Lecture Notes in Computational Vision and Biomechanics 14

Shuo Li
João Manuel R. S. Tavares *Editors*

Shape Analysis in Medical Image Analysis

Springer

Lecture Notes in Computational Vision and Biomechanics

Volume 14

For further volumes:
http://www.springer.com/series/8910

The research related to the analysis of living structures (Biomechanics) has been a source of recent research in several distinct areas of science, for example, Mathematics, Mechanical Engineering, Physics, Informatics, Medicine and Sport. However, for its successful achievement, numerous research topics should be considered, such as image processing and analysis, geometric and numerical modelling, biomechanics, experimental analysis, mechanobiology and enhanced visualization, and their application to real cases must be developed and more investigation is needed. Additionally, enhanced hardware solutions and less invasive devices are demanded.

On the other hand, Image Analysis (Computational Vision) is used for the extraction of high level information from static images or dynamic image sequences. Examples of applications involving image analysis can be the study of motion of structures from image sequences, shape reconstruction from images and medical diagnosis. As a multidisciplinary area, Computational Vision considers techniques and methods from other disciplines, such as Artificial Intelligence, Signal Processing, Mathematics, Physics and Informatics. Despite the many research projects in this area, more robust and efficient methods of Computational Imaging are still demanded in many application domains in Medicine, and their validation in real scenarios is matter of urgency.

These two important and predominant branches of Science are increasingly considered to be strongly connected and related. Hence, the main goal of the LNCV&B book series consists of the provision of a comprehensive forum for discussion on the current state-of-the-art in these fields by emphasizing their connection. The book series covers (but is not limited to):

- Applications of Computational Vision and Biomechanics
- Biometrics and Biomedical Pattern Analysis
- Cellular Imaging and Cellular Mechanics
- Clinical Biomechanics
- Computational Bioimaging and Visualization
- Computational Biology in Biomedical Imaging
- Development of Biomechanical Devices
- Device and Technique Development for Biomedical Imaging
- Digital Geometry Algorithms for Computational Vision and Visualization
- Experimental Biomechanics
- Gait & Posture Mechanics
- Multiscale Analysis in Biomechanics
- Neuromuscular Biomechanics
- Numerical Methods for Living Tissues
- Numerical Simulation
- Software Development on Computational Vision and Biomechanics
- Grid and High Performance Computing for Computational Vision and Biomechanics
- Image-based Geometric Modeling and Mesh Generation
- Image Processing and Analysis
- Image Processing and Visualization in Biofluids
- Image Understanding
- Material Models
- Mechanobiology
- Medical Image Analysis
- Molecular Mechanics
- Multi-Modal Image Systems
- Multiscale Biosensors in Biomedical Imaging
- Multiscale Devices and Biomems for Biomedical Imaging
- Musculoskeletal Biomechanics
- Sport Biomechanics
- Virtual Reality in Biomechanics
- Vision Systems

Shuo Li · João Manuel R. S. Tavares
Editors

Shape Analysis in Medical Image Analysis

Springer

Editors
Shuo Li
GE Healthcare and University
of Western Ontario
London, ON
Canada

João Manuel R. S. Tavares
Departamento de Engenharia Mecânica
Faculdade de Engenharia, Universidade
do Porto
Porto
Portugal

ISSN 2212-9391 ISSN 2212-9413 (electronic)
ISBN 978-3-319-03812-4 ISBN 978-3-319-03813-1 (eBook)
DOI 10.1007/978-3-319-03813-1
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013957553

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book presents novel and cutting-edge topics in Advances of Shape Analysis in Medical Image Analysis in order to solidify knowledge in the related fields and define their key stakeholders.

The 13 chapters included in this book were written by invited experts of international recognition and address important issues in shape analysis in medical image analysis, including: techniques for image segmentation, registration, modelling and classification and applications in biology, cardiac, brain, spine, chest, lung, and clinical practice.

The book covers the most recent advances in this area. Therefore, this book is of crucial effectiveness for researchers, students, end-users, and manufacturers from several multidisciplinary fields, as the ones related with artificial intelligence, bioengineering, biomechanics, computational mechanics, computational vision, computer sciences, human motion, mathematics, medical imaging, imaging-based intervention, medicine, pattern recognition, and physics.

The editors would like to take this opportunity to thank all invited authors for sharing their works, experiences, and knowledge, making possible their dissemination through this book.

Shuo Li
João Manuel R. S. Tavares

Contents

Part I
Methods and Models

Shape Analysis for Brain Structures

Bernard Ng, Matthew Toews, Stanley Durrleman and Yonggang Shi

Abstract Advances in magnetic resonance imaging (MRI) have enabled non-invasive examination of brain structures in unprecedented details. With increasing amount of high resolution MRI data becoming available, we are at a position to make significant clinical contributions. In this chapter, we review the main approaches to shape analysis for brain structures. The purpose of this review is to provide methodological insights for pushing forward shape analysis research, so that we can better benefit from the available high resolution data. We describe in this review point-based, mesh-based, function-based, and medial representations as well as deformetrics. Their respective advantages and disadvantages as well as the implications of increasing resolution and greater sample sizes on these shape analysis approaches are discussed.

B. Ng (✉)
PARIETAL Team, INRIA, Gif-Sur-Yvette, France
e-mail: bernardyng@gmail.com

B. Ng
FIND Lab, Stanford University, Stanford, CA, USA

M. Toews
Surgical Planning Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
e-mail: mt@bwh.harvard.edu

S. Durrleman
ARAMIS Team, INRIA, Paris, France
e-mail: stanley.durrleman@inria.fr

S. Durrleman
Institut du Cerveau et de la Moëlle épinière (ICM), Hôpital de la Pitié Salpêtrière, Paris, France

Y. Shi
Institute for Neuroimaging and Informatics, Keck School of Medicine, University of Southern California, LosAngeles, CA, USA
e-mail: yonggans@usc.edu

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_1,

1 Background

The advent of neuroimaging has opened a new era for brain research. With imaging technology, such as magnetic resonance imaging (MRI), brain structures can now be examined non-invasively in great detail. Besides being a powerful visualization tool, neuroimaging facilitates quantitative characterization of neuropathology and neurodevelopment. For instance, by statistically comparing data collected from diseased populations and matched healthy controls, one can localize specific brain areas that are affected by neurological diseases. Also, by tracking longitudinal changes, one can assess disease progression, effectiveness of treatments, and effects of aging. In this chapter, we focus on shape analysis of brain structures segmented from MRI data.

Broadly speaking, the shape of an object is the geometric information that remains after removal of positional, rotational, and scaling effects [1]. In the present context, the object of interest is a given brain structure, which needs to be segmented from the MRI brain volumes either manually or through automated means. Given segmented brain structures, one of the main goals of shape analysis is to study the variability within and across populations. For instance, to understand the effects of neurological diseases, we need to first know the normal variability within healthy population. Shape analysis has a long and rich history, and the number of methods is rapidly growing. We thus focus on the main approaches and highlight some recent advancement, as opposed to being comprehensive. Using a representation-based categorization, we divide existing methods into five types of approaches. The first type operates on landmarks or a dense set of points sampled from the boundary of a segmented brain structure. For this type of approach, shape variability is typically quantified in terms of landmark geometry, i.e. point positions or local features derived from the intensity pattern within a small neighborhood around each point. This point-based approach is discussed in Sect. 2. The second type represents the boundary of a brain structure with meshes, which naturally permits the use of intrinsic geometry for surface analysis and mapping. In particular, we will concentrate on the recent developments in spectral analysis for surface analysis. This mesh-based approach is the topic of Sect. 3. The third type is to decompose the boundary using basis functions, such as spherical harmonics and wavelets. The choice of basis governs whether global or local variability is to be captured. We discuss this function-based approach in Sect. 4. The fourth type derives a skeleton of the brain structure, which enables physically-intuitive features, such as thickness and bending, to be extracted. This medial representation approach is discussed in Sect. 5. The fifth type is deformetrics based on statistics of currents, which does not require a one-to-one point correspondence between the segmented brain structures. This approach is discussed in Sect. 6.

Central to shape analysis is the issue of correspondence. The degree to which correspondence affects results varies with the shape representation employed. For instance, traditional point-based representations generally require an accurate one-to-one point correspondence across subjects. A key methodological pursue is thus

the mitigation of this requirement. For example, a local feature-based approach has been proposed to robustly cope with missing or deformed brain areas, e.g. due to abnormality or pathology, and the deformetrics approach bypasses the need for a one-to-one point correspondence altogether. Another major complication to shape analysis is boundary noise arising from image discretization and segmentation errors. The accuracy of the segmented boundary has a major impact on correspondence creation. For instance, the branching topology of a medial representation can change dramatically with tiny boundary perturbation. Furthermore, for shape analysis to be clinically relevant, result interpretability is paramount. Point-based approaches and local function-based approaches that use wavelet basis permit quantification of local shape variation, whereas global function-based approaches that use Fourier or spherical harmonic bases capture only global variability. The advantages and disadvantages of each shape analysis approach with respect to correspondence, boundary perturbation, and result interpretation are discussed in their respective sections. Recent advances in MRI acquisition have made available increasing amount of high resolution data. The impact of higher resolution and greater sample sizes in relation to each shape representation is discussed at the end of this chapter. We hope this review on shape analysis for brain structures will provide insights into the next steps for methodological advances, so that greater clinical contributions can be made with the available high resolution data.

2 Point-Based Representation

Point-based approaches represent a brain structure as a set of points and characterize its shape based on features associated with each point. A point set may be dense (e.g. all points along a boundary or within a region) or sparse (e.g. a discrete subset of all points or manually-identified landmarks). The features may be derived based on information at a specific anatomical location or within a local neighborhood. In this section, we first provide a summary of the traditional point distribution model (PDM), which assumes the presence of a one-to-one point correspondence across subjects, possibly defined through manual means. We then describe an automated probabilistic method for extracting corresponding points across subjects, which is robust to cases where homologous brain features cannot be identified in all subjects.

2.1 Point Distribution Models

2.1.1 Formulation

Assuming we are given point sets from a group of subjects that are pre-aligned using e.g. iterative closest point [2], the traditional way for localizing significant shape variability is to apply Hotelling's T2 test to the subjects' (x_i, y_i, z_i) coordinates for

each point i [3]. Alternatively, a PDM can be built to examine the principal modes of point covariance [4]. Let $X = \{x_1, y_1, z_1, \ldots, x_N, y_N, z_N\}$ denote a vector comprising 3D Euclidean coordinates of N points sampled from the boundary of a brain structure. Assuming X follows a multivariate normal distribution, parameterized by a mean point vector $\bar{X}$ and a covariance matrix Σ, X can be decomposed as a linear combination of eigenvectors Φ_i of Σ about $\bar{X}$:

$$X = \bar{X} + \sum\nolimits_i a_i \Phi_i, \tag{1}$$

where Φ_i corresponds to the principle modes of variability and shape is represented by the set of scalar coefficients $\{a_i\}$. One can thus examine the main shape variability by retaining a subset of a_i's. We highlight that PDM is often used as constraints for segmentation. This approach is commonly referred to as the active shape model (ASM) [4] and can be augmented with appearance information [5].

The key challenge to point-based representations is that certain brain features may not be identifiable or may not even exist in all subjects. Manual means for correspondence creation by finding distinctive matching features, such as corners, are very time consuming. Automated methods, such as landmark detectors [6], provide a more repeatable means, but defining general features that are robust to inter-subject variability is non-trivial. In healthy, normal brains, the same brain structure may exhibit varying morphologies across subjects, e.g. due to different folding patterns associated with the same sulcus [7], and homologous features do not necessarily exist over the course of normal brain development, e.g. between gestational and infant stages. In the case of abnormality, pathology may be present or healthy tissue may be deformed or missing due to surgical resection.

2.1.2 Applications and Insights

The PDM has been used for aligning and segmenting subcortical structures such as the globus-pallildus, thalamus, caudate nucleus [8] and cortical structures such as sulci and gyri [9]. When using the traditional ASM, MR images must be aligned via similarity or affine transform to a common reference frame, since the model does not account for global variation in orientation. Further, homologous points along the boundary of structures to be segmented must be specified in a set of training images. A fundamental challenge is to define and identify homologous points on shapes with no distinct landmarks. This challenge may be potentially alleviated by using automatic tools for shape segmentation [10], landmark matching [11], and optimal point-based shape parameterizations, e.g. minimum description length [12]. However, inter-subject variability of the cortical surface render segmentation of cortical structures via PDM approaches non-trivial.

2.2 Local Feature-Based Approach

The idea behind local feature-based approach is to use intensity information to identify corresponding points across subjects, which can be viewed as a generalization of the traditional position-based models. A summary of commonly-used generic features is first provided, with a focus on the scale-invariant feature transform (SIFT), which has shown to be robust for many applications. Techniques for modeling brain shapes probabilistically with features extracted from training images are then described.

2.2.1 Feature Extraction

Feature-based shape modeling requires a means for robustly localizing distinctive image features reflective of the underlying brain anatomy. One approach is to construct specialized detectors to identify specific brain features, e.g. the longitudinal fissure [13], posterior tips of ventricles, etc. However, building specialized detectors for each brain structure can be quite labor demanding if whole-brain analysis is of interest. Towards this end, a number of generic feature detection methods has been proposed, which provides an automatic means for identifying distinctive features. Early methods in 2D image processing focused on finding salient patterns that can be reliably localized in the presence of arbitrary image translations and rotations, for instance, corner-like patterns that can be robustly extracted using spatial derivative operators [14]. This work was later extended to 3D volumetric brain images, which facilitates identification of a high number of generic brain landmarks [6].

A complication with spatial derivative operators is the choice of optimal spatial scale, which is typically unknown. Scale-space theory [15] formalizes the notion that landmark distinctiveness is intimately linked to the size or scale at which the image is observed. This led to the development of feature detectors designed to identify both the location and scale of distinctive image features, so-called scale-invariant feature detectors [16]. The field of image processing witnessed a number of scale-invariant feature detectors, primarily based on Gaussian derivatives in scale and in space, e.g. the SIFT method that identifies maxima in the difference-of-Gaussian (DoG) scale space [17]. SIFT has shown to provide robust matching features for brain analysis [18] in addition to various computer vision applications, and will be the focus for the remainder of this subsection.

A scale-invariant feature in 3D is a local coordinate reference frame consisting of a 3D point location $\bar{x}$, a scale parameter σ, and an orientation matrix $\Theta = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ parameterized by orthonormal axis vectors $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$. The local feature-based approach described here adopts the DoG operator in identifying feature points defined as $(\bar{x}, \sigma)$ of the DoG extrema: $\{(\bar{x}_i, \sigma_i)\} = \underset{\bar{x},\sigma}{local\,argmax} |g(\bar{x}, \kappa\sigma) - g(\bar{x}, \sigma)|$, where $g(\bar{x}, \sigma)$ represents the convolution of an image with a Gaussian kernel of variance σ^2, κ is the constant multiplicative sampling increment of scale, and

$\underset{x}{local\ argmax}\ g(x)$ denotes the set of all local maxima of $g(x)$. Although a variety of different saliency criteria could be adopted, the DoG criterion is attractive as it can be computed efficiently in $O(N \cdot logN)$ time and memory complexity in the number of voxels N using scale-space pyramids [19].

To represent brain image features in a manner invariant to global image rotation, translation, and scaling, an important step is to assign a local orientation to the feature points. Local orientation information is particularly useful for alignment, which can be estimated via highly efficient algorithms, such as the Hough transform ($O(N \cdot logN)$ time in the number of features N), to recover global similarity transforms and locally linear deformations [20]. The 3D orientation of a local feature point comprises three intrinsic parameters. These may be difficult to estimate unbiasedly due to the non-uniform joint density functions of common angular parameterizations, e.g. Euler angles, quaterions, etc. Instead, using feature-based alignment (FBA) [20] can be advantageous as it estimates direction cosine vectors of a 3×3 rotation matrix θ from spherical gradient orientation histograms, which reduces the effect of parameterization bias.

Once local orientation has been estimated, appearance features can be extracted from the image content within a local neighborhood of volume proportional to σ^3 around the point $\bar{x}$ for establishing correspondence. Local appearance may simply be represented by raw image intensities [18], but the computer vision literature has demonstrated that alternative representations, such as the gradient orientation histogram (GoH), to be superior in terms of matching distinctiveness. The local feature-based approach adopts a variant of GoH by quantizing space and gradient orientation uniformly into an $8 \times 8 = 64$ element histogram. GoH elements are rank-ordered, where each element is assigned its rank in an array sorted according to bin counts, rather than the raw histogram bin count [21]. An example comparing PDM and local feature-based approach is shown in Fig. 1, in which PDM breaks down due to missing homologous brain features in the subjects [22].

2.2.2 Probabilistic Modeling

Once scale-invariant features have been extracted from image data, they can be used as the basis for describing brain geometry and appearance probabilistically across a population. It is important to characterize variability within a normalized geometric frame of reference with variation in scale, rotation, and translation across different images removed. Let $\bar{I} = \{I_i\}$ represents a set of features extracted from an image, and let T represent a global similarity transform that maps the image to a normalized reference frame. T can be estimated using FBA [20], which employs a generative probabilistic model to learn the posterior probability of the unknown global similarity transform T conditional on the feature data $\bar{I}$:

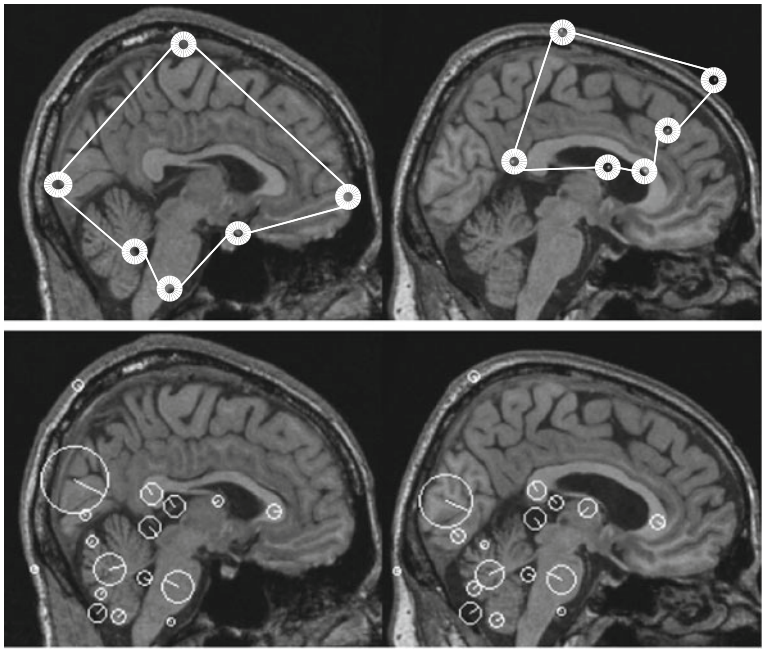


Fig. 1 PDM versus local feature-based approach. Corresponding points extracted by PDM (*upper pair*) and local feature-based approach (*lower pair*) shown. PDM fails (*upper right*) due to unexpected inter-subject variability. The local feature-based model is stable, identifying robust scale-invariant feature correspondences (*white circles*) in cortical and sub-cortical regions

$$\begin{aligned} p(T|\bar{I}) &\propto p(T)p(\bar{I}|T) = p(T)\prod_i p(I_i|T) \\ &= p(T)\prod_i \sum_j p(I_i, f_j|T) = p(T)\prod_i \sum_j p(f_j)p(I_i|f_j, T). \end{aligned} \tag{2}$$

In this formulation, $f = \{f_0, \ldots, f_K\}$ is a latent random variable taking on discrete values $f_0, \ldots, f_K$ with probability $p(f_i)$. Each value f_j indicates a distinctive anatomical pattern, whose shape and appearance are encoded locally in normalized space by the density $p(I_i|f_j, T)$ and by the occurrence probability $p(f_i)$. Note that the extracted features $\bar{I}$ are conditionally independent given transform T, and that latent random variable f is independent of the transform T. The optimal alignment solution T^* is taken as the one that maximizes the posterior probability (2). For the purpose of group analysis and classification, referred to as feature-based morphometry (FBM) [18], a random variable indicating subject group label $C = \{C_1, \ldots, C_K\}$ is incorporated into the model. Subject group can be identified by maximizing the posterior probability of C conditional on the extracted features $\bar{I}$ and T:

$$p(C|\bar{I}, T) \propto p(C)\prod_i p(I_i|C, T) = p(C)\prod_i \sum_j p(I_j|f_j, C, T)p(f_j|C). \tag{3}$$

C is assumed to be independent of T and, as in the case of FBA, the extracted features $\bar{I}$ are assumed to be conditionally independent. Unlike FBA, latent feature shape and appearance densities $p(I_i|f_j, C, T)$ and occurrence probability $p(f_j|C)$ are conditioned on the group label in FBM, and can thus be used to identify group-informative structure and to quantify group differences. Classification and analysis can both be considered as identifying the group label C^* that maximizes the Bayes decision ratio:

$$C^* = \underset{C}{argmax} \left\{ \frac{p(C)}{p(C')} \prod_i \sum_j \frac{p(I_i|f_j, C, T)p(f_j|C)}{p(I_i|f_j, C', T)p(f_j|C')} \right\}. \tag{4}$$

The decision ratio can be used to identify the optimal group label of a new subject. Note that the classification is heavily influenced by the product of likelihood ratios $p(f_j|C)/p(f_j|C')$ associated with latent model feature f_i. This likelihood ratio can be used to quantify the informativeness of the features with regard to group label, and can be used for identifying group-related anatomical structure.

The estimation of T^* and C^* require learning: the latent feature set $f = \{f_0, \ldots, f_K\}$, probabilities $p(f_j)$ and $p(f_j|C)$, and likelihoods $p(I_i|f_j, T)$ and $p(I_i|f_j, C, T)$. Here, $p(I_i|f_j, T)$ and $p(I_i|f_j, C, T)$ are assumed to be conditional Gaussian densities over the individual features I_i parameterized by mean and variance vectors. Probabilities $p(f_j)$ and $p(f_j|C)$ are parameterized by normalized feature occurrence counts. Computation of maximum likelihood estimates of the density and probability parameters can be posed as a clustering problem with cluster centers of the features extracted from training data that are similar in geometry and appearance being the estimates.

With probability and density parameters estimated, T^* and C^* can be determined by computing their respective posterior probabilities based on nearest neighbor correspondences between extracted image and latent model feature appearance vectors. T^* can thus be estimated in a manner similar to the Hough transform, and C^* is estimated in a manner similar to a Naïve Bayes classifier. An example illustrating the application of the feature-based model for alignment and group analysis is shown in Fig. 2.

2.2.3 Applications and Insights

The local feature-based approach has a variety of possible applications. The first is as a means for achieving efficient, robust alignment of difficult brain image data even in the face of missing one-to-one homology. This task is a precursor to analyzing challenging data [20, 23]. The result of alignment is not only a global similarity transform T but also a set of probable image-to-model feature correspondences that quantify local residual geometrical variation. These correspondences have been shown to be effective in initializing deformable alignment of difficult data, e.g. enlarged ventricles in infant Krabbe disease [24]. Second, latent features from the trained model can be

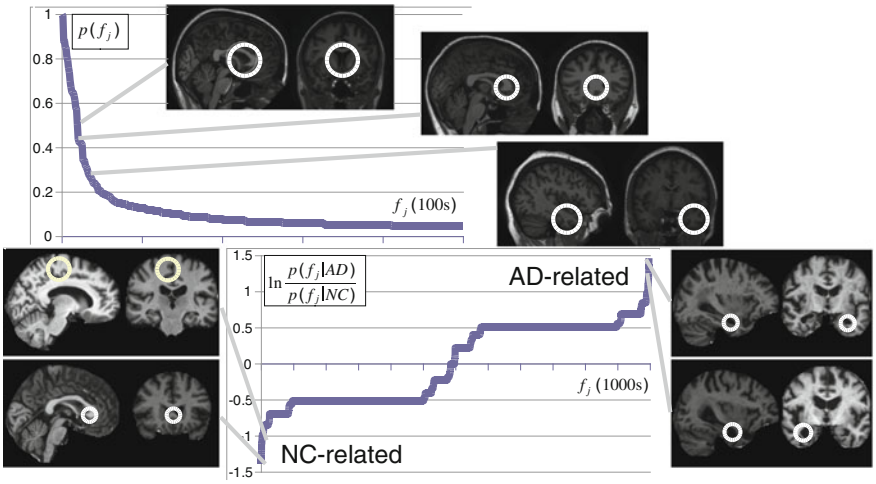


Fig. 2 Scale-invariant feature-based model for alignment and group analysis. *Above* Latent feature probability $p(f_j)$ can be used to identify stable neuroanatomical patterns across a population and quantify their occurrence frequency [20]. *Below* latent features can be related to subject group, such as normal control (NC) and Alzheimer's disease (AD), by the likelihood ratio $p(f_j|C)$ for identifying group-related neuroanatomical structure [18]

used to identify anatomical structure most representative of a population, or image structure most characteristic of specific groups. Further, this information permits prediction of group labels for new image data, e.g. in a computer-assisted diagnosis scenario [18] or to predict continuous variables, such as the age and possibly the developmental stage of the infant brain [25]. Lastly, the local feature-based approach is applicable for a variety of image modalities and can be adapted for inter-modality correspondence [23]. This approach lends itself to alignment and analysis algorithms that are highly efficient in terms of computation time and memory footprint, and is thus effective in modeling and analyzing brain structure in large image databases or across bandwidth-limited networks.

3 Mesh-Based Representation: A Spectral Analysis Approach

The boundary between tissue types contains important information for characterizing brain shape. By representing this boundary as a surface, we can exploit powerful tools from intrinsic geometry. In this section, we review works that use the spectrum of surfaces for the intrinsic surface analysis, which has led to novel ways of surface reconstruction, classification, feature extraction, and computation of conformal maps.

3.1 Laplace-Beltrami Eigen-System

Let M denote a surface in R^3. Its Laplace-Beltrami (LB) eigen-system is defined as:

$$\Delta_M f = -\lambda f, \tag{5}$$

where Δ_M is the LB operator on M and f is a smooth function defined over M. Since the LB operator is self-adjoint, its spectrum is discrete and can be ordered by the magnitude of the eigen-values as $0 \leq \lambda_0 < \lambda_1 < \lambda_2 < \ldots$ If M is a surface with boundary, we can solve the eigen-system with the Neumann boundary condition. Some special examples of the LB eigen-system are widely used in signal processing. In the Euclidean domain, the LB eign-functions are the Fourier basis. On the unit sphere, the LB eigen-functions are spherical harmonics. Due to symmetry, the spherical harmonics have multiple eigen-functions for a single eigenvalue. For the lth order spherical harmonics, there are $2l+1$ eigen-functions. This multiplicity, however, is not general. For arbitrary surfaces without such symmetry, it was proved that this is not an issue [26].

For numerical computation, a surface is often represented as triangular meshes created using the finite element method. Let $M = (V, T)$ denote the triangular mesh representation of the surface, where V is the set of vertices and T is the set of triangles. Using the weak form of (5), we can compute the eigen-system by solving:

$$Qf = \lambda U f, \tag{6}$$

where the two matrices can be derived using the finite element method:

$$Q_{ij} = \begin{cases} \frac{1}{2} \sum\limits_{V_j \in N(V_i)} \sum\limits_{T_l \in N(V_i, V_j)} \cot \theta_l^{i,j}, & \text{if } i\text{th diagonal} \\ -\frac{1}{2} \sum\limits_{T_l \in N(V_i, V_j)} \cot \theta_l^{i,j}, & \text{if } V_j \in N(V_i) \\ 0, & \text{otherwise} \end{cases}, \tag{7}$$

$$U_{ij} = \begin{cases} \frac{1}{12} \sum\limits_{V_j \in N(V_i)} \sum\limits_{T_l \in N(V_i, V_j)} Area(T_l), & \text{if } i\text{th diagonal} \\ \frac{1}{12} \sum\limits_{T_l \in N(V_i, V_j)} Area(T_l), & \text{if } V_j \in N(V_i) \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

where $N(V_i)$ is the set of vertices in the 1-ring neighborhood of V_i, $N(\mathrm{V_i}, \mathrm{V_j})$ is the set of triangles sharing the edge (V_i, V_j), $\theta_l^{i,j}$ is the angle in the triangle T_l opposite to the edge (V_i, V_j), and Area (T_l) is the area of the l^{th} triangle T_l.

As an illustration, the spectrum of a hippocampus is plotted in Fig. 3. We can see that the eigen-function becomes more oscillatory with increasing order. Thus, higher order LB eigen-functions can be considered as higher frequency basis, where frequency is intuitively the number of sign changes over the surface. This is reminiscent

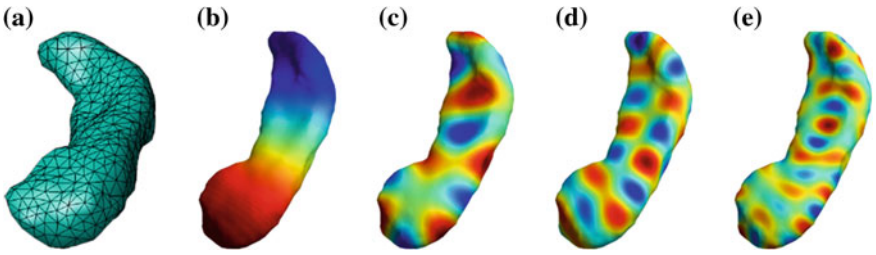


Fig. 3 LB eigen-functions on a hippocampal mesh model. *i* denotes the order of the eigen-function. **a** Mesh. **b** i = 1. **c** i = 25. **d** i = 50. **e** i = 100. With increasing order, the eigen-function becomes more oscillatory

of the Fourier basis functions on Euclidean domain. This intuition is useful for signal processing applications on the surface. In [27], the LB eigen-functions were used for denoising signal defined on surface patches. For shape analysis, the LB basis functions were used to form a subspace for outlier detection and the generation of smooth mesh representation of anatomical boundaries [28]. For different surfaces, the number of nodal domains is also different and can be used as a signature for surface classification [29], which is similar to the shape DNA concept [30]. Next we present two different ways of using the LB eigen-functions to study brain surfaces.

3.2 Reeb Graph of LB Eigen-Functions

Many brain surfaces have obvious characteristics that are invariant to the orientation of the brain. For example, the hippocampus has the tail-to-head elongated trend. The cortical surface has the superior-to-inferior, frontal-to-posterior, and medial-to-later trend. The LB eigen-functions are very effective in capturing these global characteristics of brain surfaces. One powerful tool for this purpose is the Reeb graph of the LB eigen-functions, which can transform functions into explicit graph representations.

