\_Tb.Th}/\text{Tb.Th})) * 100\,.$$

It allows a quantitative estimate of the effect of bone resorption on the shape of the trabeculae.

Stampa *et al.*[78] in their *in vivo* MRI study on finger phalanges of elderly healthy volunteers applied to each voxel a 3D plate operator and a 3D rod operator (*shape analysis*). The voxel was classified as plate-like when, in an arbitrary 2D slice of a 3D volume, more than six neighboring voxels were bone voxels. Border voxels could not be plate-like voxels. Similarly, the 3D rod operator classify a voxel as rod-like when, in an arbitrary 2D slice of a 3D volume, the voxel had two neighboring voxels aligned with itself. Voxels previously classified as plate-like could not be rod-like.

4.4. *Topological parameters*

Besides the morphological description of the trabecular bone structure another detailed analysis of the trabecular network can be achieved from inspection of its topology. In particular, the connectivity represents the main parameter measuring topology. It can be quantified by means of various methods. Already on 2D, Pugh *et al.*[79] used the *trabecular contiguity ratio* in order to describe the degree of contact between particles of granular materials (but trabeculae are not isolated particles and the application requires counting the missing trabeculae, which is even more difficult than counting the existing ones). Another suggested index is the *trabecular bone pattern factor*,[80] which is based on the measurement of the perimeter before and after a dilation operation. Some other methods based on skeletonization[81] produce results that are expressed as ratios between nodes, termini, cuts, ends and loops. Other methods for the examination of architectural features are based on node and strut analysis, which consists of measuring the density of nodes and free ends (e.g. Mellish *et al.*[82] in 2D), or on a measure of the skeleton length of the trabecular network. The *ridge number density*, proposed by Laib *et al.*[83] achieved by summing the voxels occupied by the *ridge* of trabecular elements in a gray scale image, represents an example of the latter category. This metric, applicable to *in vivo* images, was conceived as a measure of trabecular density (somewhat analogous to trabecular number). The *Trabecular Fragmentation Index* (TFI), introduced by Chevalier *et al.*,[84] is another parameter derived from 2D images able to measure the ratio between the length of the trabecular network and the number of discontinuities. Following the authors' results, a threshold of TFI = 0.195 could separate between normal and osteoporotic subjects. However, none of these ratios have any

known relation to 3D connectivity,[7] thus, from isolated two-dimensional sections, it is impossible to compute a proper measure of connectivity.

Feldkamp *et al.*[20] were among the first to recognize the inadequacy of two-dimensional approaches to characterize trabecular bone lattices, and to propose algorithms for deriving topological quantities from three-dimensional μCT images, particularly related to the connectivity. Connectivity can be interpreted as the maximum number of branches that may be removed without breaking the network into parts. In fact, connectivity measures the degree to which a structure is multiply connected, no matter the shape of the connections. Well-connected trabecular elements may be thought of as constituting closed loops. If the trabecular bone is viewed as a node and branch network, then the connectivity is the number of trabeculae minus one (Eq. (5)).

To measure connectivity, the three-dimensional Euler number, χ, was used. For an open network structure, the Euler number may be calculated from the number of nodes n and the number of branches b: $\chi = n - b$.[57] If in the network at least one path exists between any two nodes, the quantity $1 - \chi$ represents the connectivity. It is obvious that a well-connected network, such as healthy trabecular bone, has a large negative Euler number (several thousands in a volume of $1\,\mathrm{cm}^3$) that becomes less negative as connections are broken. A severely osteoporotic bone has a markedly smaller absolute value of $1 - \chi$. In fact, the connectivity has been frequently referenced as a parameter mostly affected during the progression of osteoporosis.[77,85] On a more detailed level, the breaking of a single connection will leave the trabecular network less well connected, increasing the Euler number by 1, but the addition of a connection will decrease it by 1. Being the Euler number a topological quantity, it does not carry information about positions or size of connections nor about the strength of the material comprising the connections. Hence, it cannot be used alone as an index of mechanical performance.

Feldkamp *et al.*[20] described a technique to compute the Euler number based on the inductive method of defining such number for a particular dimensionality from the definition in the next lower dimension and on a specific property. This property implies that the Euler number of the union of two sets is equal to the sum of their individual Euler numbers minus that of the set formed by their intersection. From the latter condition derives that the connectivity evaluation depends on the specimen size and the topology of its surfaces: errors tend to decrease as the examination volume increases. A solution to this problem was proposed by Odgaard and Gundersen.[86]

Starting from a discrete primitive cube constituting one of the two original solid binarized structures, the technique considers both the Euler number of a cube as unity and the cubes as connected only along cube faces. Since a set of two connected cubes also has an Euler number of unity, the Euler number of a row of cubes is equal to the number of sets of connected blocks (for further details see Feldkamp *et al.*[20]). To construct the Euler number of a two- and then of a three-dimensional array of cubes it is necessary to evaluate the Euler numbers of the constituent

rows and their intersections at first on a plane and then along the perpendicular direction. The result does not depend on the direction chosen for the initial rows and does not require attention to the appearance or disappearance of particles as the various planes are considered. The technique is very simple to apply for structures as complex as bone and it is possible to compute the Euler number of any desired subvolume in order to evaluate the uniformity of the connectivity within the specimen.

If the 3D dataset is considered as constituted by two different topological objects (a cluster of connected voxels representing the trabecular bone and another representing the marrow phase), a more general interpretation of the Euler number is possible. In fact, in a 3D structure, the Euler number may be thought[86,87] as a sum of three elements (Euler-Poincaré formula):

$$\chi = \beta 0 - \beta 1 + \beta 2 \tag{4}$$

where the β's are the first three Betti numbers, with $\beta 0$ representing the bone particles or the # of isolated parts, $\beta 1$ the connectivity or the # of redundant connections and $\beta 2$ corresponding to the marrow cavities fully surrounded by bone or the # of enclosed cavities.

Under the assumption that a complete bone does not contain isolated pieces of osseous tissue (i.e. the trabecular bone phase is one fully connected structure, $\beta 0 = 1$) and does not have isolated cavities inside osseous tissue (i.e. no cavities which do not communicate with any marrow or periosteal surface, $\beta 2 = 0$), this relationship, in general, reduces to:

$$\chi = 1 - \beta 1 = 1 - \#\text{ of trabeculae}\,. \tag{5}$$

$\beta 1$ can be evaluated in a global way by means of the Vogel algorithm.[88] In this method, the χ value (or Euler-Poincaré characteristic) in n dimensions is derived from different Euler numbers in $n-1$ dimensions. The principle is developed for the one-dimensional case and, by means of an induction mechanism, the determination in any dimension is reduced to Euler numbers of zero dimension (i.e. point counts). The method exploits the concept of the edge correction introduced by Odgaard and Gundersen[86] leading to an unbiased estimation of χ value.

Another option, which was developed in parallel to the Odgaard and Gundersen's algorithm,[86] is based on the dissector principle[87,89] consisting of a pair of closely spaced parallel sections. In this case, a property of the Euler number is utilized: in fact the Euler number of a 3D structure cannot be represented in a 2D section but the sum of the changes in the Euler number of the 2D structure seen in an imaginary plane sweeping through the 3D structure is identical to the Euler number of the structure irrespective of the direction of the sweep. This characteristic is the basis of the *ConnEulor* estimation principle which uses the difference in topology between two sections (a dissector). If complete sections through bone are evaluated, the dissector is a 3D probe which samples a volume of space created by pairs of parallel sections separated by a known distance.[90] Only three events can

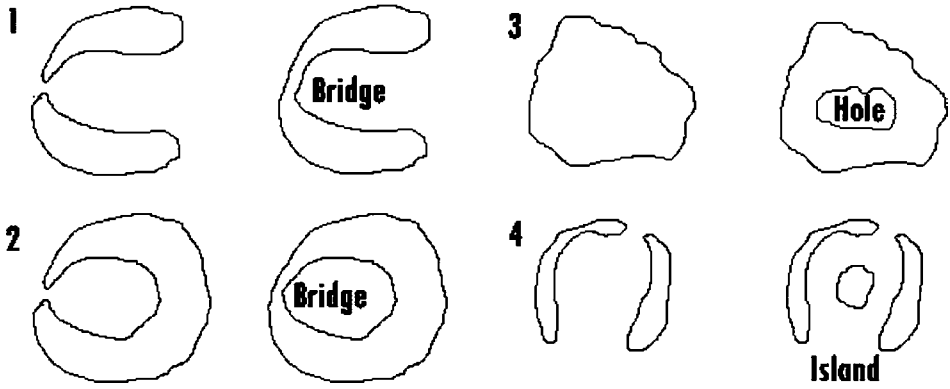


Fig. 10. Schematic representation of the possible events counted by using the ConnEulor principle: 1 left to 1 right and 2 left to 2 right: appearance of a *bridge*, i.e. a new connection between two previously unconnected bone structures; 3 left to 3 right: appearance of a *hole*; 4 left to 4 right: appearance of an *island*, i.e. a new isolated structure.

occur in terms of changes in the 2D topology of trabecular profiles: a "bridge" (B) can form (or be broken); a new "island" (I) emerges (or one disappears) and a new "hole" (H) appears (or one disappears) inside a profile (Fig. 10).

The events occur somewhere between the two sections. In a closed, complete microstructure, such as trabecular bone, connectivity is equivalent to trabecular number, because a trabecular element is defined by branching events.[91] The final estimator is:

$$\#\text{ of trabeculae/Vref} = \Sigma(B - H - I)/(2 * \text{hdis} * \text{ap} * \Sigma P) \tag{6}$$

where *hdis* is the dissector height, ΣP is the number of points hitting the reference space in one section of each dissector using a test system with an associated area per point of *ap*.

This algorithm has the advantage that edge effects can be corrected; thus, it is possible to determine the connectivity of a subset of the whole dataset. Stampa *et al.*[78] called this the *local connectivity*. For this calculation, all open ends of trabeculae touching the cortex were considered to be connected. Moreover, they defined also a 3D connectivity per unit volume calculated for the whole network, the *three-dimensional connectivity density* (GCD). For individual consecutive slices (including the cortex) of 2-mm thickness perpendicular to the long axis of the phalanx, they also determined the *local three-dimensional connectivity density* (LCD) in order to assess the spatial inhomogeneity of connectivity.

A measure of mean trabecular size as well as of the mean spacing between trabeculae can be obtained from connectivity. The first, defined as *mean trabecular volume*,[7] MTV, is given by:

$$\text{MTV} = (V * \text{TB/TV})/(\beta 1 + 1)$$

being V the volume of the specimen and $\beta 1$ the connectivity. The second, called *mean marrow space volume*,[7] MMSV may be calculated as:

$$\mathrm{MMSV} = V * (1 - \mathrm{TB/TV})/(\beta 1 + 1)\,.$$

The measure of connectivity by means of the Euler number presents some limitations.[7,76] In fact, fracture healing, bone formation and sometimes osteoporosis as well as the noise present during the segmentation process could produce isolated "bone" and "marrow" particles violating the topological assumption expressed in Eq. (5). However, it is possible to consider the complete relation (Eq. (4)) determining separately $\beta 0$ and $\beta 1$ or to remove the noise with a suitable filter.[7] Nevertheless, when the expression in Eq. (4) must be used, many different 3D architectural structures could have the same Euler value but not necessarily to be topologically equivalent. Moreover, the first Betti number ($\beta 1$) is inherently insensitive to trabecular erosion, which is known to result in the perforation of trabecular bone plates and disconnection of rod-like trabecular bone.[85,92] Therefore, the first Betti number will decrease from loss of rods, causing a reduction in the number of loops. However, it would increase as a result of the perforation of plates, which increases the number of loops. Therefore, the connectivity alone would not necessarily detect osteoporotic bone erosion.[76] Stampa *et al.*[78] suggest that together with the connectivity analysis, a shape analysis should also be used in order to classify trabecular bone as either plate-like or rod-like.

Gomberg *et al.*[93] propose a method, called digital topological analysis (DTA) that allows the unequivocal determination of the topological class pertaining to each bone voxel of the 3D trabecular network. The method is based on determining the connectivity of the neighbors in the $3 \times 3 \times 3$ neighborhood[94] of the voxel examined. The starting point of the analysis is the conversion of the 3D network to a skeletonized surface representation, which contains 1D and 2D structures only (i.e. curves and surfaces). The voxels are then assigned to the appropriate topological class (e.g. curves = rod-like structures, surfaces = plate-like structures, or junctions) in a three-step approach. The method was validated on synthetic images demonstrating its relative immunity to partial volume blurring and noise. *Surface-to-curve ratio index* is the new introduced parameter that characterizes the network topology and in the case of 3D MR microimages from distal radius has proved to be a strong predictor of Young's modulus.[93]

4.5. *Fractal dimension*

Starting from 2D sections, both from radiographs and single slices of a 3D acquisition, a different approach can be used to estimate the three-dimensional anisotropy of the trabecular bone network[95] and to achieve information pertaining to the porosity and connectivity of trabeculae. This approach exploits the fractal characteristics of trabecular bone which appear to be correlated with its biomechanical properties. In particular, the fractal analysis can be used to describe complex shapes,[96] including the overall roughness texture and the form of individual trabecular profiles.

Several methods for estimating the fractal dimension (FD) are known and computational ways to obtain it can vary considerably. Box-counting (to obtain the Kolmogorov dimension), dilatation method for the Minkowski-Bouligand dimension, sandbox for the mass-radius dimension and Hurst orientation transform[97] are the most used algorithms.[95,98–102] However, Majundar *et al.*[103] in their study on fifty-one human femur specimens used other three different fractal geometry-based techniques, namely semi-variance, surface area and Fourier transform evaluated along the three orthogonal projections.

In the box-counting method, a grid consisting of a box size ε is superimposed on the boundary of the trabecular bone network to be quantified. The number of boxes of a given size ε that contain the boundary point $N(\varepsilon)$ is computed. This procedure is repeated for different box dimensions. The FD is the absolute value of the slope in the linear portion of the plot of the log number of boxes versus the log box size ε. Podsiadlo and Stachowiak[97] have shown that the Hurst orientation transform method presents some advantages over the other techniques because it calculates a two-dimensional FD in all possible directions also providing a measure of anisotropy. Jiang *et al.*[104] have found that the Minkowski dimension is a better measure for characterizing the trabecular bone anisotropy in the X-ray images of thick specimens.

Fazzalari and Parkinson[98] in their investigation on the fractal properties of trabecular bone in severe osteoarthritis of the hip identified three straight line segments on the log-log plot, indicating a FD over three different ranges of scale. The pivot points, i.e. the box size at which the FD changes, were of similar magnitude to the trabecular thickness and trabecular separation. Subsequently, in a study on iliac crest biopsies from sixty-four postmenopausal women, the same authors[99] measured three FDs, which describe the trabecular surface texture (fractal-1), trabecular shape (fractal-2) and trabecular arrangement (fractal-3), indicating that the trabecular bone has a sectional self-similarity. The results showed that fractal-2 was significantly lower in the vertebral crush fracture group than in the nonfracture group. Recently, Parkinson and Fazzalari[105] found that trabecular bone is effectively fractal over a defined range of scale ($25\,\mu$m–$4250\,\mu$m) and that within this range there is more than one FD describing spatial structural entities. Since magnification, image orientation and threshold settings had little effect on this procedure of FD estimation, the authors concluded that the FD is a suitable model-independent method for describing the complex multifaceted structure of trabecular bone.

Majumdar *et al.*[103] reported that the fractal-based texture analysis of radiographs is technique dependent and the fractal measures correlate better with trabecular spacing and trabecular number than with the trabecular thickness. Moreover, the FD in addition to BMD improves the prediction of bone strength and elastic modulus.[103,106]

Pothuaud *et al.*[56] examined how FD of two-dimensional trabecular bone projection images could be related to the three-dimensional trabecular properties such as porosity (*phi*) or connectivity per unit volume (Cv). Significant relationships

were found between FD, *phi* and *Cv* and the authors concluded that the fractal evaluation of the trabecular bone projection has real meaning in terms of porosity and connectivity of the 3D architecture.

4.6. *Mechanical parameters — finite element modeling and analysis*

To date, invasive and destructive methods have been used to determine different mechanical properties such as bone strength and stiffness. To study the influence of density and structural anisotropy on mechanical properties and to predict bone function under various conditions, different microstructural models of trabecular bone have been introduced (see Müller[17] for a review) and computational tools, such as microstructural Finite Element (μFE) modeling and analysis have been used. Finite Element Modeling (FEM) is the division of a structure or object into discrete elementary volumetric units that have precise mathematical equations to describe their mechanical behavior. Structural Finite Element Analysis (FEA) is the calculation of the mechanical behavior (stress and strain) at any point within the structure under specific loading conditions thus simulating conventional destructive mechanical testing. The starting point of every FEM is the three-dimensional data of the object or structure. For the FE solver, the bone voxels in the data sets have to be converted into brick elements with a side length generally depending on the taking site. For example, Ulrich *et al.*[107] used a length of 28 μm for the lumbar spine, iliac crest and calcaneus samples and 56 μm for the femoral head samples; Borah *et al.*[77] converted each voxel into a structural hexahedral brick element with a resolution of 85 μm, while Newitt *et al.*[108] used equally shaped 8-node brick elements with a mass-compensated meshing technique. Even if the accuracy of the FE analysis is related to the element size[109] convergence for models with an element size of $\leq 56\,\mu$m was demonstrated.[110]

Van Lenthe *et al.*[111] assigned the same mechanical properties to each element, with an arbitrary chosen tissue modulus of 1 GPa and a Poisson's ratio of 0.3. It has been found that the tissue anisotropy has a negligible effect on the apparent Young's moduli,[112–114] and because the analyses are linear elastic, the results can be scaled to any isotropic tissue modulus with the same Poisson's ratio. Borah *et al.*[77] estimated the tissue modulus of bone to be 5,000 MPa and assigned to the Poisson's ratio the 0.3 value. To account for the arbitrary value chosen for tissue modulus the results were corrected by using suitable factors based on BV fraction and the measured applied strain. For Newitt *et al.*,[108] the tissue element properties were chosen to be linear, elastic and isotropic with a Young's modulus of 10 GPa and a Poisson's ratio of 0.3.

Using μFE technique, huge structural models with more than a million FE elements can easily result. An iterative FEA solver program can efficiently and rapidly analyze FE models of this magnitude by using some hypothesis, e.g. by assuming that each individual element has the same mechanical properties. With

this approach, all nine orthotropic elastic constants of a bone specimen can be predicted and experimental artifacts can be eliminated. Stiffness constants such as Young's moduli ($E1$, $E2$, $E3$), Poisson's ratios ($v12$, $v13$, $v23$) and shear moduli ($G12$, $G23$, $G13$) can be calculated[115] and anisotropy ratios $E1/E3$, $E2/E3$ and $E1/E2$ can be derived.

FEA has been used to study bone behavior on a microscopic and macroscopic level calculating the apparent stiffness of the architecture. By using different strain and shear deformations, it is possible to simulate mechanical tests with various types of loading conditions and to calculate the strains within the trabeculae.[115–117] In this way, it is possible to look inside the bone to see where stresses are localized and may cause fracture.

In FEA studies, the computational technique must first be validated for the specific species. In such a validation study, both experimental and computational analyses must be completed to compare the predicted and actual mechanical properties. Borah *et al.*,[65] in a μMRI study showed that the computational and experimental values of apparent modulus in the lumbar vertebra of young and mature minipigs were statistically similar. Furthermore, Ulrich *et al.*[107] found a good agreement between the results of compression tests and those of FE analysis of 3D μCT images of human trabecular bone samples.

Correlation between the obtained Young's modulus and some structural parameters has been examined: Kabel *et al.*,[118] examining 141 specimens taken at various skeletal sites, found that BV/TV was closely correlated to Young's modulus ($R^2 = 0.84$–0.94) while Ulrich *et al.*[110] found that the correlation, at the calcaneus, was lower ($R^2 = 0.52$–0.67). The latter authors also reported a R^2 of 0.59 with Tb.Sp and of 0.82 with anisotropic index concluding that the BV/TV alone does not predict enough of the Young's modulus, while other structural parameters like MIL or Tb.Sp should be used. The prediction and the best parameters changed with the sample site.

5. Conclusions

Considered the importance of osteoporosis as a social problem, an early diagnosis of this pathological status as well as the monitoring of bone quality evolution, also after pharmacological treatment, is highly desirable. The trabecular bone, as a consequence of its highest metabolic activity, is the principal tissue involved in this disease and then it should be primarily investigated and monitored. The trabecular bone of various skeletal sites has been examined and its structure has been shown to vary from plate-like to rod-like according to the location and the disease state. Because of the limits of the available imaging systems, the proximal femur and vertebrae, which are the main fracture sites, cannot be routinely investigated. Thus, other sites such as iliac crest, distal radius and calcaneus have been proposed as surrogate sites for a more accurate examination of the trabecular bone structure.

Although BMD is the main determinant of mechanical resistance, several studies have indicated also a relevant role of bone architecture. Many parameters describing the characteristics of the bone architecture have been calculated from 2D and 3D images obtained by using different techniques: histomorphometry, radiography, CT, μCT, MRI and μMRI are those described in this chapter. Each of these techniques present advantages and disadvantages particularly when they are used in the clinical practice for the evaluation of bone status in osteoporotic and non-osteoporotic patients. Non-invasiveness, high-resolution, good reproducibility are the main requirements of an ideal technique for *in vivo* studies. While pQCT seems to provide the best resolution in the X-ray field, high-resolution MRI appears to be the most promising modality, particularly for longitudinal studies since it does not require exposure of the patient to radiation.

Sophisticated analyses of 2D and 3D images, obtained with the previously mentioned techniques, are required in order to accurately estimate the fracture risk. In this chapter, structural indices derived from trabecular bone images using different approaches (morphological, topological, fractal, etc.) have been described. Several studies have shown that a strong correlation between mechanical properties and some structural parameters exists. The combined use of BMD and some structural parameters better explains the variance in mechanical properties with respect to BMD alone. BVF, MIL, trabecular thickness and Euler number seem the most useful ones. However, the most suitable parameters for a better description of the 3D architecture still need to be defined. A higher image resolution and a possible parameters selection, validation and standardization are also required.

Acknowledgment

This work was supported by grants from the Italian Ministry of Education and Research (FIRB 2001).

References

1. J. J. Pritchard, in *The Biochemistry and Physiology of Bone*, ed. G. H. Bourne **I** (Academic Press, New York, 1972), p. 1.
2. WHO: *World Health Organ. Tech. Rep.* **Ser 843** (1994).
3. K. Engelke, S. M. Song, C. C. Glüer and H. K. Genant, *J. Bone Miner. Res.* **11** (1996) 480.
4. P. Amman and R. Rizzoli, *Osteoporos. Int.* **14** (2003) S13.
5. F. W. Wehrli, P. K. Saha, B. R. Gomberg, H. K. Song, P. J. Snyder, M. Benito, A. Wright and R. Weening, *Topics Magn. Reson. Imaging* **13** (2002) 335.
6. E. R. Weibel, in *Stereological Methods*, Theoretical Foundations **2** (Academic Press, London, 1980).
7. A. Odgaard, *Bone* **20** (1997) 315.
8. A. Odgaard, E. B. Jensen and H. J. G. Gundersen, *J. Microsc.* **159** (1990) 149.
9. A. Odgaard, K. Andersen, R. Ullerup, L. H. Frich and F. Melson, *Bone* **15** (1994) 335.

10. J. D. Beck, B. L. Canfield, S. M. Haddock, T. J. Chen, M. Kothari and T. M. Keaveny, *Bone* **21** (1997) 281.
11. A. A. Postnov, A. V. Vinogradov, D. van Dyck, S. V. Saveliev and N. M. Clerck, *Physiol. Meas.* **24** (2003) 165.
12. T. M. Link, S. Majumdar, S. Grampp, G. Guglielmi, C. van Kuijk, H. Imhof, C. Gluer and J. E. Adams, *Eur. Radiol.* **9** (1999) 17810.
13. G. Luo, J. H. Kinney, J. J. Kaufman, D. Haupt, A. Chiabrera and R. S. Siffert, *Osteoporos. Int.* **9** (1999) 339.
14. R. Müller, P. Rüegsegger and P. Seitz, *Phys. Med. Biol.* **30** (1985) 401.
15. P. Rüegsegger, *Ann. Chir. Gynecol.* **77** (1988) 204.
16. R. Müller, T. Hildebrand and P. Rüegsegger, *Phys. Med. Biol.* **39** (1994) 145.
17. R. Müller, *Topics Magn. Reson. Imaging* **13** (2002) 307.
18. A. Laib, S. Hämmerle and B. Koller, *Proceedings of the 16th International Bone Densitometry Workshop* (Annecy, France, June 20–24, 2004).
19. M. J. Paulus, S. S. Gleason, S. J. Kennel, P. R. Hunsicker and D. K. Johnson, *Neoplasia* **2** (2000) 62.
20. L. A. Feldkamp, S. A. Goldstein, A. M. Parfitt, G. Jesion and M. Kleerekoper, *J. Bone Mineral. Res.* **4** (1989) 3.
21. M. Rossi, F. Casali, M. Bettuzzi, M. P. Morigi, D. Romani, S. Golovkin and V. Govorun, *Proceedings of the SPIE's 46th Annual Meeting* (2001).
22. M. Rossi, F. Casali, D. Romani and M. L. Carabini, *Proceedings of the SPIE's 46th Annual Meeting* (2001).
23. H. Guan and Y. Zhu, *Phys. Med. Biol.* **43** (1998) 2925.
24. M. F. Kijewski and P. F. Judy, *Phys. Med. Biol.* **32** (1987) 565.
25. P. Hammersberg and Månгård, *J. X-Ray Sci. Technol.* **8** (1998) 75.
26. E. van de Casteele, D. van Dyck, J. Sijbers and E. Raman, *Phys. Med. Biol.* **47** (2002) 4181.
27. M. Dasltra, F. Beckmann, F. Melsen, E. M. Hauge, S. V. Hoffmann, S. Lunt and E. Uggerhøj, *HASYLAB Annual Reports* (1999). http://www-hasylab.desy.de/science/annual_reports/1999_report/index.html.
28. M. A. Fernandez-Seara, H. K. Song and F. W. Wehrli, *Magn. Reson. Med.* **46** (2001) 103.
29. M. A. Fernandez-Seara, S. L. Wehrli and F. W. Wehrli, *Biophys. J.* **82** (2002) 522.
30. J. Ch. Ford, F. W. Wehrli and H. W. Chung, *Magn. Reson. Med.* **30** (1993) 373.
31. L. Hilaire, F. W. Wehrli and H. K. Song, *Magn. Reson. Imaging* **18** (2000) 777.
32. H. W. Chung, F. W. Wehrli, J. L. Williams and S. D. Kugelmass, *Proc. Natl. Acad. Sci. USA* **90** (1993) 10250.
33. J. Ma, F. W. Wehrli and H. K. Song, *Magn. Reson. Med.* **35** (1996) 903.
34. S. Majumdar, D. Newitt, M. Jergas, A. Gies, E. Chiu, D. Osman, J. Keltner, J. Keyak and H. Genant, *Bone* **17** (1995) 417.
35. P. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy* (Oxford University Press, New York, 1991).
36. J. Sijbers, A. J. den Dekker, J. van Audekerke, M. Verhoye and D. van Dyck, *Magn. Reson. Imaging* **16** (1998) 87.
37. H. K. Song and F. W. Wehrli, *Magn. Reson. Med.* **41** (1999) 947.
38. S. N. Hwang and F. W. Wehrli, *Magn. Reson. Med.* **47** (2002) 948.
39. A. Accardo, G. Candido, V. Jellus, R. Toffanin and F. Vittur, *IEEE Trans. Biomed. Eng.* **50** (2003) 967.
40. K. P. Pruessmann, M. Weiger, M. B. Scheidegger and P. Boesiger, *Magn. Reson. Med.* **42** (1999) 952.

41. J. A. Hipp, A. Jansujwicz, C. A. Simmons and B. Snyder, *J. Bone Miner. Res.* **11** (1996) 286.
42. N. Boutry, B. Cortet, P. Dubois, X. Marchandise and A. Cotten, *Radiology* **227** (2003) 708.
43. X. Ouyang, K. Selby, P. Lang, K. Engelke, C. Klifa, B. Fan, F. Zucconi, G. Hottya, M. Chen, S. Majumdar and H. K. Gennant, *Calcif. Tissue Int.* **60** (1997) 139.
44. S. Majumdar, H. K. Genant, S. Grampp, D. C. Newitt, V. H. Truong, J. C. Lin and A. Mathur, *J. Bone Miner. Res.* **12** (1997) 111.
45. D. L. Pham and J. L. Prince, *IEEE Trans. Med. Imaging* **18** (1999) 737.
46. S. N. Hwang and F. W. Wehrli, *Int. J. Imaging Syst. Technol.* **10** (1999) 186.
47. A. Elmoutaouakkil, F. Peyrin, J. Elkafiand and A. M. Laval-Jeantet, *IEEE Trans. Med. Imag.* **21** (2002) 354.
48. Z. Wu, H. W. Chung and F. W. Wehrli, *Magn. Reson. Med.* **31** (1994) 302.
49. S. Li, *Computer Vision* (Springler-Verlag, New York, 1995).
50. J. Wolff, *Das Gesetz der Transformation der Knochen* (Hirschwald, Berlin, 1892).
51. I. Hvid, N. C. Jensen, C. Bünger, K. Solund and J. C. Djurhuus, *Eng. Med.* **14** (1985) 79.
52. J. C. Rice, S. C. Cowin and J. A. Bowman, *J. Biomech.* **21** (1988) 155.
53. R. Muller, S. C. Gerber and W. C. Hayes, *Technol. Health Care* **6** (1998) 433.
54. T. Hildebrand, A. Laib, R. Müller, J. Dequeker and P. Rüegsegger, *J. Bone Miner. Res.* **14** (1999) 1167.
55. E. Martin-Badosa, D. Amblard, S. Nuzzo, A. Elmoutaouakkil, L. Vico and F. Peyrin, *Radiology* **229** (2003) 921.
56. L. Pothuaud, C. L. Benhamou, P. Porion, E. Lespessailles, R. Harba and P. Levitz, *J. Bone Miner. Res.* **15** (2000) 691.
57. W. J. Whitehouse, *J. Microsc.* **101** (1974) 153.
58. A. M. Parfitt, M. K. Drezner, F. H. Glorieux, J. A. Kanis, H. Malluche, P. J. Meunier, S. M. Ott and R. R. Recker, *J. Bone Miner. Res.* **2** (1987) 595.
59. W. E. Lorensen and H. E. Cline, *Comput. Graphics* **21** (1987) 163.
60. F. Guilak, *J. Microsc.* **173** (1994) 245.
61. M. Kothari, T. M. Keaveny, J. C. Lin, D. C. Newitt, H. K. Genant and S. Majumdar, *Bone* **22** (1998) 437.
62. T. Hildebrand and P. Rüegsegger, *Comput. Meth. Biomech. Biomed. Eng.* **1** (1997) 5.
63. T. Hildebrand and P. Rüegsegger, *J. Microsc.* **185** (1997) 67.
64. J. A. Goldak, Y. Xinhua, A. Knight and D. Lingxian, *J. Comput. Geom. Appl.* **1** (1991) 327.
65. B. Borah, T. E. Dufresne, M. D. Cockman, G. J. Gross, E. W. Sod, W. R. Myers, K. S. Combs, R. E. Higgins, S. A. Pierce and M. L. Stevens, *J. Bone Miner. Res.* **15** (2000) 1786.
66. A. Laib, H. J. Hauselmann and P. Ruegsegger, *Technol. Health Care* **6** (1998) 329.
67. S. N. Hwang, F. W. Wehrli and J. L. Williams, *Med. Phys.* **24** (1997) 1255.
68. P. K. Saha, F. W. Wehrli and B. R. Gomberg, *Comput. Vision Image Understanding* **86** (2002) 171.
69. P. K. Saha and F. W. Wehrli, *IEEE Trans. Med. Imaging* **23** (2004) 53.
70. M. Sing, A. R. Nagrath and P. S. Maini, *J. Bone Joint Surg. Am.* **52** (1970) 457.
71. T. P. Harrigan and R. W. Mann, *J. Mater. Sci.* **19** (1984) 761.
72. S. C. Cowin, *Mech. Mater.* **5** (1986) 251.
73. A. Odgaard, K. Andersen, F. Melsen and H. J. G. Gundersen, *J. Microsc.* **159** (1990) 335.
74. T. H. Smit, E. Schneider and A. Odgaard, *J. Microsc.* **191** (1998) 249.

75. B. R. Gomberg, P. K. Saha and F. W. Wehrli, *Med. Phys.* **30** (2003) 158.
76. P. K. Saha, B. R. Gomberg and F. W. Wehrli, *Int. J. Imaging Syst. Technol.* **11** (2000) 81.
77. B. Borah, G. J. Gross, T. E. Dufresne, T. S. Smith, M. D. Cockman, P. A. Chmielewski, M. W. Lundy, J. R. Hartke and E. W. Sod, *Anat. Rec. (New Anat)* **265** (2002) 101.
78. B. Stampa, B. Kühn, C. Liess, M. Heller and C. C. Glüer, *Topics Magn. Reson. Imaging* **13** (2002) 357.
79. J. W. Pugh, E. L. Radin and R. M. Rose, *J. Bone Joint Surg.* **56-A** (1974) 313.
80. M. Hahn, M. Vogel, M. Pompesius-Kempa and G. Delling, *Bone* **13** (1992) 327.
81. J. E. Compston, R. W. E. Mellish, P. Croucher, R. Newcombe and N. J. Garrahan, *Bone Miner.* **6** (1989) 339.
82. R. W. Mellish, M. W. Ferguson-Pell, G. V. Cochran, R. Lindsay and D. W. Dempster, *J. Bone Miner. Res.* **6** (1991) 689.
83. A. Laib, T. Hildebrand and H. J. Hauselmann, *Bone* **21** (1997) 541.
84. F. Chevalier, A. M. Laval-Jeantet, M. Laval-Jeantet and C. Bergot, *Calcif. Tissue Int.* **51** (1992) 8.
85. A. M. Parfitt, *Bone* **13** S41 (1992).
86. A. Odgaard and H. J. G. Gundersen, *Bone* **14** (1993) 173.
87. H. J. G. Gundersen, R. W. Boyce, J. R. Nyengaard and A. Odgaard, *Bone* **14** (1993) 217.
88. H. J. Vogel, *Acta Stereol.* **16** (1997) 97.
89. R. W. Boyce, T. J. Wronski, D. C. Ebert, M. L. Stevens, C. L. Paddock, T. A. Youngs and H. J. G. Gundersen, *Bone* **16** (1995) 209.
90. D. C. Sterio, *J. Microsc.* **134** (1984) 127.
91. J. R. Nyengaard and N. Marcussen, *J. Microsc.* **171** (1993) 27.
92. M. Amling, M. Posl, H. Ritzel, M. Hahn, M. Vogel, V. J. Wening and G. Delling, *Arch. Orthop. Trauma Surg.* **115** (1996) 262.
93. B. R. Gomberg, P. K. Saha and H. K. Song, *IEEE Trans. Med. Imaging* **19** (2000) 166.
94. P. K. Saha and B. B. Chaudhury, *Comput. Vision Image Understanding* **63** (1996) 418.
95. D. Chappard, E. Legrand, B. Haettich, G. Chales, B. Auvinet, J. P. Eschard, J. P. Hamelin, M. F. Basle and M. Audran, *J. Pathol.* **195** (2001) 515.
96. B. B. Maldebrot, *The Fractal Geometry of Nature* (Freeman, New York, 1983).
97. P. Podsiadlo and G. W. Stachowiak, *Med. Phys.* **29** (2002) 460.
98. N. L. Fazzalari and I. H. Parkinson, *J. Bone Miner. Res.* **12** (1997) 632.
99. N. L. Fazzalari and I. H. Parkinson, *Bone* **23** (1998) 53.
100. S. S. Cross, S. Roger, P. B. Silcocks and D. W. Cotton, *J. Pathol.* **170** (1993) 11.
101. R. S. Weinstein and S. Majumdar, *J. Bone Miner. Res.* **9** (1994) 1797.
102. H. W. Chung, C. C. Chu, M. Underweiser and F. W. Wehrli, *Med. Phys.* **21** (1994) 1535.
103. S. Majumdar, J. Lin, T. Link, J. Millard, P. Augat, X. Ouyang, D. Newitt, R. Gould, M. Kothari and H. Genant, *Med. Phys.* **26** (1999) 1330.
104. C. Jiang, R. E. Pitt, J. E. Betram and D. J. Aneshansley, *Med. Biol. Eng. Comput.* **37** (1999) 413.
105. I. H. Parkinson and N. L. Fazzalari, *J. Microsc.* **198** (2000) 134.
106. J. C. Lin, S. Grampp, T. Link, M. Kothari, D. C. Newitt, D. Felsemberg and S. Majumdar, *Osteoporos. Int.* **9** (1999) 516.

107. D. Ulrich, B. van Rietbergen, H. Weinans and P. Ruegsegger, *J. Biomech.* **31** (1998) 1187.
108. D. C. Newitt, S. Majumdar, B. van Rietbergen, G. von Ingersleben, S. T. Harris, H. K. Genant, C. Chesnut, P. Garnero and B. MacDonald, *Osteoporosis Int.* **13** (2002) 6.
109. A. J. C. Ladd and J. H. Kinney, *J. Biomech.* **31** (1998) 941.
110. D. Ulrich, B. van Rietbergen, A. Laib and P. Rüegsegger, *Bone* **25** (1999) 55.
111. G. H. van Lenthe, J. P. W. van den Bergh, A. R. M. M. Hermus and R. Huiskes, *J. Bone Min. Res.* **16** (2001) 550.
112. J. Kabel, B. van Rietbergen, M. Dalstra, A. Odgaard and R. Huiskes, *J. Biomech.* **32** (1999) 673.
113. A. J. C. Ladd, J. H. Kinney, D. L. Haupt and S. A. Goldstein, *J. Orthop. Res.* **16** (1998) 622.
114. B. van Rietbergen, J. Kabel, A. Odgaard and R. Huiskes, in *Material Identification Using Mixed Numerical Experimental Methods*, eds. H. Sol and C. W. J. Oomens (Kluwer, Amsterdam, 1997), p. 183.
115. B. van Rietbergen, A. Odgaard, J. Kabel and R. Huiskes, *J. Biomech.* **29** (1996) 1635.
116. S. J. Hollister and N. Kikuchi, *Comput. Mech.* **10** (1992) 73.
117. P. M. Suquet, in *Homogenization Techniques for Composite Media*, eds. E. Sanchez-Palencia and A. Zaoui (Springer, Berlin, 1987).
118. J. Kabel, A. Odgaard, B. van Rietbergen and R. Huiskes, *Bone* **24** (1999) 115.

This page intentionally left blank

CHAPTER 2

MEDICAL IMAGE-BASED PREFORMED TITANIUM MEMBRANES FOR BONE RECONSTRUCTION

V. PATTIJN, F. GELAUDE*, J. VANDER SLOTEN† and R. VAN AUDEKERCKE
Division of Biomechanics and Engineering Design, Katholieke Universiteit Leuven
Celestijnenlaan 300C, B-3001 Heverlee, Belgium
Tel.: +32(0)16-32-70-96; Fax: +32(0)16-32-79-94
†*jos.vandersloten@mech.kuleuven.ac.be*

The design of a personalized implant with enhanced functionality for a specific clinical problem, i.e. the reconstruction of weight-bearing bones after tumor resection surgery, is described. First, the shape of the new implant was determined using CT-scan data and strength calculations combined with material considerations. It was shown that a thin preformed titanium shell fitted on the periosteal surface of the bone and fixed with small screws provided sufficient strength for bone reconstruction in weight-bearing applications. Experimental tests were performed to determine the membrane thickness, the surface macro-structure, and the number and diameter of screws required to fix the implant to the bone, taking into account the loads acting on the bone during normal daily activities. Then, the bone-implant system was biomechanically evaluated by finite element analysis. Finally, a personalized titanium membrane was applied for the reconstruction of a distal femur after en-bloc resection of a juxta-cortical chondroma. This clinical application shows the feasibility of this new implant as well as its advantages towards the classical reconstruction methods.

Future research will focus on the expansion of titanium membranes as reconstruction devices for bone to other applications in the field of bone reconstructive surgery, e.g. large bone defects (scaffolding), acetabular reconstructions or re-enforcement of osteoporotic bone, resulting in a more automated image based design and production process.

Keywords: Titanium membrane; tumor surgery; bone reconstruction; rapid prototyping technology; custom-made implant.

1. Introduction

Large or locally aggressive benign bone tumors frequently occur in the extremities just adjacent to the joint surface. The surgical treatment consists of two stages: (1) removal of the tumor and (2) reconstruction of the defect.[1–3]

One of the main methods for removing bone tumors is intralesional excision by curettage.[4,5] The curettage starts by making a bone window as large as the lesion. After removing most of the tumor with a spoon and a large curette, a mechanical power burring of the entire cavity is necessary. Then lavage by fluids cleans away debris and allows better visualization. Further control of the tumor is sometimes

*Correspondence to: F. Gelaude; frederik.gelaude@mech.kuleuven.ac.be.

provided by the use of adjuvant agents to extend the tumor kill.[6] Especially the removal but also the control of the tumor must be performed very precisely to decrease the recurrence rate. Two other methods for bone tumor removal are resection and amputation.[7–11] These techniques remove, besides the tumor zone, also a part of the intact bone. The non-limb-saving technique, e.g. amputation, is seldomly used nowadays.

The reconstruction of the defect depends on the technique used for removing the tumor. The more drastic the removal, the more difficult the reconstruction will be. Reconstruction materials as bone grafts or bone graft substitutes are suitable for small defects in applications with a low loading. When a higher load must be withstood, then bone cement must be applied or one of the previous reconstruction materials with the addition of standard osteosynthesis devices. Finally, in the case of very radical resection or amputation prostheses are needed to reconstruct the bone and/or the adjacent joint. The choice of the appropriate treatment of each specific patient is a very complex task for the surgeon. Therefore, an evaluation and staging of the bone tumor is advisable. The Enneking staging system[12,13] is mostly preferred by orthopaedic oncologists and leads to a classification of both benign and malignant bone tumors in three stages. Each stage has a preferred treatment and helps the surgeon in choosing the most appropriate treatment for each specific clinical case. However, the surgeon has a wide variety of reconstruction techniques to his disposal. There are still some limits with respect to restoring the structural integrity and the function of the affected bone. Therefore, a new improved reconstruction method is sought that meets the flaws of the currently available techniques.

The treatment of bone tumors should try to obtain an ideal reconstruction of the affected bone, having biological affinity, resistance to infection, and sufficient strength and durability in order to achieve a long-lasting survival and function of the affected limb after reconstruction.[14] Therefore a novel implant, in particular a preformed titanium membrane, is created (Fig. 1). A membrane is a thin plate that fits on the periosteal surface of the bone and that is fixed with some small screws. This new implant is tailored to the individual bone geometry and material properties of the patient, and has enhanced functionality when compared with currently existing methods for reconstructing a tumor cavity. The use of membranes originates from reconstruction and mandibular augmentation in dental surgery.[15–17] Membranes are also applied in cranio-maxillofacial bone reconstructive surgery.[18,19]

This study will focus on the bone reconstruction after removing a giant cell tumor.[20] The technique proposed here is probably applicable for bone reconstructions of other tumoral–benign as well as malignant–defects. However, each specific application requires preliminary research.

2. Design Process

The design process of a personalized membrane consists of three consecutive steps: data acquisition, pre-operative planning and reconstruction.

Fig. 1. Titanium membrane.

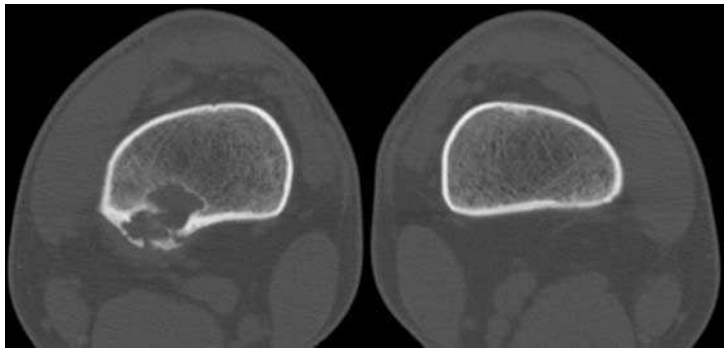


Fig. 2. CT image of affected bone and contra-lateral bone.

2.1. *Data acquisition*

A computed tomography (CT) scan of the affected region of the bone of the patient is taken, with a slice thickness of about 2 mm. The scanned region contains the tumor and a margin of approximately 20 mm above and below the tumor. Simultaneously the same region of the contra-lateral bone is scanned (Fig. 2). The CT data is imported in the software program Mimics (Materialise NV[a]). Within this program, a segmentation of the bone is performed by using a threshold value for bone and a three-dimensional model of the bone is created (Fig. 3).

2.2. *Pre-operative planning*

The surgeon indicates the approximate size of the cortical window, which will be made during surgery to remove the tumor (Fig. 4). He also determines the maximal

[a]Materialise NV, Technologielaan 15, 3001 Leuven, Belgium.

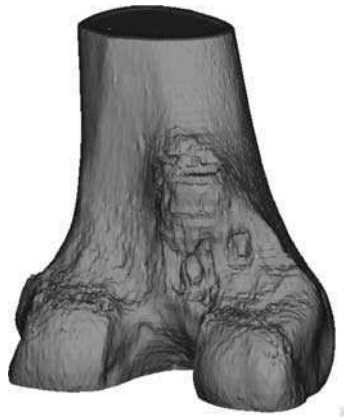


Fig. 3. 3D model of distal femur with bone tumor.

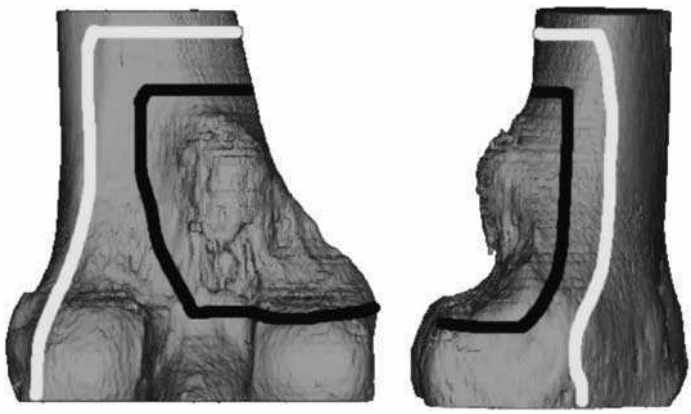


Fig. 4. Pre-operative planning: The black line indicates the approximate size of the cortical window, the white line the maximal outline of the membrane.

outline of the titanium membrane. This is the maximal size he will need to reconstruct the cortex taking into account (1) the possible per-operative decision of enlarging the cortical window and (2) a sufficient overlap between the membrane and the cortex. The maximal size of the membrane is limited by the attachment of soft tissues, e.g. ligaments and muscles.

2.3. *Reconstruction*

The reconstruction of the cortical surface in the region affected by the tumor is a very important step in the design process. It determines the shape of the membrane that must fit on the bone surrounding the tumor. Three different computer methods to restore the original shape of the cortex in the tumor zone were

explored: (1) mirror the contra-lateral bone, (2) manual editing in Mimics and (3) CAD-based technique.

2.3.1. *Mirror the contra-lateral bone*

The idea of mirroring the intact contra-lateral bone to obtain a good cortical shape for the affected bone, is based on the assumption that the skeleton is quite symmetric. Therefore, the CT images of the intact contra-lateral bone were imported in Mimics and segmented by using a threshold value for bone. Changing the orientation of the CT images within Mimics, allows the creation of a mirrored model. Both the mirrored model and the model of the affected bone were exported in stl-format and imported in the software program Magics (Materialise NV[b]). The mirrored model was repositioned to coincide with the model of the affected bone (Fig. 5). After the export of the mirrored model in the new position in stl-format, it was imported in the Mimics project of the affected bone. This allowed to check the fit between the mirrored cortex and the intact cortex of the affected bone (Fig. 6).

It seems that the human body is not as symmetric as initially thought. Figure 6 shows a bad fit between the mirrored cortical surface and the affected bone, especially in the direct environment of the tumor. This is due to the fact that the intact cortical bone around the tumor is also bulging out. In this case, mirroring the contra-lateral bone did not result in a good reconstruction of the cortical surface. The reconstructed surface did not coincide with the original cortical

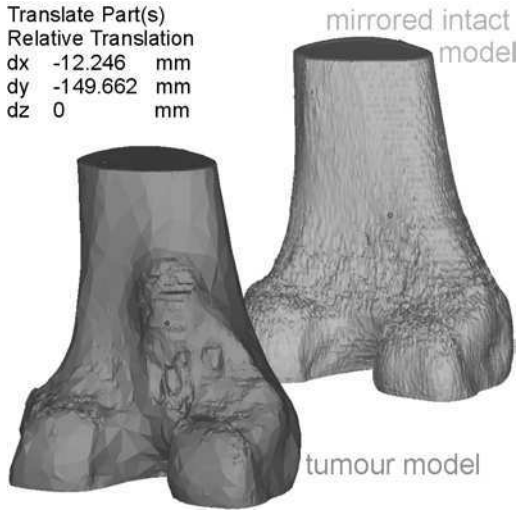


Fig. 5. Repositioning of mirrored intact model to coincide with tumor model.

[b]Materialise NV, Technologielaan 15, 3001 Leuven, Belgium.

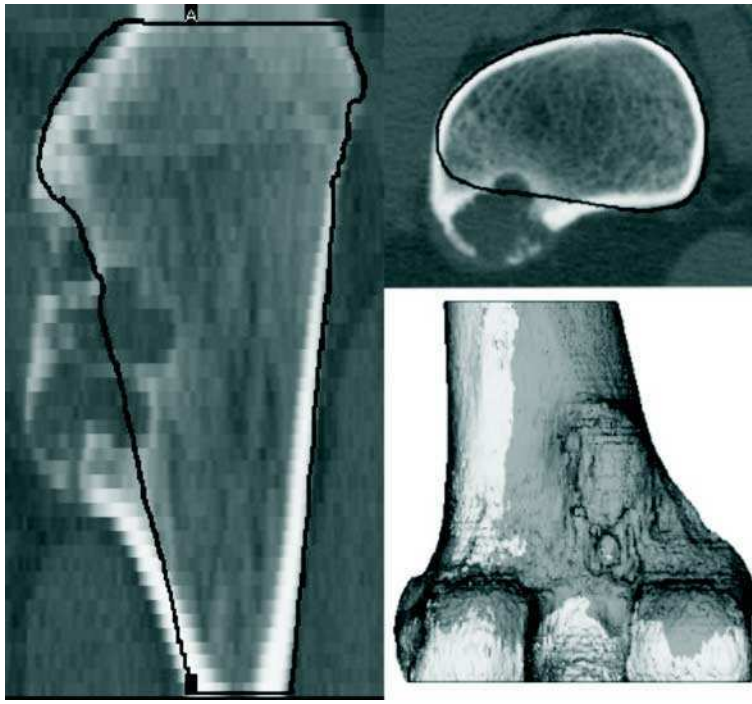


Fig. 6. Mirrored model visualized in Mimics to control the reconstructed surface with respect to the affected cortical surface.

surface unaffected by the tumor. Moreover, this method will not be of any use in cases where the images of the contra-lateral bone are not available or where the contra-lateral bone is not intact (e.g. deformed or affected by a disease). The advantage of this method is the low time consumption. There is only one extra segmentation and a repositioning of the contra-lateral bone for control purposes is needed.

2.3.2. *Manual editing in Mimics*

The software program Mimics allows to locally adapt the segmentation of an object, by a drawing and an erasing tool (Fig. 7). The segmentation of the affected bone was manually edited in each slice to reconstruct the cortical surface at the tumor zone. Since this operation had to be performed slice by slice it resulted in a wrinkled surface (Fig. 8). To obtain a smoother surface, the editing operation must be repeated. Disadvantageously, this is a very time consuming task that do not assure the creation of a smooth surface. Nevertheless, this method resulted in a reconstruction of the cortex at the tumor site, without affecting the shape of the intact bone around the tumor site. This method has the advantage that it is generally applicable, since it only uses the CT images of the affected bone.

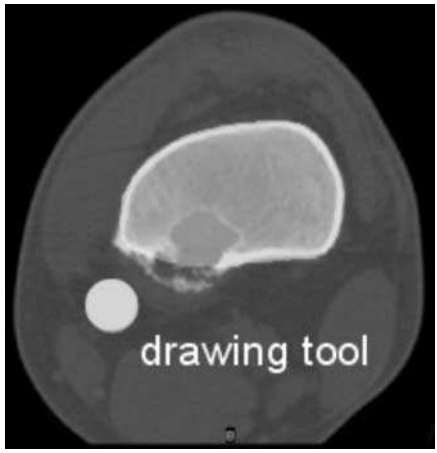


Fig. 7. The circular drawing tool within the Mimics software to manually reshape the cortex in the affected zone.

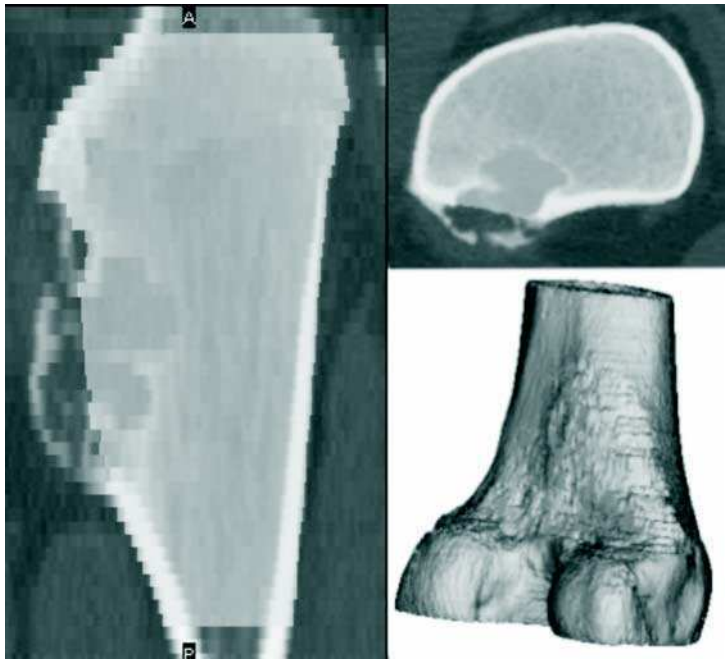


Fig. 8. Result of the manual reshaping of the cortex.

2.3.3. *CAD-based technique*

The third reconstruction method started from the segmentation of the affected bone in Mimics. The contour lines of the segmented bone in each of the CT images were exported in IGES-format (Fig. 9). These lines were imported in Mechanical Desktop

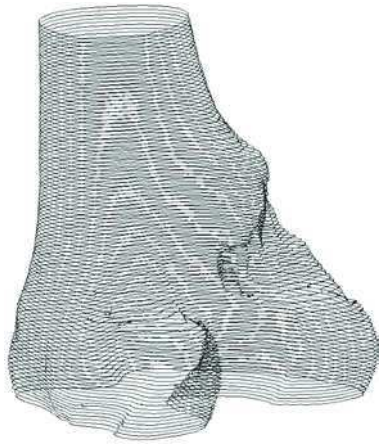


Fig. 9. IGES contours of tumor model.

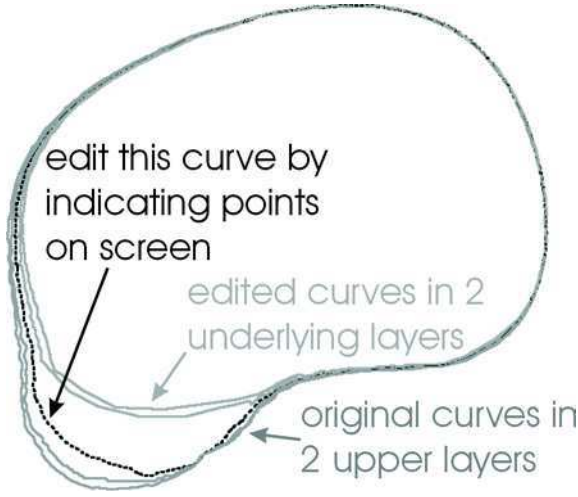


Fig. 10. Editing contours.

(Autodesk, Inc.) and edited by a routine "editcurves"[21] developed at the division BMGO[c]. This routine allowed the visualization of the contours in five successive layers. The contour in the middle layer can be adapted by indicating some points through which the curve must pass. The original curve is cut near the first and the last indicated point. The part in between is replaced by a curve fitting through the other indicated points. The already adapted curves in the two underlying layers and the original curves in the two upper layers are also shown (Fig. 10). After editing

[c]Division of Biomechanics and Engineering Design, K.U. Leuven, Celestijnenlaan 300C, 3001 Heverlee, Belgium.

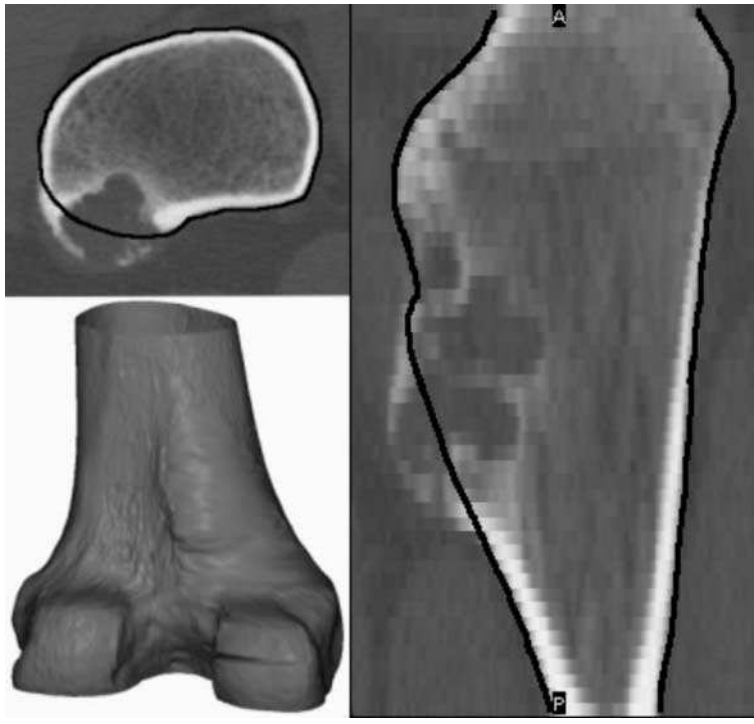


Fig. 11. Result of reshaping of cortex in CAD environment.

the curves in all the layers, a surface is created through these curves and exported in stl-format. This stl-surface is imported in the Mimics project of the affected bone for the comparison of the reconstructed surface with the original affected cortical surface (Fig. 11).

This CAD-based technique is also a generally applicable method. Moreover, it results in a smooth surface with an accurate reproduction of the unaffected cortex. Unfortunately, it requires a manual input, but it is not as time consuming as the previous method — manual editing in Mimics.

Based on the comparison of these three different reconstruction methods, it seems that the CAD-based method will be preferred to reconstruct the cortical surface at the tumor site.

After this reconstruction step a digital description of the reconstructed cortical surface of the affected bone is available in stl-format.

3. Manufacturing Process

Now that the desired shape of the titanium membrane is set, the membrane has to be manufactured. Since the titanium membrane is a personalized implant, it has

a specific shape for each individual patient. Therefore, the manufacturing process must be suited for the production of individual objects at a limited cost.

The manufacturing process exists in two steps: (1) the production of a physical model of the reconstructed cortex and (2) the drawing of the titanium membrane by using the hydroforming technique.

3.1. *Physical model*

The solid freeform fabrication (SFF) techniques, also called rapid prototyping (RP) techniques, are well suited for the manufacturing of individual objects with complex shapes. These techniques involve layer-wise fabrication of parts from a CAD representation without the use of any part-specific tooling. Some of the most widely used RP processes are stereolithography, three-dimensional printing, selective laser sintering, laminated object manufacturing, computer aided manufacturing of laminated engineering materials, and fused deposition modeling.[22]

The stereolithography technique was selected to manufacture a physical model of the reconstructed cortical surface due to its availability. Stereolithography (STL) builds a three-dimensional model from liquid photosensitive monomers that solidify when exposed to ultraviolet light. A highly focused UV laser traces out the first layer, solidifying the model's cross-section while leaving excess areas liquid. Then another layer of resin is coated on top of the cured layer, so the laser can trace the second layer atop the first. This process of curing and re-coating is repeated until the part is completed (Fig. 12). Afterwards the solid part is removed and rinsed

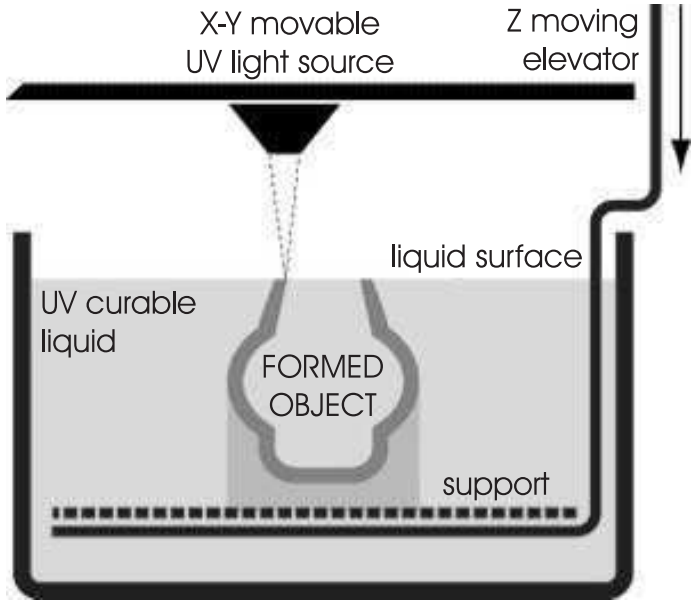


Fig. 12. Stereolithography process.

Fig. 13. Stereolithography model of reconstructed bone.

clean. Supports are broken off and the model is placed in an ultraviolet oven for complete curing.

Figure 13 shows the stereolithography model of the reconstructed cortical surface of the affected bone.

3.2. *Hydroforming process*

The hydroforming or Guerin process is a drawing process in which a rubber diaphragm backed with fluid pressure is used instead of a punch to form a flat metal sheet into the desired shape by pressing it against a die or mould (Fig. 14).[23,24]

First the mould for the hydroforming operation must be made. Therefore, the stereolithography model is positioned by the machine operator and reproduced in metal powder reinforced epoxyresin by a two step impression technique (Fig. 15). This mould is then used to press the titanium sheet (Fig. 16).

Hydroforming is a very complex manufacturing process that requires a high degree of experience and technical knowledge. For instance, the number of steps to complete the drawing process must be chosen together with the magnitude of the

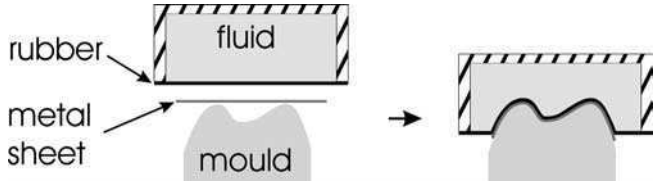


Fig. 14. Hydroforming process.

Fig. 15. Mould.

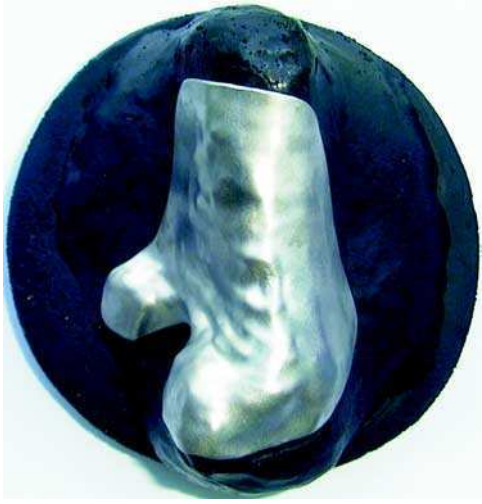


Fig. 16. Titanium membrane on mould.

fluid pressure. The fluid pressure is a critical parameter; since excessive pressure leads to tearing of the metal sheet whereas insufficient pressure results in a wrinkled surface. Especially for larger parts with a very small thickness, the failure by wrinkling is more likely. The pressure usually ranges from 30 to 500 bar.[25] Despite these difficulties, hydroforming limits the geometrical and dimensional errors, and reduces the localized thinning of the sheet. Nevertheless, the hydroforming process limits the extent of the area on which the titanium sheet can be pressed, i.e. a curvature over an angle larger than 160° cannot be reached.

4. Biomechanical Evaluation

A biomechanical evaluation of the novel implant is needed before clinical application. Moreover, some implant parameters must still be determined to assure a sufficient strength and stiffness of the bone reconstructed with a titanium membrane: the membrane thickness, the surface macrostructure, and the number and type of screws required to fix the membrane to the bone.

These implant parameters will depend on the fact whether the implant is a temporal or a lasting device. Moreover the type of filling material can have an influence on the implant parameters. A reconstruction with a titanium membrane in combination with bone graft aims at bone regeneration. Hence, the membrane should be designed to provide an optimal mechanical loading for stimulating bone regeneration on the one hand, and reducing the fracture risk of the bone on the other hand. In this case, the membrane is a temporal implant that will be removed once sufficient bone regeneration is diagnosed. When the latter does not occur, it will be a lasting implant. For a reconstruction with a titanium membrane in combination with bone cement, the membrane must be designed to only withstand the loads during normal functioning. The choice of the filling material as well as the decision whether or not remove the implant in proper time is the responsibility of the surgeon.

The biomechanical evaluation performed in this study focused on the design of a lasting implant for the specific case of a giant cell tumor in the proximal tibia. This case was chosen because of the relatively high occurrence of tumors at this site and the heavy loads acting on the tibia during normal daily activities. The latter implies a high fracture risk of the tibia after resection of the tumor and also high requirements concerning the strength and stiffness of the titanium membrane to reconstruct the defect.

4.1. *In vivo loads on tibia*

For the biomechanical evaluation of the titanium membrane used for the reconstruction of a defect in the proximal tibia, it is important to have an idea of the loads acting on the tibia during normal daily activities.

Several external forces act on the tibia: e.g. muscle forces, ligament forces, and joint forces as well as reaction forces. Figure 17 shows the lateral and the medial view of a model of the knee joint with indication of these force vectors acting on the proximal tibia.

A number of researchers[26–28] made a model of the knee joint to calculate the forces acting on the tibia. Tables 1 and 2 present an overview of some values measured or calculated for the tibiofemoral joint forces. The magnitude of these compressive and shear forces depends largely on the type of activity being performed and on the joint angle. For most of the activities, the compressive tibiofemoral joint force is an order of magnitude larger than the shear forces. Only for deep flexion and for the squat exercise, the anterior shear force has the same order of magnitude as the compressive component.

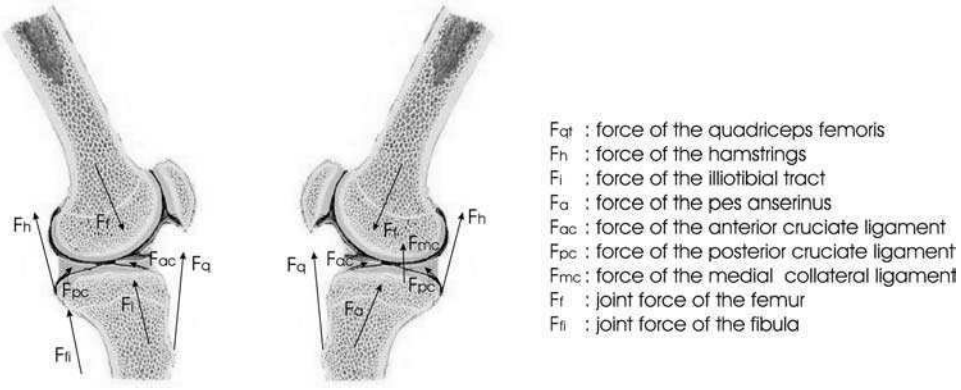


Fig. 17. Lateral and medial view of a knee joint model with indication of the forces acting on the proximal tibia.

Table 1. Tibiofemoral joint force: compressive, anterior shear, and posterior shear with corresponding knee angle.[27]

Source	Activity	Knee angle (°)	Compressive Force (xBW)	Knee Angle (°)	Ant. Shear Force (xBW)	Knee Angle (°)	Post. Shear Force (xBW)
Ericson *et al.*	Cycling	60–100	1.2	105	0.05	65	0.05
Morrison	Walking	15	3.0	5	0.4	15	0.2
Harrington	Walking	—	3.5	—	—	—	—
Smidt	Isom.[c] extension	60	3.1	—	—	30	0.4
	Isom.[c] flexion	5	3.3	45	1.1	—	—
Morrison	Up stairs	45	4.3	—	—	30	0.05
	Down stairs	60	3.8	5	0.6	15	0.1
Ellis *et al.*	Rising from chair	—	3 to 7	—	—	—	—
Kaufman	Isom.[c] exercise:						
	60 deg/sec	55	4.0	75[a]	1.7	25[b]	0.3
	180 deg/sec	55	3.8	75[a]	1.4	25[b]	0.2
Dahlkvist *et al.*	Squat-Rise	140	5.0	140	3.0	—	—
	Squat-Descent	140	5.6	140	3.6	—	—

BW = body weight, *a* = flexion, *b* = extension, *c* = isometric.

A correct implementation of the direction and the attachment points of all the forces acting on the tibia is very complex. Therefore, the model will be simplified; only the compressive component of the tibiofemoral joint force acting on the articular surfaces will be taken into account.

For the biomechanical analysis five frequent loading conditions are chosen: standing on both legs, walking, stair climbing, getting out of a chair and jumping. The values used for the compressive component of the tibiofemoral joint force for these different conditions are based on literature and are respectively 0.5, 3, 4, 6, and 24 times body weight.

Table 2. Tibiofemoral and patello-femoral joint forces during several activities.

Activity	Author	Tibio-femoral	Patello-femoral
Level walking	Morrison	3.40	0.6
Walking up ramp	Morrison	3.97	1.6
Walking down ramp	Morrison	3.95	2.6
Ascending stairs	Morrison	4.25	1.8
Descending stairs	Morrison	3.83	2.9
Getting out of a chair unaided by arms	Seedhom and Terayama	2.80	2.5
	Ellis *et al.*	4.00	3.3
Jumping	Smith	24	20

4.2. *Experiments*

Experimental tests were performed to determine the buckling strength of the membrane, the strength at the perforation holes, and the strength of the screw-bone fixation. The results of these experiments were evaluated with respect to the loads acting on the tibia during normal daily activities to determine the following parameters: the membrane thickness, the surface macrostructure, and the number and type of screws needed to fix the membrane to the bone.

4.2.1. *Buckling strength of the membrane*

An experimental set-up was made to measure the critical compression load that initiates the buckling process of the membrane (Fig. 18). Therefore, half cylinders with a diameter of approximately 70 mm were formed of commercially pure titanium grade 1 with thicknesses of 0.2, 0.3, and 0.5 mm. Besides the cylinders with a

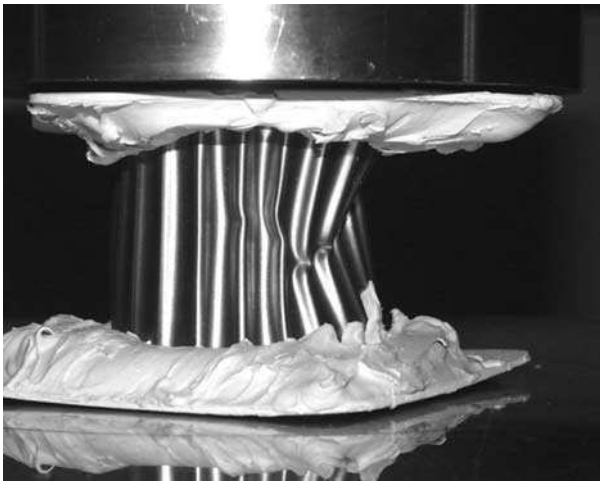


Fig. 18. Set-up to measure the buckling strength of the membrane.

flat surface, also cylinders with a wave pattern and thicknesses of 0.2 and 0.5 mm were tested. The membranes were loaded on an INSTRON machine at constant deformation rate of 1 mm/min. The load was measured as a function of time with a force cell (INSTRON, type 2511–214) with a range of 25 kN. The force versus time curve showed a few small peaks as the loading progressed and after approximately 120 seconds, i.e. about 2 mm impression, a large drop in force occurred (Fig. 19). At this last peak, the membrane buckled at the middle and always inwards.

Table 3 shows the average buckling forces for the different types of membranes.

The buckling strength was compared with the five loading conditions. To this end, a transformation of the loading condition into a force acting on the upper edge of the membrane was made based on the following assumptions:

(1) Only the cortical bone carries the load.
(2) The load is equally distributed over the contour of the tibia.

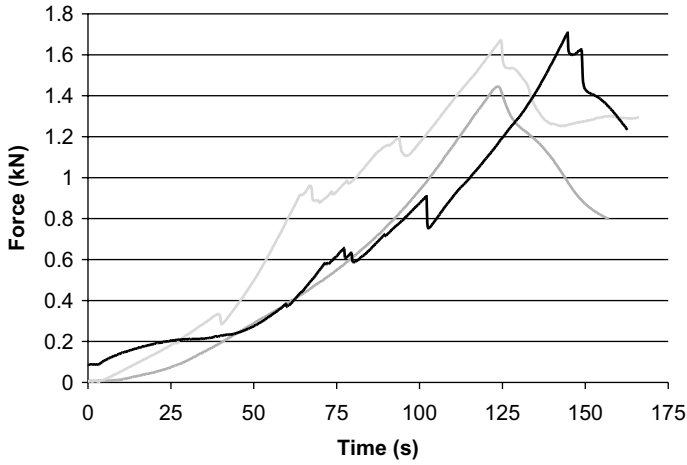


Fig. 19. Force versus time curve during buckling test of 0.2 mm thick titanium membrane without wave pattern.

Table 3. Average and standard deviation of the force that initiates the buckling process of a membrane.

Membrane Thickness (mm)	Average Buckling Force (N)	Standard Deviation (N)	Number of Membranes Tested
No wave pattern			
0.2	1608	142	3
0.3	4603	167	3
0.5	11196	1325	2
Wave pattern			
0.2	3177	246	3
0.5	12433	1924	4

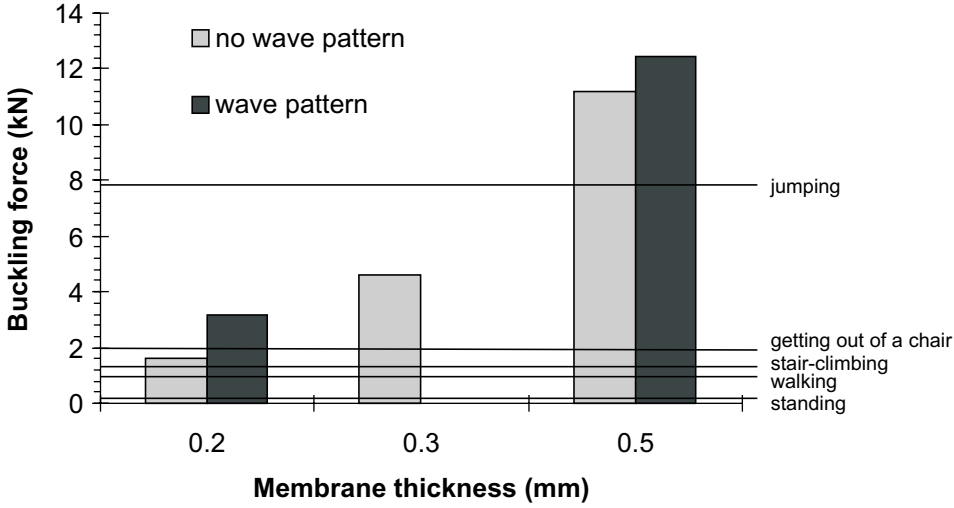


Fig. 20. Buckling force versus membrane thickness with indication of the different loading conditions.

(3) The membrane is loaded directly with the percentage of the load that is equal to the fraction of the membrane width to the tibial perimeter (approximately 100 mm to 240 mm).
(4) An average person with a body weight of 80 kg is assumed.

Figure 20 shows the average buckling strength of the different types of membranes as well as the membrane load for the five loading conditions. A minimal membrane thickness of 0.5 mm is needed to prevent buckling of the membrane during jumping. To exclude the buckling risk during limited daily activities (standing, walking, stair climbing, and getting out of a chair) a minimal thickness of 0.3 mm is needed or a wave pattern in case of a membrane of 0.2 mm thick.

4.2.2. *Strength at the perforation holes*

The maximum force at which the titanium membrane at the perforation sites ruptures was measured. The experimental set-up consisted of an INSTRON machine with a force cell (INSTRON, type 2511-317) with a range of 5 kN. The upper part of a small titanium plate was fixed in a clamp. A cylinder was inserted through a hole in the lower part of the titanium plate and fixed in another clamp (Fig. 21). A tensile deformation was applied an a rate of 1 mm/min, and the force was measured as a function of time.

A cylinder of 4.0 mm diameter was inserted through a perforation of 4.2 mm diameter in a titanium (grade 1) plate of 0.2, 0.3, and 0.5 mm thickness to investigate the influence of the membrane thickness. The influence of the perforation diameter was determined by using a cylinder with 2, 3, 3.5 or 4 mm in combination with

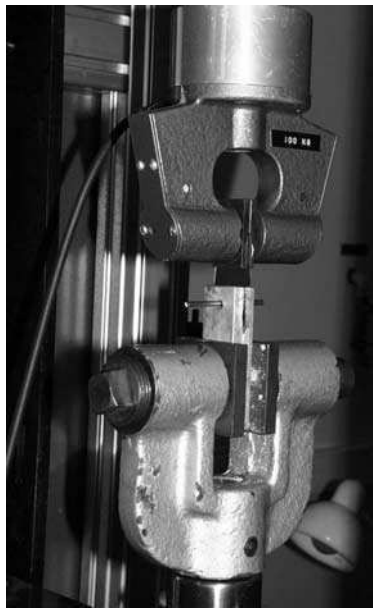


Fig. 21. Set-up for testing the strength of the perforation holes.

Table 4. Average and standard deviation of the force that initiates rupture of a perforation hole of 4.2 mm diameter in a titanium membrane of respectively 0.2, 0.3, or 0.5 mm thick.

Membrane Thickness (mm)	Average Force (N)	Standard Deviation (N)	Number of Membranes Tested
0.2	214	20	5
0.3	362	34	4
0.5	1018	133	5

Table 5. Average and standard deviation of the force that initiates rupture of a perforation hole of respectively 2.2, 3.2, 3.7 or 4.2 mm diameter in a 0.3 mm thick titanium membrane.

Diameter of Perforation Hole (mm)	Average Force (N)	Standard Deviation (N)	Number of Membranes Tested
2.2	280	15	2
3.2	323	30	6
3.7	357	62	4
4.2	367	24	4

a 0.3 mm thick titanium (grade 1) plate with perforation diameters of respectively 2.2, 3.2, 3.7, or 4.2 mm.

Tables 4 and 5 give the average forces at which the perforation sites ruptured for the different cases. The average rupture force for a perforation of 4.0 mm diameter

is exponentially correlated with the membrane thickness according to the equation $y = 75.791e^{5.1836x}$, with a correlation coefficient R^2 of 0.98. There was no correlation observed between the average rupture force and the diameter of the perforation hole for a titanium membrane of 0.3 mm thick.

4.2.3. *Strength of screw-bone fixation*

The load on the tibia, which is mainly a vertical compression, is transmitted partly to the titanium membrane through the screws by transverse forces. Therefore, it is important to know the maximal transverse load a screw can withstand without failing, i.e. breaking of the surrounding bone or breaking of the screw itself. The strength of the screw-bone fixation was experimentally determined for three types of screws: a mono-cortical screw of 2.7 mm diameter and 16 mm length, a mono-cortical screw of 3.5 mm diameter and 18 mm length, and a trabecular screw of 4.0 mm diameter and 28 mm length. A steel plate was screwed onto the cadaver tibia. The diameter of the perforation in the steel plate was 0.2 mm larger than the screw diameter. The tibia was positioned so that the screw was approximately horizontal and the plate vertical. An INSTRON machine applied a compressive load on the plate, hence a transverse load on the screw (Fig. 22). This test occurred at a constant deformation rate of 1 mm/min to avoid the effect of the deformation rate on the resulting measurement.[29] Five cadaver tibiae were available, in which each screw type was tested once or twice.

Fig. 22. Set-up for testing the strength of the screw-bone fixation: a screw is fixed in the cadaver bone and transversely loaded.

Table 6. Average and standard deviation of transverse load that a screw can withstand without failing.

Type of Screw		Average Force (N)	Standard Deviation (N)	Number of Screws Tested
Cortical screw:	2.7 mm diameter 16 mm length	305	184	9
Cortical screw:	3.5 mm diameter 18 mm length	420	257	10
Trabecular screw:	4.0 mm diameter 28 mm length	640	136	7

Table 6 presents the average transverse load measured for the three types of screws. The failing of the screw-bone fixation was always due to a fracture of the surrounding bone and not to a fracture of the screw itself. The trabecular screws are preferred due to the high load they can withstand.

4.2.4. *Conclusion*

Based on the experimental tests, it can be concluded that a bone reconstructed with a non-wave-formed titanium membrane of 0.3 mm thickness is able to withstand the loads occurring during limited daily activities. Compressive loads up to 6 times body weight are allowed. Jumping loads are forbidden, which would mean that the patient is not allowed to sport or dance.

The choice of a membrane without wave pattern was not only based on the fact that this provides already sufficient strength. A membrane with a wave-formed texture at the outer and the inner side would restrict the surgeon in the choice of the extent of the cortical window in the operating theatre. The area of the membrane with the wave pattern determines the minimal magnitude and moreover the exact position of the cortical window. In the case of a membrane with waves only at the periosteal side, there is no per-operative restriction, but there may be some discomfort for the patient. On the one hand, the membrane could be very near the skin and hence the outer wave pattern could be felt by the patient. On the other hand, the membrane could be covered by muscles, which would lead to shearing of those muscles over the waved titanium membrane. A membrane without wave texture is preferred, since it has a smooth surface and it mimics the shape of the original bone. Moreover the design and the production of a smooth membrane are easier than those of a membrane with wave pattern.

The buckling strength was compared with the loads acting on the tibia of an average person during main daily activities, being standing, walking, stair climbing, and getting out of a chair. Average values of the tibial joint force during these activities were found in literature. The tibial joint force consists of a large compressive component and a rather small shear component. Therefore, only the large compressive force was taken into account. Moreover, it was assumed that this force

was totally applied to the cortical rim. In reality the load is distributed over the tibial joint surface by the menisci, so the trabecular bone carries a part of the load. This means that the loads taken as reference for the evaluation of the buckling strength were overrated.

Trabecular screws with a diameter of 4.0 mm and a length of 28 mm are chosen to fix the implant, due to the high load they can withstand before failure of the screw-bone fixation. For this type of screw and for a titanium membrane of 0.3 mm thickness, the force that initiates failure of the screw-bone fixation is less critical than the force that starts rupture of the titanium membrane. According to the values for the latter force and the maximum load the membrane should withstand, i.e. the compression load that acts on the knee while standing up from a chair (6 times body weight), the number of trabecular screws of 4.0 mm diameter needed to fix the membrane onto the bone was chosen. For a person of 80 kg, a load of 2400 N acts on the titanium membrane, assuming that the load is fully carried by the cortex and half of the tibial cortical perimeter is replaced by the membrane. The critical force for rupturing the membrane is approximately 367 N, hence making it necessary to use at least seven screws to fix the membrane.

4.3. *Finite element analyses*

A further evaluation of the global concept, i.e. a tibia with proximal defect reconstructed with a titanium membrane, was performed by finite element analyses. This is necessary to have an idea of the stresses occurring in both the bone and the membrane during daily activities. These stresses must be limited to prevent either fracture of the bone or failure of the implant. Moreover, stress shielding must be avoided. The stresses in the bone must be above a certain lower bound to prevent bone loss.

4.3.1. *Finite element models*

A dry human cadaver tibia was selected to use as an example for the finite element modeling. First, the outer surface of the dry tibia was measured with a 3D digital coordinate measurement system. An orthopaedic surgeon (I.S.[d]) made a cortical window in the proximal tibia and resected the major part of the trabecular bone to simulate a cavity remaining after removing a giant cell tumor in the proximal tibia. Further, a CT scan with a 1 mm slice thickness was made of the dry tibia with the simulated tumor cavity. Based on these data, the following finite element models were made:

- the intact tibia;
- the tibia with the simulated tumor cavity;

[d]Ignace Samson, M.D., K.U.Leuven, UZ Pellenberg, Department of Orthopaedic Surgery, Weligerveld 1, 3212 Pellenberg, Belgium.

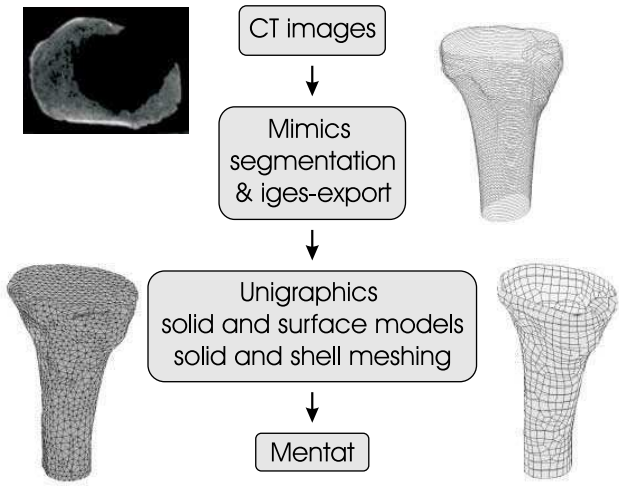


Fig. 23. Succeeding steps to create a finite element model.

- the tibia reconstructed with bone cement;
- the tibia reconstructed with a titanium membrane of 0.2 mm thick;
- the tibia reconstructed with a titanium membrane of 0.3 mm thick; and
- the tibia reconstructed with a titanium membrane of 0.5 mm thick.

Figure 23 shows the general flowchart for creating the finite element models. The first step is the acquisition of the inner and outer geometry of the bone. Therefore, the CT images of the bone were acquired and imported in the software program Mimics. After segmentating the cortical and trabecular bone, the inner and outer contours of these segmented structures were exported in IGES-format. Based on these, IGES-contours solid and 3D free from surface models were created within the software program Unigraphics (EDS). The creation of the outer surface of the intact tibia did not follow this scheme. Therefore, the point cloud resulting from the 3D digital coordinate measurement was imported in Unigraphics and used to create a surface model. Further, the solid models were meshed with the solid mesher by using tetrahedral elements and the surface models with a shell mesher using quadrilateral elements. Finally, these models were imported in Mentat (MSC.Software Corporation) for the finite element analysis.

To approximate a real tibia as good as possible a variable cortical thickness was modeled (Fig. 24). Therefore, it was necessary to use solid elements for the cortical mesh. The elements of choice for the cortex were 3D 20-node brick elements. These isoparametric, arbitrary hexahedral elements use triquadratic interpolation functions to represent the coordinates and the displacements. Hence, the strains have a linear variation within the element. The trabecular bone was built with 3D 10-node tetrahedron elements. This is a second-order isoparametric tetrahedron, which allows for an accurate representation of the strain. It has three translational

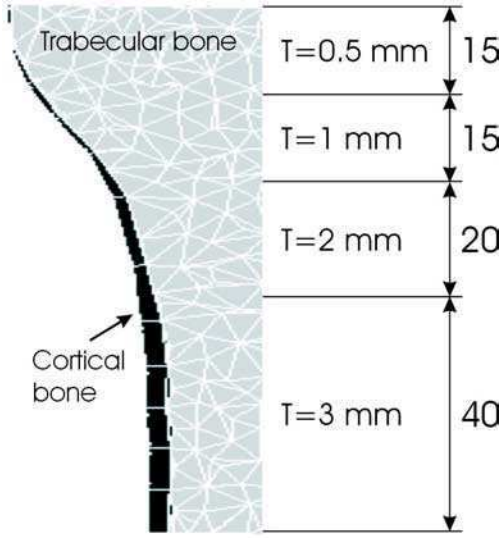


Fig. 24. Half of the intact tibia model with variable cortical thickness T.

degrees of freedom and uses four-point integration. The titanium membrane was meshed using also hexahedral elements. The filling material PMMA was built with two types of elements. The part that filled the cavity in the trabecular bone was made with tetrahedron elements, whereas the part that filled the cortical window used hexahedral elements.

For reason of simplicity, linear elastic isotropic properties were assigned to all materials. Table 7 represents the values used for the Young's modulus and the Poisson ratio. This simplification is justified since the effect of different reconstruction techniques will be analyzed by comparison, and since the fact that only compressive loads are applied to the models.

The boundary conditions applied to the finite element models were the loads, the simulation of embedding of the model to the environment, the types of contact between the different materials, and the simulation of the screw fixation.

The tibial joint force was applied as a pressure equally distributed over the entire articular surface. For a more easier implementation, the articular surface was cut, hence the surface load could be applied on a flat surface. The cortical and the

Table 7. Material properties used for the finite element analyses.

Material	Young's Modulus (MPa)	Poisson Ratio
Cortical bone	13700	0.3
Trabecular bone	1370	0.3
Titanium	110000	0.33
PMMA	2000	0.3

Table 8. Loads applied for the finite element analyses.

Loading Condition	Tibiofemoral Joint Force (N)[a]	Equivalent Face Load (MPa)[b]
Standing	400	0.1197
Walking	2400	0.7183
Stair climbing	3200	0.9577
Getting out of a chair	4800	1.4365
Jumping	19200	5.7461

a : for a body weight of 80 kg.
b : for an articular surface with an area of 3341.4 mm^2.

trabecular bone in this plane are directly loaded with an equally distributed face load. The five selected loading conditions were simulated. The compressive component of the tibiofemoral joint force of these loading conditions was transformed into an equivalent face load (Table 8). Therefore, a person with a body weight of 80 kg was assumed.

The distal part of the tibia was completely fixed to the environment; neither translation or rotation were allowed.

To oblige the materials to have the same displacement, a glue contact was applied between the PMMA and the trabecular bone, and between the PMMA and the cortical bone. This choice was made since PMMA penetrates into the pores of the bone. PMMA binds at a macroscopic level with trabecular bone and at a microscopic level with cortical bone. A touching contact was used between the titanium membrane on the one hand and the cortical bone or the PMMA on the other hand. This type of contact allows the materials to slide over each other, but inhibits the penetration into each other.

The screws to fix the membrane onto the bone were not modeled as in reality to limit the size of the finite element model and for calculation purposes. Therefore, the screws were modeled as tie-nodes by adding a link. This means that the nodes of the titanium membrane at the position of the screws were assumed to have the same displacement as the nodes on the cortical bone where the screws were inserted. The membranes were fixed with seven screws, hence seven links were added.

4.3.2. *Results*

For each of the six models the five loading conditions were simulated and the equivalent Von Mises stress was calculated. Histograms were made of the Von Mises stress in all the nodes belonging to a type of material (i.e. cortical bone, trabecular bone, PMMA, or titanium) for each model and for each loading condition. The median was taken as a representative value for the average stress, since it is not influenced by the absolute values of the inaccurate peak stresses resulting from the contact types and the modeling of the screw fixation.

For the loading condition "getting out of a chair" the histograms of the Von Mises stresses in respectively cortical and trabecular bone of the six models are

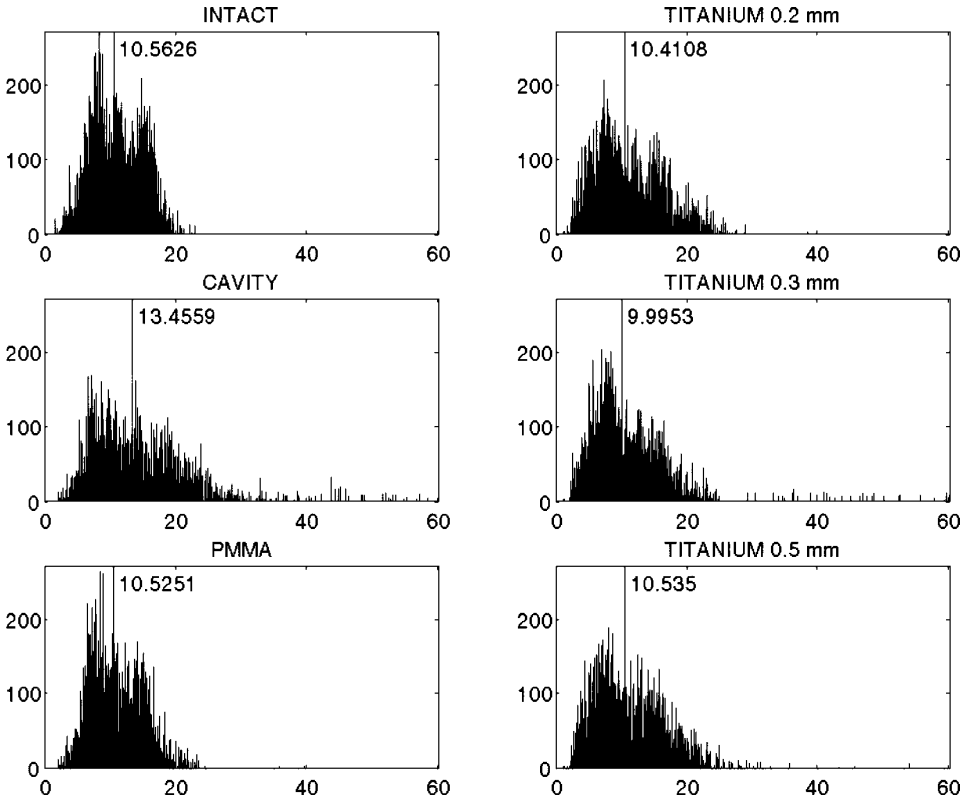


Fig. 25. Histograms of the Von Mises stresses in the cortical bone of the six models for the loading condition "getting out of a chair", with indication of the median value (MPa).

shown in Figs. 25 and 26. For the same loading condition, Fig. 27 shows the histograms of the Von Mises stress in PMMA and in titanium. These histograms were representative for all five loading conditions.

Tables 9–11 represent the values of the median of the Von Mises stress in respectively cortical bone, trabecular bone, PMMA, and titanium for the five loading conditions.

As expected, in both cortical and trabecular bone higher stresses were observed for the cavity model than for the intact model, for each of the five loading conditions. The distribution of the stresses was different for these two models. The histograms of the stresses in the intact model showed a relatively small distribution, whereas the histograms of the stresses in the cavity model showed a wider distribution with a large extension towards high stress values. This was more pronounced for the stresses in trabecular bone than those in cortical bone. Analysis of the maximal stress values revealed that these were averagely 5.5 (range 2.6 to 9) times higher than those in the intact model. For the trabecular bone, the maximal strength was exceeded in each loading case, whereas for cortical bone only in the last two

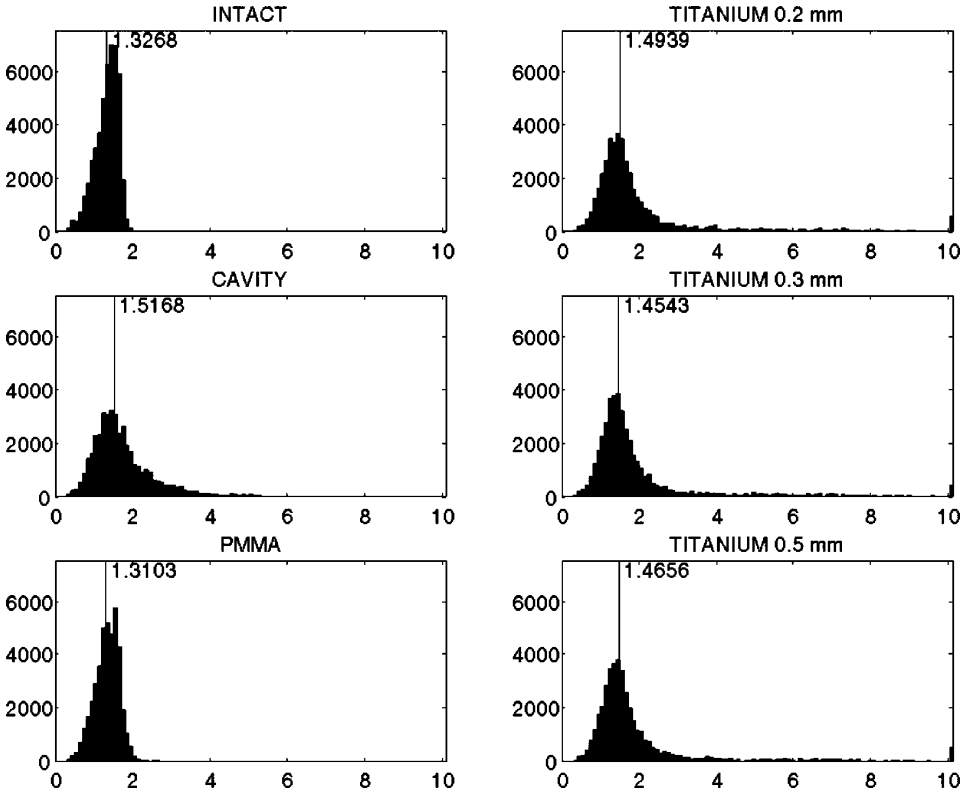


Fig. 26. Histograms of the Von Mises stresses in the trabecular bone of the six models for the loading condition "getting out of a chair", with indication of the median value (MPa).

cases (getting out of a chair and jumping). Reconstruction of the tibia with a tumor defect is certainly needed to lower the risk of fraction during normal daily activities.

The two different reconstruction techniques, i.e. PMMA as a filling material and a titanium membrane for reconstruction of the cortical window, showed an equal median stress value in cortical bone as the intact model. The maximal stress values in cortical bone were for both reconstruction techniques lower than those calculated for the cavity model. Only the PMMA reconstruction lead to similar stress distributions and approximately equal median stress values in trabecular bone as the intact model. The reconstructions with a titanium membrane showed similar stress distributions and approximately equal median stress values in trabecular bone as the cavity model. Nevertheless, the maximal stress values in trabecular bone were lower for the models reconstructed with a titanium membrane than for the cavity model. These differences could be explained by the fact that both reconstruction techniques restored the cortical window, but only the PMMA reconstruction filled the cavity in the trabecular bone.

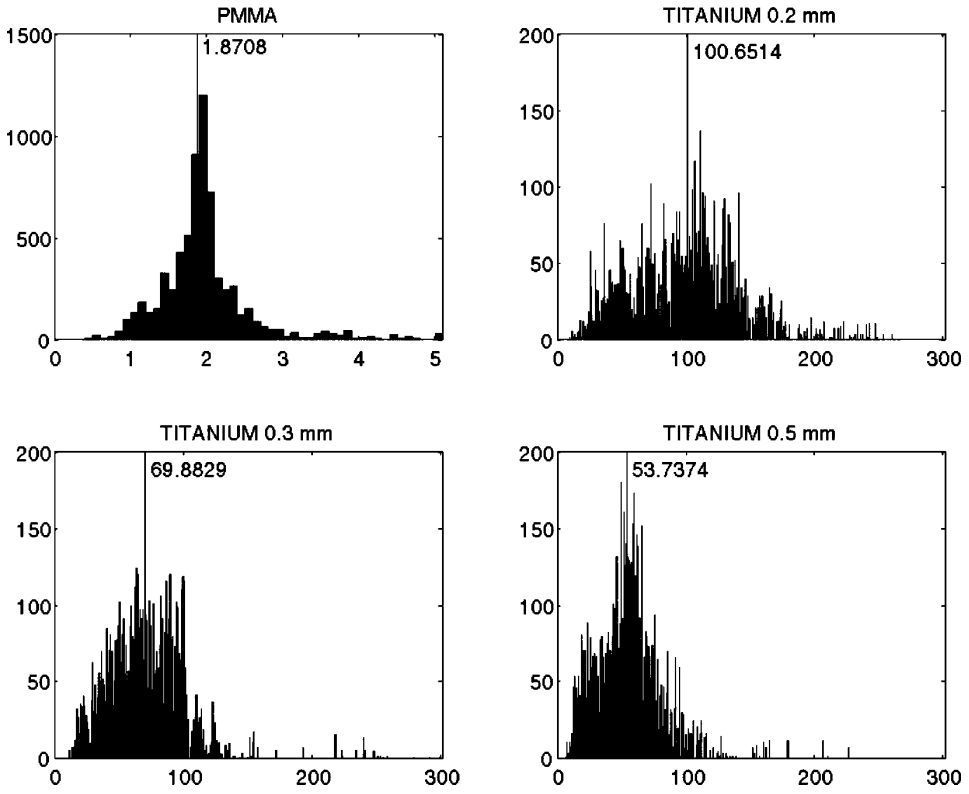


Fig. 27. Histograms of the Von Mises stresses in PMMA and in titanium for the loading condition "getting out of a chair", with indication of the median value (MPa).

Table 9. Median value of the Von Mises stress in cortical bone (MPa).

	Standing	Walking	Stair Climbing	Getting Out of a Chair	Jumping
Intact	0.9087	5.3272	7.0894	10.5626	42.0528
Cavity	1.1505	6.8006	9.0320	13.4559	53.8030
PMMA	1.0099	5.3752	7.0792	10.5251	41.8574
Titanium 0.2	0.8752	5.2109	6.9462	10.4108	42.0065
Titanium 0.3	0.8883	5.0336	6.6937	9.9953	39.5729
Titanium 0.5	0.9396	5.3313	7.0778	10.5350	42.0026

Analysis of the stresses in the PMMA filling material showed that the median stress value is a little bit higher than the corresponding value in trabecular bone, but much lower than the corresponding value in cortical bone. Only for the first loading condition (standing) the maximal stress is within the *in vivo* allowed working stress (14–17 MPa) in bone cement.[30] There is a considerable risk of fracture of the bone cement.