Given a function defined on a manifold, its Reeb graph [31] describes the neighboring relation between the level sets of the function. For a Morse function f on the mesh M, its Reeb graph is mathematically given by the following definition [31]:

Definition 1: Let $f : M \rightarrow \mathrm{R}$. The Reeb graph $R(f)$ of f is the quotient space with its topology defined through the equivalent relation $x \simeq y$ if $f(x) = f(y)$ for $\forall x, y \in M$.

As a quotient topological space derived from M, the connectivity of the elements in $R(f)$, which are the level sets of f, change topology only at critical points of f. Reeb graph is essentially a graph of critical points. If f is a Morse function [32], which means the critical points of f are non-degenerative, the Reeb graph $R(f)$ encodes the topology of M and it has g loops for a manifold of genus g. For a Morse function on a surface of genus-zero topology, its Reeb graph has a tree structure.

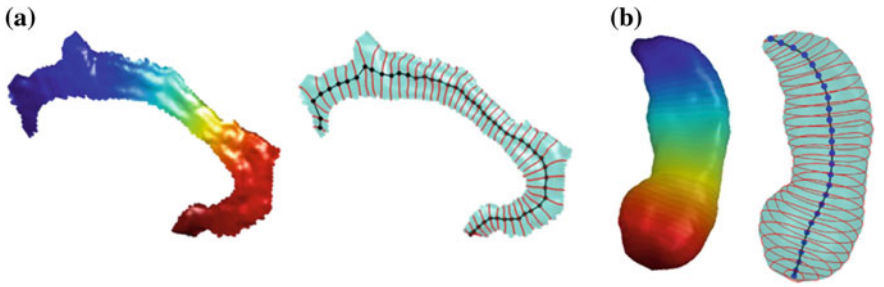


Fig. 4 Reeb graph computation via sampling level contours. **a** *Left* eigen-functions. *Right* Reeb graph as skeleton. **b** *Left* eigen-functions. *Right* Reeb graph as medial cores

For the Reeb graph to be useful in medical shape analysis, it is critical to build it from a function reflective of the underlying geometry. The height function was a popular choice in previous work, such as surface reconstruction from contour lines [33] and the analysis of terrain imaging data [34, 35], but it suffers from the drawback of being pose dependent and thus not intrinsic to surface geometry. Functions derived from geodesic distances [36–39] were proposed to construct pose invariant Reeb graphs in computer graphics. However, geodesic distances are known for their non-robustness when topological changes are involved [40, 41].

For the numerical construction of the Reeb graph, there are various approaches. Based on the intuition that Reeb graph encodes the relation of level sets on surfaces, we could simply sample a set of level contours on the surface and connect them using their neighboring relations. This approach is especially useful as a novel way of constructing skeleton or medial core of surfaces. Examples of level contours and the constructed Reeb graphs for the cingulate gyrus and hippocampus are shown in Fig. 4. For more complicated surfaces with large numbers of holes and handles, the sampling approach becomes computationally expensive because extremely dense sampling is needed to capture all the topological changes of level contours. To overcome this challenge, a novel algorithm is proposed in [42] that analyzes the topology of level contours in the neighborhood of critical points. As an illustration, the level contours in the neighborhood of a saddle point on a double torus are shown in Fig. 5b.

With increasing LB eigen-function, the level contour split into two contours. The mesh is then augmented such that the level contours become edges of the mesh. This makes it possible to use region growing on the augmented mesh to find branches of the Reeb graph. By scanning through all critical points (Fig. 5c), we can capture the topology of the level contours and build the Reeb graph (Fig. 5d). The surface partition generated by the Reeb graph ensures each surface patch is a manifold, so that further analysis of geometry and topology, such as the computation of geodesics, can be performed. For more details, see [42]. Since this method only needs to analyze the neighborhood of critical points, it can easily handle large meshes with hundreds of handles and holes. The Reeb graph provides a unified approach for topological and geometric analysis of surfaces. It has been successfully applied to develop a new

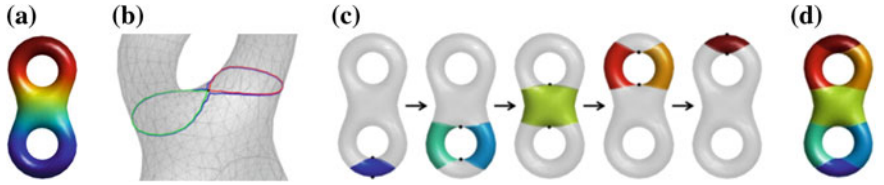


Fig. 5 Reeb graph computation as graph of critical points. **a** Eigen-function on a surface. **b** Level contours in the neighborhood of a saddle point. **c** Reeb graph construction. **d** Surface partition with its Reeb graph

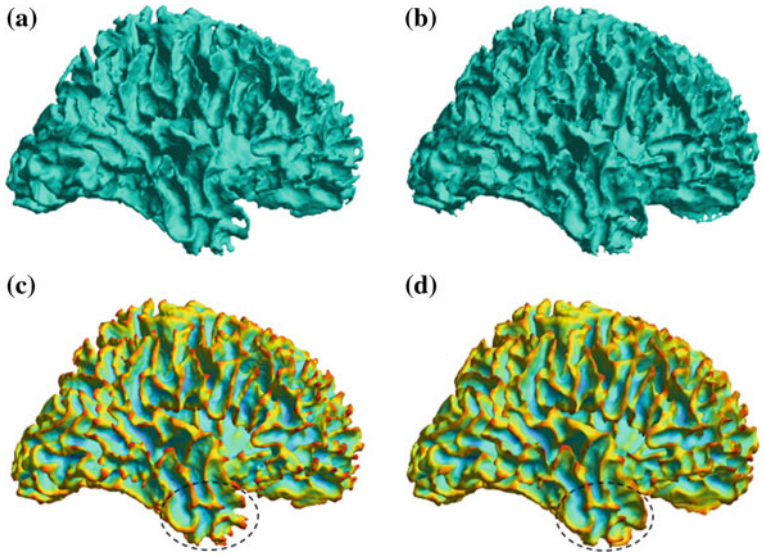


Fig. 6 Reeb analysis for unified corection of geometric and topological outliers in cortical surface reconstruction. **a** and **c** FreeSurfer reconstruction result before and after correction. **b** and **d** Reconstruction before and after Reeb-analysis based correction

way of topology and geometry correction in cortical surface reconstruction from MR images [42]. As an illustration, we show a cortical reconstruction example in Fig. 6a and c show the surface before and after correction from FreeSurfer [43]. Fig. 6b and d show the surface before and after correction generated by the Reeb analysis method. As highlighted in the circled region, clearly the latter generated better reconstruction.

3.3 LB Embedding Space

Using the LB eigen-system, we can build an embedding of the surface in the infinite dimensional l^2 space that has the advantage of being isometry invariant. This allows intrinsic comparison of surfaces and leads to novel algorithms for surface maps.

Let $\Phi = \{f_1, f_2, \ldots, \}$ a set of eigen-functions. The embedding $I_M^{\Phi} : M \rightarrow l^2$ is defined as [40]:

$$I_M^{\Phi}(x) = \left(\frac{f_1(x)}{\sqrt{\lambda_1}}, \frac{f_2(x)}{\sqrt{\lambda_2}}, \ldots, \frac{f_n(x)}{\sqrt{\lambda_n}}, \cdots \right) \forall x \in M. \tag{9}$$

For any two surfaces, (M_1, g_1) and (M_2, g_2), rigorous distance measures between them can be defined in the embedding space, e.g. the spectral l^2 distance, $d(M_1, M_2)$ [44]:

$$d(M_1, M_2) = \inf_{\substack{\Phi_1 \in \mathrm{B}(M_1), \\ \Phi_2 \in \mathrm{B}(M_2)}} \max \left(\int_{M_1} d_{\Phi_1}^{\Phi_2}(x, M_2) d_{M_1}(x), \int_{M_2} d_{\Phi_1}^{\Phi_2}(M_1, y) d_{M_2}(y) \right), \tag{10}$$

$$d_{\Phi_1}^{\Phi_2}(x, M_2) = \inf_{y \in M_2} \left\| I_{M_1}^{\Phi_1}(x) - I_{M_2}^{\Phi_2}(y) \right\|_2 \quad \forall x \in M_1 \,, \tag{11}$$

$$d_{\Phi_1}^{\Phi_2}(M_1, y) = \inf_{x \in M_1} \left\| I_{M_1}^{\Phi_1}(x) - I_{M_2}^{\Phi_2}(y) \right\|_2 \quad \forall y \in M_2 \,, \tag{12}$$

where Φ_1 and Φ_2 are any given LB orthonormal basis of M_1 and M_2, $B(M_1)$ and $B(M_2)$ denote the set of all possible LB basis on M_1 and M_2, $d_{M_1}(x)$ and $d_{M_2}(y)$ are normalized area elements, i.e., $\int_{M_1} d_{M_1}(x) = 1$ and $\int_{M_2} d_{M_2}(x) = 1$. Using this distance, intrinsic matching of surfaces can be performed. For example, it has been successfully applied to surface classification and sulcal landmark detection [44]. One critical result of the spectral l^2-distance is that it equals zero if and only if the two surfaces are isometric. This provides a new way of computing conformal surface maps, which are important tool for studying anatomical surfaces.

Given two surfaces (M_1, g_1) and (M_2, g_2), there exists a conformal metric wg_1, where $w : M_1 \rightarrow R^+$ is a positive function defined on M_1, such that the LB embedding $I_{M_1}^{\Phi_1^*}$ of (M_1, wg_1) under this new metric will be the same as the LB embedding $I_{M_2}^{\Phi_2^*}$ of M_2 because the LB embedding is completely determined by the metric, where Φ_1^* and Φ_2^* are the optimal basis that minimize the spectral l^2 distance. Since (M_1, g_1) and (M_1, wg_1) are conformal, and the two manifolds (M_1, wg_1) and (M_2, g_2) are isometric when the metric w is chosen so that the spectral l^2 distance is zero [44], we have a conformal map from (M_1, g_1) to (M_2, g_2) when we combine these maps [45]. Let Id denote the identity map from $I_{M_1}^{\Phi_1^*}$ to $I_{M_2}^{\Phi_2^*}$, the conformal map $\mu : M_1 \rightarrow M_2$ is defined as:

$$\mu(x) = \left[I_{M_2}^{\Phi_2^*} \right]^{-1} \circ Id \circ I_{M_1}^{\Phi_1^*}(x) \quad \forall x \in M_1 \,, \tag{13}$$

To numerically compute the conformal map that minimizes the spectral -distance, a metric optimization approach was developed [45] that iteratively updates the weight function w to deform the embedding such that it matches that of the target surface.

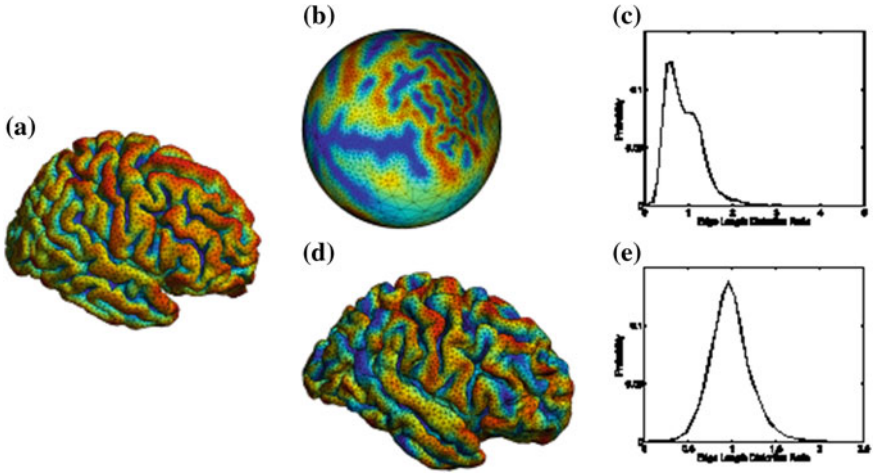


Fig. 7 A comparison of conformal maps. **a** Source surface M_1. **b** Conformal map to the unit sphere. **c** Metric distortion from M_1 to the sphere. **d** Conformal map from M_1 to a target cortical surface M_2. **e** Metric distortion from M_1 to M_2

As an illustration, we show in Fig. 7 the conformal maps from a cortical surface to the unit sphere or a target cortical surface.

The source surface shown in Fig. 7a is colored with its mean curvature. The conformal map to the unit sphere, shown in Fig. 7b, is computed with the approach in [46] that minimizes the harmonic energy. The conformal map to the target cortical surface, shown in Fig. 7d, is computed with the metric optimization approach that minimizes the spectral l^2-distance in the embedding space. The mean curvature on the source surface is projected onto the sphere and target surface using the maps. We can see large distortions of triangles in the spherical map. On the other hand, the gyral folding pattern is very well matched in the map to the target cortical surface. From the histogram plotted in Fig. 7c and e, we can see metric is much better preserved in the conformal map computed with the metric optimization approach.

Since the metric optimization approach can produce maps that align major gyral folding patterns of different cortical surfaces, a multi-atlas fusion approach has been developed to automatically parcellate the cortical surface into a set of gyral regions [45]. With metric optimization, a group-wise atlas is first computed in the embedding space. The conformal map from each atlas to the group-wise atlas is then computed. Using the property that the composition of conformal maps is still conformal, only one map to the group-wise atlas needs to be computed for an unlabeled surface. With maps to all the labeled atlases, a fusion approach can be developed to derive the labels on new surfaces. As a demonstration, we plotted in Fig. 8 the labeling results on the left and right hemispherical surfaces of two subjects. We can clearly see that excellent labeling performances have been achieved.

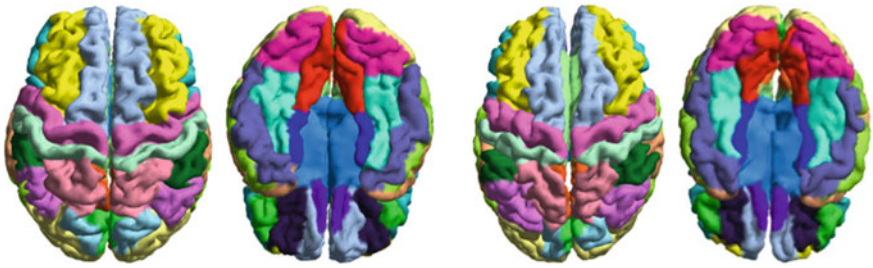


Fig. 8 Automatic labeling of cortical gyral regions

3.4 Application and Insights

The spectral analysis approach provides a general and intrinsic way of studying anatomical shapes. The LB embeddings and features derived from the spectral analysis have the advantage of being isometry invariant, which makes them robust to deformations due to variability across populations, normal development, and pathology. The mathematical foundation of the spectral analysis is applicable to shapes of arbitrary dimension and topology. With little adaptation, the spectral analysis techniques can be easily applied to many different anatomical shapes. It has been applied for mapping subcortical structures in brain mapping studies [47] and automated labeling of complicated cortical surfaces [45]. The Reeb graphs constructed from the eigen-functions could also be a general tool for topology analysis in medical imaging.

4 Function Representation

Brain shape can be described implicitly using various classes of functions, computed over the normalized image domain, i.e. images normalized with respect to similarity transform. This section discusses several methods by which this can be done: moment invariants, the distance transform, and linear projections: global and local frequency decompositions over Euclidean and spherical coordinates.

4.1 Moment Invariants

Let $I(x, y, z)$ be an image indexed by coordinates (x, y, z). Geometrical moments are defined as expectations computed across the image:

$$m_{ijk} = \sum_{x,y,z} x^i y^j z^k I(x, y, z). \tag{14}$$

Moments can be referred to by their order $n = i + j + k$, and are useful in shape description as they have intuitive geometrical interpretations. For instance, the 0th order raw moment m_{000} represents the sum of image intensities or shape volume. The 1st order raw moments $\{m_{100}, m_{010}, m_{001}\}$ normalized by m_{000} are the centers of mass of a shape: $\mu = \{\mu_x, \mu_y, \mu_z\} = \{m_{100}/m_{000}, m_{010}/m_{000}, m_{001}/m_{000}\}$. A key aspect of moments is that they can describe shape in a manner invariant to classes of image transforms that are irrelevant to shape description, e.g. similarity transform. Invariance to translation can be achieved by computing central moments about the center of mass:

$$u_{ijk} = \sum_{x,y,z} (x - u_x)^i (y - u_y)^j (z - u_z)^k I(x, y, z). \tag{15}$$

Invariance to scale changes is achieved by normalizing central moments by a suitable power of the 0th order moment:

$$n_{ijk} = \frac{u_{ijk}}{m_{000}{}^{n/3+1}}. \tag{16}$$

The 2nd order central moments correspond to spatial variance of an intensity pattern in 3D space, and 3rd order central moments provide a measure of skew. Moments are not generally invariant to rotation, but such invariance can be achieved by combining the moments in certain ways [48, 49]. A number of additional aspects of moments bear mentioning. Moments can generally be used to reconstruct the original image. Moments are related to global shape characteristics, and thus not be suitable for identifying local variations. A variety of moments exist other than geometrical moments, including Zernike moments, Legendre moments, and rotational moments.

4.2 Distance Transform

Let S be the set of all point locations along the boundary of a segmented brain structure. The distance transform of a shape S is an image $D(\bar{x})$, where the value at location $\bar{x}$ reflects the distance $d(\bar{x}, \bar{y})$ between $\bar{x}$ and the closest location $\bar{y} \in S$:

$$D(\bar{x}) = \inf\{d(\bar{x}, \bar{y}) : \bar{y} \in S\}. \tag{17}$$

There are a number of commonly-used distance functions, such as Euclidean distance and Manhattan distance. Figure 9 shows the Euclidean distance transform for a cross section of the corpus callosum. The distance transform can be considered as a local shape description, as the value of $D(\bar{x})$ is determined by the nearest boundary point in S. The value $D(\bar{x})$ can be interpreted as the radius of the largest sphere about point $\bar{x}$ lying within the boundary S. The set of ridges in $D(\bar{x})$ can be used for extracting medial axes as to be discussed in Sect. 5.

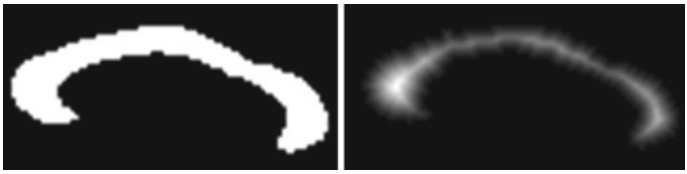


Fig. 9 Distance transform of corpus callosum

4.3 Frequency Decomposition

Shape in an image $I(x)$ can be represented implicitly by projection onto an orthogonal basis with the Fourier basis being arguably the most widely-used. Let $I(x, y, z)$ be a discrete image of size (N_x, N_y, N_z). The discrete Fourier transform (DFT) $\hat{I}(f_x, f_y, f_z)$ of an image is obtained by projection onto an orthonormal complex exponential basis:

$$\hat{I}(f_x, f_y, f_z) = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \sum_{z=0}^{N_z-1} I(x, y, z) e^{-i2\pi(xf_x/N_x + yf_y/N_y + zf_z/N_z)}. \quad (18)$$

The DFT has several properties that are of interest for shape representation. In particular, the magnitude of the power spectrum is invariant to translation and a significant portion of shape related information is embedded in the phase portion of the signal. Further, the DFT can be inverted to reconstruct the original image:

$$I(x, y, z) = \frac{1}{N_x N_y N_z} \sum_{f_x=0}^{N_x-1} \sum_{f_y=0}^{N_y-1} \sum_{f_z=0}^{N_z-1} \hat{I}(f_x, f_y, f_z) e^{i2\pi(xf_x/N_x + yf_y/N_y + zf_z/N_z)} \quad (19)$$

The DFT can be computed efficiently in $O(N \cdot logN)$ in the number of samples N using the fast Fourier transform (FFT) algorithm. The convolution theorem can be used to implement linear filtering operations efficiently in $O(N \cdot logN)$ time complexity in the Fourier domain. For natural objects, such as the brain, the power of the Fourier spectrum is concentrated in low frequency components, i.e. small values of (f_x, f_y, f_z), which arise from large-scale spatial structure.

The classical DFT operating on Cartesian image data is rarely used to characterize shape in the brain directly, as neuroanatomical structures are often represented in terms of a spherical topology. For example, the cortex is often inflated or flattened onto a sphere index by [50], and structures, such as the putamen, can be represented as a simply connected 3D surfaces [51]. Spherical harmonics (SH) generalize frequency decompositions to functions on the sphere. Let $I(\theta, \varphi)$ represent a function on the unit sphere parameterized in angular coordinates $\theta \in [0, \pi)$, $\varphi \in [0, 2\pi)$. The spherical harmonic representation is expressed as:

$$I(\theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \hat{I}(l, m) Y_{l,m}(\theta, \varphi), \tag{20}$$

where shape is represented by harmonic coefficients $\hat{I}(l, m)$ in a manner analogous to Fourier coefficients. The basis $Y_{l,m}(\theta, \varphi)$ is given by:

$$Y_{l,m}(\theta, \varphi) = k_{l,m} P_{l,m}(\cos\theta) e^{im\varphi}, \tag{21}$$

where $P_{l,m}$ is the Legendre polynomial of degree l and order m, and $k_{l,m}$ is a normalization factor. Although rotation dependent, coefficients $\hat{I}(l, m)$ can be used to compute rotation invariant shape descriptors [46]. Efficient computation of spherical harmonic coefficients is possible via fast discrete Legendre transform [52].

4.4 Local Frequency Decomposition

The aforementioned frequency decompositions are useful as global descriptors of brain structures. However, they are not suited for localizing specific areas of variability, e.g. a small fold on a surface. A collection of function-based techniques have been developed to characterize image shape locally. For encoding and reconstruction, an image may be projected onto a wavelet basis [53], consisting of translated and scaled versions of a mother wavelet function $\psi_{\bar{t},\sigma}(x, y, z)$:

$$a_{\bar{t},\sigma} = \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} \sum_{z=0}^{N_z-1} I(x, y, z) \psi_{\bar{t},\sigma}(x, y, z), \tag{22}$$

where $(\bar{t}, \sigma)$ are 3D translation and scaling parameters. Shape is captured by wavelet coefficients $a_{\bar{t},\sigma}$, where a significantly non-zero coefficient $a_{\bar{t},\sigma}$ indicates the presence of a local feature within a region of size σ about the image location $\bar{t}$. An important aspect of wavelet analysis is the design or choice of the mother wavelet function. The choice typically depends on the requirements of the image analysis task at hand, e.g. for compression and reconstruction, complete orthonormal basis is often used. However, complete wavelet basis may be overly sensitive, i.e. minor translations or scaling could result in dramatic changes in wavelet coefficients. Scale-space theory [54] shows that the Gaussian kernel is optimal for multi-scale image representation with respect to an intuitive set of axioms including translation and shift invariance, rotational symmetry, and non-creation or enhancement of local extrema. The Gaussian scale space can be considered as an overcomplete wavelet representation.

As in the case of global frequency representation, a significant research focus has been to translate local frequency decompositions from Euclidean space to the spherical topology as commonly used in brain shape analysis. Yu et al. proposed

an approach whereby wavelet coefficients are computed from the cortical surface mapped to the unit sphere by successively subdividing an icosahedron tilling [55]. Bernal-Rusiel et al. [56] compared smoothing using Gaussian and spherical wavelet techniques, finding that spherical wavelet smoothing may be better suited for preserving cortical structure in the case of spherical cortex data. Kim et al. [57] demonstrated that the wavelet transform can be generalized to an arbitrary graph over the cortical surface, which avoids sampling issues when mapping the cortical surface to a sphere.

4.5 Applications and Insights

In the seminal work of Hu [48], moment invariants were used to describe the shape of 2D image patterns. This was subsequently generalized to 3D volumetric data [49, 58]. In brain shape analysis, moment invariants have been used to describe the shape of cortical gyri [59] and regional activation patterns in functional MRI studies [60]. Principal component analysis of 2nd order moments has also been explored for describing and assigning an orientation axis to shapes, such as gyri [61]. Further, moment invariants have been employed in the context of abnormal shape variations, for instance, as a basis for image registration [62] or to identify aneurisms [63].

The distance transform has been used in segmenting brain structures, such as the corpus callosum and abnormal structure, such as tumors [64]. Also, distance transform is often used for initialization and evolution of contour-type segmentation algorithms, such as level sets. The distance transform can also be used as the basis for registration, as it remains unchanged under image translation and rotation [65].

Spherical harmonics are useful for characterizing the shape of 3D surfaces, such as ventricles [66]. It has also been applied to study shape changes of subcortical structures in patient population. In particular, global shape changes in the right caudate were observed in Schizophrenia patients [67] and shape changes in hippocampus were found to be implicated in major depression [68].

The local frequency representation finds extensive use in shape analysis of brain structures. The feature-based morphometry method characterizes shape in terms of extrema in the DoG scale-space, which correspond to distinctive, generic neuroanatomical patterns [18]. Several authors propose analyzing neuroanatomical patterns via wavelet coefficients computed across image scales, in a manner similar to voxel-based morphometry methods [69, 70]. Wavelets have been used for identifying morphological differences in the brain of healthy and diseased subjects [71]. Spherical wavelets have been used to generate a multiscale shape representation of the caudate nucleus and the hippocampus for the purpose of segmentation [72].

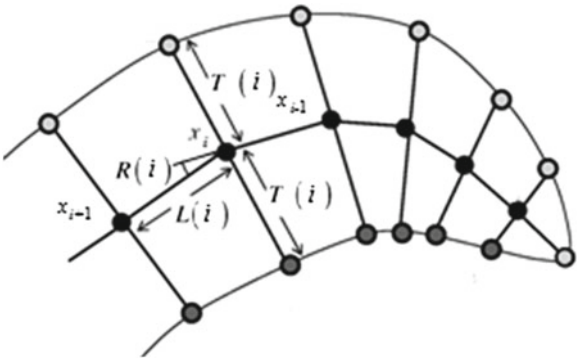


Fig. 10 Shape features associated with medial representation

5 Medial Representation

The heart of medial representation is to characterize a brain structure by generating its skeleton and describing its shape with respect to the skeleton. The skeleton is interchangeably referred to as medial axis and medial locus. We will use the term, "medial axis" from here onwards. A distinct advantage of medial representation is that it provides an intrinsic coordinate system that enables extraction of physically intuitive features, as illustrated in Fig. 10 [73].

For instance, assuming the medial axis has been discretized into a set of landmarks, x_i, which are commonly referred to as medial atoms, one can describe the local thickness of a brain structure $T(i)$, by measuring the distance from a medial atom to the boundary. Also, one can estimate the amount of local bending by examining the change in angles $R(i)$, between segments connecting consecutive pairs of medial atoms. Furthermore, one can describe local elongation based on the distance $L(i)$, between the medial atoms. The formal mathematical definitions of medial axis and medial atoms are presented in Sect. 5.1.

For studying the variability of a brain structure, a subject correspondence needs to be established. This implies all subjects' medial axes must have the same branching topology. However, medial axis extraction is extremely sensitivity to boundary noise, as illustrated in Fig. 11 [74]. Any small perturbations of the boundary can result in considerable changes in the branching topology. Hence, the extracted medial axes often do not have a clear correspondence across subjects. One common strategy is to deal with boundary noise is to prune the medial axis. Methods for extracting medial axis as well as their respective pruning strategies are discussed in Sect. 5.2.

Another strategy for establishing a subject correspondence is to predefine a branching topology that well captures the shape of a brain structure for all subjects and deform a common medial axis to each subject's data. This strategy is widely referred to as "M-Reps" [78, 79]. The key advantage of M-Reps is that it provides a one-to-one correspondence between medial atoms across subjects while only

Fig. 11 Medial axes of the corpus callosum. *First column* Segmented corpus callosum. *Second column* Conventional method based on Blum's definition of medial axis [75]. *Third column* A leading curve evolution-based method [76]. *Last column* Group-wise pruning strategy [77]

affecting the fit to the boundary mildly in practice [80]. Since medial axis lives on a Riemannian symmetric manifold, its deformation and analysis requires a special set of statistics. We describe M-Reps and its associated statistics in Sect. 5.3.

5.1 Medial Axis and Its Building Blocks

The notion of medial locus of an object was first proposed by Blum [75]. In the present context, an object would be a brain structure in either 2D or 3D space. Mathematically, if we let A be a set in R^n (n-dimensional Euclidean space) representing a brain structure ($n = 2$ or 3) and let $B_r(x)$ be a maximal inscribed ball of A defined as:

$$B_r(x) = \{y \in R^n : ||x - y|| < r\}, \quad s.t. \quad B_r(x) \subset A, \tag{23}$$

and $B_r(x) \not\subset B' \subset A$ for any other ball B', then the medial locus of A is the set of all pairs $(x, r) \in R^n \times R^+$, where R^+ is the space of positive real numbers. The medial axis (by Blum's definition) is the set of positions $\{x\}$. The intuition behind this definition of medial axis is that a maximal inscribed ball is tangent to an object at two or more locations with the centroid being equidistant from the tangent points (Fig. 12a). Thus, the union of centroids of all maximal inscribed balls would constitute the medial axis. In 2D, the medial axis corresponds to a smooth curve, with the longest curve considered as the trunk and the remaining segments referred to as branches. In 3D, the medial axis corresponds a smooth surface (Fig. 12b).

In practice, a medial axis is often represented by a discrete set of medial atoms (Fig. 13). Each atom is a tuple $(x, r, n_0, n_1) \in R^n \times R^+ \times S^{n-1} \times S^{n-1}$, where x is the position of the atom, r is the radius of the maximal inscribed ball with x being the centroid, n_0 and n_1 are unit-norm vectors pointing towards the tangent points, and S^{n-1} is a $n-1$ dimensional unit sphere. The corresponding boundary points y_0 and y_1 of a medial atom can be reconstructed by: $y_j = x + rn_j$, where $j = 0$ or 1.

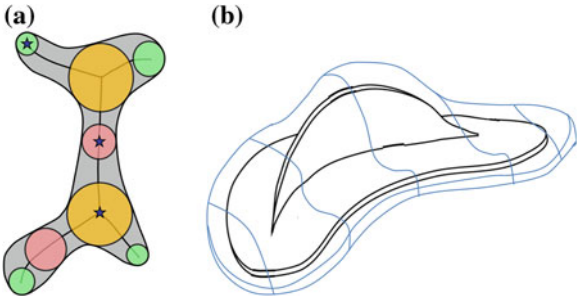


Fig. 12 Examples of medial axis in 2D and 3D. **a** 2D. **b** 3D

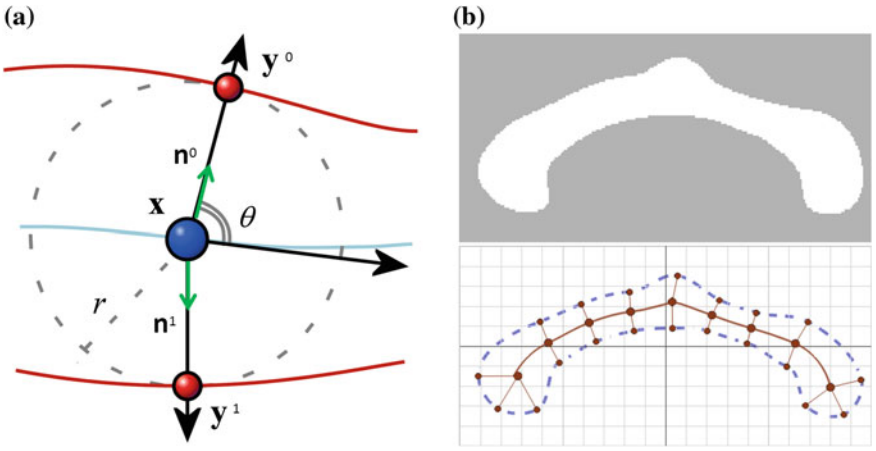


Fig. 13 Medial atom and discrete medial axis [83]. **a** Medial atom. **b** Discrete medial axis

Note that there is a continuous counterpart of discrete medial representation, see e.g. [81] and [82].

Given the medial axis of a brain structure, intuitive physical features can be easily extracted. For instance, local thickness is given by r_i of medial atom i, local elongation is given by $|x_i - x_{i-1}|$, and local bending is given by the angle between the vectors $x_i - x_{i-1}$ and $x_{i+1} - x_i$. For comparing across subjects, the features need to be normalized with respect to the intrinsic coordinate system, e.g. dividing r by $|x_i - x_{i-1}|$. If medial atoms can be correctly placed within all subjects' segmented brain structures, classical statistics can be applied to isolate the location of variability. The key is establishing a correspondence through generating medial axes that have the same branching topology across subjects. This crucial issue is discussed in the next two sections.

5.2 Medial Axis Extraction

As was shown in Fig. 11, medial axis extraction is extremely sensitive to boundary perturbations. We describe in this section three widely-used approaches for medial axis extraction and their associated strategies against boundary noise, namely Voronoi skeleton, boundary evolution, and core tracking. Group-wise strategies that ensure the extracted medial axes have the same topology across subjects are also discussed. In computer vision, morphological operators, such as erosion, are often used. The drawbacks of erosion are that it is very sensitive to the rasterization of the objects and has difficulties discerning local geometry near points of branching [79]. Detail on morphological methods for medial axis extraction can be found in e.g. [84].

5.2.1 Voronoi Skeleton

Given a set of points sampled from the boundary of a brain structure, a Voronoi skeleton is generated by first computing the Voronoi diagram of the boundary points. In 2D, a Voronoi diagram divides the space on which the boundary points live into regions such that the set of points within each region is closest to a particular boundary point than to any other boundary points. A simple example with 6 points is shown in Fig. 14a. The line segments in the Voronoi diagram correspond to loci equidistant from two boundary points and the locations at which the line segments meet are equidistant to three or more boundary points. As apparent from Fig. 14b, the line segments of the Voronoi diagram that fall completely within the boundary of a given brain structure constitute the medial axis. Medial axis generated in this manner is very sensitive to boundary perturbation, which could result in many spurious branches [83]. To obtain more robust Voronoi-based medial axis, a number of measures that captures the significance of a line segment has been introduced for pruning the medial axis [85–87]. For instance, one may remove a line segment and assess the impact on the overall match between the implied boundary and the actual boundary of a given brain structure. The choice of threshold for removing a line segment, however, is nontrivial. The procedure for generating Voronoi skeleton in 3D is similar, but more difficult to visualize and much harder to implement [85, 86].

5.2.2 Boundary Evolution

This approach is based on the grassfire analogy. Specifically, imagine a brain structure is a patch of grass. If we simultaneously light up the boundary, the fire fronts would propagate towards the inner part of the brain structure and eventually meet at some points, referred to as shocks. The set of all shocks is the medial axis. Mathematically, if we let $C(p)$ be a parameterization of the boundary of a brain structure, the evolution of $C(p)$ into a medial axis can be formulated as the following partial differential equation [88]:

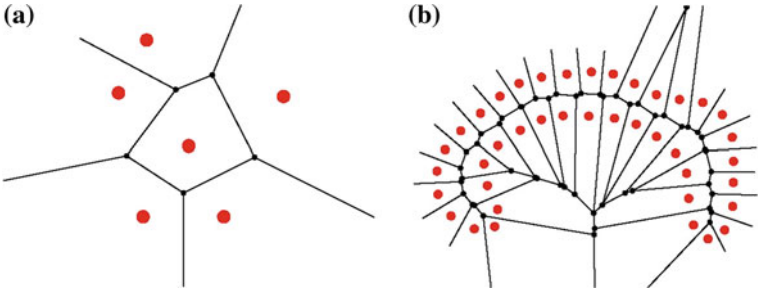


Fig. 14 Voronoi diagrams [83]. **a** Simple 6 points example. **b** Corpus callosum

$$\frac{\partial C(t,\, p)}{\partial t} = (\alpha - \beta\kappa)N, \quad C(0,\, p) = C(p), \tag{24}$$

where t is time, κ is the curvature of $C(p)$ at p, and N is the unit normal of $C(p)$ at p. α and β can be functions or simply constants. If $\beta = 0$, (24) models exactly the grassfire phenomenon with fire fronts propagating at speed α. $\beta\kappa N$ is a regularization term.

Standard numerical simulations of (24) run into problems when propagation fronts form shocks, points at which normal and curvature are not defined [89]. Level set approaches facilitate continuation of front propagation in the presence of shocks, but do not inherently provide a mechanism to detect shocks, which is crucial for extracting medial axis. A workaround to these problems is to model front propagation using a Hamilton-Jacobi formalism [89]:

$$\partial p = -\frac{\partial H}{\partial q} = (0,\ 0,\ 0), \quad \partial q = -\frac{\partial H}{\partial p} = -(D_x,\ D_y,\ D_z), \tag{25}$$

where D is the distance transform of a given brain structure's boundary, $||\nabla D||$ is the magnitude of its gradient, $H = 1 - ||\nabla D||$ is the Hamiltonian function, $q = (x, y, z)$, and $p = (D_x, D_y, D_z)$. Under this formalism, the average outward flux of the vector field ∂q about a point can be used to determine whether that point is medial or not. Specifically, non-medial points have close to zero outward flux, whereas medial points have large negative outward flux. Pruning is accomplished by removing each point one by one starting from the point with an outward flux close to zero and stopping if the removal of subsequent points alters the topology of the brain structure.

5.2.3 Core Tracking

In contrast to Voronoi skeleton and boundary evolution, core tracking can operate directly on gray scale volumes. This approach thus mitigates the need for explicitly specifying the boundary of a brain structure. But to keep the method discussion

consistent, we will describe core tracking for the case in which segmented brain structures are given. Central to core tracking is the definition of medialness that measures how well the position and orientation of the implied boundary points of a candidate medial atom match edge-like structures in a volume. The most widely-used medialness measure of a candidate atom $m = \{x, r, n_0, n_1\}$ is based on the intensity gradient of a volume I at boundary points y_0 and y_1:

$$M(m) = \nabla_\sigma I(y_0) \cdot n_0 + \nabla_\sigma I(y_1) \cdot n_1, \tag{26}$$

where $\nabla_\sigma I(y_j)$ is the intensity gradient of I at y_j, $j = 0$ or 1, convolved with an isotropic Gaussian kernel of width σ that governs the sensitivity of the extracted core to boundary noise. Core tracking proceeds by first placing an atom on a user-defined starting position and optimizing the parameters of this atom with respect to (26). A core is then tracked by placing the next atom a certain distance away from the previous one, optimizing its parameters, and repeating the process until some termination criteria are met. This algorithm provides the trunk of the medial axis, but not its branches. To find branches, one way is to use corner detectors for identifying potential branching locations and initialize new core at those locations [90].

5.2.4 Group-wise Medial Axis Extraction

Although the pruning strategies associated with the above medial axis extraction approaches help provide more robust medial axes, there is no guarantee that the extracted axes would have the same branching topology across subjects. The most widely-used strategy to deal with this problem is to enforce a strict correspondence by deforming a common predefined medial axis onto all subjects' segmented brain structures, which we will discuss in the next section. Another strategy is to jointly prune the medial axes of all subjects using group information [77]. Given the raw medial axes of all subjects, the idea is to compute a set of features that characterizes the branches and identify branches that are commonly present across subjects based on these branch features. For example, one may use the number of medial atoms within a branch, the centroid location of the branch, and the angle of the branch with respect to the trunk to characterize a branch. Correspondence can then be drawn by applying e.g. bipartite graph matching between all pairs of medial axes. Additional refinements to increase robustness are described in [77].

5.3 M-Reps

The key idea behind M-Reps is to deform a common medial axis to all subjects' brain structures, so that a subject correspondence between medial atoms is inherently established. For the 2D case, a method based on a snake-like algorithm has been proposed [91]. Specifically, if a distance transform is applied to a segmented

brain structure, the ridges of the distance map would correspond to the medial axis. Assuming a fixed representative medial axis can be defined, this method first approximates the locations of the endpoints of the medial axis based on maximum positive curvature of the boundary. Fixing the endpoint locations, a snake-like algorithm is then used to move the remaining medial atoms along the gradient of the distance map. Once the positions of these medial atoms are found, the endpoint locations are adjusted based on how well the implied boundary segments fit the original boundary.

In the general 3D case, methods based on Riemannian statistics have been proposed [78, 80, 92]. Let $M(n)$ be the space of M-reps with n medial atoms. Recall that a medial atom, m, is defined as a tuple $(x, r, n_0, n_1) \in R^3 \times R^+ \times S^2 \times S^2 = M(1)$. Thus, a M-rep model with n atoms is a point in $M(n) = M(1)^n = (R^3 \times R^+ \times S^2 \times S^2)^n$. Since R^3, R^+, and S^2 are symmetric spaces and products of symmetric spaces are also symmetric spaces, $M(n)$ is a symmetric space [80]. In $M(n)$, deforming a common medial axis to each subject's segmented brain structure is done by applying a separate similarity transform, S, on each medial atom, m:

$$S \cdot m = S.(x, r, n_0, n_1) = (sRx + w, sr, Rn_0, Rn_1), \tag{27}$$

where R is a 3×3 rotation matrix, w is a 3×1 vector for translation, and s is a scaling factor. S is optimized for each atom based on the match between the implied boundary and the segmented boundary with some tolerance to account for boundary noise [78].

Similar to point-based analysis in which the point clouds need to be rigidly aligned to account for global scale and orientation variability, the medial axes of the subjects also need to be rigidly aligned prior to analyzing their shape variability. For this, the notion of distance in $M(n)$ is required, which is based on the concept Riemannian log map. If we let $p = (0, 1, p_0, p_1)$ be the base point, where $p_j = (0, 0, 1)$, $j = 0$ or 1, the log map of a medial atom, $m = (x, r, n_0, n_1)$, is given by:

$$Log_p(m) = \left(x, \log r, Log_{p_0}(n_0), Log_{p_1}(n_1)\right), \tag{28}$$

$$Log_{p_j}(n_j) = \left(n_{j1}\frac{\theta}{\sin\theta}, n_{j2}\frac{\theta}{\sin\theta}\right), \quad j = 0 \text{ or } 1, \tag{29}$$

where $n_j = (n_{j1}, n_{j2}, n_{j3})$ and $\theta = \arccos(n_{j3})$ is the spherical distance from the base point p to n_j, $j = 0$ or 1. If we further let $u = (x, \rho, v_0, v_1) \in T_pM(1)$ be the tangent vector at point p, where x is the positional tangent, ρ is the radius tangent, and v_0 and v_1 are spherical tangents, the geodesic distance between p and m is given by [80]:

$$d(p, m) = \left\| Log_p(m) \right\| = \|u\|, \tag{30}$$

where $\|u\| = \left(\|x\|^2 + \bar{r}^2(\rho^2 + ||v^1|| + ||v^2||)\right)^{\frac{1}{2}}$. The scaling by the average radius of all atoms $\bar{r}^2$ is to account for differences in units between x, r, n_0, and n_1 [80]. Given (28) to (30), the geodesic distance between two atoms m_1 and m_2 is defined as:

$$d(m_1, m_2) = \left\| Log_{m_1}(m_2) \right\|_{m_1}. \tag{31}$$

With this notion of distance in $M(1)$, N medial axes $M_1, \ldots, M_N$, can be rigidly aligned by minimizing the cost below with respect to similarity transforms $S_1, \ldots, S_N$ [80]:

$$d(S_1, \ldots, S_N, M_1, \ldots, M_N) = \sum_{i=1}^{N} \sum_{j=1}^{i} d(S_i \cdot M_i, S_j \cdot M_j)^2, \tag{32}$$

$$d(M_i, M_j) = \sum_{k=1}^{n} d(m_{ki}, m_{kj})^2, \tag{33}$$

where m_{ki} is the kth medial atom of subject i's medial axis. Once the medial axes of the subjects are aligned, one can build a statistical model to capture the shape variability analogous to the active shape approach [4]. In $M(n)$, the mean μ is defined as:

$$\mu = \min_{M \in M(n)} \sum_{i=1}^{N} d(M, M_i)^2. \tag{34}$$

An elegant algorithm for solving (34) can be found in [80]. Given μ, The covariance Σ in $M(n)$ is defined as:

$$\Sigma = \sum_{i=1}^{N} u_i u_i T, \quad u_i = Log_{\mu}(M_i). \tag{35}$$

Eigenvalue decomposition can then be applied to Σ to identify principle modes of variability and the amount of variance. This generalization of PCA is known as principle geodesic analysis (PGA), and is often used to impose shape constraints for segmentation of new samples of previously analyzed brain structure [78]. For more details on the mathematics of medial representation, see [79].

5.4 Applications and Insights

Medial representation is widely employed in computer vision. For neuroimaging, it has mainly been used for studying subcortical structures with relatively clearer regional boundaries than cortical structures. For instance, medial representation has been applied to study shape changes in hippocampus of Schizophrenia patients [93, 94]. It has also been applied to study differences in ventricles of twins [78] as well as corpus callosum of Multiple Sclerosis patients [95]. Due to the difficulties in drawing subject correspondence between cortical structures, which tend to have unclear boundaries, medial representation is much less explored for cortical structure analysis.

6 Deformetrics: Shape Statistics with Space Deformations

Fundamentally, a one-to-one point correspondence might not exist between instances of a brain structure across subjects. For example, homologous fiber bundles extracted from diffusion MRI data of two individuals typically have different number of fibers and each fiber has a different number of points (Fig. 15). Enforcing a point correspondence across each fiber is not only nearly impossible in practice, but also not sensible from an anatomical point of view since the estimated fibers are not true neuronal fibers. The overall problem of point correspondence can be mitigated using deformetrics, which is the topic of this section. Specifically, correspondence creation can be formulated as learning a one-to-one 3D space deformation that aligns the brain structures of interest. The alignment is meant in a loose way in that the geometry of the objects should be matched, but without enforcing one point of a given object to match with another point of the same object in another individual. Instead, point correspondence is learned on the whole 3D space as an output of the deformetics approach, such that the local organization of homologous brain features is preserved across individuals. There are two key ingredients to deformetrics. The first ingredient is the metric on currents for measuring alignment accuracy without point correspondence, as described in Sect. 6.1. The second ingredient is a model for generating one-to-one smooth deformations, as described in Sect. 6.2. Combining these ingredients provides a flexible framework for aligning two instances of a structure or a structure complex of various topologies, as discussed in Sect. 6.3. Further, the framework can be extended for group analysis by drawing on concepts from Riemannian statistics, as discussed in Sect. 6.4.

6.1 Metric on Currents

The first pillar of deformetrics is the definition of similarity between instances of a brain structure or a structure complex. This measure is used for quantifying the accuracy of the alignment between homologous structures in driving the estimation of 3D space deformations between shape instances. The choice of measure should be adapted to the kind of data. If there exists homologous points across individuals, the sum of squared differences between point positions can be used as similarity measure. However, if such correspondence is absent, the metric on currents or varifolds would be more suitable [96, 97].

The idea behind metric on currents is to treat curves or surfaces in 3D as a physical object that can be excited by vector fields. The response of a surface mesh to an "excitation", namely a 3D vector field, is the flux of the vector field through the surface. For a curve, it is the path integral of the vector field. If two surfaces (or two curves) have the same response to any excitation, then they are superimposed. The maximum difference gives a quantification of the dissimilarity between the given entities. This definition requires very few assumptions about the surfaces and

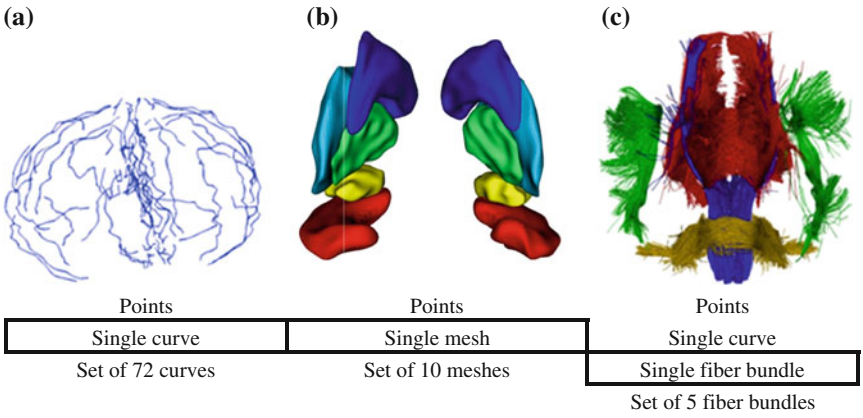


Fig. 15 Segmented brain structures exhibiting a natural hierarchy from single points to more complex objects. Homology is usually defined at a given level: **a** individual sulcal curves (central sulcus, Sylvian fissure, etc.), **b** meshes of basal ganglia (hippocampus in *red*, caudate in *blue*, etc.) or **c** fiber bundles (cortico-spinal tract in *blue*, corpus callosum in *red*, etc.). The currents metric is applicable to all different levels of homology, not just points

curves. The surfaces and curves only need to be rectifiable, i.e. integrals defined, but no requirement is made about their topology. The surfaces and curves may comprise multiple parts, may have holes, or may even be generated from different meshing or sampling.

In mathematical terms, surfaces (or curves) are seen as linear operators on a test space of vector fields W. Given a norm on the test space W, we can define the spectral norm of the operator as: $\min_{\omega=1}|S(\omega)|$, where $S(\omega)$ denotes the flux of the vector field ω through S and the minimization is over the unit ball in the test space. The key is to assume the test space of vector fields W has the structure of a reproducible kernel Hilbert space (RKHS), which comprises the convolutions between squared integrable vector fields and a smoothing kernel K^W. The kernel controls for the frequency band of the test vector fields. Specifically, the lower the band is, the slower the variations of the vector field. Also, with fast varying vector fields, the metric is more sensitive to small differences between meshes. Conversely, with slow varying vector fields, the metric is more loose and smoothes out small variations. Based on the RKHS structure, the spectral norm has a closed-form:

$$\|S\|_W^2 = \int_S \int_S K^W(x, y) n(x)^T n(y) d\sigma(x) d\sigma(y), \tag{36}$$

where $n(x)$ represents the normal to the surface S and $d\sigma(x)$ is the Lebesgue measure. K^W is e.g. a Gaussian kernel: $K^W(x, y) = exp(-||x-y||^2/\sigma_W^2)$ with σ_W^2 controlling the regularity of the test vector fields, namely their frequency bands. The norm (36) comes from an inner-product: $||S||_W^2 =< S, S >_W$, where the inner-product between two generic surfaces S and S' is given by:

$$< S, S' >_W = \int_S \int_{S'} K^W(x, y) n(x)^T n'(y) d\sigma(x) d\sigma'(y). \quad (37)$$

where $n(x)$ and $n'(x)$ denote the normal to the surface S and S'. For curves, a similar formula applies except normals are replaced by tangents and flux integrals replaced by path integrals. For unstructured point sets, normals are replaced by scalar weights and integrals replaced by sums. As evident from (36) and (37), no one-to-one point correspondence between geometric primitives nor similar sampling are required.

Currents form a vector space. Given two surfaces S and T, the sum of the flux through S and through T equals the flux through the union of both surfaces, i.e. $(S+T)(\omega) = S(\omega)+T(\omega)$. Also, flipping the orientation of the surface is equivalent to changing the current S into $-S$ (flux is opposite). Therefore, the current $S-T$ could be seen as the union of the two surfaces, where the normals of T have been flipped. This can be treated as a new surface (made of two parts), and the norm $||S - T||^2$ could be computed using (36). Equivalently, one could decompose the norm into $||S - T||^2 = ||S||^2 + ||T||^2 - 2 < S, T >$, and apply (36) and (37).

Currents norm is sensitive to translation. If S and T are two shifted versions of the same contour, the resulting norm would be large. A matching that tries to minimize the currents norm will first push one contour in the direction of the other. The norm will be further minimized by adjusting the shape of the contours. The kernel size σ_W plays an important role in this respect. Variations in shape that are much smaller than this parameter are smoothed out by the kernel, and therefore, are not taken into account for aligning shapes. If contours are shifted by a distance that is much greater than the kernel size, then the inner-product between them vanishes. In this case, the two contours are orthogonal in the space of currents. There is nearly no attraction force between them, and the optimization of the currents norm using deformation will be difficult. The value of the parameter should be selected based on the problem. Note that a sum of kernel with decreasing size is still a kernel and could be use in a multi-scale setting.

The algebraic operations permitted in the space of currents provide great flexibility in the way that similarity metric can be defined, while taking into account the "right" level of homology. Fig. 15 shows three typical examples. In Fig. 15a, sulcal curves on the cortex are labeled consistently across individuals. Denoting $L_1, \ldots, L_N$ and $L'_1, \ldots L'_N$ as the set of curves for two individuals, we can define a similarity measure as: $\Sigma_i ||L_i - L'_i||^2$ based on (36). Note that we can also add a weighting factor in the sum for balancing the contributions of different sulcal curves. This measure does not assume points on curves have a one-to-one correspondence, but does assume that the input are corresponding curves, i.e. not matching sylvian fissure with central sulcus for example. Similarly, a similarity measure between sets of sub-cortical structures is given by: $\Sigma_i ||S_i - S_i'||^2$, where the currents metric is the one for surfaces instead of for curves. For fiber bundles, the homology is at the level of the different bundles that are labeled consistently across individuals, e.g. cortico-spinal tract and arcuate fasciculi. Let B_i and B'_i be the ith bundle in two individuals, such that B_i is made of n fibers $F_{i,k}$, i.e. $B_i = \Sigma_k F_{i,k}$ (representing the union of the fibers in currents), and B'_i is made on m fibers $F'_{i,k}$, i.e. $B'_i = \Sigma_p F'_{i,p}$, where m may be

different from n. The similarity measure between homologous fiber bundles is given by: $\Sigma_i ||B_i - B'_i||^2 = \Sigma_i ||\Sigma_k F_{i,k} - \Sigma_p F'_{i,p}||^2$. Using this measure permits differences in the geometry of fiber bundles to be captured without requiring the same number of fibers within each bundle. If the union of sulcal curves, fiber bundles, and sub-cortical structures is of interest, one could combine all the previous similarity measures in a weighted sum in defining a global similarity measure on structure complexes.

In practice, surfaces are represented by meshes, and curves by polygonal lines. In this case, normals (resp. tangents) are constant over faces of the mesh (resp. segments of the line). Therefore, integrals in (36) and (37) can be efficiently approximated by a finite sum over mesh cells. Computing the norm of a mesh with N cells has a complexity of $O(N^2)$. Since the above equations are essentially convolutions, one could project normals or tangents at the nodes of a regular lattice and compute the norm using FFT, which drastically reduce computational complexity. The numerical error is in the order of $O(\Delta^2/\sigma_W^2)$, where Δ is the grid's step and σ_W is the kernel parameter [98]. Computing addition and scalar multiplication is nothing more than concatenating normal vectors or tangents and scaling them.

6.2 A Generic Model for 3D Diffeomorphisms

The second pillar of deformetrics is a model to build smooth one-to-one 3D space deformations. The deformation needs to be smooth to preserve the topology of the tissue (no shearing or tearing) and the local organization of the structures. Smooth deformations with smooth inverse are called, "diffeomorphisms," which appear naturally from integration of dynamical systems [99–103]. Inspired by mechanics and following [100, 104], a flow of diffeomorphisms, i.e. a time-indexed family of diffeomorphisms, can be defined by the integration of a time-varying velocity field. Mathematically, the flow of diffeomorphisms $x \rightarrow \phi(t, x)$ at time t (for all $x \in R^3$) is the solution of the differential equation:

$$\frac{\partial \phi(t, x)}{\partial t} = v(t, \phi(t, x)), \phi(0, x) = x, \tag{38}$$

where $v(t, x)$ is a time-varying velocity field. The trajectory of a point x_0 under (38) is given by $x(t) = \phi(t, x_0)$, which is the integral curve of the ODE: $\dot{x}(t) = v(t, x(t))$ starting at point $x(0) = x_0$. The trajectories of several points could be written in matrix form as:

$$\dot{X}(t) = v(t, X(t)), \quad X(0) = X_0, \tag{39}$$

where X_0 is the concatenation of point coordinates and $v(t, X(t))$ acts coordinate-wise. The flow of diffeomorphisms is entirely parameterized by the time-varying velocity field $v(t, x)$. In particular, the smoothness of v determines the smoothness of the diffeomorphisms. To control the regularity of the velocity field, if it is often assumed that the velocity field belongs to a RKHS with kernel K^V and has the form:

$$v(t,x) = \sum_{p=1}^{N} K^V(x, c_p(t))\alpha_p(t), \tag{40}$$

where $c_p(t)$ are the positions of the N control points at time t and $\alpha_p(t)$ are momentum vectors attached to the control points. K^V is typically set as a Gaussian function with parameter σ_V. The key motivation for the RKHS assumption is that it provides a norm for the velocity at every time t:

$$\|v(t,.)\|_V^2 = \sum_{p=1}^{N}\sum_{q=1}^{N} K^V(c_q(t), c_p(t))\alpha_p(t)^T\alpha_q(t). \tag{41}$$

Note that (40) is like applying spline interpolation on the vectors $\alpha_p(t)$ at control points $c_p(t)$. In the general case, a velocity field in the RKHS is parameterized by a continuum of control points [104]. In this approach, a finite-dimensional approximation is enforced, and this parameterization (i.e. the control point positions) will be optimized for each data set. With these assumptions, the flow of diffeomorphisms is determined by the time-varying vectors $c_p(t)$ and $\alpha_p(t)$. Assuming the control points move in space with the deformation, their positions would satisfy the ODE:

$$\dot{c}_p(t) = v(t, c_p(t)) = \sum_{q=1}^{N} K(c_p(t), c_q(t))\alpha_q(t). \tag{42}$$

Given a fixed initial condition $c_p(0)$, there is an infinite number of choices for the time-varying momentum vectors $\alpha_p(t)$ that reach the same final positions $c_p(1)$ in unit time. Following mechanical principles, one logical choice would be to select the momentum trajectories that minimize the kinetic energy needed to reach the final configuration $\int_0^1 ||v(t,\cdot)||_V^2 dt$. Such momentum trajectories satisfy the ODE:

$$\dot{\alpha}_p(t) = -\sum_{q=1}^{N} \alpha_q(t)^T\alpha_p(t)\nabla_{c_p(t)}K(c_p(t), c_q(t)). \tag{43}$$

With (43), the flow of diffeomorphisms can now be entirely parameterized by initial conditions, i.e. initial position of control points and initial momentum vectors. Given these initial conditions concatenated into a single vector S_0, control points and momentum vectors at later time t can be determined from the two sets of coupled ODEs (42) and (43). These two ODEs in a matrix form are given by: $\dot{S}(t) = F(S(t))$ with $S(0) = S_0$, where top rows of F are (42) and bottom rows are (43). Once this ODE is integrated, the trajectory $X(t)$ of any set of points X_0 is given by the integration of (39). Since the velocity field is fully parameterized by S_0, (39) can be re-written in matrix form as: $\dot{X}(t) = G(X(t), S(t))$ with $X(0) = X_0$. This construction of diffeomorphisms holds a number of important properties. First, under

the minimum energy principle, one can show that the norm of the velocity is constant in time: $||v(t,\cdot)||_V = ||v(0,\cdot)||_V$, such that the integration of the equations of motion describe a geodesic path in the group of diffeomorphisms [105], and the value of $||v(0,\cdot)||_V$ is the length of this geodesic path. The logarithm (or tangent-space representation) of the path to the identity map ($\phi(0,x) = x$) is given by the initial momentum vectors at initial control point positions. These momentum vectors live in a finite-dimensional sub-space of the RKHS. Their norm (and hence inner-product) is given by the kernel K^V, which is key to performing statistics on deformations. Second, the use of control points allows the dimension of the sub-group of diffeomorphisms to be controlled, which is important for statistical purposes when the number of available data is limited. Estimating a space deformation amounts to learning a finite-dimensional vector S_0, namely position of initial control points and their momentum vectors. Furthermore, the parameter of the kernel σ_V determines the frequency band of the velocity field. The larger σ_V is, the more regular and rigid the deformation. The intuition is that a deformation results from the combination of local deformations of patches of diameter σ_V. Motions of points at distance much smaller than σ_V are highly correlated, whereas points at larger distance may have independent trajectories.

6.3 Matching Anatomical Configurations

Let $O_1, \ldots, O_N$ be a set of N labeled structures of an individual, and $O_1', \ldots, O_N'$ the homologous structures in another individual. The goal is to find a diffeomorphism ϕ that best aligns the two sets of structures based on the metric on currents. Assuming the diffeomorphism $\phi = \phi(1,\cdot)$, which is the end point (at unit time $t = 1$) of a flow of diffeomorphisms $\phi(t,\cdot)$, the flow (and hence the deformation of interest $\phi(1,\cdot)$) is entirely parameterized by the initial positions of control points and momentum vectors, S_0. $O_1, \ldots, O_N$ can be unstructured point sets, polygonal lines or surface meshes, and they move in space to $\phi(O_1), \ldots, \phi(O_N)$. Let X_0 be the concatenation of all the vertices of all structures $\{O_i\}$, then $X(1)$ gives the vertices of all structures $\phi(O_i)$, which results from the integration of the set of two coupled ODEs at time $t = 1$:

$$\dot{X}(t) = G(X(t), S(t)), X(0) = X_0 \tag{44}$$

$$\dot{S}(t) = F(S(t)), \quad S(0) = S_0. \tag{45}$$

The position $X(1)$ depends only on S_0, since X_0 is fixed. The goal is thus to find the best S_0, so that the structures $\phi(O_i)$ are the most aligned with their homologous O_i'. Defining a currents metric as the weighted sum $\Sigma_i w_i ||\phi(O_i) - O_i'||_W^2$ and denoting it as $A(X(1))$, this metric essentially just depends on the positions of the vertices of the deformed shapes $\phi(O_i)$ and their meshing. Note that the deformation takes only the vertices of the shapes into account, while neglecting their label and structure.