Table 10. Median value of the Von Mises stress in trabecular bone (MPa).

	Standing	Walking	Stair Climbing	Getting Out of a Chair	Jumping
Intact	0.1117	0.6638	0.8846	1.3268	5.2967
Cavity	0.1281	0.7597	1.0116	1.5168	6.0671
PMMA	0.1176	0.6592	0.8765	1.3103	5.2253
Titanium 0.2	0.1255	0.7472	0.9963	1.4939	5.9727
Titanium 0.3	0.1251	0.7279	0.9700	1.4543	5.8087
Titanium 0.5	0.1260	0.7337	0.9768	1.4656	5.8433

Table 11. Median value of the Von Mises stress in PMMA and titanium (MPa).

	Standing	Walking	Stair Climbing	Getting Out of a Chair	Jumping
PMMA	0.2331	0.9463	1.2509	1.8708	7.4638
Titanium 0.2	8.5380	50.5401	67.2393	100.6514	405.0290
Titanium 0.3	7.3218	36.1553	47.5035	69.8829	266.6067
Titanium 0.5	6.6794	27.6383	36.4939	53.7374	202.8965

The stresses in the titanium membrane were much higher than those in the cortical and the trabecular bone. A significant decrease of the median stress value in titanium was observed for an increase in thickness from 0.2 mm to 0.3 mm. Further increase in thickness with 0.2 mm gave rise to a smaller gain in median stress values in the titanium. Comparison of the stress distributions shows that the larger the thickness of the membrane, the smaller the stress distribution and the extension towards high stress values. For all the loading cases except jumping the median stress values were lower than the yield strength of the five grades of commercially pure (cp) titanium (Table 12). When also considering the jumping load only cp titanium grade 4 and 5 have a yield strength higher than the median stress value. The maximal stress values (when excluding the peak stresses around the screw links) in a titanium membrane of 0.3 mm thick were lower than respectively 700 MPa for jumping and 200 MPa for the other four loading conditions.

In conclusion, it seems obvious that when dealing with a large tumoral defect, reconstruction is needed to reduce the fracture risk of the bone. The above results show that a reconstruction with a titanium membrane reduces the stresses in the

Table 12. Mechanical properties of cp titanium grade 1 to 5.

Titanium Grade	Yield Strength MPa	Tensile Strength MPa	Young's Modulus GPa
1	200	290–410	105
2	270	390–540	105
3	350	460–590	105
4	410	540–740	105
5	890	930	115

cortical bone to the level of those in the cortex of an intact bone, and also lowers the maximal stresses in the trabecular bone in comparison to those in a bone without any reconstruction. A minimal membrane thickness of 0.3 mm is needed to limit the stresses in the membrane itself during normal daily activities (except jumping).

5. First Clinical Application

The titanium membrane was first clinically applied for the specific case of a juxta-cortical cartilaginous tumor in the left distal femur, since this was the first clinical case presented at that time by the surgeon (I.S.). It was considered justified to apply the design parameters derived from the biomechanical study on the proximal tibia to this clinical case because of the following reasons.

(1) The specific example of the giant cell tumor in the proximal tibia does not prejudice the general relevance of the results with regard to the screw-bone and the screw-metal interface. These are in fact applicable to all reconstructions in epiphyseal areas of long bones.
(2) The loads acting on the distal femur are comparable with those acting on the proximal tibia. The physical loads on the femur and the tibia are both mainly axially directed and are of the same magnitude. Therefore, the reconstruction of both proximal tibia and distal femur can be simplified to a tube with reinforcement of the outer wall.
(3) The major function of the titanium membrane for the clinical case of the low-grade chondrosarcoma of the distal femur is containment of the filling material.

A male competition swimmer of 18 years complained about pain at the medial side of the left knee. A tumoral process at the posterior-medial side of the distal metaphysis of the femur was found. Figures 28 and 29 show respectively a frontal

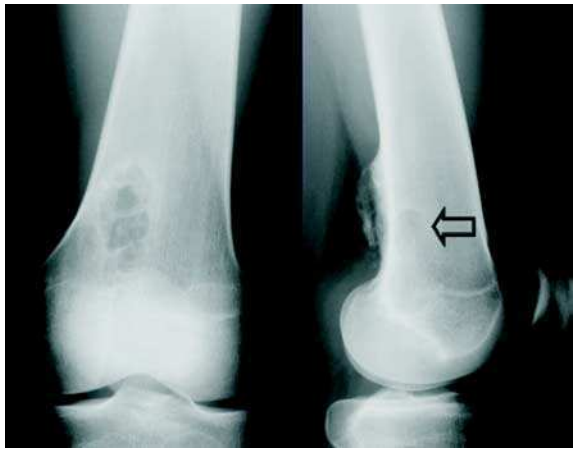


Fig. 28. Frontal and lateral radiograph.

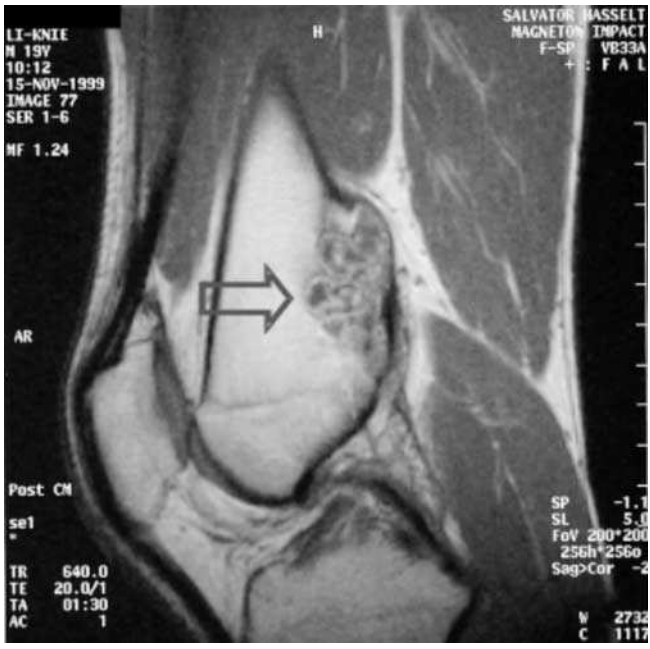


Fig. 29. MR image.

and lateral radiograph and a MR image taken four months before surgery. Due to both the benign character of this tumoral process and the young age of the patient, it is preferable to fill the cavity remaining after tumor resection with bone graft. The large size of this juxta-cortical chondroma on the one hand and its position on the other hand, make the reconstruction of the bone defect with bone graft and conventional methods very difficult. Therefore, the surgeon (I.S.) decided to use a personalized titanium membrane in combination with bone graft to reconstruct the femoral bone. A major function of the titanium membrane was to prevent the risk of drifting of the bone graft into the knee joint, which would lead to major problems for the patient. Another advantage was the possibility of compressing the bone graft to ensure a good filling of the cavity. Hence, a better contact of the filling material to the walls of the cavity is achieved, which enhances the remodeling of the bone graft.

A membrane was designed and manufactured according to the procedure given in Secs. 2 and 3. Based on the design constraints conforming to the biomechanical evalution and including a safety factor, a thickness of 0.4 mm was chosen for the titanium membrane. Because 0.4 mm thick titanium was not available when the clinical case was presented, a titanium sheet of 0.5 mm thick was selected. Afterwards the membrane was reduced to a thickness of 0.4 mm by chemical etching. This was also performed to remove impurities from the membrane resulting from the hydroforming process. The titanium membrane covered a maximum area around

the tumor. The maximum size of the membrane was determined by the limits of the hydroforming process and hence avoided any restriction of the surgeon during the surgery. Besides the titanium membrane, also a transparent membrane in Perspex was made on the same mould as a trial model for per-operative use.

The day before surgery, the titanium membrane, the Perspex trial model, the metal scissors to trim the membrane to the exact size, and the punch tool to punch the screw holes were gas sterilized.

The surgery started with making a medial incision to expose the left distal femur. After drilling some holes and chiselling the cortical bone, the tumor was resected en-bloc. Before reconstruction, the remaining tumor cavity was rinsed. Firstly, the titanium membrane was placed onto the bone to control the fit and to check whether the membrane could be used. This evaluation being positive, the transparent membrane was used to check the overlap. The contour of this Perspex membrane was trimmed to the desired shape and the positions for the screws were marked. This transparent membrane was placed on top of the titanium membrane and the outer contour was drawn onto the titanium membrane. The titanium membrane was cut with the metal scissors to the marked shape, and the holes for the screws were punched. Allogenous morselized bone grafts mixed with physiological fluid were pressed in the cavity. Because of the posterior position of the cavity these bone grafts fell out the cavity. Therefore the titanium membrane was filled with bone grafts and placed onto the bone. The holes for the screws were drilled and tapped in the bone through the punch holes in the membrane. Seven screws were used to fix the membrane to the bone. Figures 30 and 31 show respectively the per-operative result and the immediate post-operative radiographs. These show that the membrane fits perfectly in the remaining cortical bone and that the membrane is well fixed with the screws.

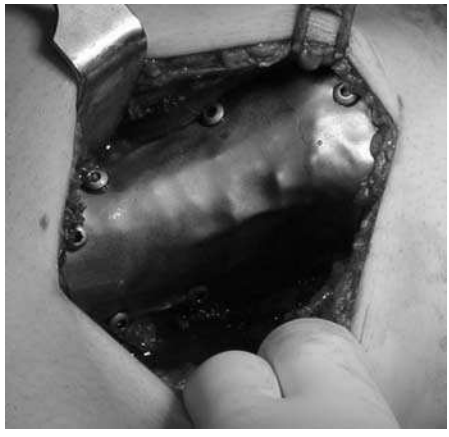


Fig. 30. Per-operative view of titanium membrane fixed onto the femoral bone.

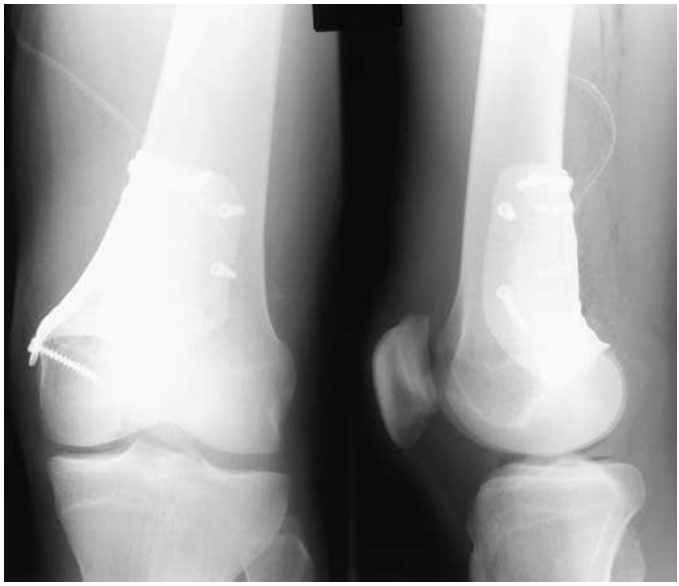


Fig. 31. Immediate post-operative X-rays.

Since the titanium membrane was designed and produced pre-operatively, the normal procedures in the operating theatre were not complicated. Moreover, there were no restrictions to the surgeon regarding the removal of the tumor because a per-operative fine-tuning of the shape and punching of the screw holes were possible.

Three months post-operative CT images showed bone regeneration, starting from the intact trabecular bone in the middle of the distal femur. However at the anterior-medial side a radiolucent line was seen at the border between trabecular bone and bone graft, which indicated a formation of soft tissue instead of bone. Six months post-operative there was no obvious change in bone regeneration visible when compared with the three months post-operative result. The patient has an excellent functioning of the bone-implant system without any restriction.

In conclusion, the first clinical use of this personalized titanium membrane for tumor reconstruction showed the feasibility of this new implant as well as its advantages towards the classical reconstruction techniques. The personalized shape of the membrane leads to a perfect fit of the membrane on the patient's bone. The use of a titanium membrane has a large intra-operative flexibility: the membrane is easily shaped to its final size and the screw holes for the fixation are easily punched during the surgery. Moreover, the membrane allows a good filling of the tumor cavity and holds the filling material in place; there is no risk of drift of the material. Finally, the membrane has a sufficient strength and stiffness to allow normal loading of the affected bone. These conclusions cannot be generalized, since they are based on one clinical trial. Further research on large numbers of reconstructions with a

personalized titanium membrane is needed to validate the general applicability as well as the clinical reliability.

6. Conclusions and Future Perspectives

The aim of this research was the design of a personalized implant with enhanced functionality for the reconstruction of weight-bearing bones after tumor resection surgery. First, the clinical problem as well as the different surgical treatments were discussed. Nevertheless, the surgeon has a wide variety of reconstruction techniques to his disposal; there are still some limits with respect to the restoration of the structural integrity and the function of the affected bone. To overcome the flaws of the current treatment methods, the use of a membrane, i.e. a thin preformed plate fitted on the periosteal surface of the bone and fixed with some small screws, is proposed. Titanium is chosen as a bio-compatible material, since it has the best behavior with respect to the material requirements and it allows radiographic follow-up. The design process as well as the manufacturing process of the titanium membrane were determined. The biomechanical fitting of the titanium membrane was performed for the specific case of a giant cell tumor in the adult proximal tibia. Experimental tests and finite element analyses were performed. These revealed the need for an 0.3 mm thick membrane (without wave pattern) for the reconstruction of a proximal tibial bone to assure withstanding of the loads occurring during daily activities. Finally, the titanium membrane was clinically used for the reconstruction of a distal femur after en-bloc resection of a juxta-cortical chondroma. This clinical application shows the feasibility of this new implant as well as its advantages towards the classical reconstruction methods.

- The personalized shape of the membrane leads to a perfect fit of the membrane on the periosteal surface of the patient's bone.
- The membrane is highly adaptable during surgery; the membrane is easily shaped to its final size and the screw holes for fixation are easily punched during the surgery.
- The membrane allows a good filling of the tumor cavity and assures the containment of the filling material.
- The bone reconstructed with a titanium membrane has a sufficient strength and stiffness to withstand the loads occurring during normal daily activities.

Personalized titanium membranes are a commercially attractive reconstruction method if the clinical outcome is equal or better than obtained with a classical reconstruction technique and the incidence of the anomaly in population is relatively high. Additionally a membrane must be low cost and swiftly to obtain.

In maxillo-facial and dental surgery personalized titanium membranes are already successfully applied (cranioplasty, bone augmentation,...). On the contrary, the conclusions presented for the tumor reconstruction of a weight-bearing bone are based on a single clinical application, and probably do not apply to any

case. A large clinical study is needed to confirm these conclusions in general and to validate the clinical reliability of a reconstruction with a personalized titanium membrane.

The scope of future research should be the expansion of titanium membranes such as reconstruction devices for bone to other applications in the field of orthopaedics, e.g. trauma surgery, maxillofacial surgery, ... The shortcomings of the currently used reconstruction techniques in the different medical fields must be evaluated to define the areas with the highest need of improvement, hence the areas in which the use of personalized implants is essential for an optimal post-operative functioning. For example, titanium membranes offer perspectives in the field of tissue engineering, a recently fast evoluting research that focuses on biological solutions for medical problems. For bone defects, regeneration of bone is the objective. Scaffolds, i.e. porous structures in which bone cells are embedded to start the growth of bone tissue are placed in the bone cavity. The mechanical parameters of the scaffold must be adapted to result in a load pattern of the bone cells that stimulates bone growth. Most currently available scaffolds are too weak to ensure the reconstruction of weight-bearing bones. Titanium membranes, as temporary fixation device, in combinations with scaffolds could be an elegant solution. A possible application for such a combination membrane-scaffold is the reconstruction of a large segmental bone defect in the metafysis of a long bone, i.e. regions of high shape complexity wherein standard equipment (plates, nails, cages, bone grafts,...) is difficult to position and would lead to stress shielding of the surrounding bone. Other applications for personalized titanium membranes are the shelling of the debris from a comminutive fracture in the proximity of a joint and reconstruction of the acetabulum in revision total hip arthroplasty.

The use of titanium membranes for tumor surgery on a regular basis is not yet possible due to the non-efficient design- and production process. First the design process of the membrane, i.e. creating the personalized shape, is a highly manual process due to the lack of computer aided design methods to reconstruct the cortical bone. Consecutive design steps such as data acquisition, filtering and reconstruction of contours are *in se* similar for different types of bones, and therefore an integrated design environment can be created. With the patient's anatomical data from the CT images, the environment automates the pre-operative planning and proposes a reconstruction to the surgeon while checking the feasibility of the implant placement. Expanding the application field of membranes to other types of bones implicates that new types of bone defects are introduced to the environment and therefore new procedures must be added to the environment. For example, a complete fracture of a long bone — compared to a tumor cavity — demands a registration of the bone pieces before reconstruction is initiated. The positioning of the fixation screws is another example. An outline of the possible screw placements on the membrane which can be calculated from thickness information in the CT images, can lower the insertion time during surgery, especially when connecting multiple bones (e.g. revision total hip arthroplasty).

Second, the production process needs a number of steps. A rapid prototyping model of the reconstructed bone is made, and afterwards reproduced in metal powder reinforced epoxyresin to be used as a mould. The hydroforming process is based on the experience of the machine operator; the more experienced the operator, the less trials are needed. The limits of the hydroforming process must be studied, since these determine the variation in producible shapes of membranes. If these are too limited further research is needed towards other manufacturing processes; if not, the manufacturing process should be optimized by e.g. using a rapid prototyping model as mould for the hydroforming process, and simulating the hydroforming in advance to reduce the number of trials.

The personalized titanium membranes designed in this study show major advantages for the reconstruction of load-bearing bones. Hence, further research to allow its commercialization and to expand their use in the medical field is of great value. Also the non load-bearing applications such as maxillofacial reconstructive surgery (e.g. cranioplasty) can benefit from an efficient design- and production environment. The latter will lower the threshold to actually choose for a membrane.

Acknowledgments

This research was performed in the Brite Euram project PISA (No. BRPR CT97-0378), partners in this project are: Materialise NV, Belgium (co-ordinator); Philips Medical Systems Nederland BV, Easy Vision Modules — Advanced Development, the Netherlands; DePuy, a Johnson & Johnson company, UK; Ceka NV, Belgium; Katholieke Universiteit Leuven, Laboratory for Medical Imaging Research and Division of Biomechanics and Engineering Design, Belgium; University of Leeds, Research School of Medicine, UK; OBL, France.

Research funded by a PhD grant of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

We are much indebted to Prof., M.D., G. Fabry, Head of the Department of Orthopaedic Surgery and to M.D., I. Samson member of the Department of Orthopaedic Surgery, K.U.Leuven, UZ Pellenberg, Weligerveld 1, 3212 Pellenberg, Belgium.

References

1. J. O. Johnston and R. Lor Randall, in *Current Diagnosis and Treatment in Orthopedics*, ed. H. B. Skinner (Mc Graw-Hill Publisher, New York, 2000), p. 247.
2. R. Coombs and G. Friedlaender, eds., *Bone Tumour Management* (Butterworths, London, 1987).
3. S. Gitelis and D. J. McDonald, in *Surgery for Bone and Soft-Tissue Tumors*, eds. M. A. Simon and D. Springfield (Lippincott-Raven Publishers, Philadelphia, 1998), p. 181.
4. S. Gitelis and D. J. McDonald, in *Surgery for Bone and Soft-Tissue Tumors*, eds. M. A. Simon and D. Springfield (Lippincott-Raven Publishers, Philadelphia, 1998), p. 133.

5. H. R. Blackley, J. S. Wunder, A. M. Davis, L. M. White, R. Kandell and R. S. Bell, *J. Bone Joint Surg. Am.* **81** (1999) 811.
6. S. Gitelis and D. J. McDonald, in *Surgery for Bone and Soft-Tissue Tumors*, eds. M. A. Simon and D. Springfield (Lippincott-Raven Publishers, Philadelphia, 1998), p. 159.
7. S. Gitelis and D. J. McDonald, in *Surgery for Bone and Soft-Tissue Tumors*, eds. M. A. Simon and D. Springfield (Lippincott-Raven Publishers, Philadelphia, 1998), p. 167.
8. B. K. Sanjay, *Indian J. Cancer* **36** (1999) 1.
9. F. F. Parish, *Clin. Orthop.* **373** (2000) 3.
10. S. Sijbrandij, *Acta Orthop. Scand.* **49** (1978) 249.
11. B. K. Sanjay, P. G. Moreau and D. A. Younge, *Int. Orthop.* **21** (1997) 291.
12. W. F. Enneking, *Clin. Orthop.* **204** (1986) 9.
13. R. E. Wolf and W. F. Enneking, *Orthop. Clin. North Am.* **27** (1996) 473.
14. H. Tsuchiya, K. Tomita, K. Minematsu, Y. Mori, N. Asada and S. Kitano, *J. Bone Joint Surg. Br.* **79** (1997) 403.
15. F. Watzinger, J. Luksch, W. Millesi, C. Schopper, J. Neugebauer, D. Moser and R. Ewers, *Br. J. Oral Maxillofac. Surg.* **38** (2000) 312.
16. R. Celleti, M. Davarpanah, D. Etienne, G. Pecora, J.-F. Tecucianu, D. Djukanovic and K. Donath, *Int. J. Periodontics Restorative Dent.* **14** (1994) 242.
17. D. van Steenberghe, C. Johansson, M. Quirynen, L. Molly, T. Albrektsson and I. Naert, *Clin. Oral Impl. Res.* **14** (2003) 63.
18. L. C. Hieu, E. Bohez, J. Vander Sloten, H. N. Phien, E. Vatcharaporn, B. H. Binh, P. V. An and P. Oris, *Rapid Prototyping Journal* **9**, 3 (2003) 175.
19. J. K. Th. Haex, H. W. C. Bijvoet and A. H. G. Dallenga, *Phidias Newsletter* **1** (1998) 4.
20. V. Pattijn, *Biomechanical Design of Medical Image-Based Personalized Implants* (Katholieke Universiteit Leuven, 2004).
21. G. Jans, *A CAD-Based Preoperative Planning Environment for Cranial Surgery* (Katholieke Universiteit Leuven, 2003).
22. A. Safari, S. C. Danforth, M. Allahverdi and N. Venkataraman, in *Encyclopedia of Materials: Science and Technology*, eds. K. H. J. Buschow, R. W. Cahn, M. C. Flemings, B. Ilschner, E. J. Kramer and S. Mahajan (Elsevier Science Ltd., Amsterdam, 2001), p. 7991.
23. J. Datsko, *Material Properties and Manufacturing Processes* (John Wiley & Sons, Inc., New York, 1966).
24. F. Vollertsen, *J. Mater. Process Tech.* **103** (2000) 424.
25. A. H. Ekholy and O. M. Al-Hawaj, *Int. J. Adv. Manuf. Tech.* **18** (2001) 79.
26. J. B. Morrison, *J. Biomech.* **3** (1970) 51.
27. K. N. An, E. Y. S. Chao and K. R. Kaufman, in *Basic Orthopaedic Biomechanics*, eds. V. C. Mow and W. C. Hayes (Raven Press, New York, 1991), p. 1.
28. N. Zheng, G. S. Fleisig, R. F. Escamilla and S. W. Barrentine, *J. Biomech.* **31** (1998) 963.
29. Y. H. An and R. A. Draughn, eds., *Mechanical Testing of Bone and the Bone-Implant Interface* (CRC Press LLC, Florida, 2000).
30. S. Saha and S. Pal, *J. Biomed. Mater. Res.* **18** (1984) 435.

CHAPTER 3

TECHNIQUES FOR TRACHEAL SEGMENTATION IN MEDICAL IMAGING

RAQUEL VALDÉS-CRISTERNA*, OSCAR YÁÑEZ-SUÁREZ
and VERÓNICA MEDINA-BAÑUELOS
Neuroimaging Laboratory
Department of Electrical Engineering
Universidad Autónoma Metropolitana – Iztapalapa
San Rafael Atlixco 186, Col. Vicentina, 09340 Mexico City, Mexico
ravc@xanum.uam.mx; Tel.: (52) 55-5804-4903 ext. 204/224

The evaluation of the upper respiratory airway obstruction is of significant importance, as its physiological effects on ventilation rapidly evolve to secondary body malfunctions. Given the nature and location of the lesions, the invasive handling of the condition with direct bronchoscopy is undesirable; an imaging approach based on computed tomography is preferred. Objective and quantitative evaluation of the obstruction therefore requires the application of image processing and analysis techniques. The majority of research conducted for airway segmentation and analysis from tomographic views has relied on region-based procedures. However, the specific problem of automatic tracheal segmentation can be approached with a continuous edge detection perspective, taking advantage of the high contrast that exists between the interior of the airway and its surrounding tissue. This chapter reviews the state-of-the-art procedures for respiratory airway segmentation and describes an active contour-based approach for this task.

Keywords: Tracheal segmentation; active contours; active surfaces; cubic splines; stenosis; CT images.

1. Human Airway Assessment from Medical Images

The respiratory system is of vital importance for human beings, given that it performs, among others, the ventilation function. Ventilation corresponds to the fundamental physiological process of gas exchange between the external medium surrounding the subject and his/her own internal medium, which is aimed at the maintenance of the acid-base equilibrium. This is why the signs or symptoms of an airway obstruction deserve immediate attention to prevent generalized secondary body malfunctions. According to published World Health Organization statistics, respiratory diseases and infections correspond to mortality indexes of 6.5% and 6.7% respectively[41]; from here the importance of accomplishing an early and precise diagnosis of the aforementioned affections. In the particular case of mild or moderate tracheal stenosis, it is not uncommon to confuse its associated signs and symptoms

*Corresponding author.

with those related to other kinds of respiratory complications such as asthma or chronic bronchitis. It is for this reason that an accurate and objective detection becomes necessary to decrease the chances of false diagnosis, thus promoting the selection of an adequate treatment on time.

1.1. *Airway anatomy and physiology*

Structurally, the respiratory apparatus is conformed of a connected, continuous series of conducts spanning the head, neck and thorax. These pipes provide access to physical paths along which the atmospheric air permeates to the blood stream, in a process named *gas exchange*. According to their body location, the respiratory airways are classified in two large-scale categories:[17]

- Upper Airways, comprising the nose, nasal cavity, paranasal sinus, nasopharynx, pharynx.
- Lower Airways, including the trachea, tracheobronchial system and lungs.

The geometry of the human airway tree follows a bipodial branching, where a parent pipe divides into two or three child branches. This pattern has important effects on the distribution of air flow in the lung. The child branches' diameter gradually decreases to reach approximately 1 to 2 mm at the bronchi level.[49] The trachea is itself divided in two distinct portions: (i) an extra-thoracic or proximal component, extending from 2 to 4 cm in length, beginning from the lower border of the cricoid cartilage and down to the thoracic entrance, projecting from 1 to 3 cm above the suprasternal nodule, and (ii) an intra-thoracic or distal extension that corresponds to the next 6 to 9 cm, reaching a total length of 10 to 13 cm. The cross-sectional diameter of the trachea in men lies between 13 and 25 mm, while in women this diameter varies from 10 to 21 mm. Figure 1 shows a schematic depiction of the respiratory airway structure, emphasizing the trachea and main bronchi.

From a transversal perspective, the internal layer of the trachea is mainly a covering epithelium (mucous), followed by a layer of connective tissue, where the tracheal glands are immersed; the next tissue layer is formed by a *C*-shaped cartilage (tracheal cartilage). The posterior open ends are complemented with smooth muscle. Tracheal distensibility and elasticity help in supporting the heart, because the upper bronchi or tracheal branches are interlaced with the pulmonary arteries that connect with it.[17]

1.2. *General airway pathology and tracheal lesions*

The incidence of laryngeal and tracheal obstructive lesions has increased in recent years; on one hand, it is observed that central airway lesions have become more common mainly due to traffic accidents, and on the other, the interventional use of inflatable tracheal probes and controlled mechanical ventilation have originated new kinds of airway trauma at the mucous and *C* rings. Tobacco consumption and

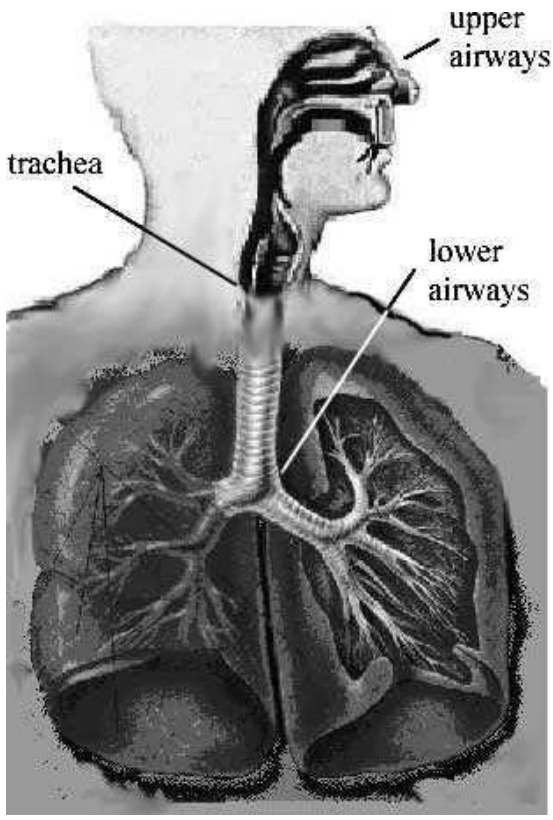


Fig. 1. Schematic display of the prominent anatomical components of the respiratory airway system.

air pollution have also lead to a rise in the incidence of upper airway carcinoma.[16,33] Miller suggests the following classification for tracheal conditions that are potentially obstructive:[16]

- developmental alterations, comprising vascular anomalies, congenital glottic stenosis or angiomas;
- infection;
- trauma-related, which represent the most common case;
- presence of foreign objects, with a large incidence in children; and
- neoplasy, whose symptoms might be confused with those of diffuse obstructive disease and treated as asthma.

Other idiopathic respiratory ailments exist that could present obstructive-like symptoms, such as osteoplastic tracheopathy (related to cartilaginous projections within the tracheal lumen), relapsing polychondritis (associated with rheumatoid arthritis), or lupus erythtematosus.[16]

Tracheal stenosis is defined as a reduction of the transversal diameter of the trachea.[51] Most cases of stenosis are developed below the sternal vertex, and thus cannot be observed in neck radiographs. In many cases, bronchoscopy represents the most appropriate assessment method for determining the location and extent of the lesion; however, the procedure requires anesthesia and, in severe cases, further airway obstruction can be induced, requiring emergency procedures.[33] There are limitations in the detection of tracheal stenosis through the application of functional tests, given that spirometry is basically insensitive to mild or moderate cases. The ratio of forced expiratory flow at 50% of vital capacity versus forced inspiratory flow at the same level has been used as a characteristic index in chronic obstruction of the upper airways, although its confidence has been questioned. The flow-volume plot is useless for obstructions below 8 mm, and fails in obstruction detection even in the presence of clear radiological evidence.[61]

The clinical signs of tracheal stenosis (like strident respiratory sounds and shortness of breathing) can hide the etiology of the condition. Chronic bronchitis and asthma have signs that appear within the stenotic signs. Symptom-free, resting cases of mild or moderate stenosis are hard to identify, and will be approached in most cases as if they were diffuse pulmonary illnesses or chronic bronchitis. In these cases, bronchoscopy and computed tomography imaging are useful diagnostic procedures. In cases of surgical therapy, the technique must be adjusted to the localization and extension of the lesion.[33,51] Acoustic analysis of stridency and acoustic rhinometry[11] have also been used in an attempt to measure airway volumes and to locate an obstruction and estimate its extent. Such a procedure has been proven sensitive and useful along the respiratory cycle, while being non-invasive and of fast realization; however, in the evaluation of severe stenosis, a 20% over-estimation of the stenotic area has been observed.[61]

1.3. *Image-based airway exploration techniques*

The use of lateral neck radiographs has been proved useful in the detection of upper tracheal deformities, but intrathoracic stenosis cannot be observed in this modality. Antero-posterior neck exposures during the emission of the "i" sound usually deliver an excellent detail of the trachea up to the carina. However, only 60% of the tracheal stenoses are detected through conventional X-ray imaging. These images suffer from structural superposition (occlusion), rendering 3D analysis of the airway tree difficult to achieve. Linear tomography might be useful to estimate the extension of the lesion although its spatial resolution is poor. These problems are overcome with computed tomography.[43]

X-ray tomographic images allow the detailed visualization of the contrast between the interior of the airways and the bronchial or tracheal wall, along with the rest of the surrounding airway tissue. Because of this, it is possible to estimate, with significant precision, not only the geometry of the pipe section, but the thickness and integrity of the walls and how this condition influences the gas

exchange process. Additionally, if this imaging technique is combined with dynamic breathing maneuvers such as Müller and Valsalva, the respiratory compromise under maximal efforts can be evaluated. The X-ray tomographic techniques most frequently used in obstruction assessment include high-resolution, helicoidal and micro-tomography.[13,15,18,21,24,32,38,39,44,48,54,64,65]

In a clinical report, LoCicero *et al.* show the results, on eleven patients, of a multi-planar study with helicoidal tomography and the associated 3D reconstruction. They describe the usefulness of the technique and state that different viewpoints of the tracheobronchial tree can be delivered.[31] Mogavero *et al.* conclude on the excellent anatomical definition quality obtained form such as 3D reconstruction, when observing two cases of central airway pathologies, a tumor in the carina and a post-surgical bronchial stenosis. Their procedure facilitated the confirmation of previous diagnosis and the design of the therapeutic approach for the patients.[37] Similar findings have been reported by Sagy *et al.* for five cases of intrathoracic airway obstruction in children, employing, once again, helicoidal tomography and 3D reconstruction of the airway.[47]

Recently, magnetic resonance imaging has been applied to obtain mono- and multi-spectral image stacks with contrast agents for the assessment of the airway condition.[1,2,29,36,63] In particular, this technique provides excellent information from the surrounding mediastinum, but still remains second to X-ray tomography in current clinical applications. Software systems have been introduced for interactive X-ray tomography analysis for airway identification and measurement.[12,56]

In contrast with all the above, there is little doubt that the technique providing the most diagnostic certainty for airway occlusion cases is direct bronchoscopy. Nonetheless, it is stressed here that its invasive condition represents a major defect, yielding a preference for image-based approaches.

2. Airway Segmentation Techniques

Image analysis is one of the most relevant applications of digital image processing. Usually, the goal of the analysis is the decomposition of the observed scene in basic information elements, well identified or classified, that would allow the construction of assertions on the studied phenomenon. In the case of medical images, the particular objective is to provide the medical specialist with information support for diagnosis or prognosis of a given disease. Respiratory illnesses as those discussed herein are no exception: bronchitis, emphysema, presence of foreign objects, stenosis, sleep apnea and asthma have been studied using image processing techniques.[21,44,48,59]

Several methods have been designed and developed with the intention of producing a realistic representation of the airway tree for diverse applications. Some of them follow an automatic or semi-automatic approach for the final tree reconstruction, based solely in the raw image information; in other cases, processed information from the images has been combined with the use of mathematical models that allow the refinement of the anatomical reconstruction of the airway structure

and the formulation of functional physiological models of air conduction through the system.[18] For instance, Summers and Cebral show that virtual bronchoscopy provides useful data to perform aerodynamic calculations in anatomical models.[52]

A frequent application in the field of respiratory imaging is virtual airway navigation, such as virtual bronchoscopy,[25,53] which is more and more required to assist the specialist with non-invasive diagnosis, or in the training of future specialists in the operation and interpretation of the bronchoscope and its images. Along this line, research efforts are focused both to the achievement of a more realistic 3D representation and to the definition of soft navigation trajectories, related to the construction of the tracheobronchial tree axis.[26,42] Kiraly and Higgins present a robust method to compute the central axes of a branching tubular structure, which is based on a detailed multi-stage refinement of a computed skeleton of a pre-segmented 3D image. Geometric properties of an assumed tubular model are used in the refinement.[26] Higgins *et al.* propose the registration of real images from a bronchoscope with those obtained from virtual bronchoscopy, for an *augmented reality* scenario that would be useful in providing views of hardly observed perspectives.[21] Three-dimensional reconstruction of the airway tree is also useful in the morphological analysis of a particular obstruction at the time of diagnosis. Moreover, the possibility of a time-course evaluation of the progress of the lesion in a specific subject, or the comparison of measures among a population, provides the medical community with better elements for the research in respiratory disease and population analysis.[42,59] In the specific case of tracheal stenosis, lesion characterization is as relevant as its 3D representation. The construction of longitudinal profiles of the airway transverse area leads to the elaboration of airflow models, as has been applied in the evaluation of the obstructive sleep apnea syndrome.[63]

For the development of the aforementioned applications, image segmentation is a fundamental component. Image segmentation refers to the algorithmic procedure through which an image under study is divided in disjoint regions that uniquely respond to a pre-established criterion, measure or property. If such a measure of aggregation is selected without the intervention of an expert or operator, the segmentation approach is considered *automatic*; if any of the parameters of the process requires the specialist's experience to be defined, the segmentation is carried out in a semi-automatic or *supervised* way.[5,20,46] The airway tree can be extracted from computed tomography images using manual identification; however interactive analysis is difficult due to time requirements and so, development of automated approaches is necessary to achieve clinical utility of quantitative airway studies.[43] In some cases, manual segmentation by experts has been performed, only with the purpose of validating semi-automatic or automatic segmentation algorithms.[3,13,59,65]

Segmentation methods, in general, can be classified into two major groups: region-based or edge-based. While the former has been heavily used in airway imaging applications, edge-based methods are feasible and convenient in specific cases such as tracheal segmentation. The following paragraphs briefly review both perspectives.

2.1. *Region-based segmentation*

This segmentation approach basically consists of the search of homogeneous characteristics of neighboring image pixels, so that meaningful aggregations can be constructed. The property used for differentiating one image region from another directly determines the specific properties of the segmentation; typical property choices are statistical, intensity-based, textural, etc.[20,46]

Region segmentation applications to thoracic computed tomography (CT) for airway detection have been carried out through gray level thresholding, as described in reports by Sagy *et al.*[47] and Mogavero *et al.*[37] Both present clinical cases evaluated with helicoidal CT aimed to 3D airway reconstruction. Segmentation of the regions of interest is performed with the adjustment of a user-selectable threshold over the gray levels, combined with a connectivity criterion. Niethammer *et al.*[40] introduced a system for the segmentation of pulmonary parenchyma based also on a thresholding criterion, established for lung contour definition and parenchymal vessel removal. In their study, the segmented images are superimposed with a perfusion image for combined analysis. Throughout the more recent reports on CT-based airway segmentation,[8,28,43,53,62] an intensive use of automatic, 3D region-growing techniques for region-based segmentation is clearly observed.

Park *et al.*[43] propose a fuzzy logic approach for individual pixel labeling on preprocessed individual CT slices. Primary paths, either corresponding to airways or vessels, are first segmented from individual CT slices by performing intensity-based region-growing and top-hat filtering. Candidate pixels are then classified, as forming or not part of an airway, with the evaluation of fuzzy features such as spatial adjacency, degree of presence of a nearby wall around the path and local gray level intensity. Finally, the airway tree is constructed from the binarized stack through shape interpolation along the z-axis. On a related paper, Liu *et al.* use the fuzzy connectedness technique to segment upper airways and their surrounding tissues from magnetic resonance image stacks.[29]

Summers *et al.*[53,55] describe a segmentation process for helicoidal CT in the context of virtual bronchoscopy, based on a 3D region-growing scheme with fixed threshold and an additional criterion of distance to the region seed; this last consideration allows their method to avoid the wrong segmentation of parenchymal voxels located around the bronchial ramification zones. In their clinical evaluation over fourteen subjects, a comparison of airway diameter measures, both from the virtual model and from direct measurement on the CT images, showed an average difference of around 0.5 mm. In the paper by Wood *et al.*,[62] the determination of the central axis of the airway is performed by iterative region-growing, where the distance to the seed is also utilized: the centroid of the voxels that have the same distance to the region seed is computed as an estimate of the central axis location at different scales. Linear adjustment of estimated centroid locations is then used to refine the position. After obtaining the complete model of the bronchial tree, the authors could perform diverse measures of branch length, ramification angle and pipe diameters; estimation errors in the order of 2 mm or 5°, respectively

were obtained. It is suggested that the vascular tree could also be segmented, by adequately adjusting the region-growing algorithm parameters.

Chiplunkar *et al.*[8] also describe an adaptive variation of the 3D region-growing scheme. As the individual selection of a region seed for each bidimensional slice had proven effective for better segmentation, their method was aimed at the substitution of manual seed selection by an adaptive approach. The growth threshold is adjusted after consideration of local image information and the knowledge of gray level relationships among voxels located at the airway wall, the pulmonary parenchyma, and air zones. Further, information from previous slices is considered for the segmentation of a new slice along the stack, by computing an initial threshold estimate from local gray level information around the estimated tracheal centroid. Seed localization at every slice is also obtained adaptively from previous estimates along the stack. This segmentation method was tested on high-resolution CT images, and validated against the manual selection of seeds and growth thresholds. Results showed that the automatic segmentation procedure generally underestimated the cross-section area with respect to the expert's estimate, while the centroid localization error, down to the fourth generation of branches, was established at around 0.75 mm.

Two powerful reasons for the extended adoption of the region-growing segmentation approach are its algorithmic simplicity and its yield of continuous airway borders for area profile evaluations. However, two underlying problematic issues could also be mentioned. First, the methods remain sensitive to image noise or structural detail that may induce the loss of pixel/voxel contiguity. Second, most implementations consume significant amounts of computational resources, which eventually becomes a practical limitation for clinical usage. Such inefficiency has been reported in a comparison between the region-growing technique and an edge detection approach using a derivative of Gaussian filter in thoracic CT.[61] The results from both methods were compared against an expert's selection of airway borders and no statistically significant differences were observed.

2.2. *Edge-based segmentation*

Segmentation through edge detection has the purpose of defining the boundaries among different anatomical structures. Various edge segmentation procedures are based on the convolution of the original image with a local kernel with specific properties. Popular kernel choices are those based on the magnitude of the gradient operator (Sobel, Prewitt, Roberts), or those using a Laplacian or derivative of Gaussian kernel.[6,35] Other methods further incorporate *a priori* information from image models to enhance edge localization.

One particularity of the filtering approaches to edge segmentation is that the contours produced as output are not necessarily continuous, a condition that is highly desirable when estimating the cross sectional area profiles for tracheal obstruction evaluation. As an alternative, the method of edge segmentation through

active contours (snakes) generates a continuous edge, out of diverse image characteristics that are combined with a pre-established dynamical model.[23]

2.3. *Hybrid methods*

Another perspective in airway characterization, that combines techniques as those previously described with the analysis of gray level profiles along a specific image line (halfmax) in high-resolution CT images has been introduced by Reinhardt, Hoffman, and Sonka.[22,45,50] Precise intra-thoracic airway measurements are achieved by modeling the scanner response and, with the aid of a reference airway model with circular cross sections, iteratively optimizing the estimates of the real airway geometry. With this approach, the authors have established that the scanner spread function has a biasing effect over the geometric estimates in airways with small cross section. The halfmax approach assumes that the image plane is perpendicular to the central axis of the airway, a condition that is not always satisfied. Segmentation of the major airways is carried out by region-growing, and extended with a rule-based scheme for the secondary tree. For the latter, a pixel-based feature vector is defined, including gray level, distance to the airway wall, and an adjacency measure. Regions are labeled as airway or non-airway by thresholding the confidence values of a fuzzy classifier. The 3D airway tree is finally constructed by the union of all the airway-labeled regions that are connected within a 3D neighborhood of the primary tree voxels. Refinements to the approach, presented by Chabat *et al.*,[7] extend the reference airway model to elliptical sections, which improves the geometric descriptions and supports a strategy for automated bronchi detection.

Aykcac *et al.*[3] have introduced a full-tree segmentation procedure based on morphological operations aimed at identifying candidate airway structures, that are later combined with transitive closures over an adjacency graph. Pretreux *et al.* provide another mathematical morphology approach for the estimation of bronchial caliber in high-resolution CT,[44] while Fan *et al.* take a topological-morphological strategy for tree reconstruction.[14] A constructive model perspective is suggested by Brown *et al.*[4,5] incorporating *a priori* knowledge of the anatomical structures of the airway tree. Analyzed portions of a thoracic CT image (described in terms of gray level, volume, shape, and location) are confronted with model primitives using an inference machine based on a semantic network. *Ad hoc* image preprocessing is used for feature extraction, a condition which heavily impacts the efficiency of the semantic segmentation. Alternatively, Law *et al.* take advantage of a genetic algorithm for optimizing a region growing segmenter.[27] Finally, Sorantin *et al.*[51] present an assessment of tracheal stenosis in 3D images, where segmentation of the laryngo-tracheal tract is carried out using fuzzy connectedness and a 3D dilation using a small 3D structuring element, with further manual refinement of the tract contour, if required. The 3D morphological skeleton of the tract is obtained through topology-preserving thinning, and used as a reference to compute a cross-sectional area profile along its extension. A comparison between the localization of stenosis

with the proposed method versus the diagnosis with direct endoscopy indicated a match in 92% of the studied cases.

3. Trachea Segmentation with Active Models

Even though most of the reports on airway segmentation utilize the region-based paradigm, the problem of automatic tracheal segmentation from CT images can be approached with a continuous edge detection perspective, taking advantage of the high contrast that exists between the interior of the airway and its surrounding tissue. For example, Swift *et al.* utilize the active contour method to determine the airway boundary for central axis estimation, a piece of information that is necessary for the definition of navigation trajectories in virtual endoscopy. These authors report favorable results from their methodology, when applied in normal and stenotic cases, stressing the speed and efficiency of their approach in contrast to other methods such as halfmax.[57,58]

The following section introduces a tracheal segmentation method, constructed over active contour or surface models based on a cubic splines interpolation. Such a model conveys an analytical description for the entire extent of the contour or surface, imposing smoothness and continuity characteristics on the adjusted solution. 3D rendering of the segmented airway path from neck and thorax CT scans using the proposed method is validated in regard to its possible use as a diagnostic tool for the characterization of tracheal stenosis. The details related to the validation of the method are presented elsewhere.[61]

3.1. *Active contour model for trachea segmentation*

In the original active contour model of Kass *et al.*,[23] the discrete formulation of the energy functional is given by the sum of internal and external energy terms for the N contour segments as

$$E^*_{snake} = \sum_{i=1}^{N} \left(E_{\text{int}}(i) + E_{\text{ext}}(i)\right)$$

In agreement with this model, it is shown in the following that an analytical formulation based on natural cubic splines interpolation is possible, and that a closed-form expression for the internal energy term can thus be derived, avoiding the classical use of finite difference approximations. The active contour or snake energy term is:

$$\begin{aligned} E_{snake}(v) &= E_{\text{int}}(v) + E_{\text{ext}}(v) \\ E_{\text{int}}(v) &= \sum_{i=1}^{N} \alpha_i(s)\left|v'_{Si}(s)\right|^2 + \sum_{i=1}^{N} \beta_i(s)\left|v''_{Si}(s)\right|^2 \\ E_{\text{ext}}(v) &= -\gamma\|G * I(v)\| \end{aligned} \tag{1}$$

where:

v	discrete contour represented by a set of control points
N	number of control points
$v_{Si}(s)$	contour segment between the ith and $(i+1)th$ control points, parametrized on s
$\alpha(s)$, $\beta(s)$	weights of internal energy term
G	bidimensional derivative of Gaussian with fixed parameters μ and Σ
$I(v)$	image intensity values evaluated on the contour v
γ	weight of external component of energy
$*$	bidimensional convolution operator.

Every segment $v_{Si}(s)$ can be interpolated with cubic splines, over a subset $\{V_{m-1}, \ldots, V_{m+2}\}$ of control points as:

$$v_{Sm}(s) = \frac{1}{6}\begin{bmatrix} s^3 & s^2 & s & 1 \end{bmatrix} \times \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -2 & -3 & 6 & -1 \\ 0 & 6 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} V_{m-1} \\ V_m \\ V_{m+1} \\ V_{m+2} \end{bmatrix} \tag{2}$$

or equivalently,

$$v_{Sm}(s) = \frac{1}{6}\mathbf{s} \times \mathbf{M} \times \mathbf{v}_m \,. \tag{3}$$

From Eq. (3), the first and second derivatives of the contour with respect to parameter s can be computed as:

$$\begin{aligned} v'_{Sm}(s) &= \frac{1}{6}\begin{bmatrix} 3s^2 & 2s & 1 & 0 \end{bmatrix} \times \mathbf{M} \times \mathbf{v}_m \\ v''_{Sm}(s) &= \frac{1}{6}\begin{bmatrix} 6s & 2 & 0 & 0 \end{bmatrix} \times \mathbf{M} \times \mathbf{v}_m \,. \end{aligned} \tag{4}$$

Using Eq. (4) in (1), it follows that

$$\begin{aligned} E_{snake}(v) = \sum_{i=1}^{N} \Bigg(& \alpha_i(s) \left| \frac{1}{6}\begin{bmatrix} 3s^2 & 2s & 1 & 0 \end{bmatrix} \times \mathbf{M} \times \mathbf{v}_i \right|^2 \\ & + \beta_i(s) \left| \frac{1}{6}\begin{bmatrix} 6s & 2 & 0 & 0 \end{bmatrix} \times \mathbf{M} \times \mathbf{v}_i \right|^2 \Bigg) - \gamma \| G * I(v) \| \,. \end{aligned} \tag{5}$$

The derivative of E_{snake} (5) with respect to a point V_k on the contour, evaluated at the start of every segment ($s = 0$), is:

$$\frac{\partial E_{snake}}{\partial V_k} = \frac{\partial E_{\text{int}}(v)}{\partial V_k} + \frac{\partial E_{\text{ext}}(v)}{\partial V_k} = \frac{\partial E_{\text{int}}(v)}{\partial V_k} - \gamma F(v)$$

where (subscripts for α and β omitted for simplicity):

$$\begin{aligned}\frac{\partial E_{\text{int}}(v)}{\partial V_k} &= \frac{1}{3}V_{k-3}(2\alpha) + \frac{1}{3}V_{k-2}(-9\alpha+36\beta) + \frac{1}{3}V_{k-1}(-18\alpha-144\beta)\\ &+ \frac{1}{3}V_k(50\alpha+216\beta) + \frac{1}{3}V_{k+1}(-18\alpha-144\beta)\\ &+ \frac{1}{3}V_{k+2}(-9\alpha+36\beta) + \frac{1}{3}V_{k+3}(2\alpha)\end{aligned} \tag{6}$$

and $F(v)$ represents the derivative of the external energy component ($\|G * I(v)\|$), which is the derivative of the image gradient magnitude calculated with a Canny filter with fixed parameters μ and Σ.[6]

If the internal energy parameters α and β in Eq. (6) are held constant, it is possible to calculate a matrix $\mathbf{A}$ that includes the simultaneous adjustment of the whole control point set. The expression for the gradient of the objective function (5) would then be:

$$\nabla E_{snake} = \mathbf{A}\underline{v} - \gamma F \tag{7}$$

where $\underline{v}$ is a vector containing the set of control points, F is the vector of the external force evaluated in $\underline{v}$, and the $\mathbf{A}$ matrix format is:

$$\mathbf{A} = \begin{bmatrix} 50\alpha+216\beta & -18\alpha-144\beta & -9\alpha+36\beta & 2\alpha & 0 & \cdots \\ -18\alpha-144\beta & 50\alpha+216\beta & -18\alpha-144\beta & -9\alpha+36\beta & 2\alpha & 0 \\ \vdots & & & \ddots & & \\ -18\alpha-144\beta & -9\alpha+36\beta & 2\alpha & 0 & \cdots & 0 \end{bmatrix}$$
$$\begin{bmatrix} 0 & 2\alpha & -9\alpha+36\beta & -18\alpha-144\beta \\ \cdots & 0 & 2\alpha & -9\alpha+36\beta \\ & & & \vdots \\ 2\alpha & -9\alpha+36\beta & -18\alpha-144\beta & 50\alpha+216\beta \end{bmatrix}. \tag{8}$$

Given that (8) is a singular matrix,[9] the minimization procedure for (5) has to be implemented as an iterative search for the optimal contour. With this purpose, a normalized gradient descent optimization is used. The final expression for the iterative procedure is:

$$\begin{aligned}\underline{v}^{(t+1)} &= \underline{v}^{(t)} + \mu\frac{(-\nabla E_{snake})}{\|\nabla E_{snake}\|}\\ \underline{v}^{(t+1)} &= \underline{v}^{(t)} - \mu\frac{\mathbf{A}\underline{v}^{(t)} - \gamma F(\underline{v}^{(t)})}{\|\mathbf{A}\underline{v}^{(t)} - \gamma F(\underline{v}^{(t)})\|}\end{aligned} \tag{9}$$

where μ represents the gradient descent step.

In summary, the trachea segmentation procedure for CT scans described implements the following algorithm:

```
Perform contour initialization (see section 3.3 below)
While contour variation is greater than a given tolerance level
      Perform contour interpolation based on cubic splines and
         resample to select N new control points
      With the new set of control points, evaluate (7)
      Perform an iteration step, evaluating (9)
      Calculate the contour variation from previous iteration
Perform a final contour interpolation
```

3.2. *Extension to active surfaces*

A general formulation for surface interpolation using a natural cubic splines basis, can be expressed in the following way:

$$V_{ij}(s,r) = \sum_{m=-1}^{2} \sum_{l=-1}^{2} P_{i+m,j+l} B_{ml}(s,r) \tag{10}$$

where $P_{i+m,j+l}$ is a control point and $B_{ml}(s,r)$ is a bi-dimensional spline basis function.

Expression (10) indicates that patch interpolation requires a 4×4 mesh of control points, as shown in Fig. 2.

The bi-dimensional basis functions are built through the tensorial product of the corresponding one-dimensional basis function, that is

$$B_{ml}(s,r) = b_m(s) b_l(r)\,.$$

The one-dimensional interpolation (3) was described as a matrix product between the parametric vector ($\mathbf{s}$), the coefficients matrix ($\mathbf{M}$) and an array of four neighboring control points $\mathbf{v}_m$, incorporating the cubic splines basis. For the surface model (10), the bidimensional basis functions are built through the tensorial product of the corresponding one-dimensional basis functions as $B_{ml}(s,r) = b_m(s)b_l(r)$. Thus, it is possible to implement the cubic splines interpolation procedure as a variation of a Kronecker product of the vectors described in (3). Let $v_i(s)$ and $v_j(r)$ be the analytical forms of the interpolation of the segments i and j in a corresponding contour, then:

$$\begin{aligned} v_i(s) &= \begin{bmatrix} s^3 & s^2 & s & 1 \end{bmatrix} \times \mathbf{M} \times \begin{bmatrix} P_{i-1} \\ P_i \\ P_{i+1} \\ P_{i+2} \end{bmatrix} = \mathbf{s} \times \mathbf{M} \times \mathbf{P}_i \\ v_j(r) &= \begin{bmatrix} r^3 & r^2 & r & 1 \end{bmatrix} \times \mathbf{M} \times \begin{bmatrix} P_{j-1} \\ P_j \\ P_{j+1} \\ P_{j+2} \end{bmatrix} = \mathbf{r} \times \mathbf{M} \times \mathbf{P}_j \end{aligned} \tag{11}$$

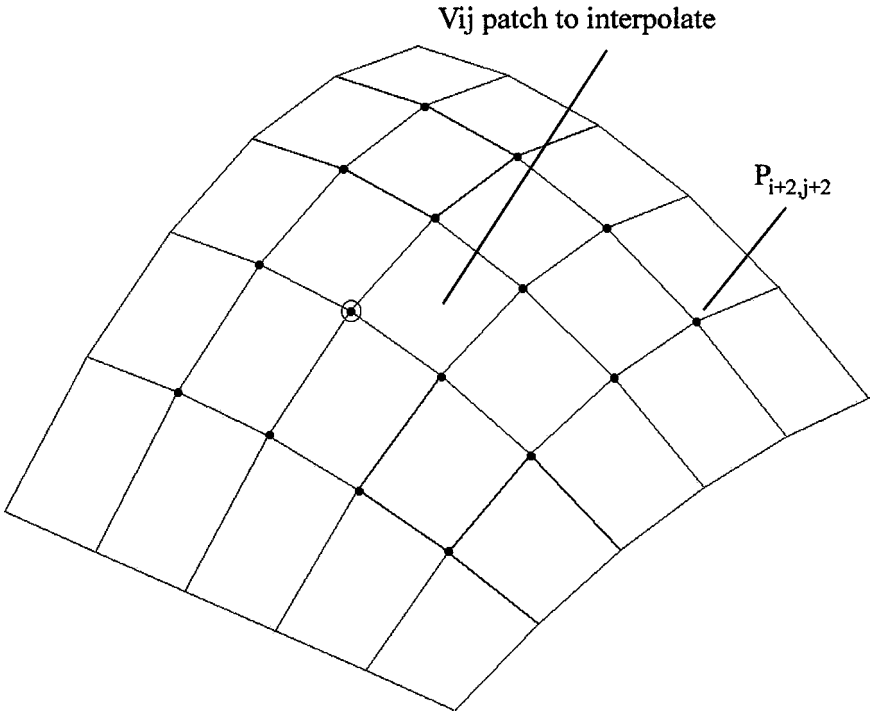


Fig. 2. Control points for active surface interpolation.

evaluating the Kronecker product between $v_i(s)$ and $v_j(r)$, we have:

$$\begin{aligned} v_i(s)^T \otimes v_j(r) &= [\mathbf{s} \times \mathbf{M} \times \mathbf{P}_i]^T \otimes [r \times \mathbf{M} \times \mathbf{P}_j] \\ &= \left[\mathbf{P}_i^T \times (\mathbf{s} \times \mathbf{M})^T\right] \otimes [r \times \mathbf{M} \times \mathbf{P}_j] \\ &= \left[\mathbf{P}_i^T \times \mathbf{M}^T \times \mathbf{s}^T\right] \otimes [r \times \mathbf{M} \times \mathbf{P}_j]\,. \end{aligned} \tag{12}$$

Then, the final expression for the surface interpolation is:

$$\begin{aligned} V_{ij}(s,r) &= \sum_{m=-1}^{2} \sum_{l=-1}^{2} P_{i+m,j+l} B_{ml}(s,r) \\ &= [\mathbf{M}^T \times [\mathbf{s}^T \otimes \mathbf{r}] \times \mathbf{M}]_{row} \times [\mathbf{P}_{ij}]_{column} \\ &= [\mathbf{M}^T \mathbf{S}_R \mathbf{M}]_{row} \times [\mathbf{P}_{ij}]_{column} \\ &= \left[\frac{1}{6}\begin{bmatrix} -1 & 3 & -2 & 0 \\ 3 & -6 & -3 & 6 \\ -3 & 3 & 6 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix}\right] \times \begin{bmatrix} s^3r^3 & s^3r^2 & s^3r & s^3 \\ s^2r^3 & s^2r^2 & s^2r & s^2 \\ sr^3 & sr^2 & sr & s \\ r^3 & r^2 & r & 1 \end{bmatrix} \end{aligned} \tag{13}$$

$$\times \frac{1}{6}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -2 & -3 & 6 & -1 \\ 0 & 0 & 6 & 0 \end{bmatrix}_{row}$$
$$\times \begin{bmatrix} P_{i-1,j-1} & P_{i-1,j} & P_{i-1,j+1} & P_{i-1,j+2} \\ P_{i,j-1} & P_{i,j} & P_{i,j+1} & P_{i,j+2} \\ P_{i+1,j-1} & P_{i+1,j} & P_{i+1,j+1} & P_{i+1,j+2} \\ P_{i+2,j-1} & P_{i+2,j} & P_{i+2,j+1} & P_{i+2,j+2} \end{bmatrix}_{column}$$

where:

$V_{ij}(s,r)$	patch to interpolate
$\mathbf{M}$	interpolation coefficients matrix
$\mathbf{S_R}$	parametric matrix (s,r), which is twice differentiable with respect to the parameters
$\mathbf{P_{ij}}$	matrix of neighboring control points.

The general equation of the active surface energy is:

$$\begin{aligned} E(v) &= E_{\text{int}}(v) + E_{\text{ext}}(v) \\ E_{\text{int}}(v) &= \sum_{i,j} w_{ij,10}(s,r)\left\|\frac{\partial V_{ij}}{\partial s}\right\|^2 + w_{ij,01}(s,r)\left\|\frac{\partial V_{ij}}{\partial r}\right\|^2 + w_{ij,11}(s,r)\left\|\frac{\partial V_{ij}}{\partial r \partial s}\right\|^2 \\ &\quad + w_{ij,20}(s,r)\left\|\frac{\partial V_{ij}}{\partial s^2}\right\|^2 + w_{ij,02}(s,r)\left\|\frac{\partial V_{ij}}{\partial r^2}\right\|^2 \\ E_{\text{ext}}(v) &= -\gamma \left\|\mathrm{G} * \mathrm{I}(v)\right\| \end{aligned} \tag{14}$$

where G represents a 3D derivative of Gaussian kernel with fixed parameters μ and Σ, and $*$ is the 3D convolution operator.

Every surface $V_{ij}(s,r)$ can be described analytically using expression (13), and its derivatives can be calculated from the derivatives of the parametric matrix S_R as follows:

$$\left\|\frac{\partial V_{ij}}{\partial s}\right\|^2 = \left\|\left[\mathbf{M}^T \frac{\partial \mathbf{S}_{\mathrm{R}}}{\partial s}\mathbf{M}\right]_{row} \times [\mathbf{P}_{ij}]_{column}\right\|^2$$
$$\left\|\frac{\partial V_{ij}}{\partial r}\right\|^2 = \left\|\left[\mathbf{M}^T \frac{\partial \mathbf{S}_{\mathrm{R}}}{\partial r}\mathbf{M}\right]_{row} \times [\mathbf{P}_{ij}]_{column}\right\|^2$$
$$\left\|\frac{\partial^2 V_{ij}}{\partial s^2}\right\|^2 = \left\|\left[\mathbf{M}^T \frac{\partial^2 \mathbf{S}_{\mathrm{R}}}{\partial s^2}\mathbf{M}\right]_{row} \times [\mathbf{P}_{ij}]_{column}\right\|^2$$
$$\left\|\frac{\partial^2 V_{ij}}{\partial r^2}\right\|^2 = \left\|\left[\mathbf{M}^T \frac{\partial^2 \mathbf{S}_{\mathrm{R}}}{\partial r^2}\mathbf{M}\right]_{row} \times [\mathbf{P}_{ij}]_{column}\right\|^2$$

$$\left\| \frac{\partial^2 V_{ij}}{\partial r \partial s} \right\|^2 = \left\| \left[\mathbf{M}^T \frac{\partial^2 \mathbf{S}_{\mathrm{R}}}{\partial r \partial s} \mathbf{M} \right]_{row} \times [\mathbf{P}_{ij}]_{column} \right\|^2 .$$

As in the 2D case, it is necessary to find the surface that nulls the first derivative of the energy term. Proceeding as in Eq. (7), the resulting expression for the gradient is:

$$\nabla E(P_{ij}) = \mathbf{A}_{row} \mathbf{P}^a_{ij\,column} - \gamma F(\mathbf{P}^a_{ij}) \tag{15}$$

where $\mathbf{A}$ is the internal energy matrix containing the w_{ij} parameters of the model, $F(\mathbf{P}^a_{ij})$ is the external energy evaluated at the augmented matrix of control points, $\mathbf{P}^a_{ij}$, corresponding to:

$$\mathbf{P}^a_{ij} = \begin{bmatrix} P_{i-2,j-2} & P_{i-2,j-1} & P_{i-2,j} & P_{i-2,j+1} & P_{i-2,j+2} & P_{i-2,j+3} & P_{i-2,j+4} \\ P_{i-1,j-2} & P_{i-1,j-1} & P_{i-1,j} & P_{i-1,j+1} & P_{i-1,j+2} & P_{i-1,j+3} & P_{i-1,j+4} \\ P_{i,j-2} & P_{i,j-1} & P_{i,j} & P_{i,j+1} & P_{i,j+2} & P_{i,j+3} & P_{i,j+4} \\ P_{i+1,j-2} & P_{i+1,j-1} & P_{i+1,j} & P_{i+1,j+1} & P_{i+1,j+2} & P_{i+1,j+3} & P_{i+1,j+4} \\ P_{i+2,j-2} & P_{i+2,j-1} & P_{i+2,j} & P_{i+2,j+1} & P_{i+2,j+2} & P_{i+2,j+3} & P_{i+2,j+4} \\ P_{i+3,j-2} & P_{i+3,j-1} & P_{i+3,j} & P_{i+3,j+1} & P_{i+3,j+2} & P_{i+3,j+3} & P_{i+3,j+4} \\ P_{i+4,j-2} & P_{i+4,j-1} & P_{i+4,j} & P_{i+4,j+1} & P_{i+4,j+2} & P_{i+4,j+3} & P_{i+4,j+4} \end{bmatrix} .$$

Proceeding as in the 2D case, the energy optimization method is a normalized gradient descent, although now modified with a momentum term to stabilize convergence. The iterative equation for the surface upgrading is:

$$\underline{v}^{(t+1)} = \underline{v}^{(t)} - \mu \frac{\mathbf{A}_{row} \mathbf{P}^a_{ij\,column} - \gamma F\left(\mathbf{P}^a_{ij}\right)}{\left\| \mathbf{A}_{row} \mathbf{P}^a_{ij\,column} - \gamma F\left(\mathbf{P}^a_{ij}\right) \right\|} + \eta\, \Delta \underline{v}^{(t)} \tag{16}$$

where $\Delta \underline{v}^{(t)}]$ is a momentum term, representing the difference between the calculated surfaces in iterations $t-1$ and t and η is the corresponding term weight.

3.3. *Model initialization*

3.3.1. *Active contours*

It is well known that the gradient descent method is highly sensitive to the initial conditions. To provide the procedure with an adequate starting point, the initial set of control points is obtained from the outcome of a Canny edge detection filter[6] with kernel size of 3×3 pixels and parameters $\mu = 0$ and $\sigma^2 = 1$. Given that this operator is the same one utilized for the estimation of the external energy term for the snake, we assume that an initial contour obtained in this way will be close to the desired image border.

Not all the pixels from the detected edge are used for the initial contour, because the initialization would remain sensitive to detection noise. In order to avoid this problem, a thresholding criteria, relative to the maximum intensity of pixels over the detected edge, is first applied to extract initialization candidates. Then all

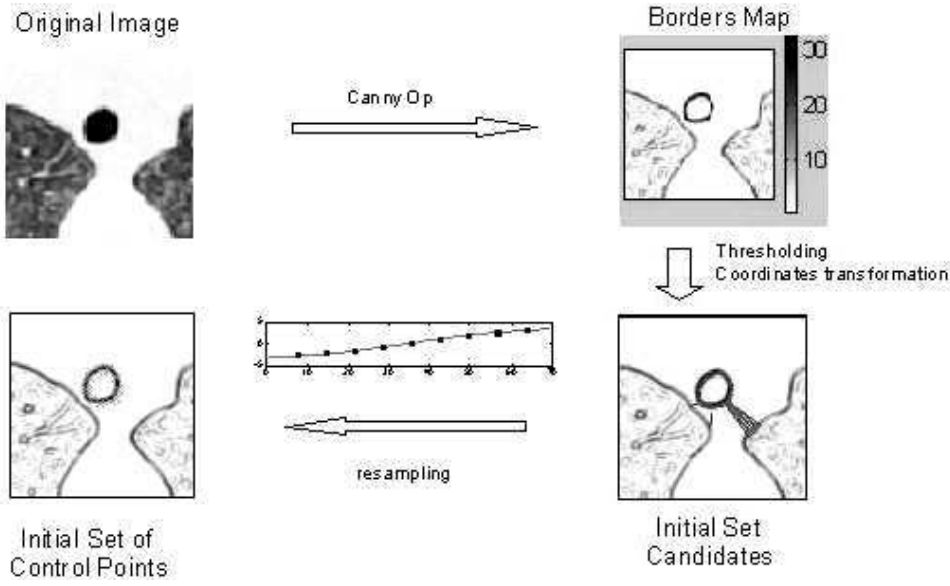


Fig. 3. Initial contour definition procedure.

candidates are subject to a spatial distribution test, that finally selects the N control points. For the trachea segmentation application, given the closely circular nature of the contours, the spatial distribution test consists of a rectangular-polar coordinate transformation, an angular sort relative to the center of mass of the candidate set, and equally-distributed angular sub-sampling. The procedure is shown in Fig. 3.

The proposed segmentation method is based on the following hypothesis:

(1) α and β in Eq. (1) remain constant; and
(2) the control points are regularly spaced.

In order to ensure the second assumption, it is necessary to interpolate the set of control points to M locations ($M \gg N$) and then perform re-sampling to build a new uniformly spaced set at every iteration.

3.3.2. *Active surfaces*

For the 3D case, the initial volume is formed from the CT image stack, and correspondingly, a 3D Canny convolution kernel is employed ($\underline{\mu} = \underline{0}$, $\mathbf{\Sigma} = \sigma^2\mathbf{I} = \mathbf{3I}$) with kernel size of $5 \times 5 \times 3$ voxels. The difference in length for the third coordinate is due to the anisotropy of voxels, that is, the difference of spatial resolution in the (x, y) plane versus the z dimension. The volume is convolved with the 3D Canny kernel and the detected borders are processed slice by slice in the same way as in the 2D case described before. Because the control point set has to be organized into a 3D mesh defining the active surface, all the resulting 2D discrete contours

are spatially matched to minimize the distance between pairs of control points in adjacent slices, in order to define four-sided square facetes with minimum torsion.

3.4. *Performance evaluation*

Convergence of active contour models to the desired image border is dependent on various factors such as the set of initial control points, the noise level and the optimization strategy selected, which in the case described consists of a normalized gradient descent algorithm to perform the energy minimization iterations. This standard procedure is guaranteed to linearly converge to a local minimum of the objective function, given that an adequate step constant is provided.[19] As the proposed initialization procedure sets the starting contour close to the border of interest, it is expected that the solution will converge to it, under controlled noise conditions. This is also true for the reference method by Kass *et al.*,[23] however, as the interpolation model imposes smoothness and continuity restrictions to the contour, the approach is capable of recovering from eventual spurious control points, as shown in Fig. 4. For the case shown in this figure, the initialization procedure sets some of the initial control points at the lung wall boundary, and the solution by Kass *et al.* converges to the wrong contour, while the proposed approach attains the correct result, even for different step constants. Nonetheless, as is the case for the other active contour models available, severe noise conditions degrade the quality and correctness of the boundary estimates. Linear convergence of the proposed procedure is evident from the results as expected. Moreover, the reference method converges faster, although to a worse solution under the same conditions.

Figure 5 shows the original images and the 3D rendering of the segmented trachea from one CT stack, along with the tracheal perimeter, equivalent circular cross-section area, and tracheal axis profiles, derived from the adjusted model information. Figure 6 presents the corresponding information for a different case. The severity of the lesion for this case is evident, as well as its localization and extension along the tracheal structure.

4. Conclusions

An alternative formulation for active contour and surface models has been presented. It is based on the inclusion of explicit analytical expressions for the contour or surface derivatives in the internal energy description associated to the active model. These expressions are constructed from an interpolation of the desired contour using natural cubic splines. Further, the application of such formulation to the problem of trachea segmentation in neck-thorax CT studies has been discussed, in the spirit of recovering quantitative information for the diagnostic description of tracheal stenosis.

The proposed procedure improves over the reference methods (Kass *et al.*,[23] Cohen *et al.*[10]) in the sense that it generates a geometrically better estimate, with

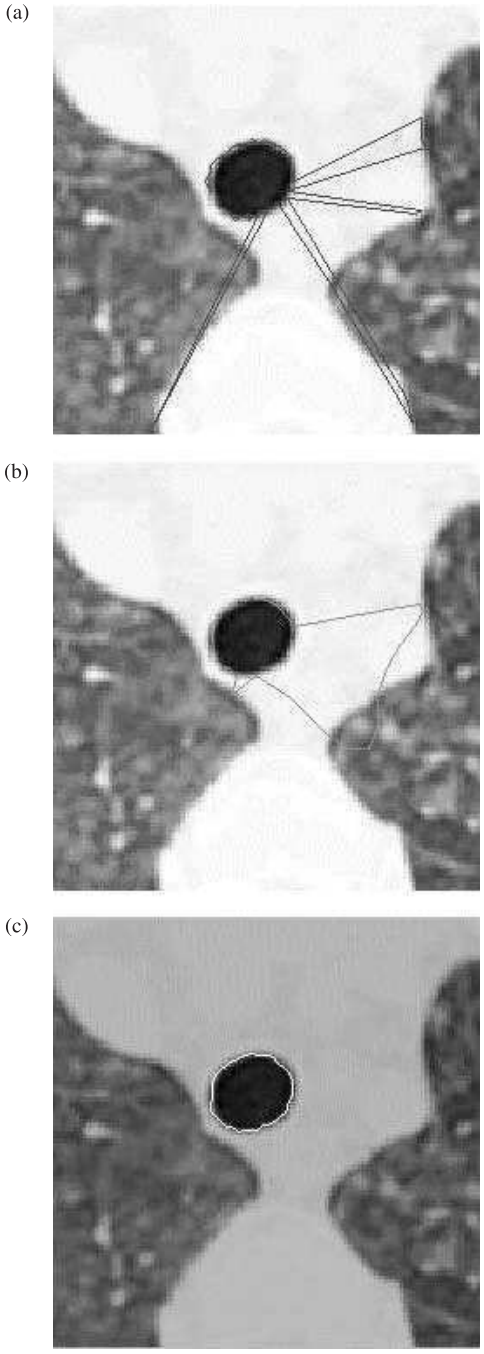


Fig. 4. Convergence analysis. (a) Initial contour. (b) Final contour by Kass *et al.* procedure. (c) Final contour by Valdés *et al.* procedure. (d) Mean Contour Difference (MCD) along iterations for (b). (e) MCD along iterations for (c).

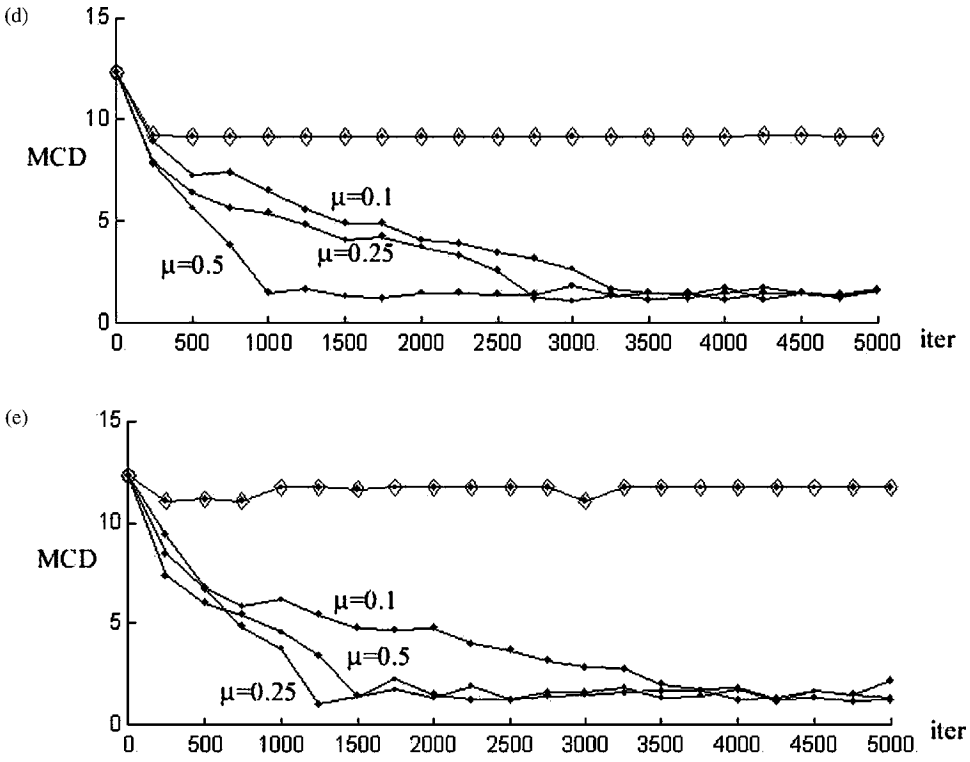
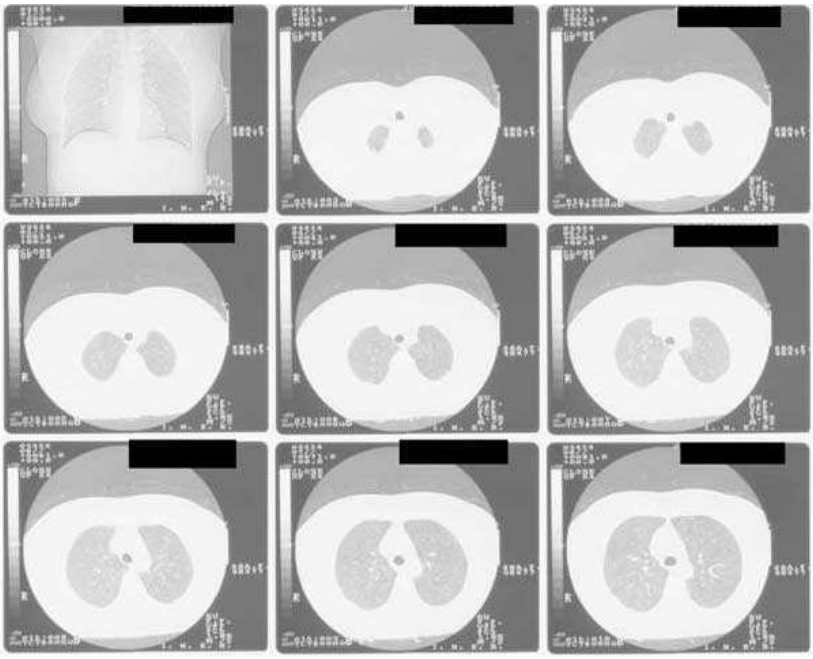


Fig. 4. (*Continued*)

good localization. Another relevant aspect of the proposed model is that in general it shows a tendency to smooth the contour or surface estimates. As long as the density of control points around a particular image region is large, the smoothing effect is minimized for the estimate in that zone. However, the larger the set of control points, the higher the computational cost of the adaptation procedure, which could yield the method impractical. Unfortunately it is hard to determine an optimal density of control points for an irregular surface, that is, one that contains intermixed portions of low and high curvature. A basic assumption of the reference active contour models and the one presented here is that the control points are equally spaced. This condition has to be relaxed in order to locally adjust the density of control points with some guided procedure that takes into account the geometry of the regions being analyzed. Work in this sense has recently been developed by Lobregt *et al.*,[30] Marín *et al.*[34] or Valdés *et al.*[60]

An issue always present in the active model applications is the definition of the initial model. For the present method, the initialization is carried out by sampling an edge enhanced version of the scan image that is obtained from filtering with a Canny detector. This sampling, together with a radial sorting procedure worked adequately for cases in which the contours are geometrically regular and



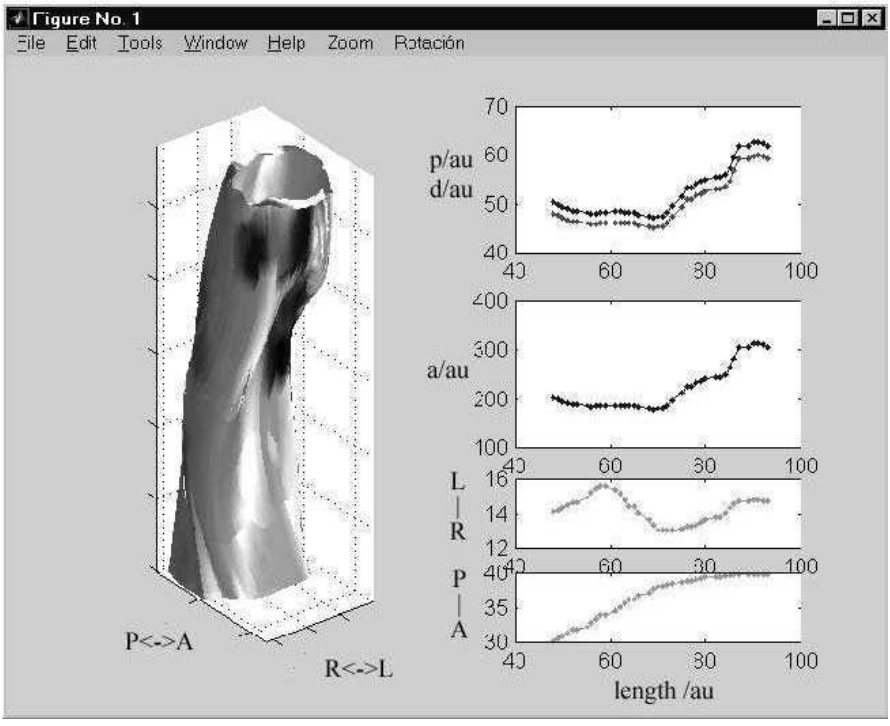


Fig. 5. Case 01: (a) Image stack (b) 3D trachea reconstruction with perimeter and area profiles. P: posterior; A: anterior; R: right; L: left; p: perimeter; d: diameter; a: area; au: arbitrary units.

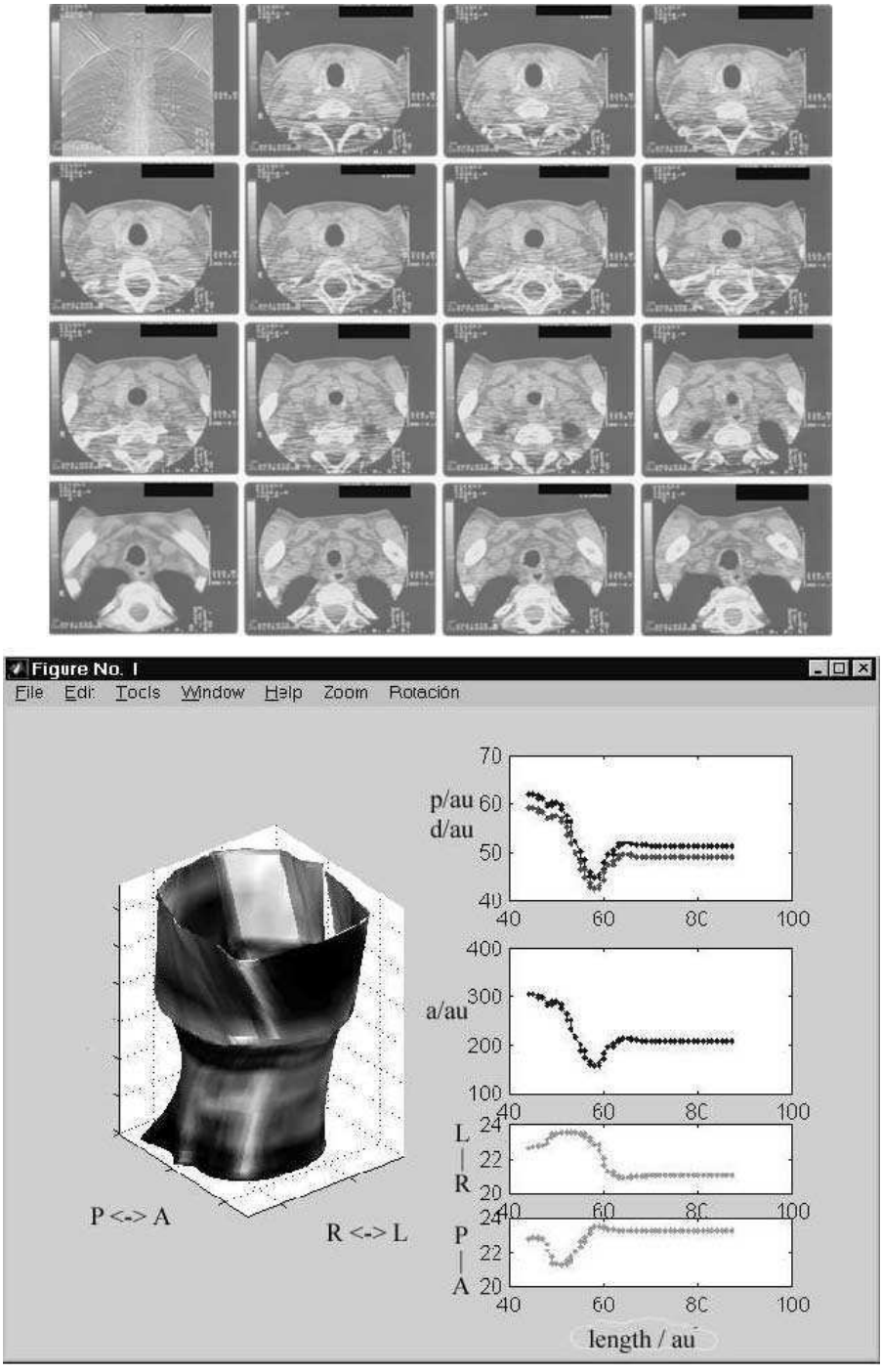


Fig. 6. Case 02: (a) Image stack (b) 3D trachea reconstruction with perimeter and area profiles. P: posterior; A: anterior; R: right; L: left; p: perimeter; d: diameter; a: area; au: arbitrary units.

convex (such as the trachea in axial CT scans), even in situations of severe stenosis. However, if the model is applied to another structure, most surely a different initialization procedure would have to be designed.

Given the adequate localization properties of the Canny detector, the contour is initialized close to the borders of interest, and therefore close to the minimum of the objective function. Further, the analytical derivatives of this function are also available through the model for direct numerical evaluation. These conditions relax the requirements for the optimization procedure, and therefore a simple approach like the gradient descent taken here is justified, not ignoring that a more efficient procedure might be devised. There is a tradeoff between linear (slower) convergence and the quality of the resulting contour estimate: the reference procedure[23] converges faster, but the interpolation-based model provides systematically better contours. The expected clinical usefulness of the procedure pushes this tradeoff towards the use of the proposed model.

Regarding the clinical application of the proposed segmentation method, the results yield a 3D mesh model of the trachea that is adequate for rendering and manipulating in a visualization system. From the model, cross-sectional area and perimeter profiles can be constructed in order to provide quantitative information for the assessment of the lesions being investigated. Implied in these measurements is the need for a calibration procedure based on a tomographic phantom.

References

1. M. B. Abbott, B. J. Dardzinski and L. F. Donnelly, Using volume segmentation of cine MR data to evaluate dynamic motion of the airway in pediatric patients, *American Journal of Radiology* **181** (2003) 857–859.
2. R. Arens, J. M. McDonough, A. M. Corbin, N. K. Rubin, M. E. Carroll, A. I. Pack, J. Liu and J. K. Udupa, Upper airway size analysis by magnetic resonance imaging of children with obstructive sleep apnea syndrome, *American Journal of Respiratory and Critical Care Medicine* **167**, 1 (2003) 65–70.
3. D. Aykac, E. A. Hoffman, G. McLennan and J. M. Reinhardt, Segmentation and analysis of the human airway tree from three-dimensional X-ray CT images, *IEEE Transactions on Medical Imaging* **22**, 8 (2003) 940–950.
4. M. S. Brown, J. G. Goldin, M. F. McNitt-Gray, L. E. Greaser, A. Sapra, K. T. Li, J. W. Sayre, K. Martin and D. R. Aberle, Knowledge-based segmentation of thoracic computed tomography images for assessment of split lung function, *Medical Physics* **27**, 3 (2000) 592–598.
5. M. S. Brown, M. F. McNitt-Gray, N. J. Mankovich, J. G. Goldin, J. Hiller, L. S. Wilson and D. R. Aberle, Method for segmenting chest CT image data using an anatomical model: Preliminary results, *IEEE Transactions on Medical Imaging* **16**, 6 (1997) 828–839.
6. J. A. Canny, A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**, 6 (1986) 679–698.
7. F. Chabat, H. Xiao-Peng, D. M. Hansell and Y. Guang-Zhong, ERS transform for the automated detection of bronchial abnormalities on CT of the lungs, *IEEE Transactions on Medical Imaging* **20**, 9 (2001) 263–273.

8. R. Chiplunkar, J. M. Reinhardt and E. A. Hoffman, Segmentation and quantification of the primary human airway tree from 3D X-ray CT, *Proceedings of SPIE Medical Imaging* **3033** (1997) 403–414.
9. J. P. Cocquerez and S. Philipp, Analyse d'images: Filtrage et segmentation, Masson Rans (1995) 191–210.
10. L. Cohen and I. Cohen, Finite element methods for active contour models and balloons for 2-D and 3-D images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-15**, 11 (1993) 1131–1147.
11. P. Dastidar, T. Heinonen and J. Numminen, Semi-automatic segmentation of computed tomographic images in volumetric estimation of nasal airway, *European Archives of Otorhinolaringology* **256**, 4 (1999) 192–198.
12. N. D. D'Souza, J. M. Reinhardt and E. A. Hoffman, ASAP: Interactive quantification of 2D airway geometry, *Proceedings of SPIE Medical Imaging* **2709** (1996) 180–196.
13. A. El-Bazl, A. A. Farag, R. Falk and R. La Rocca, Automatic identification of lung abnormalities in chest spiral CT scans, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* **2** (2003) II–261–II–264.
14. L. Fan and C. W. Chen, Reconstruction of airway tree based on topology and morphological operations, *Proceedings of SPIE Medical Imaging* **3978** (2000) 46–57.
15. C. L. Fetita and F. Preteux, Quantitative 3D CT bronchography, *Proceedings of the IEEE International Symposium on Biomedical Imaging*, 2002, pp. 221–224.
16. A. P. Fishman, *Tratado de neumología* (McGraw Hill, 1983), pp. 466–478.
17. W. Gardner and W. Osburn, *Anatomía humana*, 3rd edn., ch. 9 (Editorial Interamericana, 1985).
18. J. M. Garrity, W. P. Segars, S. B. Knisley and B. M. W. Tsui, Development of a dynamic model for the lung lobes and airway tree in the NCAT phantom, *IEEE Transactions on Nuclear Science* **50**, 3 (2003) 378–383.
19. P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization* (Academic Press, 1981), pp. 102–105.
20. R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (Addison-Wesley, 1996), pp. 447–522.
21. W. E. Higgins, J. P. Helferty and D. R. Padfield, Integrated bronchoscopic video tracking and 3D CT, *Proceedings of SPIE Medical Imaging* **5031** (2003) 80–89.
22. W. Park, J. M. Reinhardt, E. A. Hoffman and M. Sonka, Intrathoracic airway wall detection using graph search with CT scanner PSF information, *Proceedings of SPIE Medical Imaging* **3033** (1997) 93–101.
23. M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active contour models, *Proceedings of the First International Conference on Computer Vision* **1** (1987) 259–268.
24. A. P. Kiraly and W. E. Higgins, Analysis of branching tubular structures in 3D digital images, *Proceedings of the IEEE International Conference on Image Processing* **2** (2002) II–333–II–336.
25. A. P. Kiraly, W. E. Higgins, E. A. Hoffman, G. McLennan and J. M. Reinhardt, 3D human airway segmentation for virtual bronchoscopy, *Proceedings of SPIE Medical Imaging* **4683** (2002) 16–29.
26. A. P. Kiraly, W. E. Higgins, G. McLennan, E. A. Hoffman and J. M. Reinhardt, Three dimensional human airway segmentation methods for clinical virtual bronchoscopy, *Academic Radiology* **9** (2002) 1153–1168.
27. T. Y. Law and P. A. Heng, Automated extraction of bronchus from 3D CT images of lung based on genetic algorithm and 3D region growing, *Proceedings of SPIE Medical Imaging* **3979** (2000) 905–916.

28. J. Liu, J. K. Udupa, D. Odhner, J. M. McDonough and R. Arens, Upper airway segmentation and measurement in MRI using fuzzy connectedness, *Proceedings of SPIE Medical Imaging* **4683** (2002) 238–249.
29. J. G. Liu, J. K. Udupa, D. Odhner, J. M. McDonough and R. Arens, Segmentation and visualization of tissues surrounding the airway in children via MRI, *Proceedings of SPIE Medical Imaging* **5031** (2003) 298–307.
30. S. Lobregt and M. Viergever, A discrete dynamic contour model, *IEEE Transactions on Medical Imaging* **14**, 1 (1995) 12–24.
31. J. LoCicero, R. Costello, Ch. Campos, N. Francalancia, K. M. Dushay, R. C. Silvestri and J. Zibrak, Spiral CT with multiplanar and three-dimensional reconstructions accurately predicts tracheobronchial pathology, *Annals of Thoracic Surgery* **62** (1996) 811–817.
32. S. Loncaric, T. Markovinovic, T. Petrovic, D. Ramljak and E. Sorantin, Construction of virtual environment for endoscopy, *IEEE International Conference on Multimedia Computing and Systems* **1** (1999) 475–480.
33. S. Macdonell, *Asistencia respiratoria* (Editorial Salvat, 1980), pp. 274–296.
34. A. Marín-Hernández and H. Ríos-Figueroa, Eels: electric snakes, *Computación y Sistemas* **2**, 2–3 (1999) 87–94.
35. D. Marr and E. Hildreth, Theory of edge detection, *Proceedings of the Royal Society of London* **B207** (1980) 187–217.
36. E. L. McKinstry, A. C. Tooker, K. S. Hong, P. Costello, F. A. Jolesz and M. S. Albert, Airway imaging in humans using hyperpolarized helium-3 MRI, *Proceedings of the 24th IEEE Engineering in Medicine and Biology Conference* **2** (2002) 923–924.
37. N. G. Mogavero, D. J. Conces and K. K. Kopecky, Spiral CT evaluation of the trachea and bronchi, *Journal of Computer Assisted Tomography* **18**, 4 (1994) 552–554.
38. R. C. Molthen, S. T. Haworth, A. E. Heinrich and C. A. Dawson, Measuring the effect of airway pressure on pulmonary arterial diameter in the intact rat lung, *Proceedings of SPIE Medical Imaging* **5031** (2003) 195–203.
39. K. Mori, J. Hasegawwa, J. Toriwaki, H. Anno and K. Katada, Recognition of bronchus in three-dimensional X-ray CT, *Proceedings of the International Conference on Pattern Recognition*, 1996, pp. 528–532.
40. M. U. Niethammer, U. J. Schoepf, J. E. Wildberger, E. Klotz, H. Fichte and S. Schaller, A new approach to diagnosis of pulmonary embolism using multi-slice CT, *Proceedings of SPIE Medical Imaging*, 2001, pp. 244–253.
41. OMS, Informe sobre la salud en el mundo, 2003, Cuadro 2, Defunciones por causas, sexo y estratos de mortalidad en las regiones de la OMS, pp. 170–174.
42. K. Palagyi, J. Tschirren and M. Sonka, Quantitative analysis of three-dimensional tubular tree structures, *Proceedings of SPIE Medical Imaging* **5031** (2003) 277–287.
43. W. Park, E. A. Hoffman and M. Sonka, Segmentation of intrathoracic airway trees: A fuzzy logic approach, *IEEE Transactions on Medical Imaging* **17**, 4 (1998) 489–497.
44. F. Prêtreux, C. Fetita, A. Capderon and P. Grenier, Modeling, segmentation, and caliber estimation of bronchi in high resolution computerized tomography, *Journal of Electronic Imaging* **8**, 1 (1999) 36–45.
45. J. M. Reinhardt, N. D. D'Souza and E. A. Hoffman, Accurate measurement of intrathoracic airways, *IEEE Transactions on Medical Imaging* **16**, 6 (1997) 820–827.
46. J. C. Russ, *The Image Processing Handbook*, 2nd edn. (CRC Press, 1995), pp. 347–406.
47. M. Sagy, M. Poustchi-Amin, L. Nimkoff, D. Silver, M. Shikowitz and J. C. Leonidas, Spiral computed tomographic scanning of the chest with three dimensional imaging in

the diagnosis and management of paediatric intrathoracic airway obstruction, *Thorax* **51** (1996) 1005–1009.
48. J. S. Silva, A. Silva, B. S. Santos and J. Madeira, Detection and 3D representation of pulmonary air bubbles in HRCT volumes, *Proceedings of SPIE Medical Imaging* **5031** (2003) 430–439.
49. M. Sonka and L. Fitzpatrick, *Handbook of Medical Imaging, Vol. II: Medical Image Processing and Analysis* (SPIE Press, 2000).
50. M. Sonka, W. Park and E. A. Hoffman, Rule-based detection of intrathoracic airway trees, *IEEE Transactions on Medical Imaging* **15**, 3 (1996) 314–326.
51. E. Sorantin, C. Halmai, B. Erdohelyi, K. Palagyi, L. G. Nyul, K. Olle, B. Geiger, F. Lindbichler, G. Friedrich and K. Kiesler, Spiral-CT-based assessment of tracheal stenoses using 3D skeletonization, *IEEE Transactions on Medical Imaging* **21**, 3 (2002) 263–273.
52. R. M. Summers and J. R. Cebral, Tracheal and central bronchial aerodynamics using virtual bronchoscopy, *Proceedings of SPIE Medical Imaging*, 2001, pp. 22–31.
53. R. M. Summers, D. H. Feng, S. M. Holland, M. C. Sneller and J. H. Shelhamer, Virtual bronchoscopy: Segmentation method for real-time display, *Radiology* **200** (1996) 857–862.
54. R. M. Summers, R. Mullick, S. E. Finkelstein and D. S. Schrump, Confocal volume rendering of the thorax, *Proceedings of the IEEE*, 2001, pp. 297–299.
55. R. M. Summers, M. C. Sneller, C. A. Langford, J. H. Shelhamer and B. J. Wood, Improved virtual bronchoscopy using a multi-slice helical CT scanner, *Proceedings of SPIE Medical Imaging* **3978** (2000) 117–121.
56. G. Sundaramoorthy, J. D. Hoford and E. A. Hoffman, IMPROMPTU — a system for automatic 3D medical image-analysis, *Computerized Medical Imaging and Graphics* **19**, 1 (1995) 131–143.
57. R. D. Swift, A. P. Kiraly, A. J. Sherbondy, A. L. Austin, E. A. Hoffman, G. McLennan and W. E. Higgins, Automatic axis generation for virtual bronchoscopic assessment of major airway obstructions, *Computerized Medical Imaging and Graphics* **26** (2002) 103–118.
58. R. D. Swift, K. Ramaswamy and W. Higgins, Adaptive axes generation algorithm for 3D tubular structures, *Proceedings of the IEEE International Conference on Image Processing* **II** (1997) 136–139.
59. J. Tschirren, E. A. Hoffman, G. McLennan and M. Sonka, Branchpoint labeling and matching in human airway trees, *Proceedings of SPIE Medical Imaging* **5031** (2003) 187–194.
60. R. Valdés-Cristerna, V. Medina-Bañuelos and O. Yáñez-Suárez, Coupling of radial–basis network and active contour model for multispectral brain MRI segmentation, *IEEE Transactions on Biomedical Engineering* **51**, 3 (2004) 453–470.
61. R. Valdés-Cristerna and O. Yáñez-Suárez, Active contours and surfaces with cubic splines for semiautomatic tracheal segmentation, *Journal of Electronic Imaging* **12**, 1 (2003) 81–96.
62. S. A. Wood, E. A. Zerhouni, J. D. Hoford, E. A. Hoffman and W. Mitzner, Measurement of three dimensional lung tree structures by using computed tomography, *Journal of Applied Physiology* **79**, 5 (1995) 1687–1697.
63. D. M. Wooton, A. Guez, P. Vaidyanathan, D. Avinash, J. R. Stablum, J. M. McDonough, J. K. Udupa and R. Arens, Model of upper airway flow restriction in children with obstructive sleep apnea, *Proceedings of the 29th IEEE Bioengineering Conference*, 2003, pp. 217–218.

64. K. C. Yu, E. L. Ritman, A. P. Kiraly, S. Y. Wan, M. Zamir and W. E. Higgins, Toward reliable multi-generational analysis of anatomical trees in 3D high-resolution CT images, *Proceedings of SPIE Medical Imaging* **5031** (2003) 178–186.
65. L. Zhang, E. A. Hoffman and J. M. Reinhardt, Atlas-driven lung lobe segmentation in volumetric X-ray CT images, *Proceedings of SPIE Medical Imaging* **5031** (2003) 308–319.

This page intentionally left blank

CHAPTER 4

KNOWLEDGE-BASED SYSTEM FOR CONTOURING THE SPINAL CORD IN COMPUTED TOMOGRAPHY IMAGES

NECULAI ARCHIP
Brigham and Women's Hospital and Harvard Medical School
75 Francis St., Boston, MA 02115, USA
narchip@bwh.harvard.edu

PIERRE-JEAN ERARD
Computer Science Department, University of Neuchâtel
Rue Emile Argand 11, CH-2007 Neuchâtel, Switzerland
pierre-jean.erard@unine.ch

ROBERT ROHLING
Department of Electrical and Computer Engineering
University of British Columbia
2356 Main Mall, Vancouver, B.C., Canada
rohling@ece.ubc.ca

Accurate planning of radiation therapy entails the definition of treatment volumes and a clear delimitation of normal tissue of which unnecessary exposure should be prevented. The spinal cord is a radiosensitive organ, which should be precisely identified because an overexposure to radiation may lead to undesired complications for the patient such as neuronal dysfunction or paralysis. In this chapter, a knowledge-based approach to identifying the spinal cord in computer tomography images of the thorax is presented. The approach relies on a knowledge-base which consists of a so-called anatomical structures map (ASM) and a task-oriented architecture called the plan solver. The ASM contains a frame-like knowledge representation of the macro-anatomy in the human thorax. The plan solver is responsible for determining the position, orientation and size of the structures of interest to radiation therapy. The plan solver relies on a number of image processing operators. Some are so-called atomic (e.g. thresholding and snakes) whereas others are composite. The whole system has been implemented on a standard workstation. Experimental results performed on 23 patients show that the approach is reliable in spinal cord segmentation.

Keywords: Spinal cord; image interpretation; knowledge representation; medical imaging; radiotherapy.

1. Introduction

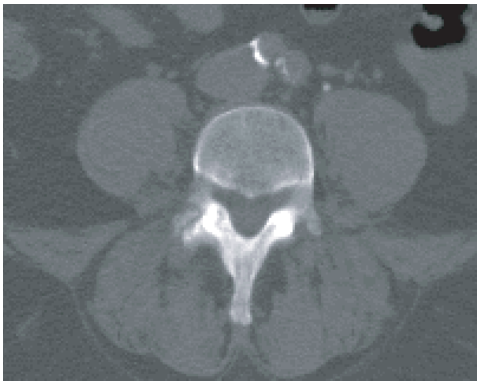
Radiotherapy is an important ingredient in the often complex treatment procedures that are initiated in order to suppress different kinds of malignant tumors.

The purpose of radiation therapy is to eradicate the tumor while minimizing the damage caused to the surrounding healthy tissues. The spinal cord is an extremely radiosensitive vital organ which should be spared as much as possible. A certain amount of exposure to radiation can induce a number of undesired neurological complications in the spinal cord (e.g. paralysis).[7,17,25] The risk of such serious complications implies much smaller dosage tolerances for the spinal cord than for the tumor.

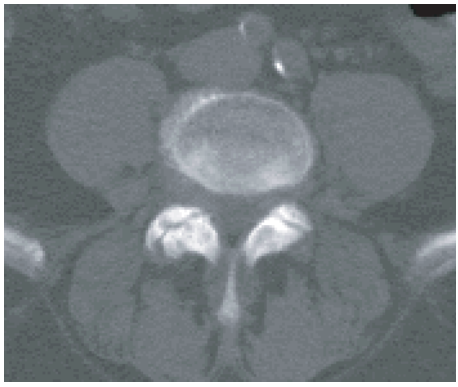
Successful radiotherapy relies on a precise planning and a thorough implementation of the radiation procedure. One of the problems of radiotherapy planning addressed here, is that it requires that the different tissues of interest, including the tumor and the surrounding (vital) organs, are to be located with high accuracy. At present, radiation therapy is being planned by a radiologist and a radiotherapist in concert, based on a careful analysis of a Computed Tomography (CT) scan that covers the tumor and the surrounding tissues. The current planning procedure entails the manual delineation of the spinal cord in each separate slice followed by an automatic reconstruction performed by the image analysis workstation that is connected with the CT-scanning device. Despite the existence of several semi-automatic approaches for the planning of repetitive radiotherapy,[8,14] automatic detection of the spinal cord in CT images remains an unresolved problem. A factor that complicates the analysis further is the occasional partial appearance of the spine around the spinal canal (Fig. 1).

Tumors are heterogeneous lesions, which exhibit growth patterns that are unique to each patient. As a consequence, the CT images cannot be acquired according to a standardized protocol but are subject to much inter-patient variation, e.g. compared with mammograms[32] or standard thorax radiographs[34] that are acquired in large numbers in Europe and North America. This rather high amount of variation in our image material impedes the application of a standard low-level image processing technique. For an image processing algorithm to be successful in our application, it should be *flexible* and also *transparent* to the radiologist and radiotherapist. A flexible approach can better cope with a high amount of inter-patient variation. Transparency of the image processing algorithms ensures that the experts can take over the image analysis, in case the automatic approach fails to give the desired result. To cope with the requirements of flexibility and transparency, we present a knowledge-based approach to automatic image analysis. The basic components of our system are the so-called *Anatomical Structures Map* (ASM)[3] and the *Plan Solver*, a task-oriented module that controls the sequence in which the subtasks are performed. The ASM and the Plan Solver are designed such that they capture parts of the anatomical and procedural knowledge that is currently being used for manual image interpretation.

The chapter is structured as follows. First, existing approaches to knowledge-based image interpretation are discussed. Then, we consider different archetypes of knowledge that are presently used to solve the spinal cord detection problem. Subsequently, we give a detailed description of the knowledge-based architecture



(a)



(b)

Fig. 1. The two types of slices: (a) spinal canal completely surrounded by bone, (b) spinal canal partially surrounded by bone.

consisting of the anatomical structures map and the plan solver. Following this description, the low-level (atomic) image operators are described in detail. In the experimental section, we report the results obtained by applying our approach to the CT images obtained from 23 patients before they underwent radiation therapy. The chapter ends with a discussion of the results and issues for future research.

2. Knowledge Guided Image Processing

2.1. *Image acquisition and interpretation*

In clinical routine, the radiotherapist performs a request containing the questions which should be resolved by the radiological examination (e.g. Where is the tumor located? How far is it from the spine? Are there other healthy tissues that will be exposed to radiation?).

The image acquisition is performed according to a standard protocol, which contains general guidelines of how CT images should be obtained for the planning of radiotherapy. The details of an acquisition are chosen such that the tumor of the particular patient is visualized in the best possible way. In general, a number of aspects should be taken into account in order to acquire CT images in such a way that the relevant findings can be established. Wegener[36] points out that there is a strong relationship between what region, organ or lesion is examined and how the image should be acquired, including imaging parameters (slice thickness, slice interval, scanning time), and contrast administration (presence/type of contrast agent, injection speed, concentration).

After CT images have been acquired, the interpretation is performed by a radiologist and a radiotherapist in concert. The image assessment relies on both morphological and densitometric findings. Grimnes mentions a number of general aspects that influence the interpretation of CT images:[16]

- the typical size and shape of the objects (organs);
- the variation in size and shape of the objects;
- the expected Hounsfield Unit (HU) value range associated with each tissue;
- the variation in the HU value range associated with each tissue;
- typical response of an organ to the contrast tracer that is used;
- organs and blood may change their expected HU range in light of disease;
- biological variation;
- and social context.

The radiological analysis results in a *synthesis* of the clinical data by relevant findings that were present in the CT images, while taking the above mentioned aspects into account. The ultimate goal of any computer system for image interpretation should be to produce such an image synthesis, either automatically or in an interactive manner, e.g. through a dialogue with the radiologist.

2.2. *Existing approaches to knowledge-based image interpretation*

The literature on computer-based image interpretation describes a large number of architectures, systems and approaches. Among the conventional approaches for image interpretation, some focus on architectural aspects of the scene (the spatial configuration composed by the objects that are present);[35] in other approaches an extensive knowledge base and an advanced reasoning strategy form the major components.[27,31,38] Probabilistic systems were also developed for knowledge-guided image interpretation.[20,22,33] Several blackboard and other knowledge-based systems were developed specifically for the interpretation of medical images: The ERNEST system has been developed for the interpretation of scintigraphic images and MR images.[23] The system VIA-RAD[30] applies four diagnostic strategies, obtained from the radiological domain, to perform image interpretation. Brown *et al.*[9] present a knowledge-based system for lung detection in CT images. A system

for object recognition in CT images of the brain is presented in Ref. 26. An architecture has been developed for the interpretation of abdominal CT images.[13] A task-based architecture to the interpretation of MR images of the brain is introduced by Gong *et al.*[15]

In computer-based systems for the interpretation of medical images, one or more of the following archetypes of knowledge may be modeled:[11]

- *structural knowledge*, which can contain information about the physical world (human anatomy, e.g. normal structures such as lungs, spinal canal, lamina, spinal cord, body contour, etc.);
- *dynamic knowledge*, which can contain information about possible normal and abnormal processes (human physiology and pathology);
- *procedural knowledge*, which divides a request (e.g. image synthesis) into a sequence of subtasks that can be performed by specific image processing algorithms.

In some applications, a satisfactory image synthesis can be obtained from solely one type of knowledge. For example, in perfusion analysis of bone tumors dynamic knowledge is sufficient for making a distinction between viable tumor and necrosis.[12] In other applications, all three types of knowledge may be a prerequisite for a successful image synthesis. Spinal cord detection and subsequent planning of radiotherapy rely primarily on structural knowledge components: where is the tumor located, the spine, etc. and on the procedural knowledge that is needed to describe how the CT images should be analyzed.[1,29]

2.3. *Knowledge representation in medical image analysis*

We will present an approach for semi-automatic image interpretation that uses a knowledge base to link different low-level image processing algorithms. For a solution of the problem addressed — spinal cord detection — a combination of *structural* and *procedural* knowledge suffice, because the pathologic growth process of the tumor does not have to be taken into account. This demarcation implies that our knowledge base should contain medical knowledge about organs and possible pathologic structures, i.e. components of the tumor. The knowledge-base is used to guide the image interpretation but also to specify the parameters of these algorithms. The model presented is inspired by *frame*[19] systems. Each anatomical structure is represented as a prototype, and its properties as *slots* as explained in greater detail later.

2.3.1. *Structural knowledge*

The core of our system is the so-called anatomical structures map (Fig. 2). (ASM), which was presented earlier in Ref. 3. A set of properties (related to shape, position, densitometric ranges) is used to characterize each of the normal structures, the

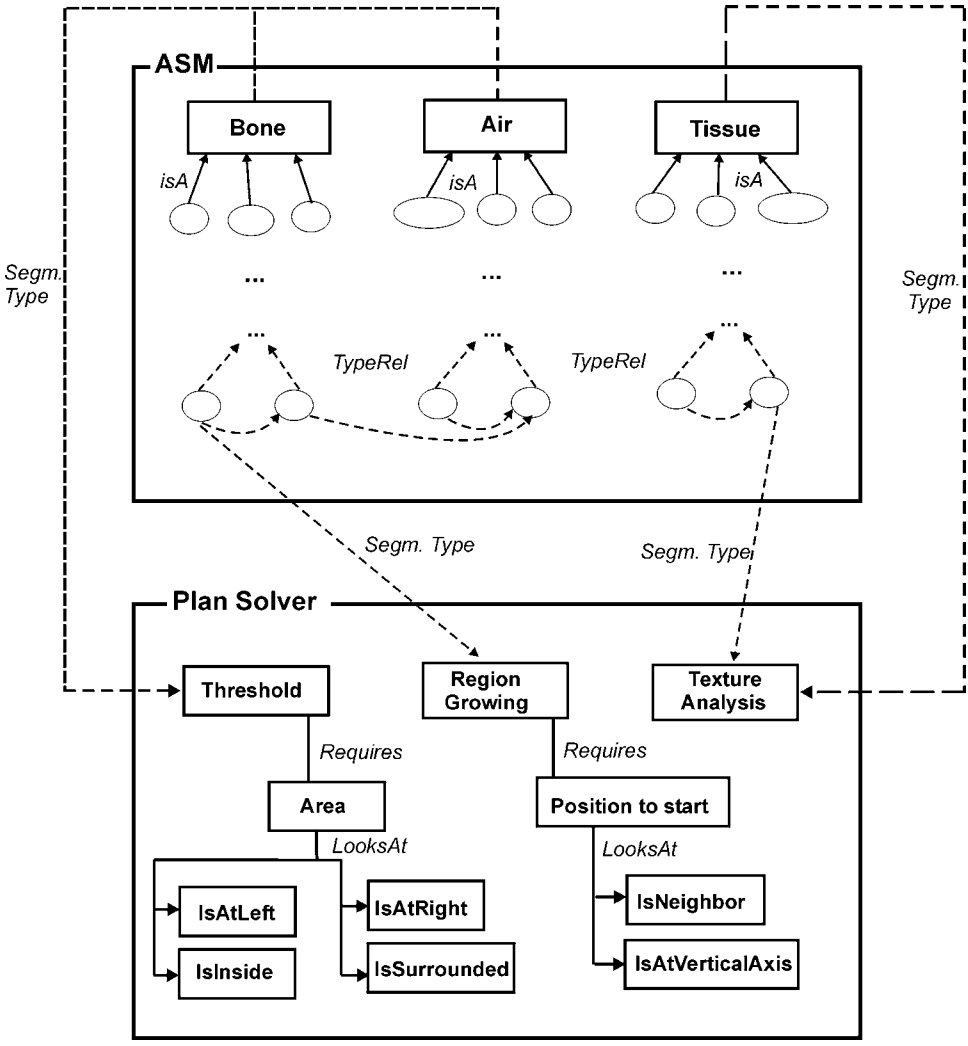


Fig. 2. Relationships in the *Anatomical Structures Map — Plan Solver*. The *ASM* presents relationships between structures, and the *Plan Solver* describes segmentation methods and associated parameters. The connections between structures and the corresponding segmentation methods are also presented.

organs, bones and the vascular system, that are represented in the ASM. The spatial arrangement of these objects is represented as a semantic network. A very simple grammar was also introduced that makes it feasible to express the semantic relations that pertain to our application, (see Table 1).

2.3.2. *Procedural knowledge*

The structural knowledge base is merged with a *task oriented* architecture, the *plan solver*, which contains the procedural knowledge that is needed to perform

Table 1.

Relation	Description
isInside	Object1 *isInside* Object2 ⇔ allmost all the pixels of Object1 are included in Object2
isSurrounded	Object1 *isSurrounded* Object2 ⇔ allmost all the pixels of Object1 are included in Object2, and Object1 is the only object which respects Object1 *isInside* Object2
isAtMedialAxis	Object1 *isAtMedialAxis* Object2 ⇔ the center of Object1 is approximately the same as the center of Object2
isNeighbor	Object1 *isNeighbor* Object2 ⇔ the Object1 has a common border with Object2
isAtLeft	Object1 *isAtLeft* Object2 ⇔ allmost all the pixels of Object1 are at the left border of Object2
isAtRight	Object1 *isAtRight* Object2 ⇔ allmost all the pixels of Object1 are at the right border of Object2

the image interpretation. The involved clinicians make use of so-called reference objects (e.g. body or lungs as in Fig. 3) to direct their focus of attention. Although the architecture of the plan solver was originally inspired by the approach followed by the involved clinicians, the task-based structure also makes it possible to recognize complex objects while benefiting from more simple (basic) object detections. Algorithms developed for the recognition of complex objects use so-called reference objects to set their initial configuration or constrain the final solution.

The task oriented architecture is responsible for running the plan solver, which dispatches a task, e.g. detect spinal canal, into sub-tasks.[15] Which sub-task should be dispatched, depends on so-called *reference objects*. $Object_x$ is reference object for $Object_y$ if:

- there is a direct, spatial relation between $Object_x$ and $Object_y$ (e.g. *isNeighbor*, *isInside*).
- $Object_x$ has an segmentation algorithm that does not depend on $Object_y$. Hence, $Object_x$ can be detected without any knowledge of $Object_y$.

When the plan-solver is called with the request *Find* $Object_x$, it identifies the sub-tasks that should be performed in order to fulfill the request, i.e. which objects are reference objects to $Object_x$. The list with reference objects found is the list with the sub-tasks to be performed. The task-planner module relies on a global positioning system (along the axes x, y, z), (Fig. 4) that maps each of the detected organs to world coordinates.

3. Anatomical Structures Map

The anatomical structures map establishes a number of spatial relations between the objects that are typically distinguished in the CT images used for the planning of radiotherapy in our clinic (see Fig. 5). The architecture of the anatomical structures

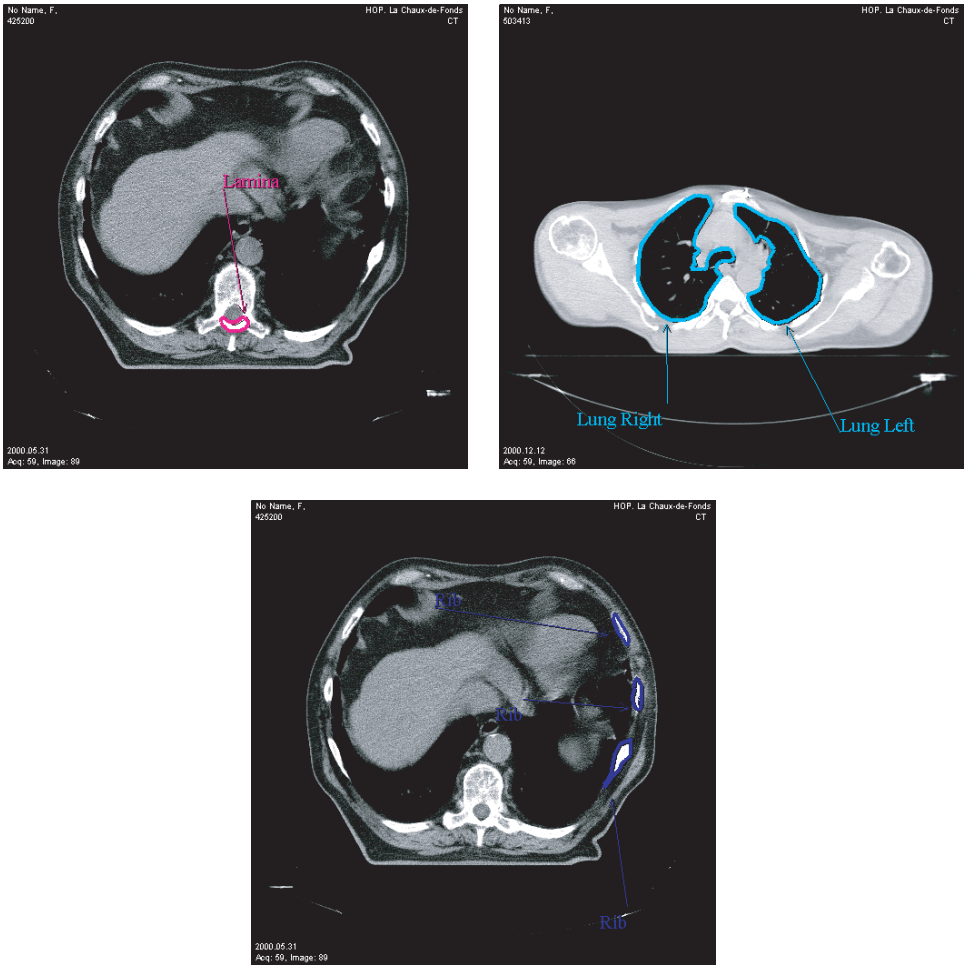


Fig. 3. Samples of reference objects.

map lends its inspiration from frame systems, a well-known concept in the artificial intelligence literature. We choose to represent the anatomical information in 2D slices. More specifically, the ASM represents spatial relations between the typical objects (e.g. spine, lamina and tumor) as well as the general category of each object: bone, air and tissues (see Fig. 2). We discern these particular categories for the following reasons. Objects belonging to the first two categories have either a very low or a very high *HU* level (e.g. the air compartment in a lung versus, e.g. bones). For these two types of objects, a threshold-based technique is in most cases sufficient for a reliable segmentation result. Tissues (e.g. organs), on the other hand, cannot be identified by thresholding within a specific *HU* range. For objects belonging to this third category, a reliable segmentation needs to be based on two

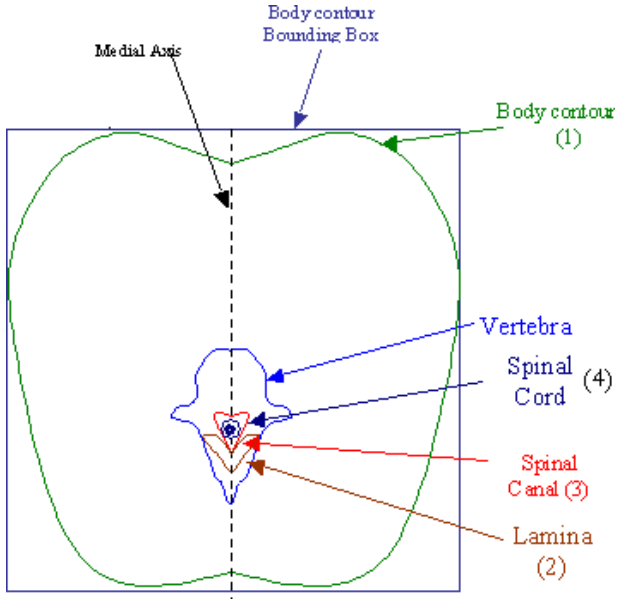


Fig. 4. Contours detected for spinal cord detection. First *Body contour* is detected, than *Lamina*, follow by *Spinal Canal* and finally *Spinal Cord*. Vertebra is added for clarity.

kinds of information: the localizations of the already detected reference objects and the results of texture segmentation.

The main object represented is the *body contour*, which comprises all the other organs. It has a so-called *independent* segmentation scheme as it is possible to detect the body by a basic image processing algorithm, in this case by thresholding.

The structures that are more difficult to segment include the spine, the lamina, the ribs and the spinal canal. The spine contains mainly bone so it has a very high *HU* and thresholding is used to detect it. All the sub-parts of the spine consist of mainly bone cortex so a threshold method is also used to detect these objects. The spinal canal consists mainly of tissue but is completely surrounded by the spine, i.e. *Spinal Canal* IsInside *Spine*. We use a region growing scheme to segment it, mainly because the border of the spinal canal has a high contrast compared with the surrounding bone (difference *HU* bone — tissue).

Finally, we represent the lung information, the ribs and any lung tumors.

4. Plan Solver

The ASM helps to partition a request (e.g. locate spine) into subtasks and further into atomic image processing tasks that are performed by dedicated routines. This hierarchical partitioning takes place in the *plan solver* module, which links the spatial relations in the anatomical structures map (see Fig. 5) with the atomic image

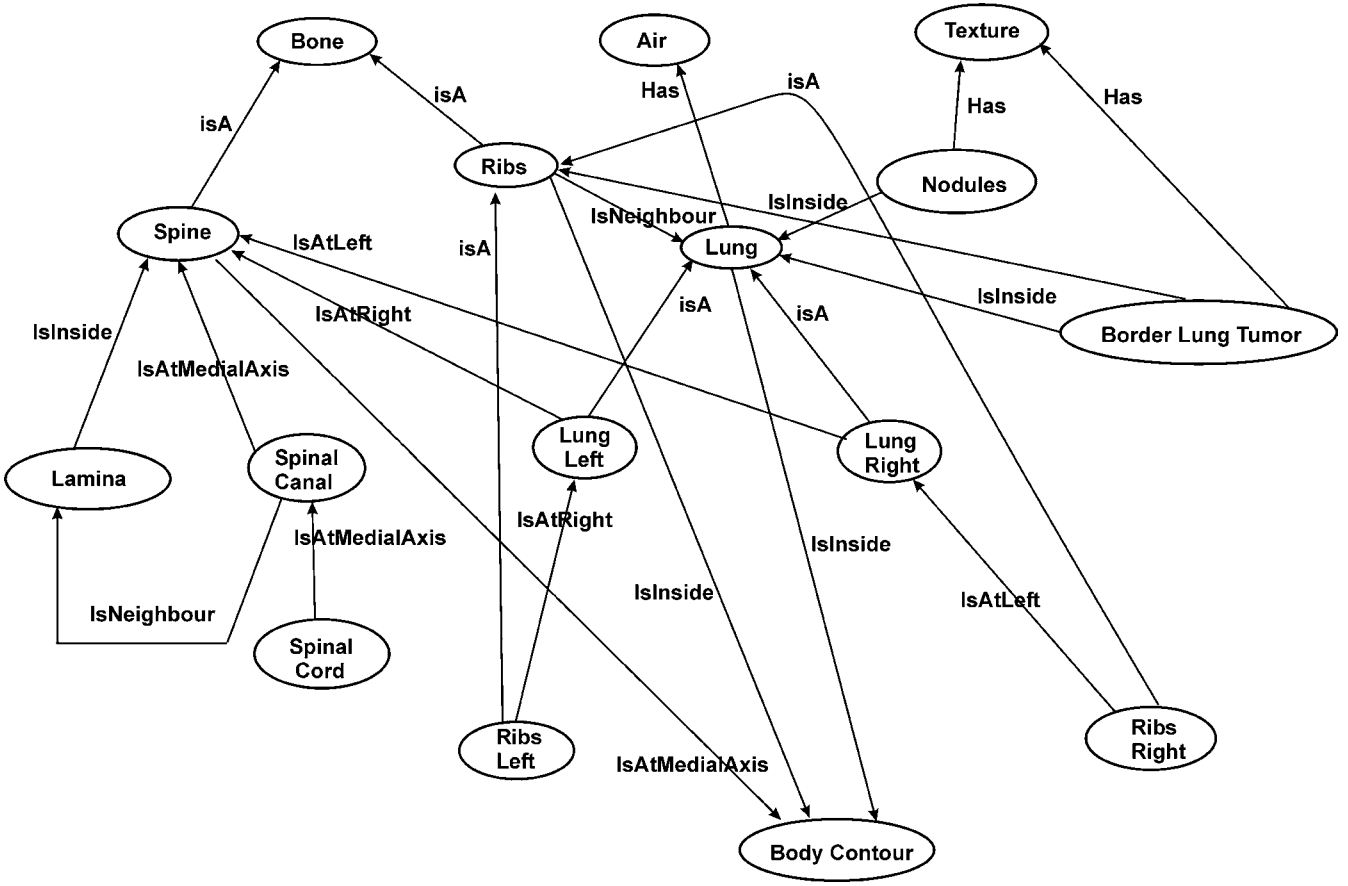


Fig. 5. Relationships between structures.

processing algorithms. The plan solver uses an *inheritance scheme* to determine the appropriate segmentation approach for a particular object or tissue (see Fig. 2). An object connected with another (basic) object by an **isA** relationship inherits the segmentation method of that basic object.

We make a distinction between different types of atomic segmentation methods that are used for object recognition in our application: the threshold-based methods (for the bones in this case, but also for the lungs in the context of lung tumors detection) and region-based methods (for the spinal canal in this case).

For the threshold-based methods, it is important to restrict the area to which they are applied. This is accomplished by using so-called reference objects. Reference objects are specified by the following relations in the anatomical structures map: **isAtLeft**, **isAtRight**, **isInside** and **isSurrounded** by a recursive top-down search.

For region based approaches, the reference objects are found between the objects with the relationship **isNeighbor** or **isVerticalAxis**. When a certain sub request *Find Object*$_x$ is dispatched, the plan solver tries to fulfill the request choosing the appropriate segmentation methods. These are either specified directly (for certain organs like the spinal canal, which is detected by region growing), or indirectly by inheritance from the reference objects by the relationship **isA**. Depending on the chosen segmentation method, the reference objects are found.

The functionality of the plan solver is illustrated by two example requests: *Find Lamina* and *Find Spinal Canal*. The first object *Lamina* does not have its own dedicated segmentation methods so Lamina is found by the inheritance structure based on the link **isA**. *Spine* also does not have its own dedicated segmentation method. Finally, *Lamina* **isA** *Bone* has a thresholding segmentation method attached. As *Lamina* is connected by the link **isA** to *bone* via *Spine*, *Lamina* is segmented by thresholding. The inference mechanism proceeds by looking for the objects linked to *Lamina* by the relation *isInside*. The only object where *Lamina* is inside, is the body (body contour), which this way becomes a reference object for *Lamina*. So the task *Find Lamina* has as sub-tasks: *Find Body Contour* and *thresholding*, the latter takes place only inside the *Spine*.

In the second example, *Find Spinal Canal*, a dedicated segmentation method is specified: region-based segmentation. So we are looking for the objects which could give us a starting point for the region growing algorithm. Thus, we are looking for the objects connected with relationships **isNeighbor**, **isAtVerticalAxis**, which are *Body Contour* and *Lamina*. *Body contour* has its own segmentation scheme, which is why it is the reference object for the *Spinal Canal*. The *Lamina*, as it is presented earlier, has as reference object *Body Contour*, which does not involve the *Spinal Canal*. So the *Lamina* is the second sub-task for the task *Find Spinal Canal*. The general architecture is presented in Fig. 6.

The Algorithm 1 connects the image processing part with the knowledge-base, using the task oriented architecture (presented in Fig. 6).

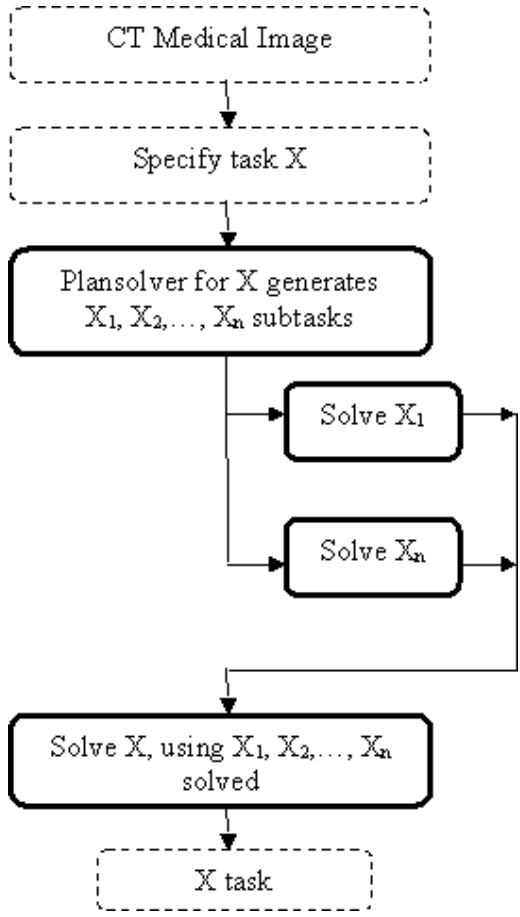


Fig. 6. Plan for a task: General Architecture. Different sub-tasks are executed to solve the original task.

Algorithm 1 DetectContourUsingASM

Require: the structure X to be found
Ensure: the contour of the structure X
1: Search X in the list with the anatomical structures represented in ASM
2: **if** X **not** $found$ **then**
3: *exit* with $failure$
4: **else**
5: **if SegmentationType(X)** = *RegionBased* **then**
6: **RefObjList** ← *Structures Connected* by **isNeighbor**, **isVerticalAxis**
7: **else**

8: **RefObjList** ← *Structures Connected* by **isAtLeft**, **isAtRight**, **isInside**, **isSurrounded**
9: **end if**
10: **for** $\forall X_i \in RefObjList$ **do**
11: **DetectContourUsingASM**(X_i)
12: **end for**
13: *Detect* X *using* $X_i \in RefObjList$
14: **return** X
15: **end if**

5. The Use of Snakes

In cases where the general schema fails, because the spinal canal is not completely surrounded by bone, the system uses a snake-based method.

Snakes were introduced in the literature by Kass,[21] and formulated as an energy-minimizing spline. Given the spline $v(s) = (x(s), y(s))$, he defined the energy function (where s is the pathlength normalized to $[0, 1]$):

$$E_{total} = \int_0^1 E_{\text{int}}(v(s)) + E_{image}(v(s)) + E_{con}(v(s))\, ds \tag{1}$$

where E_{int} represents the internal energy of the spline, composed of a first-order term controlled by $\alpha(s)$, which makes the snake act like a membrane, and the second-order term controlled by $\beta(s)$, making the snake to act like a thin plate.

$$E_{\text{int}} = (\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2)/2\,. \tag{2}$$

E_{image} is given by:

$$E_{image} = -|\nabla I(x, y)|^2 \tag{3}$$

so that the snake is attracted by the contours with large gradients.

Finally E_{con} defines to the external constraint forces given by the user. The problem of initialization of the snake is solved by using the result obtained in the previous slice. The curvature estimation plays an essential role.

To approximate the curvature, the finite difference are used. If $v_i(x_i, y_i)$ is a point on the contour, the following approximations are used:

$$\left|\frac{dv_i}{ds}\right|^2 \approx |v_i - v_{i-1}|^2 = (x_i - x_{i-1})^2 + (y_i - y_{i-1})^2$$

and

$$\left|\frac{d^2v_i}{ds^2}\right| \approx |v_{i-1} - 2v_i + v_{i+1}|^2 = (x_{i-1} - 2x_i + x_{i+1})^2 + (y_{i-1} - 2y_i + y_{i+1})^2\,.$$

In respect to these considerations, we choose a Greedy strategy as in Ref. 37. The technique is shown in Algorithm 2. The input consists of the set of points

which represent the contour to be modified. It improves the quality of the contour using intensity properties of the image. For each point a new location is found in a neighborhood (windows of size w). The new point location is the neighbor which optimizes the energy function.

The algorithm stops when either a specified number of iterations is done, or when the number of point positions changed is sufficiently small. After every iteration, the β coefficients (which control the curvature term in the energy function) are computed to detect the corners.

To find the optimum position for a point in its neighborhood Algorithm 3 is used. In each surrounding point the energy function is approximated. The point where the energy is minimal, is chosen as the new position for the ith point. A number of initial parameters for the snakes needs to be specified, which will control the behavior of the snake. These parameters are usually determined by practical experiments.

Algorithm 2 Snake segmentation

Require: Image I, snakes parameters α, β window size w, contour v, $nrPoints$
Ensure: the contour v modified
1: $G \leftarrow$ **ComputeGradient**(I)
2: $\forall i = 1, nrPoints\ \beta(i) \leftarrow 1$
3: $finish \leftarrow false$
4: **while NOT** $finish$ **do**
5: **for** $i = 1$ to $nrPoints$ **do**
6: **ModifyPoint**(i, newX, newY,α, β, w)
7: $v(i) \leftarrow (newX, newY)$
8: **Evaluate**$\beta()$
9: **end for**
10: **if ConditionsFinishOk**() **then**
11: $finish \leftarrow true$
12: **end if**
13: **end while**
14: **return** v

Algorithm 3 Modify a point on the snake

Require: Image I, the image gradient G, point i to be modified, snakes parameters α, β, γ, window size w, contour v, $nrPoints$
Ensure: newX and newY for the i^{th} point
1: $maxE_{cont} \leftarrow$ **ComputeEcontMax**(i, w)
2: $maxE_{curv} \leftarrow$ **ComputeEcurvMax**(i, w)
3: $maxGradient \leftarrow$ **ComputeMaxGradient**(i, w)

4: $distance \leftarrow \frac{1}{nrPoints} \sum_j ||v_j - v_{j+1}||$
5: $newX \leftarrow 0$
6: $newY \leftarrow 0$
7: $minE_{total} \leftarrow BIG\_VALUE$
8: **for** $\forall\ (k,l) \in wXw$ **do**
9: $E_{cont}(i,k,l) \leftarrow \frac{|distance - ||v_i, I(k,l)|||}{maxE_{cont}}$
10: $E_{curv}(i,k,l) \leftarrow$ **ComputeCurvature**$(\mathbf{v_i}, \mathbf{k}, \mathbf{l})$
11: $E_{curv}(i,k,l) \leftarrow \frac{E_{curv}(i,k,l)}{maxE_{curv}}$
12: $E_{image}(i,k,l) \leftarrow -G(i,k,l)$
13: $E_{total} \leftarrow \alpha \cdot E_{cont}(i,k,l) + \beta(i) \cdot E_{curv}(i,k,l) + \gamma \cdot E_{image}(i,k,l)$
14: **if** $E_{total} < minE_{total}$ **then**
15: $newX \leftarrow k$
16: $newY \leftarrow l$
17: $minE_{total} \leftarrow E_{total}$
18: **end if**
19: **end for**
20: **return** $newX$ and $newY$

6. 3D Processing

Our approach is based on 2D processing of each slice from a CT scanner. It uses local information found in the current slice, and/or information from adjacent slices. The 3D structures are reconstructed from the series of contours obtained in the sequence of slices. This method is similar to the one used now by radiotherapists and radiologists. They detect the first slice where the anatomical structure of interest can be precisely detected. Then, they navigate through the slices to outline the organ of interest, using the contours and information from adjacent slices.

Algorithm 4 is used for 3D structures. Let X be the object to be determined. First, we try to find the first contour of the object X in the medical examination. For this purpose, we use a 2D general procedure based on ASM and plan for a task, presented in Fig. 6. This procedure is used in all the slices until the contour found is reported to be correctly identified. The task of deciding whether or not a contour is correctly identified is performed by the **VerifyCandidate** procedure. For each structure represented in ASM, a procedure is available. It classifies a region detected as being or not being a certain structure. Specific information about the size, intensity, and position is used. A rule-based mechanism decides the correctness of the detection algorithm.

Once a contour is correctly detected in slice k, the algorithm goes into the slices i, $i = \overline{k-1,1}$. In a slice i, the contour from the slice $i+1$ is modified using a snake-based algorithm (lines 11–16).

The system continues analysis in the slices $\overline{k+1, nrTotalSlices}$. In each slice first the 2D method based on ASM (**DetectContourUsingASM**) is used.

The procedure used to determine if the contour detected in slice i ($i = \overline{k+1, nrTotalSlices}$) is correct also uses the contour obtained in the slice $i-1$. The properties of the contour i should not be too different to those of the contour $i-1$. Again, when the **DetectContourUsingASM** procedure fails, the snakes algorithm is used.

7. Problems Solved Using the ASM Approach

Two classes of problems have been solved using this approach. The first one is the detection of organs at risk in radiotherapy planning[4] (e.g. spinal cord). The second class of problems we solved is the detection of *Clinical Volume Targets* in radiotherapy treatment. In our case, lungs have been chosen for study because of the large number of patients who have lung cancer. Lung tumor detection on CT images has been an active field in recent years. Most of the methods focus on nodule identification. The method developed here goes further and another type of metastases, situated at the lung borders, is also detected. The results are published in Refs. 5 and 6. In this chapter, we focus on spinal cord segmentation, one of the most common tasks performed in oncology departments. Details are presented in the following sections.

Algorithm 4 3D detection of structure X

Require: set of 2D slices in the examination, nrTotalSlices
Ensure: the list with X contour identified in all the slices of the exam, $ListXContours$

1: **SetActiveSlice**(1)
2: $XC \leftarrow$ **DetectContourUsingASM**(X)
3: $k \leftarrow 1$
4: **while NOT VerifyCandidat**(X) **do**
5: $k \leftarrow k+1$
6: **SetActiveSlice**(k)
7: $XC \leftarrow$ **DetectContourUsingASM**(X)
8: **end while**
9: /*in the slice k, X is identified*/
10: $ListXContours(k) \leftarrow XC$
11: **for** $j = k-1$ to 1 **do**
12: **SetActiveSlice**(j)
13: $XCNew \leftarrow$ **ModifyUsingSnake** (XC)
14: $XC \leftarrow XCNew$
15: $ListXContours(j) \leftarrow XC$
16: **end for**
17: $XC \leftarrow ListXContours(k)$

18: **for** $i = k + 1$ to $nrTotalSlices$ **do**
19: **SetActiveSlice**(i)
20: $XCNew \leftarrow$ **DetectContourUsingASM**(X)
21: **if NOT VerifyCandidat**$(X, XC, XCNew)$ **then**
22: /**X not correctly identified with ASM, so use snakes*/*
23: $XCNew \leftarrow$ **ModifyUsingSnake** (XC)
24: **end if**
25: $XC \leftarrow XCNew$
26: $ListXContours(i) \leftarrow XC$
27: **end for**

8. Spinal Cord Segmentation

For the spinal cord, the occasional partial appearance of spine around the spinal canal complicates the delineation of its contour. The same 2D segmentation scheme cannot be used in all of the slices. In this section, we first present the 2D segmentation of the spinal cord, which is based on the anatomical structures map and the plan solver, applied to the slices in which the spinal canal is completely surrounded by spine. Subsequently, the procedure responsible for the detection of the 3D spinal canal contour detection process is described. Finally, the methods used in the case of the failure of the standard procedures method (in the slices where the spinal canal is not completely surrounded by spine) are presented (i.e. snakes).

8.1. *2D spinal cord detection based on the ASM*

For the task of identifying the spinal cord contour in a slice, the *plan solver* is dispatched. Its sub-tasks rely on information from the ASM. Figure 7 illustrates how the spinal cord is being detected by our knowledge-based approach. The structures that aid the detection of the spinal cord are body contour, a region of the *spine* (called the *lamina*), and the *spinal canal* (see also Fig. 4).

8.1.1. *Body contour identification*

The transition between the body (contour 1 in Fig. 4) and the outside air is very strong, which makes it rather straightforward to find the contour around the thorax. Moreover, the body is generally the only object in the image. The pixels with a gradient exceeding a threshold value ε are likely to form part of the border between body and air. Based on a correlational analysis of the HU histograms of the body and air in a pilot study, the value of ε is found. Algorithm 5 is used to delineate the contour around the body.

Because of its importance (all of the other structures are *Inside* the body contour), the body contour identification is a sub-task which is performed for every main task.

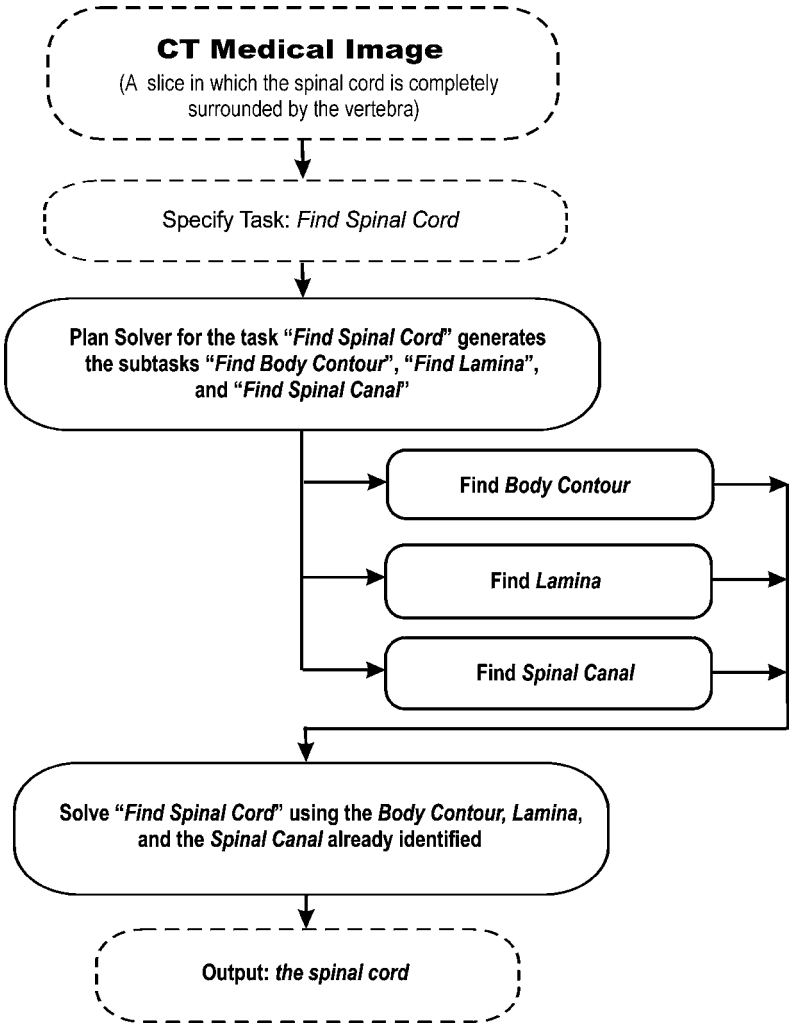


Fig. 7. Plan for a task: the scheme used for spinal cord segmentation.

Algorithm 5 Body Contour Identification

Require: Image I

Ensure: the abdomen contour

1: compute the gradient of the image using a Sobel-like operator;
2: in the middle column of the image, search the first pixel which has the gradient higher than a threshold ε;
3: this is the first point on the body contour;
4: starting from this point, follow in the clock-wise direction the high gradient, until it reaches the first point of the contour.

8.1.2. *Lamina identification*

The lamina contour (Contour 2 in Fig. 4) uses body contour as a reference object. In the ASM, the segmentation scheme associated with the lamina is a threshold operation. The threshold operator is applied to the pixels that occur inside the body contour. The lamina has a very high HU range (650–1200 HU). This is not the only structure with such a high intensity range. Other structures like the *sternum* and the *scapula* might also be detected by the application of a threshold operator. By restricting the threshold operator to a smaller region of the abdomen (centered at the medial axis), thresholding, in most cases, finds the lamina accurately. Algorithm 6 detects the lamina based on the ASM and is a specialization of the general Algorithm 1. The bounding box of the abdomen $((left, top), (right, bottom))$ is used. As the spine is in the middle of the abdomen region and in its inferior half, the threshold window is: $((middleAbdomen - wX, \frac{bottom+top}{2}), (middleAbdomen + wX, bottom))$.

All the pixels having the intensity value in lamina range interval are selected (lines 5–11). Multiple regions can be found. Those having an area larger than a minimal threshold and being positioned on the vertical axis of the abdomen are considered to be candidates for lamina (lines 12–25). Finally, the lowest candidate situated in the abdomen is considered to be lamina, conforming to the medical atlases (lines 26–30).

8.1.3. *Spinal canal detection in 2D*

The contours of the lamina and the body contour are used to detect the spinal canal (Contour 3 in Fig. 4). There is a strong transition from the lamina to the spinal canal (large HU difference bone−tissue). Consequently, a region growing algorithm is used[2]. Two problems are related to the region growing algorithm. First, the homogeneity of the pixel intensities in the region may not be guaranteed. To cope with this problem, a histogram-based method[28] is combined with the *a priori* knowledge about the typical HU range of the spinal canal. A pilot experiment has been performed to find the optimal range of HU values.

A second problem is how to set the seed point — the starting point of the region growing algorithm — automatically. This is accomplished by using the relative locations between body contour, lamina, and in relation to the spinal canal. More specifically, the spinal canal and the spine have the same medial axis (represented by the relationship **isAtMedialAxis**). So, by detection of the lamina (which is **isInside** the spine), the position of the seed point is obtained (being on the medial axis and above the top limit of the lamina). Algorithm 7 is used after Algorithm 6 detects lamina-based on the ASM. Algorithm 1 is again a specialized of Algorithm 1. The seed point is found (lines 7–13) using a smaller window inside the abdomen as a search space (at the vertical axis of the abdomen, near the lamina). The seed pixel intensity value should be in the tissue range. The gradient information is used to detect the border between bone (lamina) and tissue (spinal canal) which is very high. Finally, a region growing algorithm is used.

8.1.4. *Spinal cord detection in 2D*

The problem of spinal cord detection (Contour 4 in Fig. 4) reduces to finding the maximal inscribed circle in the polygon that represents the spinal canal (see also Fig. 8). The problem is solved by computing the medial axis of the polygon using an efficient algorithm (complexity $O(n \ \log \ n)$), which was presented in Refs. 18 and 24.

Algorithm 6 Segment Lamina

Require: Image I, image size N, the abdomen contour
Ensure: *true*, if the lamina contour can be identified and in this case the lamina contour and $false$, otherwise;

$found \leftarrow false$
$R(i,j) = 0, \forall\ i,j = 1 \ldots N$
$((left, top), (right, bottom)) \leftarrow$ the bounding box of the abdomen contour
$middleAbdomenX \leftarrow left + \frac{right-left}{2}$
for $i = middleAbdomenX - wX$ to $middleAbdomenX + wX$ **do**
  for $j = top + \frac{bottom-top}{2}$ to $bottom$ **do**
    if $I(i,j) \in [valMinLamina, valMaxLamina]$ **then**
      $R(i,j) = 1$
    end if
  end for
end for
for $i = middleAbdomenX - wX$ to $middleAbdomenX + wX$ **do**
  for $j = top + \frac{bottom-top}{2}$ to $bottom$ **do**
    if $R(i,j) = 1$ **then**
      build the contour of the region R_i starting from (i,j)
      $middleCandidate \leftarrow left(R_i) + \frac{right(R_i)-left(R_i)}{2}$
      if $air(R_i) > air_\epsilon$ **AND** $|middleCandidate - middleAbdomenX| < \epsilon$ **then**
        add R_i to the list L with candidates for the lamina
      end if
      for $\forall (k,l) InsideR_i$ **do**
        $R(i,j) = 0$
      end for
    end if
  end for
end for
if $||L|| = 0$ **then**
  $found \leftarrow false$
else
  $found \leftarrow true$

30: $LaminaContour = R_i \in L$ with $R_i.Y$ min.
31: **end if**
32: **return** ***found***

Algorithm 7 Spinal Canal Segmentation

Require: Image I, the body contour BC, the lamina contour LC
Ensure: *true*, if the SC can be identified and in this case the SC contour and $false$, otherwise;
1: **ComputeGradient G** of the image **I**
2: $((left, top), (right, bottom)) \leftarrow$ the bounding box of the LC
3: $((la, ta), (ra, ba)) \leftarrow$ the bounding box of the abdomen
4: $XS_i = la + \frac{ra-la}{2}$
5: $YS_i = top + \frac{bottom-top}{2}$
6: $found \leftarrow false$
7: **for** $SearchX = XS_i - wX$ to $XS_i + wX$ **AND not** $found$ **do**
8: **for** $SearchY = YS_i - wY$ to $YS_i + wY$ **AND not** $found$ **do**
9: **if** $I(SearchX, SearchY) \in [valMinSC, valMaxSC]$ **AND** $G(SearchX, SearchY) < \epsilon$ **then**
10: $found \leftarrow true$
11: **end if**
12: **end for**
13: **end for**
14: **if NOT** *found* **then**
15: **return** *false*
16: **else**
17: $SCContour \leftarrow$ **StartRegionGrowing**$(SearchX, SearchY)$
18: **return** *true* **AND** $SCContour$
19: **end if**

8.2. *3D spinal canal detection*

The problem of 3D spinal canal detection is based directly on the procedure for spinal canal detection in 2D, presented in Sec. 8.1.3. However, this scheme cannot be applied successfully to all slices because the spinal canal is not always surrounded by the spine. Instead, the algorithm for 3D spinal canal detection first identifies the spinal canal in each slice using Algorithm 8. Algorithm 8 is a specialization of Algorithm 4. The first step is to apply the 2D algorithm presented in the previous section to identify the spinal canal in the first slice. It uses no information about whether the spinal canal is surrounded completely by bone. A procedure verifies whether the spinal canal was identified correctly (line 4). This procedure uses information about the position, the intensity and the area of the region detected

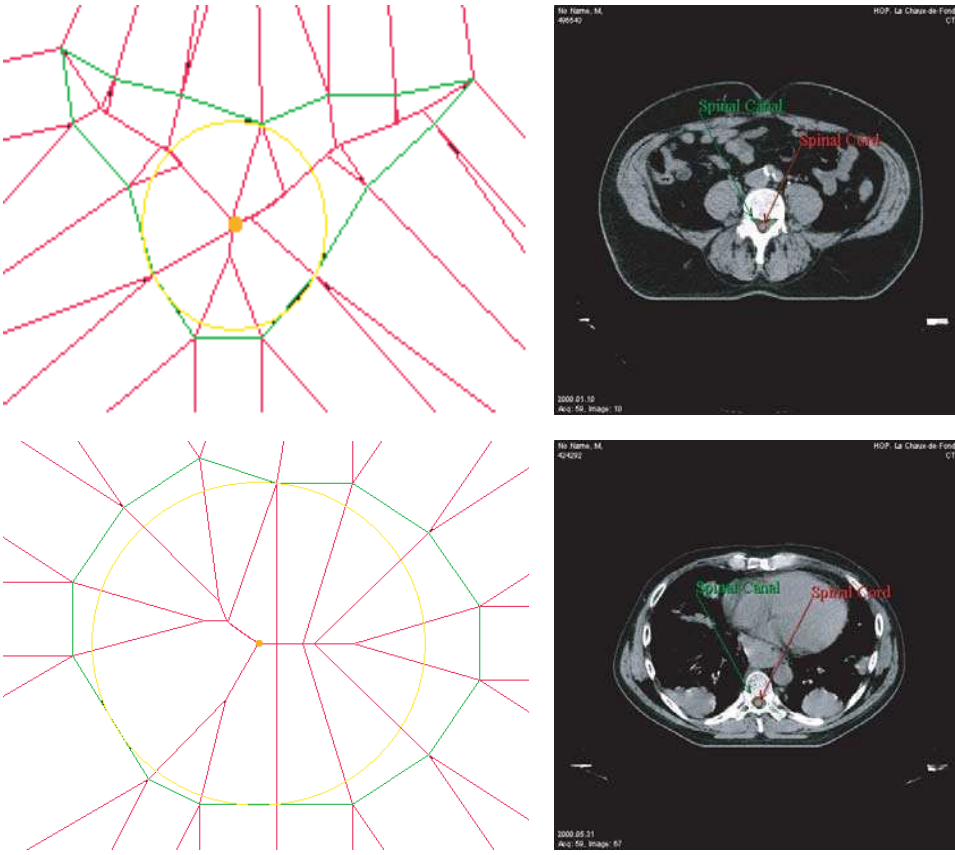


Fig. 8. Maximum inscribed circle in the spinal canal polygon.

by the 2D algorithm. If the algorithm failed to identify the spinal canal correctly in the first slice (see Fig. 9), the same 2D algorithm is applied to the next slices (lines 4–8) until it succeeds in finding the contour of the spinal canal in as many slices as possible.

Once a contour around the spinal canal has been found, the algorithm uses it as reference in the neighboring slices in two ways. First, it can be used to verify the candidate contour for the spinal canal in the immediately adjacent slices (lines 14, 28). The second way is to use a spinal canal contour as information to guide the segmentation scheme in the adjacent slice (in case the 2D algorithm fails to correctly identify the spinal canal — lines 13, 27). These two modalities of using a contour already identified are presented in the next section. In the k'th slice, the contour of the spinal canal is detected (line 10). The 3D algorithm proceeds from the $k-1 \rightarrow 1$ and $k+1 \rightarrow nrTotalSlices$ slices, applying the 2D detection algorithm. In case of failure, it chooses one of the alternative methods presented in the next section. The evolution of the algorithm in two consecutive slices is illustrated in Fig. 11.

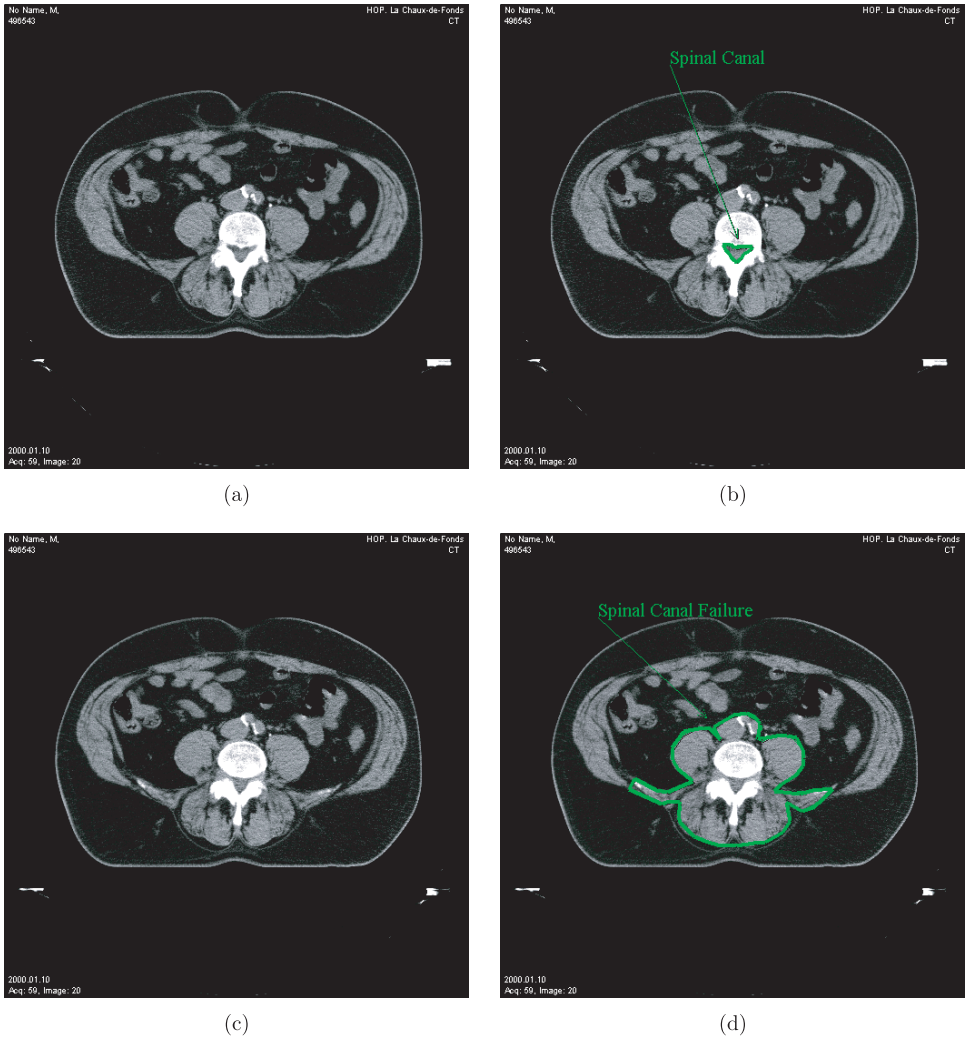


Fig. 9. The two types of slices: (a) spinal canal completely surrounded by bone; (b) the method based on ASM works well for this type of slice; (c) spinal canal partially surrounded by bone; (d) the method based on ASM fails for this type of slice.

Two procedures are used to check the results of the spinal cord detection algorithms. The first one (line 4), *VerifyCandidateSpinalCanal* (SCC), uses the *a priori* knowledge about spinal canal region: *position*, *area*, *intensity*, and *shape*. If the properties of the candidate respect the pre-defined parameters of our model, the region is recognized as *spinal canal*, otherwise it is rejected.

The second procedure (line 14), *VerifyCandidateSpinalCanal* (SCC, SCCNew), uses a contour obtained in an adjacent slice, against which it verifies the properties of the new contour detected in the current slice. If the differences between the two

Algorithm 8 3D detection of spinal canal

Require: set of 2D slices in the examination, nrTotalSlices
Ensure: a list (*ListSCContours*) of spinal canal contours (*SCC*) identified in all slices of the examination

SetActiveSlice(1)
SCC ← **DetectSpinalCanalContourUsingASM**()
$k \leftarrow 1$
while NOT VerifyCandidateSpinalCanal(SCC) **do**
  $k \leftarrow k + 1$
  SetActiveSlice(k)
  SCC ← **DetectSpinalCanalContourUsingASM**()
end while
/*in the slice k, spinal canal is identified*/
ListSCContours(k) ← *SCC*
for $j = k - 1$ to 1 **do**
  SetActiveSlice(j)
  SCCNew ← **DetectSpinalCanalContour(SCC)**
  if NOT VerifyCandidateSpinalCanal(*SCC*, *SCCNew*) **then**
    /*spinal canal not correctly identified so use snakes*/
    SCCNew ← **ModifyUsingSnake** (*SCC*)
  end if
  SCC ← *SCCNew*
  ListSCContours(j) ← *SCC*
end for
SCC ← *ListSCContours*(k)
for $i = k + 1$ to *nrTotalSlices* **do**
  SetActiveSlice(i)
  SCCNew ← **DetectSpinalCanalContourUsingASM**()
  if NOT VerifyCandidateSpinalCanal(*SCCNew*,*SCC*) **then**
    /*spinal canal not correctly identified with ASM*/
    SCCNew ← **DetectSpinalCanalContour(SCC)**
    if NOT VerifyCandidateSpinalCanal(*SCC*,*SCCNew*) **then**
      /*spinal canal not correctly identified so use again snakes*/
      SCCNew ← **ModifyUsingSnake** (*SCC*)
    end if
  end if
  SCC ← *SCCNew*
  ListSCContours(i) ← *SCC*
end for
return *ListSCContours*

contours with respect to: *position*, *area*, *intensity* and *shape* are smaller than a set of pre-specified ranges, the new candidate contour is recognized as being the *spinal canal*.

8.3. *When segmentation of the spinal canal fails*

In case the general scheme based on ASM and plan for a task fails to detect an appropriate contour around the spinal canal, the system backtracks and uses either a region-based method or snakes. Both use the (already approved) contour around the spinal canal detected in an adjacent slice for initialization.

8.3.1. *Finding the spinal canal by region growing*

Occasionally, the general scheme for detection of the spinal canal fails because it cannot identify the *lamina* region, even when the spinal canal is completely surrounded by bone cortex. In these cases, a region-based segmentation technique works well and is applied to the slice (lines 13, 27). The problem is to find the seed point for the region growing process.

We use the center of gravity of the spinal canal region identified in an adjacent slice (as in Fig. 10), thereby assuming spatial continuity of the spinal canal. To compute the center of gravity of a region given by a function f, moments (m_{pq}) are used which are defined for the continued case as:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x,y) dx\, dy \quad with\ p,q = 0,1,2,\ldots \tag{4}$$

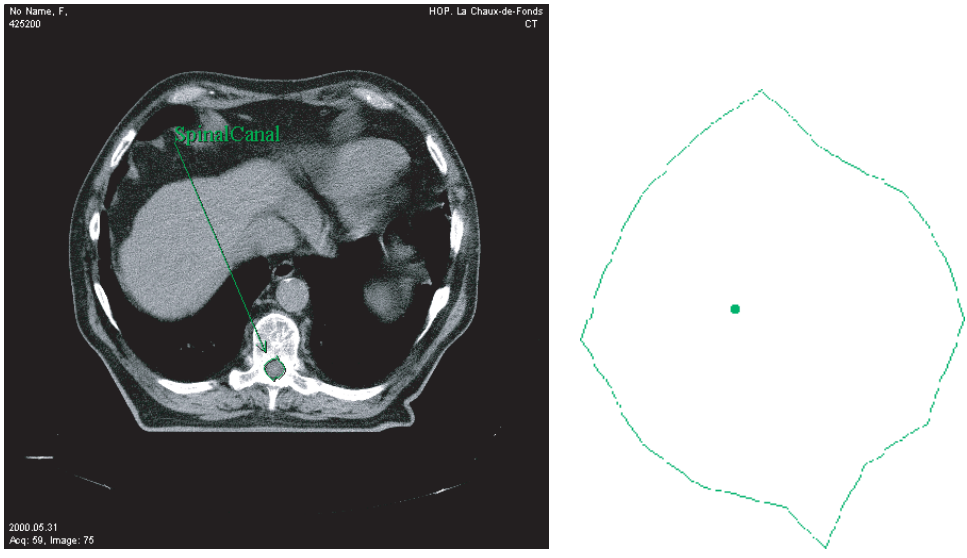


Fig. 10. The spinal cord contour and its gravity center.

and for the discrete case as:

$$m_{pq} = \sum_i \sum_j i^p j^q f(i,j) \quad with\ p,q = 0,1,2,\ldots .$$

Thus the center of gravity is defined as:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad and \quad \bar{y} = \frac{m_{01}}{m_{00}} .$$

8.3.2. *The use of snakes*

In the case where the general scheme fails to detect the spinal canal, for example because the spinal canal is not completely surrounded by spine (see Fig. 9), the system uses a snake-based method.[21]

8.4. *Experimental results*

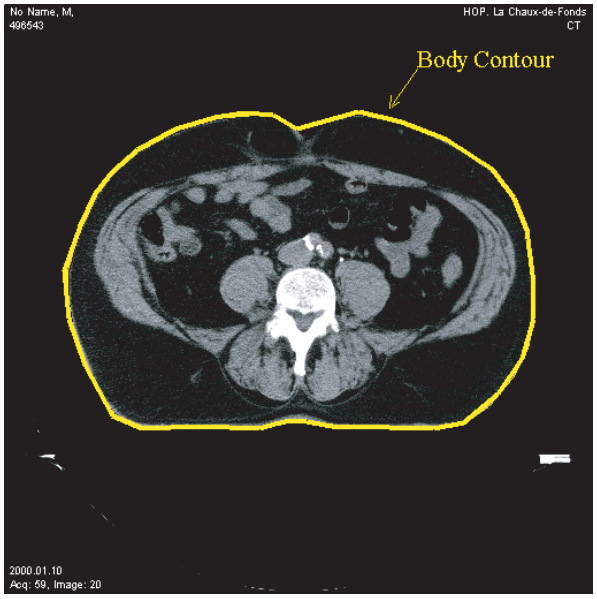
The experimental data consists of 23 patient examinations performed with a Picker CT medical scanner. The slice thickness is 3 mm, and the interslice distance is 3 mm. The data was provided by *La Chaux de Fonds* Hospital (Switzerland). The radiologists and radiotherapists of *La Chaux de Fonds* Hospital also helped with the clinical interpretation of these images. After the CT examination, each patient underwent radiotherapy in the hospital. Our population consisted of 9 male and 14 female patients. Their ages varied from 37 to 79 years with a mean of 58 years and a median of 59 years. The number of CT slices per exam varied from 9 to 97 with a mean of 45 slices and a median of 38 slices. The images were obtained from a CT scanner from Picker and were acquired with a slice thickness of 3 mm and an inter-slice distance of 3 mm.

We distinguished four types of contour: *spinal cord*, *spinal canal*, *lamina* and *body contour*. Whereas the computational time is straightforward to compute, medical expertise is needed to assess the contours that were found by our system. A radiologist skilled in radiotherapy planning was asked to accept or reject each contour in each slice among all 23 patients. In our case, evaluation was performed using a visual inspection of the contours projected on the CT image slices. The radiologist decides for each of the contours obtained with our system whether it is acceptable or not.

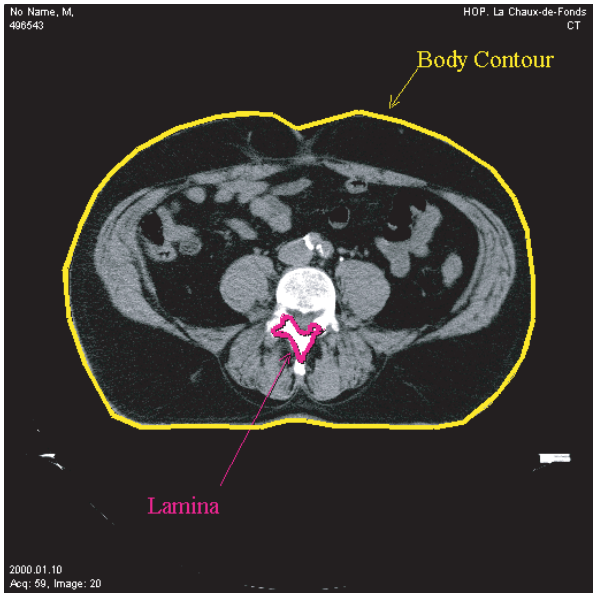
In Figs. 11 and 12, the different steps implied by the spinal cord segmentation are illustrated.

8.4.1. *Accuracy*

In Table 2, results of the experiments on the real clinical data are shown. The accuracy is percentage of the slices in the exam in which the radiologist agreed that the particular contour was located correctly. In Exam 1, for example, 91.9% of the contours were located around the spinal cord with a sufficient precision.

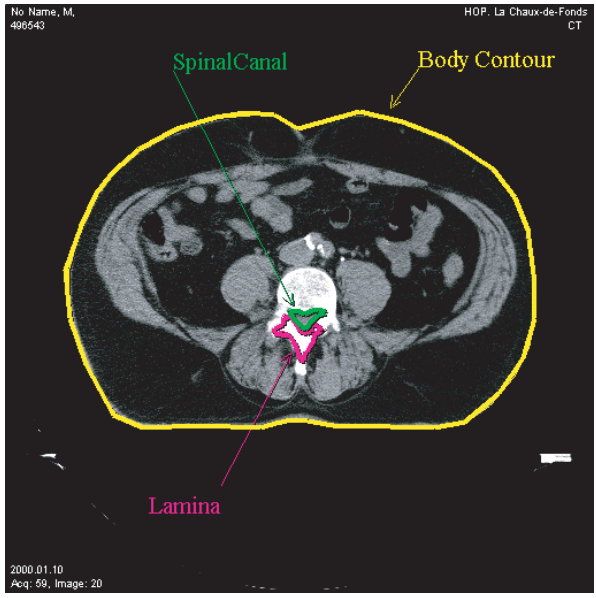


(a)

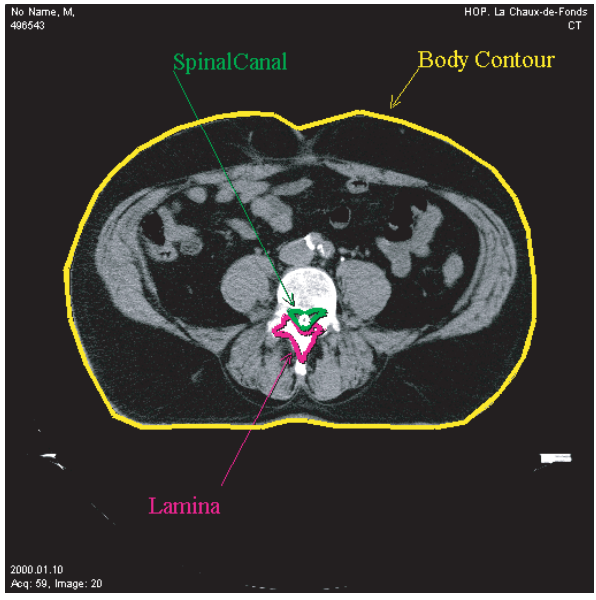


(b)

Fig. 11. The algorithm for spinal cord identification (a) in the first slice identify the body contour (b) find a *lamina* (c) using the bone part identified, find the seed for the Region Growing, which identifies the spinal canal (d) apply *Minimum Inscribed Circle* to find the spinal cord (e) propagate the spinal canal contour in the next slice and improve it using snakes (f) apply again *Minimum Inscribed Circle* to find the spinal cord.

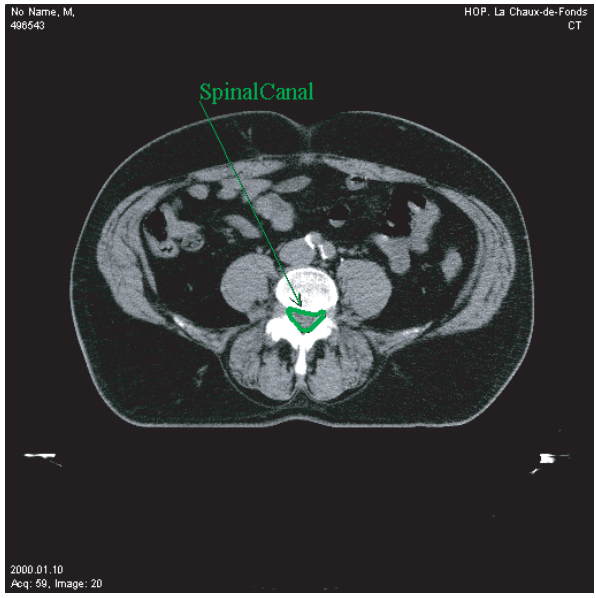


(c)

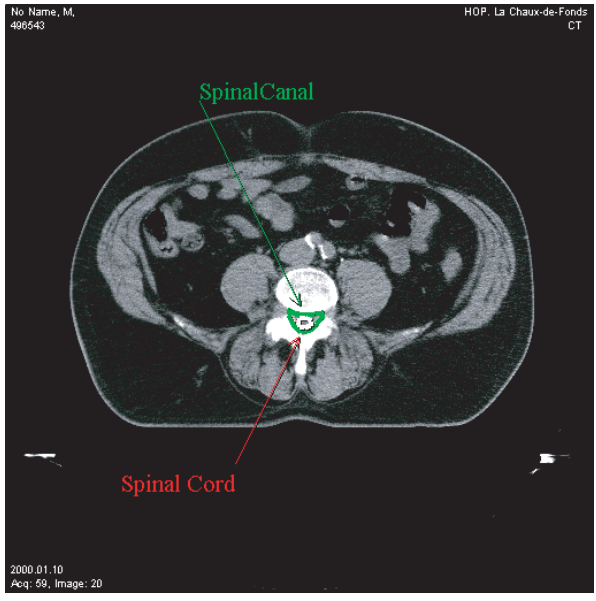


(d)

Fig. 11. (*Continued*)



(e)



(f)

Fig. 11. (*Continued*)

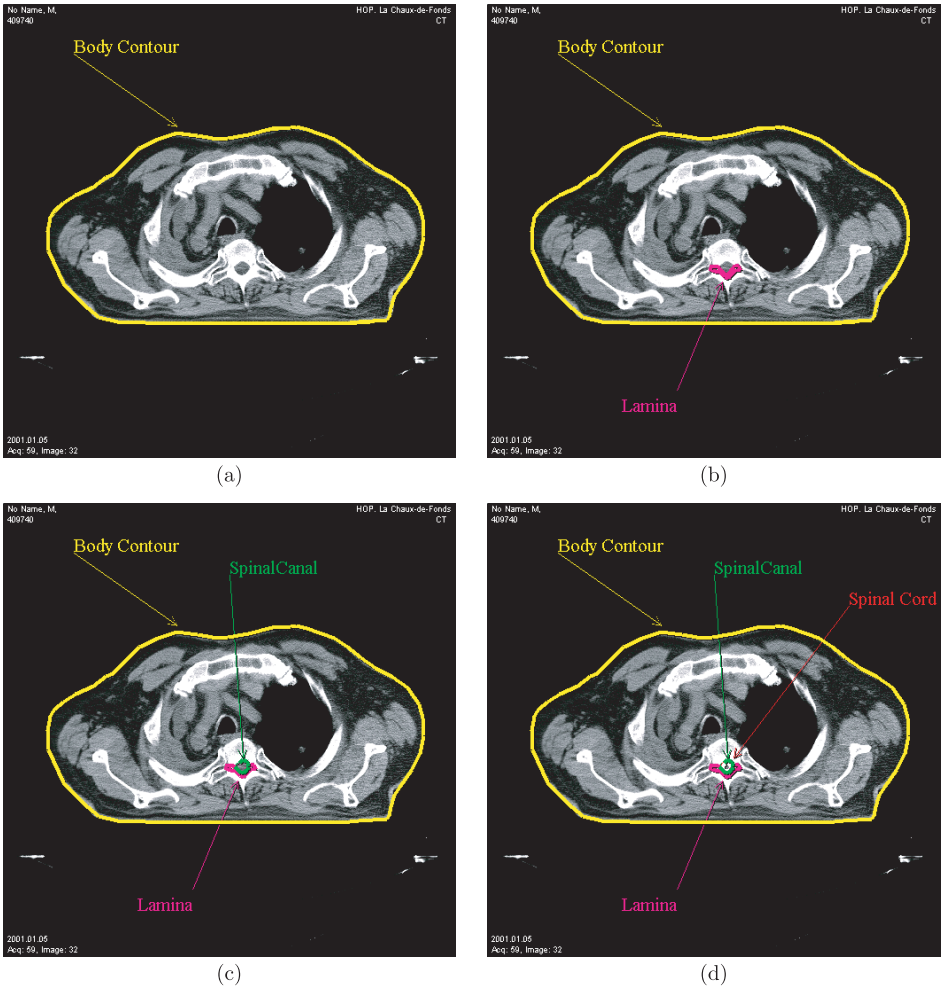
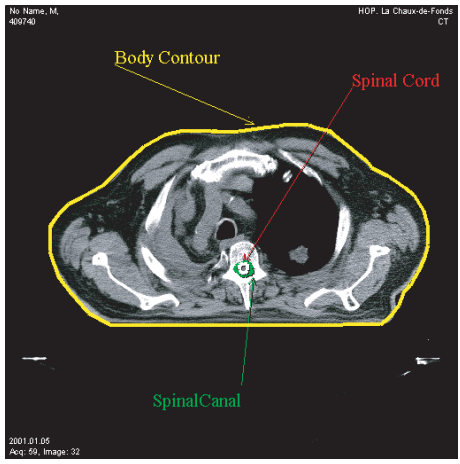


Fig. 12. The algorithm for spinal cord identification in the case when the images are affected by abnormalities (right lung not visible) (a) in the first slice identify the body contour (b) find a *spine part* (c) using the bone part identified, find the seed for the Region Growing, which identifies the spinal canal (d) apply *Minimum Inscribed Circle* to find the spinal cord (e) The process continues in the next slice.

The average accuracy of the spinal cord contours among all patients is 91.7%, the average accuracy per slice lies within the range of 80% to 100%. The spinal canal is more difficult to detect. The average detection accuracy among all patients is 85.3%, the average accuracy per slice lies within the range of 60% to 100%. The lamina is the most difficult structure to detect for our approach. The average accuracy among the 23 patients is 72.1%, and the range is 33% to 100%. Finally, the body contour is located correctly in all slices among all 23 patients. The body is easy to detect because of the sharp transition from the surrounding air to human tissue.



(e)

Fig. 12. (*Continued*)

Among these four structures, the accuracy of each contour is correlated for the spinal canal and spinal cord (0.594). The correlation coefficients between the other types of contours are all below 0.15.

In general, when contours were detected wrongly, the major cause was the mislabeling of the neighboring *reference objects*, such as the spinal canal. The problems, in most cases, arise in CT exams where the standard acquisition protocol has not been followed such that one or more unexpected objects (e.g. arms) were present in the slice. The presence of such objects affects the symmetry of the body.

8.5. *Computational cost*

Our algorithms are implemented on a PC Windows machine, with processor Pentium III 500 MHz, 512 MB RAM. As presented in Table 2, the spinal cord detection is done in one second per slice. The worst case is when the snakes are used in all the slices. The interactivity is minimal and might consist only of human correction of the errors in a couple of slices. These times are not recorded.

In the ASM, about 10 different structures are represented so the query process performed by the *plan solver* terminates quickly. The most time-consuming routine is the snake algorithm, which optimizes the location of the contour by minimizing the total energy. All other routines are performed in less than a second per slice (0.3–0.5 seconds) in our current application. The snake algorithm is only applied to slices where the spinal cord is not surrounded by the spinal canal. This happens in about half of the slices.

9. Discussion

The major contribution of this chapter is to demonstrate that a top-down knowledge-based system can provide flexible interpretation of CT images. Our

Table 2.

Medical Exam ID	No. of Slices	Age (years)	Sex	Sp. Cord		Sp. Canal		Lamina		Body Contour	
				Succ.	*Texec*	*Succ.*	*Texec*	*Succ.*	*Texec*	*Succ.*	*Texec*
Exam 1	37	68	F	91.9%	65s	91.89%	62s	48.64%	42s	100%	34s
Exam 2	87	60	M	94.26%	153s	93.10%	148s	98.85%	92s	100%	73s
Exam 3	27	58	F	88.89%	40s	85.18%	39s	74.07%	25s	100%	18s
Exam 4	70	62	M	91.43%	108s	90%	105s	62.85%	68s	100%	52s
Exam 5	87	59	F	93.11%	177s	89.65%	173s	82.75%	112s	100%	90s
Exam 6	15	61	M	86.7%	31s	80%	30s	100%	16s	100%	13s
Exam 7	31	50	F	93.55%	62s	87.09%	59s	80.64%	25s	100%	18s
Exam 8	9	54	F	100%	12s	100%	12s	100%	7s	100%	5s
Exam 9	22	51	F	95.45%	55s	95.45%	55s	77.27%	17s	100%	13s
Exam 10	97	53	F	92.78%	135s	91.75%	132s	77.31%	83s	100%	61s
Exam 11	38	48	F	92.10%	61s	84.21%	57s	78.94%	38s	100%	32s
Exam 12	22	68	F	95.45%	34s	81.81%	33s	90.90%	18s	100%	14s
Exam 13	20	52	F	95%	38s	60%	38s	80%	16s	100%	11s
Exam 14	75	63	M	80%	125s	62.66%	122s	45.33%	78s	100%	60s
Exam 15	22	66	F	95.45%	35s	90.90%	35s	86.36%	20s	100%	14s
Exam 16	50	37	F	88%	99s	82%	98s	76%	48s	100%	37s
Exam 17	18	43	F	88.88%	28s	88.88%	27s	83.33%	17s	100%	13s
Exam 18	55	62	M	98.18%	94s	92.72%	93s	40%	45s	100%	32s
Exam 19	83	70	M	93.97%	141s	90.36%	139s	36.14%	86s	100%	69s
Exam 20	45	64	M	95.55%	72s	86.67%	70s	33.33%	45s	100%	35s
Exam 21	40	56	F	87.5%	69s	80%	68s	95%	43s	100%	34s
Exam 22	44	79	M	84.09%	77s	79.54%	76s	54.54%	43s	100%	33s
Exam 23	37	55	M	86.48%	66s	78.37%	64s	56.75%	44s	100%	35s

system can cope with a large amount of inter-patient variation. Not only is the size and shape of the tumor unique to each patient, also, the acquisition parameters of the CT scan vary considerably. The ASM is a means to represent, in a compact form, important fragments of the anatomic knowledge that is used by a radiologist while interpreting the CT images. The *plan solver* contains the procedural knowledge: how to detect particular anatomical structures one-by-one. The knowledge-based architecture often makes it possible to cope with exceptional conditions which occur in a specialized clinic. Even if our approach fails, it is possible for the radiologist to "take over" and correct the wrongly placed contours of, for example, the lamina. Moreover, it is possible to identify the cause(s) of failure because of the transparent knowledge-based architecture.

The major drawback of our approach is the time it takes to formalize the anatomical and procedural knowledge that is needed to implement the ASM and the *plan solver*. This problem is well-known in the field of expert systems and has been called the *knowledge elicitation bottleneck*. Although the ASM may partly be reused, for example for automatic interpretation of MR images of the thorax, reverse engineering would be required to tailor the *plan solver* such that it applies the appropriate atomic (segmentation) algorithms. It is clear that a different image modality will, in general, require different low-level operators to find the same anatomic structures. We furthermore wish to add that the rather confined macro-anatomy of the human thorax makes it well-suited for representation in a frame-like hierarchical representation scheme such as the ASM. Representation of other regions, e.g. the human vascular system would be more difficult.

A final question addresses the transferability of the developed method to other hospitals. The *DICOM* (Digital Imaging and Communications in Medicine) standard (is used by our system). With the *DICOM* standard, the use of *HU* should work in a hospital with well-calibrated CT scanners. About the features of the system, probably the most important one is the use of a knowledge-based image processing philosophy keeping in mind that relying solely on the classical image processing algorithms is insufficient for detecting automatically anatomical structures in CT images. Another important aspect of this approach is the way the 3D processing is performed, resembling the procedure performed by the radiologist.

10. Summary and Conclusions

Radiotherapy of malignant tumors located in the vicinity of the spinal cord requires very accurate planning to avoid causing unnecessary damage in this vital organ. The spinal cord is a highly radiosensitive organ; even moderate doses of radiation can cause different complications such as paralysis of the patient. In this chapter, we present a knowledge-based approach to the interpretation of CT images. The approach is based on two closely linked knowledge bases: the *anatomical structures map* and the *plan solver*. The former represents structural (static) knowledge of the macro anatomy in the human thorax. The latter represents the procedural

knowledge — the scripts — that are used for the detection of the different objects of interest. The plan solver combines atomic and composite image processing operators using an inheritance scheme. Which (composite) operators inherit an atomic operator, say a snake algorithm, is derived from the anatomical structures map.

The knowledge-based approach was implemented on a standard PC. The system was subsequently validated on CT image data from 23 patients who were to undergo radiotherapy. The plan solver was used to locate the following four kinds of objects: the spinal cord, the spinal canal, the lamina and the body (outer thorax). The highest accuracy was obtained for the body contour, which was located correctly in all slices among the 23 patients. The spinal cord was located with an accuracy of 92%, the spinal canal with an accuracy of 85% and the lamina with an accuracy of 72%.

The major advantage of our knowledge-based system, compared with state-of-the-art low-level solutions, lies in its transparency and its flexibility. The system is transparent to the radiologist because parts of his/her medical knowledge are represented in the anatomical structures map and the plan solver. Transparency makes it easier to take over from the system in case the identification of the objects fails. Flexibility is required because the scan protocol varies among the patients, depending on the location and size of the tumor.

References

1. *Prescribing, Recording and Reporting Photon Beam Therapy*, International Commission on Radiation Units and Measurements, 7910 Woodmont Avenue Bethesda, Maryland 20814, USA, 1993.
2. R. Adams and L. Bischof, Seeded region growing, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**, 6 (1994) 641–647.
3. N. Archip and P.-J. Erard, Anatomical structures map — A way of encoding medical knowledge in a task oriented computed tomography image analysis, in *Proceedings of The 2000 International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences*, Las Vegas, Nevada, USA, June 2000, pp. 377–388.
4. N. Archip, P. J. Erard, M. Egmont-Petersen, J.-M. Haefliger and J. Germond, A knowledge-based approach for automatic detection of spinal cord in ct images, *IEEE Transactions on Medical Imaging* **21** (2002) 12.
5. N. Archip, P. J. Erard, J.-M. Haefliger and J. Germond, A computer aided diagnostic system for radiotherapy planning, *Journal of Medical Physics* **12** (2002) 252–259.
6. N. Archip, P. J. Erard, J.-M. Haefliger and J. Germond, Lung metastaces detection and visualization on ct images — A knowledge-based method, *Journal of Visualization and Computer Animation* **13** (2002) 1.
7. H. K. Awwad, *Radiobiological and Psysiological Perspectives — The Boundary Zone Between Clinical Radiotherapy and Fundamental Radiobiology and Physiology* (Kluwert Academic Publishers, 1990).
8. J. Bijhold, K. G. A. Gilhuijs, M. vanHerk and H. Meertens, Radiation-field edge-detection in portal images, *Physics in Medicine and Biology* **36**, 12 (1991) 1705–1710.

9. M. S. Brown *et al.*, Knowledge-based segmentation of thoracic computed tomography images for assessment of split lung function, *Med. Phys.* **27**, 3 (2000) 592–598.
10. K. De Geus and A. Watt, Three dimensional stylization of structures of interest from computed tomography images applied to radiotherapy planning, *Int. J. Radiation Oncology Biol. Phys.* **35** (1996) 151–159.
11. M. Egmont-Petersen, Mental models as cognitive entities, in *Proceedings of the Scandinavian Conference on Artificial Intelligence*, ed. B. Mayoh (IOS Press, Amsterdam, 1991), pp. 205–210.
12. M. Egmont-Petersen, P. C. W. Hogendoorn, R. v. d. Geest, H. A. Vrooman, H. J. v. d. Woude, J. P. Janssen, J. L. Bloem and J. H. C. Reiber, Detection of areas with viable remnant tumor in postchemotherapy patients with ewing's sarcoma by dynamic contrast-enhanced mri using pharmacokinetic modeling, *Magnetic Resonance Imaging* **15**, 5 (2000) 525–535.
13. K. Englmeier *et al.*, Model based image interpretation of spiral ct scans of the abdomen, *Artificial Intelligence in Medicine*, 1993, pp. 44–51.
14. K. G. A. Gilhuijs, P. J. H. vandeVen and M. vanHerk, Automatic three-dimensional inspection of patient setup in radiation therapy using portal images, simulator images, and computed tomography data, *Medical Physics* **23**, 3 (1996) 389–399.
15. L. Gong and C. A. Kulikowski, Composition of image analysis processes through object-centered hierarchical, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17** (1995) 10.
16. M. J. F. Grimnes, *ImageCreek: A knowledge level approach to case-based image interpretation*, PhD thesis, University of Trondheim, 1998.
17. E. J. Hall, *Radiobiology for the Radiologist* (Lippincott, Williams & Wilkins, 2000).
18. M. Held, Vroni: An engineering approach to the reliable and efficient computation of voronoi diagrams of points and line segments, *Computational Geometry* **18** (2001) 95–123.
19. P. Jackson, *Introduction to Expert Systems* (Addison-Wesley, 1999).
20. L. Jennifer, T. E. Boes, C. Weymouth and R. Meyer, Multiple organ definition in ct using a bayesian approach for 3d model fitting, in *Vision Geometry IV, Proc. SPIE 2573*, 1995, pp. 244–251.
21. M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active contour models, *International Journal of Computer Vision*, 1988, pp. 321–331.
22. V. P. Kumar and U. B. Desai, Image interpretation using bayesian networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18** (1996) 1.
23. F. Kummert *et al.*, Control and explanation in a signal understanding environment, *Signal Processing* **32** (1993) 111–145.
24. D. T. Lee, Medial axis transformation of a planar shape, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **4** (1982) 4.
25. S. A. Leibel and T. L. Philips, *Textbook of Radiation Oncology* (W.B. Saunders Company, 1998).
26. H. Li, R. Deklerck, B. De Cyper, A. Hermanus, E. Nyssen and J. Cornelius, Object recognition in brain ct-scans: Knowledge based fusion of data from multiple feature extractors, *IEEE Transactions on Medical Imaging* **14**, 2 (June 1995) 343–360.
27. W. Menhardt *et al.*, Knowledge based interpretation of cranial mr images, Technical report, Philips GmbH, Forschungslaboratorium Hamburg, POBox 540840 D-2000 Hamburg 54, 1990.
28. N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Systems, Man, and Cybernetics* **9** (1979) 1.

29. J. A. Purdy and G. Starkschall, *A Practical Guide to 3D Planning and Conformal Radiation Therapy* (Advanced Medical Publishing, Inc., 1999).
30. E. Rogers, Via-rad: A blackboard-based system for diagnostic radiology, *Artificial Intelligence in Medicine* **7** (1995) 343–360.
31. M. Sonka, G. Sundaramoorthy and E. A. Hoffman, Knowledge-based segmentation of intrathoracic airways from multidimensional high resolution ct images, *Medical Imaging 1994: Physiology and Function from Multidimensional Images*, 1994, pp. 73–85.
32. G. M. te Brake and N. Karssemeijer, Segmentation of suspicious densities in digital mammograms, *Medical Physics* **28**, 2 (2001) 259–266.
33. R. Tombropoulos *et al.*, A decision aid for diagnosis of liver lesions on mri. Technical report, Section on Medical Informatics, Stanford University School of Medicine, 1994.
34. B. van Ginneken and B. M. ter Haar Romeny, Automatic segmentation of lung fields in chest radiographs, *Medical Physics* **27**, 10 (2000) 2445–2455.
35. G. L. Vernazza *et al.*, A knowledge-based system for biomedical image processing and recognition, *IEEE Trans. Circuits and Systems* **34** (1987) 11.
36. O. Wegener, *Whole Body Computed Tomography*, 2nd edn. (Blackwell Scientific, 1992).
37. D. J. Williams and M. Shah, A fast algorithm for active contours and curvature estimation, *CVGIP: Image Understanding* **55**, 1 (1992) 14–26.
38. A. Zahalka and A. Fenster, An automated segmentation method for three-dimensional carotid ultrasound images, *Physics in Medicine and Biology* **46** (2001) 1321–1342.

CHAPTER 5

FROM GLOBAL TO LOCAL APPROACHES FOR NON-RIGID REGISTRATION

ISABELLE COROUGE*, PIERRE HELLIER and CHRISTIAN BARILLOT
IRISA/VisAGeS, CNRS, INRIA, INSERM and University of Rennes I, Project-Team, Campus de Beaulieu, 35042 Rennes Cx, France
** corouge@unc.edu*

A major issue in neuroscience is to carry out the cartography of cerebral functions under normal and pathological conditions. The establishment of such functional maps motivates the investigation of the anatomy-function relationship and the development of registration methods that map data collected from different modalities and from different subjects. In such a process of anatomical and functional fusion, the modeling of the cortical ribbon, especially of anatomical and functional landmarks such as sulci, is of great interest. An increasing number of methods have been proposed to solve this inter-individual image fusion problem. These methods are generally divided into two groups: "photometric" or global methods, that rely on the matching of voxels having the same luminance, and "geometric" or local methods that rely on the matching of appropriate landmarks. Another alternative to both methods is to use hybrid registration schemes in order to find a mid-way between the global and local approaches, taking advantage of the automaticity of the global one and the regional accuracy of the local one. In this chapter, we describe a typical example of each of these approaches. First, we present a global deformable registration method using image intensities. Second, we use a non-linear local registration framework and third, an hybrid approach able to locally control global deformations by sparse geometric landmarks. Finally, we propose a comparative evaluation methodology of registration frameworks on anatomical and functional data.

Keywords: 3D cerebral imaging; anatomical and functional atlases; cortical sulci; spatial normalization; nonlinear registration; cortical constraints; statistical shape model; evaluation.

1. Introduction

One objective of cognitive or clinical neuroscience is to carry out the cartography of cerebral functions under normal and pathological conditions. Researches are currently performed to find correlations between brain anatomical structures, essentially sulci and gyri, where neuronal activation takes place, and cerebral functions, as assessed by recordings obtained by means of various neuroimaging modalities, such as positron emission tomography (PET), functional magnetic resonance imaging (fMRI), electro-encephalography (EEG) and magneto-encephalography (MEG). One of the major issues related to the establishment of these functional maps is

*Current address: Department of Computer Science, University of North Carolina, CB #3175 Sitterson Hall, Chapel Hill, NC 27599, USA.

the mapping of data collected from different modalities and from different subjects. This matching generally relies on an anatomical substrate, usually viewed by the means of magnetic resonance images (MRI). MRI allows to obtain some 3D precise representations of cerebral anatomy and to easily visualize the cerebral cortex and its components gyri and sulci, place of (supposed) presence of many studied cognitive functions. Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modeling of anatomical structures and multi-modality fusion, make possible this ambitious goal of neuroscientists. Several arguments point out the interest of modeling cortical ribbon, especially sulci, in a process of anatomical and functional fusion:

- Cortical sulci serve as landmarks associated to major functional regions.
- Cortical sulci can be used as landmarks and so as a support for inter-individual fusion processes.
- Cortical sulci can help for the labeling of cortical structures (gyri for instance).
- Cortical sulci can be used as indices to some functional organizations.

An increasing number of methods have been proposed to solve this inter-individual image fusion problem. The reader could refer to Ref. 74 for an overall survey on that subject. These methods are generally divided into two groups:

(1) "photometric" methods, that rely generally on the matching of voxels having the same luminance, and
(2) "geometric" methods that rely on the extraction and matching of appropriate landmarks.

Geometric methods (also called local methods) dramatically depend on the extraction of features. They are generally valid near these features only, but can be very precise around the features of interest, especially due to the tight control of the topology preservation through the matching stage of the corresponding features. The application of the computed transformation, away from these features, is performed by using interpolation functions, e.g. splines, radial basis functions, etc.

In contrast, photometric methods (also called global, intensity-based or iconic methods) use the entire image intensities information and automatically estimate transformations with high degrees of freedom in an automatic way, though using highly complex optimization schemes to solve the matching problem. Intensity-based methods are more dedicated to automatic procedure, require in principle less preprocessing (e.g. segmentation) and less *a priori* information derived from human expertise. This explains why photometric methods have become so popular. In addition, it has been proved very efficient in the particular context of rigid multimodal fusion.[126]

Another alternative to both methods is to use hybrid registration schemes in order to find a mid-way between the global and local approaches, taking advantage of the automaticity of the global one and the regional accuracy of the local or geometric one.

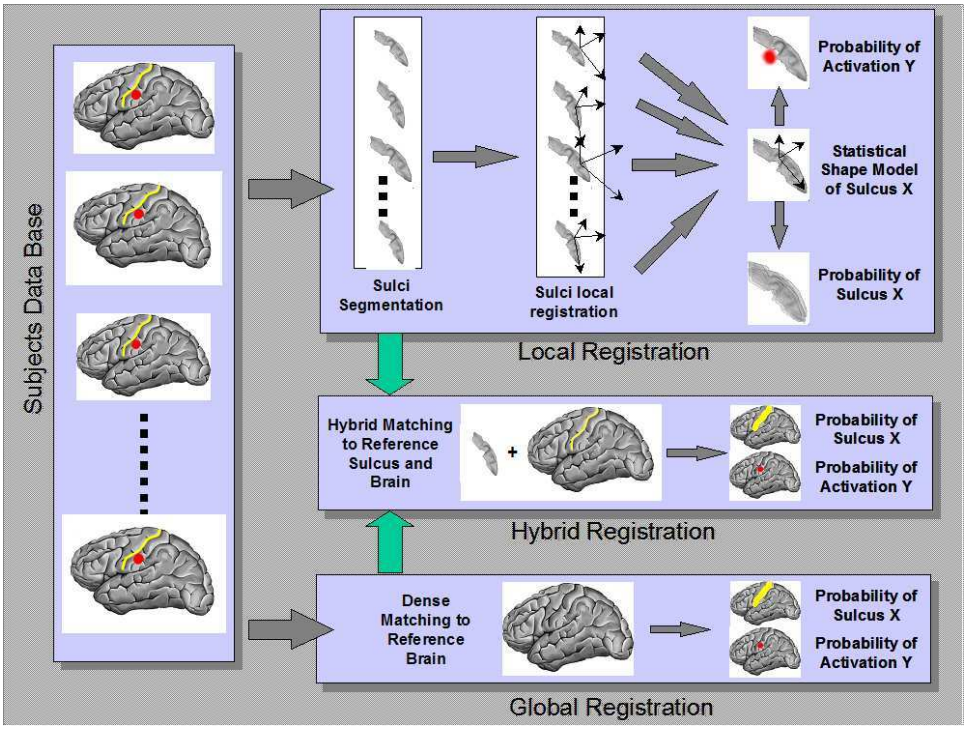


Fig. 1. Methodological framework for non-rigid registration in neuroimaging using local, global and hybrid registration methods.

Diagram in Fig. 1 summarizes this problem and illustrates how these three classes of methods can be compared based on a similar framework. The first phase consists in collecting anatomical and functional information. Then, the "Local Registration" consists in extracting structures of interest (sulci here) and to model them in 3D. These structures are then aligned on the basis of a local referential, i.e. a reference related to the studied sulci, to allow a statistical analysis of the variations in form and position of these structures. The statistical model is interpretable via its mean shape and its principal modes of deformation. Functional activations of the various subjects can then be combined in this mean statistical space by using an interpolation basis function in order to compute statistics of activation disparities. This methodology can then be compared with the "Global Deformable Registration" or the "Hybrid" frameworks.

Despite the large literature in this domain, very few works have been done so far, to propose an evaluation framework in order to compare objectively the performances of different deformable registration methods on the same database of subjects. Recently, an evaluation framework of photometric methods has been proposed,[60] based on global and local measures of the quality of the registration especially focusing on cortical regions where functional recordings generally take

place. Even though there are still limitations in such evaluation framework, conclusions of this study should entice investigators to study more carefully the real impact of the methods they use when performing inter-individual anatomical and functional brain mapping.

This chapter will describe and compare these three classes of methods. It is organized as follows: Section 2 presents a typical example of such global deformable registration method using image intensities.[61] As said before, in order to improve the performance of such global registration methods, a non-linear local registration framework (a geometric method) can be set up, capable to account for local deformations around local landmarks, e.g. cortical features, and capable to extend the estimated deformation field to nearby functional activation (in this case MEG dipoles, but could be fMRI or others).[32] This method is presented and discussed in Sec. 3. A third alternative in this framework is to use an hybrid approach able to locally control global deformation field (computed from photometric information) by sparse landmarks, i.e. geometric information. This method is presented and discussed in Sec. 4. At last, a comparative evalution of these registration frameworks is performed on anatomical and functional data in Sec. 5.

2. Global Registration Approaches

2.1. *Introduction*

In our point of view, global registration approaches have two main characteristics:

- They are dense approaches in the sense that a deformation vector is defined for each point of the discrete lattice.
- The estimation of the deformation field relies on a criterion or measure that is defined globally, that is to say identical for each point of the discrete lattice. However, the measure can express the locality (image derivatives or image moments for instance).

Global registration methods have been developed to handle complex deformations such as soft tissue deformation, evolution of lesions over time, matter appearance or dissipation and so on. More particularly, these methods have received particular attention for the development of electronic brain atlases. These atlases have emerged by overcoming some limitations of traditional paper-based atlases.[44,53,66,80,85,114] Non-rigid inter-subject registration methods have been developed in order to account for the inter-subject variability.[80]

This section will propose a rapid survey of global registration methods and will then describe the Romeo method, which is a robust elastic registration based on optical flow.

2.2. *A brief survey of global registration methods*

Registration is a very active field of research and numerous methods have been proposed. This section does not intend to propose an exhaustive list of methods

but to present generic and up-to-date methods. The interested reader will refer to Refs. 16, 51, 52, 71, 74, 79, 116, 120, 127 for a complete survey on this subject.

Methods can generally be classified according to the following criteria:

- Features that will be matched. This includes both the dimension of the data (classically from 2D to 4D) as well as the homologous structures that are chosen for matching.
- Transformation type. The transformation type can be rigid, affine, projective, etc. This item also includes the transformation domain: local or global. A transformation is called "global" when the modification of one parameter affects the entire image.
- The similarity measure. The similarity models the interaction between the data (features used for matching defined above) and the variables to be estimated (parameters of the transformation for instance).
- The regularization. The regularization can be implicit (regularized transformation model for instance) or explicit (first-order regularization for instance).
- The optimization method. Once the registration problem has been formalized, the optimization plays a crucial role in estimating the registration variables.

We have chosen here to present the following registration methods: methods that derive from the laws of continuum mechanics; methods that use cross-correlation; the demon's method; methods based on optical flow and finally methods that estimate jointly an intensity correction and a geometrical transformation.

2.2.1. *Models based on continuum mechanics*

Considering two MR images of two different subjects, the estimation of a "plausible" transformation must be sought. The notion of a "plausible" transformation in this context being particularly difficult to state, some authors have proposed to comply with the laws of continuum mechanics, either elastic (see Sec. (a)) or fluid (see Sec. (b)).

(a) Elastic models

Elastic models have been introduced by Broit[15] and extended by Bajcsy and Kovacic.[3,4] These models are nowadays used by various authors.[34,36,46,50,90,91,102,104,123] The estimated deformation field should basically obey the rule of Navier equation:

$$\mu\nabla^2 u + (\lambda + \mu)\nabla(\mathrm{div}(u)) + F = 0 \tag{1}$$

where u is the deformation field to estimate, λ and μ are the Lame coefficients and F is the sum of forces that are applied on the system. The problem is to specify the forces F that will lead to a correct registration. Bajcsy proposes to compute these forces so as to match the contours.[4] Davatzikos[36] and Peckar[90] do not compute any forces but segment the brain surface and the ventricles using two different methods. The matching of these surfaces provides boundary conditions

that make it possible to solve the problem. These two approaches are therefore sensitive to segmentation errors.

The use of elastic methods raises the following questions:

- What should be the values of Lame coefficients? The choice of these coefficients influences the deformation. Earliest work proposed that $\lambda = 0$ but it appears nowadays to be a limitation.
- This modeling cannot handle large deformations. As a matter of fact, the equation of Navier is only valid for small displacements. To solve this problem, two kinds of approaches can be used. A rigid registration can provide a good initialization, e.g. Bajcsy[4] uses principle inertia axes and Davatzikos[36] uses the stereotaxic space. Another way[90] is to solve the problem iteratively using a multiresolution approach.
- The topology of present structures will be preserved. This may be interesting in some applications but more questionable when matching brains of different subjects. Ono[89] has shown that cortical structures are not topologically equivalent among subjects indeed.