Mesh information and labels are accounted for in the current metrics, which require the computation of the normals and tangents of the meshes. Normals and tangents depend on the vertex positions $X(1)$, and we denote the gradient of the currents metric with respect to such positions by $\nabla_{X(1)}A$. This gives the direction for moving the final point positions $X(1)$ so that the distribution of normals and tangents in the $\phi(O_i)$'s becomes more similar to that of $O_i{}'$. The task now is to find an update for the initial conditions S_0, so that at time $t = 1$, the final positions $X(1)$ move in the desired direction. To this end, $E(S_0) = A(X(1)) + L(S_0)$ can be used as a matching criterion, where $L(S_0) = ||v_0||_V^2$ is the kinetic energy of the deformation (and also the length of the geodesic path connecting $\phi(1, \cdot)$ to the identity map), which is used as a regularizer. It is not tractable to use the chain rule to compute the gradient of E with respect to S_0 from the gradient of A with respect to $X(1)$. Instead, as proven in [103], the gradient $\nabla_{X(1)}A$, at time $t = 1$, can be transported back to time $t = 0$ using linear versions of the ODEs (44) and (45) to compute $\nabla_{S_0}E$ at time $t = 0$:

$$\nabla_{S_0}E = \xi(0) + \nabla_{S_0}L, \tag{46}$$

where $\xi(0)$ is computed by integration of the following ODEs (integrated backward from time $t = 1$ to $t = 0$) :

$$\dot{\theta}(t) = -(\partial_{X(t)}G(X(t), S(t)))^T\theta(t), \quad \theta(1) = \nabla_{X(1)}A, \tag{47}$$

$$\dot{\xi}(t) = -(\partial_{S(t)}G(X(t), S(t))^T\theta(t) - d_{S(t)}F^T\xi(t), \quad \xi(1) = 0. \tag{48}$$

The gradient $\nabla_{X(1)}A$ is a set vectors attached to the vertices of the deformed shape $\phi(O_i)$ at time $t = 1$, which is transported back to time $t = 0$ through a linear ODE (47). Integrated backward from time $t = 1$ to time $t = 0$, the value $\xi(0)$ interpolates this information at control point positions. It is the gradient of the data term with respect to the initial conditions. For optimizing the matching criterion, gradient descent with adaptive step size can be used. To initialize the optimization, one can set the control points at the nodes of a regular lattice whose step equals the deformation parameter σ_V and momentum vectors are set to zero, i.e. no deformation. The output of the algorithm is the optimal position of the control points and the optimal values of the momentum vectors. These values enable computation of the deformed shapes $\phi(O_i)$. Parameters that need to be set are the deformation parameter σ_V, which controls for the "rigidity" of the deformation, the currents parameters σ_W, which controls for the sensitivity of the currents metric, and the weights w_i that balance the contributions of the different structures against the regularity term and against themselves. Note that control points are not specific to particular structures. Instead, they affect a neighborhood in 3D space and drive the deformation of every part of any structures in this neighborhood. They give a unique descriptor of anatomical differences, which integrates various kinds of data such as cortical areas, fiber bundles, and sub-cortical structures. Examples illustrating structure matching by deformetrics are shown in Figs. 16 and 17. Details can be found in [106, 107].

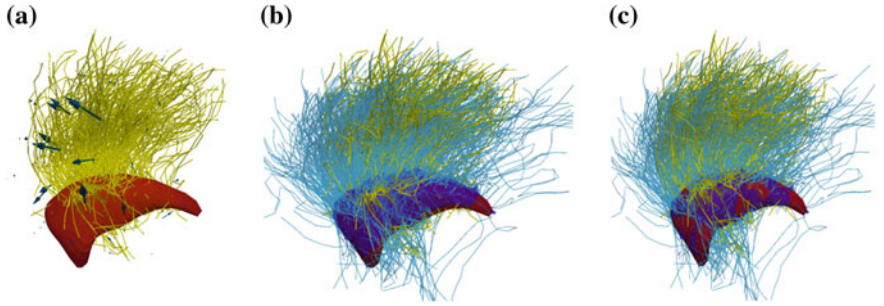


Fig. 16 Matching of a structure complex made of the surface of the caudate nucleus and the fiber bundle connecting the nucleus to the cortex. Deformation parameters are control points and momenta attached to them (*arrows* in **a**). They define a unique 3D space deformation that warps fibers and surfaces altogether. They are estimated so that they minimize, up to a regularity constraint, the sum of two currents norms, one between fiber bundles and the other between surfaces. $\sigma_V = 10\,\text{mm}$, $\sigma_W = 3\text{mm}$, $w = 0.2\text{mm}^{-2}$ (*surface*) and $\sigma_W = 2\text{mm}$, $w = 0.02$ (*bundles*). **a** Source. **b** Source and target. **c** Deformed source

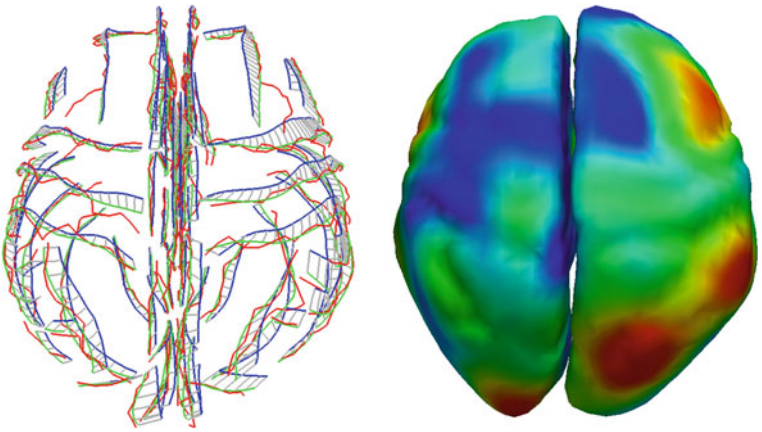


Fig. 17 Matching between 72 pairs of homologous sulcal curves. *Left* A single 3D space deformation is estimated, so that source curves (in blue) best match their homologous curves (in *red*). Deformed curves are in green. Right: Deformation is applied to the underlying cortical surface. Red corresponds to the maximum displacement and blue corresponds to zero displacement. $\sigma_V = 25\text{mm}$, $\sigma_W = 5\text{mm}$ and $w = 100$ for all curves

6.4 Statistical Shape Models Based on Deformations

Deformations between pairs of anatomical configurations can be extended for group-wise analysis. Given homologous anatomical configurations in N individuals, one can build an atlas, namely a template, and N deformations matching this template to each individual's anatomy for studying subject variability. The template should have the same number of structures as in each individual. Also, the topology of each template

structure, namely the number of points and the connectivity between them, needs to be set a priori. For closed surfaces, an ellipsoidal surface with a pre-determined number of vertices is set. For fiber bundles, one randomly picks fibers from every individual to form an initial fiber bundle. With the vertices concatenated into a vector X_0, atlas construction can be viewed as finding the optimal position of the vertices, so that the template averages out the shape features.

One way to derive a template is to use Fréchet mean, which defines the template as the anatomical configuration that minimizes the sum of squared differences between template and each individual. As a measure of distance, a good candidate is the length of the geodesic path of the diffeomorphisms that matches the template to each anatomical configuration. This amounts to minimizing $\Sigma_i ||v_0^i||^2$, subject to the diffeomorphism ϕ^i parameterized by v_0^i such that $\phi^i(X_0)$ equals the ith individual anatomical configuration. Diffeomorphisms that perfectly match two anatomical configurations do not necessarily exist and are often not desirable due to the presence of noise as well as segmentation and sampling artifacts in the data. From a statistical point of view, exact matching would lead to a statistical model that does not generalize well. Thus, a relaxed version of the Fréchet criterion is typically preferred:

$$E = \sum_i \sum_k w_k ||\phi^i(\overline{O}_k) - O_k^i||_W^2 + ||v_0^i||_V^2, \tag{49}$$

where $\bar{O}_k$ is the kth component of the template anatomical configuration (the concatenation of the vertices of all $\bar{O}_k$ being X_0), O_k^i is the homologous component in the ith individual, and ϕ^i is the ith template-to-individual deformation, which is parameterized by the velocity field v_0^i. The velocity fields are assumed to be parameterized by a set of control points $c_{0,p}$ and momentum vectors $\alpha_{0,p}^i$ associated with them. These control points and momentum vectors are defined in the template space. Control points are assumed to be the same for every subject, so that all deformations are decomposed into the same basis. Momentum vectors are specific to each subject and parameterize a different deformation of the template to each individual anatomy. Introducing $S_0^i = \{c_{0,p}, \alpha_{0,p}^i\}_p$ as the initial parameters for each deformation that become $S_0^i(t)$ at a later time t and $X^i(t)$ as the position of the template points X_0 at time t of the ith template-to-individual deformation, the criterion in (49) can be rewritten as:

$$E(S_0^i, X_0) = \sum_i A_i(X^i(1)) + L(S_0^i), \tag{50}$$

$$\dot{X}^i(t) = G(X^i(t), S^i(t)), \quad X^i(0) = X_0, \tag{51}$$

$$\dot{S}^i(t) = F(S^i(t)), \quad S^i(0) = S_0^i. \tag{52}$$

The minimization of the deformation parameters S_0^i is essentially the minimization of N independent matching criteria, with the exception that the initial control point positions are shared among individuals. The minimization over the template point

positions X_0 seems more challenging, but in actual, the estimation of the gradient does not require extra computations. The gradient of the criterion is given by:

$$\nabla_{c_p} E = \sum_i \xi^i_{c_p}(0) + \nabla_{c_p} L \ , \quad \nabla_{\alpha^i_p} E = \xi^i_{\alpha_p}(0) + \nabla_{\alpha^i_p} L \ , \tag{53}$$

$$\nabla_{X_0} E = \sum_i \theta^i(0), \tag{54}$$

where $\xi(0) = \{\xi^i_c(0), \xi^i_\alpha(0)\}$ and $\theta^i(0)$ are computed by integration of the following two ODEs (integrated backward from time $t = 1$ to $t = 0$):

$$\dot{\theta^i}(t) = -(\partial_{X^i(t)} G(X^i(t), S^i(t)))^T \theta^i(t), \quad \theta^i(1) = \nabla_{X^i(1)} A_i, \tag{55}$$

$$\dot{\xi^i}(t) = -(\partial_{S^i(t)} G(X^i(t), S^i(t))^T \theta^i(t) - d_{S^i(t)} F^T \xi^i(t), \quad \xi^i(1) = 0. \tag{56}$$

Gradient with respect to the deformation parameters can be computed in an analogous manner as in the case of matching two anatomical configurations. The gradient with respect to template point is given "for free" in the θ auxiliary variable at time $t = 0$, which is also averaged over individuals. The output of the above procedure is a template anatomical configuration encoded by the vertices X_0, a set of initial control points optimally placed in the regions of largest variability, and momentum vectors parameterizing template-to-subject deformations. Control points provide a common basis for the parameterization of template-to-individuals matching, and the momentum vectors indicate the decomposition of the deformations on this basis. See example in Fig. 18 and [103] for details. We highlight that the momentum vectors generated by the above algorithm are the tangent-space representations of the deformations. Hence, no additional computation is required to perform statistics, since the descriptors are finite-dimensional and lives in the usual Euclidean space provided by the inner-product $K = \{K(c_{0,i}, c_{0,k})\}_{i,j}$. In particular, any statistical tools, such as classification, clustering, and regression, can be directly applied. In Fig. 19, a Fisher's discriminant analysis is performed in which the direction orthogonal to the separating hyperplane is a momentum vector. This momentum vector parameterizes a deformation that can be integrated over time using (45). The deformation of the template, given by integration of (44), provides an interpretable representation of the shape features that have been detected as most discriminative.

6.5 Applications and Insights

The technique of currents and diffeomorphisms has been used so far to characterize the variability in cortical configurations [106, 108], to highlight differences in connectivity patterns between patients with Gilles de la Tourette syndrome and controls

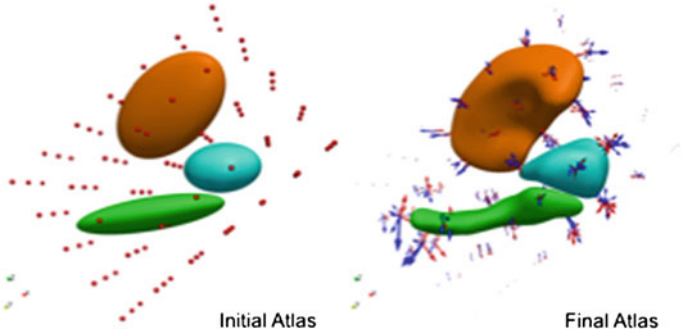


Fig. 18 Atlas construction from 16 anatomical configurations made of 3 deep brain nuclei: hippocampus (*green*), amygdala (*blue*) and putamen (*orange*) of 8 controls and 8 subjects with Down syndrome. *Left* Initial configuration given. Ellipsoidal surfaces are template shapes. *Red dots* are the initial position of control points. Momenta are set to zero. *Right* Output of deformetrics approach. Vertices of the template shapes have been updated to capture the shape invariants across the 16 individuals. Control points have moved towards the most variable parts of the shape complex. Momenta parameterize the deformation of the template to each control (*blue arrow*) and each Down syndrome subjects (*red arrows*). $\sigma_V = 10$mm, $\sigma_W = 5$mm (using the varifold representation of currents [97]), $w = 0.005\text{mm}^{-2}$ for the three structures

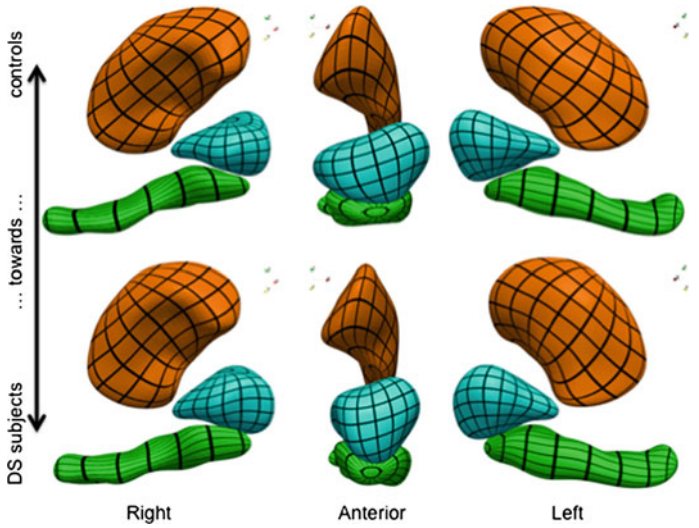


Fig. 19 Linear Discriminative Analysis (LDA) in shape space. LDA is performed on the momentum vectors from the atlas in Fig. 18. The most discriminative axis is a momenta set that parameterizes a deformation towards controls, and a deformation towards Down syndrome subjects (opposite momenta). The deformation of the template along these two directions is shown, exhibiting the most discriminative shape features captured by the atlas

[107], to measure cortical thinning of the cingluate gyrus in schizophrenia [109], to detect atypical growth patterns in longitudinal shape studies [110], and to predict cardiac remodeling in patients with Tetralogy of Fallot [111] for instance. The main assets of deformetrics are that it does not require establishing a one-to-one point correspondence between segmented structures. Also, it can handle a large variety of shapes. Moreover, it enables the analysis of anatomical shape complexes reflecting the variability in anatomical configuration in populations. The method requires only a minimal amount of preprocessing and user intervention, and opens up the possibility to use such morphometric tools routinely for large cohort studies.

Deformetrics permits analysis of complex anatomical configurations and provides a single descriptor for quantifying anatomical differences among individuals. The descriptor integrates how shape components within the structure complex co-vary, while preserving the complex organization (no collision between components). Deformetrics can therefore be used for full brain morphometry that combines cortical and sub-cortical structures with white matter fiber bundles. The control point aspect of deformetrics restrains the dimensionality of the descriptor, and hence reduces the effects of multiple comparisons. This property is particularly relevant for settings with high dimension and low sample size, which is typical for neuroimaging studies. Also, results could be displayed as deformation, which eases the result interpretation.

The main limitation of deformetrics is that it relies on segmented structures. Although the intrinsic noise parameter in the currents metric makes deformetrics robust to mesh imperfection and random local shift of boundaries, deformetrics would be affected by consistent segmentation errors occurring at the same place, or by systematic over/under estimation of structure boundaries. Further, deformetrics will not work properly if the hypothesis of homology is violated. This could arise from problems in structure identification. Presence or absence of particular structures as well as important changes in topology across individuals may be also problematic. Lastly, more attention should be paid on the numerical aspects of deformetrics. Specifically, convergence problems may arise from the simultaneous optimization of several variables with different orders of magnitude. Cautions should also be paid on the effects of parameters, such as the deformation scale, currents scale, and weights of shape components in the criterion. Multi-scale strategy and automatic estimation of such parameters in a Bayesian framework would be promising directions to explore [107, 112].

7 Future Outlooks

Advances in MRI technology have enabled acquisition down to submillimeter resolution [113]. With more high resolution MRI data becoming available, we foresee much greater clinical impact from deployment of shape analysis. However, associated with the increase in resolution and sample sizes are some caveats. For instance, approaches for brain structure segmentation need to be reconsidered. Manual segmentation is very time consuming and relies solely on expert knowledge.

Besides the problem of intra-expert and inter-expert variability [114], the increasing number of samples, coupled with the increasing number of slices, renders manual segmentation impractical. A technological push on efficient automated means is thus crucial.

A popular automated approach is to register a whole-brain atlas to new brain volumes and use the brain region labels associated with the atlas to delineate structures of interest [115, 116]. The segmentation accuracy of this approach is contingent on the registration accuracy. Another widely-used automated approach is to segment based on features, such as intensity and gradients, and constrain the segmentation with shape priors [79]. There is thus a strong interplay between shape analysis and segmentation. However, building generalizable shape priors requires a large number of segmented brain structures. Hence, there is a slight circularity problem. A potentially promising direction is to combine automated methods with semi-automated corrections [117]. This strategy would enable exploitation of expert knowledge without completely relying on it, while substantially reducing the time cost compared to pure manual segmentation.

In addition to the impact on segmentation, which is a fundamental step prior to most shape analysis approaches, higher resolution and increasing sample sizes have major implications on each of the shape analysis approaches described. For point-based representation with landmarks, the time cost would substantially increase. For dense point-based representation, errors introduced by the one-to-one point correspondence assumption as well as the problem of multiple comparisons would worsen with higher resolution. A potential workaround is to assume a one-to-one patch correspondence without enforcing the corresponding patches to have the same number of voxels/points. A useful definition of a patch is unclear, but extending ideas from the computer vision literature might be a good starting point. For mesh-based approaches, intrinsic geometry will continue to play a critical role, especially due to its robustness to natural variations in large samples. To overcome challenges arising from increased resolution and sample sizes, we can apply these methods with a multi-scale strategy to strike a balance between computational feasibility and anatomical details. For function-based representation, boundary point correspondence is not an issue, since correspondence is implicitly established at the function coefficient level. The critical factor to correspondence with function-based representation is subject alignment, which likely improves with higher resolution. Also, increasing the number of voxels provides more robust estimates of the function coefficients. Thus, function-based representation can greatly benefit from higher resolution. For medial representation, pruning to establish topology correspondence would become impractical with increasing number of samples. Also, whether a single medial axis can truly represent all samples within a population is in question due to the increased variability with greater sample sizes. Deformetrics would adapt well to increasing sample sizes, since it requires little preprocessing and is robust to small segmentation errors and mesh imperfections. However, it has high computational cost, which may be a limitation in its application to larger datasets without high performance computing resources.

The increasing resolution and sample sizes is definitely beneficial, but some work is required to extend current shape analysis techniques for data in this regime. We end this chapter with a few questions for thoughts. First, increased resolution and sample sizes would greatly complicate creation of a one-to-one point correspondence, but localization is critical for clinical applications, e.g. applying medication to specific disease-affected areas within a brain structure. Thus, how should a balance be drawn? Second, essential to clinical translation is validation. Basing validation solely on prior neuroscience knowledge is likely insufficient. Thus, what would be a good quantitative way to validate new methods on real data? Lastly, can information from other modalities, e.g. functional and diffusion MRI, be used to improve shape analysis? Investigating the relationships between brain structure and function and integrating this information into shape analysis might be a promising direction to explore.

References

1. Dryden IL, Mardia KV (1998) Statistical Shape Analysis. Wiley, Chichester
2. Besl PJ, McKay ND (1992) A method for Registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intel 14(2):239–256
3. Bookstein FL (1997) Morphometric tools for landmark Data: geometry and biology. Cambridge University Press, Cambridge
4. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models-their training and application. Comput Vis Image Underst 61(1):38–59
5. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intel 23(6):681–685
6. Rohr K (1997) On 3D differential operators for detecting point landmarks. Image Vis Comput 15(3):219–233
7. Ono M, Kubik S, Abernathy CD (1990) Atlas of the cerebral sulci. Thieme Medical, New York
8. Duta N, Sonka M (1998) Segmentation and interpretation of MR Brain images. An improved active shape model. Imaging IEEE Trans Med 17(6):1049–1062
9. Tao X, Prince JL, Davatzikos C (2002) Using a statistical shape model to extract sulcal curves on the outer cortex of the human brain. IEEE Trans Med Imaging 21(5):513–524
10. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3):1116–1128
11. Joshi S, Miller MI (2000) Landmark matching via large deformation diffeomorphisms. IEEE Trans Image Process 9(8):1357–1370
12. Davies RH, Twining CJ, Cootes TF, Waterton JC, Taylor CJ (2002) A minimum description length approach to statistical shape modeling. IEEE Trans Med Imaging 21(5):525–537
13. Brummer ME (1991) Hough transform detection of the longitudinal fissure in tomographic head images. IEEE Trans Med Imaging 10(1):74–81
14. Harris C, Stephens M (1988) A combined corner and edge detector. In: Alvey vision conference, pp 147–151
15. Witkin A (1984) Scale-space filtering: a new approach to multi-scale description. IEEE Int Conf Acoust Speech Signal Process 9:150–153
16. Lindeberg T (1998) Feature detection with automatic scale selection. Int J Comput Vis 30(2):79–116

17. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
18. Toews M, Wells WM III, Collins DL, Arbel T (2010) Feature-based morphometry: discovering group-related anatomical patterns. NeuroImage 49(3):2318–2327
19. Burt P, Adelson E (1983) The Laplacian Pyramid as a compact image code. IEEE Trans Commun 31(4):532–540
20. Toews M, Wells WM III (2013) Efficient and robust model-to-image alignment using 3D scale-invariant features. Med Image Anal 17(3):271–282
21. Toews M, Wells III WM (2009) SIFT-rank: ordinal description for invariant feature correspondence. In: IEEE conference on computer vision and pattern recognition, pp 172–177
22. Toews M, Arbel T (2007) A statistical parts-based model of anatomical variability. IEEE Trans Med Imaging 26(4):497–508
23. Toews M, Zöllei L, Wells WM III (2013) Invariant feature-based alignment of volumetric multi-modal images. In: Wells WM, Joshi S, Pohl KM (eds) IPMI 2013, LNCS, vol 7917. Springer, Berlin, pp 25–36
24. Gupta A, Toews M, Janardhana R, Rathi Y, Gilmore J, Escolar M, Styner M (2013) Fiber feature map based landmark initialization for highly deformable DTI registration. In: SPIE medical imaging, pp 866907–866907. International Society for Optics and Photonics
25. Toews M, Wells WM III, Zöllei L (2012) A feature-based developmental model of the infant brain in structural MRI. In: Ayache N, Delingette H, Golland P, Mori K (eds) MICCAI 2012, LNCS, vol 15. Springer, Berlin, pp 204–211
26. Uhlenbeck K (1976) Generic properties of eigenfunctions. Am J Math 98(4):1059–1078
27. Qiu A, Bitouk D, Miller MI (2006) Smooth functional and structural maps on the neocortex via orthonormal bases of the Laplace-Beltrami operator. IEEE Trans Med Imaging 25(10):1296–1306
28. Shi Y, Lai R, Morra J, Dinov I, Thompson P, Toga A (2010) Robust surface reconstruction via Laplace-Beltrami eigen-projection and boundary deformation. IEEE Trans Med Imaging 29(12):2009–2022
29. Lai R, Shi Y, Dinov I, Chan TF, Toga AW (2009) Laplace-Beltrami nodal counts: a new signature for 3D shape analysis. In: International symposium on biomedical imaging, pp 694–697
30. Reuter M, Wolter F, Peinecke N (2006) Laplace-Beltrami spectra as shape-DNA of surfaces and solids. Comput Aided Des 38:342–366
31. Reeb G (1946) Sur les Points Singuliers d'une Forme de Pfaff Completement Integrable ou d'une Fonction Nemérique. Comptes Rendus Acad Sci 222:847–849
32. Jost J (2001) Riemannian geometry and geometric analysis, 3rd edn. Springer, New York
33. Shinagawa Y, Kunii TL (1991) Constructing a reeb graph automatically from cross sections. IEEE Comput Graph Appl 11(6):44–51
34. Takahashi S, Ikeda T, Shinagawa Y, Kunii TL, Ueda M (1995) Algorithms for extracting correct critical points and constructing topological graphs from discrete geographical elevation data. Comput Graph Forum 14(3):181–192
35. Biasotti S, Falcidieno B, Spagnuolo M (2000) Extended reeb graphs for surface understanding and description. In: International conference on discrete geometry for computer imagery, pp 185–197
36. Lazarus F, Verroust A (1999) Level set diagrams of polyhedral objects. In: ACM symposium on solid modeling and applications, pp 130–140
37. Hilaga M, Shinagawa Y, Kohmura T, Kunii TL (2001) Topology matching for fully automatic similarity estimation of 3D shapes. In: SIGGRAPH, pp 203–212
38. Mortara M, Patané G (2002) Affine-invariant skeleton of 3D shapes. In: Shape modeling Internationl, pp 245–252
39. Tierny J, Vandeborre JP, Daoudi M (2006) Invariant high level reeb graphs of 3D polygonal meshes. In: International symposium on 3D data processing, visualization, and transmission, pp 105–112

40. Rustamov RM (2007) Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In: Eurographics symposium on geometry processing, pp 225–233
41. Ovsjanikov M, Sun J, Guibas LJ (2008) Global intrinsic symmetries of shapes. Eurograph Symp Geom Process 27:1341–1348
42. Shi Y, Lai R, Toga AW (2013) Cortical surface reconstruction via unified reeb analysis of geometric and topological outliers in magnetic resonance images. IEEE Trans Med Imaging 32(3):511–530
43. Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis I: Segmentation and surface reconstruction. NeuroImage 9:179–194
44. Lai R, Shi Y, Scheibel K, Fears S, Woods R, Toga A, Chan T (2010) Metric induced optimal embedding for intrinsic 3D shape analysis. In: International conference on computer vision pattern recognition, pp 2871–2878
45. Shi Y, Lai R, Toga AW (2013) Conformal mapping via metric optimization with application for cortical label fusion. IPMI (in press)
46. Gu X, Wang Y, Chan TF, Thompson PM, Yau ST (2004) Genus zero surface conformal mapping and its application to brain surface mapping. IEEE Trans Med Imaging 23(8):949–958
47. Gold SM, O'Connor, MF, Gill R, Kern KC, Shi Y, Henry RG, Pelletier D, Mohr DC, Sicotte NL (2012) Detection of Altered hippocampal morphology in multiple sclerosis associated depression using automated surface mesh modeling. Hum Brain Mapp. doi:10.1002/hbm.22154. [Epub ahead of print]
48. Hu MK (1962) Visual pattern recognition by moment invariants. IRE Trans Inf Theory 8(2):179–187
49. Lo CH, Don HS (1989) 3-D moment forms: their construction and application to object identification and positioning. IEEE Trans Pattern Anal Mach Intel 11(10):1053–1064
50. Fischl B, Liu A, Dale AM (2001) Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. IEEE Trans Med Imaging 20(1):70–80
51. Székely G, Kelemen A, Brechbühler C, Gerig G (1996) Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models. Med Image Anal 1(1):19–34
52. Healy DM, Rockmore DN, Kostelec PJ, Moore S (2003) FFTs for the 2-sphere-improvements and variations. J Fourier Anal Appl 9(4):341–385
53. Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intel 11(7):674–693
54. Lindeberg T (1993) Scale-space theory in computer vision. Kluwer Academic, Hingham
55. Yu P, Grant PE, Qi Y, Han X, Ségonne F, Pienaar R, Busa E, Pacheco J, Makris N, Buckner RL, Golland P, Fischl B (2007) Cortical surface shape analysis based on spherical wavelets. IEEE Trans Med Imaging 26(4):582–597
56. Bernal-Rusiel JL, Atienza M, Cantero JL (2008) Detection of focal changes in human cortical thickness: spherical wavelets versus gaussian smoothing. NeuroImage 41(4):1278–1292
57. Kim WH, Pachauri D, Hatt C, Chung MK, Johnson S, Singh V (2012) Wavelet based multiscale shape features on arbitrary surfaces for cortical thickness discrimination. In: Bartlett P (ed) NIPS 2012. LNCS, vol 25, pp 1250–1258
58. Sadjadi FA, Hall EL (1980) Three-dimensional moment invariants. IEEE Trans Pattern Anal Mach Intel 2:127–136
59. Mangin JF, Poupon F, Duchesnay E, Riviére D, Cachia A, Collins DL, Evans AC, Régis J (2004) Brain morphometry using 3D moment invariants. Med Image Anal 8(3):187–196
60. Ng B, Abugharbieh R, Huang X, McKeown MJ (2009) Spatial characterization of fMRI activation maps using invariant 3-D moment descriptors. IEEE Trans Med Imaging 28(2):261–268
61. Yang F, Kruggel F (2009) A graph matching approach for labeling brain sulci using location, orientation, and shape. Neurocomputing 73(1):179–190