(b) Fluid models

Following the same inspiration as elastic models, Christensen and Miller[23] propose to compute a deformation that obeys the rule of fluid mechanics (equation of Navier-Stokes). The major difference with the elastic modeling is the fact that the fluid continuously "forgets" about its initial position. Large displacements and complex motions are therefore much easier to handle. The equation of Navier-Stokes can be written as:

$$\frac{\partial \vec{u}}{\partial t} - \nu \triangle \vec{u} + (\vec{u} \cdot \vec{\nabla})\vec{u} + \vec{\nabla} p = 0 \tag{2}$$

where ν is the fluid viscosity, $\vec{u}$ its speed and $\vec{p}$ its pressure. This equation is highly non-linear (cross-terms) and its resolution is complex, leading to large computation times. Christensen imposes the constraint that the Jacobian be positive,[23] leading to an homeomorphic transformation.

Christensen and Johnson[22] have extended the registration approach to introduce the reversibility constraint. Given two subjects A and B, the method jointly estimates transformation from A to B and from B to A. The inverse consistency error is zero when the forward and reverse transformations are inverses of one another. Furthermore, the transformations obey the rules of continuum mechanics and are parameterized by Fourier series.

Bro-Nielsen[13] has proposed an improvement to solve the following partial differential equation:

$$\mathcal{L}v = \mu \nabla v(x) + (\lambda + \mu)\mathrm{div}(v) = f(x, u(x)) \tag{3}$$

where u is the displacement and v the instantaneous speed. For a small time change, internal forces are constant and the equation is linear. While Christensen uses a

finite element scheme, Bro-Nielsen considers the impulse response associated with operator $\mathcal{L}$. The solution is then expressed as linear combinations of eigenvectors of operator $\mathcal{L}$. This decreases significantly the computation time.

Wang and Staib[123] have also proposed two methods that obey the rule of continuum mechanics. The methods respect the properties of elastic solids or viscous fluids. A statistical shape information (sparse set of forces) is mixed with a luminance information (dense set of forces within a Bayesian framework).

2.2.2. *Correlation*

Cross-correlation is a widespread similarity measure. It has been used by popular methods such as ANIMAL[29] and Gee *et al.*[50] ANIMAL uses a multiresolution strategy to estimate local linear transformations that maximizes cross-correlation. At a resolution level σ, the regularization is based on the statement that the norm of displacement vectors should not exceed σ. Collins *et al.*[28] has extended ANIMAL so that sulcal constraints can be taken into account in the registration process.

Gee, first interested in mechanical models,[50] adopted a statistical Bayesian framework.[49] Let us note I_R the reference volume, I_T the target volume, $z = \{I_R, I_T\}$ the data and u the deformation field. The problem is then to minimize the cost functional:

$$P(z|u) \propto \exp - \left\{ \int_{x \in \Omega_T} S(I_T(x), I_R(x + u(x)))dx \right\} \tag{4}$$

where S is the similarity measure that has been chosen to be cross-correlation. The regularization follows either a membrane model $P(u) \propto \lambda \int \left(u_x^2 + u_y^2\right)dx$ or a thin plate model $P(u) \propto \lambda \int \left(u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2\right)dx$. Gee also made it possible to incorporate landmark points in the registration process. If the transformation X matches p_i with p_i', the associated potential is: $P(Z = \left(p_i, p_i'\right)|\theta = X) \propto \exp - \frac{1}{2\sigma_i^2} \|X(p_i) - p_i'\|^2$. This probabilistic approach is useful to mix mechanical regularization, photometric similarity and landmark matching. It also make it possible to experiment and compare different kinds of regularization.[49]

Cachier *et al.*[18] have proposed the Pasha algorithm where the local correlation coefficient is used. This coefficient can be efficiently computed using convolutions with a Gaussian window function. The regularization is a mixture of competitive and incremental regularization using quadratic energies.

2.2.3. *Demons*

Thirion has proposed a method well known as the Demon's algorithm.[112] At each demon's location, forces are computed so as to repulse the model toward the data. The force depends on the polarity of the point (inside or outside the model), the image difference and gradients. For small displacements, it has been shown that the demon's method and optical flow are equivalent. The method is alternated: computation of forces and regularization of the deformation field by a Gaussian

smoothing. The choice of the smoothing parameter is therefore important. The Demon's algorithm has been successfully used by Dawant *et al.*[37]

Cachier and Pennec[93] have shown that the Demon's method can be viewed as a second-order gradient descent of the SSD (Sum of Squared Differences). This amounts to a minmax problem: maximization of similarity and regularization of solution.

2.2.4. *Displaced frame difference and optical flow*

The displaced frame difference (DFD) measures the difference between voxel intensities. It can be used either directly[1,86,121] or linearized (known as optical flow).[41,62,107] The DFD is known to be highly non-linear whereas optical flow is linear. However, optical flow is only valid for small displacements and can estimate motion only in the direction of image gradient (aperture problem). In both cases, this similarity will not be valid if luminance is not conserved (this may happen because of image acquisition, acquisition systems or parameters, MR inhomogeneities, etc).

Close to mechanical approaches, Song and Leahy[107] and Devlaminck[41] have proposed to estimate the optical flow with a mechanical regularization. More specifically, when images are density images, the luminance is directly related to a physical quantity and the mass conservation hypothesis may be introduced to constraint the estimation in a plausible way.[33,107]

In the field of cardiac imaging, Reissmann *et al.*[96] have proposed to use the neuractive pyramid to register images using the optical flow. The elastic grid that is the kernel of the deformation deforms so as to reject the discontinuities at boundaries of the grid. The minimization is therefore alternated between the deformation and the optimal shape of the grid.

The SPM spatial normalization approach[2] estimates warps by matching each skull-stripped image to the skull-stripped reference. Registration involves minimizing the mean squared difference between the images, which had been previously smoothed by convolving with an isotropic 8mm FWHM Gaussian kernel. The nonrigid deformation is modeled by a linear combination of low-frequency cosine transform basis functions.[2] Displacements in each direction are parameterized by 392 basis function coefficients, making a total of 1176 parameters in total. Regularization is obtained by minimizing the membrane energy of the warps.

Vemuri[121] also uses the optical flow but models the deformation as a combination of splines similarly to Ref. 108. Finally, Musse *et al.*[86] describe a hierarchical method to estimate the deformation using the SSD (Sum of Squared Differences) criterion. The solution is sought as a combination of spines functions that ensure the regularity of the solution.

2.2.5. *Joint estimation of intensity and geometric transformations*

Many artifacts can modify the luminance of an MR image. One of them is the inhomogeneity of the magnetic field for instance.[65] As a consequence, the hypothesis

of luminance conservation might not be valid anywhere. One solution consists in using robust estimators to get rid of inconsistent data. Another solution consists in jointly estimating an intensity correction and a spatial transformation.[45,48,54]

Gupta and Prince[54] propose an affine correction model for tagged MR: $f(\mathbf{r} + \mathbf{dr}, t + dt) = m(\mathbf{r}, \mathbf{dr}, t, dt)f(\mathbf{r}, t) + c(\mathbf{r}, \mathbf{dr}, t, dt)$. The optical flow equation then becomes:

$$f(\mathbf{r}, t) + \nabla f(\mathbf{r}, t) \cdot U(\mathbf{r}, t) - f(\mathbf{r}, t)\frac{\partial m(\mathbf{r}, t)}{\partial t} - \frac{\partial c(\mathbf{r}, t)}{\partial t} = 0\,. \tag{5}$$

The equation is solved in a variational framework using a first-order regularization.

Friston[48] and Feldmar[45] propose to embed the intensity correction and the spatial transformation in the same cost functional:

$$\mathcal{C}(\mathbf{f}, g) = \sum_{M_i \in i_1} (I_2(f(M_i)) - g(I_1(M_i), M_i))^2 \tag{6}$$

where f is the 3D transformation and g is the intensity correction. Feldmar generalizes this approach and considers 3D images as 4D surfaces. The criterion becomes:

$$\mathcal{C}(\mathbf{f}, g) = \sum_{(x_j, i_j)} d((\mathbf{f}(x_j), g(x_j, i_j)), CP_{4D}(\mathbf{f}(x_j), g(x_j, i_j)))^2 \tag{7}$$

where x_j is the point of intensity i_j and CP_{4D} is the function that renders the closest point. In this sense, this method is a generalization of the Iterative Closest Point (ICP) algorithm. Functions f and g can be modeled according to the application. For instance, for an intra-subject monomodal registration, f is rigid and g is the identity. For inter-subject registration, f can be a combination of radial basis functions and f should correct acquisition artifacts.

2.3. *Romeo: Robust multigrid elastic registration based on optical flow*

2.3.1. *Introduction*

We consider the registration problem as a motion estimation problem, which has been studied by different authors.[5–9,25,62,67,88,103] Our 3D method performs a non-linear multi-modality registration of MRI acquisition of different subjects. The similarity measure incorporates robust estimators whose utility is twofold: on the one hand we want to limit the influence of the acquisition noise, on the other hand we want to cope with possible modifications of structures' topology.[58]

Many tasks in computer vision may be expressed as the minimization of a cost function. The optimization is often difficult to achieve, because the cost function is non-convex and because the optimization involves a very large number of variables. Therefore efficient iterative multigrid (or multilevel) approaches have been developed[55,81] and applied in motion estimation[43] and in early vision.[111]

To take into account large deformations, we use a multiresolution optimization scheme. Besides, at each resolution level, we use a multigrid minimization to

accelerate the algorithm and improve the quality of the estimation. Within this hierarchical approach, we designed an adaptive partition of the volume to refine the estimation on the regions of interest and avoid useless efforts elsewhere. An anatomical segmentation of the cortex is introduced and used in two ways: at each resolution level, we initialize the partition as an octree subdivision based on the segmentation, and the segmentation mask is used in the subdivision criterion which controls the refinement of the estimation.

2.3.2. *General formulation*

The optical flow hypothesis, or brightness constancy constraint, introduced by Horn and Schunck,[62] assumes that the luminance of a physical point does not vary much between the two volumes to register. It amounts to zeroing the so-called DFD (Displaced Frame Difference):

$$f(s + \boldsymbol{w}_s, t_1) - f(s, t_2) = 0 \tag{8}$$

where s is a voxel of the volume, t_1 and t_2 are the indexes of the volumes (temporal indexes for a dynamic acquisition, indexes in a database for multi-subject registration), f is the luminance function and $\boldsymbol{w}$ the expected 3D displacement field. The DFD may not be valid everywhere, because of noise and intensity inhomogeneities of MR acquisition. The robustness of the registration process with respect to acquisition artifacts will be discussed later on, in Secs. 2.3.4 and 2.3.6.

Generally, a linear expansion of this equation is preferred: $\boldsymbol{\nabla} f(s,t) \cdot \boldsymbol{w}_s + f_t(s,t) = 0$: where $\boldsymbol{\nabla} f(s,t)$ stands for the spatial gradient of luminance and $f_t(s,t)$ is the voxelwise difference between the two volumes. The resulting set of undetermined equations has to be complemented with some prior on the deformation field. Using an energy-based framework (which can be viewed either from the Bayesian point of view, or from the one of the regularization theory), the registration problem may be formulated as the minimization of the following cost function:

$$U(\boldsymbol{w}; f) = \sum_{s \in S} [\boldsymbol{\nabla} f(s,t) \cdot \boldsymbol{w}_s + f_t(s,t)]^2 + \alpha \sum_{\langle s,r \rangle \in \mathcal{C}} \|\boldsymbol{w}_s - \boldsymbol{w}_r\|^2 \tag{9}$$

where S is the voxel lattice, $\mathcal{C}$ is the set of neighboring pairs with respect to a given neighborhood system $\mathcal{V}$ on S ($\langle s,r \rangle \in \mathcal{C} \Leftrightarrow s \in \mathcal{V}(r)$), and α controls the balance between the two energy terms. The first term captures the brightness constancy constraint, thus modeling the interaction between the field (unknown variables) and the data (given variables), whereas the second term captures a simple smoothness prior. The weaknesses of this formulation are known:

(a) Due to the linearization, the optical flow constraint (OFC) is not valid in case of large displacements.
(b) The OFC might not be valid in all the regions of the volume, because of the acquisition noise, intensity non-uniformity in MRI data, and occlusions.

(c) The "real" field is not globally smooth and it probably contains discontinuities that might not be preserved because of the quadratic smoothing.

To cope with the (b) and (c) limitations, we replace the quadratic cost by robust functions. To face the problem (a), we use a multi-resolution plan and a multi-grid strategy to improve the minimization at each resolution level.

We have here introduced a simple regularization term that makes almost no assumption on the estimated deformation field. One could imagine choosing different regularizations for the different brain tissues, but that involves specific assumptions on the "real" deformation that we do not address in this chapter. However, the introduction of a robust estimator on the regularization term enables to take into account possible discontinuities on the border of structures having different physical properties.

2.3.3. *Rigid registration step*

Given two images with potentially large displacement, it seems first reasonable to estimate a rigid transformation. This step is performed by estimating a rigid transformation that maximizes mutual information.[27,122] Given two images A and B, considered as discrete random variables, let us note $p_A(a)$ and $p_B(b)$ their respective marginal probability distribution, and $p_{A,B}(a,b)$ the joint distribution. Mutual information $I(A,B)$ is then defined as:[27,122]

$$I(A,B) = \sum_{a,b} p_{A,B}(a,b) \log_2 \frac{p_{A,B}(a,b)}{p_A(a) p_B(b)} = H(A) + H(B) - H(A,B) \qquad (10)$$

with

$$H(A) = -\sum_a p_A(a) \log_2(p_A(a))$$

and

$$H(A,B) = -\sum_{a,b} p_{A,B}(a,b) \log_2(p_{A,B}(a,b)) \,. \qquad (11)$$

In some particular cases, such as brain images for instance, it is possible to define a reference coordinate system which takes into account some information about the scene (such as resolution of pixels/voxels, orientation of axes, etc.). In such cases, the two volumes to be registered are mapped in this reference coordinate system and the rigid transformation is expressed in this coordinate system. If this *a priori* information is not available, the rigid transformation is estimated in the coordinate system attached to the data.

The registration is performed through a multiresolution optimization scheme (construction of a pyramid of volumes by successive isotropic Gaussian filtering and subsampling in each direction).[56,94] At each resolution level, the similarity $I(A,T(B))$ is maximized with respect to the parameters of the transformation

using a Powell's algorithm.[95] We calculate the joint histogram on the overlapping part of A with $T(B)$ by partial volume interpolation, the latter being known to provide a smoother cost function.

2.3.4. *Robust estimators*

Cost function (9) does not make any difference between relevant data and inconsistent data, nor between neighboring pairs where the field is smooth and neighboring pairs where the field is discontinuous. Therefore, we introduce robust functions[63] and more precisely two robust M-estimators,[10] the first one on the data term and the second one on the regularization term. We do not describe in details the properties of robust M-estimators, referring the reader to Refs. 10 and 83 for further explanations. The cost function (9) can then be modified as:

$$U(\boldsymbol{w}; f) = \sum_{s \in S} \rho_1(\boldsymbol{\nabla} f(s,t) \cdot \boldsymbol{w}_s + f_t(s,t)) + \alpha \sum_{\langle s,r \rangle \in \mathcal{C}} \rho_2(\|\boldsymbol{w}_s - \boldsymbol{w}_r\|)\,. \tag{12}$$

According to some properties of robust M-estimators,[10,21] it can be shown that the minimization of U (see Eq. (9)) is equivalent to the minimization of an augmented function, noted $\overset{\star}{U}$:

$$\begin{aligned} \overset{\star}{U}(\boldsymbol{w}, \delta, \beta; f) = & \sum_{s \in S} \delta_s (\boldsymbol{\nabla} f(s,t) \cdot \boldsymbol{w}_s + f_t(s,t))^2 + \psi_1(\delta_s) \\ & + \alpha \sum_{\langle s,r \rangle \in \mathcal{C}} \beta_{sr} \|\boldsymbol{w}_s - \boldsymbol{w}_r\|^2 + \psi_2(\beta_{sr}) \end{aligned} \tag{13}$$

where δ_s and β_{sr} are auxiliary variables (acting as "weights") to be estimated. This cost function has the advantage to be quadratic with respect to $\boldsymbol{w}$. It also shows clearly that, when a discontinuity gets larger, the contribution of the pair of neighbors is limited by the reduction of the associated weight β_{sr}. The minimizers of $\overset{\star}{U}$ with respect to the auxiliary variables are obtained in closed form.[10,21] The overall minimization of such function consists in an alternated weights computation and quadratic minimizations with respect to $\boldsymbol{w}$.

2.3.5. *Multiresolution incremental computation of the optical flow*

In cases of large displacements, we use a classic incremental multiresolution procedure.[8,43] (see Fig. 2). We construct a pyramid of volumes $\{f^k\}$ with successive Gaussian smoothing and subsampling in each direction.[17] For each direction $i = \{x, y, z\}, d_i$ is the spatial resolution of a voxel (the spatial resolution of MR acquisition is around 1mm, depending on the system). We perform a Gaussian filtering using the recursive implementation proposed in Ref. 39 with a standard deviation of $2d_i$ in direction i, in order to satisfy Nyquist's criterion. This implementation allows to perform infinite impulse response filtering at a constant computation cost.

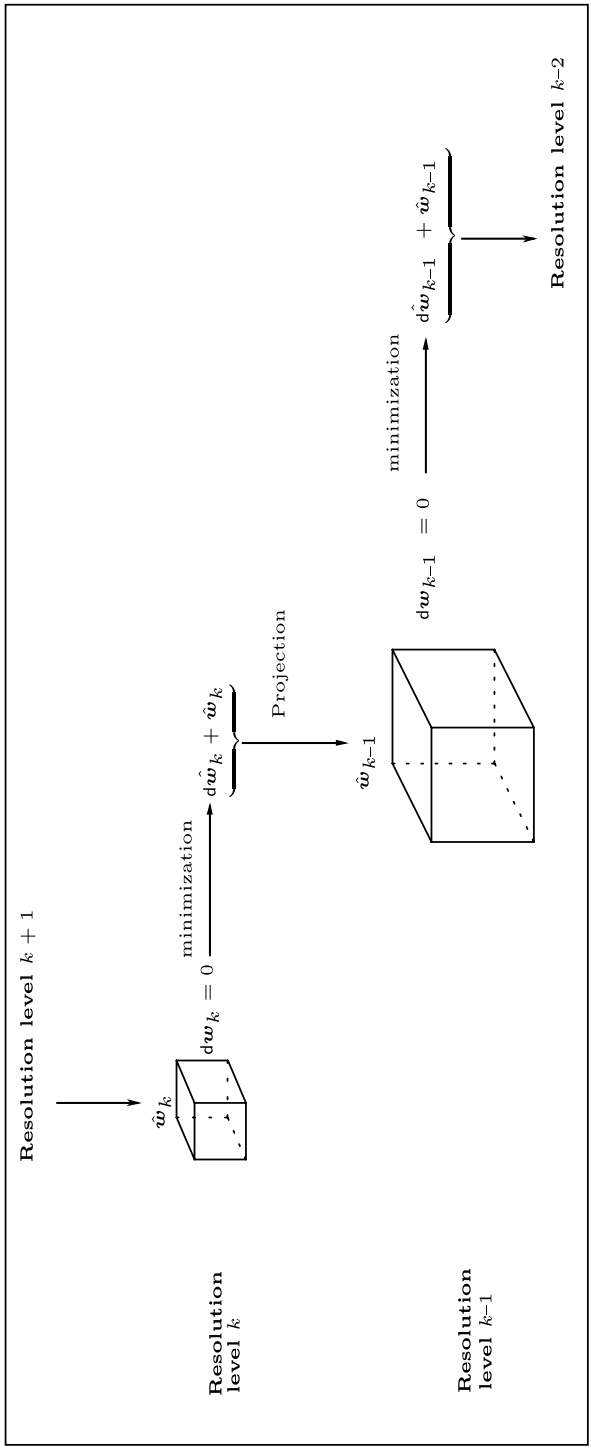


Fig. 2. Incremental estimation of the optical flow.

At the coarsest level, displacements are reduced, and cost function (13) can be used because the linearization hypothesis becomes valid. For the next resolution levels, only an increment $\mathsf{d}\boldsymbol{w}^k$ is estimated to refine the estimate $\hat{\boldsymbol{w}}^k$ obtained from the previous level. We perform the registration from resolution k_c until resolution k_f (in general $k_f = 0$). This is done using cost function (12) but with $\boldsymbol{\nabla}\tilde{f}^k(s,t) \triangleq \boldsymbol{\nabla} f^k(s + \hat{\boldsymbol{w}}_s^k, t_2)$ and $\tilde{f}_t^k(s,t) \triangleq f^k(s + \hat{\boldsymbol{w}}_s^k, t_2) - f^k(s,t_1)$ instead of $\boldsymbol{\nabla} f^k(s,t)$ and $f_t^k(s,t)$.

To compute the spatial and temporal gradients, we construct the warped volume $f^k(s + \hat{\boldsymbol{w}}_s^k, t_2)$ from volume $f^k(s,t_2)$ and the deformation field $\hat{\boldsymbol{w}}_s^k$, using trilinear interpolation. The spatial gradient is hence calculated using the recursive implementation of the derivatives of the Gaussian.[39] At each voxel, we calculate the difference between the source volume and the reconstructed volume, and the result is filtered with a Gaussian to construct the temporal gradient. As previously, these quantities come from the linearization of the constancy assumption expressed for the whole displacement $\hat{\boldsymbol{w}}_s^k + \mathsf{d}\boldsymbol{w}_s^k$. The regularization term becomes $\sum_{\langle s,r\rangle\in\mathcal{C}} \rho_2\big(\|\hat{\boldsymbol{w}}_s^k + \mathsf{d}\boldsymbol{w}_s^k - \hat{\boldsymbol{w}}_r^k - \mathsf{d}\boldsymbol{w}_r^k\|\big)$.

2.3.6. *Multigrid minimization scheme*

(a) Motivations

The direct minimization of Eq. (13) is intractable. Some iterative procedure has to be designed. Unfortunately, the propagation of information through local interaction is often very slow, leading to an extremely time-consuming algorithm. To overcome this difficulty (which is classical in computer vision when minimizing a cost function involving a large number of variables), multigrid approaches have been designed and used in the field of computer vision.[43,83,111] Multigrid minimization consists in performing the estimation through a set of nested subspaces. As the algorithm goes further, the dimension of these subspaces increases, thus leading to a more accurate estimation. In practice, the multigrid minimization usually consists in choosing a set of basis functions and estimating the projection of the "real" solution on the space spanned by these basis functions.

(b) Description

At each level of resolution, we use a multigrid minimization (see Fig. 3) based on successive partitions of the initial volume.[83] At each resolution level k, and at each grid level ℓ, corresponding to a partition of cubes, we estimate an incremental deformation field $\mathsf{d}\boldsymbol{w}^{k,\ell}$ that refines the estimate $\hat{\boldsymbol{w}}^k$, obtained from the previous resolution levels. This minimization strategy, where the starting point is provided by the previous result — which we hope to be a rough estimate of the desired solution — improves the quality and the convergence rate as compared to standard iterative solvers (such as Gauss-Seidel).

At grid level ℓ, $\Xi_\ell = \{\Xi_n, n = 1 \cdots N_\ell\}$ is the partition of the volume B into N_ℓ cubes Ξ_n. At each grid level ℓ corresponds a deformation increment $\mathsf{T}_{k,\ell}$ that is defined as follows: A 12-dimensional parametric increment deformation field is

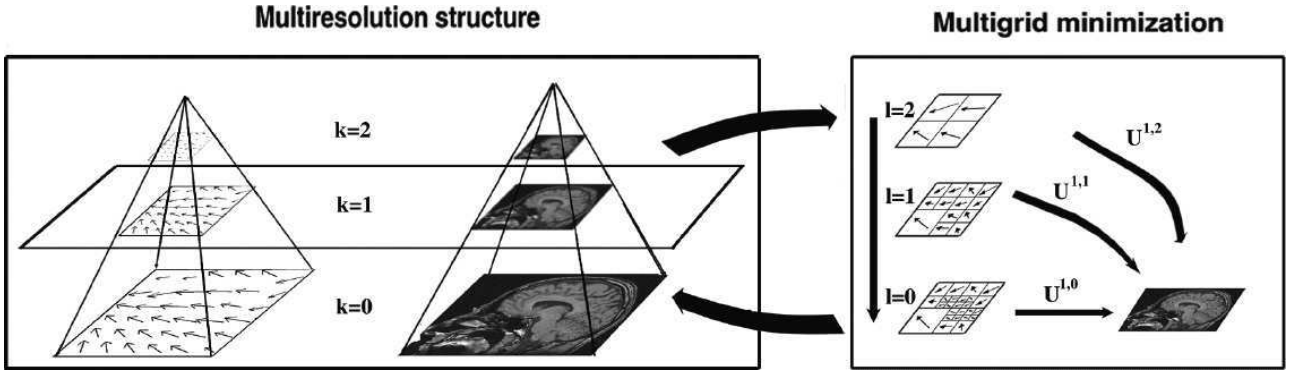


Fig. 3. Example of multiresolution/multigrid minimization. For each resolution level (on the left), a multigrid strategy (on the right) is performed. For legibility reasons, the figure is a 2D illustration of a 3D algorithm with volumetric data.

estimated on each cube Ξ_n, hence the total increment deformation field $\mathtt{d}\boldsymbol{w}^{k,\ell}$ is piecewise affine. At the beginning of each grid level, we construct a reconstructed volume with the target volume $f^k(s, t_2)$ and the field estimated previously (see Sec. 2.3.5). We compute the spatial and temporal gradients at the beginning of each grid level and the increment deformation field $\mathtt{d}\boldsymbol{w}^{k,\ell}$ is initialized to zero. The final deformation field is hence the sum of all the increments estimated at each grid level, thus expressing the hierarchical decomposition of the field.

Contrary to block-matching algorithms, we model the interaction between the cubes (see Sec. 2.3.7) of the partition, so that there is no "block-effects" in the estimation. At each resolution level k, we perform the registration from grid level ℓ_c until grid level ℓ_f. Depending on the application, it may be useless to compute the estimation until the finest grid level, i.e. $\ell_f = 0$.

(c) Adaptative partition

To initialize the partition at the coarsest grid level ℓ_c, we consider a segmentation of the brain obtained by morphological operators. After a threshold and an erosion of the initial volume, a region growing process is performed from a starting point that is manually chosen. A dilatation operation allows us to end up with a binary segmentation. At grid level ℓ_c, the partition is initialized by a single cube of the volume size. We iteratively divide each cube as long as it intersects the segmentation mask and as long as its size is superior to $2^{3\ell_c}$. We finally get an octree partition which is anatomically relevant.

When we change from grid level, each cube is adaptively divided. The subdivision criterion depends first on the segmentation mask (we want a maximum precision on the cortex), but it also depends on the local distribution of the variables δ_s (see Eq. (13)). More precisely, a cube is divided if it intersects the segmentation mask or if the mean of δ_s on the cube is less than a given threshold. As a matter of fact, δ_s indicates the adequation between the data and the estimated deformation field at voxel s. Therefore, this criterion mixes an indicator of the confidence about the estimation with a relevant anatomical information.

2.3.7. *Parametric model*

We now introduce the deformation model that is used. We chose to consider an affine 12-parameter model on each cube of the partition. That kind of model is quite usual in the field of computer vision but rarely used in medical imaging. If a cube contains less than 12 voxels, we only estimate a rigid 6-parameter model, and for cubes that contain less than 6 voxels, we estimate a translational displacement field. As we have an adaptive partition, all the cubes of a given grid level might not have the same size. Therefore we may have different parametric models, adapted to the partition.

At a given resolution level k and grid level ℓ, $\Xi_{k,\ell} = \{\Xi_n, n = 1 \cdots N_{k,\ell}\}$ is the partition of the volume into $N_{k,\ell}$ cubes Ξ_n. On each cube Ξ_n, we estimate an affine

displacement defined by the parametric vector $\Theta_n^{k,\ell}$: $\forall s = (x, y, z) \in \Xi_n, \mathsf{d}\boldsymbol{w}_s = P_s\Theta_n^{k,\ell}$, with

$$P_s = \begin{pmatrix} 1 & x & y & z & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x & y & z & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & x & y & z \end{pmatrix}. \tag{14}$$

A neighborhood system $V^{k,\ell}$ on the partition $\Xi_{k,\ell}$ derives naturally from $\mathcal{V}$: $\forall n,\ m \in \{1 \cdots N_{k,\ell}\}$, $m \in V^{k,\ell}(n) \Leftrightarrow \exists s \in \Xi_n, \exists r \in \Xi_m \backslash r \in \mathcal{V}(s)$. $\mathcal{C}$ being the set of neighboring pairs on S^k, we must now distinguish between two types of such pairs: the pairs inside one cube and the pairs between two cubes:

$$\begin{aligned} &\forall n \in \{1 \cdots N_{k,\ell}\}, \quad \langle s, r\rangle \in \mathcal{C}_n^\ell \Leftrightarrow s \in \Xi_n\,,\ \ r \in \Xi_n \ \text{ and } \ r \in \mathcal{V}(s)\,. \\ &\forall n \in \{1 \cdots N_{k,\ell}\}, \quad \forall m \in V^\ell(n)\,, \\ &\langle s, r\rangle \in \mathcal{C}_{nm}^\ell \Leftrightarrow m \in V^l(n)\,,\ \ s \in \Xi_n\,,\ r \in \Xi_m \ \text{ and } \ r \in \mathcal{V}(s)\,. \end{aligned}$$

For the sake of concision, we will now drop the resolution index k. With these notations, the cost function (13) becomes:

$$\begin{aligned} &\overset{\star}{U}{}^{\ell}(\Theta^\ell, \delta^\ell, \beta^\ell; \boldsymbol{w}, f^\ell) \\ &\quad = \sum_{n=1}^{N_\ell} \sum_{s \in \Xi_n} \delta_s^\ell \big[\boldsymbol{\nabla}\tilde{f}_s^T P_s \Theta_n^\ell + \tilde{f}_t(s,t)\big]^2 + \psi_1(\delta_s^\ell) \\ &\qquad + \alpha \sum_{n=1}^{N_\ell} \left[\sum_{m \in V^\ell(n)} \sum_{\langle s,r\rangle \in \mathcal{C}_{nm}^\ell} \beta_{sr}^\ell \|(\boldsymbol{w}_s + P_s\Theta_n^\ell) - (\boldsymbol{w}_r + P_r\Theta_m^\ell)\|^2 + \psi_2(\beta_{sr}^\ell) \right] \\ &\qquad + \alpha \sum_{n=1}^{N_\ell} \left[\sum_{\langle s,r\rangle \in \mathcal{C}_n^\ell} \beta_{sr}^\ell \|(\boldsymbol{w}_s + P_s\Theta_n^\ell) - (\boldsymbol{w}_r + P_r\Theta_n^\ell)\|^2 + \psi_2(\beta_{sr}^\ell) \right]. \end{aligned} \tag{15}$$

Considering the auxiliary variables of the robust estimators as fixed, one can easily differentiate the cost function (15) with respect to any Θ_n^ℓ and get a linear system to be solved. We use a Gauss-Seidel method to solve it for its implementation simplicity. However, any iterative solver could be used (solvers such as conjugate gradient with an adapted preconditioning would be for example more efficient). In turn, when the deformation field is "frozen", the weights are obtained in a closed form.[10,21] The minimization may therefore be naturally handled as an alternated minimization (estimation of Θ_n^ℓ and computation of the auxiliary variables). Contrary to other methods (minmax problem like the demons' algorithm for instance), that kind of minimization strategy is guaranteed to converge,[21,38,84] i.e. to converge toward a local minimum from any initialization.

Moreover, the multigrid minimization makes the method invariant to intensity inhomogeneities that are piecewise constant. As a matter of fact, if the intensity inhomogeneity is constant on a cube, the restriction of the DFD on that cube is modified by adding a constant. As a consequence, minimizing the cost function (15) gives the same estimate, whenever the cost at the optimum is zero or a constant.

3. Local Registration Approaches

3.1. *Introduction*

In this section, we describe the local registration approach illustrated in Fig. 1. It is designed to register 3D anatomical and functional data, i.e. data from various subjects acquired via various modalities, like magnetic resonance imaging (MRI) for anatomical data, magneto-encephalography (MEG) or functional magnetic resonance imaging (fMRI) for functional data, etc. While anatomical imaging visualizes brain morphology, tissues, ..., functional imaging enables the localization of cerebral activity. They thus provide additional information which can be exploited to give a functional cartography of the human brain, or to construct anatomical and functional atlases,[44,80,116] such atlases being particularly expected to account for the inter-individual variability. Their constitution requires to collect a database from which anatomical information and functional information will be extracted, then to design registration techniques between distinct subjects in order to compare them. As introduced in Sec. 1, global registration methods (see Sec. 2) and local registration approaches are fundamentally different and lead to probalistic atlases defined in different spaces. In contrast to classic global methods, local or geometric approaches explicitly exploit anatomical information by matching particular anatomical structures and are expected to be precise in the vicinity of these structures.

Our local approach is qualified "local" since it considers a particular anatomical structure, e.g. cortical sulci, and it uses a local coordinate frame, intrinsic to the considered structure. It handles the inter-subject variability via a statistical shape model of these anatomical structures of interest. Shape models not only represent the shape of an object but also the way it can vary.[31,82] Relying on a statistical approach, they achieve model and data adequation through a learning stage, and capture the variability within a class of objects. Although shape models are generally used for segmentation purposes, we have investigated their use in the context of anatomo-functional normalization. We present a generic inter-individual functional mapping scheme to register multi-subjects functional data with respect to anatomical constraints, i.e. cortical sulci, and to express them in a coordinate system linked to the sulci anatomical model. We assume that a part of functional inter-subject variability is encoded in anatomical variability and that sulci are relevant landmarks for functional analysis.[77,99,124]

Some previous studies argue in favor of an anatomo-functional correlation. Brodmann's cytoarchitectonic cortical cartography[14] shows a correspondence between some cytoarchitectonic areas and some functional areas. For example, visual areas of occipital cortex correspond to Brodmann's areas 17, 18 and 19; somatosensory areas correspond to areas 1 and 3 while motor areas correspond to areas 4 and 6. This map also shows that some limits of cytoarchitectonic areas correspond to a sulcus, suggesting that some sulci could be anatomical limits for an *a priori* localization of some functional areas. Thus, the central sulcus separates the somatosensory

cortex from the motor cortex. If this link is obvious for some primary areas, it becomes questionable when considering associative areas. For some primary areas, a functional cartography could have been established in both animals and humans. A study by Welker and Seidenstein on somatosensory areas of racoons[125] demonstrates a correspondence between various cutaneous territories and the sulco-gyral anatomy. A sulcus, if it exists, systematically separates the cortical representation of two tactiles receptors, i.e. the cortical representation of a specific receptor cannot be interrupted by a sulcus. Conversely, two cortical representations of distinct receptors are not systematically separated by a sulcus. This experiment suggests that the presence of a sulcus in one subject implies a local functional cartography valid for all the subjects, included those who do not have this sulcus. In humans, the cartography of motor and somatosensory areas by Penfield and Rasmussen[92] shows a topological correspondence between bodily parts and their cortical representation: two corporal neighboring areas project onto two cortical neighboring areas. This property is known as "somatotopy". Such properties also hold for the visual cortex and are called "properties of retinotopy". A part of the retina is represented in an ordered way on the cortical surface. The auditive cortex benefits from similar properties which are then called "tonotopy".

Accordingly, under the hypothesis of an anatomo-functional correlation, it appears sensible to consider cortical sulci as relevant anatomical landmarks for functional analysis. The proposed registration method is fully based on anatomical features and it is applied only on functional data located in a spatial neighborhood of these anatomical features. The registration process jointly uses the anatomical shape model and an interpolation scheme based on thin-plate splines.[11,12]

Section 3.2 describes the construction of the statistical model of cortical sulci, performed by learning from a set of shapes. In Sec. 3.3, we introduce the interpolation scheme and the way it can be combined with the statistical shape model for the local and non-linear registration of functional activations.

3.2. *Statistical shape model for cortical sulci*

Cortical sulci are anatomical structures whose shape may vary widely from individual to individual. We propose to model these structures by learning the variability inherent to a training set. The resulting model is related to what is commonly called a Point Distribution Model (PDM).[31] Our methodology proceeds in four steps: extraction of the features of interest, definition of an appropriate data representation, effective construction of the training set — which implies to establish correspondences between the learning shapes — and application of a statistical analysis. Here we have chosen a parametric data representation and designed a matching scheme based on the definition of a local coordinate system, from which we derive correspondences. We subsequently apply a Principal Component Analysis (PCA) to achieve a compact and meaningful representation of cortical sulci shapes and of their variations.

3.2.1. *Extraction of cortical sulci*

(a) Brief survey

Approaches for cortical sulci segmentation differ by the techniques used, but especially by the way authors consider a sulcus. For instance, a sulcus can be assimilated to its external trace, to a sulcal area or sulcal basin, or to a surface. Note that prior to sulci extraction usually takes place brain segmentation. A survey about such brain segmentation methods can be found in Ref. 26.

Caunce *et al.*[20] consider a sulcus as its external trace. From a brain segmentation, a set of points is automatically located on the sulcal fissures of the cortical envelop by mathematical morphology techniques. This representation being established for a set of sulci over a subject database, a Point Distribution Model (PDM)[31] is built. A point-to-point correspondence is performed by an Iterative Closest Point (ICP) algorithm which integrates local geometric constraints to improve its performance. New instances are then segmented via an Active Shape Models (ASM)[31] procedure. Tao *et al.*[110] also use an unidimensional sulci representation but prefer a spherical cortex representation. The curves standing for the sulci are manually extracted and projected on the unit sphere thus leading to a parametric representation of sulci. A statistical model dedicated to the segmentation of new sulci is then built on this unit sphere. Such representations appear somehow reductive to study the cortical inter-individual variability. They do not account for the buried cortical part which represents at least two thirds of the cortex. Royackkers *et al.*[101] also represent sulci as curves following their external traces but add a depth information in each point of the curve. Moreover, additional characteristics, such as position, continuity, length, orientation, describe the sulcus. However, a more complete representation of the cortical buried part can be obtained.

Rettmann *et al.*[98] propose a cortical representation into sulcal and gyral regions. Sulcal areas describe a "U" or a "V" corresponding to the inner walls of the cortical folds. The segmentation is performed by a watershed algorithm relying on a geodesic distance. The problem of sursegmentation, i.e. one region is segmented into several distincts areas, implied by this kind of algorithms is solved by a fusion step of the obtained regions. The identification of the segmented sulci is manually done at the end of this segmentation process. This approach is related to Refs. 72 and 73, where a structural description of the cortical topography is sought. This work assumes that the inter-individual sulcal variability decreases with depth. The sulci are represented by a set of sub-structures named "sulcal basins" which are volumetric regions defined as concavities in the white matter surface. Their extraction is performed by a region growing-like algorithm. According to the authors, removing grey matter accentuates the sulci in depth and leads to a better sulci localization.

An alternative to model the deep part of the cortex is the modeling of sulci by their median surface. This approach additionally leads to an easier visualization in three dimensions. Vaillant *et al.*,[117] seek a parametric representation of this surface. An active model is initialized on the external trace of the sulcus and evolves towards

its fundus. The external traces are determined by the minimum of principal curvature of the external brain surface parameterized beforehand.[35] The model evolves according to two forces which drive it toward the fundus while constraining it onto the median surface. Another median surface extraction method is proposed in Refs. 40 and 97. It is based on curvature and crest lines notions and computes these characteristics on a continuous approximation of the original volume; it thus avoids the prior brain segmentation step. Aiming at a fine structural description of the cortical network, Mangin *et al.*[76] propose a segmentation of sulci into topological elementary surfaces based on the classification proposed by Malandain.[75] These surfaces are then sur-cut at level of possible connexions of sulcal roots. This segmentation is integrated to a complex automatic labeling system based on the joint use of neuronal networks for learning and Markov fields for recognition.[78,99]

The method we use to extract cortical sulci[69] takes into account the deep cortical topography and provides a parametric representation of sulci, as "ribbons".

(b) The "active ribbon" method

Sulci are extracted from MRI volumes by a method now known as the "active ribbon" method,[69] which leads to a parametric representation of the median surface of the sulcus. This latter is considered as the surface from the fundus of the sulcus towards the outside of the brain and pseudo-normal to the external envelop of the brain. Preprocessing steps aim first at defining a region of interest (ROI) consisting of the union of gyri and sulci. Roughly, this ROI is obtained by first, segmenting the brain and second, by classifying the tissues into three classes: white matter, grey matter and cerebro-spinal fluid.[68] On the basis of this partition, a differential geometry operator, the L_{vv} operator,[47] is used to distinguish between sulci and gyri. This is a curvature extractor well adapted to a surface such as the cortex, since positive curvature characterizes the top of a gyrus, whereas negative curvature characterizes the bottom of a sulcus. After skeletonization, this sulci/gyri partition results in the superficial topography of the cortex, i.e. in the external traces of the sulci. These external traces then serve to initialize an active model extracting the surface of interest. The point is that the dimension of the active model changes from 1D (active curve) to 2D (active surface). In fact, a set of potentials are defined such that the initial curve, located at the surface of the brain, evolves toward the bottom of the sulcus. The set of successive positions of this 1D curve initializes an active surface which, once optimized, describes the median surface of the sulcus. The choice of parameterization has been fixed on a cubic B-spline surface, which is well adapted to free form object modeling.

The identification of cortical sulci is performed manually: after the skeletonization step, external traces are labeled by an expert. Note that Le Goualher *et al.* propose an automatic atlas-based labeling procedure in Ref. 70.

3.2.2. *Representation of cortical sulci*

We thus obtain a parametric representation of the shapes of interest through B-spline modeling. The spline, parameterized by its curvilinear abscissae u and

v, is described by nbp sample points and $nbc = nbc\_u * nbc\_v$ control points where nbc_u (resp. nbc_v) is the number of control points in the direction associated with parameter u (resp. v). Parametric direction u represents the length of the sulcus and direction v its depth. The number of sample points, nbp, is obtained by regularly sub-dividing the curve initialized on the external trace of the sulcus into 1 mm intervals and remains constant all along the segmentation process. The number of control points, nbc, is determined according to the finesse and smoothing degree desired for the representation. For a fine representation, the ratio nbc/nbp, called smoothing factor, will be equal or close to 1. The smaller this ratio, the smoother the surface. In practice, the smoothing factor has been chosen to $nbc/nbp = 1/24$, which is a trade-off between surface smoothing and segmentation quality.

A sulcus can be represented by the sample points of the spline modeling its median surface or by its control points. Having one of this representation, the other one can be easily obtained:

$$\mathbf{c} = B\mathbf{p} \tag{16}$$

where B is the spline matrix, $\mathbf{c}$ contains the sample points coordinates and $\mathbf{p}$ contains the control points coordinates. Giving nbc control points and knowing the spline matrix completely defines the sulcal surface. The main advantage to use control points is their ability to represent the surface in a more compact way than sample points while guaranteeing an equivalent representation.

3.2.3. *Building the training set*

The construction of the training population $\mathcal{P}$ consists in defining a random observation vector $\mathbf{x} \in \mathbb{R}^p$ whose realizations are the instances of $\mathcal{P}$, i.e a set of comparable parameters that describes the instances of $\mathcal{P}$. To do so, we design a matching scheme which leads to a point-to-point correspondence. First, we rigidly align all the shapes in a common coordinate system. We propose a local approach based on the definition of an intrinsic coordinate system to the shape "sulcus", on which the alignment is performed by a local rigid registration. Using a local scope leads to an independence towards global structures such as a brain reference for instance. Then we resample all the aligned shapes so that they have the same number of points, and eventually we establish point-to-point correspondences.

(a) Intrinsic coordinate system and local rigid registration

Initially, each sulcus is expressed in the image coordinate system, which differs from one subject to another. The principle is to express each sulcus in its own coordinate system, built so that it is common to all sulci. We call it a "local coordinate system" or "intrinsic coordinate system". Its construction is based on the inertia axis of the sulcal surface defined by the nbp sample points.

Let us consider a sulcus S. Let $\mathcal{R}(O, \mathbf{u}, \mathbf{v}, \mathbf{w})$ be the image coordinate system in which it is initially expressed and $\mathcal{R}_s(O_s, \mathbf{u_s}, \mathbf{v_s}, \mathbf{w_s})$ its local coordinate system. Axes $\mathbf{u_s}$, $\mathbf{v_s}$ and $\mathbf{w_s}$ are defined as the axes of inertia of the sulcal surface, and

are determined so that $\mathbf{u_s}$ follows the length of the sulcus, $\mathbf{v_s}$ its depth and $\mathbf{w_s}$ is normal to the "sulcal plane". This discrimination between the three axes is first carried out by considering that $\mathbf{u_s}$ (resp. $\mathbf{v_s}$) is the "most collinear" inertia axis with the *nbc_u* (resp. *nbc_v*) pseudo-parallel directions, each of them being defined by the two extremities of a sulcus' line in direction u (resp. v). The ambiguity on the direction of these vectors is raised since the sulci are always extracted in one same direction, from their external trace until their fundus. Then $\mathbf{w_s}$ is obtained by the vector product: $\mathbf{w_s} = (\mathbf{u_s} \wedge \mathbf{v_s})$. The origin O_s is the center of mass of the sulcus. Note that this latter may not lie on the surface.

To express the sulci in $\mathcal{R}_s$, we determine, for each sulcus, the matrix $\mathbf{M}$ defining the change of basis from the local coordinate system $\mathcal{R}_s$ towards the image coordinate system $\mathcal{R}(O, \mathbf{u}, \mathbf{v}, \mathbf{w})$. Let $\mathbf{R}$ and $\mathbf{t}$ be the rotation matrix and the translation vector of the inverse change of basis $\mathbf{M}^{-1}$ (i.e. from $\mathcal{R}$ towards $\mathcal{R}_s$). Then, in homogeneous coordinates:

$$\mathbf{M}^{-1} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ 0\ 0\ 0 & 1 \end{pmatrix} \text{ where } \mathbf{R} = (\,\mathbf{u_s}\ \mathbf{v_s}\ \mathbf{w_s}\,) \text{ and } \mathbf{t} = \overrightarrow{OO_s}\,. \tag{17}$$

Since $\mathbf{R}$ is orthogonal:

$$\mathbf{M} = \begin{pmatrix} \mathbf{R}^T & -\mathbf{R}^T\mathbf{t} \\ 0\ 0\ 0 & 1 \end{pmatrix}. \tag{18}$$

The rigid transformation defined by the matrix $\mathbf{M}$, defines the local rigid registration process, which will henceforth be called LR. It is computed for each subject and applied to the associate sulcus. Applying this rigid transformation to all the points of each sulcus aligns the training set with respect to the local coordinate system. Note that no homothety is performed in this LR method. In fact, as the image data is acquired to the same scale, the inter-individual size variation can thus be captured in the shape model.

(b) Resampling

In order to define each instance of the training set by a same number of points, cortical sulci are resampled. The parametric representation of the sulcal surface facilitates this resampling on each axis.

Contiguous sulci. If all the sulci in $\mathcal{P}$ are contiguous, the instance described with the maximum number of points, nbp_{max}, is sought. Each other shape is oversampled with nbp_{max} sample points. In theory, this resampling modifies the discrete shape of the surface. However, taken into account the high initial sampling (1 mm step) and the performed oversampling, initial shapes are preserved in practice.

Interrupted sulci. In other cases, we have to deal with interrupted sulci. If all the sulci in $\mathcal{P}$ are interrupted and described with the same number of segments, we resample each segment as described in the contiguous case. This is possible since each segment is labeled, for example inferior/superior for the precentral sulcus or anterior/posterior for the temporal sulcus.

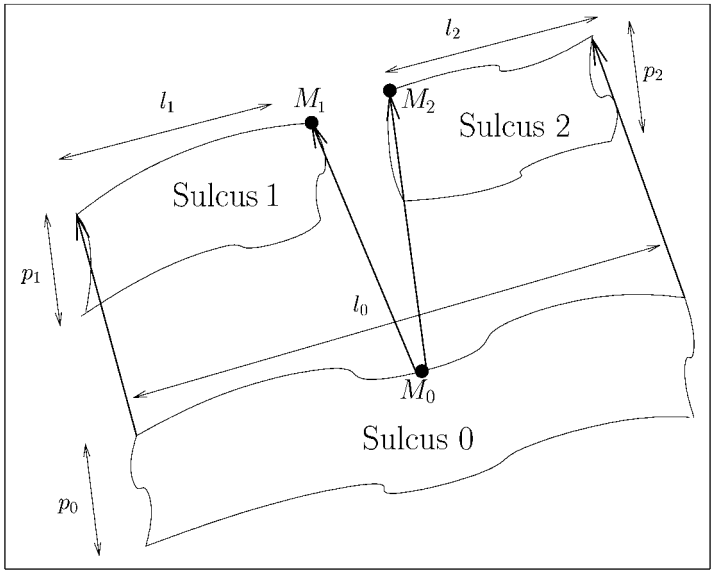


Fig. 4. Resampling and matching of two sulci, with one of them interrupted.

If $\mathcal{P}$ contains both contiguous and interrupted sulci as illustrated in Fig. 4, we proceed to a "proportional resampling". For the sake of concision, we consider only interrupted sulci in two pieces i and j, with depth p_i and p_j and with length l_i and l_j. The depth of a contiguous k sulcus is denoted p_k and its length l_k. Then:

- let p_m be the maximal depth over the training set $\mathcal{P}$, $p_m = \max_{\mathcal{P}}\{p_i, p_j, p_k\}$. We oversample the instances of $\mathcal{P}$ so that they are described by p_m sample points on the depth axis;
- let l_m be the maximal length over the training set, $l_m = \max_{\mathcal{P}}\{l_i + l_j, l_k\}$. Then,
  - if $l_m = l_i + l_j$, each contiguous sulcus k is sampled by a factor of $\frac{l_m}{l_k}$ to reach a length of l_m,
  - otherwise, for each interrupted sulci, its two segments are oversampled by a factor of $\frac{l_m}{l_i + l_j}$.

This method is generalizable to sulci having more than two segments.

(c) Matching

The instances of the training set being aligned in $\mathcal{R}_s$ and properly resampled, the associate set of control points is computed for each sulcus. The point-to-point correspondence between two sulci A and B is performed as follows: each control point of sulcus A is paired with the control point of sulcus B having the same curvilinear abscissae.

Figure 4 illustrates this matching scheme in the case where an interrupted sulcus is homologous to a contiguous sulcus.

(d) Discussion

The matching scheme we propose relies on two assumptions:

(1) a correspondence between two cortical sulci, moreover from two distinct subjects, with different topology, has sense; and
(2) such a correspondence can be established from the parameterization of the shape "sulcus".

Its anatomical, functional, physiological, etc. relevance cannot be guaranteed. Other approaches have been proposed but they suffer from the same uncertainties. Thus, Chui *et al.*[24] propose a Robust Point Matching (RPM) algorithm leading to a point-to-point correspondence between sulcal points. Caunce *et al.* use an ICP-like algorithm, Collins *et al.*[28] use a Chamfer distance to define correspondences between sulci while Vaillant *et al.*[118] rely on the curvature of the external traces of the sulci. The heterogeneity of these different techniques highlights the lack of anatomical and/or functional knowledge to rely on in order to establish the "correct" correspondence between cortical sulci from different subjects. Either based on curvature, or on distance metrics, or on parameterization, all these techniques are somehow arbitrary. To get rid off this arbitrary part, it should rely on an underlying "physiological" model that would guide the correspondence or at least that would define what is a "right" correspondence. To date and to our knowledge, such a model is still unknown. Cytoarchitectonic information might provide some hints for such purposes but this kind of microscopic information remains unaccessible *in vivo* with current scanners. Derived from the shape representation, our matching scheme has the advantage to be simple and to lead to a coherent training population for the statistical analysis.

3.2.4. *Statistical analysis of deformations*

The statistical analysis of the training set leads to a modeling of cortical sulci and of their variations. The model captures the shape variability observed within the training set. As a matter of fact, the statistical analysis reveals the main modes of variation relative to a reference shape, which is the mean shape over the training set in our case. We use a principal component analysis (PCA) which enables the representation of data on a new orthogonal basis, and which eliminates information redundancy. Moreover, this analysis allows modal approximation, performed by retaining the most significant modes.

Let $\mathcal{P}$ be the training population made up of N elements, $\mathbf{x}_i \in \mathcal{P}$ a shape, $\bar{\mathbf{x}}$ the mean shape on $\mathcal{P}$, $\mathbf{C}$ the covariance matrix. A shape $\mathbf{x}_{i,\, i=1\ldots N}$, is represented by the vector of control points of the spline which models the median surface of the sulcus:

$$\mathbf{x}_i = (x_{i_1}, y_{i_1}, z_{i_1}, \ldots, x_{i_n}, y_{i_n}, z_{i_n})^t \quad \text{with } n = nbc\,. \tag{19}$$

The mean shape, used as the reference shape, and the covariance matrix are given by:

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{20}$$

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N} d\mathbf{x}_i d\mathbf{x}_i^t \quad \text{with } d\mathbf{x}_i = \mathbf{x}_i - \bar{\mathbf{x}}\,. \tag{21}$$

Diagonalizing the covariance matrix $\mathbf{C}$ provides the new modal basis $\mathbf{\Phi}$:

$$\mathbf{C} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T \text{ where } \Lambda = diag(\lambda_1, \ldots, \lambda_{3n}) \quad \text{with } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{3n}\,. \tag{22}$$

Then any shape $\mathbf{x}$ can be written: $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi}\mathbf{b}$ where $\mathbf{b} = (b_1, \ldots, b_{3n})^t$ is the vector of modal amplitudes of deformation and $(-\mathbf{\Phi}\mathbf{b})$ corresponds to the deformation vectors from each point of $\mathbf{x}$ towards the mean shape. Since the eigenvalue λ_i is the variance explained by the i^{th} mode, a large part of the variability can be explained by retaining only the first m modes. The value m is chosen so that $\sum_{i=1}^{m}\lambda_i$, the variance explained by the first m modes, represents a desired proportion, p, of the whole variance λ_T; that is to say m is such that $\frac{\sum_{i=1}^{m}\lambda_i}{\lambda_T} \simeq p$ where $\lambda_T = \sum_{i=1}^{3n}\lambda_i$. Retaining only m modes achieves a modal approximation:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{\Phi_m}\mathbf{b_m} \tag{23}$$

$$\mathbf{b_m} = \mathbf{\Phi_m}^t(\mathbf{x} - \bar{\mathbf{x}}) \tag{24}$$

where $\mathbf{\Phi_m}$ is a submatrix of $\mathbf{\Phi}$ containing the first m eigenvectors of $\mathbf{C}$, thus defining the modal approximation basis. The vector $\mathbf{b_m} = (b_1, \ldots, b_m)^t$ represents a shape in the m-dimensional space defined by the principal components. The analysis is convenient since $\mathbf{b_m}$ provides a compact shape representation ($m < 3n$). However, $\mathbf{b_m}$ must be constrained in order to represent an "allowable" shape, i.e. a shape consistent with the learnt shapes. Given the assumption that the distribution of vectors $\mathbf{x}_i$ is gaussian, the range of variability of each $b_{i,\, i=1\ldots m}$, is typically such as:

$$-3\sqrt{\lambda_i} \leq b_i \leq +3\sqrt{\lambda_i}\,. \tag{25}$$

Given the assumption $\mathbf{x}_i \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{C})$, $b_i \sim \mathcal{N}(0, \lambda_i)$ and so $P(b_i \leq 3\sqrt{\lambda_i}) = 99.7\%$. Thus, (25) can be considered as a representativity condition of the class of objects of interest.

3.3. *Deformation field and non-linear registration*

The deformation field $(-\mathbf{\Phi_m}\mathbf{b_m})$ obtained between a given sulcus and the mean sulcus can be extended to the vicinity of the considered sulcus by using an appropriate interpolation, the thin-plate spline interpolation in our case.[11,42,100] It can then be applied to any object lying in a spatial neighborhood of this sulcus. We take advantage of this deformation field extension to register sparse functional data

(MEG dipoles in our case) located in the left central sulcus area towards a coordinate system relative to the mean left central sulcus.

3.3.1. *Interpolation using the thin-plate spline method*

The use of thin-plate spline interpolation for registration purposes in medical imaging was first proposed by Bookstein. In Ref. 11, he proposes an algebraic approach to describe deformations specified by two sets of corresponding points. This method provides an interpolation function f which maps one of the two sets of corresponding points, the source set, onto the other one, the target set. Moreover, function f is defined everywhere in Euclidean space, and particularly in a neighborhood of the source points set so that it can be applied to any point in the source space to find its homologous point in the target space.

Let $\mathcal{P} = \{P_i(x_i, y_i, z_i), i = 1, \ldots, n\}$ be the set of source points in Euclidean space, and $\mathcal{V} = \{V_i = (x'_i, y'_i, z'_i), i = 1, \ldots, n\}$ the set of target points. The set $\mathcal{P}$ describes a shape $\mathbf{x}$, expressed by $\bar{\mathbf{x}} + \mathbf{\Phi_m b_m}$ according to our model. Let $r_{ij} = |P_i - P_j|$ be the Euclidean distance between two source points P_i and P_j. Then the function f is the sum of two terms: an affine part which represents its behavior at infinity, and a second part which is asymptotically flat:

$$f(x, y, z) = a_1 + a_x x + a_y y + a_z z + \sum_{j=1}^{n} w_j U(|P_j - (x, y, z)|) \tag{26}$$

where

- the basis function U is the fundamental solution of the biharmonic equation $\Delta^2 U = \delta(0, 0)$, δ being the Kronecker function. It can be shown[12] that the equation of a thin uniform metal plate originally flat and now bent by vertical displacements is directly related to the biharmonic equation. In 3D, the function U is $U(r) = |r|$ (whereas $U(r) = r^2 \ln r$ in 2D and $U(r) = |r|^3$ in 1D);
- the coefficients $\mathbf{a} = (a_1, a_x, a_y, a_z)^t$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)^t$ are obtained by solving the linear system:

$$\begin{cases} \mathbf{Kw} + \mathbf{Pa} = \mathbf{v} \\ \mathbf{P}^t \mathbf{w} = 0 \end{cases} \quad \text{where } \mathbf{P} = \begin{pmatrix} 1 & x_1 & y_1 & z_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & y_n & z_n \end{pmatrix} \tag{27}$$

  $\mathbf{K}$ is a $n \times n$ matrix, whose general term is $(U(r_{ij}))_{1 \le i,j \le n}$, and $\mathbf{v}$ is the vector of one coordinate of the target set (e.g. $\mathbf{v} = (x'_1, \ldots, x'_n)$, which implies that (26) must be expressed for $f_x(x, y, z)$, $f_y(x, y, z)$ and $f_z(x, y, z)$).

Function f is computed from a sulcus shape $\mathbf{x}$ and the mean shape $\bar{\mathbf{x}}$, these two shapes being defined by their control points and matched point-to-point as described in Sec. 3.2.3. The shape $\mathbf{x}$ is regarded as the source set and the mean shape

as the target set. The deformation field $(-\mathbf{\Phi_m b_m})$ between $\mathbf{x}$ and $\bar{\mathbf{x}}$ is then represented by the elements of $(\mathbf{w} \mid \mathbf{a})$ and corresponds to the restriction of f to $\mathbf{x}$. But, it is henceforth defined also outside the source shape $\mathbf{x}$. We can then apply it to points in the vicinity of the source shape $\mathbf{x}$ to register them towards the mean shape $\bar{\mathbf{x}}$.

3.3.2. *Specification of the vicinity of a sulcus*

Function f is defined on $\mathbb{R}^3$, but has to be applied only in the vicinity of the constraint. In fact, we are interested in propagating the deformation all over the "influence range" of the sulcus of interest, but not all over the brain. Therefore it is desirable to explicitly limit the influence of the thin-plate spline interpolation to a pre-defined neighborhood of the concerned sulcus, which in practice would be the adjacent gyri. This can be achieved by defining a parallelepipedic bounding volume, V, specifying the neighborhood. However, in order to preserve the continuity of the transformation, we define a "fuzzy volume" through which the deformation will cancel itself out in a continuous way. This fuzzy volume can be described through a function, μ_V, defined as:

$$\begin{aligned} \mu_V &: \mathbb{R}^3 \to \mathbb{R} \\ X = (x, y, z) &\mapsto \mu_{V_x}(x).\mu_{V_y}(y).\mu_{V_z}(z) \end{aligned} \tag{28}$$

where functions $\mu_{V_\cdot}$ have the following properties:

- defined on $\mathbb{R}$,
- C^2, i.e. twice differentiable and the second order derivative is continuous,
- decreasing on $\mathbb{R}^+$,
- even (this last property is sufficient when working in the local space),
- $\exists t \in \mathbb{R}^{+*}$ such that $\mu''_{V_\cdot}(t) = 0$, $\mu''_{V_\cdot}$ being the second order derivative of $\mu_{V_\cdot}$;

where V_x, V_y and V_z specify the limits of the bounding box in directions x, y and z. For example, if $V_x = [-t_x, t_x]$ then μ_{V_x} will be chosen such that $\mu''_{V_x}(t_x) = 0$.

Given such a μ_V function, let a point $X = (x, y, z) \in \mathbb{R}^3$ and $X_f = (x_f, y_f, z_f)$ be its image point by the thin-plate spline interpolation limited to one neighborhood V. Then,

$$X_f = f(X).\mu_V(X) + X.\mu_{\bar{V}}(X) \tag{29}$$

where $\mu_{\bar{V}} = 1 - \mu_V$.

Figure 5 gives examples of possible functions $\mu_{V_\cdot}$:

(1) $\forall t \in \mathbb{R}$, $\mu^{(1)}_{V_\cdot}(t) = \dfrac{1}{1 + kt^2}$, $k > 0$

(2) $\forall t \in \mathbb{R}, \mu^{(2)}_{V_\cdot} = \begin{cases} 1 & \text{if } 0 \le |t| \le a \\ \dfrac{1}{2} - \dfrac{1}{2} sin\left[\dfrac{\Pi}{b-a}\left(|t| - \dfrac{a+b}{2}\right)\right] & \text{if } a \le |t| \le b \\ 0 & \text{if } b \le |t| \end{cases}$

with a and $b \in \mathbb{R}^+$ such that $a < b$

(3) $\forall t \in \mathbb{R}$, $\mu^{(3)}_{V_\cdot}(t) = e^{-kt^2}$, $k > 0$

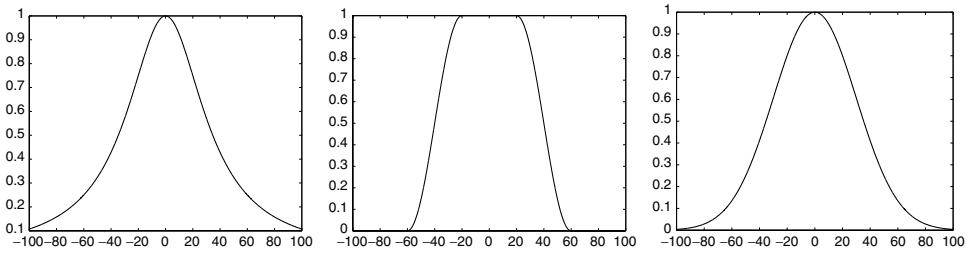


Fig. 5. Examples of $\mu_{V_\cdot}$ functions. From top to bottom: $\mu_{V_\cdot}^{(1)}$ (inverse) such that ${\mu_{V_\cdot}^{(1)}}''(20) = 0$; $\mu_{V_\cdot}^{(2)}$ (sinus) such that ${\mu_{V_\cdot}^{(2)}}''(40) = 0$; $\mu_{V_\cdot}^{(3)}$ (exponential) such that ${\mu_{V_\cdot}^{(3)}}''(30) = 0$.

We have simulated a set of source and target points (see Fig. 6(a), and computed the f function corresponding to these two sets. Figure 6(b) shows the result of the application of f on the whole grid. Although the deformation weakens far from the constraint, it remains influent. Conversely, Figs. 6(c) and 6(d) illustrate the effect of the application of a restriction of f to a specified neighborhood; Fig. 6(c) via a door function, Fig. 6(d) via one $\mu_{V_\cdot}$ function. The limitation of the deformation via a door function is efficient, but does not preserve the continuity. On the other hand, the restriction of the extrapolation to the vicinity of the constraint via one $\mu_{V_\cdot}$ function is effective and continuity preserving, as can be seen in Fig. 6(d).

The "fuzzy bounding box" will play a role if data out of the vicinity of the sulcus have to be treated or if deformation fields from different sulci are desired to be used for example. The choice of $\mu_{V_\cdot}$ as well as the tune of the limits of the box is incumbent upon an expert. However, some single rules can be designed depending on the application. For instance, one could define t_z as the thickness of a gyrus and thus consider adjacent gyri as the vicinity of the sulcus, or one could define t_z as the "mid-thickness" of a gyrus, where the mid-thickness of a gyrus could be half the distance between the gravity centers of its two bounding sulci (like central and postcentral sulci for the postcentral gyrus for example).

4. Hybrid Registration Approaches

In this section, we present a unified framework for non-rigid registration of brains, combining a global registration approach and sparse constraints. In this case, we explicitly use it with cortical sulci constraints. The section is organized as follows: we first discuss related work and then detail the hybrid registration method.

4.1. *Related work*

The cooperative approaches (photometric methods based on intensity and geometric methods based on landmarks) have received recently a growing attention.

Gee *et al.*[49] have proposed a Bayesian unified framework to this problem but without experimentations on real 3D data. Actually, the problem is to find a

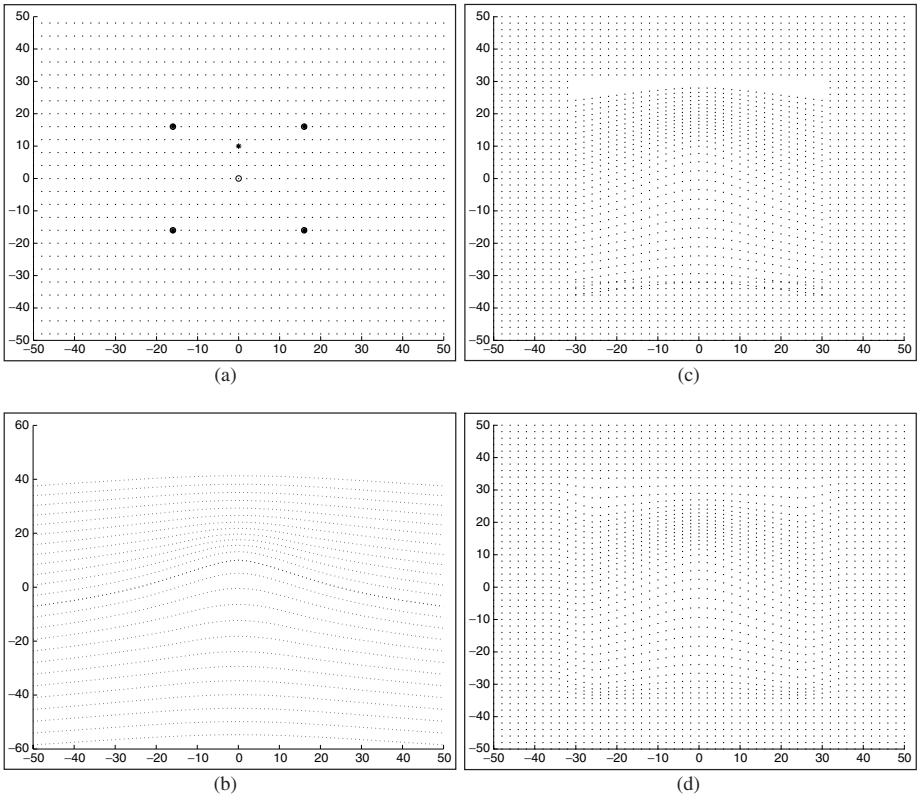


Fig. 6. (a) Simulated data: circles are source points, asterisks are target points and points represent the initial grid; (b) The initial grid is deformed by applying the f function computed from the sets of source and target points; (c) The application of the f function is restricted thanks to a non-continuous function (a door function); (d) The application of the f function is restricted thanks to a continuous function $\left(\text{the sinus function, i.e. } \mu_{V_\cdot}^{(2)}\right)$.

transformation that more or less conserves the topology of the brain while being able to adapt to local topological variations. For instance, a cortical sulcus may be present in one segment for one subject while being split in three parts for a second one, or even absent for a third subject. These variations can even be observed between two hemispheres of a single subject.

Many applications where brain warping is needed face this problem. There are some strong arguments to control non-rigid matching of brain data by local cortical landmarks. The evaluation on real data have shown first the inadequacies of methods using only image intensities,[58] and also the benefits of incorporating local cortical constraints.

Collins *et al.*[28,30] investigated the introduction of sulcal constraints, which have been introduced on the basis of a Chamfer distance between corresponding sulci of the source and target volumes. However, sulcal constraints have been introduced

without any specific formal framework, only the orthogonal projection of a sulcus onto its correspondent is used. Following the same idea, Vaillant *et al.*[118] use cortical constraints for non-rigid registration. The 3D registration is viewed as a deformation of a surface on which the sulcal constraints are defined like curves, on the brain outer surface. The elastic registration tries to match points with similar curvatures, which is a questionable assumption. This approach is rather similar to the one proposed in Ref. 113, where the surface deformation takes only into account the sulcal constraints detected on the brain external surface. Chui *et al.*[24] propose the Robust Point Matching (RPM) algorithm, which explicitly incorporates sparse constraints. The sulcal points are manually extracted by an expert, and are integrated in a minimization algorithm which first seeks a global affine transformation, then a piecewise affine one. More recently, Cachier *et al.*[19] proposed a demons-like algorithm which incorporates matching of sulcal roots. Hartkens *et al.*[128] proposed to integrate points and surfaces in a multimodal non-rigid registration algorithm. Johnson and Christensen[64] presented a consistent 2D thin-plate-spline registration method where manually-detected landmarks and intensity cooperate to compute the registration.

4.2. *Definition of the sparse constraint*

Once cortical sulci are extracted using the method described previously (see Sec. 3.2.1), we define the sparse constraint, which will be used to drive the registration process. We define the constraint as a sparse deformation field.

4.2.1. *Matching sulci*

For the registration of a subject A toward a subject B, we consider the homologous sulci between these two subjects. We then apply the resampling and matching scheme presented in Secs. (i) and (j). On the kernel of the constraint sulci of the source volume, we thus define a sparse constraint deformation field.

From Fig. 4, the curvilinear abscissa of point M_0 along the sulcus 0 length is $\frac{l_1}{l_0}$, given that this curvilinear abscissa is normalized between 0 and 1. At M_0, the constraint field corresponding to the sulci matching is not contiguous.

4.2.2. *Sparse constraint deformation field*

In every case, we finally obtain two sulci each described by N control points. The sulcus of the source volume is described by a set of control points $S_1 = \{C_S^1 \dots C_S^N\}$, and the homologous sulcus in the target volume is described by $S_2 = \{C_T^1 \dots C_T^N\}$. For each point S_1, a constraint field can be explicitly computed: $\forall k \in \{1 \dots N\}, \boldsymbol{w}_k^c = \boldsymbol{C_S^k C_T^k}$. Let us note $\mathcal{S}_c = S_1$ the kernel of the sparse constraint field.

Contrary to the matching approaches based on a distance measure (Chamfer, ICP), this algorithm matches explicitly all sulci points. This matching procedure is

questionable, since we do not know if it is anatomically plausible. Is it necessary to explicitly match the sulci extreme points? How to manage the sulci interruptions? In absence of anatomical assumptions, and at a primer stage, we have chosen this cortical mapping process. At the end of this process, we obtain a constraint field $\boldsymbol{w}^c$, which is defined on the support of source volume.

In order to reduce the sensitivity of the constraint field to sulci segmentation errors, we perform a non isotropic regularization of the field. This can be linked to a 3D adaptation of the Nagao filtering.[87] The field is smoothed with an anisotropic filter that considers voxels belonging to the kernel of the sparse constraint field. We perform the smoothing for a given number of iterations. At each step, a median filtering of the field is performed. The median filtering of the field is equivalent to a median filtering on each coordinate. This non-linear filtering is performed only for sulcal points (therefore, the kernel for the convolution is adaptive). As a consequence, the smoothed sparse constraint field is defined at the same points as the original field.

In the following, the proposed approach considers the existence of a constraint field $\boldsymbol{w}^c$, which is not, in principle, strictly limited to the incorporation of sulcal constraints. Additional landmarks, of various dimensionality, can be introduced within the same formalism.

4.3. *Integration of sparse constraints*

The sparse deformation field $\boldsymbol{w}^c$ must be integrated in the formulation of the registration problem. As in Ref. 49, this information is incorporated as a third energy term. The cost function thus becomes:

$$U(\boldsymbol{w}; f, \boldsymbol{w}^c) = \sum_{s\in S}[\boldsymbol{\nabla} f(s,t)\cdot \boldsymbol{w}_s + f_t(s,t)]^2 + \alpha \sum_{\langle s,r\rangle\in\mathcal{C}} \|\boldsymbol{w}_s - \boldsymbol{w}_r\|^2 + \alpha^c \sum_{s\in\mathcal{S}_c} \|\boldsymbol{w}_s - \boldsymbol{w}_s^c\|^2 \tag{30}$$

where α^c is a parameter that balances the weight of the sparse constraint.

The matching of local structures might not be correct for all the points. As a matter of fact, there might be some segmentation errors, and these points should not be used as a hard constraint. Furthermore, it might not be anatomically correct to assume a one-to-one correspondence between the landmarks. Therefore, we introduce a robust estimator ψ_3 on the local constraint term. The cost function is modified as:

$$\overset{\star}{U}(\boldsymbol{w}, \delta, \beta; f) = \sum_{s\in S} \delta_s(\boldsymbol{\nabla} f\cdot \boldsymbol{w}_s + f_t)^2 + \alpha \sum_{\langle s,r\rangle\in\mathcal{C}} \beta_{sr}(\|\boldsymbol{w}_s - \boldsymbol{w}_r\|)^2 + \alpha^c \sum_{s\in\mathcal{S}_c} \gamma_s(\|\boldsymbol{w}_s - \boldsymbol{w}_s^c\|)^2 + \psi_1(\delta_s) + \psi_2(\beta_{sr}) + \psi_3(\gamma_s)\,. \tag{31}$$

The sparse constraint and the associated robust function introduce two new external parameters, α^c and ψ_3. We have chosen $\alpha^c \gg \alpha$, so that the constraint is largely taken into account.

The minimization scheme is unchanged, with respect to our previous work.[58] We alternate between estimating the weights of the robust functions and estimating the deformation field. Once the weights are estimated and "frozen", the multigrid estimation of the field is performed through an iterative Gauss-Seidel scheme.

The local constraint has a relative spatial influence for two reasons. Firstly, the standard regularization term propagates the local constraint because the minimization is alternated. Additionally, the multigrid minimization, described in details in our previous work[58] makes it possible to estimate a deformation model on specified cubes. This propagates the local constraint to a large group of voxels that compose the cube.

5. Experiments and Results

We have designed a comparative evaluation framework on anatomical and functional data to assess the quality of registration methods. So far, the evaluation framework has been applied to eight global registration methods.[57,59,60] In this section, we consider the three proposed methods, i.e. the Romeo method (see Sec. 2.3), the non-linear local method (see Sec. 3) and the hybrid registration method (see Sec. 4), and two other classic global registration methods for comparison purposes. Evaluation is based on a set of meaningful anatomical and functional criteria. We assume that anatomical landmarks can be superimposed between individuals and that global registration methods should align these anatomical features. Since a major issue associated with registration is the matching of anatomy and function between individuals, we focus on the matching of cortical structures, i.e. grey and white matter, gyri and sulci. Besides, we test the ability of each method to register functional data and subsequently to reduce the functional inter-subject variability. The underlying assumption is that the functional inter-subject variability can broadly be decomposed into an anatomical variability, which might be estimated with registration methods, and a residual functional variability. To be objective, the evaluation must rely on features independent from the similarity used to drive the registration process.

5.1. *Data*

We have acquired a database of 18 subjects, male, age 35 ± 10, right handed and healthy. Each subject underwent a T1-magnetic resonance (MR) SPGR 3D study (GE 1.5T system, sagittal slices). The raw MR data have been linearly interpolated so that the voxel resolution is isotropic (the voxel resolution is 0.9375 mm, except for four subjects where the resolution is 0.976562 mm). For all subjects, the volume size is $256 \times 256 \times s$ where s is the number of slices and is subject specific.

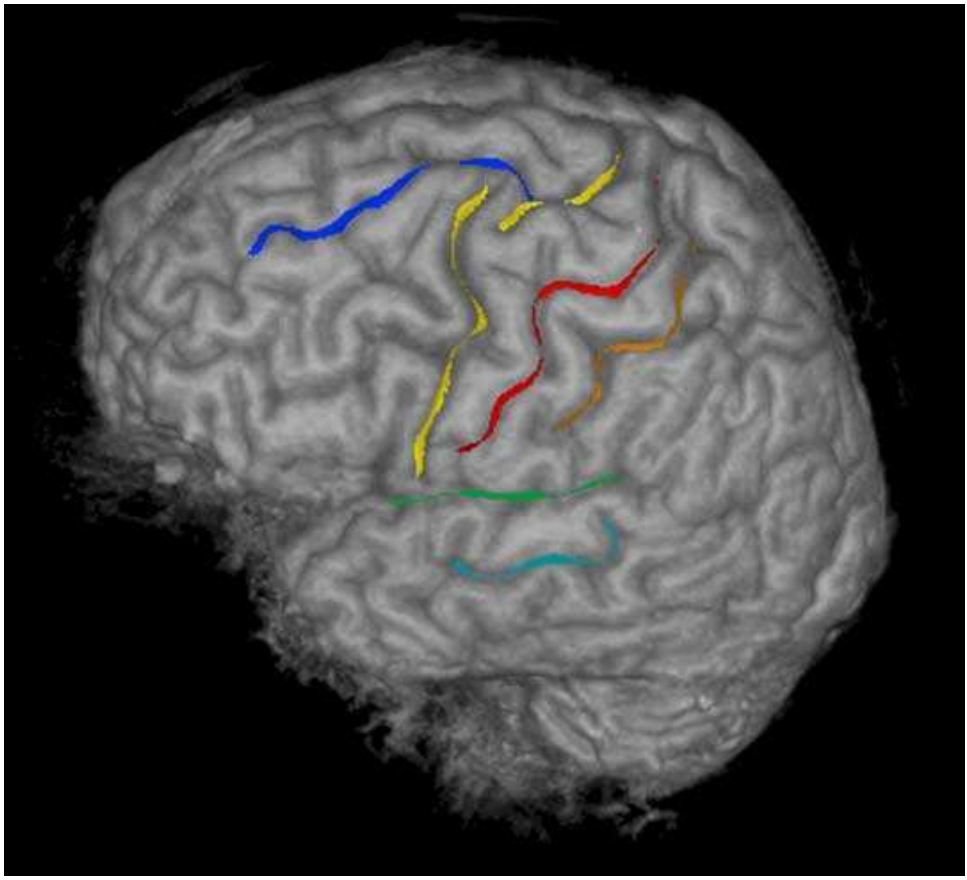


Fig. 7. Extracted sulci on the left hemisphere of one subject. From top to bottom: the superior frontal sulcus, the precentral sulcus, the central sulcus, the lateral sulcus and the superior temporal sulcus.

Additionally, for each subject and each hemisphere, six major sulci have been extracted from MR volumes by the "active ribbon" method (see Sec. (g)). They are: precentral sulcus, central sulcus, postcentral sulcus, superior frontal sulcus, superior temporal sulcus and lateral sulcus. An illustration of such extracted sulci for one subject is given in Fig. 7.

The functional data to register are MEG dipoles corresponding to a somatosensory activation of right hand fingers (thumb, index, little finger) performed by 15 volunteers out of the 18 subjects database. MEG current dipoles have been reconstructed using a spatiotemporal algorithm,[105] and selected by choosing the most significant one in the $45 +/- 15$ ms window. Thus, three dipoles, one per finger, are available for each of the 15 volunteers. The somatosensory paradigm chosen here is a very simple well-known one and is thus convenient to this evaluation, since our objective is not to explain complex physiological processes but rather to study

the behavior of such registration methods. The considered dipoles are expected to follow a somatotopic organization. First, they must be located in the left central sulcus area, and more precisely in the back of the central sulcus, i.e. in the post-central gyrus. Second, they must be identically ordered on the gyrus through the database, i.e. from the little finger to the thumb via the index according to the head-foot direction. Despite the simplicity of the protocol, reconstruction of the sources in MEG,[105] and MEG/MRI registration,[106] remain challenging tasks and the position of the MEG dipoles can be marred with errors. As a consequence, some of the dipoles are obviously mislocalized, like those situated at the wrong side of the sulcus for example. Dipoles belonging to gyri adjacent to the central sulcus are considered to be in the vicinity of the central sulcus.

5.2. *Methods*

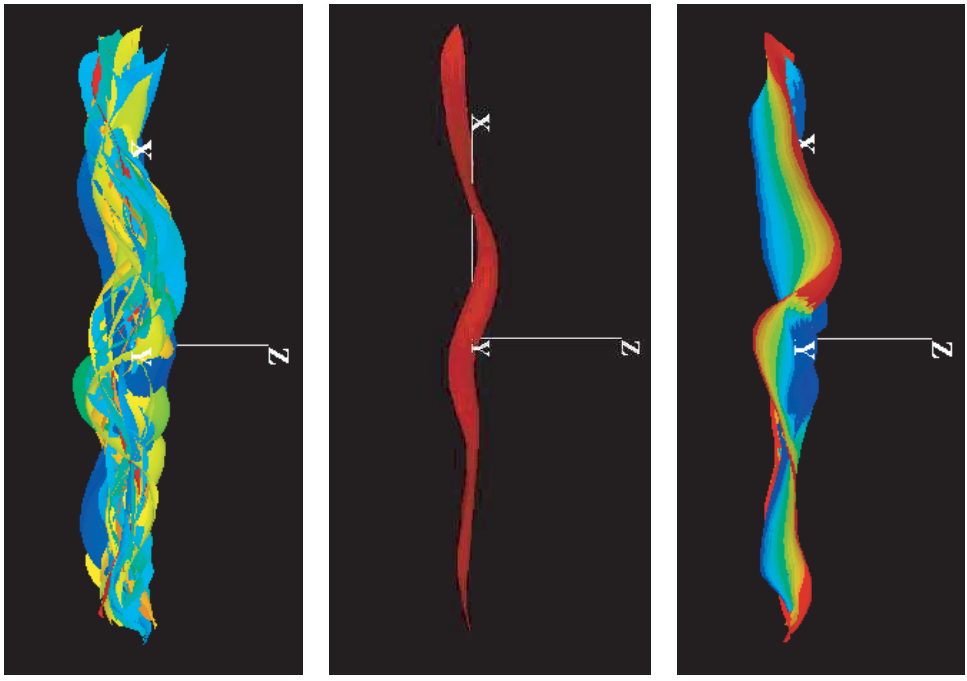
Our comparative evaluation framework considers five global methods and three local methods. The five global registration methods are:

- **MI**, a global rigid method based on maximization of mutual information.[27,122]
- **PS**, the Talairach Proportional Squaring[109] which is a global piecewise affine registration method. An affine transformation is computed on each of the 12 subvolumes defined by the Talairach Proportional Squaring.
- **S**, the Statistical Parametric Mapping (SPM) which is a template-based deformable registration.[1]
- **R**, the non-linear global "Romeo" registration method described in Sec. 2.3.
- **H**, the hybrid registration method presented in Sec. 4.

An arbitrary subject is chosen as the reference subject (see Fig. 10, top left). Given one method, every subject (source image) is registered to the reference subject (target image) so that all the registration results can be compared in the same reference frame. All extracted features can then be deformed toward the reference frame by the computed deformation field or transformation.

The three local registration methods are:

- **LR**, the local rigid method described in Sec. (h). This method is defined by a change of basis and registers each dipole, initially expressed in its image coordinate system, into the local system $\mathcal{R}_s$.
- **NLL**, the non-linear local method described in Sec. 3.3. Dipoles are first registered into $\mathcal{R}_s$ by the method LR. The statistical shape model of the left central sulcus is built as explained in Sec. 3.2. The training set, the obtained mean shape as well as the cumulative variance percentage and the variations along the first mode can be observed in Figs. 8 and 9. For each subject S, the extended deformation field $(\mathbf{w} \mid \mathbf{a})$ between $\mathbf{x}$, the left central sulcus of S, and $\bar{\mathbf{x}}$, the mean left central sulcus is computed and applied to each of the three dipoles of S. We present two tests differing by the number of modes, m, used in the construction



(a) (b) (c)

Fig. 8. Statistical shape model of the left central sulcus. The training set consists of 18 left central sulci. Directions of the axes are as follows: x = foot-head, y = toward the outside of the brain, z = anteroposterior. (a) Side view of the training set aligned in the local coordinate system, $\mathcal{R}_s$; (b) Side view of the mean left central sulcus; (c) Variations of the first mode around the mean sulcus, $-3\sqrt{\lambda_1} \leq b_1 \leq +3\sqrt{\lambda_1}$. The synthesized shapes are superimposed and a color is attributed to each of them.

of the deformation field ($-\mathbf{\Phi_m b_m}$) and so in the reconstruction of sulcus $\mathbf{x}$:

— The first one is performed with all the modes ($m = 17$) and the method NLL is then called **NLL1**.
— The second one is performed with only the five most significant modes of variation ($m = 5$). The method NLL is then called **NLL2**.

Note that the local methods are only applied on functional data (see Sec. 5.4).

5.3. *Evaluation on anatomical data*

The five global registration methods are assessed on anatomical data according to a set of criteria. In case of dense features, i.e. tissues, the evaluation criteria measure overlap after registration and are referred to as "global measures". In case of sparse features, i.e. cortical sulci, the evaluation criteria measures distances between features after registration and are referred to as "local measures".

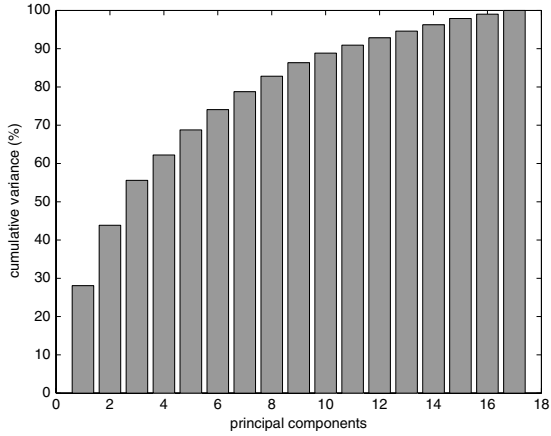


Fig. 9. Percentage, τ_p, of cumulative variance according to the number of modes retained for left and right hemisphere; $\tau_p = \sum_{i=1}^{m} \lambda_i / \lambda_T \times 100$. The training set consists of 18 left central sulci.

5.3.1. *Global measures*

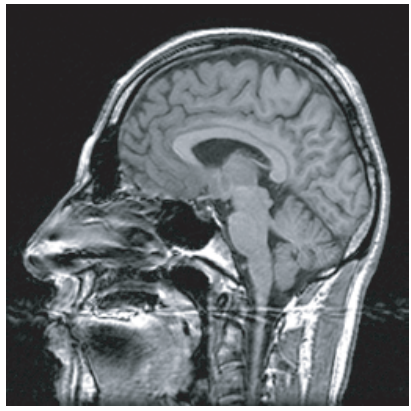
Global measures estimate how deformed source volumes overlap with the target volume. Three criteria are defined:

- Average volume. The average volume of deformed volumes is compared to the reference volume by visual observation. The mean square error between the average volume and the brain segmentation mask of the reference subject is computed. This criterion is partly related to the similarity measures used in each method but PS.
- Overlap of grey matter (GM) and white matter (WM) tissues. For each subject, grey and white matter are deformed toward the reference subject by the deformation field or transformation corresponding to each method and a trilinear interpolation. A total performance measure assesses the overlap.[119]
- Correlation of *Lvv*. Correlation between deformed *Lvv* of each subject and the *Lvv* of the reference subject is computed. The *Lvv* operator[47] has a very precise interpretation: it can be demonstrated that, when limited to the cortical region of interest, the crest of a gyrus corresponds to a negative value of the *Lvv* while a deep fold like a sulcus corresponds to its positive part.

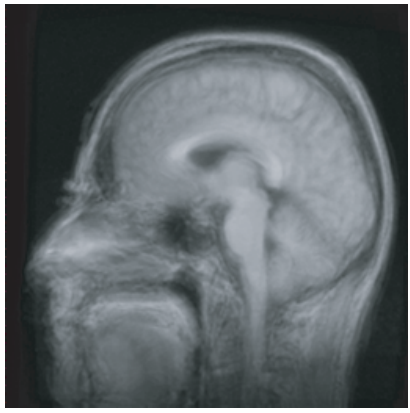
Figure 10 shows sagittal cut-planes through the average volumes as well as the mean square errors, while Table 1 gives tissue overlap and *Lvv* correlation.

5.3.2. *Local measures*

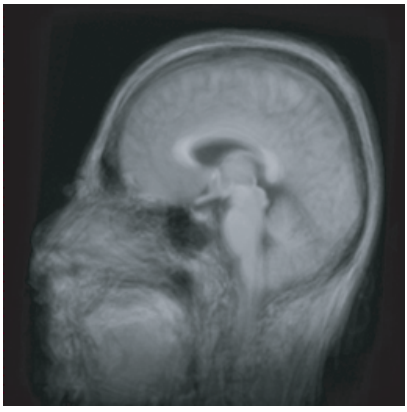
Local measures focus on anatomical sparse features and especially estimate how well cortical sulci are matched after registration. Visual observation is first performed. Figure 11 shows how the left central sulci of the 17 subjects deform toward the left central sulcus of the reference subject. Beyond visualization, we investigate two



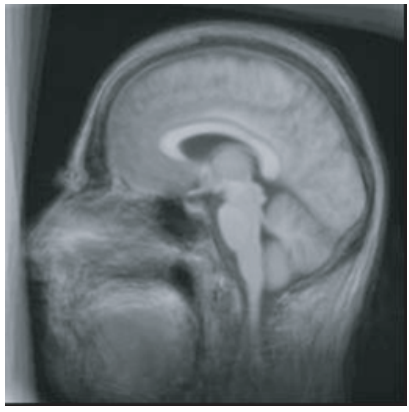
Reference subject



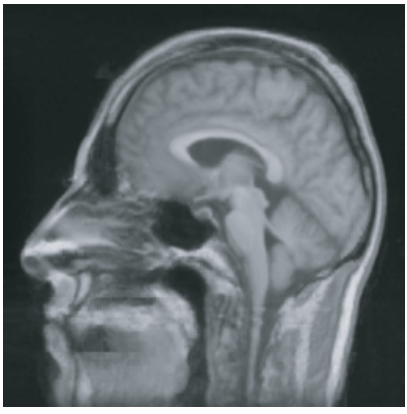
MI, MSE: 1390



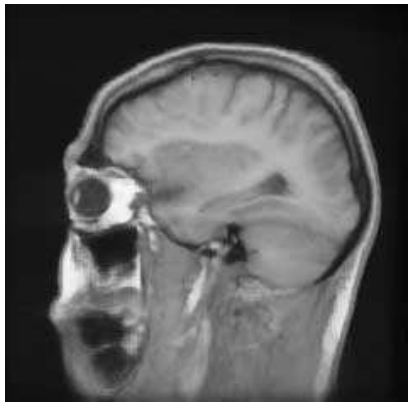
PS, MSE: 1064



S, MSE: 956



R, MSE: 624



H, MSE: 506

Fig. 10. Average volume. For each method (MI, PS, S, R and H), the average volume over the 17 deformed subjects can be compared to the reference subject (top left). The mean square error (MSE) between average and reference volumes is also given for each method.

Table 1. Tissue overlap and *Lvv* correlation. For each method, mean and standard deviation of these measures (overlap is expressed in percent) are computed over the database of subjects.

Methods	GM Overlap	WM Overlap	Correlation of *Lvv*
MI	88.8 ± 0.13	87.5 ± 0.17	0.01 ± 0.001
PS	93.5 ± 0.06	95.1 ± 0.04	0.16 ± 0.003
S	94.1 ± 0.06	95.7 ± 0.04	0.25 ± 0.003
R	93.9 ± 0.07	95.2 ± 0.07	0.32 ± 0.008
H	93.5 ± 0.05	95.3 ± 0.05	0.27 ± 0.004

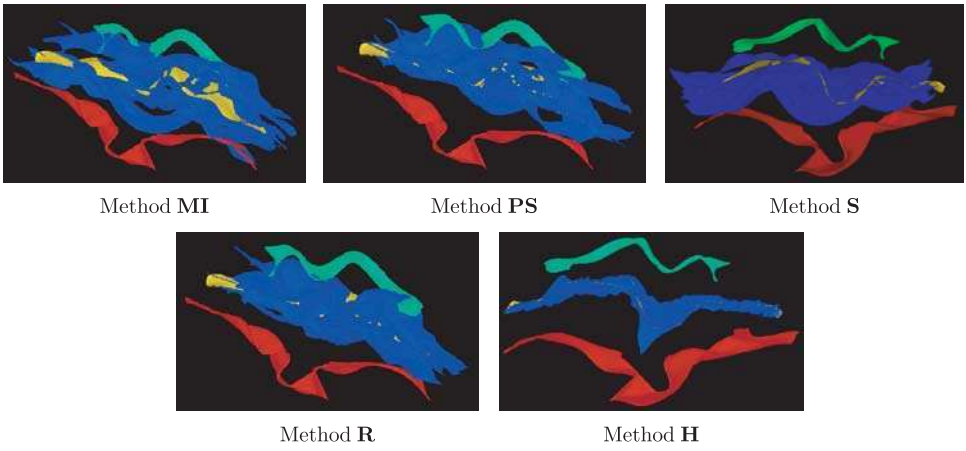


Fig. 11. Visualization of left central sulci deformed toward the reference subject. Central, precentral and postcentral sulci of the reference subject are represented in lighter grey, with the precentral sulcus at the top of each picture and the postcentral sulcus at its bottom. Darker grey sulci are the sulci deformed by each method (MI, PS, S, R and H).

measures that reflect more or less the global positioning of sulci on the one hand and the similarity of shapes on the other hand:

- Distance between registered sulci and reference sulcus. The control point-to-control point distance between a deformed sulcus and the corresponding sulcus of the reference subject is computed and averaged over all the sulci and all the subjects to provide a compact dissimilarity measure between sulci deformed by each method and the corresponding reference sulcus. Results are given in Table 2.
- Statistical shape analysis of registered sulci. For each sulcus, e.g. left central sulcus, a principal component analysis is applied on the set of deformed sulci. The reference shape is this time the reference sulcus (instead of the mean shape, see Sec. 3.2.4). We consider the trace of the covariance matrix, it reflects the dispersion of the registered sulci with respect to the corresponding reference sulcus. For each method, Table 2 compares the traces for three different sulci: left central sulcus, left frontal superior sulcus and left lateral sulcus.

Table 2. Local measures. First column: average distance (in voxels) over all sulci and all subjects, between registered sulci and the corresponding sulcus of the reference subject. Three left columns: trace of the covariance matrix (divided by the number of subjects) for three sulci: left central sulcus, superior temporal sulcus and lateral sulcus.

Method	Mean Distance	Central	Superior Frontal	Lateral
MI	11.5	621	622	1373
PS	10.7	510	859	1233
S	8.7	475	589	930
R	10.8	655	682	1052
H	1.2	87	132	119

5.3.3. *Discussion*

According to the first global criterion, we distinguish methods MI and PS on the one hand and methods S, R and H on the other hand, with method H performing even better. Methods PS, S, R and H appear quite similar in terms of tissue overlap while method MI gives the poorest results. Conversely, none of this method produces satisfactory results in regards to the correlation of Lvv. Accordingly, with respect to global measures, we note a difference in favor of non-linear global methods versus rigid methods. This observation might suggest that performance depends on the degrees of freedom of the estimated transformations. Indeed, the ranking of the methods according to their degrees of freedom and to these results would be more or less the same, with MI and PS on the one side, and S, R and H on the other side.

More surprisingly, local measures cannot reveal the superiority of one global method over the others, though method S performs slightly better. As expected, method H which integrates cortical constraints in the registration process gives significantly better results.

5.4. *Evaluation on functional data*

The five global methods and the three local methods are now compared in their ability to register functional data. In contrast with the previous section on anatomical data, the used functional data, i.e. MEG dipoles, are not involved in the registration process. In other words, this part of the evaluation is completely independent of the registration process.

We design several criteria to assess the quality of the registration. This one is mainly measured in terms of dipole dispersion since registration methods are expected to reduce dissimilarities within the original data.

We observe in Fig. 12 that the dipoles after registration with the reference sulcus superimposed for global and hybrid methods, and with the mean sulcus superimposed for local methods. Methods MI, PS and R show quite similar dispersion while methods S and H display reduced dispersion. Compared to the LR method, method NLL gathers the dipoles around the plane of the mean sulcus.

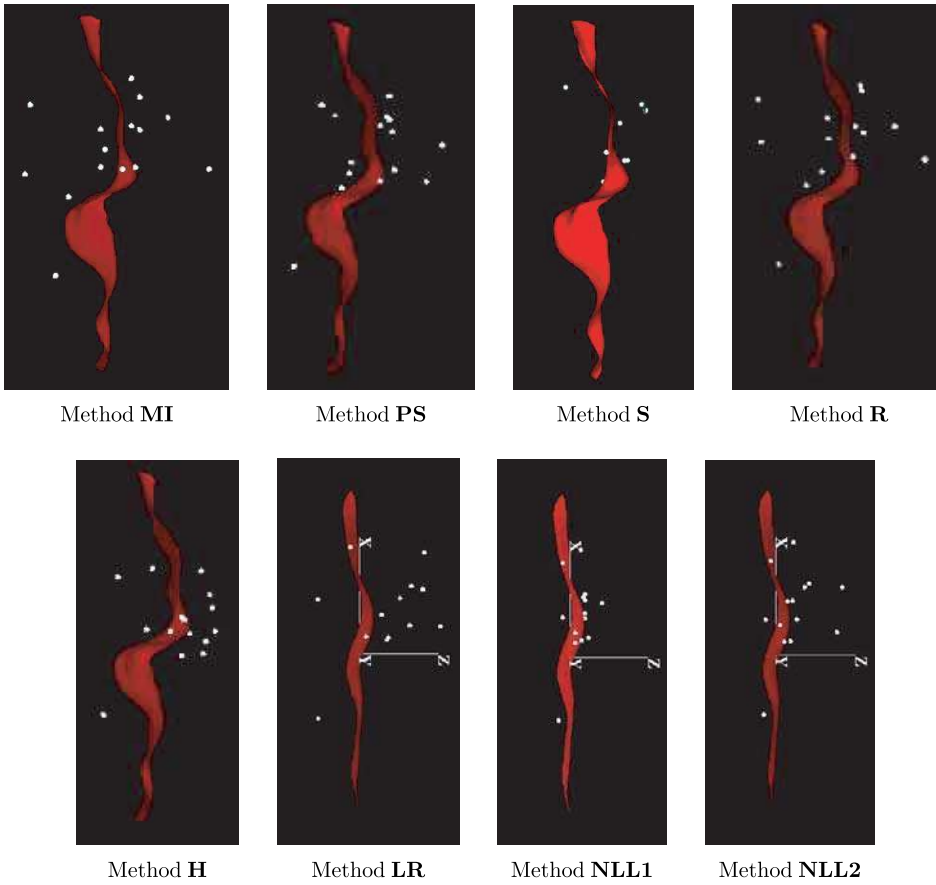


Fig. 12. Registration of MEG dipoles corresponding to a somatosensory activation of the little finger by each method. The visualized sulcus represents the left central sulcus of the reference subject for global and hybrid methods (MI, PS, S, R and H) whereas it represents the left central mean sulcus as local methods for local methods (LR, NLL1 and NLL2).

It is possible to numerically assess these observations by computing the covariance matrix and its determinant which provides objective measures of the 3D dispersion of the registered dipoles. These results are presented in Table 3. The determinant indicates the smallest dispersion for method NLL. We note that PS and S perform the best among the global methods, but they are outcast by the hybrid method which significantly reduces the dispersion in comparison with the unconstrained method R.

The flattening in the z direction by NLL method could be expected since the deformation is normal to the mean sulcus plane in this coordinate system. The observed anatomical deformation between the two sulci is indeed more important along this axis and the NLL method is specifically designed to interpolate it between the two sulci. With a limited number of modes (method NLL2), the

Table 3. Determinant of the covariance matrix of MEG dipole localizations for somatosensory activations (thumb, index finger, little finger) after global registration methods: MI, PS, S, R, hybrid registration method: H and local registration methods: LR, NLL1 and NLL2.

Methods	MI	PS	S	R
Thumb	175873	65201	79000	119487
Index finger	241181	112660	125000	245742
Little finger	268867	98649	220048	270147
Methods	H	LR	NLL1	NLL2
Thumb	33387	217671	6439	30793
Index finger	73612	292114	13479	45125
Little finger	62647	343260	18735	91921

observed dipole dispersion, in accordance with the determinant of the covariance matrix, is larger than with all the modes (method NLL1), but still lower than with LR method. This shows the interest and the relevance of a compact shape representation and this can be particularly useful to treat larger databases. Using only five modes in the construction of $(-\mathbf{\Phi_m b_m})$ can be considered as an approximation regularized by the training. This keeps the "principal shape" of the sulcus characterizing the subject and discards atypical features of the database. These features may result from processing errors (like non-systematic segmentation errors) as well as from real non-typical shape features present in the training set.

The assumption underlying this study, stating that a part of the functional inter-subject variability is encoded in the anatomical variability, also states that a part is decorrelated from the anatomy and cannot be reduced with an anatomical matching. The results show that methods integrating anatomical constraints like the hybrid method H or methods fully based on anatomical constraints like the non-linear local method NLL indeed reduce the initial functional inter-subject variability. The residual variability we can observe results from this irreducible variability plus, of course, the errors deriving from the preprocessing stages (multimodal registration, dipoles reconstruction, etc).

We state a significant difference between the hybrid and local methods and the global methods. The case of the PS method, which distinguishes itself from the other global methods, may be explained since method PS is by construction relevant and precise in the central region (it in fact relies on the location of the anatomical points AC-PC). To highlight the significant difference observed between the global methods and methods H and NLL we may note that global methods rely more on luminance information than on anatomical information, whereas the H and NLL methods explicitly use, partly or fully, anatomical constraints.