62. Zacharaki EI, Hogea CS, Shen D, Biros G, Davatzikos C (2009) Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. NeuroImage 46(3):762–774
63. Millán RD, Dempere-Marco L, Pozo JM, Cebral JR, Frangi AF (2007) Morphological characterization of intracranial aneurysms using 3-D moment invariants. IEEE Trans Med Imaging 26(9):1270–1282
64. Prastawa M, Bullitt E, Ho S, Gerig G (2004) A brain tumor segmentation framework based on outlier detection. Med Image Anal 8(3):275–283
65. Warfield SK, Kaus M, Jolesz FA, Kikinis R (2000) Adaptive, template moderated, spatially varying statistical classification. Med Image Anal 4(1):43–55
66. Gerig G, Styner M, Jones D, Weinberger D, Lieberman J (2001) Shape analysis of brain ventricles using SPHARM. In: IEEE workshop mathematical methods in biomedical image, analysis, pp 171–178
67. Levitt JJ, Styner M, Niethammer M, Bouix S, Koo MS, Voglmaier MM, Dickey CC, Niznikiewicz MA, Kikinis R, McCarley RW, Shenton ME (2009) Shape abnormalities of caudate nucleus in schizotypal personality disorder. Schizophr Res 110(1):127–139
68. Zhao Z, Taylor WD, Styner M, Steffens DC, Krishnan KR, MacFall JR (2008) Hippocampus shape analysis and late-life depression. PLoS One 3(3):e1837
69. Van De Ville D, Seghier ML, Lazeyras FO, Blu T, Unser M (2007) WSPM: wavelet-based statistical parametric mapping. NeuroImage 37(4):1205–1217
70. Canales-Rodríguez EJ, Radua J, Pomarol-Clotet E, Sarró S, Alemán-Gómez Y, Iturria-Medina Y, Salvador R (2013) Statistical analysis of brain tissue images in the wavelet domain: wavelet-based morphometry. NeuroImage 72(22):214–226
71. Hackmack K, Paul F, Weygandt M, Allefeld C, Haynes JD (2012) Multi-scale classification of disease using structural MRI and wavelet transform. NeuroImage 62(1):48–58
72. Nain D, Haker S, Bobick A, Tannenbaum A (2007) Multiscale 3-D shape representation and segmentation using spherical wavelets. IEEE Trans Med Imaging 26(4):598–618
73. Hamarneh G, Abugharbieh R, McInerney T (2004) Medial profiles for modeling deformation and statistical analysis of shape and their use in medical image segmentation. Int J Shape Model 10:187–210.
74. Ward A, Hamarneh G (2008) GMAT: The groupwise medial axis transform for fuzzy skeletonization and intelligent pruning. Technical report, School of Computing Science, Simon Fraser University, Burnaby
75. Blum H (1967) A transformation for extracting new descriptors of shape. Models for the perception of speech and visual form. MIT Press, Cambridge, pp 362–380
76. Bai X, Latecki LJ, Liu WY (2007) Skeleton pruning by contour partitioning with discrete curve evolution. IEEE Trans Pattern Anal Mach Intel 29(3):449–462
77. Ward A, Hamarneh G (2010) GMAT: the groupwise medial axis transform for fuzzy skeletonization and intelligent pruning. IEEE Trans Pattern Anal Mach Intel 32(6):1084–1096
78. Pizer SM, Fletcher PT, Joshi S, Thall A, Chen JZ, Fridman Y, Fritsch DS, Gash AG, Glotzer JM, Jiroutek MR, Lu C, Muller KE, Tracton G, Yushkevich P, Chaney EL (2003) Deformable M-reps for 3D medical image segmentation. Int J Comput Vis 55(2):85–106
79. Siddiqi K, Pizer SM (2008) Medial representations: mathematics, algorithms and applications. Springer, New York
80. Fletcher T, Lu C, Pizer SM, Joshi S (2004) Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Trans Med Imaging 23(8):995–1005
81. Yushkevich PA, Zhang H, Gee JC (2006) Continuous medial representation for anatomical structures. IEEE Trans Med Imaging 25(12):1547–1564
82. Yushkevich PA (2009) Continuous medial representation of brain structures using the Biharmonic PDE. NeuroImage 45(1):s99–s110
83. Yushkevich PA (2003) Statistical shape characterization using the medial representation. Ph.D. thesis, University of North Carolina, Chapel Hill
84. Matheron G (1988) Examples of topological properties of skeletons. In: Serra J (ed) Image analysis and mathematical morphology part II: theoretical advances. Academic Press, London, pp 217–238

85. Naf M, Kubler O, Kikinis R, Shenton M, Szekely G (1996) Characterization and recognition of 3D organ shape in medical image analysis using skeletonization. In: Workshop on mathematical methods in biomedical image analysis, pp 139–150. IEEE Computer Society
86. Styner M (2001) Combined boundary-medial shape description of variable biological objects. Ph.D. thesis, University of North Carolina, Chapel Hill
87. Katz R (2002) Form metrics for interactive rendering via figural models of perception. Ph.D. thesis, University of North Carolina, Chapel Hill
88. Siddiqi K, Ahokoufandeh A, Dickinson S, Zucker S (1998) Shock graphs and shape matching. Int Conf Comput Vis 35:13–32
89. Siddiqi K, Bouix S, Tannenbaum A, Zucker SW (1999) The Hamilton-Jacobi skeleton. Comput Vis 2:828–834. IEEE Press
90. Fridman Y, Pizer SM, Aylward S, Bullitt E (2003) Segmenting 3D branching tubular structures using cores. In: Ellis RE, Peters TM (eds) MICCAI 2003. LNCS, vol 2879, pp 570–577. Springer, New York
91. Golland P, Grimson W, Kikinis R (1999) Statistical shape analysis using fixed topology skeletons: corpus callosum study. In: Kuba A, Attila J, Samal M (eds) LNCS, vol 1613, pp 382–388. Springer, New York
92. Fletcher T (2004) Statistical variability in nonlinear spaces: application to shape analysis and DT-MRI. Ph.D. thesis, University of North Carolina, Chapel Hill
93. Styner M, Lieberman JA, Pantazis D, Gerig G (2004) Boundary and medial shape analysis of the hippocampus in schizophrenia. Med Image Anal 8(3):197–203
94. McClure RK, Styner M, Maltbie E, Lieberman JA, Gouttard S, Gerig G, Shi X, Zhu H (2013) Localized differences in caudate and hippocampal shape are associated with schizophrenia but not antipsychotic type. Psychiatry Res 211(1):1–10
95. Ishaq O, Hamarneh G, Tam R, Traboulsee A (2007) Longitudinal, regional and deformation-specific corpus callosum shape analysis for multiple sclerosis. In: IEEE international conference of engineering in medicine and biology society, pp 2110–2113
96. Vaillant M, Glaunes J (2005) Surface matching via currents. In: Christensen GE, Sonka M (eds) LNCS, vol 3565, pp 381–392. Springer, New York
97. Charon N, Trouvé A The varifold representation of non-oriented shapes for diffeomorphic registration. SIAM J Imaging Sci (to appear), eprint arXiv:1304.6108
98. Durrleman S (2010) Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Ph.D. thesis, Nice Sophia-Antipolis University, France
99. Christensen GE, Rabbitt RD, Miller MI (1994) 3D brain mapping using a deformable neuroanatomy. Phys Med Biol 39(3):609–618
100. Trouvé A (1998) Diffeomorphisms groups and pattern matching in image analysis. Int J Comput Vis 28(3):213–221
101. Dupuis P, Grenander U, Miller MI (1998) Variational problems on flows of diffeomorphisms for image matching. Q Appl Math 56(3):587–600
102. Vercauteren T, Pennec X, Perchant A, Ayache N (2009) Diffeomorphic demons: efficient non-parametric image registration. NeuroImage 45(1):S61–S72
103. Durrleman S, Prastawa M, Korenberg JR, Joshi S, Trouvé A, Gerig G (2012) Topology preserving atlas construction from shape data without correspondence using sparse parameters. In: Ayache N, Delingette H, Golland P, Mori K (eds) MICCAI 2012. LNCS, vol 7512, pp 223–230
104. Miller MI, Trouvé A, Younes L (2002) On the metrics and euler-lagrange equations of computational anatomy. Ann Rev Biomed Eng 4:375–405
105. Miller MI, Trouvé A, Younes L (2006) Geodesic shooting for computational anatomy. J Math Imaging Vis 24(2):209–228
106. Durrleman S, Pennec X, Trouvé A, Thompson P, Ayache N (2008) Inferring brain variability from diffeomorphic deformations of currents: an integrative approach. Med Image Anal 12(5):626–637
107. Medical Image Computing and Computer-Assisted Intervention—MICCAI (2013) In: Mori K, Ichiro S (eds) LNCS, vol 8149, pp 267–274

108. Auzias G, Colliot O, Glaunès JA, Perrot M, Mangin J-F, Trouvé A, Baillet S (2011) Diffeomorphic brain registration under exhaustive sulcal constraints. IEEE Trans Med Imaging 30(6):1214–1227
109. Qiu A, Younes L, Wang L, Ratnanather JT, Gillepsie SK, Kaplan G, Csernansky J, Miller MI (2007) Combining anatomical manifold information via diffeomorphic metric mappings for studying cortical thinning of the cingulated gyrus in schizophrenia. Neuroimage 37(3):821–833
110. Durrleman S, Pennec X, Trouvé A, Braga J, Gerig G, Ayache A (2013) Toward a comprehensive framework for the spatiotemporal analysis of longitudinal shape data. Int J Comput Vis 103(1):22–59
111. Mansi T, Voigt I, Leonardi B, Pennec X, Durrleman S, Sermesant M, Delingette H, Taylor AM, Boudjemline Y, Pongiglione G, Ayache N (2011) A statistical model for quantification and prediction of cardiac remodelling: application to tetralogy of fallot. IEEE Trans Med Imaging 9(30):1605–1616
112. Allassonnière S, Kuhn E (2010) Stochastic algorithm for parameter estimation for dense deformable template mixture model. ESAIM-PS 14:382–408
113. van der Kolk AG, Hendrikse J, Zwanenburg JJ, Visser F, Luijten PR (2013) Clinical applications of 7 T MRI in the brain. Eur. J. Radiol. 82(5):708–718
114. Warfield SK, Zou KH, Wells WM (2004) Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23(7):903–921
115. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33(3):341–355
116. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31(3):968–980
117. Top A, Hamarneh G, Abugharbieh R (2011) Active learning for interactive 3D image segmentation. In: Fichtinger G, Peters T (eds) MICCAI 2011. LNCS, vol 14, pp 603–610

Shape Analysis in Molecular Imaging

Fei Gao and Pengcheng Shi

Abstract Molecular imaging is a new research discipline enabling the visualization, characterization and quantification of biologic processes taking place at the cellular and subcellular levels within intact living subjects. Applications of molecular imaging techniques will benefit various clinical practices including classification and tracking of chemotherapy and treatment planning of radiotherapy, as well as drug discovery and development. Molecular imaging typically includes two or three dimensional imaging with quantification over time, and is often applied on molecular imaging modalities, such as Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT) etc. Image series acquired with spatiotemporal distribution of molecular biomarkers must be carefully analyzed to estimate the underlying physiology-related metabolic parameters. Shape analysis is one of the most powerful tools to analyze the geometrical properties from similar shapes or different groups, and can be applied to estimate both the concentration of biomarkers and interaction between biomarkers and tissue/organs. However, some limitations from molecular imaging modalities and clinical practices still hinder the quantitative accuracy of shape analysis, e.g. the low spatial and temporal resolution in PET scan, the inaccuracy of blood samplings from patients, the low Signal-to-Noise (SNR) ratio of measurement data in dynamic PET/CT scan. In this chapter, firstly, we will introduce the definition of molecular imaging, the clinical advantages and limitations of various molecular imaging modalities, secondly, we will review the challenges in data analysis based on the data processing procedure, and explain how data corrections affect the accuracy of static and dynamic PET imaging, thirdly, the general frameworks of image processing in PET and SPECT are reviewed with focus on image reconstruction, at last, we will show some recent advancements and give examples of clinical applications.

F. Gao (✉) and P. Shi
Rochester Institute of Technology, 102 Lomb Memorial Drive, Rochester, NY 14623-5608, USA
e-mail: gaofei.rit@gmail.com

P. Shi
e-mail: pengcheng.shi@rit.edu

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_2,

1 Introduction to Molecular Imaging and Molecular Imaging Modalities

Molecular imaging provides the images of molecular and cellular level activities inside the body. Molecular imaging enables doctors to measure the biological processes quantitatively and reflects the functionality of organs and tissues inside patients. According to the definition from the Society of Nuclear Medicine and Molecular Imaging (SNMMI), molecular imaging is the visualization, characterization, and measurement of biological processes at the molecular and cellular levels in humans and other living systems [58]. Molecular imaging usually involves two- or three- dimensional images and conducts quantitative analysis over the time.

Molecular imaging is a noninvasive procedure and can be used to study and diagnosis of cancer, brain diseases and disorders, cardiology, and various disorders in different organs and tissues. Molecular imaging modalities include Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), Optical imaging, Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Ultrasound (US). Of all modalities, PET and SPECT are the full functional imaging modalities while others are only with limited abilities. Based on these imaging modalities, hybrid PET/CT [90], PET/MRI [74], PET/SPECT/CT [56] further enrich the ability of molecular imaging. Different imaging modalities require different considerations of data processing methods. The following sections will introduce the major molecular imaging modalities and their data processing methods.

1.1 Positron Emission Tomography

PET as a biomedical research technique and clinical diagnostic procedure is one of the most important applications in nuclear medical imaging devices. In the past three decades, there have been significant advancements in PET scanners and image processing methods [1, 92, 94]. PET scan is a unique type of imaging test that helps doctors see how the organs and tissues inside your body are actually functioning. PET scan reveals the cellular level metabolic changes occurring in an organ or tissue. This is important and unique because disease processes often begin with functional changes at the cellular level. Currently, PET scans are most commonly used to detect cancer, heart problems, brain disorders and other central nervous system disorders. PET scan can be used to track the spread of disease inside body and patient response to drugs and therapies, which help to determine the more effective treatment plans for individual patient. PET scans can also be used to follow-up and manage ongoing cares. Quantitative dynamic PET imaging also offers good promise for personalized drug treatment by accurate pharmacokinetic analysis and will enable medicine to be tailored to each person's needs, and improve the safety, quality and effectiveness of healthcare for every patient.

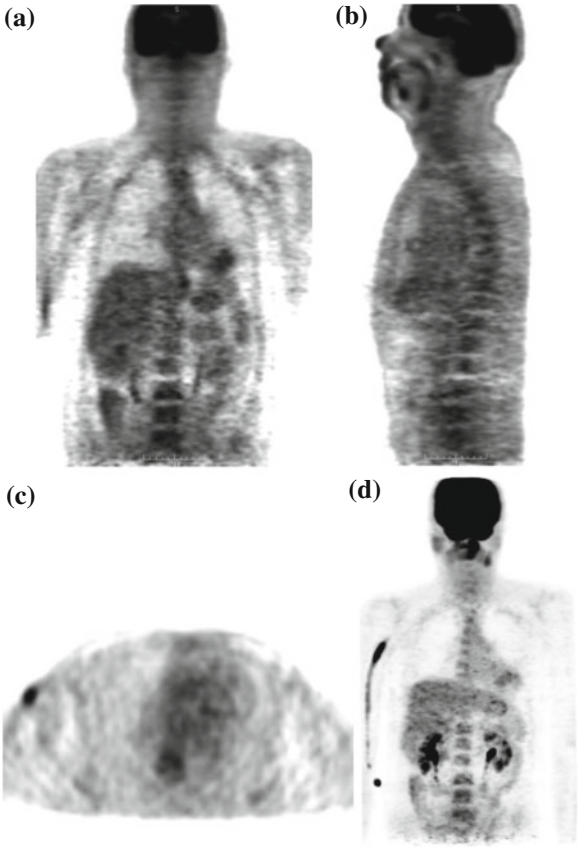


Fig. 1 PET images in different views: **a** Coronal view; **b** Sagittal view; **c** Horizontal view; **d** 3D view

PET scans rely on the injected radiotracers which circulate inside the body. PET scanner detects the pairs of gamma photons emitted from the radiotracer, which are known as positron annihilation, and creates a projection data of radiotracer distribution. The reconstructed images from projection data are used for diagnosis. PET images of different views from a volunteer are shown in Fig. 1. One currently used PET radiotracer for daily clinical routines is ^{18}F-FDG (Fluoro-2-deoxy-d-glucose), which is a compound consist of glucose and radioactive fluorine-18. When disease occurs, the activities of cells begin to change, for example, the cancer cells need more glucose and more active than normal cells, so there will be more radiolabeled ^{18}F-FDG accumulated in the cancer cells. With the PET images, it appears higher intense than surrounding tissues, which is called as a 'hot spot' indicating a high level of activity or metabolism is occurring there. Correspondingly, a 'cold spot' refers to the area of low metabolic activities indicated by lower intense in the PET images. With these PET images, doctors will be able to evaluate the working situation of organs

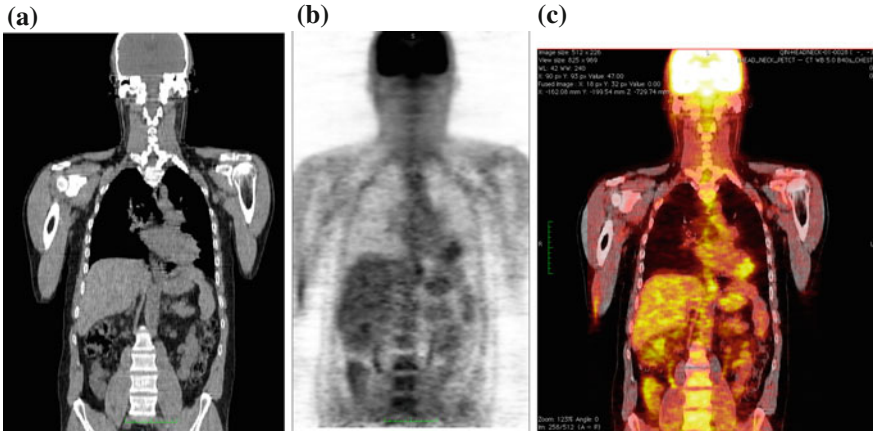


Fig. 2 PET/CT images: **a** CT image; **b** PET image; **c** Fused PET/CT image

and tissues and determine the abnormals by analyzing the hot spots and cold spots. The ability of detecting cellular level change that occurs early in the disease makes PET imaging superior to the CT and MR images which show the structural change after accumulations in cellular level. Hybrid PET/CT or PET/MRI is the combination of PET and CT or MRI, the combination of multiple imaging modalities allows both anatomical and functional images in one image set. One example of PET/CT images is shown in Fig. 2.

Different radiotracers will reveal different diseases. Besides ^{18}F-FDG mentioned before, which is widely used for cancer diagnosis, cardiology, neurology, there are many other radiotracers used in research and clinical applications. ^{18}F-FLT (3′-fluoro-3′-deoxy-l-thymidine) is developed as a PET tracer to image tumor cell proliferation [12], ^{11}C-Acetate is developed to localize prostate cancer [62], ^{13}N-ammonia is developed to quantify the myocardial blood flow [48], ^{11}C-dihydrotetrabenazine (DTBZ) is developed for brain imaging, which can be used for differentiating Alzheimer's disease from dementia and Parkinson's disease [47], ^{11}C-WIN35,428 is a cocaine analogue and sensitive to the dopamine transporters [93]. Researchers are working on labeling different drugs with radioactive ^{11}F, ^{11}C, ^{13}N etc. for PET scan, and pharmaceutical companies are especially interested in applying quantitative PET analysis on radio-labeled new drugs, this has the potential to shorten the Phase I to Phase II studies from more than 5 years to be 2 years.

Clinical PET studies include activity image reconstruction of radiotracer concentrations and parametric image reconstruction from dynamic PET scans. The quantitative accuracy of reconstructed images depends on the whole procedure: data corrections of measurement data, statistical modeling of acquisition process, and proper reconstruction methods. The data correction is the very first step, and directly determines the accuracy of following steps. The data correction consists of many specific corrections, including random correction, scatter correction, deadtime cor-

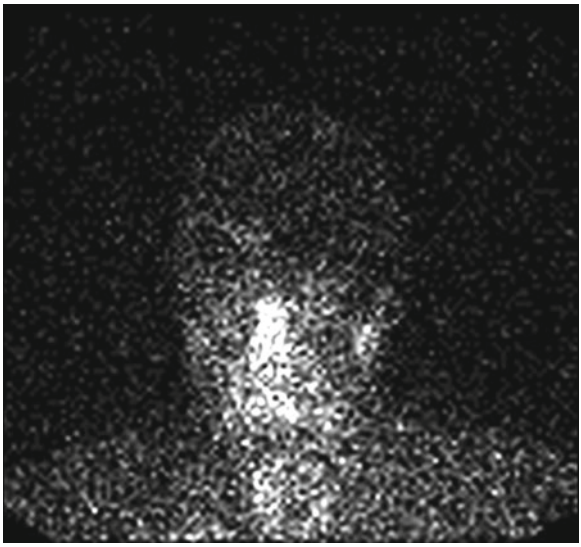


Fig. 3 One patient SPECT scan

rection, attenuation correction etc. Among all the corrections, scatter correction is the most complicated and still a very active research topic. The percentage of scatter coincidence events in measurement data is around 30 % for PET scanners from major manufactures. And their distributions are more complicated and will vary with the dosage of injected radiotracer and the status of detector system, furthermore, the scatter coincidence events can either be corrected or modeled in the system probability matrix. Additionally, the development of new full 3D PET scanner with new scintillators and Photomultiplier Tubes (PMT) [3, 6, 28, 40, 52, 87], requires corresponding new correction methods.

1.2 Single Photon Emission Computed Tomography

SPECT scan uses a gamma camera that rotates around the patient to detect the radiotracer inside body. SPECT will also produce a set of 3D images but generally have a lower resolution. The radiotracers commonly used for SPECT scan include ^{99m}Tc [59], ^{188}Re [39], ^{68}Ga [104], ^{82}Rb [23] etc. Electrocardiography (ECG)-Gated ^{82}Rb can also be used for myocardial perfusion PET [5]. One example image from SPECT scan is shown in Fig. 3. Hybrid SPECT/CT is also designed to provide more accurate anatomical and functional information [88]. SPECT scan differs from PET scan in that the tracer stays in your blood stream rather than being absorbed by surrounding tissues, therefore, SPECT scan can show how blood flows to the heart and brain are effective or not. SPECT scan is cheaper and more readily available than higher

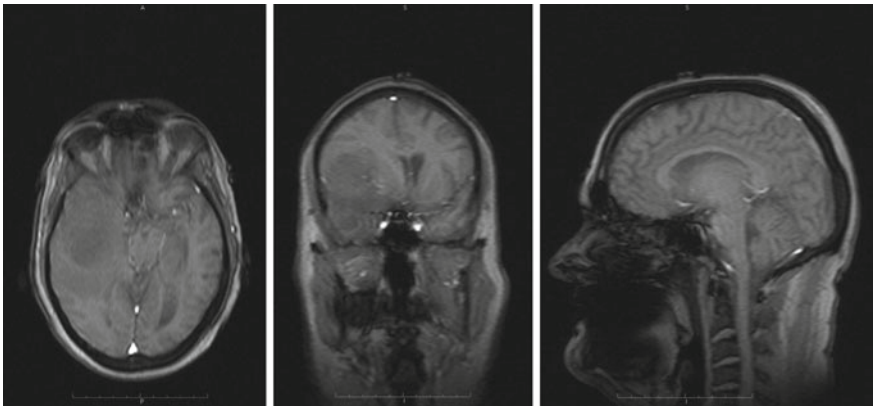


Fig. 4 One patient brain MRI scan

resolution PET scan. Tests have shown that SPECT scan might be more sensitive to brain injury than either MRI or CT scanning because it can detect reduced blood flow to injured sites. SPECT scan is also useful for presurgical evaluation of medically uncontrolled seizures and diagnosing stress fractures in the spine (spondylolysis), blood deprived (ischemic) areas of brain following a stroke, and tumors [4, 9].

1.3 Other Molecular Imaging Modalities

Other molecular imaging modalities include MRI, CT, Optical imaging, Ultrasound, which have only limited molecular imaging abilities.

1.3.1 MRI

MRI scanners produce strong magnetic fields where body tissue that contains hydrogen atoms is made to emit a radio signal which is detected by the scanner. MRI scanners produce detailed 3D images of the inside of the body. MRI scan is best for brain imaging, breast, heart and blood vessel, organs and soft tissues [13]. Contrast materials are often used to enhance MRI images. MRI has the highest spatial resolution and possibility to extract both physiological and anatomical information. However, MRI generally has low sensitivity and requires longer scan and data processing time. Samples of MRI images are shown in Fig. 4.

Functional magnetic resonance imaging or functional MRI (fMRI) is a MRI procedure that measures brain activity by detecting associated changes in blood flow. The primary form of fMRI uses the blood-oxygen-level-dependent (BOLD) contrast, and measures the change in magnetization between oxygen-rich and oxygen-poor blood, which is mostly used in brain mapping research [37].

Another MRI scan used in research is Dynamic Contrast Enhanced- MRI (DCE-MRI), which is a quantitative method that allows for tumor vascular analysis, including blood volume, perfusion, vascular leakage space. DCE-MRI uses gadolinum-based contrast agents and standard MRI scanners to provide quantitative results [34].

1.3.2 Optical Imaging

Optical imaging has the highest sensitivity, but relatively low spatial resolution. Optical imaging only provides images in limited Field of View (FOV). Optical imaging reduces patient radiation exposed significantly by using non-ionizing radiation. Optical imaging can be used to differentiate between native soft tissues and tissues labeled with either endogenous or exogenous contrast media, using their different photon absorption or scattering profiles at different wavelengths. Optical imaging is also easy to be combined with other imaging modalities. In optical imaging, Diffusive Optical Imaging (DOI) is also known as Diffuse Optical Tomography (DOT) or Optical Diffusion Tomography (ODT). DOI is used to study the functions of brains, and can provide neural activities with its time courses [51]. Optical Coherence Tomography (OCT) produces 3D images from optical scattering media and penetrates biological tissues by using long wavelength light. OCT can provide higher SNR, faster signal acquisition [78].

1.3.3 Ultrasound

Ultrasound scanners emit high frequency sound and measure the reflected sound from the patients, which varies with the organs and tissues. Ultrasound imaging can be used to detect heart problems, examine liver, kidneys, abdomens, and guide a surgeon. Ultrasound can provide a real time imaging, but with limited spatial resolution. Most scans need targeted micro-bubbles to enhance the images [21].

2 The Challenges in Molecular Imaging Data Analysis

2.1 Data Processing for PET Imaging

The previous sections explain the advantages and limitations of molecular imaging modalities. Of all molecular imaging modalities, PET and SPECT are the full functional ones. Since PET and SPECT all belong to Emission Computed Tomography (ECT), many data correction and data analysis methods can be shared. In this section, we will focus on PET to demonstrate the challenges in data analysis. As shown in Fig. 5, the measurement data of PET is first stored in listmode. The listmode data

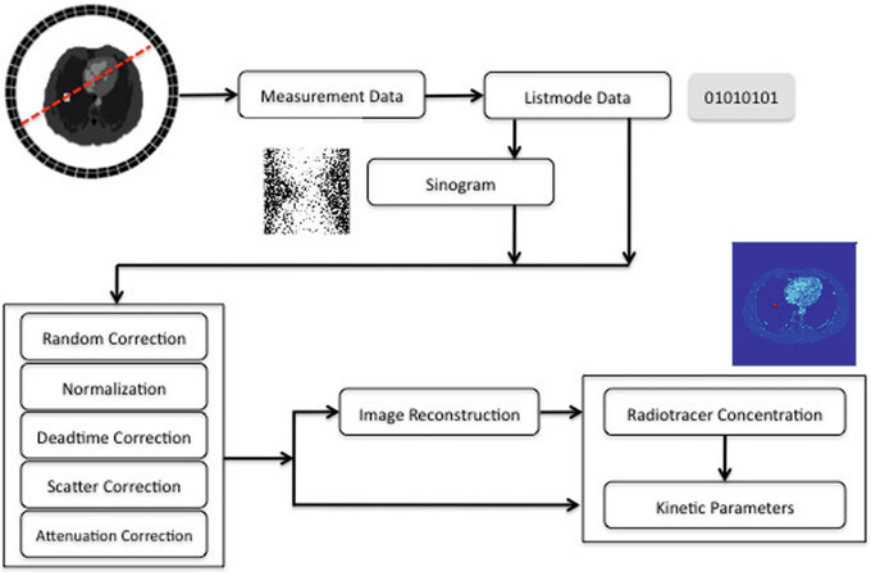


Fig. 5 The whole procedure of PET data processing

can be corrected and reconstructed directly [66, 72], or rebinned by Fourier Rebinning to sinogram, which is corrected and reconstructed to be images [18]. The data correction process generally includes random correction, normalization, deadtime correction, attenuation correction and scatter correction. The corrected data will be reconstructed by either analytical or statistical algorithms to generate both activity images which are the concentrations of radiotracers in body and parametric images which are the dynamic changes of radiotracers represented by kinetic parameters.

The challenges in PET data analysis come from the change of statistical properties of measurement data after various data corrections. The quality of results from all image reconstruction algorithms depends on the accuracy of statistical models in each data correction and image reconstruction. However, due to the complexity of PET scan, it is nearly impossible to propose a perfect model. From next subsection, we will explain how the data corrections are applied and how they affect the data analysis.

2.1.1 Data Corrections

1. Random Correction. Random coincidence events are two gamma photons from different positron annihilations are recorded as a Line of Response (LOR). The rate of random coincidences on a particular LOR is defined as $R_{ij} = 2\tau r_i r_j$, where R_{ij} is the random coincident rate on the LOR connecting detector i and j, τ is the coincidence timing window, r_i and r_j are the singles rate on detector

i and j. The random coincidence rate depends on the detector circuit response, the most common method for estimating the random coincidence events is the delayed timing window method.

2. Normalization. All the image processing methods assume the detector response are uniform and identical, however, in real clinical data acquisition, the performance of each detector will be different, the PMT gains are not exactly the same, and angle of incident of each gamma photon will also affect the detector sensitivity. The process of correcting these effects is normalization. One common clinical solution is to perform a scan using a uniform cylindrical phantom and a line source. In this scan, every possible LOR is illuminated by the same coincidence source.
3. Deadtime Correction. The detector deadtime will cause the loss of coincidence events. Especially in the high count rate cases, the detector response will be significantly delayed due to the pile-up of incident gamma photons. A common solution is to model both paralyzable and non-paralyzable components from measurements of sources of different radioactivities.
4. Attenuation Correction. Some gamma photons will be absorbed when interacting with the patients. The widely-adopted method for PET only scanner calculates the attenuation correction factors by using a rotating rod source with the object in the FOV. For hybrid PET/CT and PET/MRI scanners, the attenuation factors can be obtained by assigning predefined attenuation coefficients to the anatomical information generated by CT and MRI, which generally have a better spatial resolution.
5. Scatter Correction. The coincidence events include those that the annihilation photons have an interaction of Compton scattering and change their directional and energy information before they arrive at the detector system. They are called scatter coincidences. Scatter coincidences will decrease the contrast, resolution and SNR of reconstructed images and need to be corrected properly. Scatter correction is the most complicated correction, we will use the scatter correction as an example to show how corrections affect the quantitative accuracy in PET scan.

2.1.2 Scatter Correction

Scatter Fraction. The scatter fraction is an important parameter included in both National Electrical Manufacturers Association (NEMA) and International Electrotechnical Commission (IEC) standards, and measured in every performance evaluation report of PET scanners. The scatter fractions of PET scanners from 3 major manufactures are listed in Table 1 [28, 40, 87].

Multiple Scatter Coincidence Events. The gamma photons can be scattered multiple times before recorded. Monte Carlo simulation is the best tool to study the Compton scattering during PET acquisition. We have done similar studies with our new PET scanner, HAMAMATSU SHR74000, we retrieve all the data, and analyze the composition of simulated measurement data according to the number of

Table 1 Scatter fractions of PET scanners from three main manufactures in 3D mode

Manufacture and model	Scatter fraction (%)
Siemens TruePoint PET/CT	32
GE discovery PET/CT	33.9
Philips Gemini PET/CT	27

Table 2 The percentage of coincidence events assorted by the number of scattering of each photon

Photon2 \Photon1	0× (%)	1× (%)	2× (%)	3× (%)	4× (%)
0×	55.36	16.58	1.92	0.11	0
1×	16.56	5.81	0.68	0.04	0
2×	1.92	0.7	0.08	0	0
3×	0.11	0.04	0.005	0	0
4×	0.005	0.001	0	0	0

The data is from the simulation of a line source fixed in the transaxial center of standard phantom

Table 3 The percentage of coincidence events assorted by the number of scattering of each photon

Photon2 \Photon1	0× (%)	1× (%)	2× (%)	3× (%)	4× (%)
0×	37.66	18.87	3.7	0.39	0.03
1×	18.82	10.98	2.18	0.23	0.01
2×	3.75	2.18	0.44	0.05	0
3×	0.38	0.23	0.05	0.01	0
4×	0.02	0.02	0	0	0

The data is from the simulation of a line source fixed in the transaxial center of over-sized phantom

scattering of each photon in the coincidence photon pairs. The first data set is analyzed with the standard NEMA phantom, a 70 cm long, 20 cm diameter cylinder, the composition of scatter coincidences is summarized in Table 2. Furthermore, we have a lot of over-sized and over-weighted patients, especially in the western countries, and these patients usually take higher risk of many diseases than other patients. For these patients, their size makes more photons scattered inside their body, we correspondingly analyze the composition of scatter coincidence photons with a 70 cm long, 35 cm diameter cylinder, the results are summarized in Table 3. The number of total scatter coincidence events increase by about 20 %. If the tumor is near the body surface, the scatter fractions shown in Table 4 turn to be similar as in Table 2.

Scatter Coincidence from Activity Concentrations outside of FOV. A whole body PET scan generally needs 4–5 bed positions to complete, then gamma photons from outside of FOV will also have the possibility to be misrecorded by the detector system. Sossi et al. show the scatter coincidence events need to be accounted for when the amount of radioactivity outside of FOV is comparable to the radioactivity inside the FOV in [81]. In the case of line sources, the scatter fraction increases by about 15 % when the source is extended 4 cm outside of the FOV and 20 % when it is extended 8 cm outside of the FOV. Spinks et al. also show for a full 3D PET,

Table 4 The percentage of coincidence events assorted by the number of scattering of each photon

Photon2 \Photon1	0× (%)	1× (%)	2× (%)	3× (%)	4× (%)
0×	59.23	16.28	2.94	0.38	0.03
1×	16.29	1.18	0.1	0.01	0
2×	2.94	0.1	0	0	0
3×	0.38	0.01	0	0	0
4×	0.04	0	0	0	0

The data is from the simulation of a line source fixed in the transaxial 169 mm offset of over-sized phantom

the scatter fraction increased from 40 to 45 % with the scatter photons from outside of FOV, after introducing a fitted brain shielding, the scatter fraction decreases to be 41 % [82]. A recent paper by Ibaraki et al. show the use of the neck-shield to suppress scatters from outside FOV can improve the SNR by 8 and 19 % for $H_2\,^{15}O$ and $^{15}O_2$, respectively [38].

Scatter Correction Methods. Bailey et al. [2] and Bentourkia et al. [7] present a Convolution-Subtraction (CS) scatter correction technique for 3D PET data. The scatter distribution is estimated by iteratively convolving the photo-peak projections with a mono-exponential kernel. The method is based on measuring the scatter fraction and the scatter function at different positions in a cylinder. The method performs well on 2D measurement data and also accounts for the 3D acquisition geometry and nature of scatter by performing the scatter estimation on 2D projections.

The method is easy to set up and still applies to a lot of animal studies, where the scatter correction are usually considered not necessary. Lubberink et al. propose a non-stationary CS scatter correction with a dual-exponential scatter kernel for scatter correction in both emission and transmission data of studies of conscious monkeys using Hamamatsu SHR7700 PET scanner [54] . Kitamura et al. implement a hybrid scatter correction method, which estimates scatter components with a dual energy acquisition using a CS to estimate the true coincidence events in the upper energy window for their four layer Depth of Interaction (DOI)—PET scanner [45]. Naidoo-Variawa et al. suggest that scatter correction methods based on spatially invariant scatter functions, such as CS, may be suitable for non-human primate brain imaging in [65].

Single Scatter Simulation (SSS) is one of the most important methods for PET scatter correction, and commercial PET scanners from Siemens are using SSS derived scatter correction methods. Ollinger [67] and Watson [98] first introduced the concept of model based scatter coincidence estimation. The single scatter approximation is defined with the well accepted formulation based on Klein-Nishina equation as

$$S^{AB} = \int_v \left(\frac{\sigma_{AS}\sigma_{BS}}{4\pi R_{AS}^2 R_{BS}^2} \right) \frac{\mu}{\sigma_c} \frac{d\sigma_c}{d\Omega} [I_A + I_B], \tag{1}$$

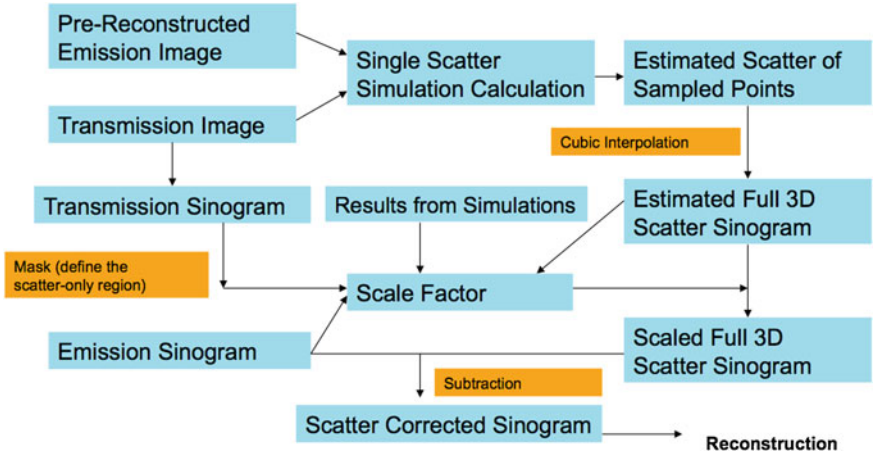


Fig. 6 The procedure of a SSS based scatter correction for clinical PET scanner

where

$$I_A = \varepsilon_{AS}(E)\varepsilon_{BS}(E')e^{-\int_S^A \mu(E,x)ds}e^{-\int_S^B \mu(E',x)ds} \times \int_S^A \lambda ds$$

$$I_B = \varepsilon_{AS}(E')\varepsilon_{BS}(E)e^{-\int_S^A \mu(E',x)ds}e^{-\int_S^B \mu(E,x)ds} \times \int_S^B \lambda ds$$

v is the sampled scatter volume, S is the scatter point, A and B are sampled detectors, so S^{AB} is the coincidence scatter rate in detectors A and B, σ_{AS} and σ_{BS} are the detector cross sections presented to the rays AS and BS, R_{AS} and R_{BS} are the distances from S to A and B. σ_c and $d\sigma_c$ are the total and differential Compton scattering cross sections, Ω is the solid angle, E is the energy of the non-scattered photon, E' is the energy of the scattered photon, $\varepsilon_{AS}(E)$ and $\varepsilon_{BS}(E)$ are the approximated detector efficiencies for gamma rays which are incident along AS and BS, $\mu(E, x)$ is the linear attenuation coefficient at the energy E and position x, and λ is the estimated activity distribution from reconstructed images.

The calculation is repeated through all the sampled scatter points from activity distribution and sampled detector blocks. The final result is the distribution of single scatter coincidence events, and will be scaled and then subtracted from the measurement data to perform scatter correction. Figure 6 demonstrates the procedure of SSS based scatter correction for clinical PET scanner.

2.1.3 Dynamic PET Imaging

Dynamic PET imaging is a combination of short interval PET scans and reflects the dynamic metabolism of injected radiotracers. For example, a dynamic acquisition can

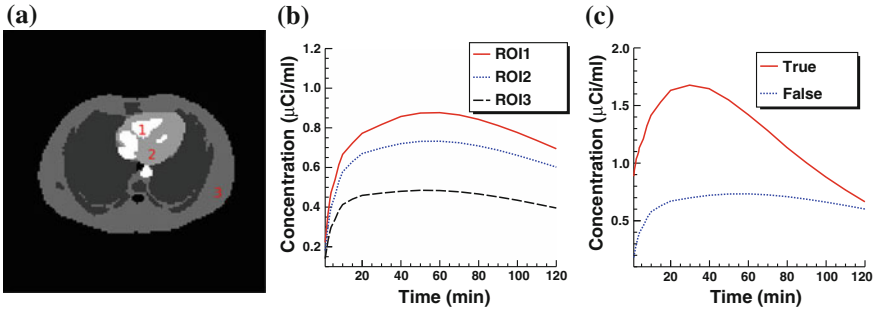


Fig. 7 **a** Standard zubal phantom; **b** TAC curves of three ROIs indicated in (**a**); **c** sample TAC curves of true lesion and false lesion

consist of 85 frames in all: 15×0.2 min, 20×0.5 min, 40×1 min, and 10×3 min. The series of acquisitions can be used to estimate the kinetic parameters which represent the metabolism of radiotracers in vivo. The difficulties in estimating kinetic parameters arise from the low count rates in the first several time frames and the low SNR. To obtain the kinetic parameters, a typical approach is first to reconstruct the activity distributions from the dynamic PET data, and then to fit the calculated time activity curve (TAC) to a predefined kinetic model. Samples of TAC curves in different organs are shown in Fig. 7. The accuracy of this kind of approaches relies on the reconstructed activity distributions. The complicated statistical noise properties, especially in the low-count dynamic PET imaging, and the uncertainties introduced by various PET data corrections will affect the activity reconstruction and lead to a suboptimal estimation of kinetic parameters [29]. There are also many efforts that try to estimate the kinetic parameters from PET projection data directly and achieve better bias and variance including both linear and nonlinear models [83, 91, 103]. One problem is that the optimization algorithms are very complicated. Kamasak et al. apply the coordinate descent algorithm for optimization but it is still limited to specific kinetic models [61]. Wang et al. apply a generalized algorithm for reconstruction of parametric images [96], however, it still lacks of estimation and analysis of individual kinetic parameter. Exactly, every kinetic parameter has its own physical meaning like radiotracer transport rate, phosphorylation rate and dephosphorylation rate (it is very sensitive to the system [35]) in FDG study, which will be critical to clinical research, drug discovery and drug development [15, 92].

3 General Framework of Shape Analysis in Molecular Imaging

3.1 Image Segmentation

In computer vision, image segmentation is the process of partitioning an image into multiple different segments (group of pixels). Especially in molecular imaging, the image segmentation is used to simplify the representation of an image and extract Region of Interest (ROI) that is more meaningful and followed by image analysis. Image segmentation is also important to find the boundaries of different regions and organs by applying different labels. Image segmentation can also be applied to 3D image stacks to help 3D image reconstruction [17, 106].

The aforementioned limitations of PET and SPECT also bring new challenges to image segmentation. Here we will introduce several basic image segmentation methods.

1. The simplest method of image segmentation is the thresholding method. Because of the simplicity and fast implementation, the thresholding method is still used in some clinical routines to identify ROI with high contrast, e.g. the lesions in lung [19, 20, 42].
2. Cluster based method is a multivariate data analysis method that uses predefined criteria to partition a large number of objects into a smaller number of clusters, in which the objects are similar to each other. Cluster based method has been applied to fMRI imaging and then dynamic PET imaging, with limited spatial resolution and SNR [101].
3. Gradient based method is to find the boundary of an object of interest with the gradient intensity observed in the image. This kind of method is fast and easy to apply, but generally works together with other method to achieve better results in molecular imaging [26].
4. Level set based method derives from snake method (active contour model), which delineates an object outline from a noisy image by attempting to minimize an energy associated to the contour as a sum of internal and external energies [57, 68]. Many methods are derived from basic level set method including deformable level set models, which have the ability to automatically handle topology [33], 3D level set methods to compute 3D contours [105] etc.
5. Kinetic model guided segmentation assumes different ROIs have different tracer kinetic properties to seprate different functional regions [11]. This method can also be used to estimate the input functions for quantitative dynamic PET data analysis [71].

3.2 Image Registration

Image registration is the process to transform different sets of data into one coordinate system. Image registration is widely used in molecular imaging, e.g. patient radiotherapy follow-up by transforming PET images from a series of studies, diagnosis by images from multiple imaging modalities [16, 32, 36, 75]. With the development of hybrid PET/CT, PET/MRI, the image registration with multiple images in one study is made easier because the motions of patients are minimized by the simultaneous data acquisition. However, images from multiple studies still need good image registrations. Mainstream image registration methods for molecular imaging include

1. Intense-base image registration. Since PET and SPECT imaging reflects the concentrations of radiotracer, intense based methods compare intensity patterns in multiple images and register the reference image and target image by defining correlation metrics [46].
2. Feature-based image registration. In a series of studies/images, common features can be extracted from the anatomical information of organs and tissues which do not change a lot and can be used as references [73]. This method can also be used for multiple imaging modalities [55].
3. Multiple modality image registration. The hybrid PET/CT and PET/MRI make the image registration focus on the deformations from patients' respirations and motions [27, 60]. The image registration can be improved by different patient preparation and pre-positioning [8], respiratory gating [10], various tracking devices etc. [76].

3.3 Image Fusion

Image fusion is the combination of relevant information from two or more images into one single image. The fused image will provide more information than any single input image. Accurate image fusion from combined PET, CT, MRI scans can significantly improve the diagnosis and provide better understandings of diseases. Image fusion generally works together and shares similar technologies with image registration [69, 89, 95].

3.4 Image Reconstruction

Image reconstruction is the process to reconstruct 2D and 3D images from acquisition data of molecular imaging modalities. Reconstruction algorithms include both analytical ones, e.g. Filtered Back Projection (FBP) and iterative ones, e.g. Maximum Likelihood (ML). Analytical algorithms are computationally fast, especially when applied to full 3D image reconstruction using 3D-FBP. FBP is also used in dynamic

PET image reconstruction, where it is believed to provide better quantitative accuracy with the extremely low count data sets. Iterative algorithms are currently the mainstream reconstruction algorithms, which use statistical assumptions and provide images of overall better qualities. Image reconstruction is one of the most important processes in data processing, other image-based processes and clinical diagnosis all depend on the accuracy of reconstructed images.

3.5 Dynamic PET Analysis

3.5.1 Clinical Requirements

The accuracy of quantitative dynamic PET studies depends on various factors including kinetic models, quantitative methods and the approximation of arterial input function from blood sampling. The most general kinetic models used are compartment model with assumptions that physiological process and molecular interactions are not influenced by injected radioligand. Current clinical adopted quantitative methods are actually semi-quantitative methods, which include methods using reference regions or calculating Standard Uptake Value (SUV). Methods using reference regions are easy to implement but have several drawbacks, e.g. the reference tissue is hard to define and has low SNR due to the low resolution of PET and SPECT scans, and the uptake of the reference tissue may change after the radiotherapy. SUV now is included in every clinical study, which is calculated as a ratio of tissue radioactivity concentration and injected dose divided by body weight, the advantage of SUV in clinical study is that the blood sampling is not required. However, the full quantitative analysis requires both dynamic PET scans and tracer concentration in the arterial blood plasma. The gold standard of blood sampling is serial arterial sampling of a superficial artery, and clinical alternative methods include venous blood sampling, image derived input function and population based input function. The drawback of the full quantitative method is only one FOV/bed position can be taken into consideration. For metastasized disease, not all lesions can be quantified simultaneously.

3.5.2 Compartment Models

Compartment models are used in many fields including pharmacokinetics, biology, engineering etc. Compartment models are the type of mathematical models to describe the way materials (radiotracers and their metabolite in PET and SPECT scan) are transmitted among the compartments (different organs and tissues). Inside each compartment, the concentration of radiotracers is assumed to be uniformly equal. Due to their simplicity and plausibility, compartment models are widely used in the dynamic PET scans to describe the tracer/drug kinetics.

Drug kinetic models include simple drug transport model, which generally contains equal or less than three compartments and can be solved directly, and compli-

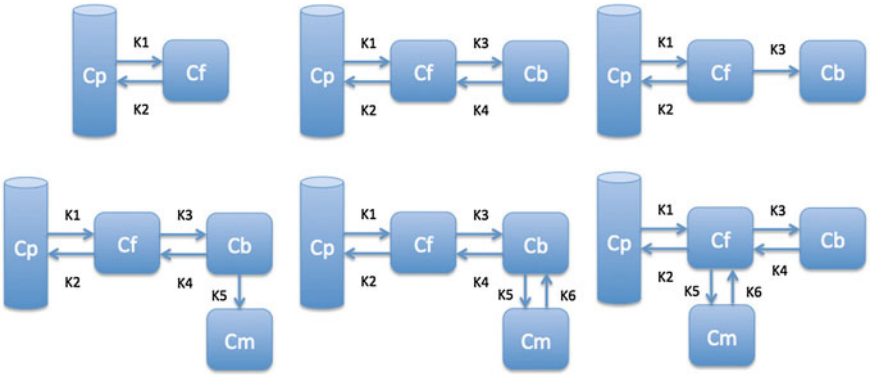


Fig. 8 Illustration of six basic compartment models

cated biological models, which can contain up to twenty compartments and generally require prior knowledge to solve [30, 31]. Most of the complicated models with many compartments can usually be decomposed into a combination of simple models with less than four compartments. The most basic compartment models used in kinetic analysis shown in Fig. 8 include two compartment blood flow model (Model 1), standard two tissue three compartment Phelps 4K model with reversible target tissue (Model 2) and Sokoloff 3K model with irreversible target tissue (Model 3), three tissue five parameter bertoldo model (Model 4), standard three tissue four compartment model (Model 5 and Model 6). More complicated models with more compartments and parallel model with multiple injection can be extended from aforementioned standard models [24, 44].

All six models can be represented by a set of differential equations with corresponding kinetic parameters $\mathrm{K} = \{k_1, k_2 \cdots k_n\}$, where n is the number of kinetic parameters. Here we utilize Model 2 as an example for demonstration. Model 2 can be represented by first-order differential equations

$$\frac{dC_F(t)}{dt} = k_1(t)C_P(t) + k_4(t)C_B(t) - (k_2(t) + k_3(t))C_F(t) \tag{2}$$

$$\frac{dC_B(t)}{dt} = k_3(t)C_F(t) - k_4(t)C_B(t) \tag{3}$$

The measurement of dPET is the combination of radiotracer in plasma C_P, non-specific binded radiotracer C_F and specific binded radiotracer C_B through

$$C_{PET} = (1 - V_b) \cdot (C_F + C_B) + V_b \cdot C_P \tag{4}$$

$$Y = DC_{PET} + e \tag{5}$$

where V_b is the blood volume fraction, Y is measured projection data , D is the system probability matrix, and e is the noises during acquisition. Equation (5) can be represented by a more general time-dependent form for all models as

$$Y(t) = DX(K, t) + e(t) \tag{6}$$

Simple models with less than three compartments generally can be solved directly, while more complicated models need simplifications and various numerical approximations.

4 Review of Recent Advancements of Shape Analysis

4.1 Mathematical Modeling and Statistical Formulation of PET Image Reconstruction

4.1.1 Statistical Image Reconstruction Criterion

The goal of mathematical modeling of data acquisition is to describe the transforms from spatial distribution of imaging objects to projection distributions on detector pairs in PET system. Denoting the spatial distribution of imaging object by a set of spatial variables $x = \{x_i | i = 1 \cdots n\} \in \mathbb{R}^n$, where n is the total number of voxels, and the expected values (means) of projection bins by $\bar{y} = \{\bar{y}_j | j = 1 \cdots m\} \in \mathbb{R}^m$, where m is the total number of bins, a mathematical expression of the transform can be obtained

$$\bar{y} = Dx + \bar{e} \tag{7}$$

where D is the system response model giving the probability matrix of mapping the transform from x to $\bar{y}$, and $\bar{e}$ is the means of background noises. A block diagram of the above procedure is shown in Fig. 9a.

With the mathematical model specified above, the problem of PET image reconstruction is to find an estimation of the imaging object $\hat{x}$ from the measurement data (projection bins) y. The image reconstruction is an ill-posed inverse problem, so an intuitive solution is to apply statistical assumptions of measurement data as regularizations. The statistical formulation tries to find an optimized relationship between measurement data y and imaging object x (or expected values of projection bins $\bar{y}$) by defining different objective functions. If denoting y_{rndsa} is the measurement data y after all data corrections and assuming all the data correction are perfectly applied, $\bar{y}$ is equal to the y_{rndsa} after all data corrections, however, in real clinical data acquisition, there will be some difference between $\bar{y}$ and y, which is accumulated from residuals of each data correction.

A block diagram of statistical image reconstruction framework is shown in Fig. 9b: the statistical properties of system noises or measurement data are first modeled based on certain statistical distributions (Gaussian, Poisson, their combination or other derivations) in block M, then inputted into system block D; the system output is generated and compared with corrected measurement data y_{rndsa} based on the predefined criteria, when the system goes convergent, estimations of the imaging

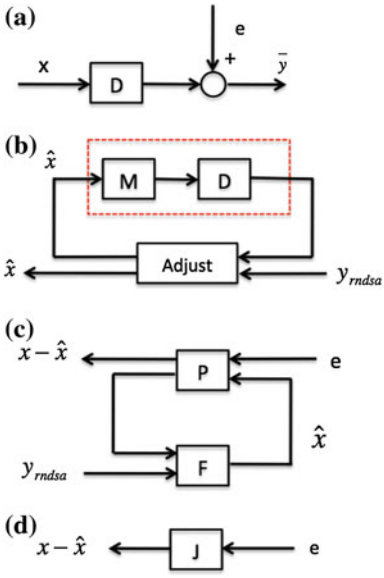


Fig. 9 Block diagrams. **a** PET data acquisition; **b** Statistical model based iterative PET reconstruction; **c** Designed system for PET image reconstruction; **d** Simplified block diagram (*black box*) of (**c**)

object $\hat{x}$ will be obtained. Three major criteria can be used are Least Square based (LS), Maximum Likelihood based (ML), Maximum A Posterior based (MAP).

LS based methods try to obtain the estimation $\hat{x}$ by minimizing the difference of fit between the predicted data by means of the modeling of the acquisition process and the measured data y_{rndsa}, the general objective function is

$$\hat{x} = arg \min_{x} ||y_{rndsa} - Dx||_2^2 \tag{8}$$

Extended algorithms are all based on the above equation and try to improve the performance by introducing different weights or penalization items.

ML based methods try to obtain the estimation $\hat{x}$ by maximizing the likelihood functions which represent the goodness of fitting statistical assumptions to measurement data. The general objective function is based on the conditional probability density of the image object with known measurement data y_{rndsa} as

$$\hat{x} = arg \max_{x} \; p(x|y_{rndsa}) \tag{9}$$

where p represents the probability density, and the above equation is a function of x. When a statistical model of measurement data or noises is defined, either Gaussian/Poisson or their combination, the objective likelihood function can be solved correspondingly. An improved method based on poisson assumption widely used is EM-SP (Shifted Possion) algorithm, which introduces two times the mean of randoms coincidences to the random precorrected data and models this sum as a Poisson random variable, and the objective function is still based on the above one [63].

Furthermore, *a priori* information can also be introduced in the form of statistical properties of the imaging object as constraints into the ML objective function.

Bayesian formulation is such a probabilistic approach, statistical information of the unknown imaging object x is introduced by adopting the probability density of x, which is the prior $p(x)$. Equation (9) can then be solved as follow:

$$p(x|y_{rndsa}) = \frac{p(x, y_{rndsa})}{p(y_{rndsa})} = \begin{cases} \frac{p(y_{rndsa}|x)}{p(y_{rndsa})} & \text{without prior} \\ \frac{p(y_{rndsa}|x)p(x)}{p(y_{rndsa})} & \text{with prior} \end{cases} \tag{10}$$

where, $p(x, y_{rndsa})$ is the joint probability density of x and y_{rndsa}. The expression with prior is just the objective function of MAP based algorithms, and different priors (e.g. image priors, independent priors) will lead to different implementations based on the above objective function. All above objective functions demonstrate the implied statistical knowledge assumptions on measurement data or noises in mathematical modelings, however, as discussed in the section of data corrections, the statistical distribution of corrected measurement data y_{rndsa} is neither single Gaussian/Poisson distribution nor their combination after various data corrections, and there will be further uncertainties during modelings. Due to the individual differences of patients during real PET scanning, the uncertainties during modeling will be more serious for over-weighted patients. With the development of new scintillators, PMTs and full 3D PET scanner, various system uncertainties and procedures of photo counting become more complicated. Additionally, there will be more scatter coincidences in measurement data. All above effects make the accurate modeling of PET measurement data an challenging issue, and it is almost impossible to establish a statistical model which can accommodate all the data errors.

4.1.2 Minimax Criterion

To deal with the differences between $\bar{y}$ and y, another solution is to adopt minimax criterion to reconstruct the measurement data y_{rndsa} from the point view of system energy. As shown in Fig. 9c, a new block F is designed to reconstruct measurement data y_{rndsa}, the output of the new system is designed to be the difference between expected values of imaging object x and estimation values $\hat{x}$, a block P contains system response D and the input is all residuals and uncertainties e. Measurement data y_{rndsa} is constant during image reconstruction, so the system can be simplified with inner loop of block F and y_{rndsa} incorporated into block P and convert to a black box J as shown in Fig. 9d. When the output (difference between current value and estimated value) goes steady, the system J reaches convergent.

With this system, the minimax criterion can be adopted intuitively, the PET image reconstruction procedure becomes to minimize the output (estimation errors $x - \hat{x}$) with maximized input (disturbance e) through system J:

$$\hat{x} = \min_{x-\hat{x}} \max_{e} J \tag{11}$$

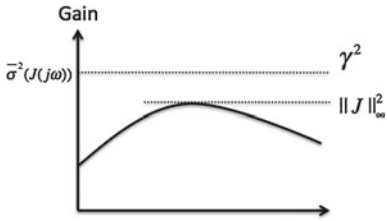


Fig. 10 Illustration of system gains and predefined upper bound

From the point view of transition system in engineering, system J can be treated as a transformation from all residuals and uncertainties e to estimation errors $x - \hat{x}$. In order to obtain a desired output, a proper upper bound for system gain of system J must be defined first, which means the transfer function of J shall also be required to comfort to the same predefined upper bound. Denoting the upper bound be γ^2, the previous description is equal to

$$\max ||J||^2 < \gamma^2 \tag{12}$$

The ∞-norm of the system can be further interpreted as the peak system gain, so if the ∞-norm of the system satisfies $||J||_\infty^2 < \gamma^2$, then all the system gains will be less than γ^2 as illustrated in Fig. 10. Here J is an element of the Hardy space, whose members consist of all stable, causal, transfer functions [49]. The continuous form of the ∞-norm $||J||_\infty$ will be

$$||J||_\infty^2 = \sup_{||e||_2 \neq 0} \frac{||x - \hat{x}||_2^2}{||e||_2^2} = \sup_\omega \bar{\sigma}^2(J) \tag{13}$$

where sup stands for supremum, and $\bar{\sigma}(J)$ is the maximum singular value of J. From Eq. (13), it is easy to obtain

$$\max ||J|| \leq ||J||_\infty^2 = \sup_{||e||_2 \neq 0} \frac{||x - \hat{x}||_2^2}{||e||_2^2} < \gamma^2 \tag{14}$$

This criterion represents a family of solutions where the peak system gain of J is less than the predefined upper bound γ^2. The problem is transformed from seeking the maximum system gain of J to calculating the ∞-norm of system gain with upper bound γ^2.

With both Eqs. (11) and (14), in order to calculate the supremum, intuitively we can minimize the numerator (estimation errors $x - \hat{x}$) while maximizing denominator (system uncertainties e). Then the objective function for the problem will be

$$\min_{x - \hat{x}} \max_e ||x - \hat{x}||_2^2 - \gamma^2 ||e||_2^2 < 0 \tag{15}$$

Seeking solutions under disturbances is a difficult problem. By introducing the ∞-norm of system gain, solutions with peak system gain will perform well under any circumstance and make the problem globally optimized. Furthermore, the minimax criterion allows one to identify a robust solution that has the best worstcase performance. The robustness of the estimator arises from the fact that it yields an energy gain less than γ^2 for all bounded energy disturbances no matter what they are.

4.1.3 Static PET Image Reconstruction Under Minimax Criterion

PET images reflect the concentration and metabolism of radiotracers *in vivo*, and the metabolic rates of radiotracers in organs or tissues may vary with time, which can be represented by differential equations with different constraint models as $\dot{x}(t) = C(x, t)$, where $C(x, t)$ models the concentrations of radiotracers with time.

In this section, we focus on the static PET image reconstruction and analysis, which assumes that the distribution of the radiotracers in the body is temporally stationary corresponding to the autoradiographic (ARG) model. The ARG assumes the equilibrium of the metabolic ratio, which can be represented by the first-order differential equation of x as:

$$\dot{x}(t) = 0 \tag{16}$$

Corresponding discrete-time form will be

$$x'(m) = H(m)x(m) + v(m) \tag{17}$$

where m represents the iterations in one discretized time frame, H is the transition matrix representing the update of estimations, and v is the possible uncertainties in measurement.