6. Conclusion

This chapter has described and evaluated one example of each registration type: global, local and hybrid.

The global method relies on a dense robust 3D estimation of the optical flow with a piecewise parametric description of the deformation field. It uses an efficient minimization framework, both multiresolution and multigrid with robust estimators. This optimization scheme is not limited to the estimation of the optical flow, but may as well be adapted to other similarity measures, leading to different registration applications. The adaptative partition of the volume accelerates the algorithm and improves the estimation in the regions of interest, tissues for instance.

The local method jointly relies on a modeling of anatomical landmarks and on a consistent way of extending their deformations in the vicinity of their initial support. The statistical modeling consists in learning the variability inherent to a training set and it is applied to relevant anatomical landmarks: the cortical sulci. It provides a compact and precise description of sulci shapes and of their variations. Such a model is independent from a reference subject, since it is relative to a local coordinate system linked to a mean shape. Under the assumption that inter-subject functional variability is partly encoded in anatomical variability, we have presented an original and consistent framework to extend the matching of cortical sulci in the vicinity of the sulci, thus enabling functional activations in MEG dipoles form, to be merged in this single local coordinate system. This registration framework is general and not restricted to MEG data.

The hybrid method takes advantage of both a photometric registration method and a landmark-based method. It naturally adapts to the energy-based formalism of the global method by adjoining the local sparse constraints. Similarly to the local method, we have chosen to incorporate sulcal constraints, since sulci are relevant brain cortical landmarks from an anatomical and functional point of view.

These three methods have been evaluated in a comparative framework in terms of their ability to register anatomical data, on the basis of global and local measures, and on their ability to register functional data. The efficiency of the global method has been proven in regards to global anatomical features, such as grey and white matter. As expected, the local method has been revealed more precise to register sparse functional activations located in the vicinity of the cortical landmarks, and it has been shown to significantly reduce the observed functional inter-subject variability. As the hybrid framework is concerned, we have demonstrated the benefit of introducing sulcal constraints for the registration of cortical areas. Moreover, compared to the global method, this framework significantly improves the registration of functional data.

Registration is a key issue in many neurosciences and neuroclinical applications. It contributes to jointly apprehend the various image data, e.g. data from multimodalities and/or from multi-subjects, and subsequently to improve their exploitation for clinical needs. It is particularly involved in tasks such as: group analysis, cross-population studies, anatomical and/or functional atlas building for normal and diseased populations. Global, local and hybrid approaches can address these issues at different levels. For example, global approaches will be more dedicated for large group analysis not requiring a particular focus on specific brain areas whereas

local approaches will be more precise for a particular brain structure of interest study or for a functional search in a specific brain area.

Future research directions could focus on a further development of statistical models built by variability learning. Indeed, such models can be used as an *a priori* knowledge on the sought transformation in each of the three presented cases and should therefore benefit to applications such as group analysis.

References

1. J. Ashburner, J. Andersson and K. Friston, Image registration using a symetric prior in three dimensions, *Human Brain Mapping* **9** (1998) 212–225.
2. J. Ashburner and K. Friston, Nonlinear spatial normalization using basis functions, *Human Brain Mapping* **7**, 4 (1999) 254–266.
3. R. Bajcsy and C. Broit, Matching of deformed images, in *Proc. Int. Conf. Pattern Recognition* **1** (IEEE, New York, 1982), pp. 351–353.
4. R. Bajcsy and S. Kovacic, Multiresolution elastic matching, *Computer Vision, Graphics, and Image Processing* **46** (1989) 1–21.
5. J. Barron, D. Fleet, S. Beauchemin and T. Burkitt, Performance of optical flow techniques, in *Proc. Computer Vision and Pattern Recognition*, 1992, pp. 236–242.
6. R. Battiti, E. Amaldi and C. Koch, Computing optical flow over multiple scales: An adaptive coarse-to-fine strategy, *Int. J. Computer Vision* **6**, 2 (1991) 133–146.
7. S. Beauchemin and J. Barron, The computation of optical flow, *ACM Computing Surveys* **27**, 3 (1995) 433–467.
8. J. Bergen, P. Anadan, K. Hanna and R. Hingorani, Hierarchical model-based motion estimation, in *Proc. Europ. Conf. Computer Vision*, 1991, pp. 5–10.
9. M. Black and P. Anandan, The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields, *Computer Vision and Image Understanding* **63**, 1 (1996) 75–104.
10. M. Black and A. Rangarajan, On the unification of line processes, outlier rejection, and robust statistics with application in early vision, *Int. J. Computer Vision* **19**, 1 (1996) 57–91.
11. F. Bookstein, Principal warps: Thin-plate splines and the decomposition of deformations, *IEEE Trans. Pattern Analysis and Machine Intelligence* **11**, 6 (1989) 567–585.
12. F. Bookstein and D. Green, A feature space for derivatives of deformations, in *Proc. Information Processing in Medical Imaging*, Lecture Notes in Computer Science **687** (1993) 1–16.
13. M. Bro-Nielsen and C. Gramkow, Fast fluid registration of medical images, eds. K. Hohne and R. Kikinis, *Proc. Visualization in Biomedical Computing*, Lecture Notes in Computer Science **1131** (Springer, 1996) 267–276.
14. K. Brodmann, *Vergleichende Lokalisationlehre de Grosshirnrinde* (Barth, Leipzig, 1909).
15. C. Broit, Optimal registration of deformed images, PhD thesis, Department of Computer and Information Science, University of Pennsylvania, 1981.
16. L. Brown, A survey of image registration techniques, *ACM: Computer Surveys* **24**, 4 (1992) 325–376.
17. J. Burt, Multiresolution image processing and analysis, chapter 2: The pyramide as a structure for efficient computation, Springer Series in Information Science **12** (Springer, 1984) 6–38.

18. P. Cachier, E. Bardinet, D. Dormont, X. Pennec and N. Ayache, Iconic feature based nonrigid registration: The PASHA algorithm, *Computer Vision and Image Understanding* **89**, 2–3 (2003) 272–299.
19. P. Cachier, J.-F. Mangin, X. Pennec, D. Rivière, D. Papadopoulos-Orfanos, J. Rigis and N. Ayache, Multisubject non-rigid registration of brain MRI using intensity and geometric features, eds. W. Niessen and M. Viergever, *Proc. Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science **2208** (Springer, 2001) 734–742.
20. A. Caunce and C. Taylor, Building 3D sulcal models using local geometry, *Medical Image Analysis* **5** (2001) 69–80.
21. P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud, Deterministic edge preserving regularization in computed imaging, *IEEE Trans. Image Processing* **6**, 2 (1997) 298–311.
22. G. Christensen and H. Johnson, Consistent image registration, *IEEE Trans. Medical Imaging* **20**, 7 (2001) 568–582.
23. G. Christensen, R. Rabbit and M. Miller, Deformable templates using large deformation kinematics, *IEEE Trans. Image Processing* **5**, 10 (1996) 1435–1447.
24. H. Chui, J. Rambo, J. Duncan, R. Schultz and A. Rangarajan, Registration of cortical anatomical structures via robust 3D point matching, eds. A. Kuba *et al.*, *Proc. Information Processing in Medical Imaging*, Lecture Notes in Computer Science **1613** (Springer, 1999) 168–181.
25. I. Cohen and I. Herlin, Optical flow and phase portrait methods for environmental satellite image sequences, in *Proc. Europ. Conf. Computer Vision*, 1996, pp. II:141–150.
26. Y. Cointepas, *Modélisation homotopique et segmentation tridimensionnelles du cortex cérébral à partir d'IRM pour la résolution des problèmes directs et inverses en EEG et MEG*. PhD thesis, École Nationale Supérieure des Télécommunications, Paris, 1999.
27. A. Collignon, D. Vandermeulen, P. Suetens and G. Marchal, 3D multi-modality medical image registration using feature space clustering, in *Proc. Computer Vision, Virtual Reality and Robotics in Medicine*, 1995, pp. 195–204.
28. D. Collins, G. Le Goualher and A. Evans, Non-linear cerebral registration with sulcal constraints, eds. W. Wells *et al.*, *Proc. Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science **1496** (Springer, 1998) 974–984.
29. L. Collins and A. Evans, Animal: Validation and applications of nonlinear registration-based segmentation, *Int. J. Pattern Rec. Artif. Intell.* **8**, 11 (1997) 1271–1294.
30. L. Collins, G. Le Goualher, R. Venugopal, A. Caramanos, A. Evans and C. Barillot, Cortical constraints for non-linear cortical registration, in *Proc. Visualization in Biomedical Computing*, *Lecture Notes in Computer Science* **1131** (1996) 307–316.
31. T. Cootes, C. Taylor, D. Cooper and J. Graham, Active shape models — Their training and application, *Computer Vision and Image Understanding* **61**, 1 (1995) 38–59.
32. I. Corouge, P. Hellier, B. Gibaud and C. Barillot, Inter-individual functional mapping: A non linear local approach, *NeuroImage* **19**, 4 (2003) 1337–1348.
33. T. Corpetti, E. Mémin and P. Pérez, Dense estimation of fluid flows, *IEEE Trans. Pattern Analysis and Machine Intelligence* **24**, 3 (2002) 365–380.
34. C. Davatzikos, Spatial transformation and registration of brain images using elastically deformable models, *Computer Vision and Image Understanding* **66**, 2 (1997) 207–222.

35. C. Davatzikos and R. Bryan, Using a deformable surface model to obtain a shape representation of the cortex, in *Proc. IEEE Computer Vision Symp.*, 1995, pp. 212–217.
36. C. Davatzikos, L. Prince and N. Bryan, Image registration based on boundary mapping, *IEEE Trans. Medical Imaging* **15**, 1 (1996) 111–115.
37. B. Dawant, S. Hartmann, J.-P. Thirion, F. Maes, D. Vandermeulen and P. Demaerel, Automatic 3D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations: Part i, methodology and validation on normal subjects, *IEEE Trans. Medical Imaging* **18**, 10 (1999) 909–916.
38. A. Delaney and Y. Bresler, Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography, *IEEE Trans. Image Processing* **7**, 2 (1998) 204–221.
39. R. Deriche, Recursively implementing the Gaussian and its derivatives, Technical Report 1893, INRIA, http://www.inria.fr/RRRT/RR-1893.html, 1993.
40. M. Desvignes, Mesures et modélisation de la variabilité en analyse d'images et reconnaissance de formes. Habilitation à diriger des recherches, Université de Caen, GREYC, 2002.
41. V. Devlaminck, A functional for compressive and incompressive elastic deformation estimation, *IEEE Signal Processing Letters* **6**, 7 (1999) 162–164.
42. J. Duchon, Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces, *RAIRO Analyse Numérique* **10** (1976) 5–12.
43. W. Enkelmann, Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences, *Computer Vision, Graphics, Image Processing* **43**, 2 (1988) 150–177.
44. A. Evans, L. Collins and B. Milner, An MRI-based stereotaxic atlas from 250 young normal subjects, in *Soc. Neuroscience Abstract* **18** (1992) 408.
45. J. Feldmar, J. Declerck, G. Malandain and N. Ayache, Extension of the ICP algorithm to nonrigid intensity based registration of 3D images, *Computer Vision and Image Understanding* **66**, 2 (1997) 193–206.
46. M. Ferrant, S. Warfield, C. Guttmann, R. Mulkern, F. Jolesz and R. Kikinis, 3D image matching using a finite element based elastic deformation model, eds. C. Taylor and A. Colchester, *Proc. of Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science **1679** (Springer, 1999) 202–209.
47. L. Florack, B. Romeny, J. Koenderink and M. Viergever, Scale and the differential structures of images, *Image and Vision Computing* **10** (1992) 376–388.
48. K. Friston, J. Ashburner, C. Frith, J. Poline, J. Heather and R. Frackowiak, Spatial registration and normalisation of images, *Human Brain Mapping* **2** (1995) 165–189.
49. J. Gee, L. Le Briquer, C. Barillot and D. Haynor, Probabilistic matching of brain images, eds. Y. Bizais *et al.*, *Proc. Information Processing in Medical Imaging* (Kluwer Academic Publisher, 1995).
50. J. C. Gee, M. Reivicj and R. Bajcsy, Elastically deforming 3D atlas to match anatomical brain images, *Journal of Computer Assisted Tomography* **17**, 2 (1993) 225–236.
51. P. Gerlot and Y. Bizais, Image registration: A review and a strategy for medical application, eds. C. De Graaf and M. Viergever, *Proc. Information Processing in Medical Imaging* (Plenum Press, 1988) 81–89.
52. A. Goshtasby, L. Staib, C. Studholme and D. Terzopoulos, eds., *Computer Vision and Image Understanding: Special Issue on Nonrigid Image Registration* **89**, 2/3 (Academic Press, 2003).

53. D. Graf von Keyserlingk, K. Niemann and J. Wasel, A quantitative approach to spatial variation of human cerebral sulci, *Acta Anatomica* **131** (1988) 127–131.
54. S. Gupta and J. Prince, On variable brightness optical flow for tagged MRI, eds. Y. Bizais *et al.*, *Proc. Information Processing in Medical Imaging* (Kluwer Academic Publisher, 1995) 323–334.
55. W. Hackbusch, *Multigrid Methods and Applications* (Springer, 1985).
56. R. He and P. A. Narayana, Global optimization of mutual information: Application to three dimensional retrospective registration of magnetic resonance images, *Computerized Medical Imaging and Graphics* **26**, 4 (2002) 277–292.
57. P. Hellier, J. Ashburner, I. Corouge, C. Barillot and K. Friston, Inter subject registration of functional and anatomical data using SPM, eds. T. Dohi and R. Kikinis, *Proc. Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science **2480** (Springer-Verlag, 2002) 590–597.
58. P. Hellier and C. Barillot, Cooperation between local and global approaches to register brain images, in *Proc. Information Processing in Medical Imaging*, Lecture Notes in Computer Science **2082** (2001) 315–328.
59. P. Hellier and C. Barillot, Coupling dense and landmark-based approaches for non rigid registration, *IEEE Trans. Medical Imaging* **22**, 2 (2003) 217–227.
60. P. Hellier, C. Barillot, I. Corouge, B. Gibaud, G. Le Goualher, D. Collins, A. Evans, G. Malandain, N. Ayache, G. Christensen and H. Johnson, Retrospective evaluation of inter-subject brain registration, *IEEE Trans. on Medical Imaging* **22**, 9 (2003) 1120–1130.
61. P. Hellier, C. Barillot, E. Mémin and P. Pérez, Hierarchical estimation of a dense deformation field for 3D robust registration, *IEEE Trans. Medical Imaging* **20**, 5 (2001) 388–402.
62. B. Horn and B. Schunck, Determining optical flow, *Artificial Intelligence* **17** (1981) 185–203.
63. P. Huber, *Robust Statistics* (Wiley, 1981).
64. H. Johnson and G. Christensen, Consistent landmark and intensity-based image registration, *IEEE Trans. Medical Imaging* **5**, 21 (2002) 450–461.
65. R. Jones and D. Witte, Signal intensity artefacts in clinical MR imaging, *Radiographics* **20** (2000) 893–901.
66. R. Kikinis, M. Shenton, D. Iosifescu, R. McCarley, P. Saiviroonporn, H. Hokama, A. Robatino, D. Metcalf, C. Wible, C. Portas, R. Donnino and F. Jolesz, A digital brain atlas for surgical planning, model driven segmentation and teaching, *IEEE Trans. on Visualization and Computer Graphics* **2**, 3 (1996) 232–241.
67. J. Konrad and E. Dubois, Bayesian estimation of motion vector fields, *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**, 9 (1992) 910–927.
68. F. Lachmann and C. Barillot, Brain tissue classification from MRI data by means of texture analysis, in *Proc. of SPIE Medical Imaging: Image Processing* **1652** (1992) 72–83.
69. G. Le Goualher, C. Barillot and Y. Bizais, Modeling cortical sulci with active ribbons, *Int. J. Pattern Rec. Artif. Intell.* **8**, 11 (1997) 1295–1315.
70. G. Le Goualher, E. Procyk, L. Collins, R. Venegopal, C. Barillot and A. Evans, Automated extraction and variability analysis of sulcal neuroanatomy, *IEEE Trans. Medical Imaging* **18**, 3 (1999) 206–217.
71. H. Lester and S. Arridge, A survey of hierarchical non-linear medical image registration, *Pattern Recognition* **32**, 1 (1999) 129–149.

72. G. Lohmann and D. von Cramon, Automatic detection and labelling of the human cortical folds in magnetic resonance data sets, eds. H. Burkhardt and B. Neumann, *Proc. Europ. Conf. Computer Vision*, Lecture Notes in Computer Science **1407** (Springer, 1998) 369–381.
73. G. Lohmann and D. von Cramon, Automatic labelling of the human cortical surface using sulcal basins, *Medical Image Analysis* **4** (2000) 179–188.
74. J. Maintz and M. Viergever, A survey of medical image registration, *Medical Image Analysis* **2**, 1 (1998) 1–36.
75. G. Malandain, G. Bertrand and N. Ayache, Topological segmentation of discrete surfaces, *Int. J. Computer Vision* **10**, 2 (1993) 183–197.
76. J. Mangin, J. Régis, I. Bloch, V. Frouin, Y. Samson and J. López-Krahe, A MRF based random graph modelling the human cortical topography, ed. N. Ayache, *Proc. Computer Vision, Virtual Reality and Robotics in Medicine*, Lecture Notes in Computer Science **905** (Springer, 1995a) 177–183.
77. J.-F. Mangin, Mise en correspondance d'images médicales 3D multi-modalités multi-individus pour la corrélation anatomo-fonctionnelle cérébrale, PhD thesis, École Nationale Supérieure des Télécommunications, 1995.
78. J.-F. Mangin, F. Tupin, V. Frouin, I. Bloch, R. Rougetet, J. Régis and J. López-Krahe, Deformable topological models for segmentation of 3D medical images, eds. Y. Bizais, C. Barillot and R. Di Paola, *Proc. Information Processing in Medical Imaging* (Kluwer Academic Publishers, 1995b) 153–164.
79. C. Maurer and J. Fitzpatrick, A review of medical image registration, in *Interactive Image Guided Neurosurgery*, American Association of Neurological Surgeons, 1993, pp. 17–44.
80. J. Mazziotta, A. Toga, A. Evans, P. Fox and J. Lancaster, A probabilistic atlas of the human brain: Theory and rationale for its development, *NeuroImage* **2** (1995) 89–101.
81. S. McCormick, Multilevel adaptive methods for partial differential equations, in *SIAM* (Philadelphia, Pennsylvania, 1989).
82. T. McInerney and D. Terzopoulos, Deformable models in medical image analysis: A survey, *Medical Image Analysis* **1**, 2 (1996) 91–108.
83. E. Mémin and P. Pérez, Dense estimation and object-based segmentation of the optical flow with robust techniques, *IEEE Trans. Image Processing* **7**, 5 (1998) 703–719.
84. E. Mémin and T. Risset, VLSI design methodology for edge-preserving image reconstruction, *Real-Time Imaging*. Electronic version: http://www.irisa.fr/EXTERNE/bibli/pi/1220/1220.html, 2000.
85. M. Miller, G. Christensen, Y. Amit and U. Grenander, Mathematical textbook of deformable neuroanatomies, *Proceedings of the National Academy of Sciences* **90**, 24 (1993) 11944–11948.
86. O. Musse, F. Heitz and J. Armspach, 3D deformable image matching using multiscale minimization of global energy functions, in *Proc. Computer Vision and Pattern Recognition* **2** (1999) 478–485.
87. M. Nagao and M. Matsuyama, Edge preserving smoothing, *Computer Vision, Graphics, Image Processing* **9** (1979) 394–407.
88. J. Odobez and P. Bouthemy, Robust multiresolution estimation of parametric motion models, *Journal of Visual Communication and Image Representation* **6**, 4 (1995) 348–365.
89. M. Ono, S. Kubik and C. Abernathey, *Atlas of the Cerebral Sulci* (Georg Thieme Verlag, 1990).

90. W. Peckar, C. Schnörr, K. Rohr and S. Stiehl, Parameter-free elastic registration of brain images using prescribed displacements. Technical report, Universitat hamburg hachbereich informatik, 1997.
91. W. Peckar, C. Schnörr, K. Rohr and S. Stiehl, Parameter-free elastic deformation approach for 2D and 3D registration using prescribed displacement, *Journal of Mathematical Imaging and Vision* **10**, 2 (1999) 143–162.
92. W. Penfield and T. Rassmussen, *The Cerebral Cortex of Man* (Macmillan, 1950).
93. X. Pennec, P. Cachier and N. Ayache, Understanding the demon's algorithm: 3D non rigid registration by gradient descent, eds. C. Taylor and A. Colchester, *Proc. of Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science **1679** (Springer, 1999) 597–605.
94. J. Pluim, J. Maintz and M. Viergever, Mutual information matching in multiresolution context, *Image and Vision Computing* **19**, 1–2 (2001) 45–52.
95. M. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, *The Computer Journal*, 1964, pp. 155–162.
96. P. Reissman and I. Magnin, The active pyramid: A new tool for modelling 3D deformable object, *Int. J. Pattern Rec. Artif. Intell.* **11**, 7 (1997) 1129–1139.
97. C. Renault, *Courbures et lignes de crête sur des images en niveau de gris, Étude comparative et application aux sillons corticaux*, PhD thesis, Université de Caen, 2001.
98. M. Rettmann, X. Han, C. Xu and J. Prince, Automated sulcal segmentation using watersheds on the cortical surface, *NeuroImage* **15** (2002) 329–344.
99. D. Rivière, Apprentissage de la variabilité inter-individuelle de l'anatomie corticale cérébrale pour la reconnaissance automatique des sillons. PhD thesis, Université d'Evry-Val d'Essonne, 2000.
100. K. Rohr, H. Stiehl, R. Sprengel, W. Beil, T. Buzug, J. Weese and M. Kuhn, Point-based elastic registration of medical image data using approximating thin-plate splines, eds. H. Hohne and R. Kikinis, *Proc. Visualization in Biomedical Computing*, Lecture Notes in Computer Science **1131** (Springer, 1996) 297–306.
101. N. Royackkers, M. Desvignes, H. Fawal and M. Revenu, Detection and statistical analysis of human cortical sulci, *NeuroImage* **10** (1999) 625–641.
102. S. Sandor and R. Leahy, Surface-based labeling of cortical anatomy using a deformable atlas, *IEEE Trans. Medical Imaging* **16**, 1 (1997) 41–54.
103. C. Schnörr, A study of convex variational diffusion approach for image segmentation and feature extraction, *J. Mathematical Imaging and Vision* **8**, 3 (1998) 271–292.
104. T. Schormann, S. Henn and K. Zilles, A new approach to fast elastic alignement with application to human brains, eds. K. Hohne and R. Kikinis, *Proc. Visualization in Biomedical Computing*, Lecture Notes in Computer Science **1131** (Springer, 1996) 337–342.
105. D. Schwartz, D. Badier, J. Bihoué and A. Bouliou, Evaluation with realistic sources of a new MEG-EEG spatio-temporal localization approach, *Brain Topography* **11**, 4 (1999) 279–289.
106. D. Schwartz, E. Poiseau, D. Lemoine and C. Barillot, Registration of MEG/EEG data with 3D MRI: Methodology and precision issues, *Brain Topography* **9**, 2 (1996).
107. S. Song and M. Leahy, Computation of 3D velocity fields from 3D cine CT images of a human heart, *IEEE Trans. Medical Imaging* **10**, 3 (1991) 295–306.
108. R. Szeliski and J. Coughlan, Hierarchical spline-based image registration, in *Proc. Computer Vision and Pattern Recognition*, 1994, pp. 194–201.
109. J. Talairach and P. Tournoux, *Co-planar Stereotaxic Atlas of the Human Brain* (Georg Thieme Verlag, Stuttgart, 1988).

110. X. Tao, J. Prince and C. Davatzikos, Using a statistical shape model to extract sulcal curves on the outer cortex of the human brain, *IEEE Trans. Medical Imaging* **21**, 5 (2002) 513–524.
111. D. Terzopoulos, Image analysis using multigrid relaxation methods, *IEEE Trans. Pattern Analysis and Machine Intelligence* **8**, 2 (1986) 129–139.
112. J.-P. Thirion, Image matching as a diffusion process: An analogy with Maxwell's demons, *Medical Image Analysis* **2**, 3 (1998) 243–260.
113. P. Thompson and A. Toga, A surface-based technique for warping three-dimensional images of the brain, *IEEE Trans. Medical Imaging* **15**, 4 (1996) 402–417.
114. P. Thompson, R. Woods, M. Mega and A. Toga, Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain, *Human Brain Mapping* **9** (2000) 81–92.
115. A. Toga, A. Coldkorn, K. Ambach, K. Chao, B. Quinn and P. Yao, Postmortem cryosectioning as an anatomic reference for human brain mapping, *Computerized Medical Imaging and Graphics* **21**, 2 (1997) 131–141.
116. A. Toga and P. Thompson, The role of image registration in brain mapping, *Image and Vision Computing* **19** (2001) 3–24.
117. M. Vaillant and C. Davatzikos, Finding parametric representations of the cortical sulci using an active contour model, *Medical Image Analysis* **1**, 4 (1996) 295–315.
118. M. Vaillant and C. Davatzikos, Hierarchical matching of cortical features for deformable brain image registration, eds. A. Kuba *et al.*, *Proc. Information Processing in Medical Imaging*, Lecture Notes in Computer Science **1613** (Springer, 1999) 182–195.
119. J. Van Bemmel and M. Musen, *Handbook of Medical Informatics* (Springer, 1997) URL : http://www.mieur.nl/mihandbook.
120. P. Van den Elsen, E. Pol and M. Viergever, Medical image matching: A review with classification, *IEEE Engineering in Medecine and Biology* **12** (1993) 26–38.
121. B. Vemuri, S. Huang, S. Sahni, C. Leonard, C. Mohr, R. Gilmore and J. Fitzsimmons, An efficient motion estimator with application to medical image registration, *Medical Image Analysis* **2**, 1 (1998) 79–98.
122. P. Viola and W. Wells, Alignment by maximization of mutual information, in *Proc. IEEE International Conference on Computer Vision*, 1995, pp. 15–23.
123. Y. Wang and L. Staib, Physical model-based non-rigid registration incorporating statistical shape information, *Medical Image Analysis* **4** (2000) 7–20.
124. W. Welker, Why does the cerebellar cortex fissure and fold? A review of determinants of gyri and sulci, in *Comparative Structure and Evolution of Cerebral Cortex* **8**, chapter 10 (1990) p. 512.
125. W. Welker and S. Seidenstein, Somatic sensory representation in the cerebral cortex of the racoon (procyon lotor), *J. Comp. Neurol.* 1959, pp. 459–501.
126. J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. Maurer, R. Kessler, R. Maciunas, C. Barillot, D. Lemoine, A. Collignon, F. Maes, P. Suetend, D. Vandermeulen, P. Van Den Elsen, S. Napel, T. Sumanaweera, B. Harkness, P. Hemler, D. Hill, D. Hawkes, C. Studholme, J. Maintz, M. Viergever, G. Malandain, X. Pennec, M. Noz, G. Maguire, M. Pollack, C. Pellizzari, R. Robb, D. Hanson and R. Woods, Comparison and evaluation of retrospective intermodality brain image registration techniques, *Journal of Computer Assisted Tomography* **21**, 4 (1997) 554–566.
127. B. Zitova and J. Flusser, Image registration methods: A survey, *Image and Vision Computing* **21** (2003) 977–1000.

128. T. Hartkens, D. L. G. Hill, A. D. Castellano-Smith, D. J. Hawkes, C. R. Maurer, Jr., A. J. Martin, W. A. Hall, H. Liu and C. L. Truwit, Using points and surfaces to improve voxel-based non-rigid registration, *Proc. of Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science **2489** (Springer-Verlag, 2003) 565–572.

This page intentionally left blank

CHAPTER 6

AUTOMATED IMAGE SEGMENTATION: ISSUES AND APPLICATIONS

ALAIN PITIOT[*,§], HERVÉ DELINGETTE[†,¶] and PAUL M. THOMPSON[‡,‖]
[*] *Mirada Solutions, Level 1, 23–38 Hythe Bridge Street*
Oxford, OX1 2EP, UK

[†] *INRIA Sophia Antipolis, Epidaure Research Project*
2004 route des Lucioles–BP 93, 06902 Sophia Antipolis Cedex, France

[‡] *Laboratory of Neuro Imaging, Department of Neurology*
David Geffen School of Medicine, UCLA, 710 Westwood Plaza
Los Angeles, CA 90095-1769, USA
[§] *apitiot@loni.ucla.edu*
[¶] *herve.delingette@sophia.inria.fr*
[‖] *thompson@loni.ucla.edu*

The explosive growth in medical imaging technologies has been matched by a tremendous increase in the number of investigations centered on the structural and functional organization of the human body. A pivotal first step towards elucidating the correlation between structure and function, accurate and robust segmentation is a major objective of computerized medicine. It is also a substantial challenge in view of the wide variety of shapes and appearances that organs, anatomical structures and tissues can exhibit in medical images.

This chapter surveys the actively expanding field of medical image segmentation. We discuss the main issues that pertain to the remarkably diverse range of proposed techniques. Among others, the characteristics of a suitable segmentation paradigm, the introduction of *a priori* knowledge, robustness and validation are detailed and illustrated with relevant techniques and applications.

Keywords: Medical imaging; segmentation; review; segmentation paradigm; *a priori* knowledge; robustness; validation.

1. Introduction

Imaging technologies have undergone fast paced developments since the early days of anatomy. Magnetic resonance imaging (MRI), computer-assisted tomography (CT), positron emission tomography (PET) and an increasing number of other techniques (see Fig. 1) now permit precise analysis of post-mortem tissue and non-invasive exploration of living organisms. They can elucidate the structures of organs and cells, observe and help understand their function, and give clinicians the means to monitor their dysfunctions, or assist in the removal of pathologies.

A deeper understanding of both the anatomical characteristics of the tissues and organs of the human body (or, more precisely, of the sub-structures we distinguish within them) and of their inter-relationships is crucial in diagnostic and

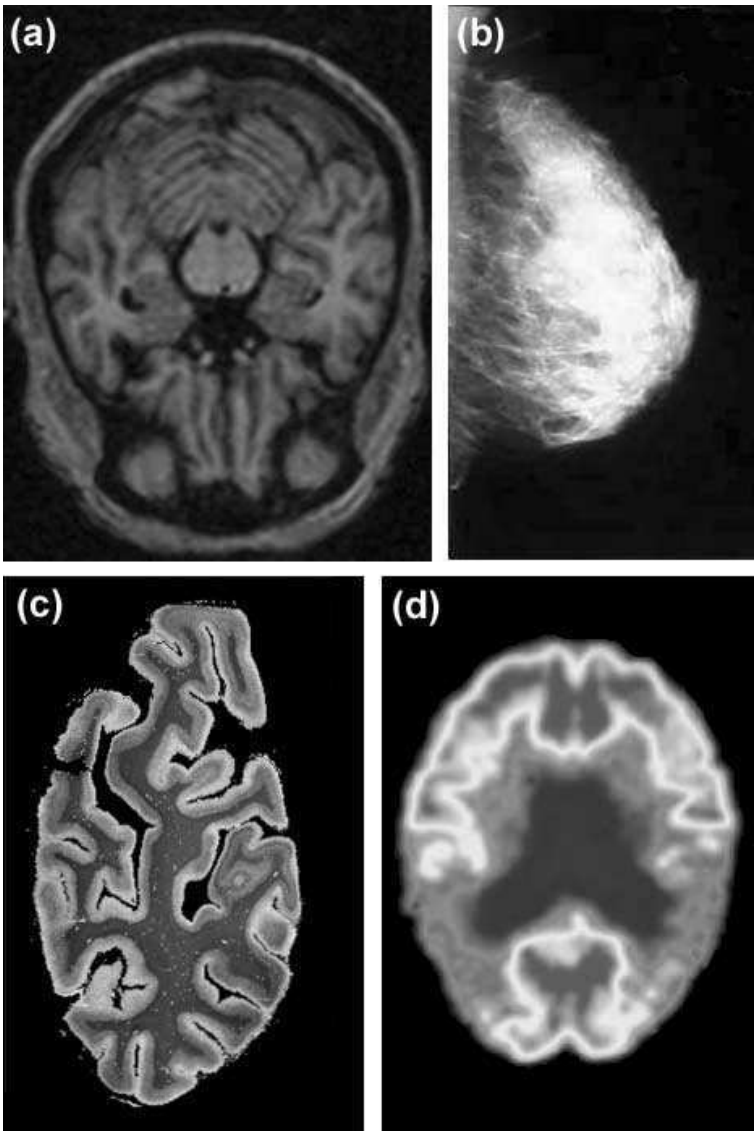


Fig. 1. A collection of imaging modalities: (a) MR image of the human brain, (b) digital mammogram, (c) myelin-stained histological section of the human visual cortex, (d) PET image of the human brain, (e) CT scan of the chest.

interventional medicine. The need, shared across many levels of description, for such correlation between structure and function is exemplified by the vast number of studies analyzing cortical structures (in populations with a particular disease,[127] through the developmental cycle[16] or comparing normal and diseased subjects[189]), quantifying tissue loss, gain or structure volumes,[55,94] or aiming for automated diagnosis of disease,[176,185] among others.

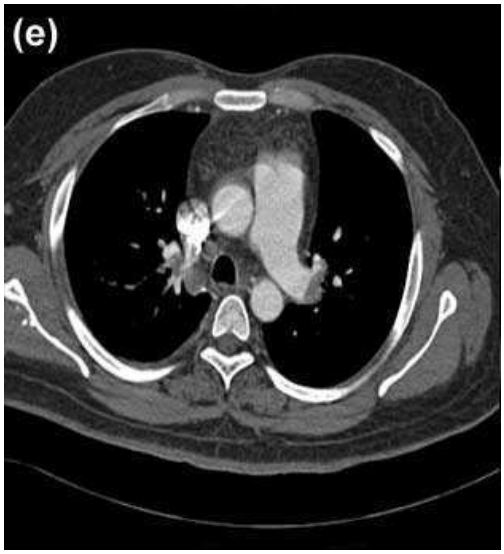


Fig. 1. (*Continued*)

While qualitative analysis may sometimes be sufficient for diagnosis, quantitative analysis for which segmentation and labeling are absolute prerequisites, is necessary for a variety of applications: longitudinal monitoring of disease progression or remission,[72,157] pre-operative evaluation and surgical planning,[6,78,90] radiotherapy treatment planning[131] or statistical analysis of anatomic variability.[38,193]

Even so, accurate segmentation of anatomical structures and tissues in medical images is especially challenging, given the wide variety of shapes, sizes and appearances they can present. Still the delineation process calls for high precision as the quality of the analysis generally depends on how accurately the various structures are identified. For instance, given the corpus callosum's key role as the primary cortical projection system, regional analysis of its structure is important in assessing several neurological disorders (Alzheimer's disease, vascular dementia, dysplasias). Nonetheless, subtle variations in shape, relative to a mean callosal delineation, are observed between and within patient and control groups, and this makes it difficult to detect and classify abnormal structural patterns. As a result, intense debate still rages on whether different callosal regions undergo selective changes in each of these disease processes and whether these are systematic differences in neuropsychiatric disorders such as autism or schizophrenia. These controversies may be alleviated by precise and reliable segmentations, applied to large image databases.

Segmentation has traditionally been tackled by human operators. However the many drawbacks of manual delineation (lack of reproducibility, *a priori* biases, lack of sufficient resources to handle ever-growing databases) favor the use of automated methods. Nonetheless, to reach the desired accuracy, many difficulties must be overcomed: input images may be noisy, poorly contrasted and full of "decoys" (many

structures are similar in shape or intensity), the target structures may be highly variable in geometry, etc.

We propose in this chapter an overview of the ever-expanding palette of automated segmentation techniques applied to biomedical images. Remarkably, the diversity of the developed techniques is more than matched by the variety of the objectives (disease diagnosis, surgical planning, atlas building, etc), of the segmentation targets and of the input imaging modalities, with substantially different hypotheses and requirements for each of them. Rather than arbitrarily privileging a particular outlook over another, we discuss the main issues associated with these application-specific parameters, and the constraints they entail. More algorithm- or application-oriented taxonomic reviews are available elsewhere.[14,32,117,138,180,224] A brief account of relevant techniques accompanies the discussion of the issues. A detailed summary of each method or application would be beyond the scope of this chapter. Instead, we provide a generic description of the main algorithmic classes and more specifically discuss their interactions with the issues we have identified.

We begin with some reflections on the definition of segmentation. Section 3 then characterizes the input images from which organs and structures must be segmented, most especially in terms of dimensionality. We also introduce the commonly used radiological modalities referred to in this chapter, considering the difficulties they create for segmentation techniques. The selection of an appropriate segmentation paradigm, which depends on the envisaged application and the imaging modality, is examined in Sec. 4. We analyze how the flexibility, locality and continuity of the model impact the segmentation performance. Section 5 discusses the introduction of *a priori* knowledge and medical expertise to guide the segmentation process towards more probable shapes. We then comment on the robustness of segmentation techniques in Sec. 6 where the difficult matters of initialization and the trade-off between genericity and application-specificity are emphasized. Validation is discussed in Sec. 7. We analyze its inherent contradictions (lack of true gold-standard due to inter/intra operator variability, conflicting error measures, application specificity) and how they bear on issues raised so far. Finally, Sec. 8 comments on the future of medical segmentation and the underexplored territories of semi-automated and manually-assisted segmentation.

2. Segmentation Criteria

Biomedical images, being digital pictures, are organized collections of values linked via a chain of treatments and sensors to some underlying physical measures (radiation absorption, acoustic pressure, radiofrequency responses, etc). These measures are related to the physical characteristics of the imaged tissues (density, chemical composition, cellular architecture). In that respect, these images are a means to analyze otherwise undecipherable raw measurements. The necessity to understand the measured values then becomes that of extracting meaningful information from the associated images, that is, to establish a relationship between the acquired data

and the associated physiological phenomena. Segmentation is the first step on the path towards understanding these complex inter-relationships.

Irrespective of the envisioned application, segmentation essentially consists of partitioning the input image into a number of disjoint regions, so as to satisfy in each region a given criterion. This criterion may be derived from the intensity distribution of the underlying voxels in the image, or from the morphology of the region, among other choices.

The first difficulty in designing an adequate criterion is characterizing the objectives of the segmentation process, which may vary quite subtly from one application to the next and often depend on the object being segmented. For instance, the segmentation accuracy required to construct anatomical atlases (built by averaging the shapes of several segmented instances of the structures in the atlas) may be somewhat less than that required to quantify gray matter loss in a neurological study.[172] Indeed, missing the true boundary between gray and white matter might significantly bias the analysis in the latter case, whereas the errors introduced by the atlas shape averaging process are often as significant as those of a standard segmentation step. Therefore, even though correctly dealing at the criterion level with partial volume effect voxels (voxels to which multiple tissues contribute, which result in blurred edges at the surface of structures) may not be so crucial for atlas building, it is pivotal for accurate quantification of subtle tissue changes.

The second difficulty is linked to estimating the effects of noise and of the plethora of artifacts that plague the input images (bias fields, projection errors, variability in tracer uptake times, etc). Together with more structure-specific parameters (contrast with respect to surrounding tissues, shape variability), these are bound to influence the choice of a segmentation criterion, most especially in terms of its leniency and permissiveness (and false positives and negatives).

In turn, the characteristics of the chosen criterion will affect the efficiency of the optimization or evolution process to which most segmentation applications can be reduced (see Sec. 6.3). Evaluating this efficiency, that is, assessing the performance of the segmentation system, is however particularly difficult in the absence of a satisfactory ground truth (we comment on this validation issue in Sec. 7).

2.1. *Hard segmentation*

From a mathematical point of view, image segmentation is an injective process that maps sets of voxels (low-level numerical bits of information) to annotated regions (high-level semantic information). These annotations may be actual labels (in reference to some dictionary of labels) or simply ordinal numbers to differentiate the regions.

More formally, let I be an input image, defined by its value (usually a scalar intensity value, but sometimes a real-valued vector or even a tensor) at each point of its domain Ω. Let ω be a non-empty subset of Ω. Then let Λ be a predicate (the above mentioned criterion restricted to a single target organ or structure) which

assigns the value true or false to ω. A segmentation of image I for predicate Λ is a partition of Ω into n disjoint non-empty subsets $\{\omega_i\}_{i=1}^{n}$ such that:

- $\forall\ i,\ 1 \leq i \leq n,\ \Lambda(\omega_i) = \text{true}$; and
- $\forall\ i, j,\ 1 \leq i, j \leq n,\ i \neq j,\ \Lambda(\omega_i \cup \omega_j) = \text{false}$.

For example, in thresholding approaches (see Sec. 4.2.2), perhaps the simplest image segmentation approach, regions in I are defined by creating a partition of the image intensities.[160] If we restrict I to take values in $[0, 1]$, a choice for Λ could be:

$$\Lambda : \Omega \rightarrow \{\text{true, false}\}$$

$$\omega \mapsto \begin{cases} \text{true} & \text{if } \forall\ x \in \omega,\ \ I(x) > \theta \\ \text{false} & \text{otherwise} \end{cases}$$

where $\theta \in [0, 1]$ is a threshold, usually determined from the image intensity histogram (see Fig. 2).

Note that Λ often relies on a neighborhood of ω_i to assign it a boolean value (see Sec. 4.2): in other words, it is not a uniformity predicate.

2.2. *Soft segmentation*

Fuzzy segmentation[91] generalizes this predicate to a membership function, providing an efficient means to deal with partial volume effects.

Given a set of K tissue or target classes $C = \{c_1, \ldots, c_K\}$, K membership functions are designed: $\forall\ k \in [1, K]$, $\mu_k : \Omega \rightarrow [0, 1]$, subject to the constraint

$$\forall\ x \in \Omega,\ \sum_{k=1}^{K} \mu_k(x) = 1\,.$$

They represent the contribution of each tissue (i.e. volume fraction) at every voxel in the input image.

A number of fuzzy clustering techniques have been developed to automate the computation of these membership functions.[199] In the context of brain tissue segmentation, these classes may represent gray matter, white matter or cerebrospinal fluid for example.[76,137] Probability densities can also be substituted for fuzzy membership functions within a classical *a posteriori* maximization/expectation-maximization framework.[97,212]

Yet, apart from probabilistic and fuzzy atlases[191,193] which directly use probability densities of tissue memberships, most clinical applications require the actual boundaries of the segmentation targets to be accurately determined. Consequently, suitable cut-off and thresholds have to be selected to turn soft segmentations into hard ones, a non-trivial problem in itself.

3. Input Data

Each imaging modality comes with a specific set of characteristics (structural or functional measures, actual spatial and temporal resolution, effective field of view,

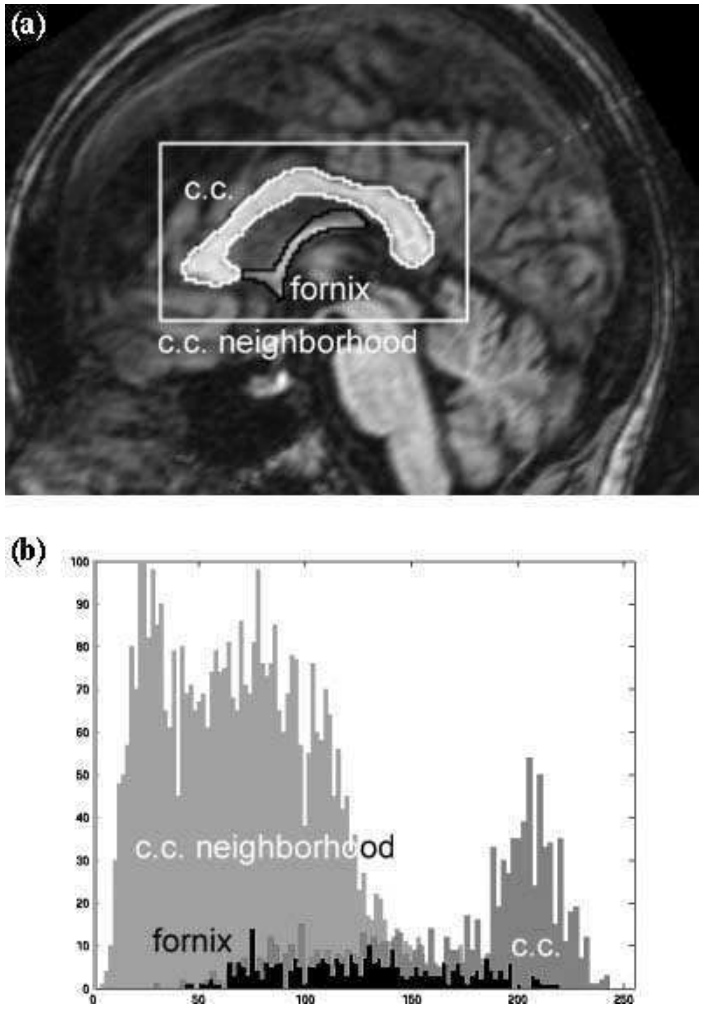


Fig. 2. The corpus callosum (c.c) and its neighbors in a $T1$-weighted MRI section (a) and their associated intensity distributions (b).

signal to noise ratio, etc) and sheds a different light on the studied pathology or organ. Ideally, segmentation algorithms should be fed images acquired from the full battery of existing modalities: MRIs to act as anatomical references, histologic sections for precise pathological tissue analysis, PET/SPECT or functional MRI data to reveal metabolic or functional relationships, etc. However, practical considerations beg for compromises to be found. Imaging resources may be unavailable (not only are radiological machines very expensive, but each acquisition is also costly), there may be potential health hazards linked to the invasiveness of data acquisition (X-ray CT or PET radiations must be used with caution, histology is a *post-mortem* analysis), etc. Often, the availability of a set of imaging modalities will

determine the medical objectives which can be reasonably achieved. Conversely, the envisaged applications (or standard diagnostic protocols) condition the acquisition of appropriate modalities. In any case, segmentation systems have to be designed, or adapted, accordingly.

3.1. *Image characteristics*

Since biomedical images are located at the interface between physical measurements and human cognitive processes (namely, image interpretation and analysis), their characteristics are intimately linked to those of both the imaging equipment that acquired them and of the mathematical model that allows their algorithmical manipulation. These two sets of features are intricately related. For instance, while X-ray films are inherently continuous in space, segmentation systems manipulate them as a discrete set of pixels, once digitized. Conversely, most deformable model approaches (simplex meshes,[51] medial representations,[143] etc see Sec. 4) operate on discrete input images, such as MRIs or CT scans, represented as continuous functions by using interpolation techniques, thereby achieving sub-voxel accuracy. Images can then be considered either as continuous functions observed at a continuous or discretized positions in space, or as a set of discrete intensity values organized on a regular or irregular lattice. They may also be treated as observations of a random vector (where each component is a random variable associated to a site in a lattice) or even as the realization of a stochastic process (for Markov random field approaches).

As much a characteristic of the input images as one of the conceived applications, dimensionality (the dimension of the space in which the segmentation algorithm operates) may also significantly affect the choice, implementation and behavior of a segmentation system. For instance, in one of the corpus callosum statistical variability studies mentioned above,[190] only the mid-sagittal sections of the input MRI's were selected for delineation, whereas Narr *et al.*[127] used 3 additional 1 mm thick slices on each side. Clearly, even if a fairly simple 2D algorithm should be sufficient to automate the callosal segmentations in the first case, the segmentation of actual surfaces would be better handled in a true 3D segmentation system in the second case: the callosal surface obtained would be smoother and more globally coherent than those resulting from concatenating successively segmented 2D sections.

Incidentally, while some modalities are inherently 3D (MR for instance, even though images are acquired slice by slice), or inherently 2D (histological slices), others are so only artificially. X-ray images for instance are 2D projections of a 3D object; conversely, CT scans are 3D volumes reconstructed from a series of 2D projections (see Fuchs *et al.*[61] for a review). Projections often make for more difficult segmentations as the target boundaries of structures may be substantially distorted and the contours of other organs may be intercepted by the projection and may pollute the image, further umpairing precise target localization. Furthermore, 3D reconstruction induces many artifacts which may reduce the signal to noise ratio.[60,69] Note that 2D segmentation techniques are sometimes applied to

3D data, one 2D slice at a time.[7,101] This could be for complexity reasons (real-time constraints, limited memory resources), and algorithmic reasons (thresholding techniques, for instance, are not affected by the dimensionality of the data space as they do not rely on neighborhood considerations: only the intensity at the considered voxel is taken into account) or application-oriented reasons (the corpus callosum may be easier to segment in a series of coronal slices on each side of the mid-sagittal plane than in a 3D MRI taken as a volume).

A hybrid dimensionality case, brain cortex parcellation deals with the segmentation of patches on 2D surfaces folded in 3D space (2D manifolds) and often requires specialized segmentation systems[152,160,195] where the expected geometry and topology of the manifolds have to be woven into the contour-finding algorithms.

3.2. *Modalities*

We consider in this chapter only the most commonly used radiological modalities. Macovski[107] and Sprawls[174] provide in-depth introductions to the underlying physical phenomena.

Arguably the most common modality, radiography encodes in each voxel of the generated image the accumulated density of the structures intercepted by a beam of ionizing radiations (X-rays). This is a fast acquisition process which yields a particularly high contrast between hard structures (such as bones) and soft tissues (organs). Unfortunately, radiography is also substantially invasive, suffers from the projection issues mentioned above (poor localization, decoy structures), and provides only limited information about soft tissues. In the related fluoroscopy, a contrast agents is injected into the patient and a moving X-ray beam enables the observation of structures *in vivo* and in real time. Among other applications, digital mammography has proved invaluable for the early detection of tumors and microcalcification clusters. In view of the poor contrast of structures in breast images, the robust detection of tumors is more important than the accuracy of their segmentation.

Another X-ray-based modality, computed tomography (CT) alleviates most of the projection issues of planar radiography. It provides excellent soft tissue constrast and allows the 3D visualization of deep internal structures.

In ultrasound imaging, high frequency sound waves replace the ionizing radiations of radiographic techniques. The sound waves emitted by a transducer moved over the patient skin by an operator, are reflected back to the transducer at the interfaces between the traversed organs or tissues. An image of the variations of accoustic impendance can subsequently be reconstructed. Ultrasound systems are completely non-invasive, operate in real time and allow multi-planar imaging. Their low cost has ensured a wide dissemination. They can observe static organs and follow their dynamic evolution. They are, however, plagued with high level of speckling. Furthermore, bones and air act as opaque screens and prevent the visualization of deep structures.

The modality of choice for segmentation systems, magnetic resonance imaging (MRI) records the radio-frequency signal emitted by the protons of water molecules

after excitation by a strong magnetic field. It provides excellent soft tissue contrast (most especially as it can be tuned by using an appropriate pulse sequence), high signal-to-noise ratio and good spatial resolution (commonly 1 mm^3) to the detriment of the temporal resolution, unfortunately (20 minutes for a standard examination). Besides, a number of intensity inhomogeneities and artifacts[169,171] complexify the segmentation task. A large number of dedicated overviews are available.[14,32,224]

In scintigraphy, radioisotopes are injected into the patient and a series of cameras correlate the emitted beams to reconstruct a 3D map of increased or decreased activity. It is an inherently functional modality, which unfortunately suffers from poor spatial resolution.

4. Segmentation Paradigm

A segmentation paradigm encompasses under a single umbrella the many considerations about the nature of the segmentation application (statistical variability analysis, CAD, tumor tracking), the associated operational constraints, the algorithmic class of the selected segmentation technique and its working hypotheses, among others. As such, it depends on the envisioned application and on the imaging modality employed. For instance, segmentation of gray and white matter in a cerebral MRI induces vastly different constraints from that of a vertebrae in an X-ray of the vertebral column, in terms of target topology, prior knowledge, choice of target representation, signal to noise ratio, and dimensionality of the input data. The selection of an adequate segmentation paradigm is therefore pivotal as it affects how efficiently the segmentation system can deal with the target organ or structure, and conditions its accuracy and robustness. We detail below the foremost compromises and parameters that should shape an educated choice.

4.1. *Bottom-up versus top-down*

Reminiscent of the bipolar character of the couple image/application, segmentation involves extracting from the input image the relevant features (texture patches, edges, etc) associated with the target structure and immersing these into a higher-order model of the target (surface mesh, *m*-rep, etc). These are then passed to the application for analysis. Not surprisingly, this dual nature is reflected in the dichotomy between feature-extraction algorithms (bottom-up) and model-based approaches (top-down).

Bottom-up strategies can usually be decomposed into three stages: first, features are extracted, then they are grouped into several regions or contours, which, finally, serve to identify the structure's boundaries. However, since these techniques usually consider only local neighborhoods without a higher order comprehension of the nature of the image,[a] they are prone to generating invalid outlines. For instance,

[a]That is, they operate on values attached to the image voxels without necessarily establishing a relationship with the reality that they represent.

in edge detection, all extracted contours do not correspond to the boundaries of the target structure: some of them may merely follow decoys or noise artifacts. These problems are rooted in the inherently numerical nature of the data manipulated by these low-level *model-free* algorithms, and aggravated by the underconstrained nature of the segmentation of highly variable structures. As such, image-level segmentation techniques (region growing, edge detection, etc) tend to operate adequately only under substantial expert guidance.

On the other hand, high-level *model-based* approaches (top-down strategies) operate on semantic concepts (shape, appearance, relative position with respect to surrounding structures, etc) associated with a representation of the actual segmentation target, extracted from the image. They are linked to the interpretation and understanding of the input data and can overcome many of these limitations.[b] As such, top-down strategies provide a convenient framework to include *a priori* knowledge and medical expertise (see Sec. 5). Briefly, they also consist of three stages: model building (based on prior knowledge or on structures segmented *a priori*), model initialization (initialization of the parameters that control the model's shape, position, etc), and model matching (adaptation of the parameters to the input image). By considering the target boundaries as a whole, they become a lot more robust to noise and imaging artifacts than bottom-up techniques. Unfortunately, they are also potentially less accurate. Indeed, the reference *a priori* models which act as shape, intensity, or distance constraints effectively prevent the segmentation of what they identify as noisy or invalid contours. When these correspond to actual noise, robustness is increased. However, when they correspond to the true contours of the target structure, accuracy decreases (Section 6 comments on this trade-off between accuracy and robustness, which also relates to locality, flexibility and continuity).

Often, a segmentation system will implement a mixture of bottom-up and top-down approaches, either explicitly[18] or implicitly (as with deformable models, see Sec. 4.3).

4.2. *Locality*

Locality refers to the extent of the neighborhoods around the voxels of the input image considered by the segmentation process. It is inherently linked to the segmentation strategy discussed above. Namely, very local techniques are usually considered bottom-up approaches (the size of the neighborhood being too small for

[b]Incidentally, whereas model-based approaches usually require a training set of segmented contours as an input, low-level feature extraction is generally performed without reference to an *a priori* contour model. Such distinction between the problem of contour modeling and that of edge extraction is characteristic of Marr's vision paradigm.[114] The interest of that dichotomy lies in its ability to decompose the segmentation problem into independent and manageable tasks. Unfortunately, it may also result in a unidirectional and somewhat irreversible cascade of errors. Furthermore, due to image noise and the image projection process, local model-free edge extraction is an ill-posed problem with no unique solution.[145]

an actual model to be used) whereas global techniques are ideally suited for introducing shape or appearance models. Local segmentation methods tend to be very fast, owing to the small number of parameters considered at each voxel.[74] In the absence of a model, they are more accurate (immune as they are from constraints linked to the geometrical arrangement of voxels or to their intensity distribution) but they are also more sensitive to noise and inhomogeneities. Conversely, larger neighborhoods increase noise robustness at the expense of accuracy.

In the absence of high-level models, local techniques are effectively a special case of classification algorithms. Segmentation then consists in deciding, for every voxel in the input image, whether or not it belongs to the target structures, based on attributes (intensity values, geometric descriptors or other statistics) collected in its immediate neighborhood.

On the other end of the locality spectrum, we find most of the model-based (top-down) approaches. These use the maximum amount of information that can be gathered from the underlying image to fit the parameters of the models they rely on and to guide the segmentation process (the considered neighborhood is then often extended to the voxels in the vicinity of the entire model surface, or even to the whole image).

After a generic description of voxel classification methods, we detail more specific local approaches in Sec. 4.2.2. The main global techniques, deformable models and atlas warping approaches are discussed later in Secs. 4.3 and 5.2.

4.2.1. *The classification paradigm*

Classification techniques associate every voxel in the input image with a unique label, or class, where classes are built from the voxel attributes[54] (usually, one class per segmentation target, plus a "non-target" class). In the simplest general case, voxels are classified independently of each other as the criteria employed, often based on distances between attribute vectors, and do not take into account their geometric arrangements. As such, classification is really only a pre-processing step. Connected components must be extracted from the classification map to effectively segment the input image.

From a pattern recognition point of view, classification techniques aim to partition the multidimensional feature space formed by the voxel attributes. This could be a supervised or unsupervised process.

Supervised classification. Supervised methods consist of two phases: a training phase in which a learning set of *a priori* segmented structures help adjust the classifier parameters, and a classification phase where a previously unknown input image is processed with the trained classifier. A large number of supervised techniques are available in the literature (see Duda and Hart[54] for a review).

In view of the difficulty of modeling the probability distribution of the target voxel attributes, non-parametric techniques, which make no hypothesis about the

class distributions, have proved popular. Nearest-neighbor classifiers for example assign a voxel to the same class as the voxel in the training set which is closest in terms of attribute distance (often, the class of the learning set voxel with the closest intensity). A generalization of this straightforward approach, a k-nearest-neighbor classifier[79] assigns classes based on a majority vote of the k closest voxels in the learning set. Parzen window techniques extend the majority vote to a rectangular or spherical neighborhood of the attribute space centered on the considered voxel.

When the distribution of the attribute values is better behaved, parametric classifiers may increase the segmentation performances.[54] Often, voxel attributes are assumed to be drawn from a mixture of Gaussian distributions, as with the maximum likelihood classifier (Bayes method). Training such a classifier requires estimating the means and standard deviations of the Gaussian distributions and their mixing coefficients from the learning set of *a priori* segmented structures. In the classification phase, voxels are then assigned to the classes which maximize the posterior probability.

Unsupervised classification. When no *a priori* learning set is available to train the classifier, unsupervised classification (clustering) become a more suitable alternative. In the absence of an initial parameter fitting phase, clustering techniques often maintain, for each class, a model of the characteristics of their attribute distribution. These models are iteratively updated during the classification process, which usually alternates between classification and model fitting. The training phase is consequently distributed over the entire course of the classification phase.

The unsupervised version of the k-nearest-neighbor algorithm, the k-means clustering algorithm[87] models each class by its mean attribute vector and partitions the voxels in the input image by assigning them to the class whose mean is closest. By introducing fuzzy membership functions into the classification step, the fuzzy c-means algorithm[99] allows for a soft segmentation of the input image. A parametric unsupervised technique, the expectation-maximization algorithm (EM) assumes a Gaussian mixture model for the voxel attributes and iterates between the computation of the posterior probabilities associated to each class and the maximum likelihood estimation of the model parameters (means, covariances, mixing coefficients).[119] Note that unsupervised techniques are often more sensitive to the initial values of their parameters than supervised classifiers.[54]

Feature selection. Often, the voxel intensity alone is considered as an attribute. However, when multiple attributes are available, an "optimal" subset of attributes, as discriminating as possible, must be selected while keeping the number of selected attributes reasonably small. This selection task (also called feature reduction) is a fundamental problem in statistical pattern recognition. Indeed, reducing the number of attributes saves computing resources by discarding irrelevant or redundant features. It also alleviates the effects of the so-called

curse of dimensionality,[83] which links the ratio: sample size (in the learning set)/dimensionality of the feature vector to the classification performances.[c]

Given an objective function, which evaluates the performance of the classifier on an *a priori* set of measurements, the feature selection problem then boils down to a search problem in the combinatorial space defined by the voxel attributes. Trying out, in a brute force manner, every possible combination of features can be prohibitively costly when the number of attributes is high (although, as argued by Cover,[43] traversing the entire search space is the necessary condition to an optimal selection). Specific sub-optimal selection strategies have therefore been suggested in the literature. They either rely on *a priori* knowledge about the classification problem at hand to discard features, or use generic optimization heuristics when no domain-specific information is available (or when it is too difficult to exploit). Algorithms as diverse as stochastic annealing, genetic algorithms, max-min pruning, principal component analysis or neural network node pruning have been introduced (see Jain *et al.*[84] for a taxonomy of feature selection algorithms).

Applications. In view of the tremendous shape and topological variability of the human cortex, brain tissue segmentation in MRI is a natural application for classification techniques.[102] The many intensity inhomogeneities and bias fields mentioned in Sec. 3.2 tend to favor unsupervised clustering approaches[71,103,137,158] even though *a priori* intensity models of the cerebral tissues may also prove adequate.[23,212] However, due to its highly convoluted morphology, the gray matter ribbon comprises a large proportion of partial volume effect voxels in typical T1-weighted MRIs. There are often better handled by fuzzy approaches.[30,156] Classifiers are also ideally suited to extract lesions and tumors in digital mammography[150] or in MRI,[31] using a combination of intensity and texture attributes.

4.2.2. *A few classification approaches*

We detail in this section three commonly used classification techniques in decreasing order of locality (thresholding, region growing and Markov random fields) and briefly comment on more exotic classification techniques at the end.

Thresholding. As far "left" as could be on the locality spectrum, thresholding algorithms (see Sankur *et al.*,[163] Sahoo *et al.*[160] or Lee *et al.*[96] for broad overviews) do not consider any neighborhood in the image *per se*: the classification is based solely on comparing the voxel's intensity and an intensity threshold that is set in advance, or determined globally from the intensity histogram of the input image. A partition of the input image is therefore obtained by partitioning the image intensities. When the histogram is multi-modal, several thresholds can be computed.

[c]Namely, when the dimensionality of the feature space increases, more parameters must be estimated, which enhances the risk of overfitting the model: adding new descriptors then usually first increases the classification performance, which attains a peak value, before decreasing as more descriptors are added (overfitting phenomena).

As could be expected from a local technique, the effectiveness of a thresholding algorithm depends on the contrast between the target structure or tissue and the surrounding ones. Several improvements have been made to increase the segmentation robustness. In Lee *et al.*[95] for instance, connectivity constraints were added to separate regions that would otherwise be incorrectly merged. A number of local histogram thresholding techniques are also available (iterative Bayesian classification,[113] dynamic thresholding,[126] etc).

Thresholding techniques have been mostly applied to the segmentation of digital mammograms, either as a first-stage segmentation tool to provide a second-stage classifier with candidate pathologies,[81,147] or to determine with enhanced accuracy the location of masses in previously pathologically labeled images.[70] Highly contrasted structures such as bones in CT[68,227] or the cavities of the left ventricle in cardiac MR[50,170] can also readily be segmented by thresholding techniques. In Zimmerand *et al.*,[226] ovarian cysts were extracted from ultrasound images with an attribute vector consisting of intensity and texture descriptors.

Region growing. The simplest form of neighborhood information is exploited by region growing approaches that rely on the hypotheses that adjacent pixels have similar characteristics, and in particular, comparable intensity values. Region growing is then an iterative process that produces connected regions. First, a number of seeds are selected in the input image to form single voxel regions. Then, at each iteration, the neighboring voxels of those in the regions are tested against a similarity criterion and those that pass the test are added to the corresponding region. This process is repeated until no more voxels can be added or until a stopping condition is satisfied. Among the many similarity criteria, usually based on features computed on the regions, we find geometric ones (convexity, size, shape of region) and radiometric ones (intensity, texture).[73] However, the fundamental assumption of feature consistency makes region growing techniques very sensitive to noise and imaging artifacts. Homotopic region growing techniques[111] have been developed to constrain the shape of the regions in the face of the potential topological changes (holes, fusion of previously disconnected regions, etc) induced by imaging inhomogeneities. Furthermore, the placement of the initial seeds is particularly difficult to automate (although seed-invariant approaches are available[207]) and depends on the segmentation application (see Sec. 6.2).

In view of its many drawbacks, region growing, like thresholding techniques, often require post-processing. Its main applications are segmenting tumors and lesions in digital mammography[66,92] or in MRI.[146] Figure 3 shows four steps of the segmentation of a human hippocampus with a heavily regularized region growing algorithm initialized with two seeds.

Markov Random Fields. A favored means to model images in the presence of noise and artifact,[75] Markov random fields (MRF) are particularly well-suited to capturing local intensity and textural characteristics of images as they provide a consistent way to model context-dependent entities such as image voxels and

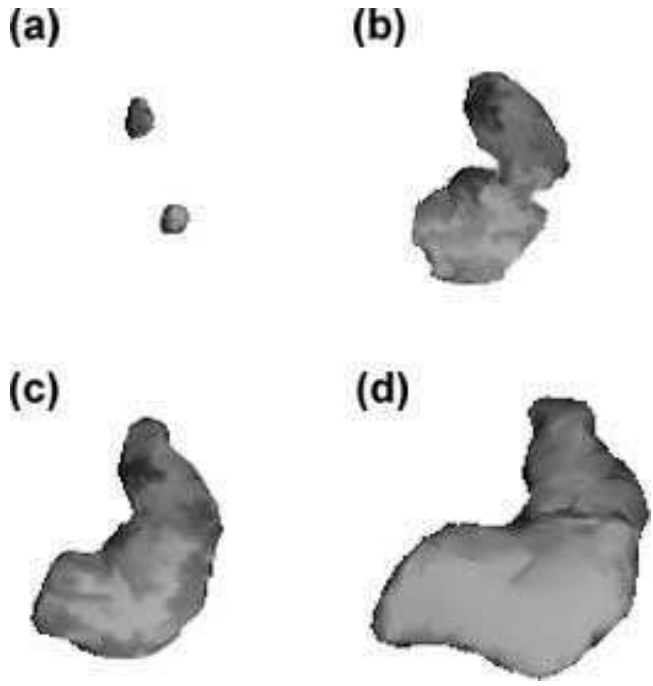


Fig. 3. Four steps of the region growing segmentation of a human hippocampus from a $1\,\mathrm{mm}^3$ T1-weighted MRI of the head.

correlated features.[9,49] They rely on the assumption that the intensity of any given voxel partially dependent on those of its neighbors (that is, that neighbor voxels belong to the same class, or that they must belong to an *a priori* defined class, e.g. voxels from the amygdala class are not allowed to be posterior to voxels from the hippocampus class). In the context of medical image segmentation, the hypothesis becomes that of a low probability for a single voxel of a given class to occur inside an homogeneous group of voxels from another class.

A defining characteristic of an MRF system is the shape and size of the neighborhood system imposed on the voxels of the input image. Inside these neighborhoods, cliques (subset of sites in which every pair of distinct sites are neighbors) are often used to define the conditional MRF distributions that characterize the mutual influences between pixels and textures.

In accordance with the Hammersley-Clifford theorem,[13] an MRF can also be represented by a Gibbs distribution

$$P(x) = Z^{-1} e^{-U(x)}$$

where $Z = \sum_{x \in X} e^{-U(x)}$ is the so-called partition function that acts as a normalizing constant, and $U(x)$ is the associated energy function, which is usually much easier to manipulate.

In this framework, segmentation consists of estimating a label process Λ from the realization of a voxel process Π. Rather than directly modeling $P(\Lambda = \lambda|\Pi = p)$, a Bayesian approach is generally used to compute the conditional probability from a fixed probability law imposed on Λ and an estimation of $P(\Pi = p|\Lambda = \lambda)$. λ can then be iteratively approximated by maximizing this *a posteriori* probability, which boils down to minimizing the compound energy U associated with the MRF model. Among the many optimization approaches, we find iterated conditional modes[13] and stochastic simulated annealing.[64]

In spite of their high computational demands, MRF techniques have been successfully applied to a variety of medical image segmentation tasks. Their ability to handle local inhomogeneities and topologically complex voxel arrangements makes them ideally suited for brain tissue segmentation in MRI.[75,153] Other MRI applications include knee image labeling,[28] cortical sulci[159] and magnetic resonance angiograms segmentation.[200] MRF texture models have proved useful in the segmentation of lung in chest radiographs,[205] bones in CT[135] and, of course, pathological masses in digital mammography.[26,57,101]

Exotic classifiers. In addition to the above mentioned techniques, less standard classification approaches are also available.[54] Neural network, in particular, are worth detailing.

Artificial neural networks (ANN) are mathematical tools that mimic the densely interconnected and parallel structure of the brain, and the adaptive biological processes of the nervous system, both in terms of learning and of information processing. They are composed of a large number of processing elements (so-called neurons) linked together by weighted connections (analogous to synapses). Each neuron receives activation signals from other neurons and outputs a non-linear function of the weighted sum of these activations. This nonlinear mapping function (ϕ) is called the activation function. A commonly used activation function is the sigmoid function (hyperbolic tangent for instance). Output from a neuron $neuron_i$ is then written as:

$$neuron_i(x) = \phi\big(w_i^t x\big) + w_{i,0}$$

where x is the d_i dimensional vector of input signals, w_i is the weight vector (or vector of synaptic weights), and $w_{i,0}$ is a constant bias.

A neural network is then a graph of interconnected neurons, whose connectivity matrix defines the connection pattern. Given a learning set of labeled samples, training an ANN then essentially consists of modifying the weights of its neurons so as to minimize the overall difference between the output values of the network and the target values from the learning set. The most popular learning technique is the so-called back-propagation algorithm, which is based on a straightforward gradient descent technique. Various more sophisticated learning techniques have been developed. Please refer to Ref. 214 for a broad overview.

Neural networks have been mostly used to segment tissues and structures in MR images.[14,155,208,216] Hall *et al.*[71] also used them to segment tumors and

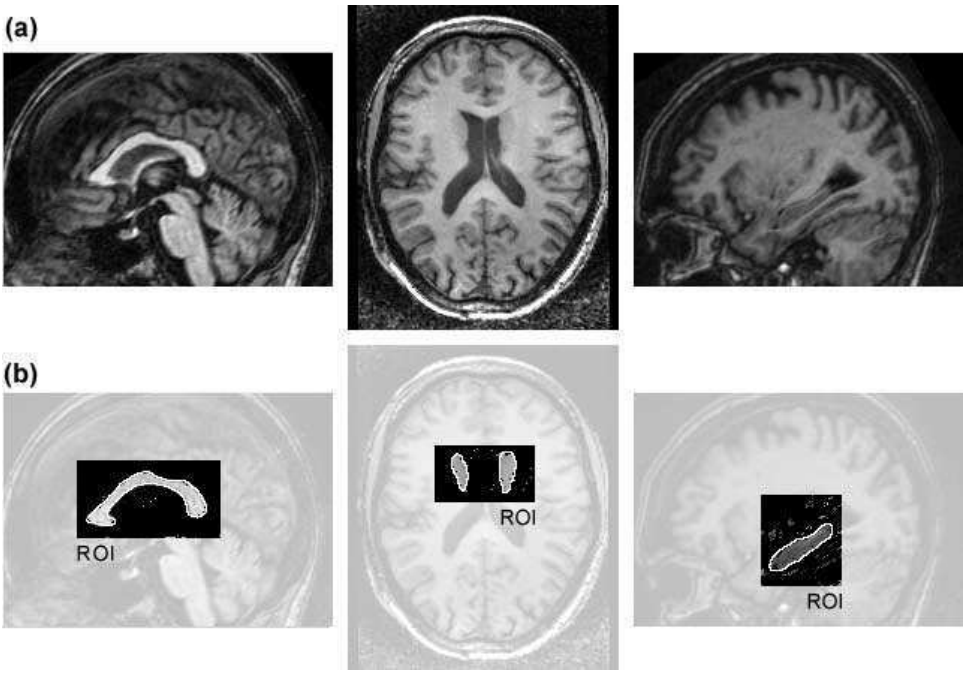


Fig. 4. Neural classification of corpus callosum, caudate nucleus and hippocampus: (a) input MRI; (b) extracted structures (after thresholding of classification map) with true outlines superimposed. The classifier was trained and applied only to the unshaded regions of interests (ROI).

edema. They can also be employed as a pre-processing stage for other segmentation algorithms.[56,140,204] Figure 4 displays the classification results for three cortical structures (corpus callosum, caudate nuclei and hippocampus) obtained with the two stage neural network presented in Pitiot *et al.*[139]

4.3. *Flexibility*

In this chapter, flexibility stands for (1) the actual geometrical and topological flexibility of the segmentation process and of the underlying shape or intensity models if available, as well as (2) their combined expressivity, that is, their ability to represent a variety of shapes and appearances with a minimal number of parameters. Because of their high accuracy, local techniques (classification approaches in particular) are geometrically and topologically very flexible. Clearly, their locality enables them to segment arbitrarily complex arrangements of voxels. However, they are not compact in that they often require as many parameters as there are voxels in the input image to represent the extracted structure or tissue.

We submit that the use of a segmentation model is a necessary condition to achieve true flexibility. We therefore focus our discussion of this issue on deformable models, and in particular on how they are formulated: explicitly or implicitly.

Deformable models[117] are model-based (top-down) segmentation approaches that evolve parametric surfaces or curves in fashions inspired by mechanics and physics.[d] They are characterized both by their surface representation (continuous or discrete, explicit or implicit) and by their evolution law, which determines the space of available shapes (see Montagnat *et al.*[123] for a thorough taxonomy). Once initialized reasonably close to the segmentation target (in terms of position and of shape), they often deform via iterative relaxation of a compound functional E. Classically, E is made up of three terms:

- an internal (or regularization) energy $E_{internal}$ which characterizes the possible deformations of the deformable surface;
- an image coupling energy E_{image} which couples the model to the image; and
- a constraint energy $E_{constraint}$ which regroups the various available constraints (shape, appearance, etc).

We get:

$$E = \alpha \cdot E_{internal} + \beta \cdot E_{image} + \gamma \cdot E_{constraint}$$

with α, β, $\gamma \in \mathbb{R}$.

Typically, the internal energy measures the amount of bending and stretching undergone by the deformable model as it evolves. A large number of image forces are also available.[124] They can be based on the gradient of the input image,[123] on a smoothed version of its associated edge-image,[142] on intensity profiles,[27] etc.

Alternatively, the evolution of the deformable model can be controlled by a dynamic equation within a Lagrangian framework (following the reasoning detailed in Sec. 6.3), or within a Bayesian probabilistic framework.[117]

When the deformable surface is described by coordinate functions that depend on a vector of shape parameters, the model is explicit. Alternatively, implicit formulations model the surface with implicit equation. At a glance, explicit parametric models are the most frugal in terms of parameters, while implicit models, level sets or atlas registration (see Sec. 5.2), win the palm of flexibility.

Explicit parametric models are especially interesting in medical image segmentation for the following reasons. First, as detailed below, they can adequately handle the various discontinuities that sampling artifacts and noise create on the boundaries of the target structures. Also, they compactly describe a wide variety of shapes while minimizing the overall number of parameters or masking these behind a small and easily manageable set of physical principles. They often provide a local, if not global, analytical model of the structure once segmented, which makes it easier to analyze subsequently. Finally, *a priori* knowledge about the shape, location, or appearance of the target structure can guide the deformation process: deformable models are then the framework of choice to mix

[d]As the name indicates, deformable models generally behave like elastic bodies, within a Lagrangian dynamics framework.

bottom-up constraints computed from the input images with *a priori* top-down medical knowledge.

Despite these advantages, explicit models raise several practical concerns, most of which are linked to the somewhat delicate balance between the contributions of the internal and external forces or energies. Clearly, as image coupling forces may drive the models towards the wrong boundaries (especially in the absence of prior knowledge constraints), the regularization constraints must limit the geometrical flexibility of the models. The extent of this limitation is not trivial to determine *a priori*, and will often depend on the variability of the segmentation target and on the input image characteristics. As a result, explicit models often exhibit significant difficulties in reaching into concave boundaries. Balloon forces,[35] gradient vector flow[217] or dynamic structural changes (subdivision of model basis functions,[122] densification of control points in regions which undergo dramatic shape changes[201] or which are too far from the closest potential boundary[142]) are a few of the many techniques developed to control the accuracy of the segmentation scheme and address this restriction. Furthermore, most models cannot accommodate topological changes since these are usually not coded into the model parameters. Still, a few topologically flexible approaches are available in the literature that can adapt the topology of the deformable surface as it evolves.[52,109,118,164,201,215] Finally, they are also notoriously sensitive to initialization (we tackle this particular issue in Sec. 6.2).

A popular implicit formulation, level sets model deformable surfaces[180] using a higher dimensional signed function whose zero level corresponds to the actual surface. From the desired properties of the surface evolution process, an adequate flow equation can be derived for the embedding signed function.

Initially proposed by Sethian and Osher[130,167] to track moving interfaces in fluid mechanics and combustion simulations, level sets alleviate both the parameter granularity issue of explicit approaches (namely, which sampling strategy to choose to parameterize the deformable surface) and their difficulties in handling topological changes. Their main drawbacks are linked to their inherently implicit formulation, which makes it difficult to analyze the segmented surface, once extracted from the input image, in the form of an unstructured set of voxels. It also makes it substantially awkward to incorporate prior medical expertise into them. Finally, like explicit models, they are quite sensitive to initialization. Other implicit formulations include algebraic surfaces,[184] superquadrics[10] and hyperquadrics.[34]

Not surprisingly, deformable models (explicit and implicit) have been mostly applied to the segmentation and tracking of anatomical structures.

Aside from the cerebral cortex[45,63] and the ventricles,[168] they have been widely employed in the segmentation of most of the deep gray cortical structures (corpus callosum,[140,202] hippocampus,[144,202] caudate nuclei,[140,168] etc). A natural segmentation target in X-ray images, the extraction of bones has also proved amenable to the use of deformable surfaces.[134,186,220] Soft tissue organs such as the liver,[62,125] the kidney,[86] the stomach[110] or the heart[25,118,165] have also been targeted.

Because of their dynamic nature, deformable models have been immensely popular in motion tracking applications, most especially in ultrasound images.[108,121]

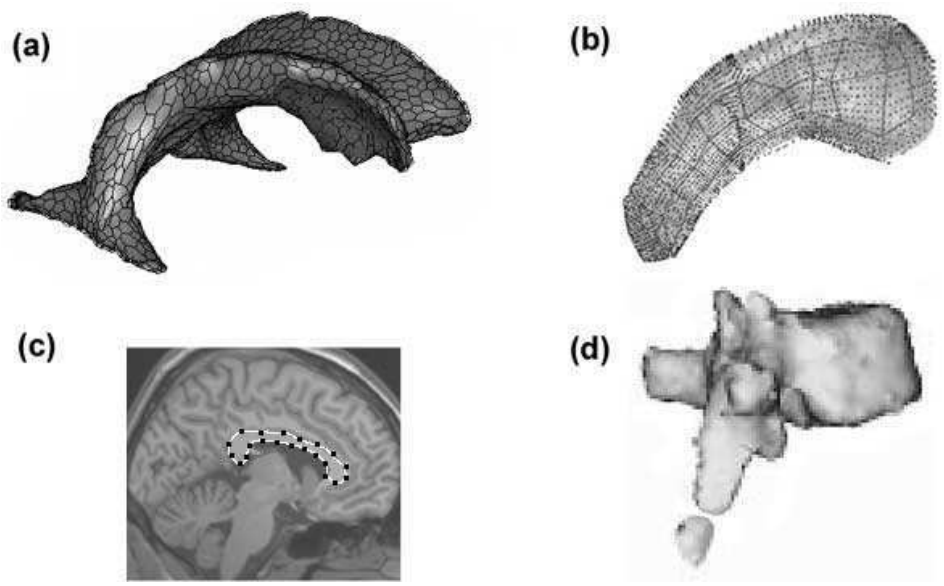


Fig. 5. Four types of deformable models: (a) simplex mesh of ventricles; (b) m-rep of hippocampus (courtesy of Pizer *et al.*); (c) B-spline representation of the mid-sagittal slice of the corpus callosum; (d) level-set of vertebra (courtesy of Leventon *et al.*).

Finally, they have also been applied to the delineation of a variety of lesions and tumors: brain tumors in MRI,[77] cysts in ultrasound breast images,[219] etc. Please refer to McInerney *et al.*[117] for additional applications and Fig. 5 for an illustration.

4.4. *Continuity*

In view of the various irregularities that sampling artifacts or noise induce along the target contours in the input image, boundary continuity is another constraint that substantially affects the analysis of the segmented target. For instance, statistically comparing the shape variability of anatomical structures across diseased populations or between normal subjects and disease groups,[127] or computing an average pattern of gray matter growth across a population of patients,[172] is much more easily and accurately performed when continuous surfaces (discrete meshes in these cases) are used to represent the target tissues or structures instead of the unstructured and quite probably topologically complex sets of voxels that level set techniques would produce.[e] This allows noisy and possibly sparse local image features to be structurally linked inside a coherent and compact model.

[e]Note that, as demonstrated in Leventon *et al.*[100] or Golland *et al.*,[67] average anatomical shapes can still be computed from level set representations by averaging their associated distance maps. However, this approach makes the strong assumption, that pixels at the same location across the set of level set segmentation are homologous which certainly does not hold when topological changes occur, even when relaxed by the influence of the diffusion factor of the distance computation.

Use of a continuous segmentation model (deformable models, for instance) in conjunction with a continuous representation of the image space (via interpolation techniques[98]) also enables sub-voxel accuracy. This is especially interesting when the radius of the target structure is not large with respect to the resolution of the input image (for example, for the segmentation of small lesions in MRI, or of organs in standard $2 \times 2 \times 4$ mm PET scans). The segmentation results are however particularly difficult to validate in this case as little to no information is available to compare, inside each voxel, the location of the segmented contours to those of the actual structures.

On the other hand, continuity can be a hindrance to accurate segmentation. For instance, the quantification of cortical gray matter loss in Alzheimer disease[188] does not require that voxels classified as gray matter form a single connected component. Ensuring the extraction of the cortical gray layer by a single continuous surface or by a single voxel connected component would only prevent difficult to reach regions from being segmented. As an illustration, the inferior horns of the human brain ventricles are connected to the rest of the lateral ventricles via a group of partial volume voxels that are exceedingly difficult to segment in standard $T1$-weighted $1\,\text{mm}^3$ resolution MRIs (to the point where a lot of manual delineation protocols will just exclude them for fear of introducing an artificially high shape variability). Consequently, a region growing approach initialized with seeds inside the main body of the ventricles only would most probably stop at the PVE voxels and discard the inferior horns altogether. Likewise for deformable model techniques, the image coupling energy, often linked to the image gradient, would most likely prevent the model from reaching as far as the inferior horns. Clearly, this difficulty would not impede other classification techniques even though these might incorrectly exclude some of the PVE voxels. Similar considerations apply to the segmentation of the tail of the caudate nucleus in brain MRI (Fig. 7).

4.5. *Surface versus volume*

Aside maybe from the cortical gray matter layer which can be handled as a surface in typical $1\,\text{mm}^3$ resolution brain MRIs, most segmentation targets correspond to actual anatomical volumes, and behave like deformable solids. When the target boundaries have to be extracted from a single static image though, choosing between a surface or a volume representation is arguably a mere matter of taste since most of the features of volume segmentation systems useful on static images have their counterparts in surface segmentation systems and vice versa. However, in dynamic segmentation cases (organ tracking, for instance) much can be gained by using a volumetric framework within which physiologically or anatomically motivated constraints between structures or tissues are more easily incorporated.

Volumetric approaches are especially interesting in cardiac motion tracking applications[117,136,166] as they can model the thick-walled left ventricles as a whole instead of requiring difficult to synchronize endocardial and an epicardial surface models.[1]

For these dynamic segmentation problems (which are often functional in nature), volumetric biomechanical models[116,132] could help increase the overall robustness, owing to their ability to encode the dynamics of the shape, appearances and relative positions of structures in an anatomically accurate and mechanically plausible way.

Undoubtedly, the choice of a surface or volume representation is in part dictated by how easily the extracted segmentation target can be analyzed in the envisaged application (average shapes are more easily computed on sets of surfaces than volumes, tissue loss is often simpler to assess from sets of voxels, etc) and in part determined by the characteristics of the target itself embedded inside the input image (volumetric frameworks have typically proved more efficient than dual surface approaches in preventing thick walls from collapsing in the presence of heavy noise).

5. Expert Knowledge (*a priori* Information)

However variable in shape and appearance the target structures or tissues may be, their general morphology, contrast and relative position with respect to surrounding tissues and neighborhood structures is often known. In view of the many intensity inhomogeneities and the plethora of artifacts and decoys present in the input image, this *a priori* medical knowledge is an invaluable tool in the search for the best trade-off between accuracy and robustness. In addition to facilitating the actual segmentation process, shape, appearance and position models can also significantly assist the subsequent analysis of the segmented target. Clearly, compact models are more easily interpreted and compared (between themselves or against statistically built average models and modes of variation) than tediously long lists of vertices or voxels.

The available corpus of medical information can be leveraged in essentially two ways: implicitly (computationally) and explicitly. Given a learning set of *a priori* segmented instances of the segmentation target, implicit knowledge algorithms have to automatically discover the relationships and functional dependencies of the various parameters of the model. However, explicit information about the target is often available, in the form of medical expertise. For instance, the relative positions of most of the deep gray nuclei in the brain is fairly consistent across individuals, anatomical structures do not usually intersect, etc. From these observations, a series of rules can be derived to better drive the segmentation process. Broadly speaking, explicit knowledge approaches can be seen as a special case of implicit knowledge algorithms where the additional medical expertise provides short cuts in the search for the target structure or tissue.

As mentioned earlier, model-based (top-down) methods are more amenable to the introduction of medical knowledge. Nonetheless, bottom-up techniques such as thresholding and region growing can also benefit from intensity models built from *a priori* observations.

We review below a selection of approaches organized according to the type of knowledge that they model: shape and appearance in Sec. 5.1 and position in

Sec. 5.2. In each case, we propose a number of implicit and explicit knowledge techniques.

5.1. *Modeling shape and appearance*

5.1.1. *Implicit models*

Even though a given structure can present a wide variety of forms, the notion of biological shape seems reasonably well explained by a statistical description over a large population of instances. Consequently, statistical approaches have attracted considerable attention.[40,41,175] A deformable model is then constrained not only by the number of degrees of freedom imposed by its geometric representation, but also in that it must be a valid instance of the shape model. Developed by Cootes and Taylor,[41] active shape models are represented by both a set of boundary/landmark points and a series of relationships established between these points from the different instances of the training set. New shapes are modeled by combining in a linear fashion the eigenvectors of the variations from the mean shape.[198] These eigenvectors encode the modes of variation of the shape, and define the characteristic pattern of a shape class. The shape parameter space serves as a means to enforce limits and constraints on the admissible shapes, and insure that the final extracted shape presents some similarity with the shape class, as established from the training set. Many variants have been presented in the literature.[80,209] They either introduce more constraints or decrease the control over admissible shapes. In particular, active appearance models[40] incorporate both a statistical model of the shape of the target, and a description of the statistical distribution of the gray-level intensities of the structure. A similar PCA approach was applied to the signed functions embedding level set representations in Leventon *et al.*[100]

Blum *et al.*[17] introduced the medical representation as a model of biological growth and a natural geometry for biological shapes. Pizer *et al.*[143] derived a sampled medial model. Joshi *et al.*[86] used it within a Bayesian framework to incorporate prior knowledge of anatomical variations. A multi-scale medial representation was used to build the template examples needed to obtain prior information about the geometry and shape of the target anatomical structure. Within this framework, the anatomical variability of a structure corresponds to the distribution of the admissible transformations of the shape model. Medial representations are however difficult to build for non-symmetrical shapes and are notoriously plagued by topological difficulties.

Staib *et al.*[175] used a similar Bayesian scheme to control the coefficients of an elliptic Fourier decomposition of the boundary of a deformable template. They introduced a likelihood functional, which encoded the spatial probability distribution of each model, to be maximized under a Bayesian framework. The distribution of the model parameters was derived from a learning set of instances of the target object, and served to constrain the deformable template towards the most likely shapes. Székely *et al.*[182] added an elastic property to a Fourier decomposition to create elastically deformable Fourier surface models. A mean shape and its associated

modes of variation were extracted via statistical analysis of a learning set of Fourier decomposed instances of the target structure. The elastic fit of the mean model in the shape space was used as a regularization constraint.

Styner *et al.*[179] combined a fine-scale spherical harmonics boundary description with a coarse-scale sampled medial description. The SPHARM description, introduced by Brechbühler[20] is a global, fine scale parameterized description which represents shapes of *spherical* topology. It uses spherical harmonics as a basis function. Styner's medial models were computed automatically from a predefined shape space using pruned 3D Voronoï skeletons to determine the stable medial branching topology.

Metaxas *et al.*[120] devised deformable superquadrics which combined the global shape parameters of a conventional superellipsoid with the local degrees of freedom of a membrane spline. The relatively small number of parameters of the superellipsoid captured the overall shape of the target structure while the local spline component allowed flexible shape deformation in a Lagrangian dynamics formulation. Vemuri *et al.*[203] used the properties of an orthonormal wavelet basis to formulate a deformable superquadric model with the ability to continuously transform from local to global shape deformations. Such model can continuously span a large range of possible deformations: from highly constrained with very few parameters, to underconstrained with a variable degree of freedom. Here again, a Bayesian framework biased the deformable models towards a range of admissible shapes.

Poupon *et al.*[149] proposed the use of 3D moment invariants as a way to embed shape distributions into deformable templates. They devised a framework capable of dealing with several simultaneously deforming templates, thanks to their fairly low updating cost, with the goal of segmenting deep grey nuclei in 3D MRI. The remarkable stability of the invariant moments allowed them to study the anatomical variability of the deep gray nucleir in brain MRI.

A given instance of the target structure may not always exhibit homogeneous intensity distribution along its boundaries. Yet, the intensity may be locally characteristic. The intensity profile, computed along the border of the structure models, then provides an efficient means to introduce *a priori* knowledge. Cootes *et al.*,[39] for instance, modeled the statistical distribution of the intensity profile on each side of the structure surface. The mean profile, established from a structure learning set, was compared against the image data to determine the cost of a particular configuration of the model and guide the segmentation process. Brejl *et al.*[21] used a somewhat similar border appearance model to automatically design cost functions that served as a basis for the segmentation criteria of edge-based segmentation methods.

Note that most of these statistical or Bayesian approaches require that correspondences between the shapes in the learning set be available *a priori*, a non-trivial problem in itself.[33,46,48,59,89,141,192,196,209]

5.1.2. *Explicit models*

An even larger variety of explicit knowledge techniques is available in the literature. These approaches tend to be more heterogeneous as they usually combine shape

and intensity descriptions in the same framework. Often, explicit information is complemented or generalized by implicit information (for instance, a purely explicit position rule can be made more robust as a fuzzy condition, which however introduces non-explicit elements: the α parameter of the cut-off, the amount of diffusion, etc).

Since the seminal work on spring loaded templates by Fischler *et al.*,[58] many explicit knowledge approaches have been proposed in the literature to incorporate computationally extracted medical expertise about the shape or appearance of a structure.

Early work frequently relied on highly specific hand-crafted models. Yuille *et al.*[221] chose to use circles and parabola to retrieve eye and mouth patterns in face pictures. Noticing the elliptical shape of the vertebra in axial cross section images of the spine, Lipson *et al.*[104] used deformable ellipsoidal templates to extract their contours. These methods present the advantage of providing a very economical description of the shape in terms of the number of required parameters but lack genericity in that a new model with new parameters has to be developed with each new object.

Even though ASM can handle disconnected shapes it is easier to partition a complex shape (i.e. the vertebral column), into simpler and more manageable elements (the vertebrae). Nothing this, Bernard *et al.*[12] devised a two-level hierarchical scheme to model both the shape and the topology of the resulting complex model. Each individual structure was controlled by its own ASM, subject to an overall global ASM encoding the relative positions and orientations of the set of components.

Amit and Kong[3] used a complex graph of landmarks, automatically chosen from the input images as a topological model, to guide the registration process of X-ray images of the hand. A dynamic programming algorithm was used on decomposable subgraphs of the template graph to find the optimal match to a subset of the candidate points.

As it can represent and merge uncertain or imprecise statements, fuzzy theory is particularly well-suited to model shape. Among others, Chang *et al.* developed a fuzzy-controlled rule-based system capable of segmenting MR images of diseased human brains into physiologically and pathologically meaningful regions by incorporating expert knowledge about both brain structures and lesions. They used the distance between pixels and the ventricular boundary as a fuzzy property of periventricular hyperintensities to help diagnose the studied disease. Barra and Boiré[11] used information fusion to combine medical expertise with fuzzy maps of morphological, topological, and tissue constitution data to segment anatomical structures in brain MRIs. For instance, they encoded expert information about the relative position of two structures as a fuzzy distance map. Wen *et al.*[213] used fuzzy-embedded human expert knowledge to evaluate the confidence level of two matching points using their multiple local image properties such as gradient direction and curvature. Studholme *et al.*[177] merged region labeling information with

classic iconic image registration algorithm via information fusion to align MR and PET images of the pelvis.

When anatomic knowledge can be captured by a series of simple positional, geometric or intensity rules, expert systems provide a convenient framework to assist in segmentation tasks. Ardizzone[5] for instance developed a descriptive language to express the geometric features and spatial relationships among areas of images. Reference 115 also used a rule-based system to organize and classify features (such as brightness, area, neighborhood, etc) for regions that had been automatically extracted via region growing and they segmented scalp, gray and white matter, CSF and strokes. In Brown *et al.*,[22] lung boundaries were segmented in chest X-ray images by matching an anatomical model to the image edges using parametric features guided by a series of rules. Li *et al.*[101] described a knowledge-based image interpretation system to segment and label a series of 2D brain X-ray CT-scans. Their model contained both analogical and propositional knowledge on the brain structures, which helped interpret the image primitive information produced by different low-level vision techniques. Finally, Poupon *et al.*[149] used 3D moment invariants to embed shape distributions in deformable templates. They devised a framework that could deal with several simultaneously deforming templates, with a fairly low updating cost, to segment deep gray nuclei in 3D MRI. We presented in Pitiot *et al.*[140] an expert-knowledge guided system which evolved, in parallel, a number of deformable models (one per target structure). These evolutions were supervised by a series of rules and meta-rules derived from *a priori* analysis of the model's dynamics and from medical experience. The templates were also constrained by knowledge on the expected textural and shape properties of the target structures (caudate nuclei, ventricles, corpus callosum and hippocampus in T1-weighted MRIs).

5.2. *Position*

Often, the positions (and shapes) of nearby anatomical structures are not independent of each other. For instance in the human brain, the caudate nuclei are juxtaposed to the lateral ventricles, so any change in the shape or position of one will affect the other. Information about the respective position of structures can then dramatically help the segmentation process. Positional knowledge can be either relative (with respect to neighborhood structures) or absolute (with respect to an anatomical atlas or standardized coordinate system).

5.2.1. *Distance constraints*

Relative positional knowledge often takes the form of distance constraints. In Barra and Boiré[11] for instance, fuzzy logic was used to express both distance and positional relationships between structures. In Tsai *et al.*,[197] a series of parametric models, built via principal component analysis of multiple signed distance functions, enabled the concurrent segmentation of anatomical structures, via minimization of

a mutual information criterion. Inter-object distance constraints were also used in Yang *et al.*[218] where a maximum *a posteriori* estimator for anatomical shapes helped constrain the evolution of level set functions.

In Pitiot *et al.*,[140] we also chose distance maps as they can model distance constraints accuratly and robustly (guaranteeing non-intersection, for instance). Given a deformable model Π^0 (a simplex mesh[51]), we wished to impose on it a distance constraint with respect to another model Π^1. We first computed the distance map D^1 associated with a discrete sampling of Π^1. We used a classical Chamfer map[19] algorithm to compute a signed distance map, positive outside the discrete sampling of Π^1 and negative inside. At each vertex P_i^0 of Π^0, we then computed a "distance force" $f_{distance}$ whose magnitude depended on the value of the distance map at the considered vertex. We derived two types of constraints. For some pairs of structures, we wanted the force to attract the vertex, along the direction of the gradient of the distance map, up to an exact distance d_{target} of the target mesh: For other pairs, we only wished to enforce that this same vertex remained at distance greater than d_{target} (to prevent intersections between structures for instance). Note that these forces could also be applied to a subset of the mesh vertices (so-called "zones") to enforce more local constraints.

5.2.2. *Atlas warping*

A hybrid shape/position explicit knowledge approach, atlas registration or warping[8,36,183] enables the concurrent segmentation and labeling of several target structures. Prior anatomical expertise about the shape, orientation and position of the target structure is projected onto a 3D atlas, usually modeled as an elastically (or fluidly) deformable object, to be used as anatomical reference. Segmenting the target structures then boils down to registering the *a priori* labeled atlas to the input image. In effect, this transforms a model-to-region matching problem (the initial segmentation task) into an intensity to intensity matching one (iconic registration of two images). As a result, the effectiveness of the process relies on the assumption that the actual target structures in the input image are only marginally different in shape, orientation and location from the ones in the atlas, a reasonable hypothesis in non pathological cases.

Atlas techniques usually consists of two phases. First, the atlas is initialized over the input image with a rigid or affine registration algorithm.[4,93] Then, a non-linear registration stage corrects for the finer anatomical differences.[29,36,44,162] Not surprisingly, the main drawbacks of this approach are classical registration issues. To begin with, a poor linear initialization will undoubtedly yield an even worse non-linear registration. Second, because of their tessellated nature, biomedical images are packed with decoy structures which look surprisingly close to the target ones and may fool the registration process. Finally, because of the high variability of the segmentation targets, the non-linear registration process may not be flexible enough to adapt the atlas contours to the convoluted boundaries of the actual structures, all

the more since the regularization parameters of the non-linear registration have to be kept fairly high to prevent warping to incorrect boundaries and to avoid producing self-intersecting boundaries. One way to alleviate this issue is to restrict atlas registration to only the initialization step of the segmentation process (see Sec. 6.2). Another work-around consists in using preprocessing techniques. In[161] for instance, 3D edge detection and a series of morphological operators were used to extract from the input MR images the main cortical sulci. These helped increase the convergence speed of the atlas warping procedure by providing a smooth representation of the cortical surface.

Atlas warping has been mostly applied to the segmentation of cerebral structures in MR images of the brain.[194] Aside from the segmentation *per se*, it also provides a standard reference space in which to study the morphometric properties of structures and organs.[47,85]

6. Robustness

A segmentation system can arguably never be robust enough, as exemplified by the variety of techniques discussed in the literature to cope both with the high variability of the target structures and with the noise characteristics of the input images. However, as already mentioned above, increased robustness often comes at a cost, that of decreased accuracy. As always, trade-offs have to be found, which we discuss in this section, along with the two main robustness factors: initialization (Sec. 6.2) and the optimization framework (Sec. 6.3).

6.1. *Generic versus specific*

In the absence of a single segmentation algorithm capable of effectively handling all segmentation applications with satisfactory accuracy and robustness, most segmentation approaches have to deal with the delicate balance between genericity and specificity. On the one hand, generic techniques perform reasonably well over a large number of applications, mostly due to a high robustness to noise and imaging artifacts. On the other hand, application-specific methods are more accurate, the use of adapted prior knowledge increases their robustness and they can deal optimally with artifacts associated with the images they have been specifically trained on. In between these extremes, application-tailored approaches provide the user with the means to adapt an otherwise generic segmentation technique to the application at hand. For instance, the statistical shape constraint techniques we reviewed in Sec. 5 effectively adapt generic deformable model formulations to segment specific target structures (or a specific class of target structures).

Specialization is all the more attractive since several optimization tricks can be applied to improve the segmentation performance and speed when the application is restricted to a limited domain. In motion tracking[117] for instance, the boundaries of the segmentation target extracted at a given time may serve to initialize

the segmentation of the same target at the next time instant, a tactic that relies on the assumption that the target exhibits only small changes in shape and position between time instants. Although initially developed in the context of computer vision,[88,187] the most popular motion tracking application is probably the analysis of the dynamic behavior of the human heart, the left ventricle in particular.[117] The multi-channel capabilities of MR systems also motivate the increasing specialization of algorithms. Indeed, a variety of MR pulse sequences are available, where each sequence yield a different distribution of the tissue contrast characteristics. In the event where a segmentation system should be applied to images acquired with the same sequence on a single scanner, a careful study of the imaging characteristics of the sequence would most probably favor a combination of highly specific bottom-up strategies and specifically tailored generic approaches. Conversely, it is sometimes possible to determine the optimal pulse sequence for a particular segmentation target or application.[151,173] The optimized MR acquisition processes are then specifically tuned to maximize the contrast between the tissues underlying the segmentation target and their surroundings, thereby facilitating the segmentation process.

Furthermore, real time issues and other resource constraints (CPU power, memory occupation) may severely impede the adaptation of a segmentation system from one application to another. Sophisticated model-based techniques are not particularly fast for instance and must be optimized in speed at the expense of great efforts if they are to be used in the surgical arena.[210]

At any rate, segmentation systems will most probably require a large amount of specialization to become fully automated.

6.2. *Initialization*

As discussed throughout this chapter, the amount of noise present in the input images, the intensity inhomogeneities and imaging artifacts that plague them and the variability of the segmentation targets all contribute to a poorly structured and highly non-convex space that the segmentation system must traverse in search for the target's boundaries. Most approaches would only lead to weak sub-optimal solutions (where the deformation model adapts to noise or decoys or maybe only follows parts of the desired boundaries) if the search space were not drastically reduced by assuming that a good approximation to the solution was available. This can be either in the form of a set of pose parameters (position, orientation, scale) or shape and appearance descriptors.

Various approaches have been presented in the literature to overcome this robustness issue. Some are specific to a particular segmentation technique (histogram peak detection for region growing for instance), others (such as atlas registration) are applicable across a wider ranger of segmentation strategies.