Similarly, the discrete-time form of PET measurement equation will be

$$y_{rndsa} = Dx(m) + e(m) \tag{18}$$

Then the transition system J is extended to include uncertainties v as input together with e. The initialization of x can also affect the reconstruct accuracy and speed, so it is also included in the input as a source of uncertainties. Finally, the minimax criterion can be derived from Eq. (14) as

$$||J||_{\infty}^2 = \sup \frac{\sum_m \|x(m) - \hat{x}(m)\|_{Z(m)}^2}{\|x(0) - \hat{x}(0)\|_{p_0^{-1}}^2 + \sum_m (\|v(m)\|_{V(m)^{-1}}^2 + \|e(m)\|_{E(m)^{-1}}^2)} \tag{19}$$

$Z(m)$, p_0^{-1}, $V(m)^{-1}$ and $E(m)^{-1}$ are the weighting matrices at iteration m to make the criterion more extensible. After setting the upper bound γ^2, the final objective function of minimax criterion for PET image reconstruction can be derived based

on Eqs. (15) and (19), which minimizes the estimation errors $z - \hat{z}$ with maximized initial disturbance $x(0) - \hat{x}(0)$, measurement disturbance e and v:

$$\min_{z(m)-\hat{z}(m)} \max_{v,e,x(0)-\hat{x}(0)} ||J||^2 = \sum_m \|x(m) - \hat{x}(m)\|^2_{Z(m)} - \gamma^2 \|x(0) - \hat{x}(0)\|^2_{p_o^{-1}} - \gamma^2 \sum_m (\|v(m)\|^2_{V(m)^{-1}} + \|e(m)\|^2_{E(m)^{-1}}) \quad (20)$$

4.1.4 Solutions and Optimization

The final objective function Eq. (20) of minimax criterion for static PET image reconstruction is established from the point view of transition system in engineering. The objective function calculates the system energy instead of building sophisticated statistical assumption of measurement data, and this makes the minimax criterion more robust.

Of many methods that can be adopted to solve the above objective function, the well-validated H_∞ filter is the best to optimize this ∞-norm problem [49, 79]. A specific H_∞ filter has been modified based on Eq. (20). The H_∞ filter represents a typical minimax problem where the worst situation is first induced by the disturbances and then the estimator is introduced for improvement, in other words, the H_∞ filter is in fact a two-person game between the external disturbances and the estimator. This solution is just like optimization by using a game theoretic algorithm which can be implemented through recursive updating of the filter gain $K(m)$, the Riccati difference equation solution $P(m)$, and the state estimates $\hat{x}(m)$ as follow :

$$K(m) = H(m)P(m)S(m)D^T E(m)^{-1} \quad (21)$$

$$P(m+1) = H(m)P(m)S(m)H(m)^T + V(m) \quad (22)$$

$$\hat{x}(m+1) = H(m)\hat{x}(m) + K(m)(y_{rndsa} - D\hat{x}(m)) \quad (23)$$

$$S(m) = (I - \gamma^{-2} Z(m)P(m) + D^T E(m)^{-1} DP(m))^{-1}$$

$$P(0) = p_0$$

where H_∞ gain $K(m)$ indicates the system gain (correspondingly shows convergence of the state estimation). Convergent estimations of the imaging object will be obtained when the gain $K(m)$ goes steadily. m still represents the number of iterations. From the solution procedure, it can be noticed that the H_2 norm filter for this objective function is just the Kalman filter widely used. Detailed proofs of above solution can be found in [79].

4.1.5 Computation Issues

Design of H_∞ filter consists of choosing the weighting matrices Z, E, V, p_o and the performance bound γ^2. When there are accurate modelings of some effects of PET acquisition, corresponding E, V, p_o can be initialized by the covariance matrices of measurement disturbance e, state transition disturbance v and initial condition $\hat{x}(0)$. If there is no modeling or one does not want to use current models, E, V, p_o can be simply initialized by identity matrix. Since we generally assume the estimation results after convergence are just what we desired, so Z is initialized by identity matrix, when scanning over-weighted patient and suffering obvious underestimation, weighting matrix Z can be scaled by a scaler less than 1. γ^2 is the predefined upper bound of performance, theoretically, the smaller the γ^2 value, the smaller the estimation error, however, the selection of γ^2 must make the Riccati equation have a positive definite solution. So firstly, we define and iteratively update a residual matrix $R(m+1)^{-1}$ through

$$R(0) = (P(0)^{-1} - \gamma^{-2}Q(0))^{-1} \tag{24}$$

$$\begin{aligned} R(m+1)^{-1} = &[H(m)(R(m)^{-1} + D^T E(m)^{-1} D)^{-1} H(m)^T \\ &+ V(m)]^{-1} - \gamma^{-2} Z(m) \end{aligned} \tag{25}$$

As a result, the optimal γ value can be determined as:

$$\begin{aligned} &[H(m)(R(m)^{-1} + D^T E(m)^{-1} D)^{-1} H(m)^T + V(m)]^{-1} - \gamma^{-2} Z(m) > 0 \\ \rightarrow\ &[H(m)(R(m)^{-1} + D^T E(m)^{-1} D)^{-1} H(m)^T + V(m)]^{-1} > \gamma^{-2} I \\ \rightarrow\ &\gamma^2 I > H(m)(R(m)^{-1} + D^T E(m)^{-1} D)^{-1} H(m)^T + V(m) \\ \rightarrow\ &\gamma^2 > max\{eig[H(m)(R(m)^{-1} + D^T E(m)^{-1} D)^{-1} H(m)^T + V(m)]\} \\ \rightarrow\ &\gamma = \xi max \left\{ eig[H(m)(R(m)^{-1} + D^T E(m)^{-1} D)^{-1} H(m)^T + V(m)] \right\}^{0.5} \end{aligned} \tag{26}$$

where $max\{eig(A)\}$ denotes the maximum eigenvalue of the matrix A, and ξ is a constant larger than 1 to ensure that γ is always greater than certain optimal performance level. If the γ value is too close to the optimal performance level, *i.e.* $\xi \approx 1$, it might lead to numerical errors because the matrix $R(m)$ is now close to a singular matrix. The matrix inverse is required in every time step in the conventional H_∞ filter in order to calculate the H_∞ gain. Generally, inversion of small matrices is fairly easy, but the inversion of a large matrix will require more computational costs in a practical implementation. The steady state H_∞ filter is much easier to be implemented in a system in which real-time computational effort or code size is a serious consideration [80].

The minimax objective function is given as Eq. (20), where the parameter matrices N, V and Q are symmetric positive definite matrices chosen by the designer based

on the specific problem. Since the designed parameters of the underlying system can be treated as fixed values for input, then the steady state solution to the minimax problem can be obtained. Referring to H_∞ filter, the steady state solution will be

$$\begin{aligned} K &= PSD^T N^{-1} && (27)\\ P &= HPSH^T + V && (28)\\ \hat{x}(m+1) &= H\hat{x}(m) + HK(m)(y(m) - D\hat{x}(m)) && (29)\\ S &= (I - \gamma^{-2}\bar{Q}P + D^T N^{-1} DP)^{-1} \\ \bar{Q} &= \mathscr{F}^T Q \mathscr{F} \end{aligned}$$

In order to have a solution to the problem, the following condition must be hold:

$$P^{-1} - \gamma^{-2}\bar{Q} + D^T N^{-1} D > 0 \tag{30}$$

If γ^{-2}, $\mathscr{F}$, N or Q is too large, or D is too small, the H_∞ estimator will have no solution. After the conditions above satisfied, Eq. (28) can be written as

$$P = H[P^{-1} - \gamma^{-2}\bar{Q} + D^T N^{-1} D]^{-1} H^T + V \tag{31}$$

Applying the matrix inversion lemma to the inverse of the above expression, we can get

$$\begin{aligned} P &= HP - P[P + (-\gamma^{-2}\bar{Q} + D^T N^{-1} D)^{-1}]^{-1} PH^T + V \\ &= HPH^T - HP[P + (-\gamma^{-2}\bar{Q} + D^T N^{-1} D)^{-1}]^{-1} PH^T + V \end{aligned} \tag{32}$$

Equation (32) is a discrete-time algebraic Riccati equation that can be solved by control system software or numerically iterating Eq. (28) until it converges to a steady state value.

The disadvantage of the steady state H_∞ filter is that theoretically it does not perform as well as the time-varying filter. However, the reduced performance that is seen in the steady state H_∞ solution is often a small fraction of the optimal performance, whereas the computational saving can be significant [80].

4.2 Dynamic PET Image Reconstruction

4.2.1 Radioisotope Decay Constrained Dynamic PET Imaging

Both PET and SPECT use radiotracers, which decay with time. The natural decay property of the radioisotope, can be introduced into the objective function as the temporal guidance for multi-frame image sequence reconstruction, and can also be used to separate multiple radiotracers in dynamic imaging. The projection equation

of dynamic PET imaging can be formulated through an affine transform between the projection data and emission object as:

$$y(t) = Dx(t) + r(t) + s(t) \tag{33}$$

where the emission sinogram data is represented by a vector y, and the activity of emission object is represented by x. D is system probability matrix, which gives the probability of a photon emitted from ith voxel being detected in projection jth bin. t is the time frame. r and s are the contribution of random coincidence events and scatter coincidence events. After the conventional online delayed-window random correction, the Eq. (33) can be rewritten as:

$$y(t) = Dx(t) + e(t) \tag{34}$$

here e is an error vector, which represents *unknown* measurement uncertainties including scatter coincidence events.

In the section of static PET imaging reconstruction, the distribution of the radioisotopes in the body is assumed to be temporally stationary corresponding to the autoradiographic model, however, in the real situation, the radioisotope will decay with time, and its activity at time t should be

$$x = X_0 e^{\frac{ln(0.5)}{T}t} \tag{35}$$

here X_0 is the initial activity distribution, and T is the half life of the radioisotope. So the real-time change of radioisotope can be represented as

$$\frac{dx}{dt} = X_0 \frac{ln(0.5)}{T} e^{\frac{ln(0.5)}{T}t} \tag{36}$$

then the dynamic change of radioisotope from one frame to the next can be obtained from the integral of Eq. (36). A general representation of state transition will be

$$x(t+1) = H(t)x(t) + v(t) \tag{37}$$

where $x(t)$ is the radioactivity concentration at time frame t, and $H(t)$ is a coefficient matrix for state transition at time frame t. $v(t)$ represents the uncertainties during state transition. With the introduction of decay model shown as Eq. (37), we are able to make use of the radioisotope's own temporal properties as constraints to guide our reconstruction.

4.2.2 Dynamic PET Image Reconstruction with Minimax Criterion

Minimax criterion, which allows one to identify a robust solution as one that has the best worstcase performance can also be applied to dynamic PET reconstruction. In

general, a robust discrete optimization problem can be formulated as follows. Let X be the set of all solutions, E be the set of uncertainties of measurement in single time frame, and M be the set of uncertainties for state transition among time frames, performance of a solution $x \in X$ under uncertainties $e \in E$ and $v \in M$ is $F(x, e, v)$. The problem is to find the solution that has the best worst-case performance, which is the same as minimizing (over all solutions) the maximum (over all uncertainties) performance:

$$\min_{x \in X} \max_{e \in E, v \in M} F(x, e, v) \tag{38}$$

from the description of Eqs. (34) and (37), the estimation of activity distribution $x(t)$ at time t is not only computed based on measurement $y(t)$, but also affected by previous estimations, so we define a linear combination of $x(t)$ as

$$z(t) = \mathscr{F}x(t) = g(x(k), H(k), v(k)) \quad k = 1, 2...t \tag{39}$$

so the measure of performance $F(x, e, v)$ is given by

$$J = \frac{\sum \|z(t) - \hat{z}(t)\|^2_{Q(t)}}{\|x(0) - \hat{x}(0)\|^2_{p_o^{-1}} + \sum (\|v(t)\|^2_{V(t)^{-1}} + \|e(t)\|^2_{N(t)^{-1}})} \tag{40}$$

where the notation $\|x\|^2_G$ is defined as the square of the weighted (by G) L_2 norm of x (*i.e.* $\|x\|^2_G = x^T Gx$). $N(t)$, $V(t)$, $Q(t)$ and p_o are the weighting matrices for the uncertainties of measurement in single time frame, the uncertainties of state transition among time frames, the estimation error, and the initial conditions respectively. $\hat{x}(0)$ is the initial estimate of the state. The optimal estimate $z(t)$ among all possible $\hat{z}(t)$ should satisfy:

$$\|J\|_\infty = sup\ J \ < \ \gamma^2 \tag{41}$$

where $\gamma^2 > 0$ is a prescribed level of disturbances. It is assumed that the L_2 norms of $e(t)$ and $v(t)$ exist. Then the minimax performance criterion of Eq. (40) where the estimator strategy $z(t)$ playing against the exogenous inputs $e(t)$, $v(t)$ and the initial state $x(0)$ becomes

$$\begin{aligned} \min_{z(t) - \hat{z}(t)} \max_{v, e, x(0)} J = \sum \|z(t) - \hat{z}(t)\|^2_{Q(t)} - \gamma^2 \|x(0) - \hat{x}(0)\|^2_{p_o^{-1}} \\ -\gamma^2 \sum (\|v(t)\|^2_{V(t)^{-1}} + \|e(t)\|^2_{N(t)^{-1}}) \end{aligned} \tag{42}$$

Now the problem becomes to solve the above objective function, and the kinect model (e.g. the decay model in this section) is successfully incorporated in it. Same solution framework as static PET image reconstruction can be used here.

4.3 Kinetic Parameter Estimation

From the section of compartment models, a two-tissue three-compartment model (Model2) is general enough to describe regional tracer kinetics as shown in Fig. 8, where C_P(pmol/ml) is arterial concentration of radiotracer, C_F and C_B (pmol/ml) are the concentrations of non–specific binding and specific binding tracers in tissues. Parameters k_1, k_2, k_3 and $k_4(min^{-1})$ specify radiotracer transport rates. The time variation of kinetic model in voxel i can be denoted by first-order differential equations as:

$$\frac{dC_{Fi}(t)}{dt} = k_{1i}(t)C_{Pi}(t) + k_{4i}(t)C_{Bi}(t) - (k_{2i}(t) + k_{3i}(t))C_{Fi}(t) \tag{43}$$

$$\frac{dC_{Bi}(t)}{dt} = k_{3i}(t)C_{Fi}(t) - k_{4i}(t)C_{Bi}(t) \tag{44}$$

Here this model will be used to derive the solutions framework of kinetic parameter estimation.

4.3.1 Modeling of Dynamic PET Measurement with Tracer Kinetics

Dynamic PET imaging involves a sequence of contiguous acquisition with different temporal resolutions, which can be formulated as a projection transform:

$$y(t) = Dx(t) + e(t) \tag{45}$$

Here, $y(t)$ is the projection data and $x(t) = \{x_i(t)|i = 1, \cdots, n\}^T$ is the activity concentration at time frame t. n is the total number of voxels. D is the system probability matrix. $e(t)$ is the overall measurement uncertainties. Here we will transform Eq. (45) to accommodate kinetic models. Firstly, activity concentration x will be the combination of C_F and C_B, then Eq. (45) will be

$$y(t) = \begin{bmatrix} D & D \end{bmatrix} \begin{bmatrix} C_F(t) \\ C_B(t) \end{bmatrix} + e(t) \tag{46}$$

where $C_F(t) = \{C_{Fi}(t)|i = 1, \cdots, n\}^T$ and $C_B(t) = [C_{Bi}(t)|i = 1, \cdots, n\}^T$. After the dynamic change of measurement $\frac{dy_i(t)}{dt}$ being deduced, we substitute the differential equations Eqs. (2) and (3) and do a simple transformation to arrange four kinetic parameters (k_1, k_2, k_3, k_4) in a column vector will yield

$$\frac{dy_i(t)}{dt} = \begin{bmatrix} D & D \end{bmatrix} \begin{bmatrix} \vdots \\ \frac{dC_{Fi}(t)}{dt} \\ \vdots \\ \frac{dC_{Bi}(t)}{dt} \\ \vdots \end{bmatrix} + e_i^{'}(t)$$

$$= \begin{bmatrix} D & D \end{bmatrix} \begin{bmatrix} \vdots & & & \\ C_{Pi}(t) & -C_{Fi}(t) & -C_{Fi}(t) & C_{Bi}(t) \\ \vdots & & & \\ 0 & 0 & C_{Fi}(t) & -C_{Bi}(t) \\ \vdots & & & \end{bmatrix} \begin{bmatrix} k_{1i}(t) \\ k_{2i}(t) \\ k_{3i}(t) \\ k_{4i}(t) \end{bmatrix} + e_i^{'}(t) \quad (47)$$

By denoting $R_i(t) = \begin{bmatrix} \vdots & & & \\ C_{Pi}(t) & -C_{Fi}(t) & -C_{Fi}(t) & C_{Bi}(t) \\ \vdots & & & \\ 0 & 0 & C_{Fi}(t) & -C_{Bi}(t) \\ \vdots & & & \end{bmatrix}$ and $S_i(t) = \begin{bmatrix} k_{1i}(t) \\ k_{2i}(t) \\ k_{3i}(t) \\ k_{4i}(t) \end{bmatrix}$,
we can get the dynamic change of total measurement data from all voxels as

$$\frac{dy(t)}{dt} = \sum_{i=1}^{n} \frac{dy_i(t)}{dt} = \begin{bmatrix} D & D \end{bmatrix} \begin{bmatrix} R_1(t) \cdots R_i(t) \cdots R_n(t) \end{bmatrix} \begin{bmatrix} S_1(t) \\ \vdots \\ S_i(t) \\ \vdots \\ S_n(t) \end{bmatrix} + e^{'}(t)$$

$$= \begin{bmatrix} D & D \end{bmatrix} R(t)S(t) + e^{'}(t) \quad (48)$$

Now we have set up the relationship between the change of measurement data and kinetic parameters directly by Eq. (48).

4.3.2 Solution Under Minimax Criterion

The minimax solution framework can also be transform to estimate kinetic parameter. No statistical assumptions needed makes the minimax criterion robust to the poor statistical properties in low count acquisition and system noises. Since kinetic parameters are generally assumed to be constant, we can set:

$$S(t+1) = S(t) + v(t) \tag{49}$$

Here, v is possible disturbances. With Eqs. (48) and (49), the corresponding minimax performance equation will be "$\min_{S\in L} \max_{e\in E, v\in V} F(S, e, v)$", where L, E and V are the sets of solutions, uncertainties of measurement and state transition. As an iterative solution, we also define a linear combination of $S(t)$ as "$z(t) = g(S(m), v(m))$ where $m = 1, 2...t$", then objective function J will be

$$J = \frac{\sum \|z(t) - \hat{z}(t)\|^2_{Q(t)}}{\|S(0) - \hat{S}(0)\|^2_{p_o^{-1}} + \sum (\|v(t)\|^2_{V(t)^{-1}} + \|e(t)\|^2_{E(t)^{-1}})} \tag{50}$$

where the notation $\|x\|^2_G$ is defined as the square of the weighted (by G) L_2 norm of x. p_o, $E(t)$, $V(t)$ and $Q(t)$ are weighting matrices. $\hat{S}(0)$ is the initialization of x. More detailed settings and initialization of parameters can be found in [80].

5 Clinical Practices and Future Directions

5.1 Personalized Drug Metabolism Analysis

Dynamic PET is a molecular imaging technique that is used to monitor the spatiotemporal distribution of a radiotracer in vivo and enables cellular level metabolism analysis in clinical routine. dPET provides a good promise for quantitative lesion metabolism analysis to help identify lesions. However, due to poor statistical properties of the measurement data in low count dynamic PET acquisition and disturbances from surrounding tissues, identifying small lesions inside the human body is still a challenging issue. Furthermore, the mismatch between general purpose models and patient size/motions makes the situation even worse. Quantitative kinetic analysis of radiotracer uptakes requires the reconstruction of kinetic parameters [25, 77, 96]. The mainstream is statistical reconstruction algorithms, however, whose quality is determined by the accuracy of sophisticated system probability matrix (SM). Many efforts have been devoted to improve the accuracy of SM [64, 70, 85, 107]. However, the ideal SM is almost impossible to obtain under practical conditions. The general purpose SM also could not compensate different sizes of patients and the motions during acquisition, which will decrease the accuracy of reconstructions. Furthermore, the reconstruction of dynamic PET image sequences, whose poor temporal resolution, insufficient photon counts, more complicated data corrections and poor statistical properties of measurement data also requires a more accurate SM.

To improve the personalized lesion metabolism analysis, we can generate a patient adaptive SM using machine learning techniques. Both patient size information and potential small lesion information are incorporated by simulations of phantoms of different sizes and individual point source responses [50, 53, 102]. Training data

set from simulations of 90 studies is conducted using 15 phantoms of different sizes based on Zubal thorax phantom. Each simulation has randomly generated motions and lesions in lung. The personalized SM should be able to differentiate true lesion and false lesion, and further deal with input functions of different accuracies.

5.1.1 System Matrix Derived from Supervised Learning

Statistical reconstruction requires a well modeled SM, which directly determines the accuracy of reconstruction results. The SM D is extended to include 2 parts, D_1 is a SM generated from geometry information and physical phenomena, and will account for sizes and motions of different patients, D_2 is an additional SM generated from point source responses. D_1 and D_2 are full size SM, and combined together by weighting matrices w_1 and w_2 according to the anatomical information of patients. This effort makes the SM more patient adaptive. The measurement equation is extended to be

$$Y(t) = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} X(\kappa, t) + e(t) \tag{51}$$

w_1, w_2, D_1, D_2 are updated by supervised learning. Training sets are provided by Monte Carlo simulations using GATE toolbox [41]. Correspondingly, two series of simulations are performed, one is performed with human thorax phantom of different sizes, and the other is done by point source response inside a thorax phantom of normal size. Denoting the activity concentrations as $X = \{x_1, x_2, \cdots x_n\}$ and measurement datasets as $Y = \{y_1, y_2, \cdots x_n\}$. n is the number of training sets, and every dataset is a dynamic data sequence related to time t. For simplification of expression, the PET measurement equation is written as $Y(t) = D'X(k, t) + e(t)$. Since ADALINE has been proved to be simple yet successful for updating SM in [85], we also adopt ADALINE for our SM training here. The initialization of D_1 and D_2 are the SMs generated with uniform cylindrical phantom. The update procedure by ADALINE using back-propagation and least mean square error is:

$$\hat{y}_m(t) = D'_m X(k, t) + e_m(t) \qquad \delta_k(t) = Y(t) - \hat{y}_m(t) \tag{52}$$

$$D'_{m+1}(t) = D'_m(t) + 2L\delta_m X^T(t) \qquad e_{m+1}(t) = e_m(t) + 2L\delta_m(t) \tag{53}$$

where m is the iteration step of training, and L is the learning rate. After defining a precision level of learning ε,

$$D' \quad \textbf{subject to} \quad \begin{cases} Y(t) - \hat{y}_m(t) < \varepsilon \\ \hat{y}_m(t) - Y(t) < \varepsilon \end{cases} \tag{54}$$

the weighting matrices w_1 and w_2 will be obtained when convergence is achieved (Fig. 11).

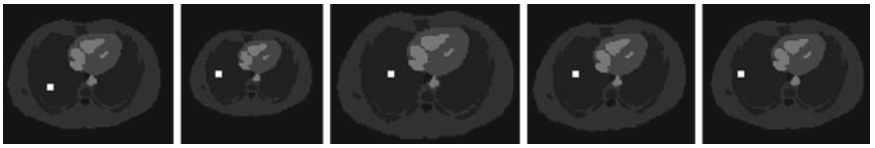


Fig. 11 Demonstration of five phantoms from the phantom library used to generate personalized SM

5.1.2 Parameter Reconstruction of Dynamic PET

The kinetic model and image reconstruction are combined in one equation, the log likelihood function can be derived with measurement data $y(t)$ as

$$\mathscr{L}(y|\kappa) = \sum_t y(t) \log \bar{y}(\kappa, t) - \bar{y}(\kappa, t) \tag{55}$$

$$\textbf{where} \quad \bar{y}(\kappa, t) = D' x(\kappa, t) + e(t) \tag{56}$$

$$\hat{\kappa} = \arg\max \Phi(\kappa) \tag{57}$$

$$\Phi(\kappa) = \mathscr{L}(y|k) - \beta \mathscr{U}(k) \tag{58}$$

where $\mathscr{U}$ is the penalty regularization term with parameter β controlling resolution/noise tradeoff. Equation (57) can be solved by a paraboloidal surrogates algorithm in [22]. The above solution is global convergent, however, the kinetic parameter reconstruction has a higher data dimensionality/freedom, so we also define the evaluation of kinetic parameters through images by using student's t-distribution hypothesis test to determine their statistical differences among iterations. By selecting Region of Interest (ROI), calculate

$$t = \frac{|\bar{x}_m - \bar{x}_{m+1}|}{\sigma} \tag{59}$$

where

$$\sigma = \left(\frac{var_m + var_{m+1} - 2cov_{m,m+1}}{N} \right)^{0.5} \tag{60}$$

and

$$cov_{m,m+1} = \frac{1}{N-1} \sum_{i=1}^{N} (x_{m,i} - \bar{x}_m)(x_{m+1,i} - \bar{x}_{m+1}) \tag{61}$$

$\bar{x}_m$ and $\bar{x}_{m+1}$ are the means in ROI at iteration m and $m+1$, var is the corresponding variances across the image elements. cov is the covariance across the two iterations. t is calculated until less than $t_{0.05}$ in the t-table to show a confidence level of 95 % that the difference between images is small enough.

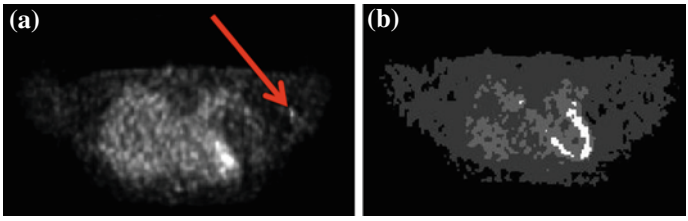


Fig. 12 **a** Lesion in 40th slice; **b** 32nd slice as reference

Table 5 Summary of experiments

	Group1	Group2	Group3	Group4	Group5	Group6
Lesion type	True	True	True	False	False	False
Input function	1	2	3	1	2	3
Successful estimation/Total studies						
Generic SM	13/15	7/15	5/15	11/15	7/15	6/15
Personalized SM	14/15	13/15	7/15	12/15	12/15	7/15

5.1.3 Clinical Results

The clinical patient data in this study was a dynamic PET scan acquired from a 28-year-old, 75 kg male volunteer. 10 mCi ^{18}F-FDG was injected and a dynamic acquisition of the thoracic cavity started just after injection. The acquisition consists of 40 time frames: 20×0.5 min, 15×1 min, and 5×2 min. All corrections are performed properly with the software provided by the scanner. The input function is estimated by the image-derived method. Figure 12a shows a lesion region by a red arrow in the 40th slice. We calculate the influx rates of the lesion and compare them with the heart muscles in the 32nd slice. The lesion metabolism calculated by the new personalized SM is closer to the muscles than that by generic SM, and the lesion is confirmed by the doctor as a false lesion with temporarily increased metabolism than muscles. Results from the new personalized SM show potential improvements in diagnosis. Table 5 shows the statistical results from all 90 studies, which shows the ratio of successful identification of lesions using different input functions. Input Function 1 is perfect input function (equivalent to perfect blood sampling with less than 5 % error), Input Function 2 is good input function (equivalent to a disturbed blood sampling with about 20 % error), Input Function 3 is an Image Derived Input Function (IDIF). Both methods performs well by using Input Function 1. However, with Input Function 2 (with disturbances), the accuracy of generic SM decreases, but the personalized SM still provides good results. The results show the improvement in identifying lesions by personalized SM, and the reduction of requirement of accurate input function.

Table 6 Influx rate of patient study

Muscles	Generic SM	Personalized SM
0.0054	0.0085	0.0071

5.2 Model Selection

In drug discovery and development, the procedure of drug selection is full of challenging issues. Kelloff et al. show that more than 90 % of all new oncology drugs fail in the late stages of development because of inadequate activity and difficulties in determining their efficacy [43]. Quantitative pharmacokinetic analysis with dynamic PET imaging now plays a promising role as determinants of in vivo drug action to help select drug candidate. Fast and accurate pharmacokinetic analysis with rapid information feedback in the early stage of drug discovery and development is critical to obtain the in vitro and in vivo drug properties [14, 100].

A typical procedure of pharmacokinetic analysis by dPET imaging includes, firstly, setting up a working hypothesis of the target enzyme or receptor for a particular disease, secondly, establishing suitable models (or surrogate markers) to test biological activities, and at last, screening the new drug molecules for biological activities. In this procedure, model selection by dPET has seldom been studied because of various scientific challenges, for example, (1) the kinetic models for drugs are generally very complicated, when facing a new biomarker (new drug), it is hard to determine which model will work best, (2) accurately solving these complicated models always needs special mathematical considerations, (3) although we can always use more complicated models to represent certain biological activity, the computational cost increase significantly due to the complex of the model, which cause a burden for the early drug discovery. (4) measurement data from dPET suffers from poor spatial and temporal resolutions, especially the first several time frames, (5) blood sampling is required in pharmacokinetic analysis but it is very hard to generate an accurate one [31].

5.2.1 Temporal-Difference Reinforcement Learning

Machine learning in image processing and analysis is growing rapidly [97]. Of various machine learning methods, reinforcement learning is meant to be a general approach to learn from interactions [86]. It is a control method which presents a robust mechanism for goal directed decision making. Unlike supervised learning methods, no examples of desired behaviors are provided during training, instead, behavior is guided through positive or negative reinforcements [99]. So this method do not require a large training dataset, and is especially suitable for preclinical drug selection and pharmacokinetic analysis with only limited data sets [84]. Additionally, as a control mechanism, the method can help solve the complicated kinetic models with noisy dPET acquisition data. At the same time, the method can inherently deal with disturbances during blood sampling. Therefore, in this section we will intro-

duce a reinforcement learning based method which combines model selection and parameter estimation for pharmacokinetic analysis by dPET.