Classification techniques often require *ad hoc* initializations. When only limited *a priori* knowledge about the characteristics of the target voxel attributes is

available, in PET tumor or lesion detection applications for instance, the salient peaks in a histogram of the voxel attribute values can be used to seed region growing algorithms. Other techniques ensure relative insensitivity to seed position.[206] Note that the closely related split and merge algorithms effectively avoid this seed positioning difficulty.[106,112,133] On the other hand, when the intensity characteristics of the target structure or tissue can be statistically estimated, they can help initialize the various Gaussian means, variances and mixing coefficients of EM and Bayesian classification approaches, to ensure better classification performance.[97]

In view of their inherent complexity, model-based approaches are certainly even more sensitive to initial parameters. As pointed out by Xu and Prince,[217] the initial distance between the model and the target structure (both in terms of actual Euclidean distance and of morphological difference) and the ability to reach into concave boundaries are the two key difficulties of parametric deformable models. These have been tackled by numerous authors. Blake *et al.*[15] for instance implemented a coarse to fine strategy, the Graduated Non-Convexity Algorithm, where a scalar parameter controlled the amount of "local" convexity. To resolve the issue of the capture range of the segmentation target within a highly cluttered and tessellated environment, the models can also be initialized at a number of locations and evolved in sequence: the deformed model with the best final match is then selected. In Pitiot *et al.*,[142] a hybrid evolutionary algorithm controlled a family of deformable templates that were evolved simultaneously to explore the search space in a robust fashion. A pre-processing stage involving filtering, thresholding or morphological techniques may also be useful.[223]

Yet another series of techniques utilizes linear or non-linear registration to initialize the deformable models or the seed points of region growing approaches or level set methods reasonably close to their expected positions.[162]

In Pitiot *et al.*,[140] we selected an MRI brain dataset for its "standard" appearance (the reference MRI), and the target structures were carefully segmented in it (see Fig. 6(a)) following established anatomic delineation protocols. Given an input MRI to be processed, the first step consisted of registering the reference MRI to it with a non-linear registration algorithm with an elastic prior (the MAMAN algorithm[24]). The transform obtained was then applied to the meshes segmented in the reference MRI. These transformed meshes served as initial guesses for the segmentation of the target structures (Fig. 6(b)).

6.3. *Optimization scheme*

While the segmentation model determines the structure of the space of extracted shapes (see Sec. 4), the optimization scheme conditions the ability of the model to traverse this space in search of the correct boundaries.

Most of the segmentation approaches we have mentioned in this chapter can be cast as an optimization problem whereby the search for the target boundaries corresponds to the search for the global minimum of a given functional. The difficulties

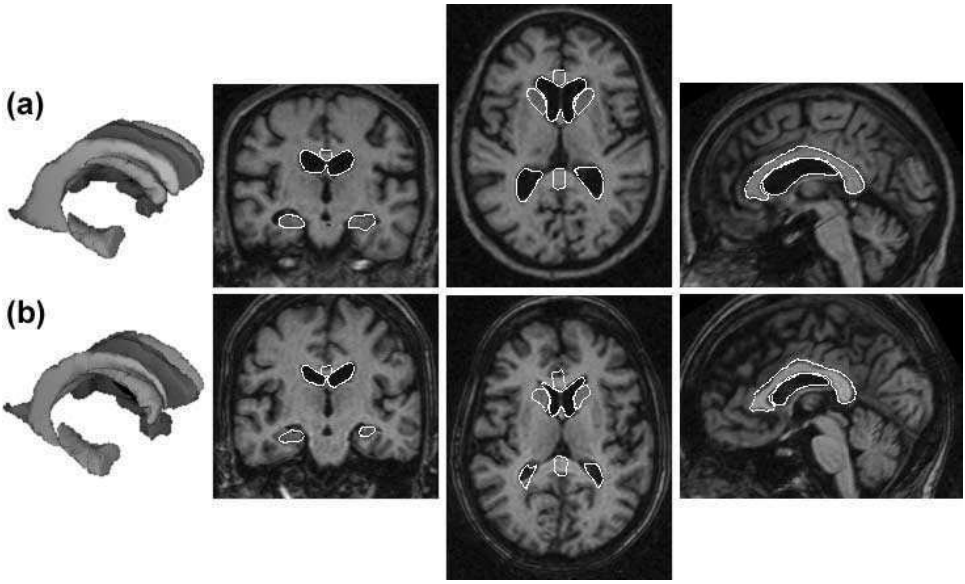


Fig. 6. (a) Reference MRI with manually delineated structures and (b) reference MRI registered to an input MRI and initialized structures.

linked to the choice of a suitable segmentation criterion (which we identified in the introduction to Sec. 2) carry over to the determination of an adequate objective function. Namely, the remarkable variability in shape and appearance of the segmentation targets, the intensity characteristics of the input images (noise distribution, artifacts, contrasts between structures, etc) and the varying objectives of the envisioned applications all contribute to the burden of an appropriate design. However in return, formulating the segmentating process as an optimization problem clearly states its objective: finding the global minimum of the functional.[f]

As with deformable models, the objective function to be minimized can be thought of as an energy, or as a sum of energies $E = \sum_i E_i$. A necessary condition to minimize E is the zero crossing of its first derivative: $\nabla(E) = 0$, which effectively can be read as a balance of forces (with one force per energy component).

When it is not possible to minimize E in a static or algebraic way, either theoretically or practically, a dynamical system can be built whose evolution to equilibrium yields the same minimum.[124] An estimation of the target boundaries can then be obtained iteratively. In fact, it is sometimes easier to directly design the laws of

[f]In view of the somewhat irregular nature and lack of convexity of the landscape of the optimization functional, it would be illusory to expect that the global minimum would effectively coincide with the target boundaries, if it can be found at all. Yet, in practice, the many *a priori* knowledge constraints and *a priori* information imposed on the segmentation process increase the probability that a good approximation of the true boundaries coincides with a reachable local minimum not too far away from the global one.

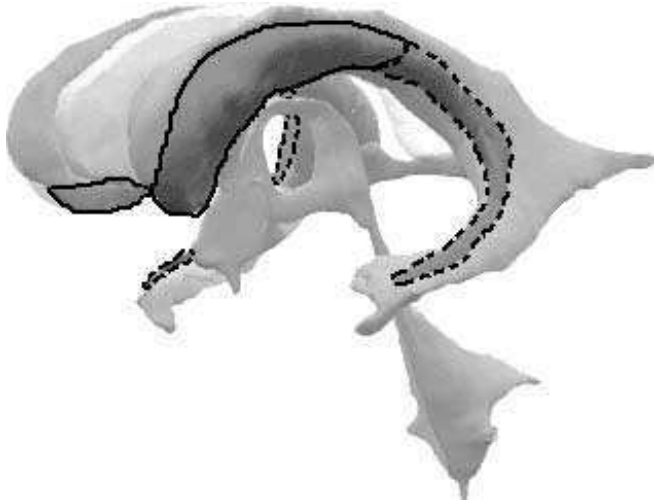


Fig. 7. Anatomically correct caudate nucleus (continuous and dotted black line) and manually segmented caudate nucleus (continuous black line) as obtained from most delineation protocols: the caudate tail (dotted line) is explicitly cut to reduce delineation variability. The nearby ventricles and corpus callosum are rendered in light gray.

evolution that control such a system, most especially when some of its terms do not derive from an energy, as is often the case with medical knowledge-based constraints. An added benefit is the offered possibility to converge towards the local minimum closest to the initial position. This proves useful in semi-automated segmentation systems where a manual estimate of the target boundary can be refined automatically.

However invaluable they can be in increasing the robustness to noise and imaging artifacts, these shape or appearance models induce a poorly structured optimization space when they serve as constraints on the deformation process. The matter is made worse by the high variability of the segmentation target and the tessellated nature of medical images. All in all, we are left with a very difficult minimization problem.

Several algorithms have been developed to remedy this difficulty,[124] most of them coarse-to-fine strategies. In a multiscale framework[128,187] for instance, the segmentation is first performed on a smoothed downsampled version of the input image and successively refined at higher resolutions. With multiresolution techniques (pyramidal schemes), the segmentation model itself is subjected to changes in resolution (multi-scale pyramid of basis functions in Székely *et al.*,[182] octree-spline in Széliski *et al.*,[181] and dynamic mesh decimation in Lötjönen *et al.*[105]). Among other optimization strategies, dynamic programming was used by Amini *et al.*[2] or Coughlan *et al.*[42] to increase the spectrum of the search for the global minima. Poon *et al.*[148] selected simulated annealing, owing to its ability to reach the global minimum and to incorporate non-differentiable constraints.

In Pitiot *et al.*,[140] most of the parameters controlling the segmentation process were dynamically modified, along with the deformation of the models. The overall robustness was increased without sacrificing too much accuracy by dynamically controlling the balance between the two and adapting it to the segmentation problem at hand.

7. Validation

Characterizing, both qualitatively and quantitatively, the performances of an automated segmentation system is all the more pivotal since the available algorithms have only limited precision and accuracy. In other words, since segmentation approaches make compromises, the validity of the underlying assumptions must be checked against the envisaged applications. Yet, medical image segmentation validation is still an open problem plagued with challenges and conflicting objectives.

The foremost difficulty stems from the absence of ground truth. Given the current resolution of the most common imaging modalities and the artifacts that afflict them, even human operators cannot establish the actual boundaries of the segmentation targets from the acquired data with sufficient accuracy. The availability of physical or simulated phantoms[37,178] partially alleviates that issue, but their limited sophistication prevents them from accurately modeling either the large anatomical complexity of organs and structures (for physical phantoms) or the full flavor of imaging characteristics (noise, inhomogeneities, partial volume effects) both in normal and pathological cases (for simulated phantoms).

Often, the target boundaries extracted by an automated algorithm are compared against those manually delineated by a series of experts, under the assumption that the expert delineations are a good approximation of the actual boundaries. However, as demonstrated by many intra- and inter- operator variability results,[127,129,154,228] manual segmentation accuracy is often poor when target structures are difficult to segment, sometimes to the point where the manual delineation protocols must explicitly discard certain anatomical parts from the target structures to limit the delineation variability and avoid introducing spurious outlines (as an illustration, the segmented caudate nuclei reported in Pitiot *et al.*[140] have a very short tail, and the inferior horns of the ventricles are missing; see also Fig. 7). Approaches to establish the true boundary from a series of manual delineations are being investigated.[211]

What is more, the *modus operandi* of the validation studies must be adapted to the objectives of the application for which the structures are segmented in the first place. For instance, longitudinal studies typically rely more on reproductibility than actual accuracy: systematic segmentation biases may be tolerable so long as the segmentation targets are outlined consistently across a large population of input images. The ability to adequately handle pathological cases is also an application parameter that may influence the design of a segmentation approach. Consequently, although assessing the behavior of a system tuned for standard cases on pathological

data is certainly informative, it would seem unfair to penalize it on such a ground. Clearly, as mentioned above, the increased robustness yielded by introducing prior medical knowledge is often counterbalanced by decreased accuracy, especially with shape models that tend to forbid the segmentation of non-standard boundaries.

The variety of segmentation objectives is clearly reflected in the diversity of validation metrics (see Zhang[222] for a broad overview). Some authors use volume overlap,[53,82] or ratios of misclassified voxels,[75] others the Hausdorff distance between contours.[65] Agreement measures have also been developed[225] as a means to draw a most probable decision from a set of expert ones. Unfortunately, from a practical point of view, how accurately and easily a given validation metrics can be implemented often depends on the chosen segmentation model. This also introduces errors in the quantitative validation. For instance, the computation of the Hausdorff distance between two continuous curves or surfaces runs into quantization problems as the curves must be discretized. Furthermore, manual outlines are often obtained in the form of a set of voxels, at a somewhat coarse resolution for most radiographical modalities, which is again likely to introduce artificial errors in the validation measures.

All in all, a consensus still has to emerge as to which validation strategy to apply in each case.

8. Concluding Remarks

With the advent of increasingly powerful and gradually less invasive modalities, medical imaging has become an invaluable tool for clinical use as well as research applications. While the quest for even better acquisition techniques continues, the attention of the medical imaging community has now shifted to the extraction of meaningful anatomical and physiological information out of the ever growing databases of medical images. Before the function, morphology or inter-relationship of organs, structures and tissues can be fully investigated, these must be isolated, that is segmented, from their embedding images. Mirroring the great variety of modalities and medical or research objectives, an even larger variety of segmentation systems has therefore been, and is still being, developed.

Yet, accurate and robust segmentation remains a challenge beset by a number of issues that we discussed throughout this chapter. Clearly, in view of the complexity of the segmentation problem, there are no general prescriptions for selecting a "good" segmentation method. This choice must not only be driven by the characteristics of the input image (imaging artifacts, signal-to-noise ratio, contrast of the segmentation target with respect to surrounding image features, etc) but also by the possible usage constraints (algorithmic complexity with respect to available memory/CPU resources, time limits if real-time applications are envisioned, etc) and of course by the downstream treatments that follow this segmentation step (diagnosis, morphometric analysis, shape recognition, etc).

However helpful automated segmentation systems could be in replacing manual operators, their limited accuracy and, more importantly, their inadequate robustness still prevent their widespread application.

In a research environment where time is less of a premium than the quality of the overall analysis, the parameters of a segmentation system can always be set to specifically fit the requirements of the envisoned application, in terms of accuracy or robustness. Furthermore, the segmentation results can be, and usually are, thoroughly checked before the actual analysis takes place. Even so, the rapid growth of image databases presses the need for fully automated tools. Because of the sheer size of the image collections to be processed, database applications are more forgiving with regard to the shortcomings of segmentation systems. Clearly, invaluable information can always be extracted even when the algorithm employed suffer from statistical biases, as long as these are consistent.

In a clinical setting though, time is a precious resource that has to be managed tightly. To be useful, a segmentation system must exhibit maximum accuracy with the standard "out of the box" set of parameters. To gain physicians' trust, it must also be sufficiently robust not to require any exhaustive and tedious quality checking phase.

To achieve this long-term goal, substantial progress is still to be made. In the meantime, semi-automated segmentation is likely to remain the favored means to assist in the labor intensive tasks of medical imaging analysis. By assisting manual operators rather than replacing them, partial automation effectively decreases the segmentation burden without compromising the trust placed in the quality of the final results. Semi-automated delineation tools that can complete delineations based on prior shape knowledge, atlas registration assisted segmentation systems, or expert system controlled segmentation approaches that communicate with the operator to attract his attention to potential problems, are set to obtain an increasing share of the limelight.

References

1. A. A. Amini and J. S. Duncan, Bending and stretching models for LV wall motion analysis from curves and surfaces, *Image and Vision Computing* **10**, 6 (1992) 418–430.
2. A. A. Amini, T. E. Weymouth and R. C. Jain, Using dynamic programming for solving variational problems in vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**, 9 (1990) 855–867.
3. Y. Amit and A. Kong, Graphical templates for model registration, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**, 3 (1996) 225–236.
4. N. C. Andreasen, R. Rajarethinam, T. Cizadlo, S. Arndt, V. W. Swayze II, L. A. Flashman, D. S. O'Leary, J. C. Ehrhardt and W. T. C. Yuh, Automatic atlas-based volume estimation of human brain regions from MR images, *Journal of Computer Assisted Tomography* **20**, 1 (1996) 98–106.
5. E. Ardizzone, D. Peri, R. Pirrone, A. Palma and G. Peri, Knowledge-based approach to intelligent data analysis of medical images, in *Proc. of Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP'02)*, 2001.

6. N. Ayache, P. Cinquin, I. Cohen, L. Cohen, F. Leitner and O. Monga, Segmentation of complex three dimensional medical objects: A challenge and a requirement for computer-assisted surgery planning and performance, in *Computer Integrated Surgery: Technology and Clinical Applications*, 1996, pp. 59–74.
7. K. T. Bae, M. L. Giger, C. Chen and C. E. Kahn, Automatic segmentation of liver structure in CT images, *Medical Physics* **20**, 1 (1993) 71–78.
8. R. Bajcsy and S. Kovacic, Multiresolution elastic matching, *Computer Vision, Graphics, and Image Processing* **46** (1989) 1–21.
9. P. Baldi and S. Brunak, *Bioinformatics. The Machine Learning Approach* (MIT Press, 1998).
10. E. Bardinet, L. Cohen and N. Ayache, Tracking and motion analysis of the left ventricle with deformable superquadrics, *Medical Imaging Analysis* **1**, 2 (1996) 129–149.
11. V. Barra and J. Y. Boire, Automatic segmentation of subcortical brain structures in MR images using information fusion, *IEEE Transactions on Medical Imaging* **20**, 7 (2001) 549–558.
12. R. Bernard, B. Lika and F. Pernus, Segmenting articulated structures by hierarchical statistical modeling of shape, appearance, and topology, in *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI'01)*, 2001, pp. 499–506.
13. J. Besag, Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society* **36** (1974) 344–348.
14. J. C. Bezdek, L. O. Hall and L. P. Clarke, Review of MR image segmentation techniques using pattern recognition, *Medical Physics* **20** (1993) 1033–1048.
15. A. Blake and A. Zisserman, *Visual Reconstruction* (MIT Press, 1987).
16. R. E. Blanton, J. L. Levitt, P. M. Thompson, L. F. Capetillo-Cunliffe, T. Sadoun, T. Williams, J. T. McCracken and A. W. Toga, Mapping cortical variability and complexity patterns in the developing human brain, *Psychiatry Research* **107**, 1 (2001) 29–43.
17. T. O. Blum, *Transformation for Extracting New Descriptors of Shape. Models for the Perception of Speech and Visual Form* (MIT Press, 1967).
18. E. Borenstein, E. Sharon and S. Ullman, Combining top-down and bottom-up segmentation, in *Proc. of Computer Vision and Pattern Recognition (CVPR'04)*, 2004.
19. G. Borgefors, Distance transformations in arbitrary dimensions. *Computer Vision, Graphics, and Image Processing* **27** (1984) 321–345.
20. C. Brechbühler, *Description and analysis of 3D shapes by parameterization of closed surfaces*, PhD thesis, IKT/BIWI, ETH ETH Zürich, 1995.
21. M. Brejl and M. Sonka, Object localization and border detection criteria design in edge-based image segmentation: Automated learning from examples, *IEEE Transactions on Medical Imaging* **19**, 10 (2000) 973–985.
22. M. S. Brown, L. S. Wilson, B. D. Doust, R. W. Gill and C. Sun, Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images, *Computerized Medical Imaging and Graphics* **22** (1998) 463–477.
23. E. Bullmore, M. Brammer, G. Rouleau, B. Everitt and A. Simmons, Computerized brain tissue classification of magnetic resonance images: A new approach to the problem of partial volume artifact, *Neuroimage* **2** (1995) 133–147.
24. P. Cachier, E. Bardinet, D. Dormont, X. Pennec and N. Ayache, Iconic feature based nonrigid registration: The PASHA algorithm, *CVIU — Special Issue on Nonrigid Registration* **89**, 2–3 (February–March 2003) 272–298.

25. 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.
26. C. H. Chen and G. G. Lee, On digital mammogram segmentation and microcalcification detection using multiresolution wavelet analysis, *Graphical Models and Image Processing* **59**, 5 (1997) 349–364.
27. Y. Chen, H. D. Tagare, M. Rado, D. Wilson and E. A. Geiser, Using prior shape and intensity profile in medical image segmentation, in *Proc. of International Conference on Computer Vision (ICCV'03)*, 2003, pp. 1117–11123.
28. S. M. Choi, J. E. Lee, J. Kim and M. H. Kim, Volumetric objection reconstruction using the 3D-MRF model-based segmentation, *IEEE Transactions on Medical Imaging* **16** (1997) 887–892.
29. G. E. Christensen, S. C. Joshi and M. I. Miller, Volumetric transformation of brain anatomy, *IEEE Transactions on Medical Imaging* **16** (1997) 864–877.
30. M. C. Clark, L. O. Hall, D. B. Goldgof, L. P. Clarke, R. P. Velthuizen and M. S. Silbiger, MRI segmentation using fuzzy clustering techniques, *IEEE Engineering in Medicine and Biology Magazine*, 1994, pp. 730–742.
31. M. C. Clark, L. O. Hall, D. B. Goldgof, R. Velthuizen and F. R. Murtagh, Automatic tumor segmentation using knowledge-based techniques, *IEEE Transactions on Medical Imaging*, 1998, pp. 187–201.
32. L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine and M. Vaidyanathan, MRI segmentation: Methods and applications, *Magnetic Resonance Imaging* **12**, 3 (1995) 343–368.
33. I. Cohen, N. Ayache and P. Sulget, Tracking points on deformable objects using curvature information, in *Proc. of European Conference on Computer Vision (ECCV'92)*, 1992, pp. 458–466.
34. I. Cohen and L. D. Cohen, Hyperquadric model for 2D and 3D data fitting, in *Proc. of International Conference on Pattern Recognition (ICPR'94)*, 1994, pp. 403–405.
35. L. D. Cohen and I. Cohen, Finite element methods for active contour models and balloons for 2D and 3D images, *Transactions on Pattern Analysis and Machine Intelligence* **15**, 11 (1993) 1131–1147.
36. D. Collins, C. Holmes, T. Peters and A. Evans, Automatic 3D model-based neuroanatomical segmentation, *Human Brain Mapping* **3**, 3 (1995) 190–208.
37. D. Collins, A. Zijdenbos, V. Kollokian, J. Sled, N. Kabani, C. Holmes and A. Evans, Design and construction of a realistic digital brain phantom, *IEEE Transactions on Medical Imaging* **26**, 3 (1998) 463–468.
38. D. L. Collins, A. P. Zijdenbos, T. Paus and A. C. Evans, Use of registration for cohort studies, eds. J. Hajnal, D. Hawkes and D. Hill, *Medical Image Registration* (Kluwer, 2003).
39. T. F. Cootes, D. H. Cooper, C. J. Taylor and J. Graham, Trainable method of parametric shape description, *Image Vision Computing* **10**, 5 (1992) 289–294.
40. T. F. Cootes, G. J. Edwards and C. J. Taylor, Active appearance models, in *Proc. of European Conference on Computer Vision (ECCV'98)*, 1998, pp. 484–498.
41. T. F. Cootes, A. Hill, C. J. Taylor and J. Haslam, Use of active shape models for locating structures in medical images, *Image and Vision Computing* **12**, 6 (1994) 355–366.
42. J. Coughlan, A. Yuille, C. English and D. Snow, Efficient deformable template detection and localization without user initialization, *Computer Vision and Image Understanding* **78**, 3 (2000) 303–319.

43. T. M. Cover and J. M. Van Campenhout, On the possible orderings in the measurement selection problem, *IEEE Transactions on Systems Man and Cybernetics* **7** (1977) 657–661.
44. C. Davatzikos, Spatial normalization of 3D images using deformable models, *Journal of Computer Assisted Tomography* **20**, 4 (1996) 656–665.
45. C. Davatzikos and R. N. Bryan, Using a deformable surface model to obtain a shape representation of the cortex, *IEEE Transactions on Medical Imaging* **15**, 6 (1996) 785–795.
46. C. Davatzikos, J. Prince and N. Bryan, Image registration based on boundary mapping, *IEEE Transactions on Medical Imaging* **15**, 1 (1996) 212–215.
47. C. Davatzikos, M. Vaillant, S. Resnick, J. L. Prince, S. Letovsky and R. N. Bryan, Computerized method for morphological analysis of the corpus callosum, *Journal of Computer Assisted Tomography* **20** (1998) 88–97.
48. R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton and C. J. Taylor, A minimum description length approach to statistical shape modelling, *IEEE Transactions on Medical Imaging* **21**, 5 (2002) 525–537.
49. A. C. Davison, *Statistical Models* (Cambridge Press, 2003).
50. R. Deklerck, J. Cornelis and M. Bister, Segmentation of medical images, *Image and Vision Computing* **11**, 8 (1993) 486–503.
51. H. Delingette, General object reconstruction based on simplex meshes, *International Journal of Computer Vision* **32**, 2 (1999) 111–146.
52. H. Delingette and J. Montagnat, Shape and topology constraints on parametric active contours, *Computer Vision and Image Understanding* **83**, 2 (2001) 140–171.
53. L. R. Dice, Measures of the amount of ecologic association between species, *Ecology* **26**, 3 (1945) 297–302.
54. R. O. Duda, E. Hart and D. G. Stork, *Pattern Classification*, 2nd edn. (Wiley-Interscience, 2000).
55. J. S. Duncan, X. Papademetris, J. Yang, M. Jackowski, X. Zeng and L. H. Staib, Geometric strategies for neuroanatomic analysis from MRI, *Neuroimage*, 2004 (to appear).
56. M. Egmont-Petersen, D. de Ridder and H. Handels, Image processing with neural networks — A review, *Pattern Recognition* **35**, 10 (2002) 2279–2301.
57. B. Fischl, D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen and A. M. Dale, Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain, *Neuron* **33** (2002) 341–355.
58. M. Fischler and R. Elschlager, The representation and matching of pictorial structures, *IEEE Transactions on Computers* **22**, 1 (1973) 67–92.
59. M. Fleuté, S. Lavallée and R. Julliard, Incorporating a statistically based shape model into a system for computer-assisted anterior cruciate ligament surgery, *Medical Image Analysis* **3**, 3 (1999) 209–222.
60. A. S. Frangakis, A. Stoschek and R. Hegerl, Wavelet transform filtering and nonlinear anisotropic diffusion assessed for signal reconstruction performance on multidimensional biomedical data, *IEEE Transactions on Biomedical Engineering* **48**, 2 (2001) 213–222.
61. T. O. J. Fuchs, M. Kachelriess and W. A. Kalender, Fast volume scanning approaches by X-ray-computed tomography, *Proceedings of the IEEE* **91**, 10 (2003) 1492–1502.
62. J. Gao, A. Kosaka and A. Kak, A deformable model for human organ extraction, in *Proc. of International Conference on Image Processing (ICIP'98)* **3** (1998) 323–327.

63. Y. Ge, J. M. Fitzpatrick, B. M. Dawant, J. Bao, R. M. Kessler and R. Margolin, Accurate localization of cortical convolutions in MR brain images, *IEEE Transactions on Medical Imaging* **15** (1996) 418–428.
64. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** (1984) 721–741.
65. G. Gerig, M. Jomier and M. Chakos, Valmet: A new validation tool for assessing and improving 3D object segmentation, in *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI'01)*, 2001, pp. 516–528.
66. P. Gibbs, D. L. Buckley, S. J. Blackband and A. Horsman, Tumour volume detection from MR images by morphological segmentation, *Physics in Medicine and Biology* **41**, 11 (1996) 2437–2446.
67. P. Golland, W. E. L. Grimson, M. E. Shenton and R. Kikinis, Deformation analysis for shape based classification, in *Proc. of Information Processing in Medical Imaging (IPMI'01)*, 2001, pp. 517–530.
68. C. L. Gordon, C. E. Webber, J. D. Adachi and N. Christoforou, *In vivo* assessment of trabecular bone structure at the distal radius from high-resolution computed tomography images, *Physics in Medicine and Biology* **41** (1996) 495–508.
69. J. Green, *Medical Image Processing: The Mathematics of Medical Imaging* (Greenwood Research, 2001).
70. R. Gupta and P. E. Undrill, The use of texture analysis to delineate suspicious masses in mammograhy, *Physics in Medicine and Biology* **40**, 5 (1995) 835–855.
71. L. O. Hall, A. M. Bensaid and L. P. Clarke, A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain, *Neural Networks* **3**, 5 (1992) 672–682.
72. S. M. Haney, P. M. Thompson, T. F. Cloughesy, J. R. Alger and A. W. Toga, Tracking tumor growth rates in patients with malignant gliomas: A test of two algorithms, *American Journal of Neuroradiology* **22**, 1 (2001) 73–82.
73. R. M. Haralick and L. G. Shapiro, Image segmentation techniques, *Computer Vision, Graphics and Image Processing* **29** (1985) 100–132.
74. T. J. Hebert, Fast iterative segmentation of high resolution medical images, *IEEE Transactions on Nuclear Science* **44** (1997) 1363–1367.
75. K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis and H. Müller-Gärtner, Markov random field segmentation of brain MR images, *IEEE Transactions on Medical Imaging* **16**, 6 (1997).
76. R. C. Herndon, J. L. Lancaster, A. W. Toga and P. T. Fox, Quantification of white matter and gray matter volumes from T1 parametric images using fuzzy classifiers, *Journal of Magnetic Resonance Imaging* **6** (1996) 425–435.
77. S. Ho, E. Bullitt and G. Gerig, Level set evolution with region competition: Automatic 3D segmentation of brain tumors, in *Proc. of International Conference on Pattern Recognition (ICPR'02)*, 2002, pp. 532–535.
78. L. T. Holly and K. T. Foley, Intraoperative spinal navigation, *Spine* **28**, 15 (2003) 554–561.
79. L. Holmstrom, P. Koistinen, J. Laaksonen and E. Oja, Neural and statistical classifiers-taxonomy and two case studies, *IEEE Transactions on Neural Networks* **8**, 1 (1997) 5–17.
80. J. Hug, C. Brechbühler and G. Székely, Model-based initialisation for segmentation, in *Proceedings of European Conference on Computer Vision (ECCV'00)*, 2000.

81. N. Ibrahim, H. Fujita, T. Hara and T. Endo, Automated detection of clustered microcalcifications on mammograms: CAD system application to MIAS database, *Physics in Medicine and Biology* **42** (1997) 2577–2589.
82. P. Jaccard, The distribution of flora in the alpine zone, *New Phytologist* **11** (1912) 37–50.
83. A. K. Jain and B. Chandrasekaran, Dimensionality and sample size considerations, eds. P. R. Krishnaiah and L. N. Kanal, *Pattern Recognition Practice* **2** (1982) 835–855.
84. A. K. Jain and D. Zongker, Feature selection: Evaluation, application, and small sample performance, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, 2 (1997) 153–158.
85. S. Joshi, M. Miller and U. Grenander, On the geometry and shape of brain submanifolds, *Pattern Recognition and Artificial Intelligence* **11** (1997) 1317–1343.
86. S. Joshi, S. Pizer, P. T. Fletcher, A. Thall and G. Tracton, Multi-scale 3D deformable model segmentation based on medial description, in *Proc. of Information Processing in Medical Imaging (IPMI'01)*, 2001, pp. 64–77.
87. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman and A. Y. Wu, An efficient k-means clustering algorithm: Analysis and implementation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**, 7 (2002) 881–892.
88. M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active contour models, *International Journal of Computer Vision* **1**, 4 (1988) 321–331.
89. A. Kelemen, G. Székely and G. Gerig, Three-dimensional model-based segmentation of brain MRI, *IEEE Transactions on Medical Imaging* **18**, 10 (1999) 838–849.
90. V. S. Khoo, D. P. Dearnaley, D. J. Finnigan, A. Padhani, S. F. Tanner and M. O. Leach, Magnetic resonance imaging (MRI): Considerations and applications in radiotherapy treatment planning, *Radiotherapy and Oncology* **42**, 1 (1997) 1–15.
91. R. Kruse, J. Gebhart and F. Klawonn, *Foundations of Fuzzy Systems* (Wiley, 1994).
92. M. A. Kupinski and M. L. Giger, Automated seeded lesion segmentation on digital mammograms, *IEEE Transactions on Medical Imaging* **17** (1998) 510–517.
93. J. L. Lancaster, L. H. Rainey, J. L. Summerlin, C. S. Freitas, P. T. Fox, A. C. Evans, A. W. Toga and J. C. Mazziotta, Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method, *Human Brain Mapping* **5** (1997) 238–242.
94. S. M. Lawrie and S. S. Abukmeil, Brain abnormality in schizophrenia: A systematic and quantitative review of volumetric magnetic resonance image studies, *Journal of Psychiatry* **172** (1998) 110–120.
95. C. Lee, S. Hun, T. A. Ketter and M. Unser, Unsupervised connectivity-based thresholding segmentation of midsagittal brain MR images, *Computer Biology Medicine* **28** (1998) 309–338.
96. S. U. Lee, S. Y. Chung and R. H. Park, A comparative performance study of several global thresholding techniques for segmentation, *Computer Vision, Graphics, and Image Processing* **52** (1990) 171–190.
97. K. Van Leemput, F. Maes, D. Vandermeulen and P. Suetens, Automated model-based tissue classification of MR images of the brain, *IEEE Transactions on Medical Imaging* **18**, 10 (1999) 897–908.
98. T. M. Lehmann, C. Gonner and K. Spitzer, Survey: Interpolation methods in medical image processing, *IEEE Transactions on Medical Imaging* **18**, 11 (1999) 1049–1075.
99. J. M. Leski, Generalized weighted conditional fuzzy clustering, *IEEE Transactions on Fuzzy Systems* **11**, 6 (2003) 709–715.

100. M. Leventon, E. Grimson and O. Faugeras, Statistical shape influence in geodesic active contours, in *Proc. of Computer Vision and Pattern Recognition (CVPR'00)*, 2000, pp. 4–11.
101. H. Li, R. Deklerck, B. De Cuyper, A. Hermanus, E. Nyssen and J. Cornelis, Object recognition in brain CT-scans: Knowledge-based fusion of data from multiple feature extractors, *IEEE Transactions on Medical Imaging* **14**, 2 (1995) 212–229.
102. Z. Liang, Tissue classification and segmentation of MR images, *IEEE Engineering in Medicine and Biology Magazine* **12**, 1 (1993) 81–85.
103. Z. Liang, J. R. MacFall and D. P. Harrington, Parameter estimation and tissue segmentation from multispectral MR images, *IEEE Transactions on Medical Imaging* **13** (1994) 441–449.
104. P. Lipson, A. L. Yuille, D. O'Keefe, J. Cavanaugh, J. Taaffe and D. Rosenthal, Deformable templates for feature extraction from medical images, in *Proceedings of First European Conference on Computer Vision (ECCV'90)*, 1990, pp. 477–484.
105. J. Lotjonen, P.-J. Reissman, I. Magnin and T. Katila, Model extraction from magnetic resonance volume data using the deformable pyramid, *Medical Image Analysis* **3**, 4 (1999) 387–406.
106. Y. L. Lu, T. Z. Jiang and Y. F. Zang, Split-merge based region growing method for fMRI activation detection, *Human Brain Mapping*, 2004 (in press).
107. A. Macovski, *Medical Imaging Systems* (Prentice-Hall, 1938).
108. S. Malassiotis and M. G. Strintzis, Tracking the left ventricle in echocardiographic images by learning heart dynamics, *IEEE Transactions on Medical Imaging* **18**, 2 (1999) 282–290.
109. R. Malladi, J. Sethian and B. C. Vemuri, Shape modeling with front propagation: A level set approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17**, 2 (1995) 158–175.
110. R. Malladi, J. A. Sethian and B. C. Vemuri, A fast level set based algorithm for topology-independent shape modeling, *Journal of Mathematical Imaging and Vision* **6** (1996) 269–289.
111. J. F. Mangin, V. Frouin, I. Bloch, J. Regis and J. Lopez-Krahe, From 3D magnetic resonance images to structural representations of the cortex topography using topology preserving deformations, *Journal of Mathematical Imaging and Vision* **5** (1995) 297–318.
112. I. N. Manousakas, P. E. Undrill, G. G. Cameron and T. W. Redpath, Split-and-merge segmentation of magnetic resonance medical images: Performance evaluation and extension to three dimensions, *Computers and Biomedical Research* **31**, 6 (1998) 393–412.
113. K. V. Mardia and T. J. Hainsworth, A spatial thresholding method for image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**, 6 (1988) 919–927.
114. D. Marr and H. K. Nishihara, Visual information processing: Artificial intelligence and the sensorium of sight, *Technology Review* **81**, 1 (1978) 2–23.
115. M. Matesin, S. Loncaric and D. Petravic, A rule-based approach to stroke lesion analysis from CT brain images, in *Proc. of Second International Symposium on Image and Signal Processing and Analysis*, 2001, pp. 219–223.
116. W. Maurel, Y. Wu, N. Magnenat Thalmann and D. Thalmann, *Biomechanical Models for Soft Tissue Simulation* (Springer Verlag, 1998).
117. T. McInerney and D. Terzopoulos, Deformable models in medical image analysis: A survey, *Medical Image Analysis* **1**, 2 (1996) 91–108.

118. T. McInerney and D. Terzopoulos, T-snakes: Topology adaptive snakes, *Medical Image Analysis* **4** (2000) 73–91.
119. G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions* (John Wiley and Sons, 1997).
120. D. Metaxas and D. Terzopoulos, Shape and nonrigid motion estimation through physics-based synthesis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**, 6 (1993) 580–591.
121. I. Mikic, S. Krucinski and J. D. Thomas, Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates, *IEEE Transactions on Medical Imaging* **17** (1998) 274–284.
122. J. V. Miller, D. E. Breen, W. E. Lorensen, R. N. O'Bara and M. J. Wozny, Geometrically deformed models: A method for extracting closed geometric models from volume data, in *Proc. of SIGGRAPH'91*, 1991, pp. 217–226.
123. J. Montagnat and H. Delingette, Globally constrained deformable models for 3D object reconstruction, *Signal Processing* **71**, 2 (1998) 173–186.
124. J. Montagnat, H. Delingette and N. Ayache, A review of deformable surfaces: Topology, geometry and deformation, *Image and Vision Computing* **19** (2001) 1023–1040.
125. J. Montagnat, H. Delingette, N. Ayache, J. M. Clément, C. Roy, Y. Russier, V. Tassetti and J. Marescaux, Liver segmentation in contrast enhanced helical ct-scans, in *Proc. of World Congress on Medical Physics and Biomedical Engineering*, 1997.
126. Y. Nakagawa and A. Rosenfeld, Some experiments on variable thresholding, *Pattern Recognition* **11** (1979) 191–204.
127. K. L. Narr, P. M. Thompson, T. Sharma, J. Moussai, A. F. Cannestra and A. W. Toga, Mapping morphology of the corpus callosum in schizophrenia, *Cerebral Cortex* **10**, 1 (2000) 40–49.
128. W. J. Niessen, K. L. Vincken, J. Weickert, B. M. Ter Haar Romeny and M. A. Viergever, Multiscale segmentation of three-dimensional MR brain images, *International Journal of Computer Vision* **31**, 2 (1999) 185–202.
129. A. Obenaus, C. J. Yong-Hing, K. A. Tong and G. E. Sarty, A reliable method for measurement and normalization of pediatric hippocampal volumes, *Pediatric Research* **50** (2001) 124–132.
130. S. Osher and J. A. Sethian, Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton–Jacobi formulations, *Journal of Computational Physics* **79** (1988) 12–49.
131. B. O'Sullivan and J. Shah, New TNM staging criteria for head and neck tumors, *Semin. Surf. Oncol.* **21**, 1 (2003) 30–42.
132. X. Papademetris, A. J. Sinusas, D. P. Dione, R. T. Constable and J. S. Duncan, Estimation of 3D left ventricular deformation from medical images using biomechanical models, *IEEE Transactions on Medical Imaging* **21**, 7 (2002) 786–800.
133. J. M. Pardo, D. Cabello and J. Heras, Automatic 3D shape reconstruction of bones, in *Proc. of Engineering in Medicine and Biology Society*, 1995, pp. 387–388.
134. J. M. Pardo, D. Cabello and J. Heras, A snake for model-based segmentation of biomedical images, *Pattern Recognition Letters* **18** (1997) 1529–1538.
135. J. M. Pardo-Lopez, D. Cabello, J. Heras and J. Couceiro, A Markov random field model for bony tissue classification, *Computerized Medical Imaging and Graphics* **22** (1998) 169–178.

136. J. Park, D. Metaxas and L. Axel, Analysis of left ventricular wall motion based on volumetric deformable models and MRI-SPAMM, *Medical Image Analysis* **1**, 1 (1996).
137. D. L. Pham and J. L. Prince, An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensity inhomogeneities, *Pattern Recognition Letters* 1999, pp. 57–68.
138. D. L. Pham, C. Xu and J. L. Prince, Current methods in medical image segmentation, *Annual Review of Biomedical Engineering, Annual Reviews* **2** (1996) 315–337.
139. A. Pitiot, Automated segmentation of cerebral structures incorporating explicit knowledge, PhD thesis, Ecole des Mines de Paris, 2003.
140. A. Pitiot, H. Delingette, P. M. Thompson and N. Ayache, Expert knowledge guided segmentation system for brain MRI, *Neuroimage*, 2004 (to appear).
141. A. Pitiot, H. Delingette, A. Toga and P. Thompson, Learning object correspondences with the observed transport shape measure, in *Proc. of Information Processing in Medical Imaging IPMI'03*, 2003.
142. A. Pitiot, A. W. Toga and P. M. Thompson, Adaptive elastic segmentation of brain MRI via shape-model-guided evolutionary programming, *IEEE Transactions on Medical Imaging* **21**, 8 (2002) 910–923.
143. S. M. Pizer, D. S. Fritsch, P. Yushkevich, V. Johnson and E. L. Chaney, Segmentation, registration, and measurement of shape variation via image object shape, *IEEE Transactions on Medical Imaging* **10**, 18 (1999) 851–865.
144. S. M. Pizer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, E. L. Chaney, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash and J. M. Glotzer, Deformable M-reps for 3D medical image segmentation, *International Journal of Computer Vision* **55**, 2 (2003).
145. T. Poggio and V. Torra, Ill-posed problems and regularization analysis in early vision, in *Proceedings of AARPA Image Understanding Workshop*, 1984, pp. 257–263.
146. S. Pohlman, K. A. Powell, N. A. Obuchowski, W. A. Chilcote and S. Grundfest-Broniatowski, Quantitative classification of breast tumors in digitized mammograms, *Medical Physics* **23** (1996) 1337–1345.
147. W. E. Polakowski, D. A. Cournoyer, S. K. Rogers, M. P. DeSimio, D. W. Ruck, J. W. Hoffmeister and R. A. Raines, Computer-aided breast cancer detection and diagnosis of masses using difference of Gaussians and derivative-based feature saliency, *IEEE Transactions on Medical Imaging* **16** (1997) 811–819.
148. C. S. Poon, M. Braun, R. Fahrig, A. Ginige and A. Dorrell, Segmentation of medical images using an active contour model incorporating region-based image features, ed. R. A. Robb, *Visualisation in Biomedical Computing*, 1994, pp. 90–97.
149. F. Poupon, J.-F. Mangin, D. Hasboun and V. Frouin, Multi-object deformable templates dedicated to the segmentation of brain deep structures, in *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, 1998, pp. 1134–1143.
150. C. E. Priebe, D. J. Marchette and G. W. Rogers, Segmentation of random fields via borrowed strength density estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19** (1997) 494–499.
151. J. L. Prince, Q. Tan and D. Pham, Optimization of MR pulse sequences for Bayesian image segmentation, *Medical Physics* **22** (1995) 1651–1656.
152. J. Rademacher, A. M. Galaburda, D. N. Kennedy, P. A. Filipek and V. S. Caviness, Human cerebral cortex: Localization, parcellation and morphometry with magnetic resonance imaging, *Journal of Cognitive Neuroscience* **4** (1992) 352–374.

153. J. C. Rajapakse, J. N. Giedd and J. L. Rapoport, Statistical approach to segmentation of single-channel cerebral MR images, *IEEE Transactions on Medical Imaging* **16** (1997) 176–186.
154. R. A. Rauch and J. R. Jinkins, Variability of corpus callosal area measurements from midsagittal MR images: Effect of subject placement within the scanner, *American Journal of Neuroradiology* **17**, 1 (1996) 27–28.
155. W. E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin and R. J. Deaton, Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks, *IEEE Transactions on Medical Imaging* **16** (1997) 911–918.
156. A. L. Reiss, J. G. Hennessey, M. Rubin, L. Beach, M. T. Abrams, I. Warsofsky, M.-C. Liu and J. M. Links, Reliability and validity of an algorithm for fuzzy tissue segmentation of MRI, *Journal of Computer Assisted Tomography* **22** (1998) 471–479.
157. D. Rey, G. Subsol, H. Delingette and N. Ayache, Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis, *Medical Image Analysis* **6**, 2 (2002) 163–179.
158. M. R. Rezaee, P. M. J. van der Zwet, B. P. E. Lelieveldt, R. J. van der Geest and J. H. C. Reiber, A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering, *IEEE Transactions on Image Processing* **9**, 7 (2000) 1238–1248.
159. D. Rivière, J.-F. Mangin, D. Papadopoulos-Orfanos, J.-M. Martinez, V. Frouin and J. Régis, Automatic recognition of cortical sulci of the human brain using a congregation of neural networks, *Medical Imaging Analysis* **6**, 2 (2002) 77–92.
160. P. K. Sahoo, S. Soltani and A. K. C. Wong, A survey of thresholding techniques, *Computer Vision, Graphics and Imaging Processing* **41** (1988) 233–260.
161. S. Sandor and R. Leahy, Towards automatic labelling of the cerebral cortex using a deformable model, in *Proc. of Information Processing in Medical Imaging (IPMI'95)*, 1995.
162. S. Sandor and R. Leahy, Surface-based labeling of cortical anatomy using a deformable atlas, *IEEE Transactions on Medical Imaging* **16** (1997) 41–54.
163. B. Sankur and M. Sezgin, A survey over image thresholding techniques and quantitative performance evaluation, *Journal of Electronic Imaging* **13**, 1 (2004) 146–165.
164. G. Sapiro, R. Kimmel and V. Caselles, Object detection and measurements in medical images via geodesic deformable contours, *Vision Geometry IV*, 1995, pp. 366–378.
165. A. Sebbahi, A. Herment, A. deCesare and E. Mousseaux, Multimodality cardiovascular image segmentation using a deformable contour model, *Computerized Medical Imaging and Graphics* **21** (1997) 79–89.
166. M. Sermesant, C. Forest, X. Pennec, H. Delingette and N. Ayache, Deformable biomechanical models: Application to 4d cardiac image analysis, *Medical Image Analysis* **7**, 4 (December 2003) 475–488.
167. J. A. Sethian, *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Sciences* (Cambridge University Press, 1996).
168. D. Shen, E. H. Herskovits and C. Davatzikos, An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures, *IEEE Transactions on Medical Imaging* **20**, 4 (2001) 257–270.
169. A. Simmons, P. S. Tofts, G. J. Barker and S. R. Arridge, Sources of intensity nonuniformity in spin echo images at 1.5T, *Magnetic Resonance Medicine* **32** (1994) 121–128.

170. H. R. Singleton and G. M. Pohost, Automatic cardiac MR image segmentation using edge detection by tissue classification in pixel neighborhoods, *Magnetic Resonance Medicine* **37** (1997) 418–424.
171. J. G. Sled and G. B. Pike, Standing-wave and RF penetration artifacts caused by elliptic geometry: An electrodynamic analysis of MRI, *IEEE Transactions on Medical Imaging* **17** (1998) 653–662.
172. S. M. Smith, N. De Stefano, M. Jenkinson and P. M. Matthews, Normalised accurate measurement of longitudinal brain change, *Journal of Computer Assisted Tomography* **25**, 3 (2001) 466–475.
173. H. Soltanian-Zadeh, R. Saigal, J. P. Waindham, A. E. Yagle and D. O. Hearshen, Optimization of MRI protocols and pulse sequence parameters for eigenimage filtering, *IEEE Transactions on Medical Imaging* **13**, 1 (1994) 161–175.
174. P. Sprawls, *Physical Principles of Medical Imaging* (Aspen Publishers, 1993).
175. L. H. Staib and J. S. Duncan, Boundary finding with parametrically deformable models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**, 11 (1992) 1061–1075.
176. J. Stoeckel, G. Malandain, O. Migneco, P. M. Koulibaly, P. Robert, N. Ayache and J. Darcourt, Classification of SPECT images of normal subjects versus images of Alzheimer's disease patients, in *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI'01)*, 2001, pp. 666–674.
177. C. Studholme, D. L. G. Hill and D. J. Hawkes, Incorporating connected region labelling into automated image registration using mutual information, in *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'96)* (IEEE Computer Society Press, 1996), pp. 23–31.
178. M. Styner, C. Brechbuhler, G. Székely and G. Gerig, Parametric estimate of intensity inhomogeneities applied to MRI, *IEEE Transactions on Medical Imaging* **19** (2000) 152–165.
179. M. Styner and G. Gerig, Medial models incorporating object variability for 3D shape analysis, in *Proceedings of Information Processing in Medical Imaging (IPMI'01)*, 2001, pp. 502–516.
180. J. S. Suri, L. Kecheng, S. Singh, S. N. Laxminarayan, Z. Xialoan and L. Reden, Shape recovery algorithms using level sets in 2-D/3-D medical imagery: A state-of-the-art review, *IEEE Transactions on Information Technology in Biomedicine* **6**, 1 (2002) 8–28.
181. R. Szeliski and S. Lavallee, Matching 3D anatomical surfaces with non-rigid deformations using octree-splines, *International Journal of Computer Vision* **18**, 2 (1996) 171–186.
182. G. Székely, A. Kelemen, C. Brechbühler and G. Gerig, Segmentation of 2D and 3D objects from MRI volume data using constrained elastic deformations of flexible Fourier surface models, *Medical Image Analysis* **1**, 1 (1996) 19–34.
183. J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System: An Approach to Cerebral Imaging* (Thieme Medical Publisher, Stuttgart, 1988).
184. G. Taubin, C. Fernando, S. Sullivan, J. Ponce and D. J. Kriegman, Parameterized families of polynomials for bounded algebraic curve and surface fitting, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**, 3 (1994) 287–303.
185. P. Taylor, Computer aids for decision-making in diagnostic radiology — A literature review, *British Journal of Radiology* **68** (1995) 945–957.
186. H. Tek and B. B. Kimia, Volumetric segmentation of medical images by three-dimensional bubbles, *Computer Vision and Image Understanding* **65** (1997) 246–258.

187. D. Terzopoulos and K. Fleischer, Deformable models, *The Visual Computer* **4**, 6 (1988) 306–331.
188. P. M. Thompson, M. S. Mega, R. P. Woods, C. I. Zoumalan, C. J. Lindshield, R. E. Blanton, J. Moussai, C. J. Holmes, J. L. Cummings and A. W. Toga, Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas, *Cerebral Cortex* **11**, 1 (2001) 1–16.
189. P. M. Thompson, J. Moussai, S. Zohoori, A. Goldkorn, A. A Khan, M. S. Mega, G. W. Small, J. L. Cummings and A. W. Toga, Cortical variability and asymmetry in normal aging and Alzheimer's disease, *Cerebral Cortex* **8**, 6 (1998) 492–509.
190. P. M. Thompson, K. L. Narr, R. E. Blanton and A. W. Toga, Mapping structural alterations of the corpus callosum during brain development and degeneration, eds. E. Zaidel and M. Iacoboni, *The Corpus Callosum* (MIT Press, 2003).
191. P. M. Thompson, C. Schwartz and A. W. Toga, High-resolution random mesh algorithms for creating a probabilistic 3D surface atlas of the human brain, *Neuroimage* **3**, 1 (1996) 19–34.
192. P. M. Thompson and A. W. Toga, Detection, visualisation and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations, *Medical Image Analysis* **1**, 4 (1997) 271–294.
193. P. M. Thompson, R. P. Woods, M. S. Mega and A. W. Toga, Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain, *Human Brain Mapping* **9**, 2 (2000) 81–92.
194. A. W. Toga, *Brain Warping* (Academic Press, 1998).
195. D. Tosun, M. E. Rettmann, X. Han, X. Tao, C. Xu, S. M. Resnick, D. Pham and J. L. Prince, Cortical surface segmentation and mapping, *Neuroimage*, 2004 (to appear).
196. A. Trouvé and L. Younes, Diffeomorphic matching problems in one dimension: Designing and minimizing matching functionals, in *Proc. of European Conference on Computer Vision (ECCV'00)*, 2000, pp. 573–587.
197. A. Tsai, W. Wells, C. Tempany, E. Grimson and A. Willsky, Coupled multi-shape model and mutual information for medical image segmentation, in *Proceedings of Information Processing in Medical Imaging (IPMI'03)*, 2003, pp. 185–197.
198. M. Turk and A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience* **3**, 1 (1991) 71–86.
199. J. K. Udupa and P. K. Saha, Fuzzy connectedness and image segmentation, *Proceedings of the IEEE* **91**, 10 (2003) 1649–1669.
200. D. Vandermeulen, R. Verbeeck, L. Berben, D. Delaere, P. Suetens and G. Marchal, Continuous voxel classification by stochastic relaxation: Theory and application to MR imaging and MR angiography, *Image and Vision Computing* **12** (1994) 559–572.
201. M. Vasilescu and D. Terzopoulos, Adaptive meshes and shells: Irregular triangulation, discontinuities and hierarchical subdivision, in *Proc. of Computer Vision and Pattern Recognition (CVPR'92)*, 1992, pp. 829–832.
202. B. C. Vemuri, Y. Guo, C. M. Leonard and S.-H. Lai, Fast numerical algorithms for fitting multiresolution hybrid shape models to brain MRI, *Medical Image Analysis* **1**, 1 (1997) 343–362.
203. B. C. Vemuri, A. Radisavljevic and C. Leonard, Multiresolution 3D stochastic hybrid shape models for image segmentation, in *Proceedings of Information Processing in Medical Imaging (IPMI'93)*, 1993, pp. 62–76.
204. D. L. Vilarino, V. M. Brea, D. Cabello and J. M. Pardo, Discrete-time CNN for image segmentation by active contours, *Pattern Recognition Letters* **19** (1998) 721–734.

205. N. F. Vittitoe, R. Vargas-Voracek and C. E. Floyd, Identification of lung regions in chest radiographs using Markov random field modeling, *Medical Physics* **25** (1998) 976–985.
206. S.-Y. Wan and W. E. Higgins, Symmetric region growing, *IEEE Transactions on Image Processing* **12**, 9 (2003) 1007–1015.
207. S.-Y. Wan and E. Nung, Seed-invariant region growing: Its properties and applications to 3D medical CT images, in *Proc. of International Conference on Image Processing (ICIP'01)*, 2001.
208. Y. Wang, T. Adali and S. Y. Kung, Quantification and segmentation of brain tissues from MR images — A probabilistic neural network approach, *IEEE Transactions on Image Processing* **7**, 8 (1998) 1165–1181.
209. Y. Wang, B. S. Peterson and L. H. Staib, Shape-based 3D surface correspondence using geodesics and local geometry, in *Proc. of Computer Vision and Pattern Recognition (CVPR'00)*, 2000, pp. 644–651.
210. S. K. Warfield, F. A. Jolesz and R. Kikinis, Real-time image segmentation for image-guided surgery, in *Proc. of Conference on High Performance Networking and Computing*, 1998, p. 42.
211. S. K. Warfield, K. H. Zou and W. M. Wells, Validation of image segmentation and expert quality with an expectation-minimization algorithm, in *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI'02)*, 2002, pp. 298–306.
212. W. M. Wells, W. E. L. Grimson, R. Kikins and F. A. Jolesz, Adaptive segmentation of MRI data, *IEEE Transactions on Medical Imaging* **15** (1996) 429–442.
213. H. H. Wen, W. C. Lin and C. T. Chen, Knowledge-based medical image registration, in *Proceedings of the IEEE Engineering in Medicine and Biology Society*, 1996, pp. 1200–1201.
214. P. J. Werbos, *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks and Political Forecasting* (Wiley, 1994).
215. R. Whitaker, Volumetric deformable models, in *Proc. of Visualization in Biomedical Computing (VBC'94)*, 1994.
216. A. J. Worth and D. N. Kennedy, Segmentation of magnetic resonance brain images using analogue constraint satisfaction neural networks, *Image Vision Computing* **12**, 6 (1994) 345–354.
217. C. Xu and J. Prince, Snakes, shapes and gradient vector flow, *IEEE Transactions on Image Processing* **7**, 3 (1998) 359–369.
218. J. Yang, L. H. Staib and J. S. Duncan, Neighbor-constrained segmentation with 3D deformable models, in *Proceedings of Information Processing in Medical Imaging (IPMI'03)*, 2003, pp. 198–209.
219. A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver and A. Tannenbaum, A geometric snake model for segmentation of medical imagery, *IEEE Transactions on Medical Imaging* **16** (1997) 199–209.
220. Z. Yue, A. Goshtasby and L. V. Ackerman, Automatic detection of rib borders in chest radiographs, *IEEE Transactions on Medical Imaging* **14** (1995) 525–536.
221. A. L. Yuille, P. W. Hallinan and D. S. Cohen, Feature extraction from faces using deformable templates, *International Journal of Computer Vision* **8** (1992) 99–111.
222. Y. J. Zhang, A survey on evaluation methods for image segmentation, *Pattern Recognition Letters* **29** (1996) 1335–1346.
223. L. Zhukov, K. Museth, D. Breen, R. Whitaker and A. H. Barr, Level set modeling and segmentation of DT-MRI brain data, *Journal of Electronic Imaging* **12**, 1 (2003) 125–133.

224. A. P. Zijdenbos and B. M. Dawant, Brain segmentation and white matter lesion detection in MR images, *Critical Reviews in Biomedical Engineering* **22** (1994) 401–465.
225. A. P. Zijdenbos, B. M. Dawant, R. A. Margolin and A. C. Palmer, Morphometric analysis of white matter lesions in MR images: Method and validation, *IEEE Transactions on Medical Imaging* **13** (1994) 716–724.
226. Y. Zimmerand, R. Tepper and S. Akselrod, A two-dimensional extension of minimum cross entropy thresholding for the segmentation of ultrasound images, *Ultrasound in Medicine and Biology* **22** (1996) 1183–1190.
227. R. A. Zoroofi, Y. Sato, T. Sasama, T. Nishii, N. Sugano, K. Yonenobu, H. Yoshikawa, T. Ochi and S. Tamura, Automated segmentation of acetabulum and femoral head from 3D CT images, *IEEE Transactions on Information Technology in Biomedicine* **7**, 4 (2003) 329–343.
228. K. H. Zou, W. M. Wells, M. Kaus, R. Kikinis, F. A. Jolesz and S. K. Warfield, Statistical validation of automated probabilistic segmentation against composite latent expert ground truth in MR imaging of brain tumors, in *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI'02)*, 2002, pp. 315–322.

This page intentionally left blank

CHAPTER 7

BRAIN SHIFT ESTIMATION FOR IMAGE GUIDED SURGERY BASED ON AN INTEGRATED RANGE-SENSING, SEGMENTATION AND REGISTRATION FRAMEWORK

MICHEL A. AUDETTE
National Institute of Advanced Industrial Science & Technology-AIST
Namiki 1-2, Tsukuba, 305-8564, Japan
m.audette@aist.go.jp

KALEEM SIDDIQI* and FRANK P. FERRIE†
Center for Intelligent Machines, School of Computer Science
McGill University, 3480 University Street Montreal, PQ, Canada H3A 2A7
**siddiqi@cim.mcgill.ca*
†*ferrie@cim.mcgill.ca*

TERRY M. PETERS
Imaging Research Laboratories, Robarts Research Institute
University of Western Ontario, 100 Perth Drive
P. O. Box 5015, London, Ontario N6A 5K8, Canada
tpeters@imaging.robarts.ca

Image-guided surgery (IGS) is a technique for localizing anatomical structures on the basis of volumetric image data and for determining the optimal surgical path to reach these structures, by the means of a localization device, or probe, whose position is tracked over time. The usefulness of this technology hinges on the accuracy of the transformation between the image volume and the space occupied by the patient anatomy and spanned by the probe. Unfortunately, in neurosurgery this transformation can be degraded by intra-surgical brain shift, which often measures more than 10 mm and can exceed 25 mm. We propose a method for characterizing brain shift that is based on non-rigid surface registration, and can be combined with a constitutively realistic finite element approach for volumetric displacement estimation.

The proposed registration method integrates in a unified framework all of the stages required to estimate the movement of the cortical surface in the operating room: model-based segmentation of the pre-operative brain surface in magnetic resonance image data, range-sensing of the cortex in the OR, range-MR rigid transformation computation, and range-based non-rigid brain motion estimation. The brain segmentation technique is an adaptation of the surface evolution model. Its convergence to the brain boundary is the result of a speed term restricted to white and grey matter voxels made explicit by a classifier, and the final result is post-processed to yield a *Closest Point Map* of the brain surface in MR space. In turn, this Closest Point Map is used to produce the homologous pairs required to determine a highly efficient, 2D spline-based, Iterative Closest Point (ICP) non-rigid surface registration. The baseline for computing intra-operative brain displacement, as well as the initial starting point of the non-rigid ICP registration, is determined by a very good rigid range-MR transformation, produced by a simple

procedure for relating the range coordinate system to that of the probe, and ultimately to that of the MR volume.

Keywords: Image guidance; brain shift; level set models; non-rigid surface registration; range-sensing.

1. Introduction

1.1. *Preliminaries: Image-guided surgery*

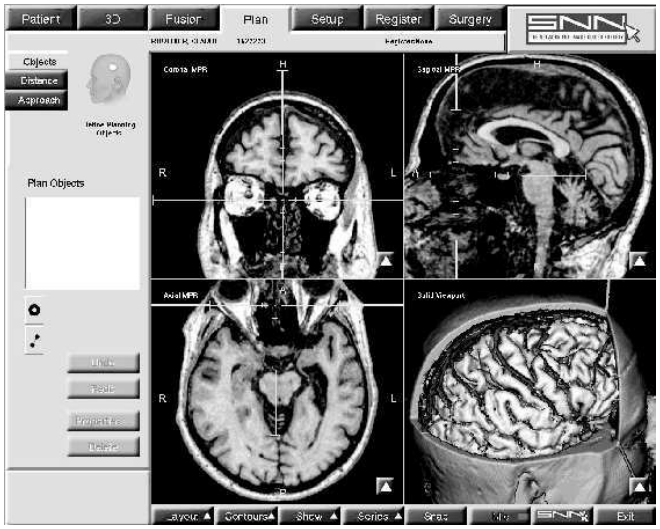
Image-guided surgery (IGS) describes a class of techniques for localizing anatomical structures on the basis of volumetric image data and for determining the optimal surgical path to reach these structures. With the advent of tomographic image volumes produced by computed tomography (CT) and magnetic resonance (MR), it has become possible to elaborate a 3D image model of the patient, whose internal coordinate system ("*image space*") can be related to that of a *localization device* spanning the physical space occupied by the patient ("*patient space*"). Brain IGS applications include the treatment of movement and convulsive disorders and of arteriovenous malformations as well as the biopsy, surgical resection, radiotherapy and brachytherapy of tumors (for a survey, see Ref. 48).

The image model typically consists of the tomographic volume itself, displayed in tri-planar view format, as illustrated in Fig. 1, and may include triangulated anatomical boundary surfaces identified from the tomographic scan and rendered onto a computer screen. The position of the point of interest within the 3D image model is conveyed on the basis of:

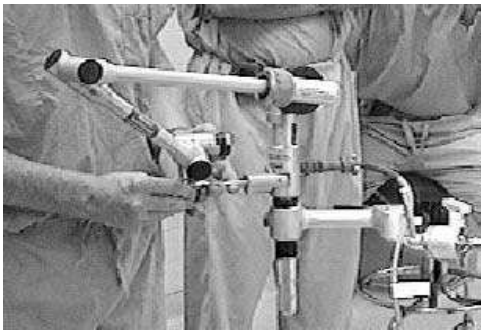
- the choice of tri-planar view of the underlying tomographic data and the position of the cursor overlaid on each plane as well as
- the position of a rendered virtual pointer overlaid at the appropriate position and orientation on the displayed anatomical surfaces.

There are two broad categories of localization technique, *frame-based*[21] and *frameless*,[27,44,61] with the latter increasingly more prevalent and illustrated in Figs. 1(b) and (c). Frameless localization is generally characterized by a handheld, tracked pointing device, consisting of a *probe* and perhaps a *probe-holder*, whose coordinates are calculated with respect to a reference object, which is generally affixed to a Mayfield clamp immobilizing the head. Assuming that the geometry relating the probe tip to the probe holder is known, and *following a registration procedure that maps image space to the physical space spanned by the tracked probe, the anatomical position of the probe tip is reflected by that of the 3D image cursor at all times*. Frameless localization also uses landmarks for determining the image-patient transformation which are extrinsic to the localization device; i.e. fiducials glued to the scalp or imbedded in the skull.

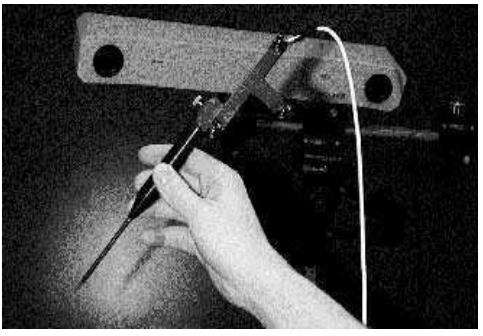
The usefulness of image-guided surgery hinges on the accuracy of the transformation between the image and patient spaces. This accuracy is dependent on



(a)



(b)



(c)

Fig. 1. Image-guided surgery: (a) 3D image model of patient as presented in typical commercial IGS system, namely in tri-planar and 3D view formats, courtesy of Lahbib Soualmi of the Montreal Neurological Institute; frameless localization devices: (b) the FARO arm, during probe calibration, and (c) the Northern Digital Polaris optical tracking system.

two conditions:

- first, an accurate initial registration, which in turn presumes accurate tracking equipment,[28,46] distortion-free MR/CT data[41] and accurate and consistent fiducial localization;[21,40,61] and
- second, a static relationship between patient and image model.

The most important source of error in IGS is the violation of the assumption of static anatomy, caused by intra-surgical brain shift. The amount of intra-surgical brain shift, as documented by numerous authors[38,45] often measures more than 10 mm and can exceed 25 mm, and can be abscribed to "the progressive sinking and displacement of the cortical surface as a result of cortico-spinal fluid drainage and progressive mesial displacement of the brain due to gravity".[45]

Existing methods for characterizing brain shift in general exploit either a combination of *sparse* cortical surface displacement data and biomechanical model,[39,42] or a combination of volumetric displacement information from tomographic data, namely intra-operative ultrasound[16] or MRI,[20] with some form of volumetric interpolation. Several authors have proposed realistic finite element (FE)[20,42] or mass-spring[52] models of the brain.

1.2. *A new approach to estimating brain shift*

In this chapter, we propose a method for characterizing brain shift that is based on non-rigid surface registration, and can be combined with a constitutively realistic finite element approach for volumetric displacement estimation.[20,42,52] As shown in Fig. 2, our registration method integrates in a unified framework all of the stages required to estimate the movement of the cortical surface in the operating room (OR):

- semi-automatic identification (segmentation) of relevant anatomical surfaces within the MRI volume;
- range-sensing of the exposed brain surface in the OR;
- computation of the range-MRI transformation, based on a calibration procedure and sensor base tracking; and
- non-rigid motion tracking, over time, of the range image of the brain.

Our philosophy for characterizing brain shift is based on the *fast* acquisition and processing of *dense* displacement information, using a sensor that could easily be deployed to, and retracted from, a scanning position, with an appropriate mounting framework. This perspective is dictated by the spatially and temporally underdetermined nature of non-rigid motion estimation. In addition, the semi-automatic surface identification stage presented here can be explicitly integrated with meshing software to make a finite element model *patient-specific* with little user interaction. Finally, a range-sensor is less expensive than both ultrasound and intra-operative MR, as well as easier to use.

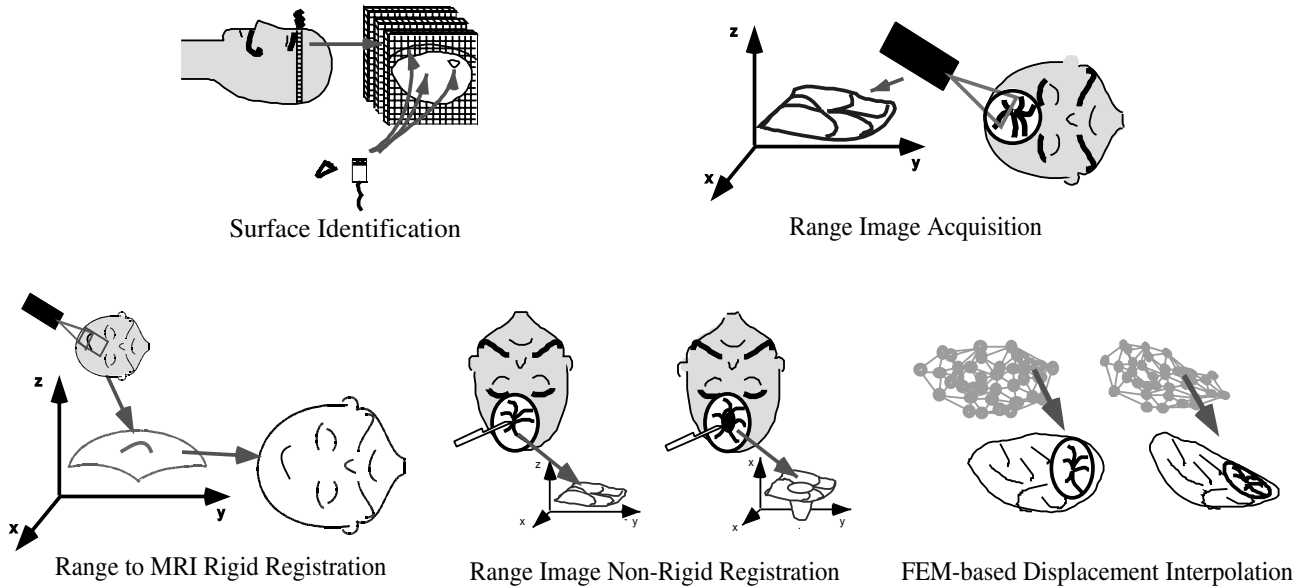


Fig. 2. Brain shift estimation framework, featuring the four stages comprising our non-rigid surface registration framework, which can be combined with a FE-based volumetric displacement estimation stage.

Our brain segmentation technique is an adaptation of surface evolution models.[12,36] Its convergence to the brain boundary is the result of a speed term restricted to white and gray matter voxels made explicit by a classifier,[18] and the final result is post-processed to yield a Closest Point Map of the brain surface in MR space. In turn, this Closest Point Map is used to produce the homologous pairs required to determine a highly efficient, 2D spline-based, Iterative Closest Point (ICP)[8,31] non-rigid surface registration. In particular, given an initial range-MR transformation, this Map associates with each MR voxel the label of its closest brain surface point, so that for any transformed cortical range point coinciding with that voxel, its closest MR homologue is immediately looked up. Finally, the baseline for computing intra-operative brain displacement, as well as the initial starting point of the non-rigid ICP registration, is determined by a very good rigid range-MR transformation, produced by a simple procedure for relating the range coordinate system to that of the probe, and ultimately to that of the MR volume. For a survey of anatomical surface registration techniques, the reader is referred to Ref. 4.

The rest of this chapter presents details of our method for computing the brain surface displacement as follows. Section 2.1 examines how laser-based range-sensing can provide 3D brain shape data in the operating room, as well as presents our procedure for computing the rigid transformations which relates the range coordinate system with the probe and MR coordinates systems. Next, the non-rigid registration of a time sequence of range images of the brain, to estimate the intra-surgical cortical surface displacement over time, is the subject of Sec. 2.3. Finally, the identification of the cortex in the MR volume is addressed in Sec. 2.4. This section discusses our contributions to making surface evolution models better adapted to identifying the brain surface and to post-processing the brain surface in order to improve the efficiency of the the subsequent registration with range data. Our validation study and a discussion of each stage is then presented in Sec. 3.

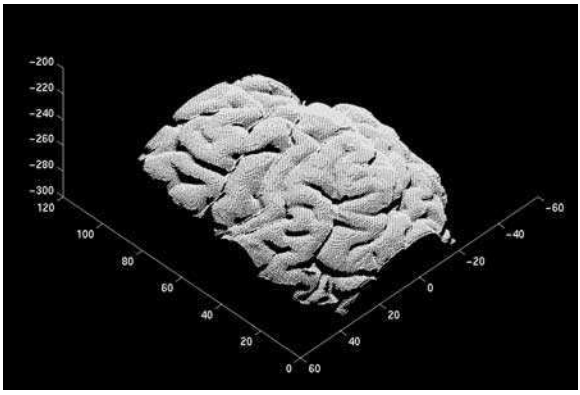
2. Materials and Methods

2.1. *Laser range-sensing and calibration-based range-MR transformation*

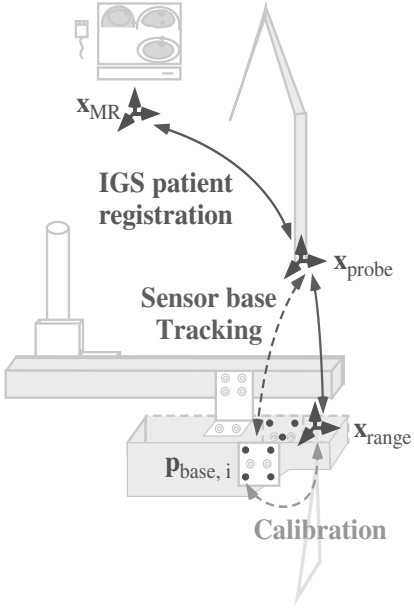
In order to characterize brain shift, one must first find a simple-to-use, robust means of measuring the 3D shape of the brain exposed in the OR. We have chosen to measure the external surface shape, rather than use a tomographic method, because of the ease of use of a surface tracking and FE approach. Moreover, tomographic studies have indicated that much of the brain movement is concentrated near the brain surface exposed by the craniotomy,[15,38] so that a good characterization of the volumetric displacement function is possible with a surface tracking and FEM approach. Even in the presence of deep structure motion, our assumption is that brain shift is not discontinuous over brain depth, that the combination of surface motion and of a realistic biomechanical model (in both constitutive properties and

boundary conditions) will accurately predict this motion, and that the effect of tissue resection can be assessed independently.[51]

Laser-based range-sensing is the industrial standard for quickly and accurately acquiring dense, quantitative information on 3D shape. Figure 3(a) illustrates the dense shape characterization of the cortical surface achievable with this technology. It is worth noting that other 3D surface shape estimation methods could have been



(a)



(b)

Fig. 3. (a) Range image of the left hemisphere of a human brain. (b) Range to MR transformation geometry, illustrating the range, probe and MR referentials and the sequence of procedures for relating them to each other.

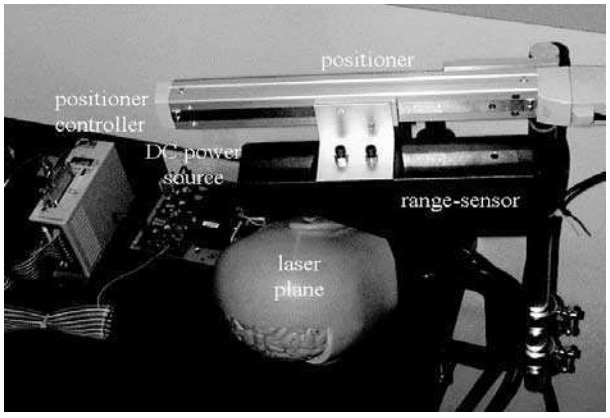
used, such as shape-from-stereo, particularly if integrated with a binocular surgical microscope. Moreover, there are two registration problems that are relevant to our method for characterizing brain shift: *relating range data to MR data*, and *tracking the non-rigid cortical motion over time* implicit in the time sequence of range images. The first problem could perhaps have been solved by computing the MR-range transformation on the basis of a fully autonomous registration of cranial skin surfaces produced by the two modalities. However, our technique, inspired by Comeau's ultrasound-to-MR registration method,[16] uses the tracked probe to establish this transformation, as illustrated in Fig. 3(b). It features a *calibration* stage that relates the internal range coordinate system to external plates, featuring milled divots, bolted to the range-sensor, a *sensor base tracking* stage that determines the range-probe transformation in the OR, and an *IGS patient registration* stage, whose MR-probe transformation is finally multiplied by the range-probe transformation to relate the range and MR coordinate systems.

2.2. *Laser range-sensing*

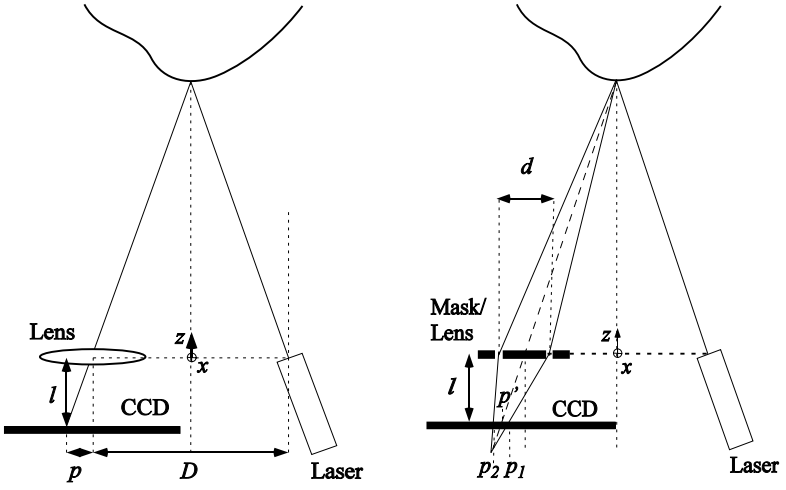
The three-dimensional coordinates of the exposed cortical surface are computed by a commercial range-sensor made by Vitana Corp.[57] (Ottawa, Canada), based on the Canadian National Research Council's Biris design.[9] This sensor uses both laser-based *triangulation* and *defocusing* techniques to estimate range. The sensor is mounted on a commercial linear positioner, produced by Intelligent Actuator.[23] The range-sensing configuration is featured in Fig. 4(a).

Laser-based triangulation involves projecting a thin ray of laser light onto an object, capturing the reflected light by a charge-coupled device (CCD), and inferring depth from the pixel position of the image point and from the geometric relationship between the position of the laser source and the CCD, as shown in Fig. 4(b). Moreover, the Biris sensor allows the laser light to pass through a mask featuring two apertures at a predetermined distance d apart, rather than one, as shown in Fig. 4(c). The sensor sees two laser spot images at p_1 and p_2, instead of one. Triangulation in this sensor is implemented by considering the *average position* of the two spots $p = \frac{p_1+p_2}{2}$. Moreover, defocusing allows the range value to be determined independently from the *space* $p_2 - p_1$ between the two laser spot images (for details, see Ref. 9). Finally, this sensor dispenses with one sweep axis by optically spreading the ray into a *laser plane*. The intersection of this plane with the imaged object instantaneously produces a 2D range-map of a "slice" of the surface, or *laser profile*, on the CCD. The final *range profile* coincides with depth z evaluated across all 256 CCD columns.

A 3D surface is then obtained by linearly sweeping the sensor in a direction normal to the laser plane, across the anatomy of interest, while taking range profiles at regular intervals. The final 3D surface is composed of 256 such range profiles. The coordinate y is given by the linear positioner, and it is orthogonal to the $x - z$ plane spanned by the range profile. Finally, the sensor possesses a depth of field



(a)



(b) (c)

Fig. 4. Biris laser range-sensing geometry: (a) range-sensor and positioner configuration in the laboratory, (b) triangulation; (c) defocusing.

of 135 mm and a field of view of 100–120 mm, which translates to a pixel accuracy of 0.4–0.5 mm along x and of 0.5 mm along z. The regular sampling nature of the CCD and the constant-rate data acquisition along the y-axis result in the collection of uniformly spaced samples.

2.2.1. *A procedure for relating range to MR space*

The baseline for cortical movement is implemented by computing the rigid-body transformation between range and MRI spaces and applying this transformation to

the identified cortex surface in the scan. In the literature, computing an arbitrarily large transformation between two surfaces is generally approached with features based on the differential geometry of the surfaces,[22,55] or global descriptions[11,25] of surface shape. However, an autonomous registration procedure would have to rely on a small skin patch, surrounded by surgical draping, and neither kind of approach would likely be sufficiently discriminating. Our technique uses the IGS tracked probe to establish the position in probe space, and ultimately in MR space, of the *sensor base*, coinciding with a configuration of milled holes in two external plates bolted to the sensor. All that is needed to complete the picture is a *calibration* procedure to relate the sensor base to the sensor internal reference. An important point to be made here is that a very accurate range-MR transformation is achievable from the automatic detection of implanted fiducials,[40,58] but we chose to proceed otherwise because of the practice at our institute of not using implantable fiducials.

The geometry of the calibration, sensor base tracking and IGS registration is seen in greater detail in Figs. 5 and 6. The objective of the overall procedure is to find a way to relate range coordinates to MRI coordinates. The goal of the calibration, which can take place outside the OR and is illustrated in Figs. 5(a), (b) and 6, is to relate the inner range-sensor reference to the milled divots on its outer side-plates, or sensor base. The goal of the sensor base tracking, as shown in Fig. 5(c), is to establish the position of these sensor base points in probe coordinates at time t_o in the OR, and use this information along with the calibration results to relate the inner range-sensor reference to probe space. The probe-image transformation, produced by the IGS patient registration, can then be used to relate range to MR coordinates, as featured in Fig. 5(d). To simplify the formalism adopted for this section, we use the letters R, P and M to designate *range*, *probe* and *MR* coordinates, while t_c and t_o correspond to time instants coinciding with *calibration* and the intervention in the *OR*.

The calibration procedure (refer to Figs. 5 and 6) consists of the following steps, assuming that the sensor is brought to *home position*, $y = 0$:

(1) *Linear positioner calibration.* The y-coordinate is calibrated with a long plate featuring a row of machined divots at a 10 mm $\pm$.01 mm spacing.
(2) *Range-sensing of a "wedding cake" non-coplanar calibration tool*, featuring an array of milled holes of 2 mm radius.
(3) *Automatic identification of hemispherical pits* $\mathbf{p}_{T,j}^{R,t_c}$ *within the range image of sensor calibration tool* by *template matching.*[47]
(4) *Probe-space identification of pits* in the tool, $\mathbf{p}_{T,j}^{P,t_c}$, by insertion of the probe tip.
(5) *Registration of range pits* $\mathbf{p}_{T,j}^{R,t_c}$ *with probe-space coordinates of the pits* $\mathbf{p}_{T,j}^{P,t_c}$, producing a probe-range transformation $\mathbf{T}_{P,t_c}^{R,t_c}$ at time t_c.[3]
(6) *Probe-space identification of sensor base points*, i.e. the 9 milled divots in the outer side-plates bolted to the sensor: $\mathbf{p}_{B,i}^{P,t_c}$.

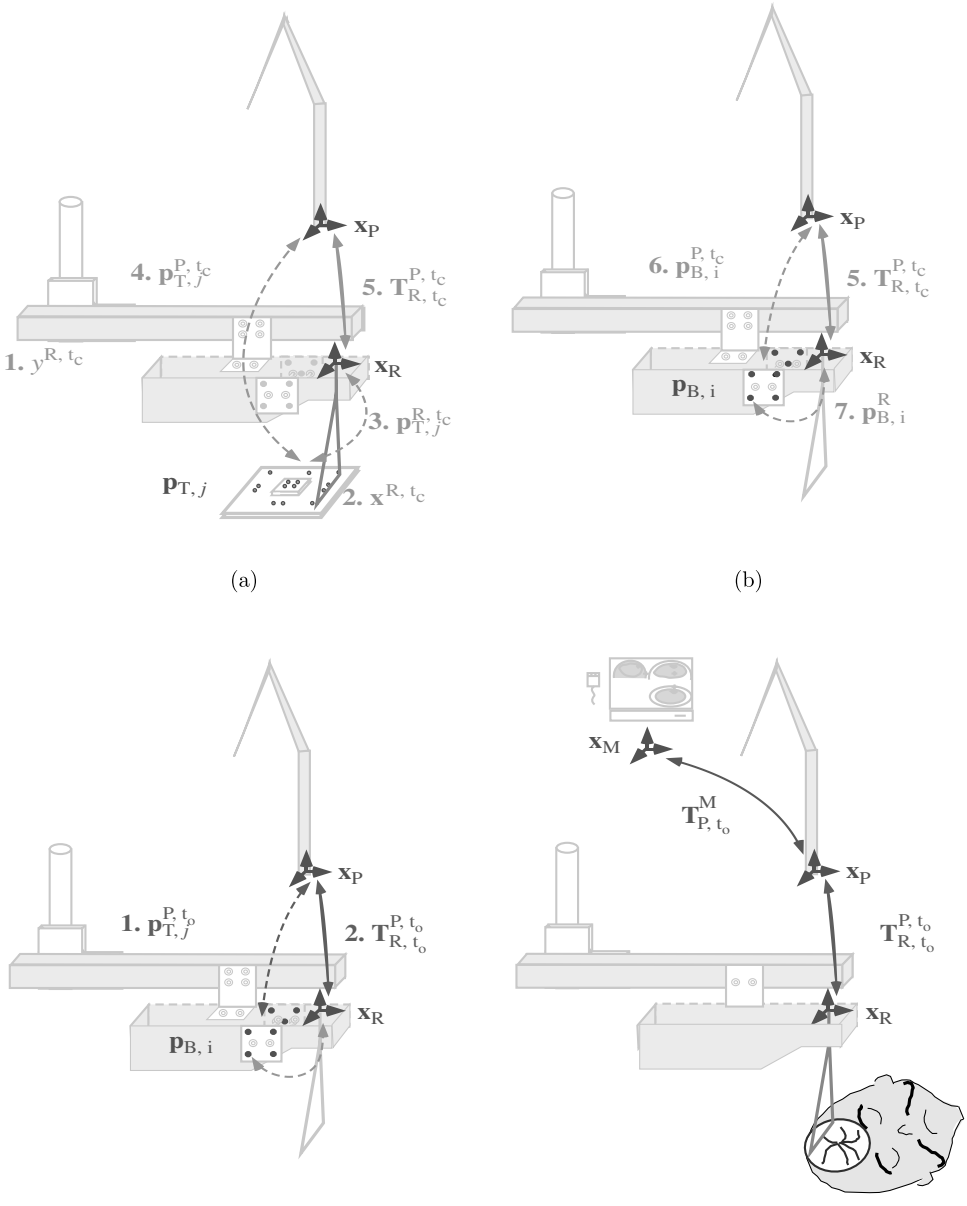


Fig. 5. Range to MRI transformation computation. At $t = t_c$, the calibration procedure first relates range to probe space (a), then determines the sensor base points in range space (b). (c) Sensor base tracking determines sensor base points in probe space at $t = t_o$ and finds the range-probe transformation in the OR. (d) The final step uses this and the probe-MR transformation to map range to MR.

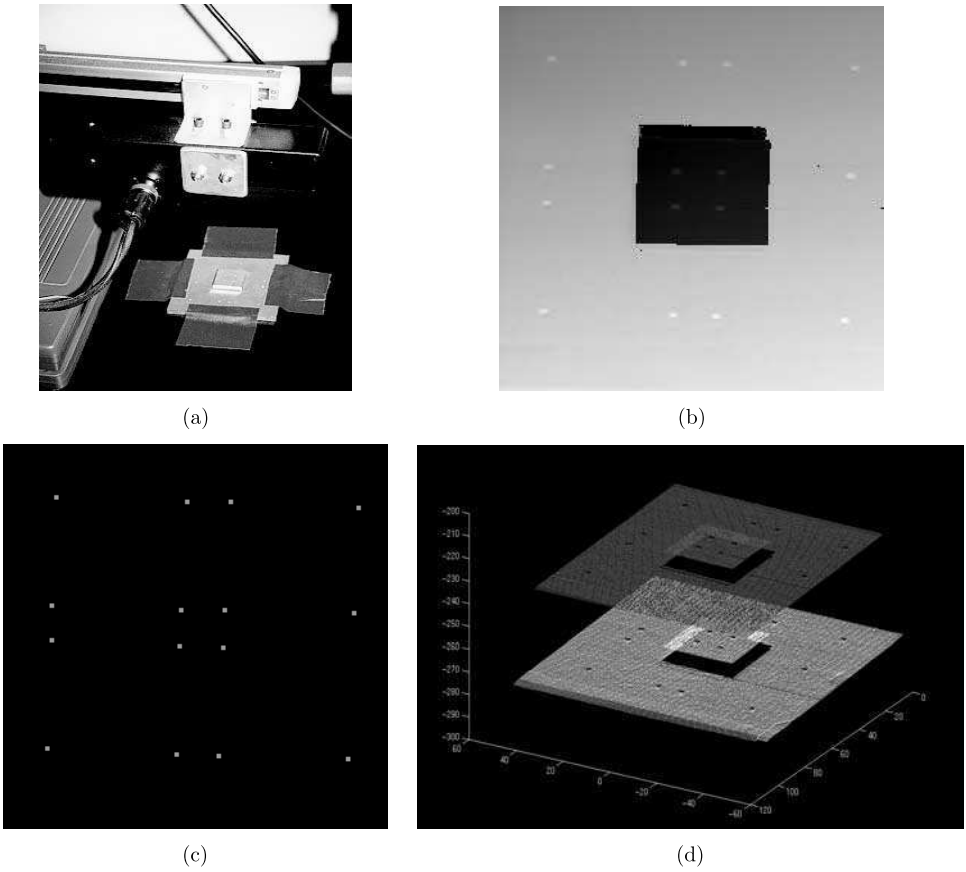


Fig. 6. Illustration of calibration procedure: (a) calibration plate and sensor, (b) range image of calibration plate, displayed as a gray-level; (c) detected pits in range image; (d) range images comprising a 2-level calibration procedure.

(7) *Apply probe-range transformation to sensor base points*:

$$\mathbf{p}_{B,i}^{R} = \mathbf{T}_{P,t_c}^{R,t_c}\ \mathbf{p}_{B,i}^{P,t_c} \tag{1}$$

relating the sensor base points to the range referential.
Steps (2) to (5) can be repeated at more than one positions for the calibration plate, as illustrated in Fig. 6(d), with the sensor home position remaining static.

The sensor base tracking procedure in the OR at t_o is as follows:

(1) *Identify the location of pits in sensor side-plates in probe space*, to characterize the new sensor base position: $\mathbf{p}_{B,i}^{P,t_o}$.
(2) *Match probe-space pits with the points* $\mathbf{p}_{B,i}^{R}$ to determine the probe reference-range transformation $\mathbf{T}_{R,t_o}^{P,t_o}$ for the configuration at time t_o.

Finally, the range data comprising surface S_B are then related to MRI-space anatomical surfaces by considering the IGS probe-MRI and the probe-range transformations:

$$\mathbf{x}^M_{S_B,t_o} = \mathbf{T}^M_{P,t_o}\mathbf{T}^{P,t_o}_{R,t_o}\mathbf{x}^R_{S_B,t_o} = \mathbf{T}^M_{R,t_o}\mathbf{x}^R_{S_B,t_o}\,. \tag{2}$$

The final result can be further refined with a rigid-body Iterative Closest Point registration, based on the skin surface identified in MR/CT data, and imaged in the OR by the range-sensor. The sensor base tracking procedure steps 1 and 2 can be performed in less than a minute, together with the range data acquisition, following the deployment of the sensor to a scanning position. The deployment of the sensor, on a locking articulated arm or on a programmable robot, could also be achieved within a minute or so, although the prototype described here was not tested in the OR. A dense 256×256 grid of range points, referred to MR space, is therefore achievable in considerably less time than it takes to manually acquire a sparse, probe-based cloud of cortical surface points.

2.3. *Surface non-rigid displacement estimation*

2.3.1. *Motivation and overview of non-rigid cortical motion estimation*

Brain shift at the exposed cortical surface, measured with respect to the rigidly transformed pre-operative brain surface identified in MRI, can be characterized as non-rigid surface motion captured by the initial range image, as well as those displacements captured in subsequent range images. This surface motion is estimated with a non-rigid Iterative Closest Point[8,31] registration technique. ICP techniques start from a rough initial alignment between two surfaces S_A and S_B, and iteratively use the set of *closest point pairs* between the two surfaces to determine a new incremental transformation. The incremental transformation $\mathbf{T}_k$ corresponding to the kth iteration is then applied to one of the surfaces (which may already be transformed), namely $\mathbf{T}_{k-1}S_B$, gradually improving its alignment with S_A. In most applications, the transformation $\mathbf{T}_k$ consists of an incremental rotation $\mathbf{R}_k$ and translation $\mathbf{t}_k$.

In our non-rigid registration application, $\mathbf{T}_k$ may also represent a 2D local piecewise polynomial function: a *smoothing surface spline*. The choice of a surface spline-based ICP technique is justified by its suitability to non-rigid motion estimation, in contrast with approaches based on features and global shape. Moreover, this technique is better suited than an active surface model to tracking an open range image.[4,5] The main drawback of such an iterative technique is the requirement of a rough initial alignment, but this is provided by the range-MR transformation procedure.

Prior to providing implementation details of our registration, we present a new elastically deformable brain-shaped[7] validation phantom, as shown in Fig. 7, made from *Polyvinyl Alcohol Cryogel* (PVA-C).[13] PVA-C is a relatively *viscous liquid*, which upon freezing and thawing, becomes an *elastic solid*. Furthermore, PVA-C

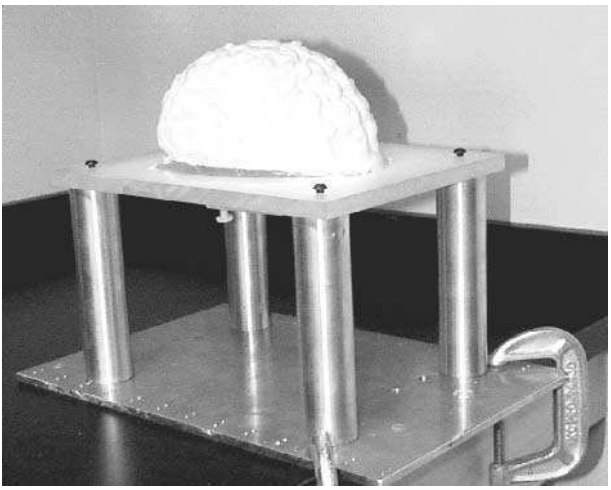


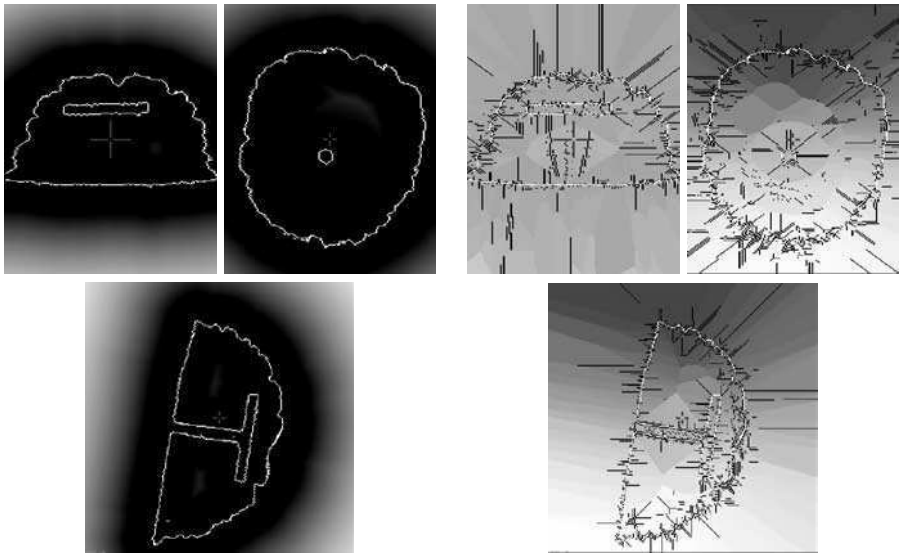
Fig. 7. Elastic PVA-C brain phantom.

can produce material properties comparable to values published in the literature for gray and white matter.[60] This phantom rests on a support plate of plexiglass (usable within an MR scanner), and features a moving assembly consisting of a small disk and rod, also of plexiglass, imbedded within the elastic material. The movement of this imbedded disk, in response to changes in the position of set-screws, triggers material deformation.

2.3.2. *ICP homologous pairs by look-up: The closest point map*

The iterative registration technique proposed here bears comparison to that of Lavallée and Szeliski,[31,53] which is characterized by the preprocessing, with an octree-based distance map, of the volume spanning one of the surfaces, to accelerate its registration with the second surface. When a second surface falls within the mapped volume, each point on it inherits a distance value to its closest homolog of the first surface. Because this preprocessing produces a distance map, but does not provide homologous pairs, the subsequent registration must still resort to an optimization method to determine the best transformation.

In contrast, our method incorporates a processing stage applied to the identified brain surface in the MRI volume, based on a propagative distance-to-surface algorithm called the *Fast Marching* (FM) method,[49] and which is modified to produce a *Closest Point Map*[5] as well. The Fast Marching method is used to compute distances from a given surface, such as shown in Fig. 8(a) for tri-planar images of the elastic brain phantom. The Closest Point Map is an adaptation of this method, whereby the *label of the closest surface point*, from which the distance is computed, is also stored. Moreover, as shown in Sec. 2.4, the FM method is also used to initialize the surface model with which we identify the outer brain surface. In other words, the same numerical estimation framework serves to both improve the efficiency of



(a) (b)

Fig. 8. (a) Distance and (b) Closest Point Maps computed from the surface of the elastic phantom (white), and shown in tri-planar views, mapped to a gray-level. In (b) ambiguous voxels, possessing more than one closest surface point, are shown in black.

the registration and to initialize the brain segmentation algorithm. To every voxel in MRI space, the Closest Point Map associates the label of the brain (or phantom) surface point closest to it. This concept is shown in Fig. 8(b) for tri-planar images of the elastic brain phantom, with labels mapped to a gray-level and with ambiguous voxels (more than one closest point) identified in black.

The characterization of non-rigid motion becomes much more efficient if its starting point is a dense set of homologous pairs, as opposed to a distance map which imposes a costly minimization at each ICP iteration. Each rigid transformation iteration can exploit a closed-form computation method,[3] any of which requires explicit point-pairs, not distances between them. Each non-rigid iteration can also make use of a dense vector displacement function obtained by look-up, and as will be seen in the next section, a highly efficient numerical scheme for smoothing this function with surface splines. The justification for emphasizing computational efficiency here is two-pronged: *clinical acceptability* and the inherent *temporal underdetermination* in estimating non-rigid motion.

2.3.3. *Modeling non-rigid surface displacement with 2D recursive spline filtering*

The regular spacing of the range data is exploited by using extremely efficient *recursive smoothing splines*[56] to smooth the vector displacement field. These surface splines express the fitting of interpolating, smoothing or approximating splines

as *recursive filters*, provided that the data samples exist at regular positions. As mentioned earlier, this is the case for the range surface points.

For this work, we adopt a smoothing spline approach, whose first stage is a convolution with a smoothing kernel[a]

$$S_\lambda^n(Z) = 1/\left(B_1^n(Z) + \lambda(-Z + 2 - Z^{-1})^{\frac{n+1}{2}}\right) \tag{3}$$

where $B_1^n(Z)$ is the *indirect B-spline transform* given by Unser,[56] n is the order of the spline fitting and λ is the regularization parameter of the smoothing. This stage produces B-spline coefficients and is followed by a convolution with the indirect transform $B_1^n(Z)$ to yield the smoothed output. The filters can simply be cascaded to implement smoothing in 2D. A first-order fit is chosen, i.e. $n = 1$, to limit overshoots due to a higher-order continuity constraint. This first-order recursive IIR filter can be broken down into factors in Z and Z^{-1}, and implemented as separate anti-causal and causal filters respectively, which translates into local operations involving the *current and immediately neighboring* pixels, independently of the smoothing parameter λ (for further details, see Ref. 56). The effect of λ is illustrated in Fig. 9.

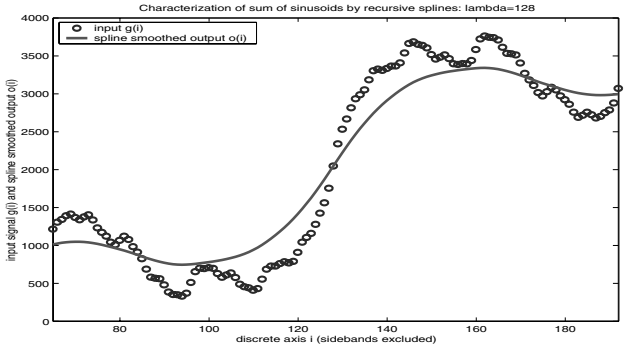
In order to make the non-rigid motion estimation well-behaved, the registration features the following stages:

- A *rigid registration* stage: a series of instantaneous rigid ICP iterations. This stage is halted when the registration error is no longer reduced significantly.
- A *non-rigid registration* stage: a series of non-rigid ICP iterations where the smoothing parameter λ is initially set very high, and is gradually lowered as the process iterates, progressively resolving finer-level motion detail.

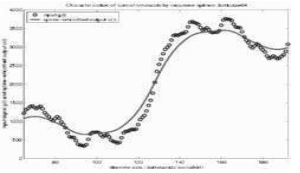
2.3.4. *Special considerations for range-MR non-rigid surface registration*

Each non-rigid registration iteration may require some additional consideration, in comparison to a rigid-body stage, as illustrated in Figs. 10 to 13. At some pixels in the range-sensor CCD, a low signal can occur, resulting in a gap or an unreliable range point. Such a low signal value can be detected from the z-value or intensity profile of the range image and pruned. Furthermore, because the range data are somewhat denser than the MR-based cortical surface points, the range-MR homologous pairs are *many-to-one*. Consideration of all of the homologous pairs would lead to an erroneous tangential displacement component. This issue suggests refining these many-to-one homologous pairs, resulting in a set of *mutually closest pairs*, followed by a proximity-weighted displacement propagation. In the neighborhood of gaps, the non-rigid motion estimation is more reliant on this displacement propagation. While it is geometrically feasible for a voxel to have more than one closest surface point, simple disambiguation and validation of each homologous pair

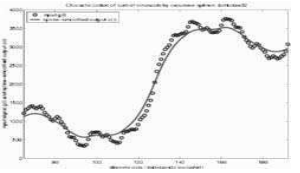
[a]Note: here Z relates to the Z-transform.



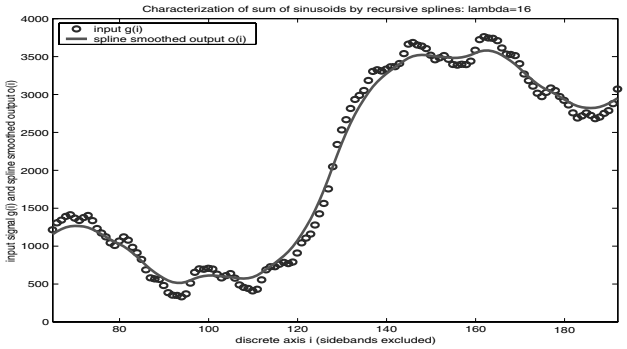
(a) $\lambda = 128$



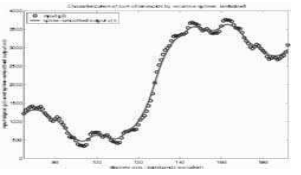
(b) $\lambda = 64$



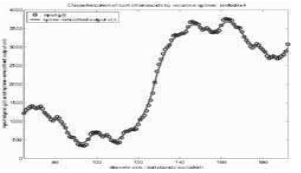
(c) $\lambda = 32$



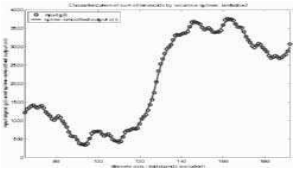
(d) $\lambda = 16$



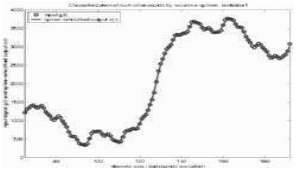
(e) $\lambda = 8$



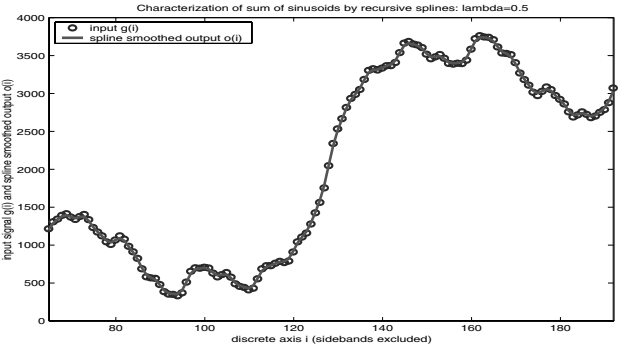
(f) $\lambda = 4$



(g) $\lambda = 2$



(h) $\lambda = 1$



(i) $\lambda = 0.5$

Fig. 9. Illustration of recursive spline filter response, given an input consisting of a sum of a ramp and 7 harmonic sinusoids, as λ decreases from 128 to 0.5.

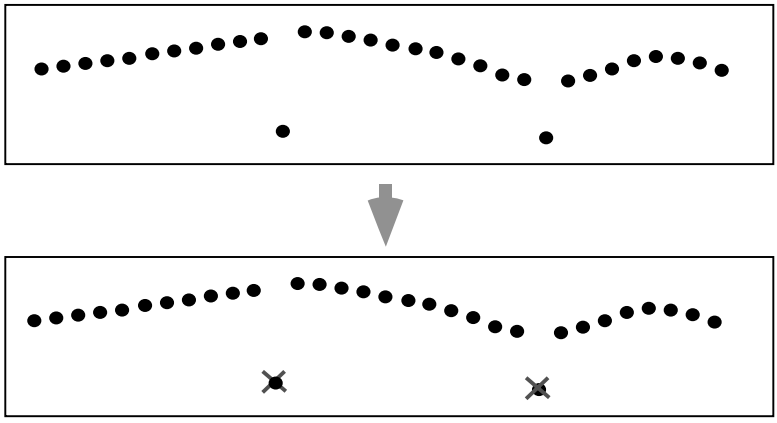


Fig. 10. Range data featuring unreliable, low-signal pixels which are identified and eliminated from further consideration.

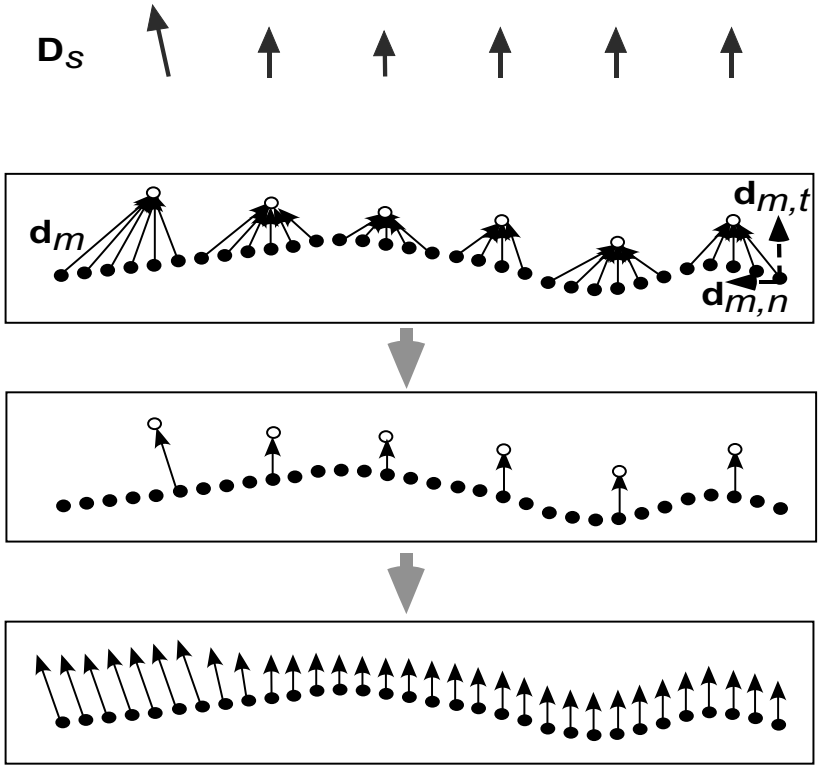


Fig. 11. Dealing with many-to-one range-MR matches arising from unequal data density (MR points in white, range points in black). Top: true surface displacement $\mathbf{D}_S$; second row: displacement field implicit in many-to-one homologous pairs; third row: disambiguation by pruning all but mutually closest homologous pairs; and bottom: distance-weighted propagation.

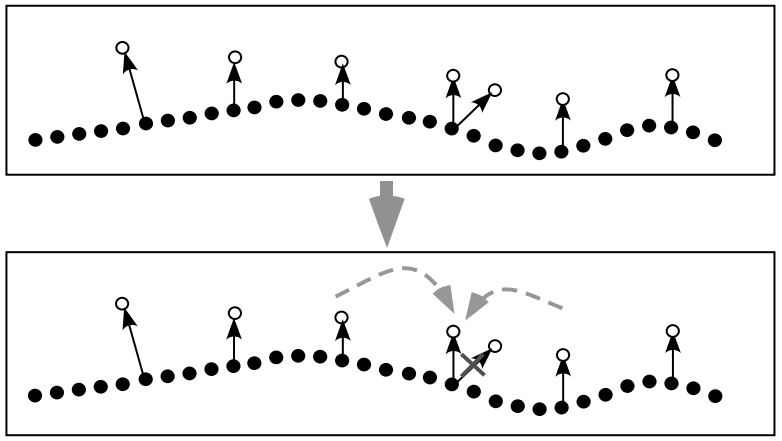


Fig. 12. Multiple or spurious MR closest point: disambiguation from continuous displacement assumption, by neighborhood estimate.

are possible, based on the assumption of *continuous displacement*.[37] This test is a comparison between the displacement vector determined by each homologous pair, and an average displacement computed from neighboring points.

Finally, the result of the procedure described so far is a *forward range-MR transformation* $\mathbf{T}_R^M$, composed of a final multiplicative rigid transformation $\mathbf{T}_{R,r}^M$ which is the product of successive ICP transformation matrices,[b] followed by an additive spline-based non-rigid displacement function $\mathbf{T}_{R,n}^M$:

$$\begin{aligned} &\mathbf{x}'_{R,i} = \mathbf{T}_R^M\left(\mathbf{x}_{R,i}\right) \equiv \mathbf{T}_{R,n}^M\left(\mathbf{T}_{R,r}^M \mathbf{x}_{R,i}\right) \quad \text{where} \\ &\mathbf{T}_{R,r}^M = \mathbf{T}_{R,r,K}^M \cdots \mathbf{T}_{R,r,1}^M \quad \text{and} \quad \mathbf{T}_{R,n}^M(\mathbf{x}) \equiv \mathbf{T}_{R,n,L}^M \cdots \left(\mathbf{T}_{R,n,K+1}^M\left(\mathbf{x}\right)\right). \end{aligned} \tag{4}$$

The indices K and L indicate the final iterations of rigid (multiplicative) and non-rigid (additive) ICP registration. In turn, the following recursive expression further defines the lth non-rigid transformation

$$\mathbf{T}_{R,n,l}^M(\mathbf{x}) \equiv \tilde{\mathbf{d}}_l\left(\mathbf{x} + \tilde{\mathbf{d}}_{l-1}\right) + \mathbf{T}_{R,n,l-1} \tag{5}$$

where $\tilde{\mathbf{d}}_l$ represents the smoothed displacement at iteration l, evaluated at a transformed range point. In other words, the method described so far maps the range surface data $\{\mathbf{x}_{R,i}$, where $\mathbf{x}_{R,i} \in S_R\}$ to MR space, in a manner that conforms to the pre-operative MR cortical surface data $\{\mathbf{x}_{M,j}$ where $\mathbf{x}_{M,j} \in S_M\}$.

What is needed to characterize brain shift, however, is the *inverse transformation* $\mathbf{T}_M^R$ which non-rigidly maps the pre-operative MR brain surface data $\mathbf{x}_{M,j}$ to conform to the intra-operative cortical range data. The inverse transformation of $\mathbf{T}_R^M$ is the negative of the overall range-MR non-rigid vector displacement field

[b]Homogeneous coordinates reduce the application of successive rigid body transformations to a sequence of matrix multiplications.[17]

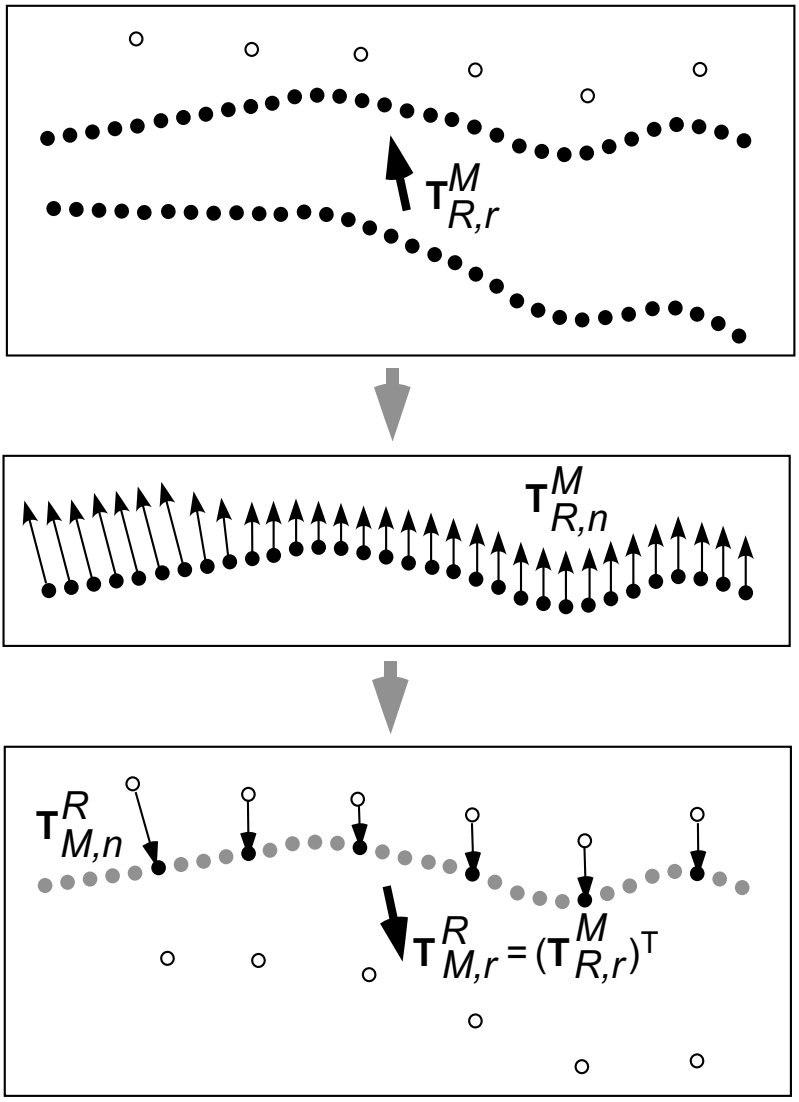


Fig. 13. Generating the MR-range transformation $\mathbf{T}_M^R$ from $\mathbf{T}_R^M$. Top: overall rigid range-MR transformation $\mathbf{T}_{R,r}^M$; middle row: overall non-rigid range-MR displacement $\mathbf{T}_{R,n}^M$. Bottom: the non-rigid and rigid components of the transformation which maps the MR brain surface to the range image are as follows: $\mathbf{T}_{M,n}^R$ is the negative of $\mathbf{T}_{R,n}^M$, and $\mathbf{T}_{M,r}^R$ is the transpose of $\mathbf{T}_{R,r}^M$.

applied to pre-operative MR data, followed by the application of the inverse, or transpose, of the overall range-MR rigid-body transformation, as shown in Fig. 13:

$$\begin{aligned} \mathbf{x}'_{M,j} = \mathbf{T}_M^R \mathbf{x}_{M,j} = \mathbf{T}_{M,r}^R \mathbf{T}_{M,n}^R \left(\mathbf{x}_{M,j}\right) \quad \text{where} \\ \mathbf{T}_{M,n}^R(\mathbf{x}) = -\mathbf{T}_{R,n}^M(\mathbf{x}) \quad \text{and} \quad \mathbf{T}_{M,r}^R = \mathbf{T}_{R,r}^{M^{-1}} = \mathbf{T}_{R,r}^{M^T}. \end{aligned} \tag{6}$$

The computation of the inverse transformation in this manner pre-supposes the smoothness of the forward non-rigid mapping, in a manner sufficient for the inverse to be well-defined, and the absence of significant occlusions.

2.4. *Model-based brain surface segmentation*

Currently, the traditional technique for identifying anatomical surfaces in IGS is a labor-intensive slice-by-slice approach where a "seed" is selected, corresponding to the tissue type of interest, and a software program identifies the rest of the contiguous voxels which are of similar intensity.[24] The contiguous labeled voxels within these slices together comprise a volume of interest, coinciding with brain tissue for example, the outer surface of which constitutes the boundary of interest. In contrast, a semi-automatic segmentation technique in general requires far less user interaction, has better reproducibility than a manual approach, can incorporate prior knowledge into the model to avoid common pitfalls, and can easily process image volume information in a fully three-dimensional manner. Surface models can be categorized into those based on an explicit representation and on a physical equilibrium between virtual internal properties of the model and image-based external forces,[26,54] and those based on the evolution of an implicit surface expressed as a partial differential equation in time, which imbeds the problem into one defined over the complete image domain[12,36] (for a survey, see Ref. 43). We adopt a surface evolution model because of its topological adaptability and simple formulation in 3D.

2.4.1. *Surface evolution model with classifier-based speed term*

The anatomical surface identification technique presented here is known as a surface evolution model,[12,36] which is characterized by imbedding the identification of a 3D surface $\mathcal{S}$ into the evaluation of a function Ψ whose domain spans the image volume:

$$\left\{X \in \Re^3 \text{ such that } \mathcal{S}(t) \equiv X : \Psi(X,t) = 0\right\}. \tag{7}$$

$$\frac{\partial \Psi}{\partial t} = F(x,y,z)\|\nabla\Psi\| \left[\operatorname{div}\left(\frac{\nabla\Psi}{\|\nabla\Psi\|}\right) + \nu\right] + \nabla F \cdot \nabla\Psi \tag{8}$$

where $\operatorname{div}(\nabla\Psi/\|\nabla\Psi\|)$ represents the mean curvature H of an isosurface of the function Ψ. This model features a *diffusive* smoothing term $\|\nabla\Psi\|\operatorname{div}(\nabla\Psi/\|\nabla\Psi\|)$, a *hyperbolic* motion term $\|\nabla\Psi\|\nu$, and two image terms: a *speed* function F and a *doublet* term $\nabla F \cdot \nabla\Psi$ which serve to bind the evolving surface to the anatomical boundary of interest. Moreover, the model is initialized by one or more initial user-defined surfaces $\{\Psi(X,0) = 0\}$ which fully contain, or are fully contained by, the anatomical boundary.

The model evolves in a manner which nudges the zero-level isosurface $\{\Psi(X,t) = 0\}$ inwards or outwards until the image terms bind it to the boundary. The imbedding function Ψ is initialized as a *signed distance map* from the initial user-defined surface(s). For the sake of efficiency, we restrict the computation of our model to a narrow band,[1] whereby the model is computed only within

a thin volumetric shell close to the evolving $\{\Psi(t) = 0\}$ isosurface. As mentioned earlier, we choose the Fast Marching (FM) technique in order to compute distances from the $\{\Psi(X,t) = 0\}$ isosurface, to initialize $\Psi(X,t)$ over its whole domain and to restrict the computation to a narrow band near the $\{\Psi(X,t) = 0\}$ isosurface.

In most existing implementations of surface evolution models, the speed term is a function of the image gradient, i.e.

$$F = \frac{1}{1 + \|\nabla \tilde{I}(x,y,z)\|^n} \tag{9}$$

where typically $n = 1$, 2 or 3 and $\tilde{I}(x,y,z)$ is usually a Gaussian-filtered image (MR) volume. However, this approach has some limitations when it comes to identifying the brain surface, such as a lack of $T1$ contrast between gray matter and the sagittal sinus and between gray matter and muscle tissue, possibly entailing a bleeding effect outside the brain surface. To alleviate this problem, the model is endowed with higher-level anatomic information, namely by first running a Minimum Distance (MD) *tissue classification* algorithm[18] on the image volume(s), and to compute a speed function which restricts the surface model to expand only within voxels classified as white matter (WM) and *certain* gray matter (GM) voxels, according to the following discrete relation:

$$\begin{aligned} F(x,y,z) &= 1.0 \text{ if } \{(x,y,z) \in WM\} \text{ or } \{(x,y,z) \in GM \text{ and near } WM\} \\ &= 0 \text{ otherwise}. \end{aligned} \tag{10}$$

Other classification techniques were considered, but the MD technique performs well with a small training set (the presence of pathology may preclude warping a large, generic training data set to patient data) and with $T1$-weighted data alone. Tissue classification is the thrust of ongoing research of ours[6] and is a complex issue in itself, particularly if it must cope with pathological data and with a relatively small set of training points. While it would appear feasible to simply consider the outside surface of GM points (with the Marching Cubes algorithm[33] for example) in practice, false GM positives that are iso-intense with GM (e.g. muscle) do arise and would undermine such an approach, particularly if only $T1$ data were available. The classifier-based surface model also offers an elegant means of integrating multi-spectral data, such as $T1$-, $T2$- and PD-weighted MR, if available.

2.4.2. *Surface model implementation*

Our implementation is a two-stage procedure which first identifies the outer WM boundary, then, based on a flexible coupling implemented with the *WM/GM Distance Function* (WDF),[5] the outer brain surface is identified with the position of the WM boundary as a constraint. The effect of this coupling on the speed term computation is illustrated in Fig. 14. One advantage of our method over more rigidly coupled surface models[34,59] is that it will tolerate embedded pathologies. Moreover, the threshold that limits the motion of the brain surface through GM voxels varies

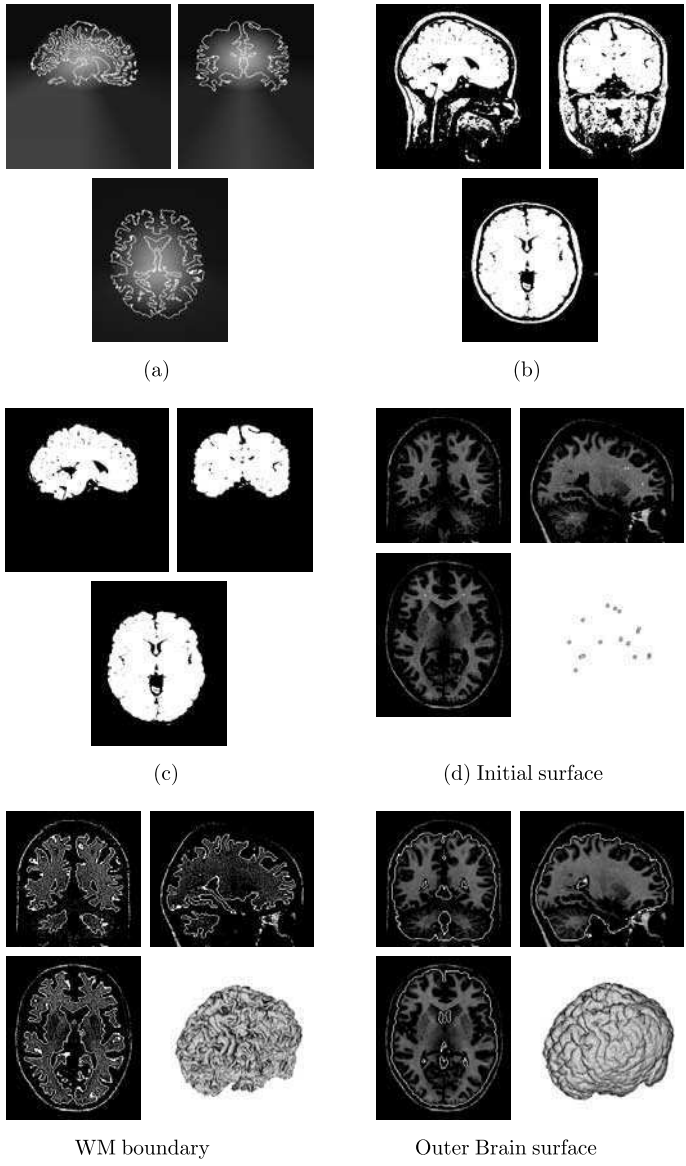


Fig. 14. WM/GM surface coupling: (a) WDF mapped to a gray-level intensity with WM contours in white; (b) speed function based on WM or GM tissue classes from $T1$ alone; (c) speed function, as in (b), but excluding GM voxels exceeding WDF shown in (a). (d) Surface model applied to ICBM subject 00100.

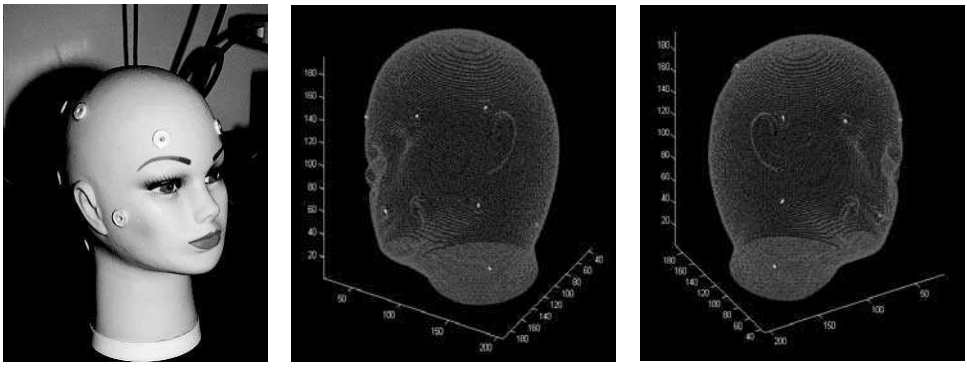
according to spherical coordinates, so that the coupling between white and gray matter takes into account prior assumptions about the thickness of gray matter, which is much thinner in the cerebral cortex than in the cerebellum or lower frontal area, for example.

The surface model computation culminates in a post-processing of the outer brain surface to produce a *Closest Point Map*, for the subsequent non-rigid registration with the intra-operative range-based surface. This is simply the Fast Marching method applied not to computing a distance map from the evolving $\{\Psi(X,t)=0\}$ isosurface within a narrow band, but to computing distance and closest point label from the final brain surface everywhere in the volume. *As a result of its propagative nature,*[49] *the FM algorithm is easily modified not only to estimate the distance of a surface to any voxel in the image volume, but also to store the identity of the particular surface point which is closest.* The motivation for using the FM algorithm is its computational efficiency and adaptability for storing Closest Point labels. However, there exist other methods for estimating distances which could serve the same purpose.

3. Validation and Discussion

3.1. *Validation of the range-MR transformation procedure*

Validation of all procedures involved in relating range to MR is based on the skin phantom shown in Figs. 15 and 16. This phantom features 12 glued-on torus-shaped fiducials, which are visible under both CT and MR. Since the phantom was solid and contained a metallic core, its construction precluded a study based on Magnetic Resonance imaging, and we employed a CT validation. The dimensions of the CT voxels are 0.89 mm in the x and y directions, and 1 mm in z. The localizing device used in this study is a FARO arm,[61] discussed in Sec. 1. The overall range-CT transformation makes use of the prior IGS probe-CT transformation based on 7 pairs of fiducials $\{\mathbf{x}^C_{F,k} \leftrightarrow \mathbf{x}^P_{F,k}\}$ accessible in probe space. The IGS transformation results



(a) (b) (c)

Fig. 15. (a) Skin phantom. 7 homologous point pairs used for left (b) and right (c) side studies: CT space (light gray) and probe space fiducials transformed to CT space (white), overlaid on outer skin phantom voxels.

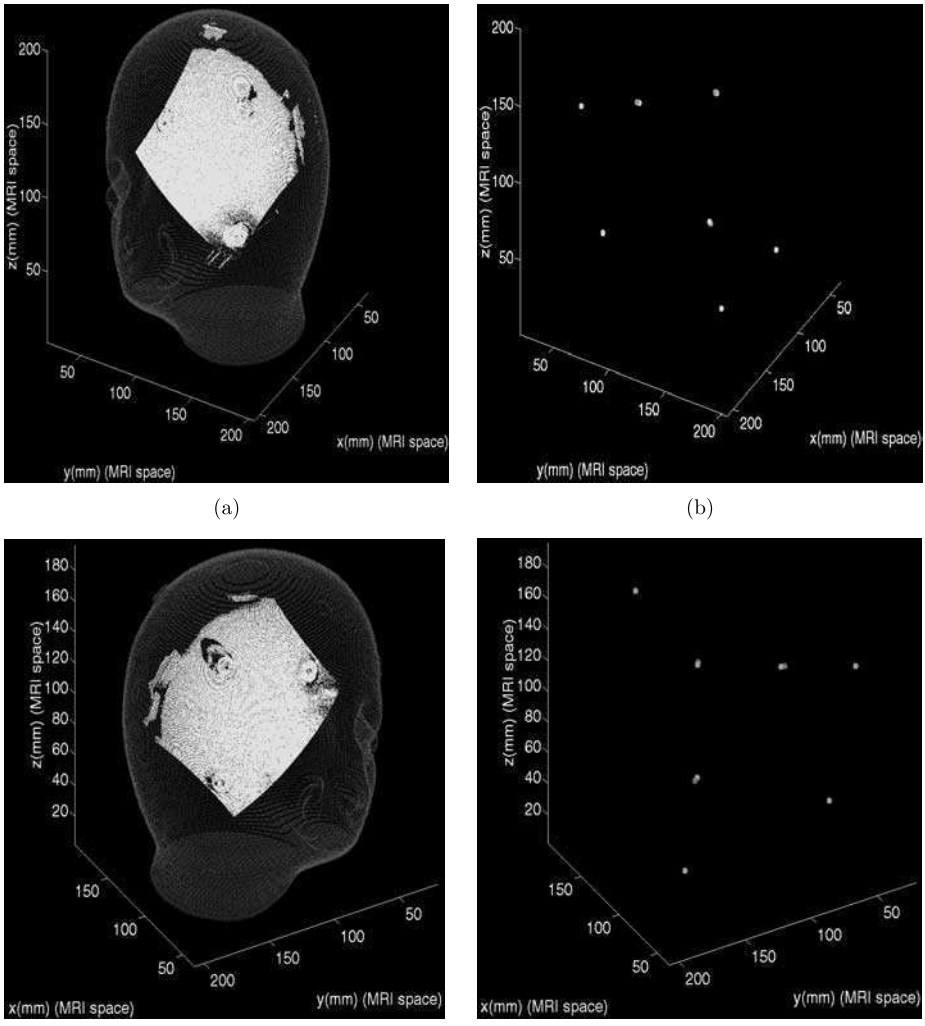


Fig. 16. Range-CT transformation results (trial 1): (a)–(b) left side: (a) range data transformed to CT space overlaid on CT data; (b) CT (gray) and transformed range space (white) fiducials; (c)–(d) right side: (c) transformed right side range data overlaid on CT data; (d) CT space and transformed range space fiducials.

for the left and right side studies are characterized by errors of 1.4 mm and 1.6 mm respectively. The validation study consists of a comparison of

- *computed CT fiducial positions*: manually identified fiducials in the range image, transformed to CT space with the procedure presented here, against
- *manually identified fiducial positions in CT.*

Table 1. Trial results for calibration and sensor base tracking procedure. Rigid body transformation by least-squares and by robust regression, with and without ICP refinement based on the skin identified in CT and range spaces. The value ϵ_{rms} represents the RMS error (in mm) between CT fiducial positions and their homologues in the range image transformed to CT space by $\mathbf{T}^{\mathbf{CT}}_{\mathbf{range}}$. The value d_{surf} is the RMS distance (in mm) between the transformed range points and the CT skin surface. The symbols $\|\mathbf{t}_c^o\|$ and θ_c^o indicate the magnitude of the translation and rotation between the sensor base positions at t_c and t_o.

		Left Side Study				
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
$\|\mathbf{t}_{cal}^{OR}\|$ (mm)/θ_{cal}^{OR} (deg)		293.4/11.0	139.1/11.1	201.6/12.9	265.5/53.8	167.0/14.4
least squares	ϵ_{rms} (mm)	2.48	2.48	2.98	2.57	1.99
	d_{surf} (mm)	0.97	0.88	0.99	0.79	0.61
robust	ϵ_{rms} (mm)	2.01	2.41	4.31	2.69	1.74
	d_{surf} (mm)	0.91	0.91	1.53	1.01	0.67
least squares, ICP	ϵ_{rms} (mm)	2.44	2.47	2.71	2.57	2.09
	d_{surf} (mm)	0.46	0.47	0.47	0.47	0.46
robust, ICP	ϵ_{rms} (mm)	2.11	2.16	3.90	2.51	2.01
	d_{surf} (mm)	0.46	0.46	0.49	0.46	0.45
		Right Side Study				
$\|\mathbf{t}_{cal}^{OR}\|$ (mm)/θ_{cal}^{OR} (deg)		266.7/1.5	141.9/1.0	231.1/2.7	243.5/43.8	191.2/24.4
least squares	ϵ_{rms} (mm)	2.14	2.48	3.39	2.43	2.19
	d_{surf} (mm)	0.97	0.52	0.89	0.91	0.63
robust	ϵ_{rms} (mm)	2.11	3.08	5.40	2.71	3.08
	d_{surf} (mm)	0.66	0.92	1.65	0.86	0.72
least squares, ICP	ϵ_{rms} (mm)	2.12	2.12	2.92	2.33	2.13
	d_{surf} (mm)	0.41	0.41	0.45	0.42	0.41
robust, ICP	ϵ_{rms} (mm)	2.20	2.49	4.27	2.42	2.78
	d_{surf} (mm)	0.41	0.40	0.45	0.41	0.40

While an automatic localization of the fiducials in both spaces may have been desirable, to do so reliably was difficult to achieve, due to self-occlusion in the range data.

Table 1 sums up the average values for disparities between identified and computed fiducial positions in CT space for the left and right side studies.

Also featured in Table 1 is a description of the magnitude of the transformation between the two sensor base positions at t_c and at t_o (during validation), from side-plate probe measurements. The transformation is expressed as two values: the magnitude of the calibration-OR translation vector $\|\mathbf{t}_c^o\|$ in mm and the angle of rotation θ_c^o in degrees, about the rotation axis.[c]

These results are encouraging and they are consistent with the localization technology used for implementing the IGS registration. In particular, Rohling[46] noted

[c]Recall that a rotation can be expressed in terms of Euler Symmetric Parameters, i.e. as an axis of rotation (k_x, k_y, k_z) and an angle of rotation θ about this axis (see Ref. 17, pp. 55–56).

that the FARO estimates distances to within 0.5 mm, whereas the Optotrak is generally within 0.15 mm, and Alp *et al.* have shown that torus-shaped fiducials are identified less precisely than other fiducials.[2] Nevertheless, the goal of this procedure should be to achieve a characterization of the range-MR transformation to the nearest millimeter.

These results can be interpreted in the context of the formal relationship between the *fiducial localization error* (FLE) and the *target registration error* (TRE) explored by Maurer, Fitzpatrick *et al.*[40] Furthermore, Langron and Collins,[32] based on prior work of Sibson,[50] showed that fiducial registration errors (FRE) add in quadrature. A lower bound on the final TRE of the calibration and sensor base tracking procedure can be estimated on the basis of the FLE or TRE associated with each step. Assuming a unity $TRE \approx TLE$ relationship (for the sake of simplicity), a probe FLE_P of 0.5 mm,[46] and a range $FLE_R \approx \sqrt{(0.5^2 + 0.5^2 + 0.5^2)} = 0.86\,\text{mm}$, where a value of 0.5 mm is also presumed for the y-axis (due to limitations of the linear positioner calibration procedure), we assess the accuracy of the calibration plate localization as $\sqrt{(0.5^2 + 0.86^2)} \approx 1.0\,\text{mm}$. The accuracy with which the sensor base divots can be located with respect to the range reference at t_c, is then $\sqrt{(1.0^2 + 0.5^2)} \approx 1.12\,\text{mm}$. The accuracy of the sensor base tracking procedure, from t_c to t_o, is given by:

$$TRE_{R-P,t_o} \approx FLE_{tot,p_B} \approx \sqrt{(1.12^2 + 0.5^2)} \approx 1.22\,\text{mm}\,. \tag{11}$$

If one factors in the TRE of the probe-CT transformation, which for this study is in the 1.5 mm range, then a reasonable lower bound for the final TRE is

$$TRE_{R-M,t_o} = \sqrt{(TRE^2_{R-P,t_o} + TRE^2_{P-M})} \approx \sqrt{(1.22^2 + 1.5^2)} \approx 1.93\,\text{mm}\,. \tag{12}$$

This value is compatible with the values observed in Table 1, particularly for a torus-shaped fiducial. Further analysis of these results indicates that the accuracy of the registration does not improve appreciably with robust statistics or with a subsequent skin-based ICP registration, and that there does not appear to be a correlation between the registration error and the size of the transformation.[5] The former finding suggests that the homologous pairs appear to preclude a severe outlier, as well as underscores the importance of their geometric configuration,[40] since the systematic removal of information by the robust method tends to make the registration degrade. The latter finding is consistent with most of registration error being tangential to the surface. One explanation could be a FLE in the range data which is more important than the first thought. A pixel error of $(\delta i, \delta j)$ in the fiducial localization results in a localization error whose largest components are along x_R and y_R, for a brain surface which for all intents and purposes is normal to z_R.

3.2. *Non-rigid registration validation*

3.2.1. *Validation study with brain shift simulation by realistic elastic brain phantom*

In order to assess non-rigid surface tracking, the set-screws under the phantom support plate are tightened, triggering a deformation of up to 15 mm at the top, and the phantom surface is then imaged by the range-sensor. A rough manual alignment is provided, based on finding the points on the MR surface corresponding to the corners of the range domain and adding a vertical offset. The gradual improvement of the registration of the range and MR surfaces of the elastic phantom, over successive iterations and progressively more local characterization, can be illustrated as follows.

- *Distance statistics*: average and standard deviation values computed from the distance between homologous pairs, as shown in Fig. 17.
- *Spatial distribution of the distance between two surfaces*: this distance function can easily be visualized as a gray-level point whose position is that of a surface point in range or MR space and whose gray-level is based on a distance-gray scale, as in Fig. 18. The top and bottom halves of this figure respectively

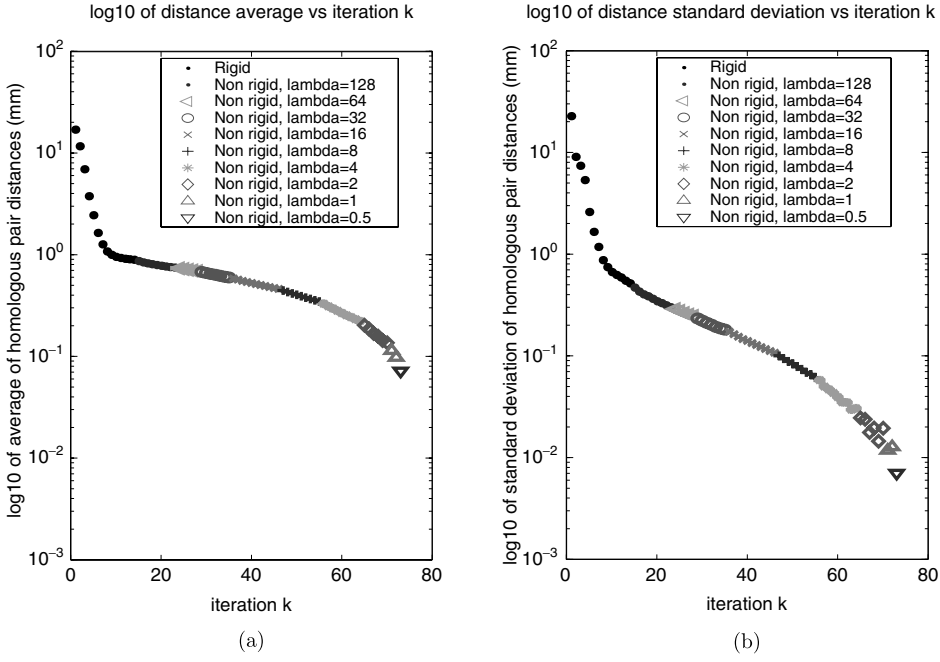
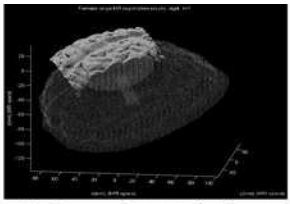
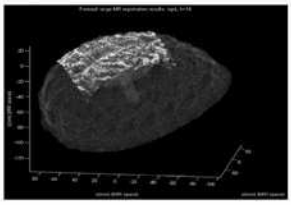


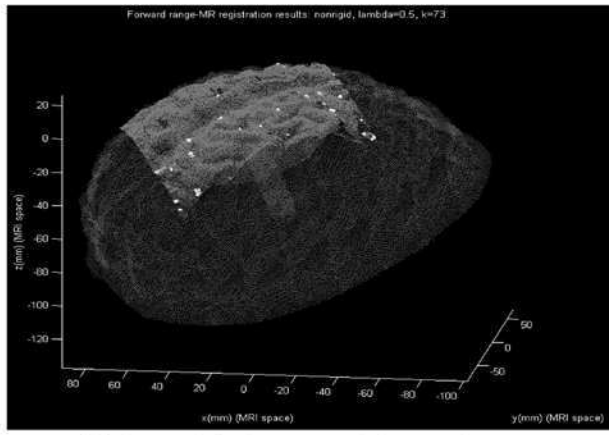
Fig. 17. Statistics of distances between homologous points, assuming an overall transformation $\mathbf{T}_R^M$ (based on expressions 4 evaluated up to iteration $K = k$ or $L = k$) applied to range points: (a) log_{10} of distance average; (b) log_{10} of distance standard deviation.



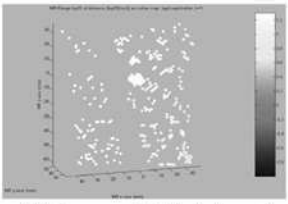
(a) forward initial: $k = 1$



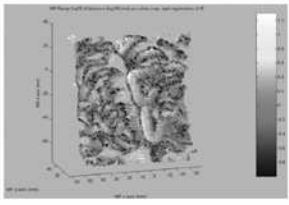
(b) forward final rigid: $k = 14$



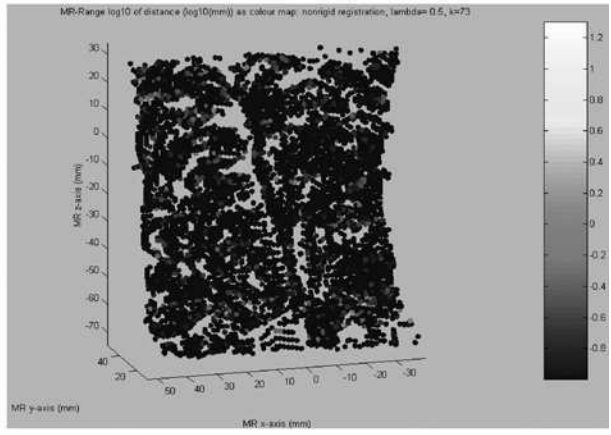
(c) forward final non-rigid: $k = 73$ $\lambda = 0.5$



(d) inverse initial: $k = 1$



(e) inverse final rigid: $k = 14$

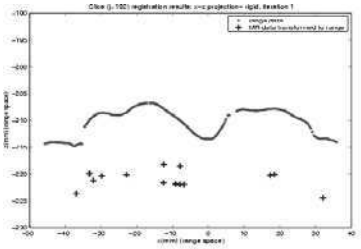


(f) inverse final non-rigid: $k = 73$ $\lambda = 0.5$

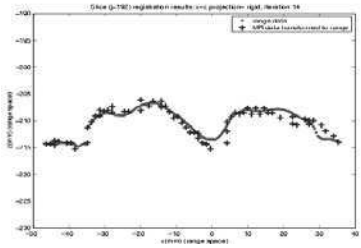
Fig. 18. Forward and inverse transformation results, featuring initial, final rigid and final non-rigid transformations. The gray-level scheme of the data indicates the distance between the range to MR brain surfaces. (a)–(c): Forward-transformed range data: medium gray (lighter than MR data): $d < 0.5$ mm, light gray: $0.5 \leq d < 2.0$ mm, white: $d \geq 2.0$ mm. (d)–(f): Inverse-transformed MR data; logarithmic scale: black corresponds to -1.0 (0.1 mm) and white to 1.3 (20 mm).

illustrate the *forward* and *inverse transformation* results, including the rough manual starting point, the final rigid transformation, and the final non-rigid transformation.

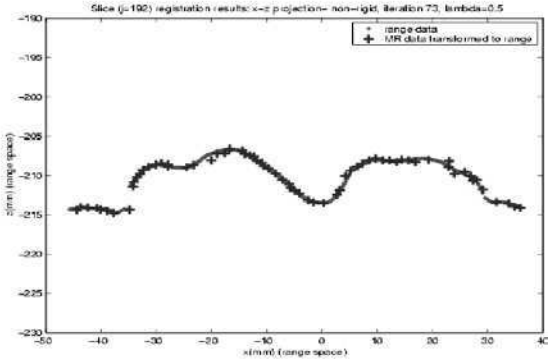
- *2D slice illustration*: a qualitative illustration of the surface registration as it evolves over k and progressively more local characterization is provided by the consideration of any given 2D slice of the range image, such as shown in Fig. 19 for the slice $y_R = 79.2$.



(a) rigid: $k = 1$



(b) final rigid: $k = 23$



(c) final non-rigid: $k = 73\ \lambda = 0.5$

Fig. 19. Inverse transformation results, featuring initial, final rigid and final non-rigid transformations, for the *slice* $y_R = 79.2$ ($j = 192$). Shown here are homologous range points and MR brain points mapped back to range space according to $\mathbf{T}^R_{M,k}$ and projected to the $x_R - z_R$ plane.

3.2.2. *Validation study with an analytical displacement applied to an anthropomorphic digital brain surface*

This validation study of the non-rigid surface registration procedure compares the vector displacement function applied to synthetic range image of the brain, obtained from an anthropomorphic digital phantom, with the displacement computed by the registration procedure. As shown in illustration 20, the clinical situation is simulated by taking a patch of the known brain surface of this phantom, of isotropic 1 mm sampling (refer to Collins[14] for further details), and interpolating it in order to replicate the density of the range-sensor CCD. To each synthetic range image obtained in this manner, we apply a "downward" analytical displacement function consisting of a constant plus a slowly varying sinusoid. This study features two synthetic range images, as shown in Fig. 20: a top patch, isolated from the most distal brain surface points of the phantom in a well-centered square area, whose displacement is essentially proximal, as well as a left patch, from the left-most brain surface points in a similarly well-centered square area, whose displacement is towards the right. A measure of the accuracy of the registration procedure is then computed from the vector difference between the *pre-deformation* synthetic range patch, and the *post-deformation, post-registration* range patch. Figure 21 and Table 2 illustrate the gradual improvement in displacement estimation, over successive iterations. The deformation function is of the type

$$\mathbf{d} \equiv d_3 = [A + B\cos(2\pi f(x_1 - \bar{x}_1))\cos(2\pi f(x_2 - \bar{x}_2))]\,\hat{x}_3 \tag{13}$$

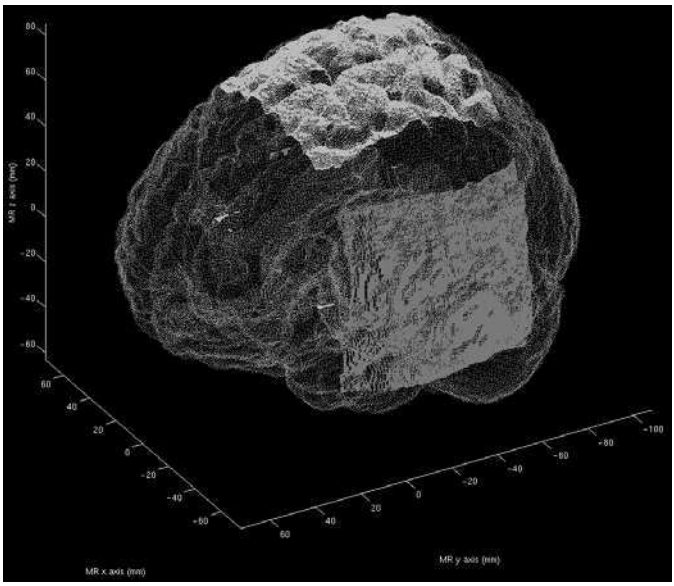
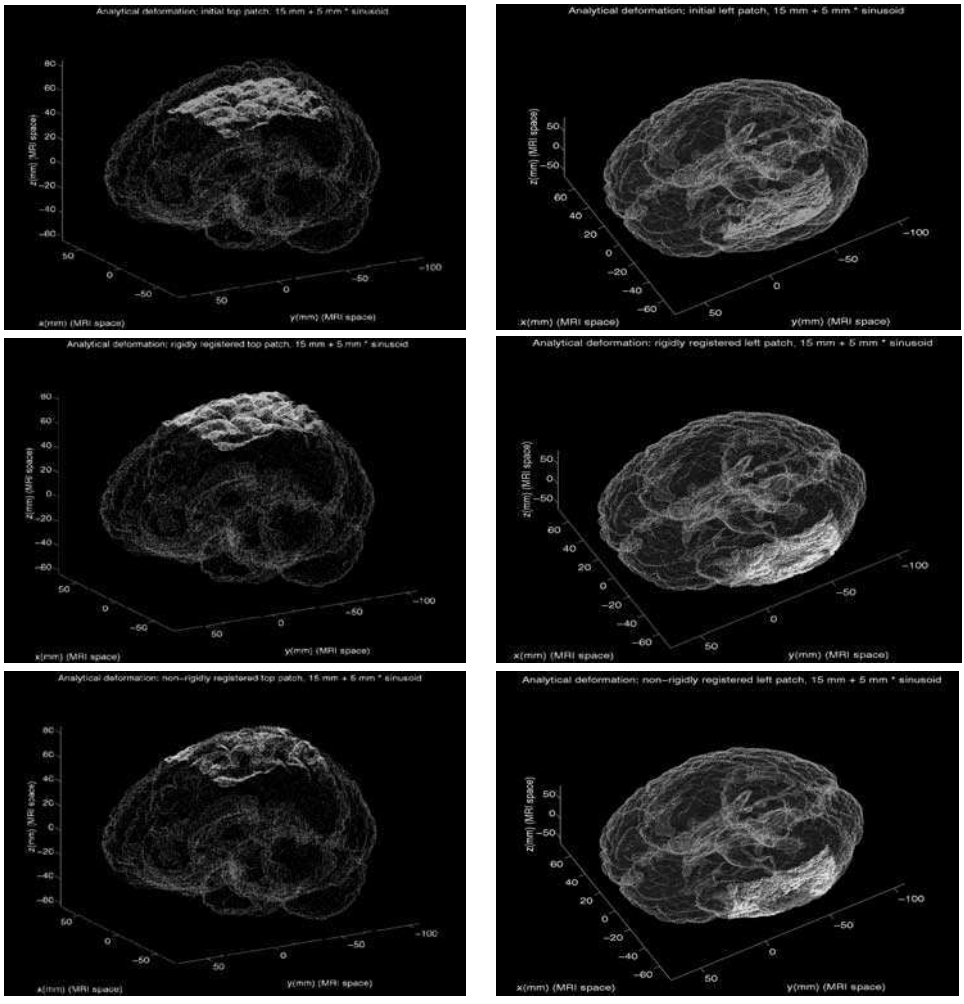


Fig. 20. Two synthetic range images overlaid on original phantom brain surface data, from which they were obtained by interpolation.

where x_1 and x_2 coincide with x and y for the top patch study, and with y and z for the left patch study. The period of the sinusoid is chosen to be approximately the width of the patch, and a few combinations of A and B are documented in Table 2. At the finer scales, small improvements in surface-to-surface distances may no longer be reflected in an improvement in the displacement estimation, and there is a risk that spline-based motion characterization could warp the surface too aggressively, essentially unfolding the sulci. To alleviate this risk, the progression of the motion characterization from coarse to fine is gradual, and each non-rigid iteration is in fact a composition of a rigid-body iteration and a spline-based non-rigid motion characterization. As a consequence, the iterative minimization of surface-to-surface distance correlates well with the minimization of the error in a non-rigid displacement estimation.

Estimation error statistics of the analytical deformation study are presented in Table 2, along with the distance between two surfaces. In all cases, the final non-rigid registration provides, on average, displacement estimation on the order of 1 mm or less. Nevertheless, this error is significantly larger than the final surface-to-surface distance. Given that the average distance between the two surfaces is relatively small, much of the displacement error is likely to be orthogonal to the direction of the true displacement. In no way is any assumption of a strong gravitational component exploited in the deformation estimation, because even if the movement were exclusively downward, one could not assume that the sensor $x - y$ axis were normal to the "downward" direction. Therefore, attempting to model the movement



(a) (b)

Fig. 21. Analytical deformation results: (a) top study and (b) left study. Top row: initial deformation $A + B\cos(2\pi f(x_1 - x_{1,c}))\cos(2\pi f(x_2 - x_{2,c}))$, where $A = 15\,\text{mm}$ and $B = 5\,\text{mm}$; middle row: final rigid registration; bottom row: final non-rigid registration. Medium gray (lighter than MR data) signifies that the deformation is estimated with error $\leq 1.0\,\text{mm}$, light gray: $1.0\,\text{mm} <$ error $\leq 2.0\,\text{mm}$, and white: error $> 2.0\,\text{mm}$.

exclusively along z would not characterize it adequately. However, the assumption of a dominant downward component is manifested in the validation displacement function which is along z for the top patch study, and along x for the left patch study. The anomalous result of the second top patch study stems from an early exit of the rigid registration stage, which could be prevented by more tolerant logic, but even then the non-rigid stage, consisting of a sequence of compositions of rigid transformation and spline-based displacement, provided reasonable results.

Table 2. Analytical deformation statistics. Left column: deformation parameters (in mm, mm and mm^{-1}), as per expression (13). Middle column: magnitude of vector difference between estimated and analytical displacement (at the start, after the final rigid iteration and after the final non-rigid iteration). Right column: distance between mutually closest points of the synthetic range and MR surfaces.

Deformation Parameters			Estimation Error (mm)			Inter-Surface Distance (mm)		
Left Patch			Start	Rigid	**Nonrigid**	Start	Rigid	Nonrigid
A = 5	B = 5	f = 1/80	4.2 ± 2.2	1.4 ± 0.6	$\mathbf{0.7 \pm 0.2}$	3.6 ± 5.4	0.7 ± 0.3	0.09 ± 0.01
A = 10	B = 5	f = 1/80	9.3 ± 2.1	1.6 ± 0.5	$\mathbf{0.8 \pm 0.2}$	4.0 ± 12.2	0.8 ± 0.3	0.09 ± 0.01
A = 15	B = 5	f = 1/80	14.1 ± 2.0	1.8 ± 0.6	$\mathbf{0.9 \pm 0.3}$	3.7 ± 13.9	0.8 ± 0.3	0.08 ± 0.01
A = 5	B = 5	f = 1/60	3.3 ± 3.2	1.7 ± 0.9	$\mathbf{0.9 \pm 0.3}$	3.0 ± 3.8	0.9 ± 0.4	0.09 ± 0.01
Top Patch			Start	Rigid	**Nonrigid**	Start	Rigid	Nonrigid
A = 5	B = 5	f = 1/80	5.2 ± 2.1	1.2 ± 0.6	$\mathbf{0.6 \pm 0.2}$	2.4 ± 2.5	0.7 ± 0.3	0.10 ± 0.01
A = 10	B = 5	f = 1/80	10.6 ± 2.3	9.1 ± 2.3	$\mathbf{1.0 \pm 0.5}$	2.8 ± 4.5	0.8 ± 0.3	0.09 ± 0.01
A = 15	B = 5	f = 1/80	15.8 ± 2.5	1.6 ± 0.4	$\mathbf{0.6 \pm 0.2}$	2.9 ± 5.5	0.8 ± 0.3	0.09 ± 0.01
A = 5	B = 5	f = 1/60	4.1 ± 3.2	1.5 ± 1.0	$\mathbf{0.7 \pm 0.3}$	2.2 ± 1.8	0.8 ± 0.4	0.10 ± 0.01

In view of the final non-rigid results in the bottom row of Fig. 21, the spatial distribution of the points, many of whose estimation error lies between 1.0 and 2.0 mm, suggests a possible limitation of using an Iterative Closest Point scheme to model brain deformation. The estimation is better at ridges and peaks, coinciding with gyri, and conversely is worse in crevasses coinciding with sulci, as well as in sloped areas. One possible interpretation is that the closest point pair constitutes a good homologous set when the surface normals at both points are roughly aligned with the true direction of the deformation, as is the case with anatomically equivalent points from a gyrus. However, a sulcus is characterized by a crevasse surrounded by points on a slope: for a given crevasse point on surface A, the closest point is likely to be on a slope, rather than the corresponding crevasse point on surface B. This geometric limitation is also why the ICP method cannot resolve motion having a strong component tangential to the either surface normal. For example, while this method can resolve a radial increase in scale of a spherical surface, it cannot deal with a pure rotation.

3.2.3. *Nonrigid registration validation conclusions*

This section featured innovations related to the validation of non-rigid registration, through the use of a constitutively realistic brain-shaped deformable phantom, whose deformation is triggered from within, and an analytical deformation function applied to a synthetic cortical range surface, obtained by interpolating the surface coinciding with the outer gray matter boundary of the digital head phantom. The registration results are consistent with other, less efficient ICP techniques, characterized by a gradual reduction of the distance between two surfaces.

The study based on a known deformation suggests that overall the ICP registration technique accurately estimates non-rigid motion, but performs slightly less

well in highly sloped areas, such as those in the deep recesses of sulci. This consideration leads to a refinement of the ICP method similar to that proposed by Feldmar and Ayache,[19] where only those points which are both close in space and of similar shape are considered homologous. However, it is less obvious how to adapt this shape-based approach to a Closest Point Map.

3.3. *Surface model validation*

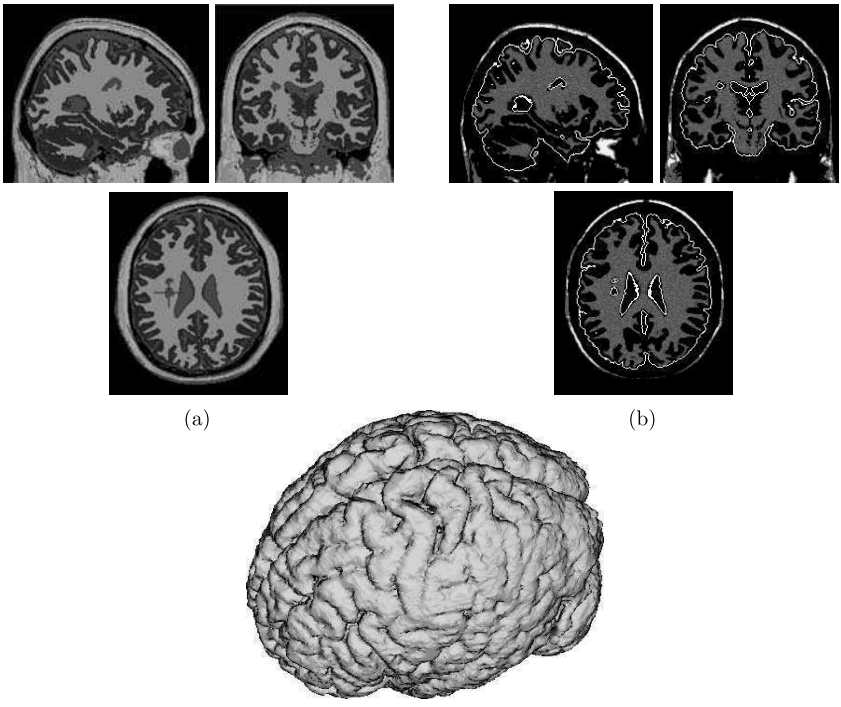
The accuracy of our surface-based brain shift estimation framework presupposes an accurate brain segmentation algorithm. We have validated the brain surface identification technique presented so far with a quantitative approach based on realistic synthetic data, where ground truth is known, and qualitative tests with *in vivo* subject data. The latter data comes in the form of healthy subject scans obtained from the ICBM[d] database, as well as Montreal Neurological Institute (MNI) patient scans featuring pathologies. This validation is presented in Ref. 5.

The quantitative validation stage presented in this section makes use of two software tools which were developed at the MNI: a MRI simulator specifically designed and implemented by Kwan *et al.*[30] for the purpose of image analysis evaluation, and a high-resolution, volumetric, anthropomorphic digital brain phantom elaborated by Collins *et al.*[14] In turn, the healthy-subject digital phantom is the starting point for a second phantom featuring pathological tissue.[10] The accuracy of the segmentation of the brain surface within simulated $T1$-weighted MR data can be assessed from the known boundary of the brain within the digital phantom, as shown in Fig. 22 (for more details, see Ref. 5).

The evolution of the zero-level isosurface, from spheres imbedded in white matter, through the white matter boundary, to the outer brain surface, is illustrated for healthy subject 00100 from the ICBM database in Fig. 14. Other examples, featuring real patient data with pathologies present, appear in Fig. 23. These examples represent the worse case, from a classification standpoint: $T1$ data considered alone, as one would expect the tissue discrimination ability of the classifier to improve with the consideration of $T2$ and PD-weighted data.

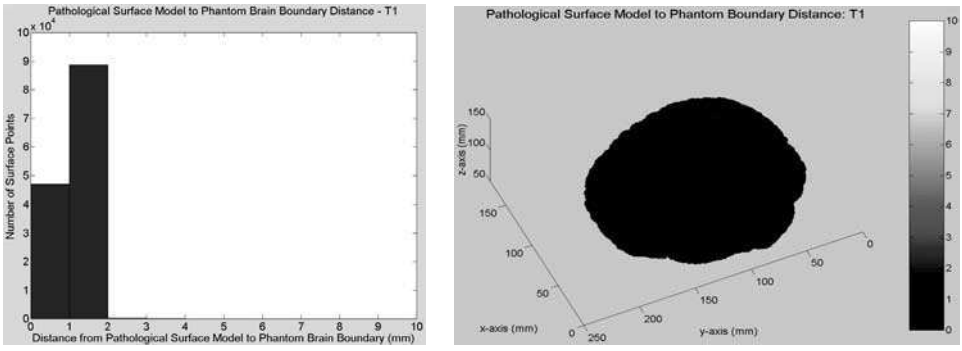
This section presented a brief overview of the validation study of our surface evolution model. Our study makes use of data from real subjects both healthy and with pathologies, as well as synthetic data with simulated pathological tissue. One point not discussed so far is that the model as currently implemented assumes a voxel-based boundary, in particular because the FM technique employed at the time used a voxel-based surface as a starting point. However, FM techniques for triangulated surfaces have since been published, and it should be feasible to adapt the Closest Point Map computation to triangulated surfaces as well. This improvement would be significant, as it would produce a Distance and Closest Point Map of a surface established with sub-voxel accuracy. The overall accuracy of the surface

[d]International Brain Mapping Consortium.



(a) (b)

(c)



(d) (e)

Fig. 22. Surface model behavior with synthetic pathological data: (a) original pathological phantom and (b)–(e) brain surface identified with surface model, from simulated $T1$-weighted scan: (b) tri-planar view; (c) rendered view; (d) distribution of distance between brain surface in scan and phantom brain boundary; (e) error between brain surface identified in scan and underlying brain boundary in phantom, shown as gray-level plot: avg. $= 0.68 \pm 0.51$ mm.

tracking method proposed here is dependent on the accuracy of all its constituent stages, so that an overall objective of 1 mm presupposes segmentation accuracy well within 1 mm, which in turn will probably only accrue from a sub-voxel, triangulated method.

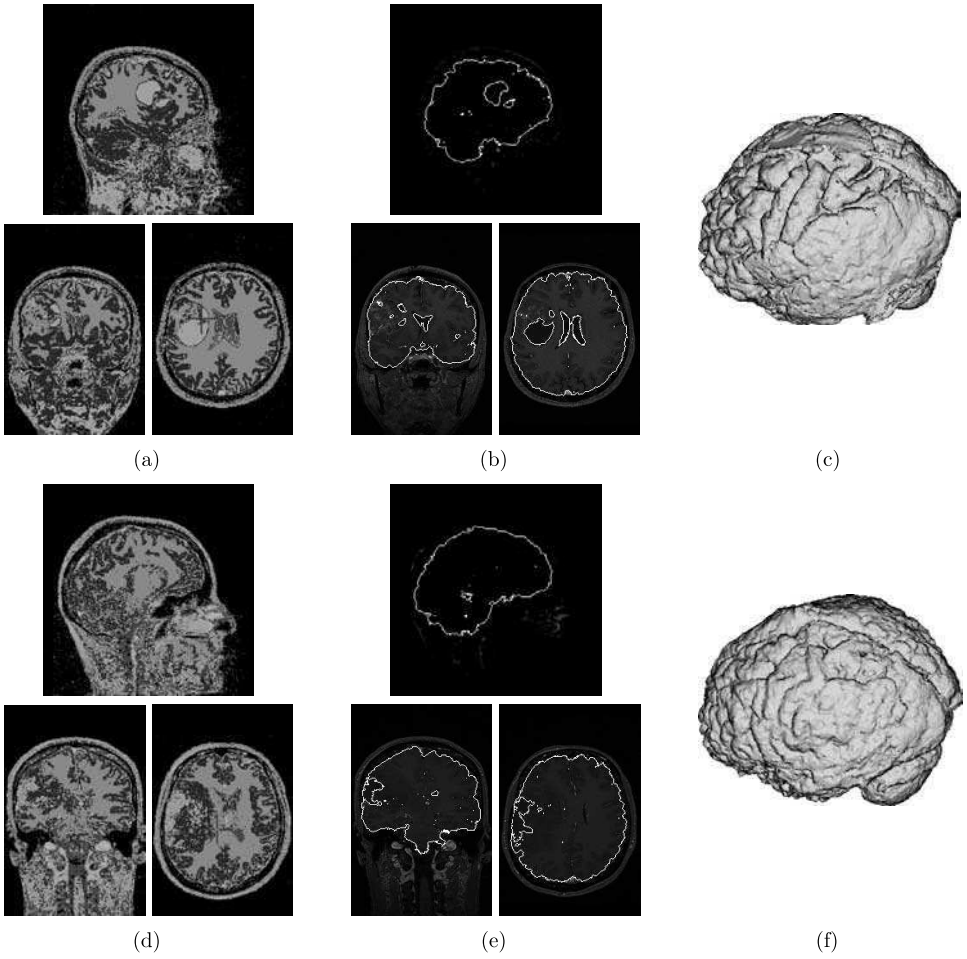


Fig. 23. Surface model behavior with real pathological $T1$-weighted data; (a)–(c) patient 1: (a) classification results, featuring 7 classes; (b) tri-planar view; (c) rendered view; (d)–(f) patient 2: (d) classification results, featuring 7 classes; (e) tri-planar view; (f) rendered view.

4. Conclusion

This chapter addressed the problem of intra-surgical brain deformation, and suggested a non-rigid surface registration framework for characterizing this deformation. The objective of this framework was to produce a smoothed, dense vector displacement function from the undeformed preoperative brain surface, semi-automatically identified in MR data, to the exposed intra-operative brain surface characterized with a laser range-sensor and referred to the MR coordinate system. This vector displacement function could then be interpolated everywhere in the brain volume, on the basis of an ulterior volumetric, constitutively realistic FE model.

Furthermore, the methods proposed here were conceived from assumptions based on practical clinical requirements. First, the semi-automatic method for identifying the brain surface in MR should be able to cope with the presence of imbedded pathology, even very large tumors. Next, the 3D surface capture and surface registration methods should be fast, producing results virtually at the instant they are required in the OR, with no user intervention except to initiate the processes. And finally, the method for relating the MR and range spaces must function despite the presence of large swaths of surgical draping, possibly occluding one or more intra-surgical fiducials, which suggests a calibration-based solution.

The underlying assumption of this chapter is that an accurate volumetric displacement characterization is possible with a displacement map at the exposed cortical surface and with a suitable finite-element model. The documented use of volumetric FE models, with sparse surface displacement data and with surface models fit to intra-operative MR data, suggests that at the very least, the accurate volumetric estimation of intra-surgical brain displacement with this framework is highly feasible. A future validation study which would support this hypothesis is the comparison of the position of beads imbedded in our elastic brain-shaped phantom, as detected within a MR volume, against the position predicted on the basis of a surface registration and volumetric interpolation approach. Furthermore, the method presented here is entirely compatible with the integration of probe-based information about the volume of resected tissue,[51] whose interaction with the finite-element brain model could be inspired from cutting models used in surgical simulation.[35]

References

1. D. Adelsteinsson and J. A. Sethian, A fast level set method for propagating interfaces, *J. Computational Physics* **118** (1995) 269–277.
2. M. S. Alp *et al.*, Head registration techniques for image-guided surgery, *Neurol. Res.* **20** (1998) 31–37.
3. K. S. Arun, T. S. Huang and S. D. Blostein, Least-squares fitting of two 3-D point sets, *IEEE Trans. PAMI* **9**, 5 (May 1987) 698–700.
4. M. A. Audette, F. P. Ferrie and T. M. Peters, An algorithmic overview of surface registration techniques for medical imaging, *Med. Imag. Anal.* **4**, 3 (2000) 201–217.
5. M. A. Audette, Surface-based characterization of intra-surgical brain deformations in image-guided surgery, PhD thesis, Department Biomedical Engineering, McGill University, Montreal, Canada, 2002.
6. M. A. Audette, Global structure-preserving voxel classification for patient-specific surgical simulation, *Proc. IEEE Conf. EMBS*, Houston, 2002.
7. www.barebonesstore.com/products/default.asp?Prod=86.
8. P. J. Besl and N. D. McKay, A method for registration of 3D shapes, *IEEE Trans. Pattern Analysis & Machine Intelligence* **14**, 2 (1992) 239–256.
9. F. Blais and J. A. Beraldin, Calibration of an anamorphic laser based 3D range sensor, *SPIE Proc. Videometrics V* **3174** (1997) 113–122.
10. P. A. Bottomley *et al.*, A review of ^{1}H nuclear magnetic resonance relaxation in pathology: Are T_1 and T_2 diagnostic?, *Med. Phys.* **14**, 1 (January/February 1987) 1–37.

11. R. J. Campbell and P. J. Flynn, Eigenshapes for 3D object recognition in range data, *Proc. IEEE Conf. Computer Vision & Pattern Recognition*, 1999.
12. V. Caselles *et al.*, Minimal surfaces: A three dimensional segmentation approach, *Technion EE Pub. No. 973*, 1995.
13. K. C. Chu and B. K. Rutt, Polyvinyl alcohol cryogel: An ideal phantom material for MR studies of arterial flow and elasticity, *Magnetic Resonance in Medicine* 37 (1997) 314–319.
14. D. L. Collins *et al.*, Design and construction of a realistic digital brain phantom, *IEEE Trans. Medical Imaging* **17**, 3 (June 1998) 463–468.
15. R. M. Comeau, Intraoperative Ultrasound Imaging for the Detection and Correction of Tissue Movement in Image-guided Surgery, PhD Thesis, Department Biomedical Engineering, McGill University, Montreal, 2000.
16. R. M. Comeau *et al.*, Intra-operative ultrasound for guidance and tissue shift correction in image-guided neurosurgery, *Med. Phys.* **27**, 4 (April 2000).
17. J. J. Craig, *Introduction to Robotics: Mechanics and Control* (Addison-Wesley, 1989).
18. R. O. Duda and P. E. Hart, *Pattern Analysis and Scene Analysis* (Wiley Interscience, 1973).
19. J. Feldmar and N. Ayache, Rigid and affine registration of free-form surfaces, using differential properties, *Proc. Euro. Conf. Comp. Vision*, 1994, 397–406.
20. M. Ferrant *et al.*, Registration of 3D intra-operative MR images of the brain using a finite element biomechanical model, *MICCAI*, 2000, 19–27.
21. R. L. Galloway, R. J. Maciunas and J. W. Latimer, The accuracies of four stereotactic frame systems: An independent assessment, *Biomed. Instr. & Tech*, November/December 1991, 457–460.
22. A. Guéziec and N. Ayache, Smoothing and matching of 3D space curves, *Int. J. Comp. Vision* **12**, 1 (1994) 79–104.
23. Intelligent Actuator (IAI), www.intelligentactuator.com.
24. ISG Technologies Inc., Viewing Wand Image-Guided Surgical System-Operator's Guide, 1997.
25. A. E. Johnson and M. Hebert, Surface matching for object recognition in complex three-dimensional scenes, *Image and Vision Computing* 16 (1998) 635–651.
26. M. Kass, A. Witkin and D. Terzopoulos, Snakes: active contour models, *Intl. J. Computer Vision*, 1987, 321–331.
27. P. J. Kelly *et al.*, Computer-assisted stereotaxic laser resection of intra-axial brain neoplasms, *J. Neurosurg.* 64 (1986) 427–439.
28. R. Khadem *et al.*, Comparative tracking error analysis of five different optical tracking systems, *Comput. Aided Surg.* **5** (2000) 98–107.
29. B. B. Kimia and K. Siddiqi, Geometric heat equation and nonlinear diffusion of shapes and images, *Computer Vision and Image Understanding* **64**, 3 (November 1996) 305–322.
30. R. K.-S. Kwan, A. C. Evans and G. B. Pike, MRI simulation-based evaluation of image-processing and classification methods, *IEEE Trans. Medical Image Analysis* **18**, 11, November 1999.
31. S. Lavallée and R. Szeliski, Recovering the position and orientation of free-form objects from image contours using 3D distance maps, *IEEE Trans. Pattern Analysis & Machine Intelligence* **17**, 4 (1995) 378–390.
32. S. P. Langron and A. J. Collins, Pertubation theory for generalized procrustes analysis, *J. Royal Stat. Soc. B* **47** (1985) 277–284.
33. W. E. Lorensen and H. E. Cline, Marching cubes: A high resolution 3D surface construction algorithm, *Computer Graphics* **21**, 3 (July 1987) 163–169.

34. D. MacDonald, D. Avis and A. C. Evans, Proximity constraints in deformable models for cortical surface identification, *Medical Image Computing and Computer-Assisted Intervention- MICCAI'98*, 1998, 650–659.
35. M. Mahvash and V. Hayward, Haptic rendering of cutting: A fracture mechanics approach, **2**, 3, *www.Haptics-e.org*.
36. R. Malladi, J. A. Sethian and B. C. Vemuri, Shape modeling with front propagation: A level set approach, *IEEE Trans. Pattern Analysis & Machine Intelligence* **17**, 2, February 1995.
37. D. Marr, *Vision* (W. H. Freeman and Co., 1982).
38. C. R. Maurer *et al.*, Measurement of intra-operative brain deformation using a 1.5 tesla interventional MR system: Preliminary results, *IEEE Trans. Med. Imag.* **17**, 5 (October 1998) 817–825.
39. C. R. Maurer *et al.*, Measurement of intra-operative brain surface deformation under a craneotomy, *MICCAI*, 1998.
40. C. R. Maurer *et al.*, Registration of head volume images using implantable fiducial markers, *IEEE Trans. Med. Imag.* **16**, 4, August 1997.
41. J. Michiels *et al.*, On the problem of geometric distortion in magnetic resonance images for stereotactic neurosurgery, *Magn. Reson. Imag.* **12**, 5 (1994) 749–765.
42. M. I. Miga *et al.*, Updated neuroimaging using intra-operative brain modeling and sparse data, *Stereotac. & Func. Neurosurg.* **72** (1999) 103–106.
43. J. Montagnat *et al.*, Representation, shape, topology and evolution of deformable surfaces, Application to 3D medical segmentation, *Image & Vis. Comput.* **19**, 14 (December 2001) 1023–1040.
44. Northern Digital Inc., www.ndigital.com.
45. A. Olivier *et al.*, Image-guided surgery of epilepsy, *Clinical Frontiers of Interactive Image-Guided Neurosurgery*, ed. R. J. Maciunas, *Neurosurg. Clin. of North America* **7**, 2 (April 1996) 245–266.
46. R. Rohling *et al.*, Comparison of relative accuracy between a mechanical and an optical position tracker for image-guided neurosurgery, *J. Imag. Guid. Surg.* **1** (1995) 30–34.
47. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd edn. (Academic Press, 1982).
48. H. Schiffbauer, Neuronavigation in brain tumor surgery-clinical beta-phase of the Oulu neuronavigator system, Academic Dissertation, Faculty of Medicine, Univ. Münster, available online at http://herkules.oulu.fi/isbn9514252543.
49. J. A. Sethian, Fast marching level set method for monotonically advancing fronts, *Proc. Nat. Acad. Sci. USA* **93** (February 1996) 1591–1595.
50. R. Sibson, Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling, *J. Royal Stat. Soc. B* **41** (1979) 217–229.
51. M. Sinasac, *Use of 3D deformable models for intra-operative visualization and quantification of cerebral tissue resection*, Master's Thesis, Department Electrical and Computer Engineering, McGill University, 1999.
52. O. Škrinjar, D. Spencer and J. Duncan, Brain shift modeling for use in neurosurgery, *MICCAI*, 1998, 641–649.
53. R. Szeliski and S. Lavallee, Matching 3D anatomical surfaces with non-rigid deformations using octree-splines, *Int. J. Computer Vision* **18**, 2 (1996) 171–186.
54. D. Terzopoulos and D. Metaxas, Dynamic 3D models with local and global deformations: Deformable superquadrics, *IEEE Trans. Pattern Analysis & Machine Intelligence* **13**, 7 (1991) 703–714.

55. J.-Ph. Thirion, Extremal points: Definition and application for 3D image registration, *Proc. IEEE Conf. Computer Vision & Pattern Recognition*, 1994, 587–592.
56. M. Unser, A. Aldroubi and M. Eden, B-spline signal processing: Part I- Theory, *IEEE Trans. Signal Proc.* **41**, 2 (February 1993) 821–833.
57. Vitana Corporation, *Shape Grabber Reference Manual*, 1997.
58. M. Y. Wang *et al.*, An automatic technique for finding and localizing externally attached markers in CT and MR volume images of the head, *IEEE Trans. Biomed. Eng.* **43**, 6, June 1996.
59. X. Zeng *et al.*, Segmentation and measurements of the cortex from 3D MR images, *Medical Image Computing and Computer-Assisted Intervention- MICCAI'98*, 1998, 519–530.
60. C. Zhou, T. B. Khalil and A. I. King, A new model comparing impact responses of the homogeneous and inhomogeneous human brain, *Society of Automotive Engineers, Inc. report #952714*, 1995.
61. S. J. Zinreich *et al.*, Frameless stereotaxic integration of CT imaging data: Accuracy and Initial Applications, *Radiol.* **188**, 3 (September 1993) 735–742.

CHAPTER 8

IMAGE REGISTRATION AND FUSION FOR INTERVENTIONAL MRI-GUIDED TREATMENT OF PROSTATE CANCER

BAOWEI FEI* and DAVID L. WILSON†
Departments of Biomedical Engineering and Radiology
Case Western Reserve University and University Hospitals of Cleveland
10900 Euclid Ave, Cleveland, OH 44106-7207, USA
* *baowei.fei@case.edu*
† *david.wilson@case.edu*

We are investigating interventional MRI (iMRI) guided radiofrequency thermal ablation for the minimally invasive treatment of prostate cancer. Nuclear medicine can detect and localize tumor in the prostate not reliably seen in MRI. We intend to combine the advantages of functional images such as nuclear medicine SPECT with iMRI-guided treatments. Our concept is to first register the low-resolution SPECT with a high resolution MRI volume. Then by registering the high-resolution MR image with live-time iMRI acquisitions, we can, in turn, map the functional data and high-resolution anatomic information to live-time iMRI images for improved tumor targeting. For the first step, we used a three-dimensional mutual information registration method. For the latter, we developed a robust slice-to-volume (SV) registration algorithm with special features. The concept was tested using image data from three patients and three volunteers. The SV registration accuracy was $0.4\,\text{mm} \pm 0.2\,\text{mm}$ as compared to a volume-to-volume registration that was previously shown to be quite accurate for these image pairs. With our image registration and fusion software, simulation experiments show that it is quite feasible to incorporate SPECT and high resolution MRI into the iMRI-guided minimally invasive treatment procedures.

Keywords: Image registration; image fusion; interventional magnetic resonance imaging (iMRI); nuclear medicine; image guided therapy; minimally invasive treatment; radiofrequency thermal ablation.

1. Introduction

We use an interventional magnetic resonance imaging (iMRI) system to guide minimally invasive treatments, including the radiofrequency (RF) thermal ablation of abdominal cancers.[1–3] The iMRI system consists of a 0.2 T, clinical *C*-arm open MRI scanner, an in-room RF-shielded liquid crystal monitor, an MR compatible mouse, a foot pedal, and a RF device. We are currently investigating the extension of these techniques to the treatment of prostate cancer. Since MRI does not reliably show prostate tumors, we intend to incorporate nuclear medicine or MR spectroscopy images with higher sensitivity for detecting and localizing prostate tumors.[4,5] We will first register the low-resolution functional images with a high-resolution MRI volume.[6,7] Then by registering the high-resolution MR volume

with live-time iMRI acquisitions, we can, in turn, map the functional data and high-resolution anatomic information to live-time iMRI images for improved tumor targeting. As discussed later, since live-time iMRI is used for device guidance, the accuracy requirements for registering these supplemental images might be less strict than required in some other applications.

We previously described a rigid body volume-to-volume (VV) registration method for the pelvic and prostate MR images that was accurate when images were acquired under similar conditions.[8] We used bony landmarks and three-dimensional (3D) centroids of segmented prostates to evaluate VV registration. For volume pairs acquired over a short time span from a supine subject with legs flat on the table, registration accuracy of both the prostate centroid (typically < 1 mm) and bony landmarks (average 1.6 mm) was on the order of a voxel (≈ 1.4 mm). The centroid error was slightly smaller because the prostate was at the volume center and rotation errors had less effect on it. The localization error in finding 3D points from bony landmarks is probably greater than that of finding centroids of relatively large prostate volumes where segmentation errors average out. We obtained somewhat larger prostate registration errors of about 3.0 mm when volume pairs were obtained under very different conditions that would be avoided in patient studies, e.g. legs flat and legs raised.

To incorporate image data from other sources in a live-time iMRI procedure, we intend to register two-dimensional (2D) slice images quickly acquired on the iMRI scanner in live-time with a previously acquired MR volume. We call this slice-to-volume (SV) registration. Because of our success with VV prostate registration, we can determine SV accuracy by comparing results to VV registration for volume pairs having low VV registration error.

To incorporate an image volume from another modality, it can be registered with the full MR volume. Thus, to incorporate SPECT in an iMRI procedure, we will first register the SPECT image volume with a high-resolution MR volume; then, when we register iMRI slice images to the high-resolution MR volume, we can also map them to the SPECT functional image data. If this procedure is successful, then a variety of potential visualization tools can help the physician appropriately localize and apply treatments. The live-time iMRI images will be used for guidance, and very probably any small misregistration errors can be mentally corrected by the physician. To possibly improve the slice-to-volume (SV) registration step, we intend to use MR images acquired with similar pulse sequences.

The application of SV registration to iMRI-guided treatment of prostate cancer raises several challenges. First, a single slice has much less voxels than an entire volume for voxel based matching. Second, iMRI images often have lower signal to noise ratio (SNR) than diagnostic MR images because of the emphasis on fast imaging and because of the typically lower field strength of open iMRI magnets. Third, the normal prostate is a small organ; when healthy, it measures only ≈ 3.8 cm in its widest dimension.[9] The small prostate is located below the much larger bladder that can change its shape and size during imaging. Fourth, the non-homogenous receive

coil response can change from one imaging session to the next. Finally, times for registration and algorithm robustness are of particular concern for this treatment application.

Previously reported methods for SV registration were mainly applied to the brain for applications of functional MRI,[10] postmortem pathology studies,[11] and anatomical modeling.[12] Voxel-based methods, particularly those based upon mutual information (MI), are robust, require no segmentation that can be prone to error, are suitable for multi-modality registration, and are highly accurate for many applications.[3,8,10,13–15] However, the MI method has the problem of interpolation artifacts, which can be especially serious in the case of down sampling in a multi-resolution approach.[16] Other similarity measures such as the correlation coefficient (CC) can reduce the presence of local minima.[17]

In the next sections, we will report algorithms and results for the slice-to-volume registration between an iMRI thick slice and a high-resolution MRI volume, the three-dimensional registration of SPECT and high resolution MRI volumes, and the fusion of the three modalities for potential applications in iMRI-guided thermal ablation of the prostate.

2. Registration Algorithm

2.1. *Similarity measurements*

We used two similarity measures, mutual information and correlation coefficient, in our registration. Suppose one image R is the reference, and the other F is floating. Their mutual information $MI(R, F)$ is given below.[18]

$$MI(R,F) = \sum_{r,f} p_{RF}(r,f) \log \frac{p_{RF}(r,f)}{p_R(r) \cdot p_F(f)} .$$

The joint probability $p_{RF}(r, f)$ and the marginal probabilities $p_R(r)$ of the reference image and $p_F(f)$ of the floating image, can be estimated from the normalized joint intensity histogram. The correlation coefficient $CC(R, F)$ is given below.[19]

$$CC(R,F) = \frac{\sum (R(r) - \bar{R}(r))(F(f) - \bar{F}(f))}{\sqrt{\sum (R(r) - \bar{R}(r))^2 \sum (F(f) - \bar{F}(f))^2}} .$$

Here $\bar{R}(r), \bar{F}(f)$ denote the average intensities of the reference and floating images and the summation includes all voxels within the overlap of both images.

We compared the two similarity measures at different resolutions in order to determine their suitability for SV registration. At 1/4 resolution, we resampled images so as to give 1/4 number of the voxels along each linear dimension. At full resolution, we used the full number of voxels. In Figs. 1 and 2, we plot the two similarity measures as a function of two translation parameters. After two typical high-resolution MR volumes were registered,[8] values were plotted with the origin as the optimal transformation. We calculated CC and MI values while moving the simulated iMRI image relative to the high-resolution MR image along coronal

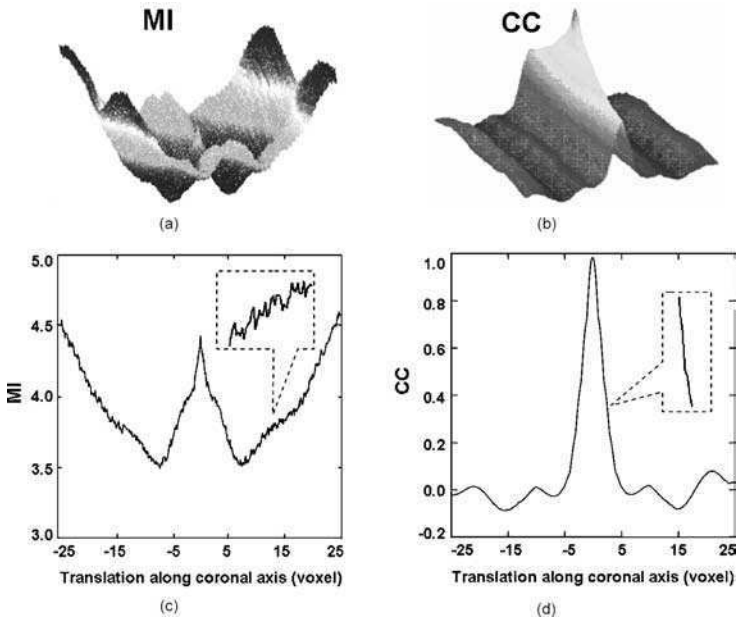


Fig. 1. Similarity functions are plotted as a function of translations at the lowest resolution in the multi-resolution registration process. Two high-resolution MRI volumes were registered. From the optimal parameters, we computed the similarity of the simulated iMRI and MRI images as a function of translations along the coronal (anterior-posterior) and sagittal (left-right) axes. MI is plotted in (a) and (c); CC is plotted in (b) and (d). Graphs (a) and (b) are three-dimensional (3D) plots for translations along the coronal and sagittal axis. Graphs (c) and (d) are two-dimensional (2D) plots for translations about the coronal axis. The small insets in (c) and (d) are magnified curves showing noise having local maxima in (c). A false global maximum for MI occurred at +25 voxels. Images are from volunteer $S2$, and they are down sampled by 1/4 along each linear dimension, giving a distance between voxel centers of ≈ 5.5 mm.

(anterior-posterior) and sagittal (left-right) axes. The simulated iMRI image was obtained as described later in Sec. 3.4.

Features of MI and CC demonstrate their suitability at high and low resolutions, respectively. At 1/4 resolution, CC surfaces are much smoother than MI, which is noisy and contains many local maxima as shown in Fig. 1(a) and Fig. 1(c). In fact, there is a false global maximum at +25 voxels. At full resolution, Figs. 2(a) and 2(c) shows that MI has a much sharper peak than CC, but once again there is high frequency noise in the MI curves, far from the optimum, that gives rise to local maxima that must be avoided. From these figures, we infer that CC is better at low resolution and MI is better at full resolution, when one is close to the optimum value. As described next, our registration algorithm makes use of these features.

2.2. *Slice-to-volume registration algorithm with special features*

The algorithm includes special features to improve robustness for registration of MR prostate images. Suppose the iMRI image slice is the reference slice, the matching

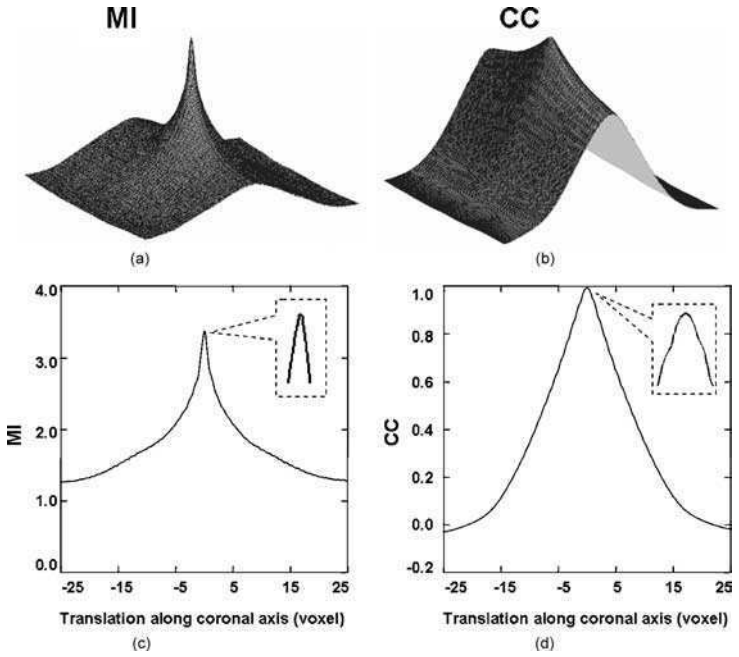


Fig. 2. Similarity functions are plotted as a function of translations at full resolution. Many details are given in the legend of Fig. 1. Again, MI is plotted in (a) and (c); CC is plotted in (b) and (d). MI in (a) and (c) has a much sharper peak than CC in (b) and (d). The voxel is isotropic with 1.4 mm on a side. Image data are the same used in Fig. 1.

slice extracted from the high-resolution MRI volume is the reformatted slice, and the final reformatted slice is the registered slice. We use a multi-resolution approach and perform registration from low to high resolution. We use CC at the two lower resolutions because it gives fewer local maxima and because it can be calculated faster than MI. We use MI at full resolution because of its peaked surface. To avoid local maxima, we include a restarting feature where registration is restarted with randomly perturbed parameters obtained from a uniform distribution about the initial transformation values at the current resolution being used. The algorithm restarts until the absolute CC is above a threshold of 0.5 as determined later or the maximum number of restarts is reached. Absolute CC is used rather than MI because it has a well-defined range between 0 and 1 and because it provides an independent check of the MI result at the highest resolution.

We record all important results following an optimization cycle including the CC and/or MI values and the transformation parameters. At the end of processing at a lower resolution, we always select the transformation parameters having the maximum CC value. We then scale the translation parameters appropriately and assign the new parameters to be initial values at the next higher resolution. At the highest resolution, MI instead of CC is the similarity measure, and we select the final transformation parameters to be those having the maximum MI value.

Additional algorithm details are now described for slice-to-volume registration algorithm. For registration, we use rigid body transformation (three translations and three rotations) and trilinear interpolation. For optimization, we use the downhill simplex method of Nelder and Mead[20] or the Powell method.[21] Optimization of similarity ends either when the maximum number of calculations is reached (typically 500) or the fractional change in the similarity function is smaller than a tolerance (typically 0.001). The input MRI volume is a 3D MR acquisition giving $256 \times 256 \times 128$ nearly isotropic voxels over a field of view covering the whole pelvis. We create isotropic voxels of about 1.4 mm on a side using 3D linear interpolation. We use IDL (Interactive Data Language, Research System Inc., USA) as the programming language.

Typical parameter values are now described. We use an initial guess assuming an identity transformation, i.e. all initial translation and rotation parameters are zero, because the patient is normally oriented approximately the same way from one scan to the next. We set the maximum numbers of restarts at 10, 5, and 3, from low to high resolution, respectively.

2.3. *Registration of SPECT and high-resolution MRI volume*

The mutual information algorithm was used to register MRI and SPECT volume images because of its ability to align multi-modality images.[13] Registration of SPECT and MR images is challenging because the two image types have different spatial resolutions and image features. The radiotracer used for SPECT imaging was ProstaScint® (Cytogen Corporation, Princeton, NJ), a monoclonal antibody that binds to prostate-specific membrane antigen (PSMA). Before registration, both SPECT and MRI volumes were resized using trilinear interpolation to create volumes matrix of $128 \times 128 \times 128$ with 3 mm isotropic voxels, a voxel size between that of the two scans. The standard parameter set for automatic registration included: 256 intensity levels for each volume, the entire 2D joint histogram, the full field of view of $128 \times 128 \times 128$ voxels for both volumes, and no masking or cropping of either volume. Phantom data were preprocessed in a similar fashion.

3. Experimental Methods

3.1. *High-resolution MR image acquisition*

High-resolution MRI volumes were acquired using a 1.5 T Siemens MRI system (Magnetom Symphony, Siemens Medical Systems, Erlangen, Germany). An 8-element phased-array body coil was used to ensure coverage of the prostate with a uniform sensitivity. Typically two anterior and two posterior elements were enabled for signal acquisition. We used two different MR sequences.

First, we used a 3D rapid gradient echo sequence (PSIF) designed to acquire the spin-echo component of the steady state response, rather than the free induction decay. The spin echo component forms immediately prior to the RF pulse; it is

shifted toward the prior RF pulse through appropriate gradient waveform design. The sequence with 9.4/5.0/60 (TR/TE/flip) yields $160 \times 256 \times 128$ voxels over a $219 \times 350 \times 192$-mm rectangular FOV and $1.4 \times 1.4 \times 1.5$-mm voxels oriented to give the highest resolution for transverse slices. There is over sampling at 31% in the slice direction to reduce aliasing artifacts. The acquisition time is 4 min and 15 sec. This sequence gave excellent image contrast for the prostate and its surroundings. It was used to acquire volumes for three volunteers.

Second, we used a 3D RF spoiled gradient echo steady state pulse sequence (FLASH) with TR/TE/flip parameters of 12/5.0/60 which give $256 \times 256 \times 128$ voxels over a $330 \times 330 \times 256$-mm field of view (FOV) to yield $1.3 \times 1.3 \times 2.0$-mm voxels oriented to give the highest resolution for transverse slices. The acquisition time is 5 min and 38 sec. This sequence is good for pelvic imaging but is not ideal for the prostate. It was used to acquire volumes for five patients.

When acquiring high-resolution MR volumes, volunteers laid supine in a manner similar to the diagnostic position in routine MR scanning. Between volume acquisitions, volunteers got up from the MR table, stretched, and walked around to ensure that they would assume a different position when they laid back on the table. The coil array was centered on the prostate. We acquired three volumes from each of the three volunteers. For five patients, we acquired nine MRI volumes and each patient with at least one volume.

3.2. *Interventional MRI image acquisition*

We acquired iMRI images using a clinical 0.2 T C-arm open MR scanner (Siemens Open Symphony, Erlangen, Germany) modified for interventional MRI procedures and in this paper referred to as the iMRI system. We used a two-dimensional PSIF with 15.2/7.4/45 (TR/TE/FA) for image slice acquisitions. The iMRI slices were 128×128 with in-plane pixel size of 2.8×2.8 mm and with effective slice thickness of 5 mm.

We acquired iMRI images under the conditions simulating the treatment application. The volunteer was supine, and his legs were supported at 30°–60° relative to the horizon and separated in a "V" with an angle of 60°–90° between two legs. This is similar to the lithotomy position used in prostate therapies, and it should provide access for needle insertion in brachytherapy or RF thermal ablation. We call this the treatment position. For each of three volunteers, we acquired 50 iMRI image slices covering the prostate. They included 30 transverse, 10 coronal, and 10 sagittal image slices. We call these images "actual" iMRI images to differentiate them from "simulated" images as described in Sec. 3.4.

3.3. *SPECT image acquisition*

The study included five patients with either high Gleason scores (> 5) from biopsy or rising PSA level (> 10 mcg/L, prostate specific antigen) or palpation staging beyond stage $T1$. After patient eligibility was established, patients gave informed

consent. The Institutional Review Board of the University Hospitals of Cleveland approved the imaging protocol.

Approximately four days after injecting 5 mCi ProstaScint®, the abdominal and pelvic regions were scanned using a two-head Siemens E.CAM+ camera (Siemens Medical System, Inc., Hoffman Estates, Illinois, USA). ProstaScint® is an [In-111]-labeled monoclonal antibody capromab penditide (111In MoAb 7E11.C5) used for imaging prostate cancer. The evening before scanning, patients performed a bowel prep with Fleet R Prep Kit #3 (Fleet Pharmaceuticals, Lynchburg, VA). Images were acquired with a medium energy collimator and 15% energy window. The acquisition parameters included a step-and-shoot motion, a 128×128 pixel matrix for each projection, an imaging time of 25 sec per stop, and a total of 120 stops over a full 360° rotation. The field of view of was $53.3 \times 38.7\,\text{cm}^2$. The Ordered Subsets Expectation Maximization (OSEM) algorithm was used for image reconstruction.[43] SPECT images consisted of $4.795 \times 4.795 \times 4.795$-mm isotropic voxels. Each patient had one SPECT scan of the pelvis.

To analyze and validate registration of high-resolution MRI and SPECT under a controlled situation, an acrylic phantom of the pelvis and lower abdomen was used. Spheres of proportional size representing portions of the bladder, acetabula, rectum, and the prostate gland were placed in appropriate positions in the torso phantom. The spheres of acetabulum were filled with potassium phosphate. Other spheres were filled with water. The torso phantom was filled with a small amount of copper sulfate dissolved in deionized water. The SPECT scan was conducted after injecting all spheres with [In-111]-DTPA at relative concentrations comparable to those detected in human scans. The water in the torso was given a background activity of $1\,\mu\text{Ci/ml}$ such as to mimic the background in human SPECT scans.

3.4. *Simulation of iMRI image slices*

In experiments, we used high-resolution MRI volumes to simulate iMRI image slices, which are thicker, noisier, and degraded by receive coil inhomogeneity. Clinically, we typically use an iMRI slice thickness of 4.0–6.0 mm. We used trilinear interpolation to create isotropic high-resolution MRI volumes with voxel size of $1.4 \times 1.4 \times 1.4$ mm. From the isotropic high-resolution MRI volume, we averaged three 1.4 mm adjacent thin slices to create a 4.2 mm thick slice. MR noise in a magnitude is described by the Rician distribution.[22] At SNR values of greater than approximately five, the noise can be approximated as being Gaussian white noise.[23] We measured typical signal and noise values on our iMRI system using a homogenous phantom, and volunteer images in the region of the prostate with methods described elsewhere.[24–25] In all cases, image SNR was greater than 10 in all tissues including the prostate. With this justification, we added Gaussian noise to the simulated iMRI image slices either to match the measured SNR or to give much greater noise to further stress registration. We report noise experiments using the SNR of the simulated image

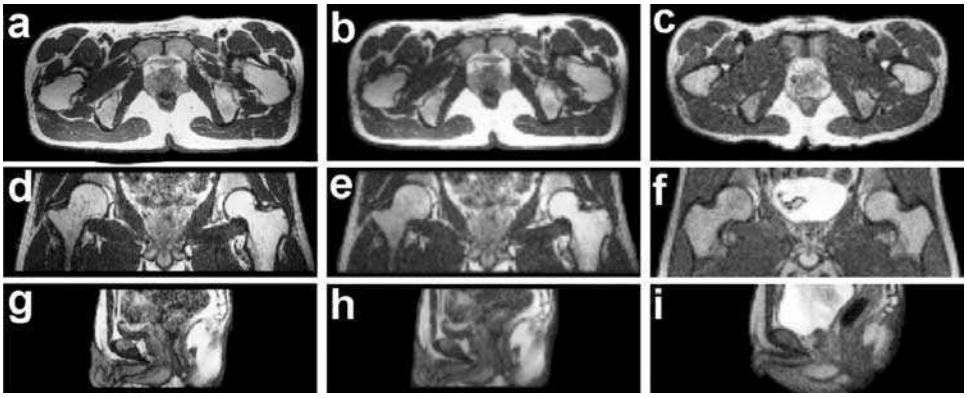


Fig. 3. High-resolution MR images, simulated and actual iMRI image slices. Images on the left column, (a), (d) and (g), are the original high-resolution MR images from the 1.5 T scanner in the transverse, coronal, and sagittal planes, respectively. Images in the middle column are the corresponding, simulated thick iMRI images with noise added to give SNR = 15 and with sensitivity fall off from a belt coil. Images on the right panel are actual iMRI slices (0.2 T scanner) from similar spatial locations. The actual iMRI slices seem blurred because of nearly doubled pixel size. Images are from volunteer $S2$.

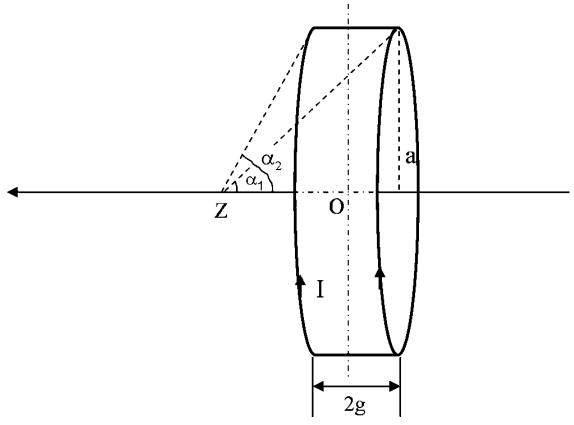


Fig. 4. Geometry of solenoidal receive coil. Model parameters are defined in the figure. The axial line is along the cranial-caudal direction of the patient.

slices. Figure 3 shows high-resolution MR images as well as simulated and actual iMRI image slices.

We simulated receive coil inhomogeneity from a belt coil used in our clinical iMRI acquisitions. The coil is modeled as a solenoid with parameters shown in Fig. 4. Coil parameters are a, the radius of the coil, $2g$, the length of the coil, I, the current, μ_0, the permeability of free space, n, the turns, and the z-axis, the

axis along the center line of the coil. The magnetic field in the xy plane can be approximated as[26]:

$$(B_1)_{xy} = \frac{\mu_0 n}{2} \frac{I}{[a^2 + g^2]^{1/2}} .$$

The z-component of the field is given by[27]:

$$(B_1)_z = \frac{nI}{2g}(\cos\alpha_1 + \cos\alpha_2)$$

where the definition of the angles α_1 and α_2 are given in Fig. 4. The magnetic field is highest at the coil center and falls off along the axial direction. According to the Biot-Savart law,[28] this model also accounts for the spatial sensitivity of the coil to MR signal sources. Figure 5 shows a coronal image with simulated inhomogeneity along the axis (head-foot) direction.

Because a needle will often be present during an iMRI intervention, we tested the effect of simulated needles on registration. We used artifact sizes from a previous report on the effects of pulse sequence design and magnetic field orientation on needle artifacts in MR-guided biopsy and aspiration.[29] Figure 6 shows sagittal images with and without a simulated needle artifact. The simulated artifacts in Fig. 6(b) appeared as straight noisy bars 2-mm in width.

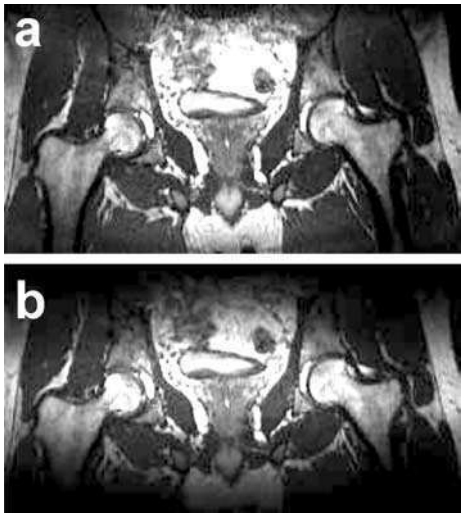


Fig. 5. Simulated signal changes due to receive coil inhomogeneity. The original image (a) is acquired using a phased array coil on a conventional 1.5 T MRI system. Using a belt coil model with a diameter of 350 mm and a width of 50 mm, the simulated iMRI image is shown in (b). The image intensity is highest at the center and decreases along the axial direction.

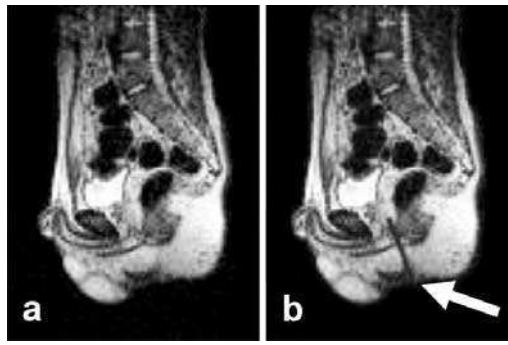


Fig. 6. Synthetic image with simulated needle artifact. Image (a) is the sagittal slice acquired from the 0.2 T iMRI system without a needle artifact. Image (b) is obtained from image (a) with a simulated needle artifact (white arrow) for an RF needle probe inserted into the prostate. Images are from volunteer S3.

3.5. *Registration experiments*

3.5.1. *Registration experiments using simulated iMRI images*

We used 12 pairs of high-resolution MR volumes to perform registration experiments. For each volume pair, we extracted data from one volume to simulate thick iMRI image slices; and then we registered the simulated image slices to the other volume. We desire an iMRI slice image acquisition method that gives robust, accurate registrations and is relatively insensitive to acquisition parameters. Hence, we performed experiments to determine the dependence on slice orientation (transverse, Sagittal, and coronal), on slice position relative to the prostate (above, centered, and below), on image noise from fast imaging techniques, and on the inhomogeneous sensitivity response from a belt coil.

3.5.2. *Registration experiments using actual iMRI image slices*

We also performed two types of SV registration experiments using the actual iMRI images. First, we registered actual iMRI image slices with high-resolution (1.5 T system) MR volumes and visually evaluated results. For each volunteer, there were three high-resolution MR volumes and 50 iMRI image slices giving 150 SV registration experiments, and a total of 450 experiments. Second, we registered thick slices simulated from the volume of image data obtained on the iMRI scanner with the corresponding high-resolution (1.5 T scanner) MR volume. In this case, we compared results to VV registration obtained by registering the volume from the iMRI system with the high-resolution volume (1.5 T scanner). We investigated the effect of iMRI slice thickness by averaging 1–10 contiguous image slices to create a thick slice and registering it to the high-resolution volume. The original actual iMRI volumes have a slice thickness of 1.4 mm and in-slice dimensions of 1.3×1.3 mm. We used trilinear interpolation to create isotropic actual iMRI volumes with voxel size of $1.3 \times 1.3 \times 1.3$ mm. Thus, thick slices simulated from actual iMRI volumes are 1.3 mm to 13 mm.

3.5.3. *Registration experiments of SPECT and high-resolution MRI*

A number of technical issues were examined for MI registration of MRI and ProstaScint® SPECT prostate images. Firstly, MRI acquisition, by varying the MR imaging pulse sequence, various structures can be emphasized or suppressed. Several different acquisition sequences were tested and its effect on registration accuracy and robustness was determined. Secondly, because of the different dynamic ranges between MR and SPECT images, intensity scaling was studied for its effect on registration. This is prompted by a recent study showing that scaling images to 16 gray levels gives better results than 256 gray levels when registering muscle fiber images.[44] Thirdly, because of the sparseness in the histogram, the use of a portion or a section rather than the full joint histogram was evaluated. This effectively restricted the registration to particular intensity ranges. Fourthly, the multi-resolution approach was examined for its ability to expedite the automated search algorithm. Fifthly, the use of spatial masking was investigated to see whether it facilitates the registration of partially overlapping volumes. In all cases, registration experiments were performed with and without these modifications to determine their effect on the success of registration. Success was determined by comparing the results of these experiments to those of manual registration of the same images as described in the next section. Experiments with these parameters should provide insight into improving registration of MR and SPECT prostate images. We performed registration experiments using the SPECT and MRI image volumes from five patients.

3.6. *Registration evaluation*

3.6.1. *Visual inspection*

We evaluated registration experiments by visual inspection. We used RegViz, a program created in IDL in our laboratory with multiple visualization and analysis methods. First, we manually segmented prostate boundaries in image slices and copied them to corresponding slices. This enabled visual determination of the overlap of prostate boundaries over the entire volume. Second, color overlay displays were used to evaluate overlap of structures. One image was rendered in gray and the other in the "hot-iron" color scheme available in IDL. To visualize potential differences, it was quite useful to interactively change the contribution of each image using the transparency scale. Third, we used a sector display, which divided the reference and registered images into rectangular sectors and created an output image by alternating sectors from the two input images. Even subtle shifts of edges would be clearly seen.

3.6.2. *Volume-to-volume registration standard*

Our standard evaluation method for slice-to-volume registration was to compare SV and VV registration. The VV registration accuracy was previously evaluated.[8] For volume pairs acquired over a short time span from a supine subject with legs flat on

the table, prostates were well aligned and prostate centroid displacements were typically < 1 mm. The registration accuracy as determined from displacements of pelvic bony landmarks was $1.6\,\text{mm} \pm 0.2\,\text{mm}$. This error might be overestimated because it includes the uncertainty of locating the bony landmarks. From our success with VV prostate registration, we decided that we could obtain SV accuracy by comparing to VV registration for those volume pairs having low VV registration error.

To compare SV and VV registration results, we defined a rectangular volume of interest (VOI) just covering the prostate over which to calculate registration error. To voxels within the VOI, we applied the transformations obtained by the VV and by SV registrations. We then calculated the 3D displacements between the transformed voxels. The mean voxel distance over the VOI was used as our metric of SV registration error. For the evaluation of algorithm robustness, we defined the SV registration as being successful when the mean 3D displacement was less than 2.0 mm.

3.6.3. *Evaluation of SPECT and MRI registration*

The success of computer registration of SPECT and MRI volumes was determined by comparing results to manual registration. Manual registration was done by two board-certified nuclear medicine radiologists blinded to the automatic registration results. Manual registration was done using a software package with a graphical user interface (GUI) developed in–house, which allows graphical manipulation of volumes with six degrees of freedom in a rigid body registration. A color overlay was used to assess registration quality.

Two radiologists with a nuclear medicine specialty aligned the image volumes, and whenever there was a discrepancy, they reached a consensus for a single transformation. This painstaking cross-validation was a time-consuming process and certainly would not be a routine procedure, but the results served as the gold standard for the automated method. We defined a successful automatic registration to be obtained when all displacements were < 2 voxels (6 mm) in the x, y, and z directions and angle differences were < 2 degree for all angles about each of the three axes.

Although manual registration is difficult and somewhat operator dependent, it is the only acceptable option for an independent registration on the patient SPECT and MRI volumes. Skin fiducials would be of limited value in the pelvis, and there are no good identifiable point anatomical landmarks in the SPECT images.

We simulated the iMRI-guided procedures using our image registration and fusion software that are specially designed for this application. Before treatment, we acquired SPECT and high resolution MRI volumes from the same patients. Second, we registered the two images and transferred the pair of aligned data sets to a workstation that was used for the slice to volume registration. Third, we connected the workstation to the iMRI scanner and obtained iMRI image slices from the scanner. Fourth, we performed the slice to volume registration. Finally, the software created fused images of the three modalities as would be done for image guidance. All registrations and image fusions are automatic.

4. Results

4.1. *Registration results with simulated iMRI images*

As described in Sec. 3.4, we obtained relatively low noise, high-resolution MR images and simulated SV registration results. These data sets allowed us to test effects of noise and receive coil inhomogeneity in a controlled fashion. And, because we had substantial previous experience showing the accuracy of VV registration under comparable conditions, we could easily determine SV error by comparing results to VV registration.

In Fig. 7, the sector display shows a simulated image slice registered with a high-resolution image volume. The simulated image slice was obtained at a transverse orientation near the center of the prostate. The sector display shows close alignment at this position. Other transverse images were also well aligned indicating that the registration was successful in three dimensions.

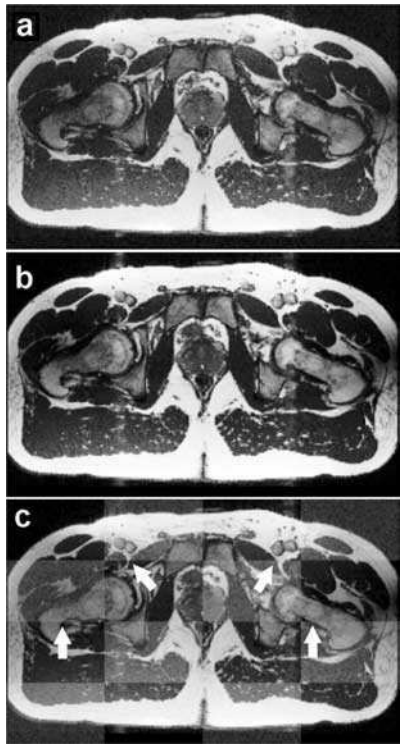


Fig. 7. Sector display showing quality of SV registration. Transverse slices are shown for simulated iMRI (a) and high-resolution MRI (b) images. In the sector display (c), a checker board pattern is created where image sections from (a) and (b) are alternated. Square sections from (a) are made brighter in order to show the boundaries. As indicated by the arrows, the boundaries of bones and other structures are continuous across the sections indicating excellent registration. The prostate registered very well.

We determined SV registration results for slices near the prostate in the three standard orthogonal orientations. Comparing to VV, mean and standard deviation registration errors across 12 volume pairs and 60 SV registration experiments were $0.4\,\text{mm} \pm 0.2\,\text{mm}$, $0.5\,\text{mm} \pm 0.2\,\text{mm}$, and $2.6\,\text{mm} \pm 1.6\,\text{mm}$ for transverse, coronal and sagittal slices covering the prostate, respectively. Transverse slices worked best because they contain many relatively rigid anatomical structures (See Fig. 3). We further found that transverse slices centered on the prostate produced better results than those above or below the prostate. Image slices above included the deformable bladder that could give an inconsistent structure from one volume to the next. Image slices below the prostate mainly contained muscle and fatty regions from the hips that could deform, again giving inconsistent image data. Coronal slices worked next best. Sagittal slices gave the largest error because they contained a large portion of the deformable bladder and rectum.

Simulation experiments showed SV registration to be very insensitive to noise. We performed over 150 registration experiments with noise added to give SNR's ranging from 20 to 5. Using the slice configurations recommended above (transverse slices near the prostate center), we obtained 100% successful registrations (an error $< 2.0\,\text{mm}$) for SNR's ≈ 10, a value much worse than the clinical SNR value of ≈ 25 on our iMRI system.

Receive coil inhomogeneity also had little effect on registration. Registration again was 100% successful for all volume pairs under all receive coil configurations, even when the coil for the slice acquisition was displaced up to 200 mm towards the head from the prostate center, the position of the coil for the volume acquisition.

4.2. *Registration results with actual iMRI images*

Figure 8 shows results for an SV registration of actual iMRI image slices with a high-resolution MR volume. The contours overlap and overlay images show that the prostate matches very well. Other visual inspection techniques also demonstrate excellent registration. Note that a single iMRI image was used to produce this registration result.

Figure 9 shows SV registration error as a function of slice thickness. As described previously, we first registered each volume from the iMRI scanner with the corresponding high-resolution MRI volume (1.5 T scanner) using rigid body voxel-based registration[8] and used the result as the gold standard for calculating the SV error. Each thick slice image was obtained by averaging several contiguous slices from the actual iMRI volume. As the slice thickness increases from $1 \times 1.3\,\text{mm}$ to $4 \times 1.3\,\text{mm}$, the registration error decreases, possibly because of improved signal to noise ratio and/or because of the inclusion of more features. Error increases with thicker slices, probably because of the inconsistency of image features between the thick slice and more finely sampled volume.

In Fig. 10, we evaluated SV registration for thick slices at different orientations. The evaluation method was the same as that used in Fig. 9, and the slices were 5 mm

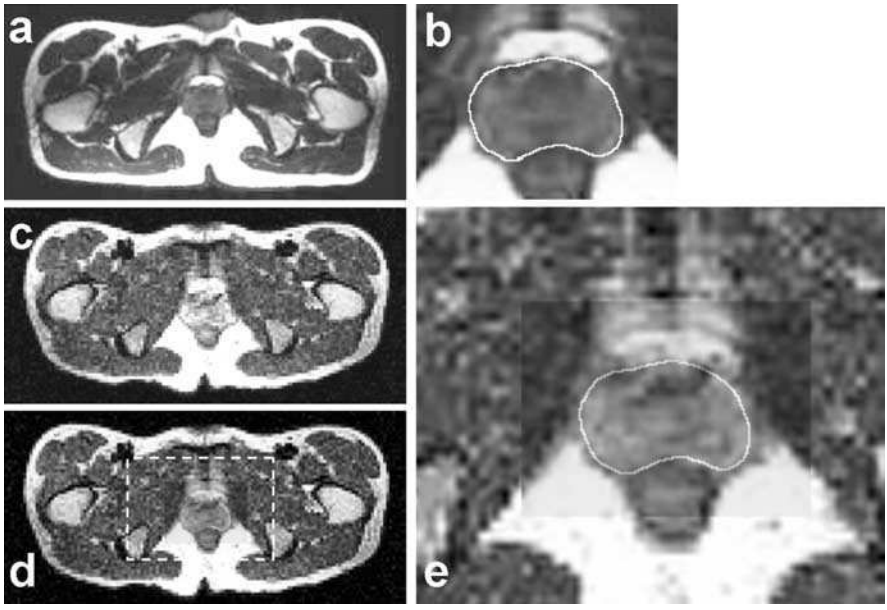


Fig. 8. Images after SV registration of actual iMRI slices from a 0.2 T open MR system. Image (a) is a transverse slice from a high-resolution MR volume (1.5 T scanner). The prostate is segmented and magnified in image (b). Image (c) is the actual iMRI slice (0.2 T scanner). Images (c) and (b) are displayed together in an overlay in image (d), and the white rectangular region is magnified in image (e). The segmented prostate boundary from the high resolution MR image is copied to the actual iMRI image where it closely matches the prostate in the actual iMRI image slice indicating excellent registration.

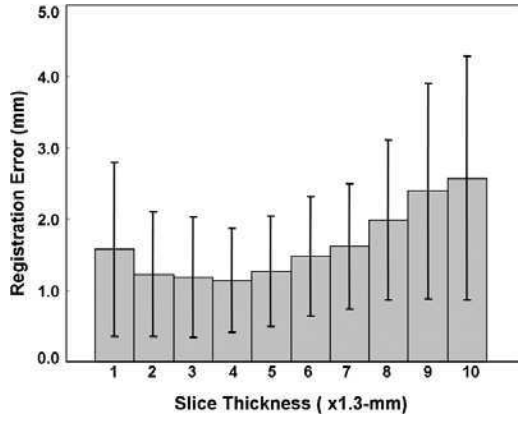


Fig. 9. SV registration using images with different slice thickness. The error metric is the average voxel displacement between the SV and VV registrations. Plotted are mean errors as well as standard deviation from a rectangular VOI surrounding the prostate. One typical data sets of high-resolution MRI volume and actual iMRI slices of volunteer $S1$ are used for the registration experiments. For each thickness, ten registration experiments were conducted using 10 different simulated iMRI transverse slices that intersected the prostate with different distances. Thick iMRI slices were obtained by averaging 1–10 iMRI image slices.

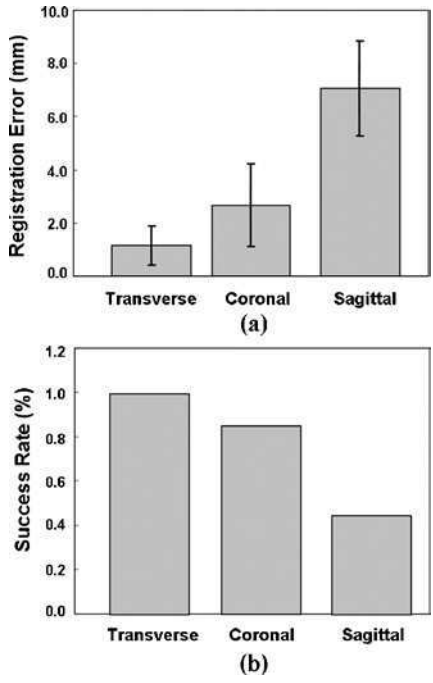


Fig. 10. SV registration error and robustness for iMRI images in the three standard orientations. In (a), registration error relative to VV registration is plotted as a function of image slice orientation. In (b), success rate is also plotted as a function of orientation where registration is successful when the error is < 2.0 mm. For volunteer $S2$, one high-resolution volume and one volume from the iMRI scanner were used in these experiments. Data were extracted from the iMRI volume to simulate iMRI slices with a thickness of about 5 mm. Fifteen transverse, coronal, and sagittal slices from the prostate center were used for SV registration, respectively.

thick and intersected the volume near the prostate center. Results were consistent with those from the previous simulation experiments. Transverse slices worked best with an average VOI displacement of only $1.1\,\text{mm} \pm 0.7\,\text{mm}$ and a success rate of 100%. The coronal images gave a reasonable average error, but the success rate dropped to 86%. The sagittal orientation gave the worst result.

Needle artifacts had little effect on the SV registration. In each of the 30 experiments, we registered a high-resolution volume with an actual iMRI image slice containing or not containing a simulated needle artifact. Visual inspection, the correlation coefficient and mutual information values of registered images showed little effect of the needle artifact. The success rate was 100% in both cases.

4.3. *Registration results of SPECT and high-resolution MRI volumes*

An example of a successful automatic registration is shown in Fig. 11. All anatomical features including the bone marrow in the femur and pubic symphysis are well

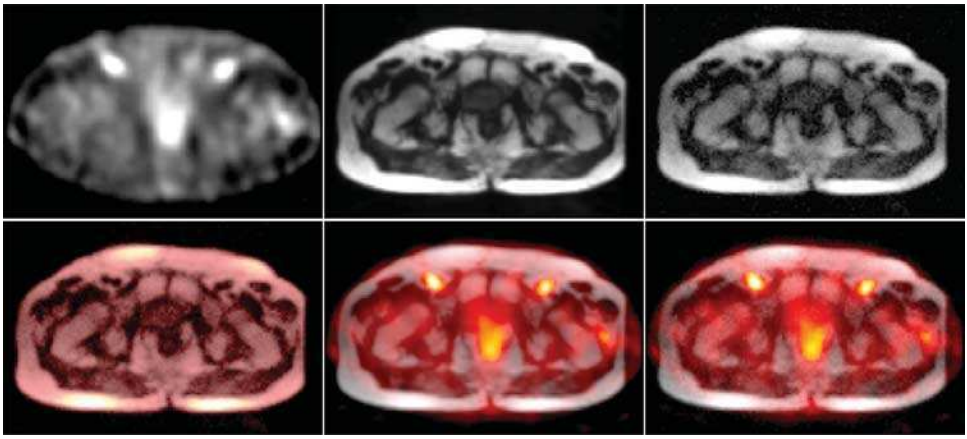


Fig. 11. Registration results of patient data. The top three images show corresponding registered SPECT, high resolution MRI, and simulated iMRI images, respectively. The bottom three windows show the fused images of the three modalities, from left to right, iMRI/MRI, SPECT/MRI, and SPECT/iMRI, respectively.

aligned in the color overlay. This MR-SPECT volume pair and four others were successfully registered according to the criteria defined earlier. Standard algorithm parameters (Sec. 2.3) were used with the lower-left quadrant of the joint histogram used for calculating MI. Successful image registration was obtained with images from three patients. There were four other MR-SPECT volume pairs obtained from two other patients that were not successfully registered with our program. In all the four cases, the MR images were not acquired using our final, optimized MR sequence (Sec. 3.1). When we used the optimized sequence with full anatomical coverage, registration was always successful. We believe that automated SPECT-MRI registration will be feasible on many patients' images.

We now report the registration results of SPECT and high resolution MRI images of the phantom. Registrations of the phantom images were carried out by displacing the aligned image pair with known rotation and translations. All orientations, axial, sagittal, and coronal, were successfully registered. Other experiments showed that intensity scaling and multi-resolution could not improve the registration ability for both phantom and human data.

4.4. *Image fusion and visualization*

We created image registration and fusion software for the potential applications in iMRI-guided procedures. In Fig. 12, we demonstrate the image fusion visualization software in a simulation of clinical usage. SPECT and high resolution MR images were acquired, transferred to a workstation, and registered prior to the "simulated" procedure. We then simulate acquiring thick iMRI slices, register them to the high resolution volume, and prepare the visualization in Fig. 12. In this figure, one can see

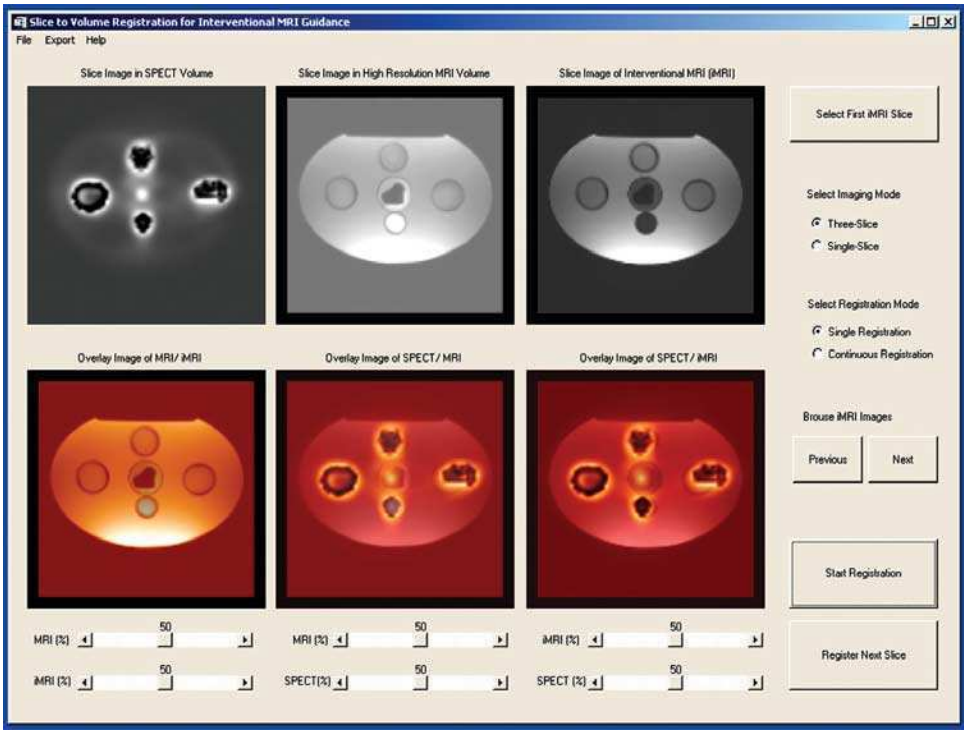


Fig. 12. Simulation experiments with phantom using registration and fusion software. The top three windows from left to right show corresponding registered SPECT, high resolution MRI, and iMRI images, respectively. The bottom three windows from left to right show the fused images, iMRI/MRI, SPECT/MRI, and SPECT/iMRI, respectively. Other buttons and sliders control the configuration and registration.

all. The registered images are shown in the three windows at the top line (Fig. 12). After registration, the program creates fused images as displayed at the bottom.

4.5. *Algorithmic robustness and implementation*

The slice-to-volume registration algorithm was quite robust for transverse slices covering the prostate. Using simulated iMRI slices from high-resolution MRI volume pairs of four volunteers, the algorithm never failed for any transverse slice covering the prostate. In addition, the final registration result was insensitive to initial guesses within a very large range, $[-60, +60]$ mm for translations and $[-20, +20]$ degrees for rotations. With the restarting algorithm, we even successfully registered slices as much as 80 mm from the optimum. This working range should be quite sufficient for clinical applications where we can ensure good starting values. Using the pelvic bones as markers and device localization methods,[29] we should be able to position the prostate within about ± 20 mm in the imaging field. In addition, the patient normally lies supine in the MR bed with very little rotation ($< \pm 5$ degrees).

Using CC and MI at different resolutions was an important feature that increased robustness. MI registrations at low resolution sometimes gave false maxima (Figs. 1(a) and 1(c)), and only 60% success was achieved when MI was used at all resolutions. The interpolation artifacts at low resolutions often caused failures and required more restarts.[16] CC performed well and gave fewer local maxima at the lower resolutions (Figs. 1(b) and 1(d)), but MI was more accurate than CC at the highest resolution due to the sharper peak of the MI surface (Figs. 2(a) and 2(c)).[8] Our registration algorithm thus combined advantages from the two similarity measures.

The multi-resolution approach improved algorithmic robustness and speed. When we used only MI at full resolution, registration was 70% successful compared to the 100% of the full algorithm. This failure of MI was also reported by others.[13,17] The multi-resolution approach enabled the program to quickly approach the final value because of the reduced number of calculations at low resolutions. For a typical image pair, iterations at 1/4 resolution were approximately 4 and 25 times faster than that at 1/2 and full resolution respectively.

Restarting was important for image pairs with large translations and/or rotations from the optimum. In our experience with over 800 slice-to-volume registration experiments, restarting occurred in about 5% of them. For an example pair with an 80 mm displacement, the number of restarts was 3, 1, and 0 at 1/4, 1/2, and full resolutions, respectively. Without restarting, we found that registrations sometimes failed in cases of volumes with a large mismatch of 54 mm and high noise. The algorithm was insensitive to the CC threshold for restarting. When we decreased the threshold from 0.8 to 0.5 with an interval of 0.05, we found little change in the number of restarts and no change in the final registrations. We set the threshold at 0.5 to avoid only the most obvious local maxima.

We now describe some aspects of the implementation. The time for an SV registration was typically about 15 sec on a Pentium IV, 1.8 GHz CPU, with 1 Gbytes of memory. The algorithm was written in IDL and could probably be made much faster in a lower level language such as C. A call to the Simplex optimization typically resulted in 50 to 105 similarity evaluations before the tolerance value (0.001) was reached. The simplex optimization method worked about 1.5–2.0 times faster than the Powell method in our implementation. We used the Simplex method for our experiments in this study.

5. Discussion and Conclusion

5.1. *Applicability for iMRI guidance*

This preliminary study has shown promising algorithmic results for bringing nuclear medicine, functional images into the interventional MRI suite. Automatic registration of SPECT and MRI volumes was always successful with "good" MRI volumes obtained using the optimized acquisition sequence and covering all anatomy of interest. Slice-to-volume automatic registration was even more successful with

highly accurate, robust registration obtained. Putting these two steps together, a patient's SPECT images can be registered to a high resolution MRI volume prior to an iMRI procedure; live-time iMRI slice images can be registered to the MRI volume; and, finally, one can then display the live-time iMRI slice image with the appropriately reformatted, fused image from the SPECT and high resolution MRI image volumes.

The required registration accuracy is lowered because live-time images are available. The live-time iMRI image obtained in the plane of the advancing needle will always be used for guiding a needle for intervention or biopsy. The corresponding fused SPECT-MRI and/or high resolution MRI images will be used as a planning guide. With proper visualization tools, interventional radiologists should be able to mentally account for small registration errors. In addition, there is often image evidence of cancer in MR prostate images that can perhaps be identified with the aid of the functional images. Such MR-visible lesions can then become the markers for tumor targeting. Any potential gross registration errors should be easily recognized resulting in a failure to include the functional image data in the iMRI suite but not in a catastrophic misguidance of the therapy needle.

To minimize registration error, we recommend that image data are obtained under comparable conditions by keeping a similar posture and by taking clinical measures to reduce rectal and bladder filling. Warping registration method may be useful to correct significant deformations at the expense of additional complexity, time, and possibly robustness.[38,39]

Finally, we believe that it is quite feasible to include previously acquired nuclear medicine SPECT images and high-resolution MRI data into iMRI-guided minimally invasive treatment procedures. We are beginning to explore this application in animal experiments.

5.2. *Robustness of slice-to-volume registration*

Despite complications such as image noise, receive coil inhomogeneity, a limited number of voxels, and needle artifacts, slice-to-volume voxel-based registration can be quite robust and accurate. For transverse slices covering the prostate, registration results agreed very favorably with volume-to-volume results. Below, we further discuss the algorithm and its practicality.

There are probably several reasons why mutual information does not work well at low resolution. First, the similarity curve is noisy with periodic oscillations from the so-called interpolation artifact[8,16] that is accentuated at reduced resolutions.[30] As a result, there are many local maxima in Figs. 1(a) and 1(c) that can trap the optimization; and a similar result was reported for brain registration.[13] In additional experiments, we decreased the number of bins for both images to 256, 128, 64 and 32 and plotted mutual information values as a function of translation. With a larger number of bins, we got no discernable effect of bin size. When the number of bins was reduced to 32, the MI surface was degraded. Others showed that Gaussian blurring of images before registration did not improve performance at low

resolutions and that there was little difference between standard and normalized mutual information.[40] Second, when images are of low resolution and there is only a small region of overlap, the mutual information function can even contain incorrect global maxima[30] as found in Fig. 1(a). This false result was obtained at very large displacements where the slice to volume overlap was reduced. This occurs because MI is not only a function of how well the images match in the overlap, but also by how much information is provided by the two images in the overlap.[31,32,35] As shown above, using both mutual information and correlation coefficient at different resolutions was an important feature that increased robustness.

5.3. *Registration accuracy*

Essentially, we found that SV is of similar accuracy to VV registration, with an average voxel displacement difference of only 0.4 mm in the prostate for the simulated images and about 1 mm for actual iMRI image data. Hence, the accuracy of the best SV method is essentially the same as that previously reported for VV registration.[8]

More analysis of registration error is possible. The overall registration error of placing a SPECT image with a live-time iMRI image depends upon both SPECT-MRI and MRI-MRI slice-to-volume errors. The slice-to-volume registration error for voxels near the prostate is ≈ 1.4 mm, as argued elsewhere.[41,45] The SPECT-MRI registration error is larger than the SV registration as would be expected from the low resolution and reduced number of features with SPECT. The error is comparable to the uncertainty of manual registration (± 6 mm and ± 2 degrees). (After all, that is how we specified the requirements for "acceptability.") Despite such uncertainty, ProstaScint SPECT images have been routinely registered with CT and MR images at our institution to use for diagnostic studies of prostate cancer. Hence, we predict the overall error for registering live-time iMRI slices to SPECT to be dominated by the SPECT/high-resolution MRI error.

The automatic slice-to-volume registration provides sufficient accuracy for many potential iMRI applications. As compared to a typical SPECT and/or iMRI slice thickness of ≥ 3.0 mm, SV registration is quite accurate. MR spectroscopy also is done at limited resolution. If one were to use functional or high-resolution MR images directly for targeting, the requirements for registration accuracy would be great. However, fused image data will not be used blindly. Rather, these visualizations will be used as a guide. Physicians will always use the live-time iMRI images for needle guidance. With the aid of visualization tools, they should be able to account for small registration errors.

We recommend that image data are obtained under comparable conditions by keeping a similar posture and by taking clinical measures to reduce rectal and bladder filling. We see no reason to suspect that SV registration will be inaccurate when such conditions are met. When images were acquired under much different

conditions, such as legs flat and legs raised, rigid body registration could result in prostate centroid errors as much as 3.4 mm. Another effect may be the tissue deformation from insertion of the RF needle. From our previous experience observing *in vivo* needle insertion in both animal models and clinical trials with real-time MRI, the amount of tissue deformation that occurs with insertion of a sharp bevel tip needle is minimal and transient in tissues with normal interstitial pressure. In certain lesions, such as cysts or necrotic tumor, persistent deformation is possible; however, we can see such deformations in the live-time interventional MRI images and very probably mentally correct the registered, fused images. We previously reported a warping registration method[38,39] that can correct deformations at the expense of additional complexity, time, and possibly robustness.

5.4. *Practicality and application*

The registration experiments presented here provided fairly comprehensive tests for the potential application in iMRI-guided RF thermal ablation of the prostate. Simulation provided an efficient way to extensively evaluate registration performance. The algorithm was extremely robust to noise levels, far beyond those encountered in clinical iMRI applications. Similarly, the inhomogeneity seen with a belt coil was not problematic for transverse images, probably due to coil inhomogeneity simply scaling the grayscale values, an operation that should not affect MI or CC similarity measures. Needle artifacts had little effect, probably because they occupy relatively few voxels. The actual iMRI images acquired under more realistic conditions further tested practicality. Images from the iMRI system contained more noise and had less contrast than those from the 1.5 T scanner. Registration quality was comparable to that of simulation experiments. Registration time can probably be improved considerably using optimized C code rather than IDL. If registration is done in the background in a seamless way, the time for registration is probably quite acceptable. Although we normally used $T2$-weighted image pairs, the registration worked well for pairs of $T1$-weighted and $T2$-weighted images.

We conclude that the automatic slice-to-volume registration algorithm is quite robust for transverse image slices covering the prostate and that the registration provides sufficient accuracy to aid image-guided therapy. From previous reports of MR-PET or MR-SPECT registration accuracy,[6,7] it appears feasible to combine functional images to aid iMRI-guided procedures. We are beginning to explore this application in animal experiments.

Acknowledgments

The algorithm developed in this research was supported by DOD grant DAMD17-02-1-0230 to Baowei Fei and NIH grant R01-CA84433-01 to David L. Wilson. Imaging techniques were developed under the support of NIH grant R33-CA88144-01 to Jeffery L Duerk.

References

1. J. S. Lewin, C. F. Connell, J. L. Duerk, Y. C. Chung, M. E. Clampitt, J. Spisak, G. S. Gazelle and J. R. Haaga, Interactive MRI-guided radiofrequency interstitial thermal ablation of abdominal tumors: Clinical trial for evaluation of safety and feasibility, *J. Magn. Reson. Imaging* **8** (1998) 40–47.
2. D. L. Wilson, A. Carrillo, L. Zheng, A. Genc, J. L. Duerk and J. S. Lewin, Evaluation of 3D image registration as applied to MR-guided thermal treatment of liver cancer, *J. Magn. Reson. Imaging* **8** (1998) 77–84.
3. A. Carrillo, J. L. Duerk, J. S. Lewin and D. L. Wilson, Semiautomatic 3-D image registration as applied to interventional MRI liver cancer treatment, *IEEE Trans. Med. Imag.* **19** (2000) 175–185.
4. D. B. Sodee, N. Malguria, P. Faulhaber, M. I. Resnick, J. Albert and G. Bakale, Multicenter ProstaScint imaging findings in 2154 patients with prostate cancer, *Urology* **56** (2000) 988–993.
5. J. Scheidler, H. Hricak, D. B. Vigneron, K. K. Yu, D. L. Sokolov, L. R. Huang, C. J. Zaloudek, S. J. Nelson, P. R. Carroll and J. Kurhanewicz, Prostate cancer: Localization with three-dimensional proton MR spectroscopic imaging — Clinicopathologic study, *Radiol.* **213** (1999) 473–480.
6. D. B. Sodee, G. T. MacLennan, M. I. Resnick, P. F. Faulhaber, Z. Lee, A. D. Nelson, J. P. Molter and G. Bakale, Comparison of CT- or MRI-fused PET-FDG and SPECT-ProstaScint (R) imaging of prostate cancer with the gold standard of histology, *J. Nucl. Med.* **42** (2001) 1222.
7. Z. Lee, D. B. Sodee, J. L. Duerk, A. D. Nelson and M. S. Berridge, Automatic registration of SPECT-MRI in the pelvis, *J. Nucl. Med.* **41** (2000) 232.
8. B. W. Fei, A. Wheaton, Z. Lee, J. L. Duerk and D. L. Wilson, Automatic MR volume registration and its evaluation for the pelvis and prostate, *Phys. Med. Biol.* **47** (2002) 823–838.
9. H. Gray, *Anatomy, Descriptive and Surgical* (The classic collector's edition) (Gramercy Books, New York, 1977), p. 1010.
10. B. Kim, J. L. Boes, P. H. Bland, T. L. Chenevert and C. R. Meyer, Motion correction in fMRI via registration of individual slices into an anatomical volume, *Magnetic Resonance in Medicine* **41** (1999) 964–972.
11. T. S. Kim, M. Singh, W. Sungkarat, C. Zarow and H. Chui, Automatic registration of postmortem brain slices to MRI reference volume, *IEEE Trans. Nucl. Science* **47** (2000) 1607–1613.
12. J. Zhengping and P. H. Mowforth, Mapping between MR brain images and voxel model, *Medical Informatics* **16** (1991) 183–193.
13. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, Multimodality image registration by maximization of mutual information, *IEEE Trans. Med. Imag.* **16** (1997) 187–198.
14. Y. M. Zhu, Volume image registration by cross-entropy optimization, *IEEE Trans. Med. Imag.* **21** (2002) 174–180.
15. R. Shekhar and V. Zagrodsky, Mutual information-based rigid and nonrigid registration of ultrasound volumes, *IEEE Trans. Med. Imag.* **21** (2002) 9–22.
16. J. P. W. Pluim, J. B. A. Maintz and M. A. Viergever, Interpolation artefacts in mutual information-based image registration, *Comput. Vis. Image Understanding* **77** (2000) 211–232.
17. D. L. G. Hill, L. A. Langsaeter, P. N. Poynter-Smith, P. E. Summers, S. F. Keevil, R. Walsh, D. J. Hawkes and M. J. Gleeson, Feasibility study of magnetic resonance imaging-guided intranasal flexible microendoscopy, *Computer Aided Surgery* **2** (1997) 264–275.

18. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens and G. Marchal, Automated multimodality image registration using information theory, Information Processing in Medical Imaging: *Proc. 14th International Conference (IPMI'95), Computational Imaging and Vision*, 1995, pp. 287–298.
19. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vellerling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd edn. (Cambridge University Press, New York, 1993).
20. J. Nelder and R. A. Mead, A simplex method for function minimization, *Comp. J.* **7** (1965) 308–313.
21. M. J. D. Powell, An iterative method for finding stationary values of a function of several variables, *Comp. J.* **5** (1962) 147–151.
22. A. Macovski, Noise in MRI, *Magnetic Resonance in Medicine* **36** (1996) 494–497.
23. R. C. Gregg and R. D. Nowak, Noise removal methods for high resolution MRI, *IEEE Nuclear Science Symposium* **2** (1997) 1117–1121.
24. L. Kaufman, D. M. Kramer, L. E. Crooks and D. A. Ortendahl, Measuring signal-to-noise ratios in MR imaging, *Radiol* **173** (1989) 265–267.
25. R. M. Henkelman, Measurement of signal intensities in the presence of noise in MR images, *Med. Phys.* **12** (1985) 232–233.
26. D. I. Hoult and R. E. Richards, The signal-to-noise ratio of the nuclear magnetic resonance experiment, *Journal of Magnetic Resonance* **24** (1976) 71–85.
27. 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).
28. L. K. Forbes, S. Crozier and D. M. Doddrell, Rapid computation of static fields produced by thick circular solenoids, *IEEE Trans. Magnetics* **33** (1997) 4405–4410.
29. J. S. Lewin, J. L. Duerk, V. R. Jain, C. A. Petersilge, C. P. Chao and J. R. Haaga, Needle localization in MR-guided biopsy and aspiration: Effects of field strength, sequence design, and magnetic field orientation, *American Journal of Roentgenology* **166** (1996) 1337–1345.
30. J. P. W. Pluim, J. B. A. Maintz and M. A. Viergever, Image registration by maximization of combined mutual information and gradient information, *IEEE Trans. Med. Imag.* **19** (2000) 809–814.
31. C. Studholme, D. L. G. Hill and D. J. Hawkes, Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures, *Med. Phys.* **24** (1997) 25–35.
32. D. Rueckert, M. J. Clarke, D. L. G. Hill and D. J. Hawkes, Non-rigid registration using higher-order mutual information, ed. K. M. Hanson, *Proc. SPIE Medical Imaging 2000: Image Processing* **3979** (2000) 438–447.
33. F. Maes, Segmentation and registration of multimodal images: From theory, implementation and validation to a useful tool in clinical practice, PhD Thesis, Katholieke Universiteit Leuven, Belgium, 1996.
34. C. Studholme, D. L. G. Hill and D. J. Hawkes, An overlap entropy measure of 3D medical image alignment, *Pattern Recognition* **32** (1999) 71–86.
35. P. A. Viola, Alignment by maximization of mutual information, PhD Thesis, Massachusetts Institute of Technology, 1995.
36. J. M. Fitzpatrick, J. B. West and C. R. Maurer, Predicting error in rigid-body point-based registration, *IEEE Trans. Med. Imag.* **17** (1998) 694–702.
37. M. V. Herk, A. Bruce, A. P. G. Kroes, T. Shouman, A. Touw and J. V. Lebesque, Quantification of organ motion during conformal radiotherapy of the prostate by three dimensional image registration, *Int. J. Radiat. Oncol. Biol. Phys.* **33** (1995) 1311–1320.

38. B. W. Fei, C. Kemper and D. L. Wilson, Three-dimensional warping registration of the pelvis and prostate, eds. M. Sonka and J. M. Fitzpatrick, *Proc. of SPIE Medical Imaging 2002: Image Processing* **4684** (2002) 528–537.
39. B. W. Fei, C. Kemper and D. L. Wilson, A comparative study of warping and rigid body registration for the prostate and pelvic MR volumes, *Computerized Medical Imaging and Graphics* **27** (2003) 267–281.
40. J. P. W. Pluim, J. B. A. Maintz and M. A. Viergever, Mutual information matching in multiresolution contexts, *Image and Vision Computing* **19** (2001) 45–52.
41. B. W. Fei, J. L. Duerk, D. T. Boll, J. S. Lewin and D. L. Wilson, Slice to volume registration and its potential application to interventional MRI guided radiofrequency thermal ablation of prostate cancer, *IEEE Transactions on Medical Imaging* **22**, 4 (2003) 515–525.
42. B. W. Fei, Z. Lee, J. L. Duerk, J. S. Lewin, D. B. Sodee and D. L. Wilson, Registration and fusion of SPECT, high resolution MRI, and interventional MRI for thermal ablation of the prostate cancer, *IEEE Transactions on Nuclear Science* **51**, 1 (2004) 177–183.
43. H. M. Hudson and R. S. Larkin, Accelerated image-reconstruction using ordered subsets of projection data, *IEEE Transactions on Medical Imaging* **13** (1994) 601–609.
44. B. Likear and F. Pernus, Registration of serial transverse sections of muscle fibers, *Cytometry* **37** (1999) 93–106.
45. B. W. Fei, J. L. Duerk and D. L. Wilson, Automatic 3D registration for interventional MRI-guided treatment of prostate cancer, *Computer Aided Surgery* **7** (2002) 257–267.

CHAPTER 9

DETECTION AND SEGMENTATION OF DRUSEN DEPOSITS ON RETINA IMAGES

RAPANTZIKOS KONSTANTINOS
Department of Electrical & Computer Engineering
National Technical University of Athens
Athens, GR-73100, Greece
rap@image.ece.ntua.gr

ZERVAKIS MICHALIS* and BALAS KONSTANTINOS†
Department of Electronic & Computer Engineering
Technical University of Crete
Chania, GR-73100, Greece
**michalis@danai.systems.tuc.gr*
†*balas@electronics.tuc.gr*

Computer assisted prognosis and diagnosis of retinal diseases gained the interest of many researchers during the last years. Much attention has been drawn on Age-related Macular Degeneration (AMD), which is the leading cause of irreversible vision among the elderly in developed countries. Assessment of the risk for the development of AMD requires reliable detection and quantitative mapping of retinal abnormalities that are considered as precursors of the disease. Typical signs for the latter are the so-called drusen that appear as abnormal white-yellow deposits on the retina. Segmentation of these features using conventional image analysis methods is quite complicated mainly due to the non-uniform illumination and the variability of the pigmentation of the background tissue. This chapter presents a brief overview of the area and a novel segmentation algorithm for the automatic detection and mapping of drusen in retina images acquired with the aid of a digital Fundus camera. We employ a modified adaptive histogram equalization, namely the *MultiLevel histogram Equalization (MLE)* scheme, for enhancing local intensity structures. For the detection of drusen in retina images, we develop a novel segmentation technique, the *Histogram-based Adaptive Local Thresholding (HALT)*, which extracts the useful information from an image without being affected by the presence of other structures. We provide experimental results from the application of our technique to real images, where certain abnormalities (drusen) have slightly different characteristics from the background. The performance of the algorithm is established through statistical analysis of the results. This analysis indicates that the proposed drusen detector gives reliable detection accuracy in both position and mass size.

Keywords: Age-related macular degeneration; drusen detection; segmentation algorithm; quantitative mapping of retinal abnormalities.

1. Introduction

Age-related macular degeneration (AMD) is the leading cause of irreversible vision loss among the elderly in developed countries. Many studies have confirmed that

the presence of the drusen, identified as gray-yellow deposits that build up in or around the macula of the retina, represents a significant risk factor for the development of visual loss from AMD.[2,8,10,27] Drusen are deposited by-products of rod and cone metabolism located just beneath the Retinal Pigment Epithelial (RPE) cell layer.[4] It is believed that they may signal the presence of an altered pathophysiology of the retinal pigment epithelium and consequently they may be a marker for the degree of diffused RPE dysfunction in patients with AMD.[5] The existing strong indications for the correlation between AMD and drusen development characteristics suggest that the clinical assessment of the latter might have predictive value in determining if and when a patient will suffer visual loss from AMD. Additionally, it could facilitate the development of efficient, fast and accurate clinical tests for the evaluation of the effectiveness of different treatment modalities.

Routinely, drusen characteristics are evaluated by inspecting the retina with the aid of an optical imaging apparatus known as Fundus camera. In some cases and in order to assist the evaluation of features of diagnostic importance, slides or digital images of the retina are submitted to medical centers, where specialized professionals assess the drusen characteristics. In other clinical studies, this assessment is performed with the aid of comparisons with standard photographs or with templates.[2,12,17,27,28] While the use of such procedures provides important data toward the standardization of the diagnostic procedure, their precision is relatively low.

Besides the subjectivity and the lack of reproducibility, visual assessment is not efficient in analyzing and classifying complex morphological patterns. Drusen vary in size from a few microns in diameter to large confluent complexes, which may extend to hundreds or even thousands of microns.[10] Moreover, their color appearance varies notably even within the same eye, depending on the amount of the deposited by-products beneath the RPE in each spatial location. Their color appearance is also affected by the color of the overlaid RPE, which also varies as a function of the location within the same eye, while it is strongly affected by several factors such as blood vasculature, race etc. The appearance of the retinal features is also degraded by the non-uniform transfer function of the illumination-imaging optics of the Fundus camera. These variables affect randomly the perceived contrast between drusen and background, which makes the attempt for the automatic drusen extraction a demanding image analysis task.

The contribution of image processing to the diagnosis of pathologies related to the eye may be divided into the following three groups: *image enhancement*, *mass screening* and *monitoring of the disease*. Image enhancement is generally needed due to the poor contrast and the noisy nature of the acquired images. Contrast enhancement and noise reduction may aid the human interpretation of the images and provide a meaningful first step towards automatic analysis of the fundus images. Mass screening is very important for the early diagnosis and treatment of several retinal diseases. Treatment would be more efficient if an early diagnosis could be attained. Monitoring is related to the assessment of the evolution of the disease that

has to be made by experts in order to evaluate the efficiency of new therapeutics or observe the development of specific symptoms.

1.1. *State of the art*

The problem of fully- or semi-automated drusen detection has received considerable attention over the last decade by various research groups.[11,16,21] However, acceptable performance is still an issue mainly due to the inability to compensate satisfactorily for factors that result in poor contrast among various retinal features. Drusen segmentation can actually be seen in an abstract level as a two-stage problem. Detection of the optic disk is the first important stage, since the disk shares similar attributes, in terms of brightness, color and contrast with the drusen. Additionally, correct identification of the optical disk can be seen as a landmark and it can be used for registration purposes when different retinal images of the same patient are to be compared. Its diameter is usually employed as a reference length for measuring distances and sizes. The second step is the detection of other retinal features including drusen or exudates in general.

Most of the recent research focuses on the simultaneous detection of optic disk and possible exudates, since these two procedures are complementary. In Ref. 29, the optic disk is localized exploiting its gray level variation. The approach seems to work well if there are only few exudates present that appear bright and well contrasted. Although not recent, the work of Tamura *et al.*[31] and Akita *et al.*[30] are representative of possible approaches towards detecting the disk. In Ref. 31, an area threshold is used to detect the optic disk. The final contours are detected by the circular Hough transform. This approach is time consuming and relies on conditions that are not always met. A different approach is proposed by Akita *et al.*[30] They localize the disk by backtracking the vessels to their origin. This is one of the safest ways to localize the disk but it has to rely on successful vessel detection. In Mendels *et al.*, morphological filtering techniques and active contours are used, while in Walter *et al.*,[32] an area threshold is combined with the watershed transformation in order to find the desired boundary of the disk.

Osareh *et al.*[35] propose a method to identify exudates automatically that also partially extracts the optic disk as candidate exudate region due to color similarity. This method is based on color normalization, contrast enhancement and color segmentation based on Fuzzy C-Means (FCM) clustering. This partial localization of the optic disk requires further processing to isolate it. In Osareh *et al.*,[34] the selection of candidate optic disk regions amongst the exudates is performed via boundary analysis and the optic disk centre and radius are estimated using minimum boundary arc lengths. This has been shown to work well, but relies on the local contrast enhancement that has been introduced in Osareh *et al.*[35] Their method amplifies noise, particularly in areas of only a few features. Shin *et al.*[26] propose an automated but supervised Fundus image analysis technique. They are facing the problems of retinal images (non-uniform illumination, poor contrast) in

two algorithmic steps, namely preprocessing and segmentation. Their segmentation scheme analyzes each pixel of the preprocessed image as part of a square area varying in size from 20 to 100 pixels. Skewness greater than a threshold signifies the presence of drusen. The main drawback of this technique is the requirement of close supervision by experts to achieve adequate accuracy and robustness.

Motivated by the work in Shin *et al.*,[26] Rapantzikos *et al.*[36,37] expanded histogram-based operators and improved the accuracy of drusen detection moving towards an unsupervised tool for the detection and mapping of AMD symptoms. They examine several enhancement techniques and propose a robust multilevel scheme that can effectively operate without supervision. In the segmentation step, the local histograms' shape is thoroughly analyzed by employing more descriptors than the skewness alone, so as to derive robust and accurate thresholding results through the *Histogram Adaptive Local Thresholding* (*HALT)* operator. Morphological operators are applied afterwards to compensate for possible deficiencies. The purpose of this chapter is to analyze and test a complete system based on Rapantzikos *et al.*,[37] for the detection of drusen and illuminate several aspects of the whole process.

The chapter proceeds as follows. Section 2 reviews conventional adaptive contrast enhancement and segmentation algorithms and establishes a novel scheme for image enhancement. Section 3 introduces the HALT operator for drusen detection as the main contribution of this chapter and considers step by step the application of the proposed AMD detection algorithm. Section 4 discusses the experimental results on representative images of macular degeneration and the conclusions are presented in Sec. 5.

2. Processing Tools and Methods

2.1. *Image enhancement*

Drusen are roughly distinguished visually from their background by means of their brightness, morphology and yellowish color. However, the color by itself does not convey consistent information for discrimination. Thus, in order to evaluate the contribution of color to the characterization of the symptoms, an experimental analysis was initially performed considering the RGB (Red-Green-Blue), HSI (Hue-Saturation-Intensity), CMYK (Cyan-Magenta-Yellow-Black), G/R & R/B (Green/Red, Red/Blue bands) and CIElab color spaces. Several references for color processing can be found that use different color bands for enhancement purposes. We studied drusen visibility in various color spaces and concluded that the gain in visual improvement is less than or almost the same as that of the green band of the RGB space. In this space the red band provides information for reflectance in the image and therefore is strongly affected by the non-uniform illumination, whereas the blue band contains almost no useful information for drusen. The green band is more informative and less affected from the overall variation of illumination. Our

empirical observations also agree with the selection of the green channel,[26] as the channel with the maximum contrast. Another issue related to illumination concerns the normalization of surfaces to light exposure and reflection. When irregular surfaces are illuminated, the amount of light reflected back to the camera from each region is a function of its orientation with respect to the source of light and the camera. The shape irregularity of the retina produces variable shading across the field of view when illuminated with a bright source, as in the Fundus camera.[26] For illumination compensation, a simple technique such as homomorphic filtering[24] provides good results and is recommended for our application domain. A more complicated finite element analysis of light distribution is computationally inefficient.

Following the illumination compensation, the next processing step aims at enhancing the contrast of the retina's image. Towards this direction, histogram equalization and its adaptive versions are quite promising, since they can spread out the modes of a histogram. Rapantzikos *et al.*[36,37] use histogram equalization as the core of the enhancement method. Global histogram techniques, like contrast stretching and histogram equalization are widely used to achieve contrast enhancement. Although they are simple to implement, global schemes are affected by the overall distribution in the image and they only stretch illumination differences that are widely spread within the image. Actually, they only separate strong concentrations in the histogram distribution of the image as demonstrated in Fig. 1(b). Such techniques are more effective in our case when applied in small windows as local transforms after the non-uniform illumination compensation. Such enhancement operators may be used in a hierarchical form so as to stretch local histogram distributions and enhance the contrast of the image by taking into consideration both global and local variations.

In order to standardize the enhancement of retina images and overcome the need for selecting different parameters for each image considered, a MultiLevel histogram Equalization (MLE) technique is developed that is based on sequential application of histogram equalization. In fact, MLE is a multilevel (hierarchical) scheme that progresses from the entire image to smaller regions defined via windows. Due to the expected intensity similarity in small areas, the windows considered are non-overlapping. Compared with a sliding window approach, this scheme results in smaller computational complexity and larger speed of operation, without

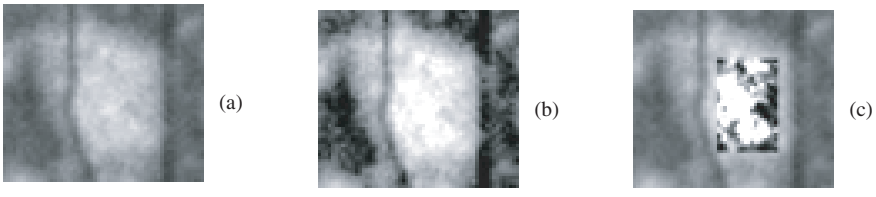


Fig. 1. (a) Original window containing one relative large drusen; (b) Histogram equalization using the entire window; (c) Histogram equalization using a smaller window inside the drusen's area.

compromising on the local enhancement ability owing to its multilevel nature. A potential problem could arise using windows that are small enough to fit inside a drusen's region. Similar to adaptive histogram modification algorithms, it can produce non-desirable misleading contrast variations within a drusen, as shown in Fig. 1(c). This problem is only experienced when using small windows and forms the opposite drawback (over-enhancement) from that of global techniques. To avoid such effects, we limit the size of windows considered up to the expected size of any drusen in the image.

Considering all these constraints, the MLE enhancement algorithm proceeds as follows: The first stage of equalization uses a window equal to the size of the image (global). The second stage splits the image into nine non-overlapping windows and applies the same operation to each sub-block of the previous result. At any stage $\boldsymbol{i}$, a window $\boldsymbol{w}^i$ is segmented and the segments are labeled as to estimate the mean size of the drusen involved. This window is further processed by nine smaller non-overlapping windows if and only if it involves smaller drusen. More specifically the algorithm proceeds to the $\boldsymbol{i+1}$ stage for a specific window $\boldsymbol{w}^i$ if the size of the largest label in $\boldsymbol{w}^i$ is smaller than $1/9^{\text{th}}$ the size of $\boldsymbol{w}^i$.

An example of the MLE operation is presented in Fig. 2. The first "pass" is responsible for enhancing the brightest parts of the image, including small, bright drusen and central parts of larger drusen (Fig. 2(c)). However, vague anomalies and dark areas that belong to spread drusen must be further enhanced, in order to be detected. The second stage of equalization, as shown in Fig. 2(d), contributes in generating more distance between those "hidden" anomalies and their surrounding areas. In our application we always proceed to the second stage of equalization. Nevertheless, due to the relatively large drusen experienced in all images tested, further window splitting and enhancement is not necessary. Figures 2(e) and 2(f) demonstrate the additional enhancement achieved by the second stage of MLE.

2.2. *Threshold-based drusen detection: histogram properties*

Segmenting the drusen in the enhanced image is an intriguing task. Parts of the drusen are difficult to distinguish from the background because of brightness similarities; especially when encountering drusen near to vessels. In order to efficiently address the problem of region segmentation, two general approaches are widely used, namely the stochastic classification[20] of pixels into object classes and the histogram thresholding for clustering similar-intensity pixels into compact objects. In this work, we adopt the second approach, i.e. histogram-based thresholding, and particularly focus on the analysis of local histogram as a stochastic mixture density that models small-included objects within the background distribution. This analysis leads to the definition of the HALT operator, presented in detail in the next section. The HALT operator as a region-based threshold selection scheme fits well with the region-based MLE approach developed for enhancement.

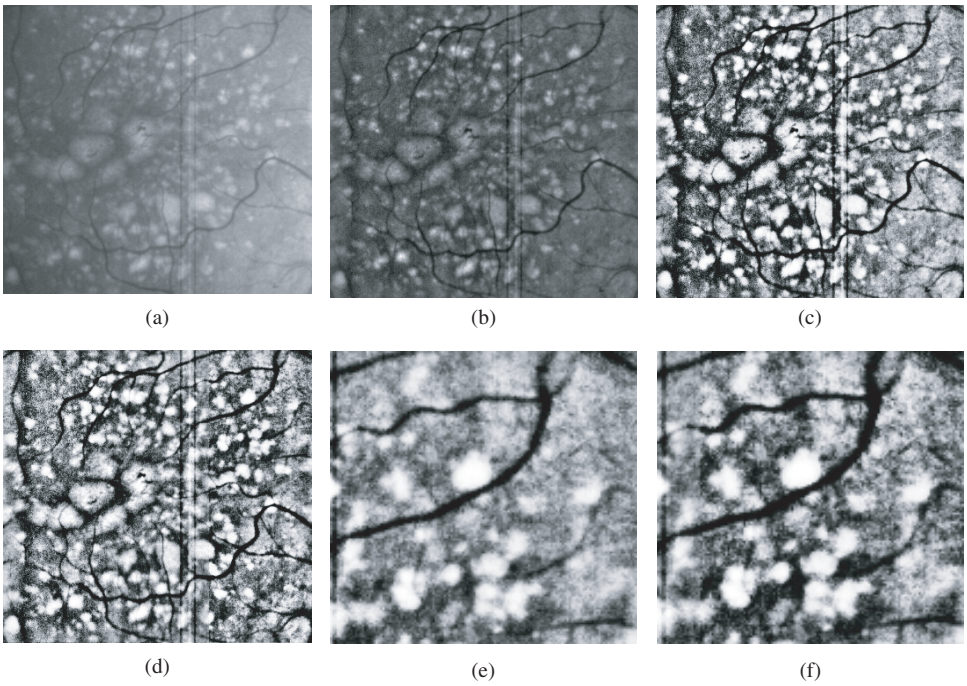


Fig. 2. (a) Original image; (b) image after non-uniform illumination compensation; (c) first level of histogram equalization (global) applied to entire image; (d) second level of histogram equalization applied to regions of previous result; (e) enlarged section of upper left corner in (c); (f) enlarged section of upper left corner in (d).

The observed value of a random variable (such as pixel intensity) can provide significant information with respect to the stochastic distribution of the variable. In fact, under the ergodicity assumption, statistical measures of the distribution can be accurately inferred from the sample measures as the size of the observed data increases.[19] The statistics that can be easily computed from the observations (data set) of a random variable fall into several categories, with the most important being the following:

(a) *Measures of central tendency* — statistics that describe the center of the data, including the mean, the median and the mode.
(b) *Measures of spread* — statistics that describe the dispersion of the data, including the variance, standard deviation, range, and inter-quartile range.
(c) *Measures of shape* — statistics that compare the shape of the data to that of a normal distribution, including the skewness and kurtosis.

The mean, median and mode are normally close to each other. These three statistics measure the center of the data in somewhat different ways. The mean is the value that minimizes the mean square distance from the data set. The median

is the value that minimizes the absolute distance and the mode is the one that minimizes the H_∞(norm) distance from the data set. The case that these statistics are very close to each other indicates that the data is probably symmetric. As the distribution of the data becomes skewed, the sample mean moves away from the sample median in the direction of the longer tail. Generally, the sample mean is affected by extreme values (outliers), while the median tends to stay with the main body of the data.[7,9,22] Thus, the median is often located in between the mode and the mean of a distribution.

A measure of the center of a distribution is much more useful if there is a corresponding measure of *dispersion* that indicates how the distribution is spread out with respect to the center.[1,18,25] For the median, a natural measure of dispersion can be obtained from the lower and upper *quartiles*. The lower quartile is the value that integrates 1/4 of the data set and the upper quartile is the value at 3/4 of the data set. For the mean value these limits are usually obtained at the 10% and the 90% of the cumulative distribution.

Distinguish between background and drusen regions by means of thresholding the local histogram can be achieved by exploiting the above statistics. Our goal is to separate the drusen, without being affected by intensity variations caused by vessels, noise and uncompensated non-uniform illumination. Zooming into each local intensity area reveals different shapes of the histogram for each of these regions and different relative distributions of the drusen and the background. Thus, in order to determine an efficient threshold for each neighborhood, the local histogram in terms of its central tendency, symmetry and shape tendency is taken into consideration. More specifically, the symmetry of a distribution via two quotients is considered. The first quotient, namely the $|mean - median|$ difference, is a first measure of symmetry based on local statistics, as indicated before. The second quotient, namely the $|mean - mode|$ difference, is chosen as a measure of histogram's main lobe spread. If both of them are small, smaller than $1/3\sigma_b$, then the distribution is considered symmetric. Otherwise, the distribution is labeled asymmetric.

Subsequently, the skewness in conjunction with the kurtosis are used as measures of the histogram's tendency. Let b and $P(b)$ denote the variable (intensity) and its distribution (histogram), with σ_b and $\bar{b}$ representing its standard deviation and mean, respectively. Skewness is defined as $S_S = \frac{1}{\sigma_b^3}\sum_{b=0}^{L-1}(b-\bar{b})^3 P(b)$ and kurtosis as $S_K = \frac{1}{\sigma_b^4}\left[\sum_{b=0}^{L-1}(b-\bar{b})^4 P(b)\right] - 3$. A distribution is skewed if one of its tails is longer than the other. Positive skew indicates a long tail in the positive direction, while negative skew indicates a long tail in the negative direction. Kurtosis is based on the size of a distribution's tails. Distributions with relatively small tails (sharp-peaked, $S_K > 0$) are called "leptokurtic"; those with large tails (flat-topped, widely spread, $S_K < 0$) are called "platykurtic". A distribution with the same kurtosis as the normal distribution ($S_K = 0$) is called "mesocurtic". These measures can increase the confidence with which drusen (outliers on the assumed normal distribution of the background) are detected and they are used in the definition of the HALT operator, which is analyzed and tested within the context of the next section.

3. Methodology for Drusen Detection — HALT Operator

This section outlines a complete algorithm for automatic segmentation and detection of drusen based on the previous analysis. The algorithmic steps are shown in Fig. 3 and explained in the following. The homomorphic filter is applied at the front end of the algorithm to compensate for illumination irregularities. The second step involves the enhancement operation that is responsible for stretching intensity differences characterizing drusen and background. The MLE approach succeeds in enhancing most drusen, being insensitive to small brightness variations that are caused e.g. from remaining non-uniform illumination and noise. Small and bright drusen are extracted successfully, whereas large and spread drusen that tend to be darker near the edges are also identified.

As a result of the previous operators, sharp abnormalities in intensity (candidate drusen) are strongly enhanced and differentiated from the background. Such intense drusen can be readily detected by thresholding techniques. A global threshold is capable of removing darker parts that belong to the drusen's surrounding areas (background). For this purpose the global Otsu's[18,25] thresholding technique is employed. A single threshold, however, cannot identify small intensity differences that often discriminate vague abnormalities hidden in bright background areas. Thus, a two-stage histogram thresholding approach is designed. The first stage applies the global Otsu threshold to provide an initial segmentation map. This level of thresholding cannot discriminate vague abnormalities hidden in the local regions of background. It only detects and preserves regions of evident abnormalities that are crisply separated from their background, as well as background regions mixed with vague abnormalities. The second stage of thresholding that refines the segmentation map operates on a local level defining a different threshold for each local region of interest. For this particular stage, a novel local thresholding operator (HALT) is designed and analyzed by Rapantzikos *et al.*[37] and is described in this section. The HALT operator checks the local histogram for general symmetry or asymmetry and uses shape tendency indicators for assessing regions as drusen or actual background.

A morphological dilation operator precedes the HALT operator. The morphological dilation (disk shaped structuring element, 3-pixels in diameter) expands the regions that are not removed by global thresholding. If this expansion occurs in background areas, there is no serious effect, since the following segmentation step is capable of removing these expanded regions completely. The main advantage of dilation is appreciated in areas that contain only one or two large drusen without background, where direct application of any local threshold would completely eliminate the drusen area. The morphological expansion reconstructs some of the background and recovers different intensity areas in the local histogram. In other words, it forces better distinction between bright areas and their darker surroundings at the corresponding local histogram. To achieve this operation, the dilation operator is embedded into the global thresholding operator, such that the overall global thresholding mask is obtained by dilating the initial threshold mask of the

Otsu operator. Thus, the dilation is applied on the binary map of thresholded drusen areas to expand their extent.

The HALT operator applies different thresholds to regions of the image, depending on the properties of the corresponding histogram. As in the case of the MLE operator, the image is split into nine non-overlapping windows, where the HALT operator is applied. If needed, each window is further split into nine sub-windows, in order to refine the segmentation. Within each window, the HALT operator checks the statistics of local histogram and assigns the appropriate threshold. The background is composed of a noise process superimposed on a deterministic smoothly varying terrain. A symmetric Gaussian distribution efficiently characterizes this overall background process. Using the ergodicity assumption, any realization of the stochastic process or any acquired image from this process is also characterized by this Gaussian distribution. Thus, in case of a window in pure background regions, it is expected that by thresholding the distribution at its cumulative 90% level and preserving only values above this 90% threshold, we preserve only isolated pixels randomly distributed along the spatial extent of the image. These pixels are easily removed by median filtering. A symmetric distribution, however, may not always characterize background alone but can also characterize certain combinations of drusen and background distributions. This case requires thorough examination. By similar means, a positively skewed distribution indicates the presence of drusen influencing the higher part of the intensity distribution. Otsu's threshold is most appropriate in this case; if there is strong evidence that drusen is the cause of this positive influence. Negatively skewed distributions are most likely to describe areas of background, since drusen abnormalities affect the higher end of the histogram (bias towards bright values). So, by setting 90% as a threshold would also remove such regions.

Organizing these potential distributions, the HALT operator first classifies the local histogram into two distinct cases, depending on its symmetry properties, as described by the two symmetry quotients in Sec. 2.2. Subsequently, the appropriate threshold is determined according to the measures of spread and shape, as follows:

(a) Histogram totally or almost symmetric (Table 1)

- A totally symmetric gray level distribution signifies areas that are mainly occupied by background regions. However, small drusen may be present, so setting the point of 90% of the cumulative distribution as threshold would be adequate to remove background and preserve compact anomalies.
- The class of platykurtic distributions may be misleading. Generally, symmetric distributions signify background areas. Nevertheless, the platykurtic feature signifies interaction of distributions that jointly preserve symmetry. For example, if background and symptoms' gray levels are normally and equally distributed, the total histogram still appears symmetric. In this case, to avoid removal of drusen, Otsu thresholding is employed.

Table 1. The HALT operator in symmetric distributions.

Kurtosis \ Skewness	< 0	≈ 0	> 0
Platykurtic	Mainly background ⇓ 90%	Possible combination of two or more distributions ⇓ Otsu	⇓ Otsu
Mesokurtic	Mainly background ⇓ 90%	Mainly background and maybe some drusen or just large drusen (one distribution) ⇓ 90%	Drusen and background are hard to distinguish ⇓ Application of HALT in smaller regions
Leptokurtic	Mainly background ⇓ 90%	Almost constant background ⇓ 90%	Can signify the case of only a small portion of drusen: segment small drusen of high intensity ⇓ 90%

- In the case of sharp-peaked (leptokurtic) almost symmetric histograms we observe high concentration of pixels around the mean value. These regions appear with almost uniform background. Such leptokuric background distributions may only allow the existence of small drusen as outliers that do not alter the general uniformity of intensities. Using Otsu thresholding, that is obtaining a threshold value close to mean, would retain anomalies and a large part of the background. Alternatively, setting 90% as threshold would remove background areas and retain, if existing, small compact drusen areas.
- The case of a mesokurtic and positively skewed histogram requires particular attention. The mesokurtic characteristic most likely arises from the background distribution. The positive skewness indicates interaction with another distribution, which is observable but not as significant as to alter drastically the background statistics. This second distribution is detected at high intensity values indicating the existence of object(s), whose intensity however interacts with that of the background. Thus, the direct segmentation of object and background may be inefficient. Using Otsu's threshold may leave large areas of the background, whereas using the 90% threshold may delete a good portion of the object's structure. Thus, an additional step of local thresholding is used, which is actually the application of HALT method focused on smaller areas of the first level's region. This helps in obtaining better distinction of anomalies and background at corresponding second level histograms.

(b) Histogram totally or almost asymmetric (Table 2)

- A positively skewed distribution of this class notifies the presence of many small or a few large drusen. In fact, bright gray levels that generally characterize anomalies

Table 2. The HALT operator in asymmetric distributions.

Kurtosis \ Skewness	< 0	≈ 0	> 0
Platykurtic	Mainly background ⇓ 90%	Drusen are present ⇓ Otsu	Drusen & background are almost equally distributed ⇓ Otsu
Mesokurtic	Mainly background ⇓ 90%	Drusen & background are almost equally distributed ⇓ Otsu	Mainly drusen ⇓ Otsu
Leptokurtic	Mainly background ⇓ 90%	Mostly background, less drusen ⇓ Otsu	Drusen & background ⇓ Otsu

dominate the histogram. Otsu thresholding is best suited to this case, since the distinction of bright and darker areas (background) is obvious.

- An asymmetric non-skewed distribution signifies the presence of drusen. This distribution results as a combination of similar distributions, characterizing background and abnormalities (drusen). Thus, Otsu thresholding is appropriate for segmenting the drusen in such regions.

The exact process for selecting the threshold in the HALT operator is outlined in Tables 1 and 2. The HALT operator is succeeded by a median filter that eliminates sparse pixels that cause false "alarms" during presence of anomalies. In this way, the HALT preserves as drusen only those pixels that appear compactly distributed into regions. The median filter is necessary to remove sparse pixels preserved by the application of the 90% threshold in background regions.

The block-wise application of the HALT operator may produce undesirable segmentation results when a region's histogram appears to be almost symmetric (small skew and mesokurtic). Applying a 90% threshold on this region's histogram, followed by a median filter, may preserve isolated bright groups of pixels. If these small bright areas are very close to each other, then they possibly belong to the same larger drusen and must be expanded so as to capture the entire drusen region. A morphological closing with a small structuring element (disk shaped, 3-pixels in diameter) applied locally within such regions can join together the neighboring groups of pixels into a single drusen. It is emphasized here that this selective expansion process is only applied on small sub-blocks of the image that possess symmetric small skewed and mesokurtic histogram, as have been identified by the HALT operator. The proposed algorithm for drusen detection and segmentation is summarized in detail in Fig. 3.

In order to demonstrate the efficiency of the HALT approach over the localized Otsu[18] and the Shin *et al.*[26] methods for threshold selection, two representative

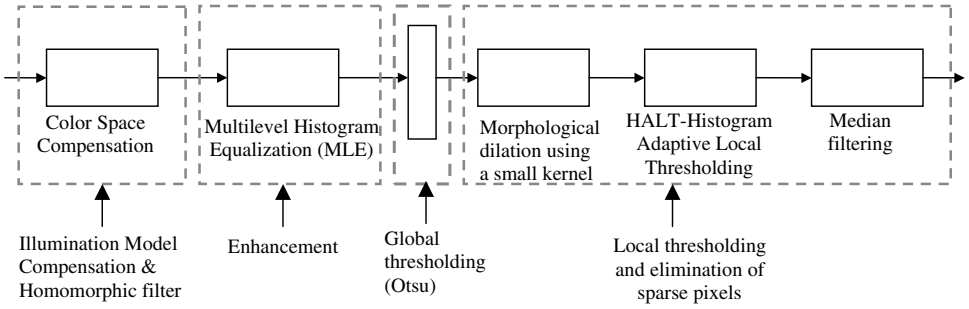


Fig. 3. Proposed algorithm for the detection of anomalies in human eye's retina.

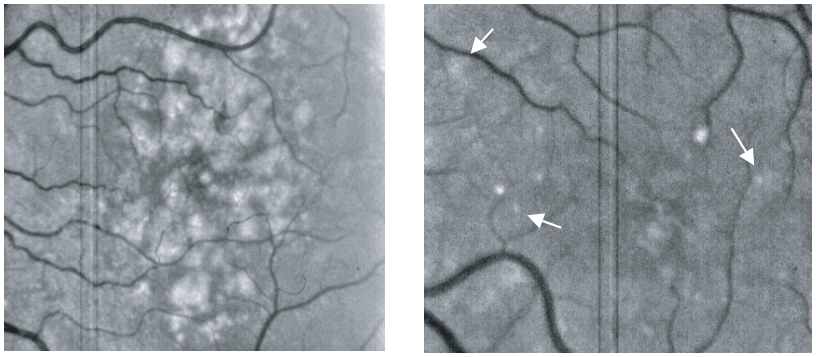


Fig. 4. (a) Image with large dense drusen; (b) image with small sparse drusen.

examples are shown. One with large drusen dominating extensive areas (Fig. 4(a)) and one with few small and vaguely defined drusen (Fig. 4(b)). Both images are enhanced using multilevel histogram equalization (MLE) and then thresholded using local Otsu,[18] Shin *et al.*[26] and HALT techniques. A median filter is applied afterwards to remove isolated pixels. The results are presented in Fig. 5.

Otsu's localized thresholding scheme works well in regions dominated by drusen (brighter areas), since the distinction between them and the background is evident. This is demonstrated in Fig. 5(a), where drusen at the central part of the image are correctly distinguished from the surrounding areas. However, the algorithm is strongly affected by regions that do not contain abnormalities, like those regions at the sides of the image. Due to remaining effects of non-uniform illumination, parts of these regions are brighter and are misclassified as anomalies. Figure 5(b) brings out another disadvantage of the local Otsu scheme. Vaguely defined drusen, which are either small or located inside bright background regions are not segmented. The algorithm detects the most obvious drusen (two of them are easily conceived), but fails to isolate and detect "hidden" anomalies; arrows indicate some of those.[26]

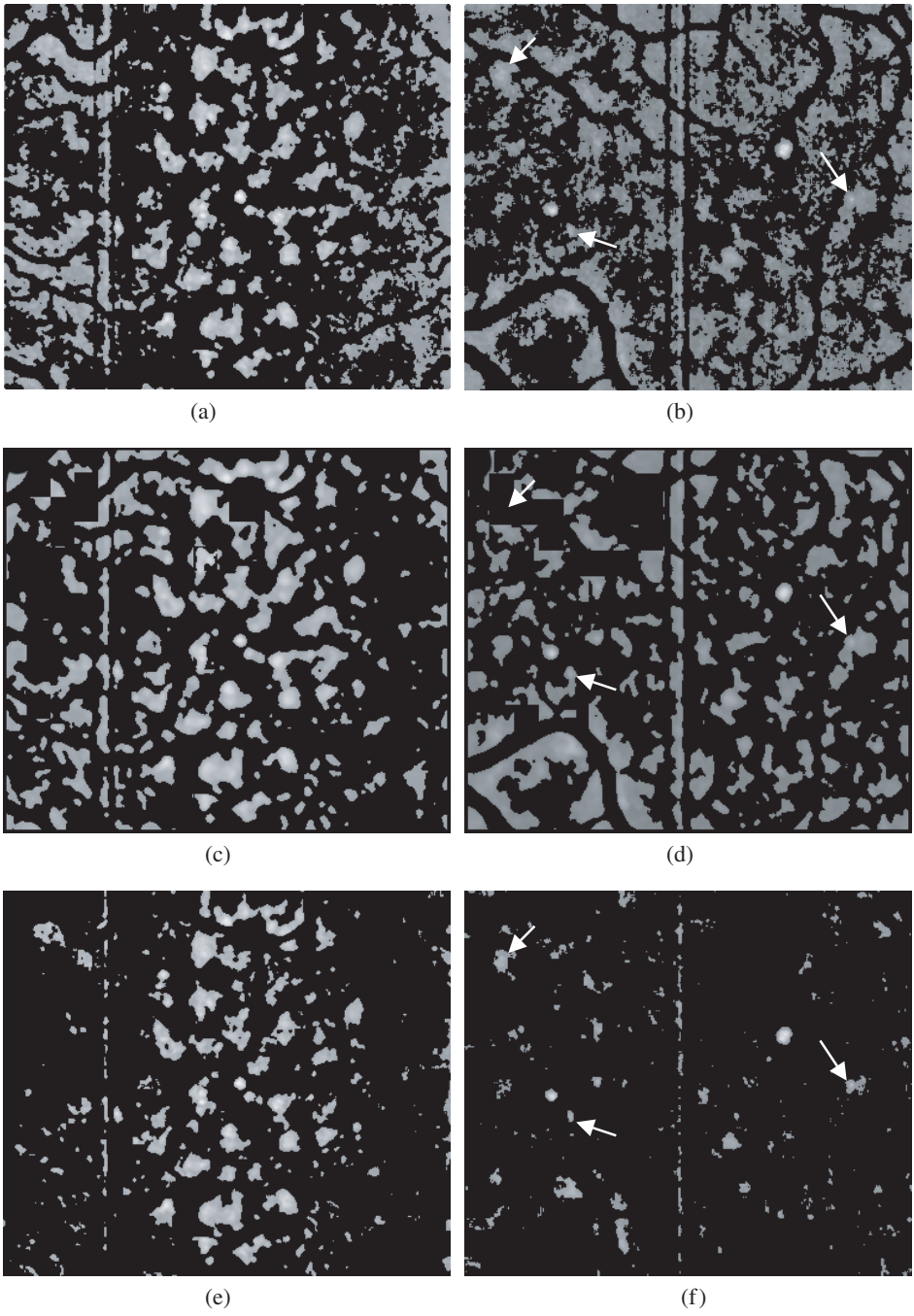


Fig. 5. Comparison of segmentation techniques; (a) (b) local Otsu; (c) (d) Shin *et al.*'s technique; (e) (f) HALT operator.

The segmentation of Shin *et al.* (Figs. 5(c) and 5(d)) tends to spread and overestimate the area of drusen especially around vessels. Although the most obvious drusen of the first image are detected by this method, supervision is required in order to remove many incorrectly segmented areas. Using the same parameters in the second test image, this method produces the result of Fig. 5(d) expressing an inability to accurately isolate small and vaguely defined drusen. It is emphasized here that the Shin *et al.*'s[26] technique can give improved results with a proper selection of its parameters for each image. This need for parameter selection specifically for each test image renders the method inappropriate for the automatic and unsupervised segmentation of drusen.

On the contrary, the HALT technique removes most of the background in both cases, as shown at Figs. 5(e) and 5(f). Even the most hard-to-see drusen are segmented without loosing their actual size and shape. Some sparse false negatives generated by the existence of noise can be easily removed through simple median filtering. Notice that the parameters of our algorithm are set once and remain fixed for the ensemble of images tested.

4. Results

We tested our algorithm using a set of 23 images, acquired by the Fundus camera. Eight pairs of them were actually captured from the left and right eye of patients. We focused in the central part of the retina by defining a rectangle at the right or left side of the optical nerve (right or left eye respectively). Figure 6 presents examples of gray-scale versions (green band) of the original color images. Drusen show up as bright blobs, but it is evident that the automatic extraction of these pathological features is difficult, since drusen vary strongly in shape and size and they tend to spread (varying brightness) around their location. Additionally, small bright regions of the background tend to create larger areas that can be mistaken as large drusen. The results of the proposed algorithm for the detection of defect regions (drusen) inside the human retina are presented in this section.

Figure 7 demonstrates the detection potential of the proposed algorithm on three representative images from the available set of retinal images; one image with small and large drusen, a second one with large drusen and a third one with small sparse drusen. The segmentation results are more than satisfactory since in all cases the background areas are segmented out and the symptoms are almost correctly isolated. A qualitative evaluation from experts is presented at the end of the current section. Figure 8 presents step-by-step the results and shows the effectiveness and robustness of the proposed operators.

Another example of an image that requires expansion of some regions after the HALT operator is shown at Fig. 9. This image contains large drusen that consist of bright and darker parts. In Fig. 9(b) it is obvious that after the HALT we are left with areas that must be joined together or expanded, so as to recover missing parts of anomalies. As shown in Fig. 9(c), after the proposed morphological closing the

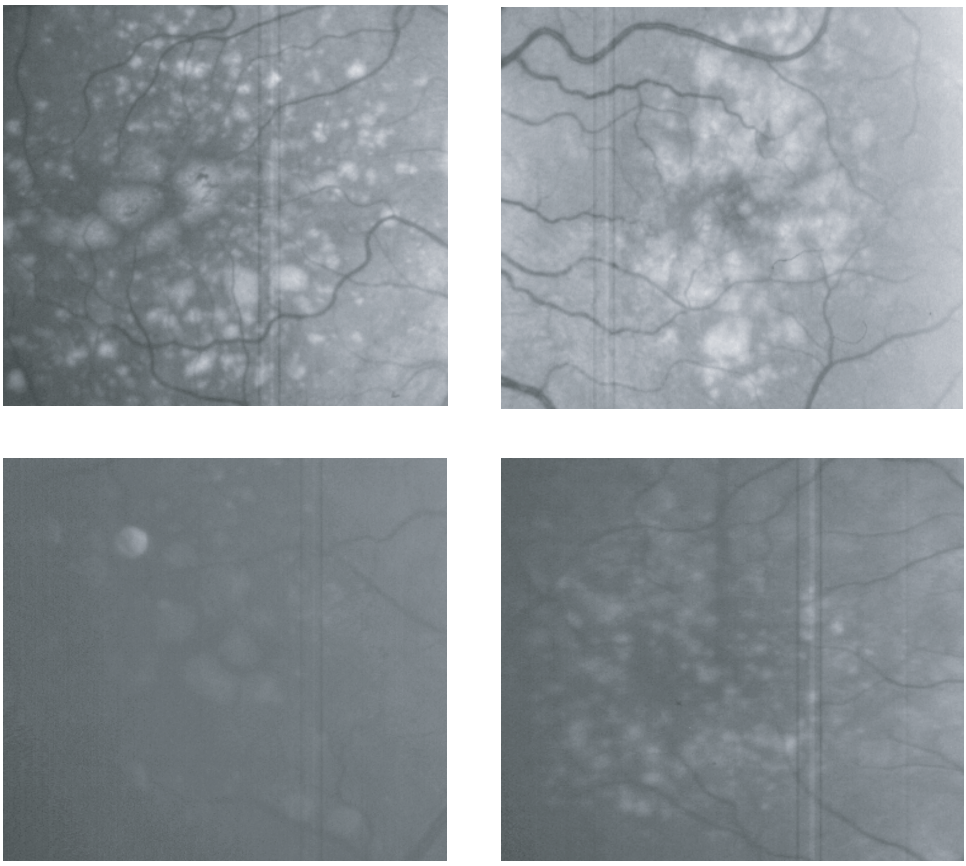


Fig. 6. Examples of test images.

upper areas that appeared "cracked" are joined together and form a single region that covers almost entirely the actual anomaly area.

A hard to enhance image is shown in Fig. 10. The presence of noise is strong, as it is detected at background regions. In addition, large drusen do not differ sufficiently from the background. Except the circular bright drusen, all others are noisy and intermixed with surrounding areas. Even in this case, our algorithm detects correctly all small drusen and loses only few parts of larger ones, which appear at the central part of Fig. 10(c).

In general, the presence of vessels and their interaction in intensity with drusen pose serious problems even in manual drusen detection. The proposed algorithm overcomes this problem and does not experience false detection, in the entire test set of images. This is due to the appropriate consideration of features in local areas that can separate drusen from vessel distributions. Overall, the proposed algorithm performs quite efficiently in the entire set of macular degeneration images tested.

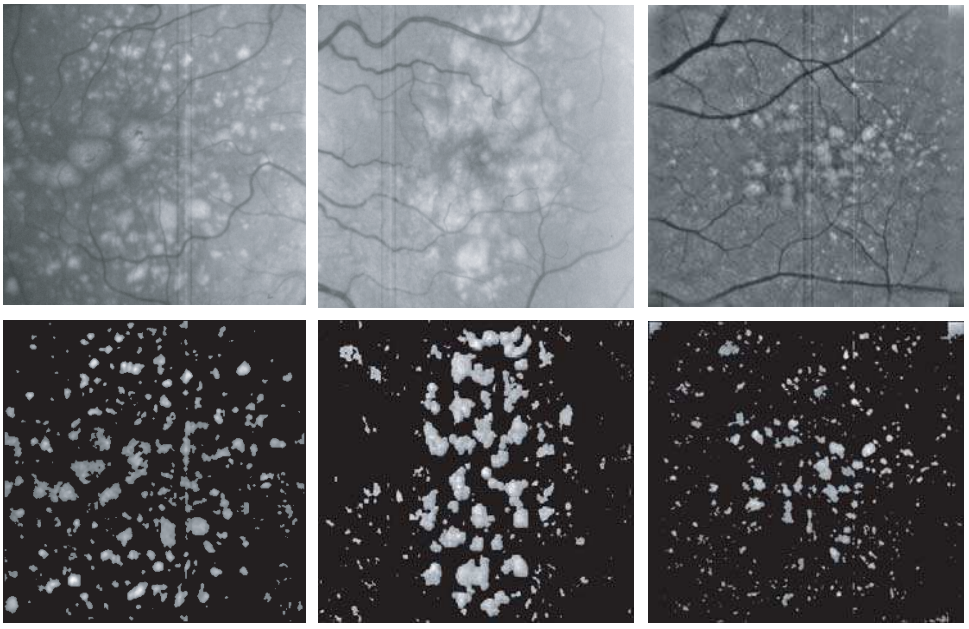


Fig. 7. Results of drusen extraction in retinas of (a)–(c) different persons. The first row shows the original images and the second row presents the segmented ones.

This set of images covers a wide range of possible drusen sizes and formations, including vague, non-canonical shaped and thin blobs.

In order to provide a statistical analysis of the algorithm's performance, we asked for experts' assistance in determining the actual drusen areas. Notice that all images reflect actual test cases without any prior information on the status and extent of AMD. Thus, for testing the algorithm's classification (drusen versus normal background) against an "actual" state, we are based on clinical evaluations performed by the experts. Two experts have extensively studied the retinal images and all the areas that are considered drusen by the doctors have been manually segmented. Their intersections, i.e. the areas that are classified as drusen by both experts, are considered as "true" drusen areas. Thus, our statistical tests give priority to "correct detection" than to "false alarm". Statistical measures, such as the rate of true positive detection (sensitivity or TPR), false positive detection (1-specificity or FPR) and false negative detection (FNR) have been computed (Table 3), in order to establish the performance of the algorithm. As mentioned before, we tested our algorithm using a set of 23 images. 8 pairs of them were actually captured from the left and right eye of patients (indicated by a, b in Table 3). The sensitivity and specificity of the algorithm exceed 96% for almost all test cases. Only in one case the sensitivity falls around 88% due to noise (test case corresponds to Fig. 10). The FNR statistic reveals that the algorithm underestimates drusen

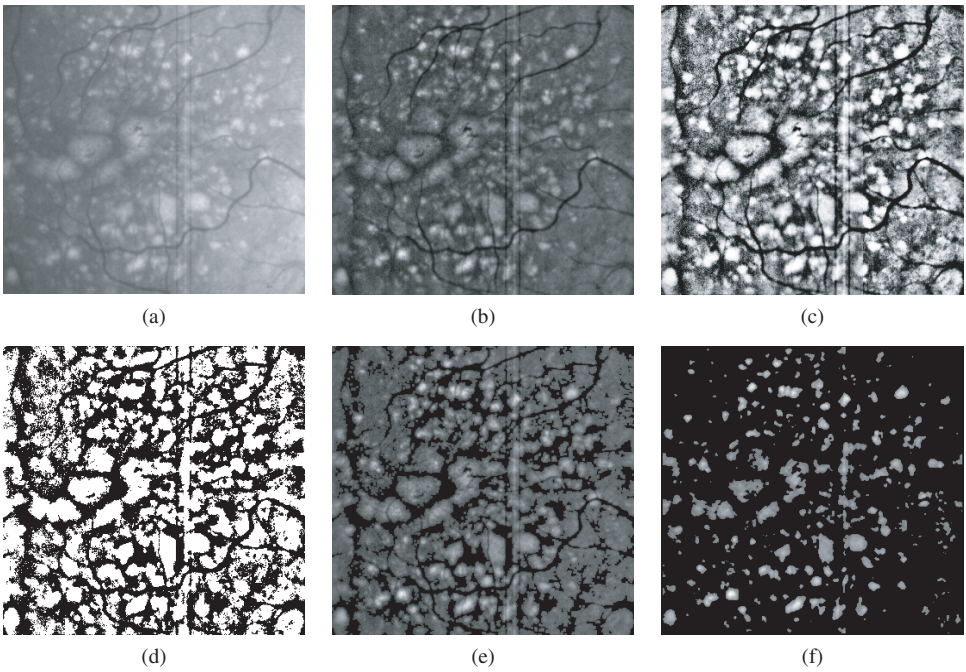


Fig. 8. Results of each step of our proposed algorithm: (a) original image; (b) non-uniform illumination correction; (c) enhancement; (d) global thresholding; (e) morphological dilation; (f) HALT and median filtering.

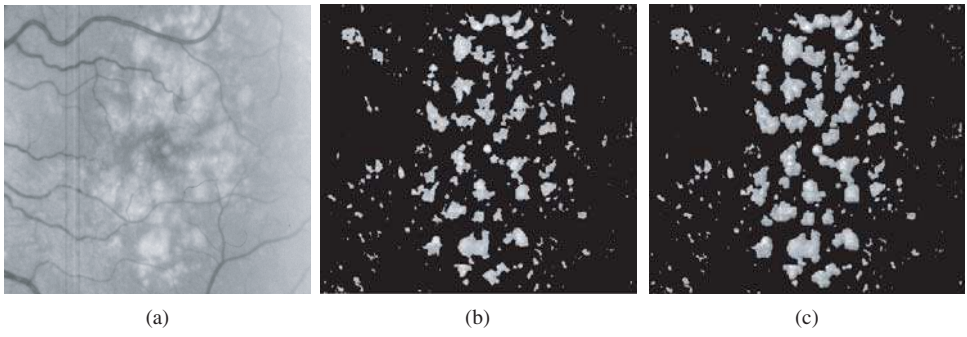


Fig. 9. (a) original image; (b) HALT and median filtering; (c) expansion of problematic areas.

areas in this case. The overall performance of the proposed algorithm on the entire set of images tested is presented in the last row of Table 3.

Further demonstrating the efficiency of the proposed algorithm, the results are subtracted from the original images, so that the detected regions appear black. Parts of the drusen that are not detected should appear bright, retaining their original gray level. Figure 11 illustrates the experts' comments on representative cases. The

(a) (b) (c)

Fig. 10. (a) original image; (b) HALT and median filtering; (c) expansion of problematic areas.

Table 3. Statistical analysis of the results.

Image	% Sensitivity (TPR)	% Specificity (1-FPR)	% FNR
1a	96.54	100	3.46
1b	100	100	0
2a	100	98.86	0
2b	100	99.75	0
3a	98.51	99.64	1.49
3b	100	99.77	0
4a	100	100	0
4b	100	99.52	0
5a	88.27	100	11.73
5b	100	100	0
6a	100	97.87	0
6b	100	98.95	0
7a	99.44	97.94	0.56
7b	99.09	99.14	0.91
8a	100	100	0
8b	100	99.78	0
9a	97.86	98.82	2.14
10a	96.15	99.19	3.85
11a	100	98.64	0
12a	97.6	100	2.4
13a	100	100	0
14a	100	96.42	0
15a	100	100	0
Overall	98.846	99.317	1.154

areas inside the solid lines are underestimated (in size), while those inside dotted lines are overestimated. Overestimation is experienced mainly at the borders, due to the different lighting model from the center to the edges of the image. These false-alarm areas can be easily rejected by the doctor inspecting the results and do not pose any problems to the early detection of AMD cases. In general, the drusen of interest in AMD examination are located inside or around the macula, the central

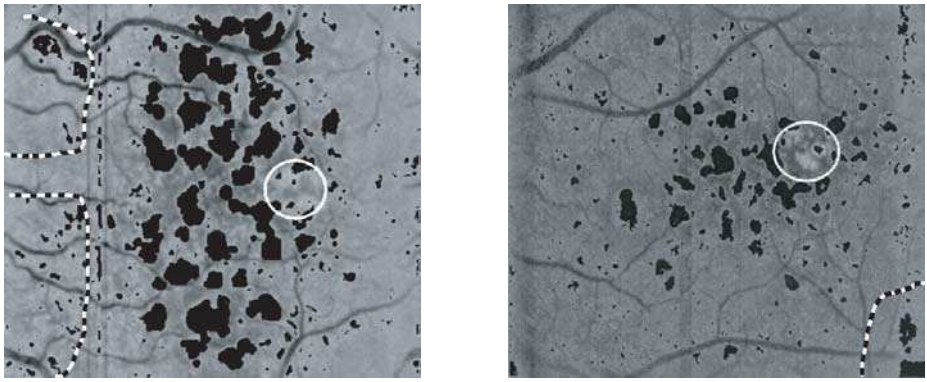


Fig. 11. Evaluation of problems judged by experts. Solid lines represent area underestimation and dotted lines represent overestimation.

part of the eye. In these areas, our proposed methodology does not produce false alarms.

Underestimation of drusen areas is a more severe problem. It should be emphasized that the underestimation of area experienced in some cases does not imply complete miss of the drusen, but only partial segmentation of it. In these cases, our methodology provides a diagnosis aid for indicating drusen presence for further examination by the doctor, who will anyway be responsible for reaching the final diagnosis.

5. Conclusions

This chapter considers histogram-based techniques for the problem of automatic AMD evaluation. The detection of anomalies in human eye's retina is a biomedical problem appropriate for image processing and automated segmentation, whose solution is intended to help the ophthalmologists in their decision making process. Use of the proposed detector may reduce false negatives and give reliable detection accuracy in both position and mass size.

We introduce and test a histogram-based enhancement technique (MLE), which uses histogram equalization as its core operator and a histogram-based segmentation technique (HALT) to segment areas that differ slightly from their background regions. Furthermore, we establish an unsupervised and non-parametric method for drusen extraction and consider its effectiveness through several examples. The proposed method is able to detect actual drusen in various cases tested. Even in hard-to-diagnose cases, where many small and vague drusen exist, our method succeeds in isolating them from the background. The proposed algorithm extends the work in Shin[26] towards the development of robust, unsupervised detection and reliable quantitative mapping of drusen abnormalities.

References

1. S. Boukharouba, J. M. Rebordao and P. L. Wendel, An amplitude segmentation method based on the distribution function of an image, *Computer Vision Graphics, and Image Processing* **29** (1985) 47–59.
2. S. B. Bressler, M. G. Maguire, N. M. Bressler and S. L. Fine, The macular photocoagulation study group, Relationship of drusen and abnormalities of the retinal pigment epithelium to the prognosis of neovascular macular degeneration, *Arch Ophthalmology* **108** (1990) 1442–1447.
3. G. Deng, L. W. Cahill and G. R. Tobin, The study of logarithmic image processing model and its application to image enhancement.
4. T. R. Friberg, Macular degeneration and its possible relationship to the microcirculation, in *Ocular Microcirculation*, eds. R. N. Weinerb, W. L. Joyner and L. A. Wheeler (Elsevier Science Publisher, Amsterdam, Netherlands, Excepta Medica, 1992), pp. 173–178.
5. T. R. Friberg, Discussion, surgical repair of full-thickness idiopathic macular holes associated with significant macular drusen, *Ophthalmology* **107** (2000) 2233–2239.
6. R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (1993).
7. F. Hampel, E. Ronchetti, P. Rousseeuw and W. Stahel, *Robust Statistics* (John Wiley, 1986).
8. F. G. Holz, T. J. Wolfensberger, B. Piguet, M. Gross-Jendroska, J. A. Wells, D. C. Minassian, I. H. Chisholm and A. C. Bird, Bilateral macular drusen in age-related macular degeneration, *Ophthalmology* **101** (1994) 1522–1528.
9. P. S. Huber, *Robust Statistics* (John Wiley, 1983).
10. International ARM Edipemiological Study Group, A classification and grading system for age-related macular degeneration, *Surv. Opthal.* **39**, 5 (1995) 368–374.
11. J. N. P. Kirkpatrick, T. Spencer, A. Manivannan *et al.*, Quantitative image analysis of macular drusen from fundus photographs and scanning laser ophtalmoscope images, *Eye* **9** (1995) 48–55.
12. R. Klein, B. E. K. Kelin, S. C. Jensen and S. M. Meuer, The five-year incidence and progression of age-related maculopathy: The beaver Dam eye study, *Ophthalmology* **104** (1997) 7–21.
13. R. Klein, M. D. Davis, Y. L. Magli, P. Sgal, B. E. K. Klein and L. Hubbard, The Wisconsin age-related maculopathy grading system, *Ophthalmology* **98** (1991) 1128–1134.
14. J. S. Lee, Digital image enhancement and noise filtering by use of local statistics.
15. Macular Photocoagulation Study Group, Risk factors for choroidal neovascularization in the second eye of patients with juxtafoveal or subfoveal choroidal neovascularization secondary to age-related macular degeneration, *Arch. Opthalmol.* **115** (1997) 741–747.
16. W. H. Morgan, R. L. Cooper, I. J. Constable and R. H. Eikelboom, Automated extraction and quantification of macular drusen from fundal photographs, *Aust. N. Z. J. Ophthalmology* **22** (1994) 7–12.
17. R. J. Olk, T. R. Friberg, K. L. Stickney, L. Akduman, K. L. Wong, M. C. Chen, M. H. Levy, C. A. Garcia and L. S. Morse, Therapeutic benefits of infrared (810 nm) diode laser macular grid photocoagulation in prophylactic treatment of non-exudative age-related macular degeneration, Two-year results of a randomized pilot study, *Ophthalmology* **106** (1999) 2082–2090.
18. N. Otsu, A threshold selection method from gray level histograms, *IEEE Transactions on Systems, Man and Cybernetics* **SMC-9**, 1 (1979).
19. A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, 1984).

20. T. Pappas, An adaptive clustering algorithm for image segmentation, *IEEE Transactions on Signal Processing* **40**, 4 (1992).
21. E. Peli and M. Lahav, Drusen measurements from fundus photographs using computer image analysis, *Ophthalmology* **93** (1986) 1575–1580.
22. I. Pitas and A. N. Venetsanopoulos, *Nonlinear Digital Filters* (Kluwer Academic Publishers, 1990).
23. W. K. Pratt, *Digital Image Processing* (1991).
24. J. C. Russ, *The Image Processing Handbook* (1999).
25. P. K. Sahoo, S. Soltani, K. C. Wong and Y. C. Chen, *A Survey of Thresholding Techniques* (Academic Press, 1988).
26. D. S. Shin, N. B. Javornik and J. W. Berger, Computer-assisted, interactive fundus image processing for macular drusen quantitation, *Ophthalmology* **106**, 6 (1999).
27. W. E. Smiddy and S. L. Fine, Prognosis of patients with bilateral macular drusen, *Ophthalmology* **91** (1984) 271–277.
28. The Choroidal Neovascularization Prevent Trial Research Group, Laser treatment in eyes with large drusen: Short-term effects seen in a pilot randomized clinical trial, *Ophthalmology* **105** (1998) 11–23.
29. C. Sinthanayothi, J. F. Boyce, H. L. Cook and T. H. Williamson, Automated localization of the optic disk, fovea and retinal blood vessels from digital color fundus images, *Br. J. Ophtalmol.* **83** (August 1999) 231–238.
30. K. Akita and H. Kuga, A computer method for understanding ocular images, *Pattern. Recogn.* **15**, 6 (1982) 431–443.
31. S. Tamura and Y. Okamoto, Zero-crossing interval correction in tracing eye-fundus blood vessels, *Pattern. Recogn.* **21**, 3 (1988) 227–233.
32. T. Walter and J.-C. Klein, Segmentation of color fundus images of the human retina: Detection of the optic disk and the vascular tree using morphological techniques, *Proc. 2nd Int. Symp. Medical Data Analysis (ISMDA 2001)*, October 2001, pp. 282–287.
33. F. Mendels, C. Heneghan and J.-P. Thiran, Identification of the optic disk boundary in retinal images using active contours, *IMVIP'99*, September 1999, pp. 103–115.
34. A. Osareh, M. Mirmeddi, B. Thomas and R. Markham, Locating the optic disk in retinal images, *2nd International Workshop on Computer Assisted Fundus Image Analysis*, ed. Bjarne Ersboll, TU Denmark, October 2001, pp. 35–35.
35. A. Osareh, M. Mirmeddi, B. Thomas and R. Markham, Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks, *Proc. Medical Image Understanding Analysis Conf.* **3** (July 2001) 49–52.
36. K. Rapantzikos and M. Zervakis, Nonlinear enhancement and segmentation algorithm for the detection of age-related macular degeneration (AMD) in human eye's retina, *ICIP'01*, 2001, pp. 1055–1058.
37. K. Rapantzikos, M. Zervakis and K. Balas, Detection and segmentation of drusen deposits on human retina: Potential in the diagnosis of age-related macular degeneration, *Medical Image Analysis* **7** (2003) 95–108.

INDEX