Temporal Difference reinforcement learning (TD Learning) is a combination of Monte Carlo ideas and dynamic programing [86, 99]. Therefore, like Monte Carlo methods, TD Learning can learn from raw experiences without pre-defined models, and like dynamic programing, TD learning can update its estimations based on a part of learning outcomes rather than the final outcome. These features are especially suitable for model selection and noisy dPET data. Regardless of model, when we have an initial K, we define an action set a, which contains $2n$ components, $\{a_1^+, a_1^-, a_2^+, a_2^-, \cdots a_n^+, a_n^-, \}$, with the subscript corresponds to the index of kinetic parameters, and the superscript represents increasing (+) or decreasing (−) that kinetic parameter by certain amount during estimation. We derive the TD Learning algorithm for model selection and parameter estimation from the classic one as shown in $Algorithm$ 1. Details will be shown in next subsection.

5.2.2 Model Selection

As shown in the algorithm, reinforcement learning acts according to the rewards, we define three rewards based on physical constraints for model selection. The combination of three rewards is able to exclude non-matching models fast, which can improve the computational efficiency, and reduce the disturbances from the noises in low count dPET data. By denoting K' as the estimated kinetic parameters after selecting an action from a,

1. Reward 1: We compare the measured total counts in each time frame of measurement data Y and estimated total counts with K'.

$$MSE(TotalCounts(Y(t)), TotalCounts(DX(K', t))) < Threshold1 \tag{62}$$

2. Reward 2: We compare the first order difference of Time Activity Curve (TAC) from measurement data Y and TAC curve estimated with K'.

$$MSE(Difference(TAC(Y)), Difference(TAC(DX(K', t))) < Threshold2 \tag{63}$$

3. Reward 3: This is an optional reward, if there is priori knowledge from clinical data available, they are learned together through a 2-hidden layer neural network (NN) as shown in Sect. 5.1.1, and then used as a reference for estimated K'.

$$MSE(NN(Y), K') < Threshold3 \tag{64}$$

Where MSE is the operation to calculate the mean squared error. When each reward criterion is met, we have a reward ($rew = +1$), otherwise, ($rew = -1$). Then we accumulate all rewards by eligible trace Q-Learning [86] as shown in the $Algorithm 1$

$$Q(K,a) \leftarrow Q(K,a) + \alpha[rew_{m+1} + \gamma Q(K_{m+1}, a_{m+1}) - Q(K,a)] \tag{65}$$

Where α, γ are learning parameters, which control the width and depth of learning. Proofs in [86] show that α and γ mostly affect the convergence speed, and have only limited effect on learning accuracy after convergency. Q is the value function in the Q-Learning, which stores all the rewards, and m represents iteration steps.

Then the algorithm is applied to every model in the model bank, with each estimated K' from the maximum in Q, we calculate the $Bias$ between estimated TAC (TAC_e) and true TAC (TAC_m) from measurement data for each model by $Bias = \frac{1}{T}\sum \frac{\|TAC_m - TAC_e\|}{TAC_m}$, where T is number of time frames. The model with the lowest $Bias$ in the model bank with be the selected model by the proposed method.

5.2.3 Parameter Estimation

When using **Algorithm 1** to choose model, simultaneously calculated K' will be the initial parameter for that model. And the kinetic parameter can also be further calculated with a refined action set a_{ref} containing smaller increasing or decreasing amount using $Algorithm$ 1.

Algorithm 1 Model Selection and Parameter Estimation by TD Learning

```
Initialize Q(K,a) arbitrarily
Initialize K
Repeat until convergency
    Randomly choose one action from a
    Repeat for all steps
        Take action a, generate K′, observe reward rew using defined criteria
        Choose a′ from K′ derived from Q
        Accumulate all rewards using Eq. 65
        K ← K′
    End
End
Select the Maximum in Q, Corresponding K is the estimated K
```

5.2.4 Clinical Results

We study three cases of real patient dPET scans. Figure 13 shows the three cases, the first case is the scan of patient thorax, a ROI is defined in the normal muscle region,

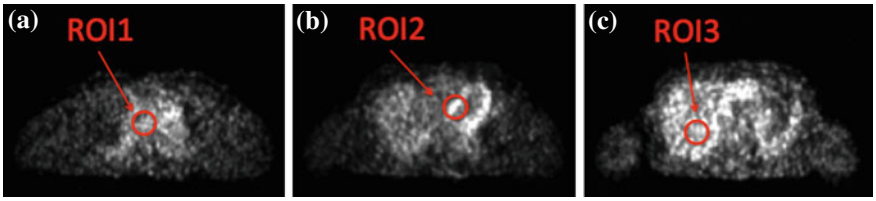


Fig. 13 **a** Case 1; **b** Case 2; **c** Case 3

Table 7 Model selection results

	Model1	Model2	Model3	Model4	Model5	Model6
Case 1	1.5901	1.4248	1.1735	1.6549	1.5558	NA
Case 2	0.7963	NA	NA	0.9482	NA	NA
Case 3	1.3809	1.0721	1.6025	1.2169	NA	NA

NA is "Not Applicable"

Table 8 Estimated kinetic parameters for three cases

	K1	K2	K3	K4	K5	K6
Case 1	0.0960	0.1540	0.0375	NA	NA	NA
Case 2	0.1080	0.0838	NA	NA	NA	NA
Case 3	0.0400	0.0620	0.0145	0.0015	NA	NA

NA is "Not Applicable"

the second case studies a ROI in the heart region, and the third case is with a ROI in the liver. The dynamic PET scans are performed on our PET scanner, the dynamic data set consists of 40 time frames: 20×0.5 min, 15×1 min and 5×2 min. The input function is estimated by fitting the reconstructed dynamic images. This input function is equal to a disturbed one affected by noises in reconstructed images. The model bank used is the compartment models shown in Fig. 8. The model selection results is shown Table 7. The results of model selections are consistent with suggestions from clinical studies. ROI 1 is normal tissue and Model3 is mostly used. ROI 2 is near the left ventricular and highly affected by the blood flow, so the blood flow model (Model 1) is most suitable. For ROI 3, clinical results had shown the necessity and importance of estimation of K4 in liver cancer, and the proposed method correctly choose the right Model2 (Phelps 4K model). And the non-applicable models are excluded successfully. The estimated kinetic parameters are shown in Table 8.

5.3 Future Directions

In this chapter, we review the shape analysis and application in molecular imaging. Molecular imaging is a relatively new but fast-developing area for both research and

clinical applications. Although with some limitations, molecular imaging modalities especially PET show the superior ability to quantitatively measure the biologic processes in a functional way at the cellular and subcellular level within living subjects. All the data processing and analysis of molecular imaging depend on the physical natures of the molecular imaging modalities. The emerging new scanner systems with new detectors will further enhance their abilities, and bring new challenges in data correction and image analysis at the same time. The data corrections algorithms need to be adjusted with the properties of new system design, and new features in detector system correspondingly. Monte Carlo simulation is always a good way to study the new design and provide references for validation of new data correction methods. And image reconstruction with disease specified statistical model instead of the generic models can improve the image qualities of certain disease. With the advancement of data correction and image reconstruction methods, the image post-processing including image registration and image fusion must adopt related changes. Researchers are also actively using machine learning methods to extract applicable pathological models from a series of patient studies and apply the strategy to personalized treatment. Pharmaceutical companies are also interested in the accurate quantitative pharmacokinetic parameter estimation using PET to study the metabolism of new drugs, which has the potential to shorten the drug development cycle and save tons of money for the industry and patients. With the evolution of both image pre-processing and post-processing methods, molecular imaging is believed to be able to study more complicated diseases currently in the unknown area, for example, the origins of Alzheimer's disease and dementia.

References

1. Bailey D, Townsend D, Valk P, Maisey M (2005) Positron emission tomography: basic sciences. Springer, London
2. Bailey DL, Meikle SR (1994) A convolution-subtraction scatter correction method for 3D PET. Phys Med Biol 39(3):411
3. Bao Q, Newport D, Chen M, Stout DB, Chatziioannou AF (2009) Performance evaluation of the inveon dedicated PET preclinical tomograph based on the NEMA NU-4 standards. J Nucl Med 50:401–408
4. Bartenstein P, Minoshima S, Hirsch C, Buch K, Willoch F, Mösch D, Schad D, Schwaiger M, Kurz A et al (1997) Quantitative assessment of cerebral blood flow in patients with Alzheimer's disease by SPECT. J Nucl Med 38(7):1095
5. Bateman TM, Heller GV, McGhie AI, Friedman JD, Case JA, Bryngelson JR, Hertenstein GK, Moutray KL, Reid K, Cullom SJ (2006) Diagnostic accuracy of rest/stress ECG-gated Rb-82 myocardial perfusion PET: comparison with ECG-gated Tc-99m sestamibi SPECT. J Nucl Cardiol 13(1):24–33
6. Bendriem B, Townsend DW (1998) Theor Pract 3D PET. Kluwer Academic/Plenum Publishers, New York
7. Bentourkia M, Msaki P, Cadorette J, Lecomte R (1995) Assessment of scatter components in high-resolution PET: correction by nonstationary convolution subtraction. J Nucl Med 36(1):121–130

8. Beyer T, Tellmann L, Nickel I, Pietrzyk U (2005) On the use of positioning aids to reduce misregistration in the head and neck in whole-body PET/CT studies. J Nucl Med 46(4):596–602
9. Bonte FJ, Weiner MF, Bigio EH, White C et al (1997) Brain blood flow in the dementias: SPECT with histopathologic correlation in 54 patients. Radiology 202(3):793–797
10. Boucher L, Rodrigue S, Lecomte R, Bénard F (2004) Respiratory gating for 3-dimensional PET of the thorax: feasibility and initial results. J Nucl Med 45(2):214–219
11. Brankov JG, Galatsanos NP, Yang Y, Wernick MN (2003) Segmentation of dynamic PET or fMRI images based on a similarity metric. IEEE Trans Nucl Sci 50(5):1410–1414
12. Buck AK, Halter G, Schirrmeister H, Kotzerke J, Wurziger I, Glatting G, Mattfeldt T, Neumaier B, Reske SN, Hetzel M (2003) Imaging proliferation in lung tumors with PET: 18F-FLT Versus 18F-FDG. J Nucl Med 44(9):1426–1431
13. Bushong SC (1988) Magnetic resonance imaging. CV Mosby Company, St. Louis
14. Catafau M, Bullich S (2013) Molecular imaging PET and SPECT approaches for improving productivity of antipsychotic drug discovery and development. Curr Med Chem 20(3):378–388
15. Cobelli C, Foster D, Toffolo G (2000) Tracer kinetics in biomedical research: from data to model. Kluwer Academic/Plenum Publishers, New York
16. Crum W, Hartkens T, Hill D (2004) Non-rigid image registration: theory and practice. Br J Radiol 77(Suppl 2):S140–S153
17. Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Grégoire V (2003) Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. Radiother Oncol 69(3):247–250
18. Defrise M, Casey ME, Michel C, Conti M (2005) Fourier rebinning of time-of-flight PET data. Phys Med Biol 50(12):2749
19. Drever L, Roa W, McEwan A, Robinson D (2007) Iterative threshold segmentation for PET target volume delineation. Med Phys 34:1253
20. Erd, YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn, R, Humm JL (1997), Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer 80(S12) 2505–2509.
21. Fenster A, Downey DB (1996) 3-D ultrasound imaging: a review. IEEE Eng Med Biol Mag 15(6):41–51
22. Fessler J Erdogan H (1998) A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction. In: Nuclear science symposium, 1998. Conference record. 1998 IEEE, vol 2, pp 1132–1135.
23. Freedman N, Schechter D, Klein M, Marciano R, Rozenman Y, Chisin R (2000) SPECT attenuation artifacts in normal and overweight persons: insights from a retrospective comparison of Rb-82 positron emission tomography and Tl-201 SPECT myocardial perfusion imaging. Clin Nucl Med 25(12):1019–1023
24. Gao F, Liu H, Jian Y, Shi P (2009) Dynamic dual-tracer PET reconstruction. In: Information processing in medical imaging (IPMI 2009). LNCS, Vol 5636. Springer, Berlin, Heidelberg, pp 38–49
25. Gao F, Liu H, Shi P (2011) Robust estimation of kinetic parameters in dynamic PET imaging. In: Medical image computing and computer-assisted intervention–MICCAI 2011, vol 6891. pp 492–499.
26. Geets X, Lee JA, Bol A, Lonneux M, Grégoire V (2007) A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur J Nucl Med Mol Imaging 34(9):1427–1438
27. Goerres GW, Kamel E, Heidelberg TNH, Schwitter MR, Burger C, von Schulthess GK (2002) PET-CT image co-registration in the thorax: influence of respiration. Eur J Nucl Med Mol Imaging 29(3):351–360
28. Gregory R, Partridge M, Flower M (2006) Performance evaluation of the Philips Gemini PET/CT system. IEEE Trans Nucl Sci 53(1):93–101

29. Gunn R, Gunn S, Turkheimer F, Aston J, Cunningham V (2002) Tracer kinetic modeling via basis pursuit. In: Senda M, Kimura Y, Herscovitch P (eds) Brain Imaging using PET. Academic Press, San Diego
30. Gunn RN, Gunn SR, Cunningham VJ (2001) Positron emission tomography compartmental models. J Cereb Blood Flow Metab 21(6):635–652
31. Gunn RN, Gunn SR, Turkheimer FE, Aston JA, Cunningham VJ (2002) Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling. J Cereb Blood Flow and Metab 22(12):1425–1439
32. Hajnal JV (2001) Medical image registration. CRC Press LLC, New York
33. Han X, Xu C, Prince JL (2003) A topology preserving level set method for geometric deformable models. IEEE Trans Pattern Anal Mach Intell 25(6):755–768
34. Hara N, Okuizumi M, Koike H, Kawaguchi M, Bilim V (2005) Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modality for the precise detection and staging of early prostate cancer. The Prostate 62(2):140–147
35. Hu Z, Shi P (2010) Sensitivity analysis for biomedical models. IEEE Trans Med Imaging 29(11):1870–1881
36. Huang X, Hill NA, Ren J, Guiraudon G, Boughner D, Peters TM (2005) Dynamic 3D ultrasound and MR image registration of the beating heart. In: Medical image computing and computer-assisted intervention-MICCAI 2005. Springer, pp 171–178.
37. Huettel SA, Song AW, McCarthy G (2004) Functional magnetic resonance imaging. Sinauer Associates, Sunderland
38. Ibaraki M, Sugawara S, Nakamura K, Kinoshita F, Kinoshita T (2011) The effect of activity outside the field-of-view on image signal-to-noise ratio for 3D PET with 15O. Phys Med Biol 56(10):3061
39. Iznaga-Escobar N (1998) ^{188}Re-direct labeling of monoclonal antibodies for radioimmunotherapy of solid tumors: biodistribution, normal organ dosimetry, and toxicology. Nuc Med Biol 25(5):441–447
40. Jakoby B, Bercier Y, Watson C, Bendriem B, Townsend D (2009) Performance characteristics of a new LSO PET/CT scanner with extended axial field-of-view and PSF reconstruction. IEEE Trans Nucl Sci 56(3):633–639
41. Jan S, Santin G, Strul D, Staelens S, Assie K, Autret D, Avner S, Barbier R, Bardies M, Bloomfield P et al (2004) GATE: a simulation toolkit for PET and SPECT. Phys Med Biol 49(19):4543
42. Jentzen W, Freudenberg L, Eising EG, Heinze M, Brandau W, Bockisch A (2007) Segmentation of PET volumes by iterative image thresholding. J Nucl Med 48(1):108–114
43. Kelloff GJ, Sigman CC (2012) Cancer biomarkers: selecting the right drug for the right patient. Nat Rev Drug Discov 11(3):201–214
44. Kelly CJ, Brady M (2006) A model to simulate tumour oxygenation and dynamic [18F]-Fmiso PET data. Phys Med Biol 51(22):5859
45. Kitamura K, Ishikawa A, Mizuta T, Yamaya T, Yoshida E, Murayama H (2007) Detector normalization and scatter correction for the jPET-D4: a 4-layer depth-of-interaction PET scanner. Nuc Instrum Methods Phys Res Sect A Accelerators Spectrom Detect Assoc Equip 571(1–2):231–234
46. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP et al (2010) Elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 29(1):196–205
47. Koeppe RA, Gilman S, Junck L, Wernette K, Frey KA et al (2008) Differentiating Alzheimer's disease from dementia with Lewy bodies and Parkinson's disease with (+)-[11C] dihydrotetrabenazine positron emission tomography. Alzheimer's and dementia J Alzheimer's Assoc 4(Suppl 1):S67
48. Krivokapich J, Smith G, Huang SC, Hoffman E, Ratib O, Phelps M, Schelbert H (1989) 13N ammonia myocardial imaging at rest and with exercise in normal volunteers. quantification of absolute myocardial perfusion with dynamic positron emission tomography. Circulation 80(5):1328–1337

49. Kwakernaak H (1993) Robust control and H_∞-optimization-tutorial paper. Automatica 29(2):255–273
50. Laffon E, de Clermont H, Vernejoux JM, Jougon J, Marthan R (2011) Feasibility of assessing [18F]FDG lung metabolism with late dynamic PET imaging. Mol Imaging Biol 13(2):378–384
51. Leff DR, Warren OJ, Enfield LC, Gibson A, Athanasiou T, Patten DK, Hebden J, Yang GZ, Darzi A (2008) Diffuse optical imaging of the healthy and diseased breast: a systematic review. Breast Cancer Res Treat 108(1):9–22
52. Lewellen T (2008) Recent developments in PET detector technology. Phys Med Biol 53:R287
53. Li Z, Li Q, Yu X, Conti P, Leahy R (2009) Lesion detection in dynamic FDG-PET using matched subspace detection. IEEE Trans Med Imaging 28(2):230–240
54. Lubberink M, Kosugi T, Schneider H, Ohba H, Bergström M (2004) Non-stationary convolution subtraction scatter correction with a dual-exponential scatter kernel for the Hamamatsu SHR-7700 animal PET scanner. Phys Med Biol 49(5):833
55. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P (1997) Multimodality image registration by maximization of mutual information. IEEE Trans Med Imaging 16(2):187–198
56. Magota K, Kubo N, Kuge Y, Nishijima KI, Zhao S, Tamaki N (2011) Performance characterization of the Inveon preclinical small-animal PET/SPECT/CT system for multimodality imaging. Eur J Nuc Med Mol Imaging 38(4):742–752
57. Malladi R, Sethian JA, Vemuri BC (1995) Shape modeling with front propagation: a level set approach. IEEE Trans Pattern Anal Mach Intell 17(2):158–175
58. Mankoff DA (2007) A definition of molecular imaging. J Nucl Med 48(6):18N–21N
59. Maruyama A, Hasegawa S, Paul AK, Xiuli M, Yoshioka J, Maruyama K, Hori M, Nishimura T (2002) Myocardial viability assessment with gated SPECT Tc-99m tetrofosmin wall thickening: comparison with F-18 FDG-PET. Ann Nucl Med 16(1):25–32
60. Mattes D, Haynor DR, Vesselle H, Lewellen TK, Eubank W (2003) PET-CT image registration in the chest using free-form deformations. IEEE Trans Med Imaging 22(1):120–128
61. Kamasak ME, Bouman CA, Morris ED, Sauer K (2005) Direct reconstruction of kinetic parameter images from dynamic PET data. IEEE Trans Med Imaging 25:636–650
62. Mena E, Turkbey B, Mani H, Adler S, Valera VA, Bernardo M, Shah V, Pohida T, McKinney Y, Kwarteng G, Daar D, Lindenberg ML, Eclarinal P, Wade R, Linehan WM, Merino MJ, Pinto PA, Choyke PL, Kurdziel KA (2012) 11C-Acetate PET/CT in localized prostate cancer: a study with MRI and histopathologic correlation. J Nucl Med 53(4):538–545
63. Michel C, Sibomana M, Boi A, Bernard X, Lonneux M, Defrise M, Comtat C, Kinahan P, Townsend D (1998) Preserving poisson characteristics of PET data with weighted OSEM reconstruction. In: Nuclear science symposium, 1998. Conference record. 1998 IEEE, vol. 2. pp. 1323–1329
64. Moehrs S, Defrise M, Belcari N, Guerra AD, Bartoli A, Fabbri S, Zanetti G (2008) Multi-ray-based system matrix generation for 3D PET reconstruction. Phys Med Biol 53(23):6925
65. Naidoo-Variawa S, Lehnert W, Banati RB, Meikle SR (2011) Scatter correction for large non-human primate brain imaging using microPET. Phys Med Biol 56(7):2131
66. Nichols TE, Qi J, Asma E, Leahy RM (2002) Spatiotemporal reconstruction of list-mode PET data. IEEE Trans Med Imaging 21(4):396–404
67. Ollinger JM (1996) Model-based scatter correction for fully 3D PET. Phys Med Biol 41:153–176
68. Osher S, Fedkiw RP (2001) Level set methods: an overview and some recent results. J Comput Phys 169(2):463–502
69. Pajares G, Manuel de la Cruz J (2004) A wavelet-based image fusion tutorial. Pattern Recogn 37(9):1855–1872
70. Panin V, Kehren F, Rothfuss H, Hu D, Michel C, Casey M (2006) PET reconstruction with system matrix derived from point source measurements. IEEE Trans Nucl Sci 53(1):152–159
71. Parker BJ, Feng D (2005) Graph-based mumford-shah segmentation of dynamic PET with application to input function estimation. IEEE Trans Nucl Sci 52(1):79–89

72. Parra L, Barrett HH (1998) List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET. IEEE Trans Med Imaging 17(2):228–235
73. Pellizzari C, Levin DN, Chen GT, Chen CT (1993) Image registration based on anatomic surface matching. In Maciunas RJ (ed) Interactive image-guided neurosurgery. AANS, Park Ridge, pp 47–62
74. Pichler BJ, Kolb A, Nagele T, Schlemmer HP (2010) PET/MRI: paving the way for the next generation of clinical multimodality imaging applications. J Nucl Med 51(3):333–336
75. Pietrzyk U, KarlHerholz G, AndreasJacobs R (1994) Image registration: Validation for PET, SPECT, MRI and CT brain studies. J Nucl Med 35(12):2011–2018
76. Rahmim A, Rousset O, Zaidi H (2007) Strategies for motion tracking and correction in PET. PET Clin 2(2):251–266
77. Rahmim A, Tang J, Zaidi H (2009) Four-dimensional (4D) image reconstruction strategies in dynamic PET: beyond conventional independent frame reconstruction. Med Phys 36(8):3654
78. Schmitt JM (1999) Optical coherence tomography (OCT): a review. IEEE J Sel Top Quantum Electron 5(4):1205–1215
79. Shen X, Deng L (1997) Game theory approach to discrete H_∞ filter design. IEEE Trans Sig Process 45:1092–1095
80. Simon D (2006) Optimal state estimation: Kalman, H_∞ and nonlinear approaches. Wiley, Hoboken, New Jersey
81. Sossi V, Barney J, Harrison R, Ruth T (1995) Effect of scatter from radioactivity outside of the field of view in 3D PET. IEEE Trans Nucl Sci 42(4):1157–1161
82. Spinks TJ, Miller MP, Bailey DL, Bloomfield PM, Livieratos L, Jones T (1998) The effect of activity outside the direct field of view in a 3D-only whole-body positron tomograph. Phys Med Biol 43(4):895
83. Meikle SR, Matthews JC, Cunningham VJ, Bailey DL, Livieratos L, Jones T, Price P (1998) Parametric image reconstruction using spectral analysis of PET projection data. Phyis Med Biol 43 651–666
84. Strauss LG, Pan L, Cheng C, Haberkorn U, Dimitrakopoulou-Strauss A (2011) Shortened acquisition protocols for the quantitative assessment of the 2-tissue-compartment model using dynamic PET/CT 18F-FDG studies. J Nucl Med 52(3):379–385
85. Su KH, Wu LC, Lee JS, Liu RS, Chen JC (2009) A novel method to improve image quality for 2-D small animal PET reconstruction by correcting a monte carlo-simulated system matrix using an artificial neural network. IEEE Trans Nucl Sci 56(3):704–714
86. Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge University Press, Cambridge
87. Teras M, Tolvanen T, Johansson J, Williams J, Knuuti J (2007) Performance of the new generation of whole-body PET/CT scanners: Discovery STE and Discovery VCT. Eur J Nucl Med Mol Imaging 34(10):1683–1692
88. Tharp K, Israel O, Hausmann J, Bettman L, Martin W, Daitzchman M, Sandler M, Delbeke D (2004) Impact of 131I-SPECT/CT images obtained with an integrated system in the follow-up of patients with thyroid carcinoma. Eur J Nucl Med Mol Imaging 31(10):1435–1442
89. Townsend DW, Beyer T (2002) A combined PET/CT scanner: the path to true image fusion. Br J Radiol 75(suppl 9):S24–S30
90. Townsend DW, Carney JP, Yap JT, Hall NC (2004) PET/CT today and tomorrow. J Nucl Med 45(1 suppl):4S–14S
91. Tsoumpas C, Turkheimer F, Thielemans K (2008) A survey of approaches for direct parametric image reconstruction in emission tomography. Med Phys 35:3963
92. Valk P, Delbeke D, Bailey D, Townsend D, Maisey M (2006) Positron emission tomography: clinical practice. Springer, London
93. Villemagne VL, Rothman RB, Yokoi F, Rice KC, Matecka D, Dannals RF, Wong DF (1999) Doses of GBR12909 that suppress cocaine self-administration in non-human primates substantially occupy dopamine transporters as measured by [11C] WIN35, 428 PET scans. Synapse 32(1):44–50

94. Wahl RL, Buchanan JW (2002) Principles and practice of positron emission tomography. Lippincott Williams and Wilkins, Philadelphia
95. Wahl RL, Quint LE, Cieslak RD, Aisen AM, Koeppe RA, Meyer CR et al (1993) "Anatometabolic" tumor imaging: fusion of FDG PET with CT or MRI to localize foci of increased activity. J Nucl Med 34(7):1190
96. Wang G, Qi J (2009) Generalized algorithms for direct reconstruction of parametric images from dynamic PET data. IEEE Trans Med Imaging 28(11):1717–1726
97. Wang S, Summers R (2012) Machine learning and radiology. Med Image Anal 16:933–951
98. Watson C (2000) New, faster, image-based scatter correction for 3D PET. IEEE Trans Nucl Sci 47:1587–1594
99. Wiering M, van Otterlo M (2012) Reinforcement learning: state-of-the-art, vol 12. Springer, London
100. Willmann JK, Van Bruggen N, Dinkelborg LM, Gambhir SS (2008) Molecular imaging in drug development. Nat Rev Drug Discov 7(7):591–607
101. Wong KP, Feng D, Meikle SR, Fulham MJ (2002) Segmentation of dynamic PET images using cluster analysis. IEEE Trans Nucl Sci 49(1):200–207
102. Wu H, Pal DO, Sullivan J, Tai YC (2008) A feasibility study of a prototype PET insert device to convert a general-purpose animal PET scanner to higher resolution. J Nucl Med 49(1):79–87
103. Yan, J, Planeta-Wilson B, Gallezot J, Carson R (2010) Initial evaluation of direct 4D parametric reconstruction with human PET data. In: Nuclear science symposium conference record (NSS/MIC), 2009 IEEE. pp 2503–2506
104. Yang DJ, Azhdarinia A, Kim EE (2005) Tumor specific imaging using Tc-99m and Ga-68 labeled radiopharmaceuticals. Curr Med Imaging Rev 1(1):25–34
105. Yang J, Staib LH, Duncan JS (2004) Neighbor-constrained segmentation with level set based 3-D deformable models. IEEE Trans Med Imaging 23(8):940–948
106. Zaidi H, El Naqa I (2010) PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging 37:2165–2187
107. Zhang L, Staelens S, Holen RV, Beenhouwer JD, Verhaeghe J, Kawrakow I, Vandenberghe S (2010) Fast and memory-efficient monte carlo-based image reconstruction for whole-body PET. Med Phys 37(7):3667

Variational Shape Representation for Modeling, Elastic Registration and Segmentation

Amal A. Farag, Ahmed Shalaby, Hossam Abd El Munim and Aly Farag

Abstract Shapes describe objects in terms of information invariant to scale, translation and rotation. Depending of the data source, shapes may be represented by object contours or representation/transformations that sustain the objects characteristics, such as the signed distance function. Biomedical objects have inherent plasticity due to movement and changes over time. Elastic registration is a fundamental image analysis step for tracking anatomical structures, diseases, progress of treatment and in image-guided interventions. Variational level set methods (LSM) represent objects' contours through an implicit function that enables tracking the objects' topologies. This chapter provides an overview of variational shape modeling as applied to the registration and segmentation problems. The chapter evaluates similarity/dissimilarity measures and common energy functional representations used in elastic shape registration. Common numerical methods to solve the optimization involved are studied. In addition, the chapter discusses clinical applications for which shape-based models enable robust performance with respect to occlusion and other image degradation.

A. A. Farag (✉)
Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, NIH, Bethesda, MD, USA
e-mail: aly.farag@louisville.edu

A. Shalaby · A. Farag
Department of Electrical and Computer Engineering, University of Louisville, Louisville, KY, USA
e-mail: www.cvip.uofl.edu

H. Abd El Munim
Computer and Systems Engineering Department, Ain Shams University, Cairo, Egypt
e-mail: hossamabdelmunim@gmail.com

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*, Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_3,

1 Introduction

In this chapter, we summarize and expand upon our work on variational shape modeling for segmentation and registration (e.g., [1–10]). Specifically, we represent shapes using vector distance functions (VDF). We use the VDF as shape prior for both shape-based segmentation and elastic shape registration. We derive the energy formulation for elastic registration and shape-based segmentation. We highlight the algorithms and the optimization technique used for solving the energy function. Finally, we apply the methodologies for various biomedical image analysis problems.

Shapes are represented either explicitly or implicitly [1, 9]. 2D/3D shape boundary points can be used directly/explicitly to deal with shapes (e.g., applications of alignment and retrieval) where the shape points are used directly to compute shape geometric properties and features. In the implicit shape representation, the shape boundary points can be computed by solving the zero level equation of the implicit shape function. This representation can be in either scalar or vector form. In this chapter, we use the VDF shape representation as a similarity measure in the shape registration process. More general transformations with inhomogeneous scaling, rotation, and translation parameters will be incorporated. The use of such vector functions results in a more adequate energy function which is optimized to achieve the transformation parameters both in the global and local registration schemes. A variational framework for the registration process is formulated. The gradient descent optimization criterion is used to handle the global registration similar to that in [11]. The local deformations are covered using the incremental free form deformations. The gradient descent optimization is not used to estimate the positions of the control points where the number of deformation parameters are large compared to the global alignment case. We demonstrate the nonrigid registration problem in vector implicit spaces as well.

Following our latest results (e.g., [9, 10]), we adopt a closed form solution for computing the elastic registration parameters which provides a large time reduction in comparison to the large number of iterations required by the gradient descent approach. We propose a quadratic energy function in terms of the control points positions (i.e., unknowns). Hence, the objective function is convex which leads to a single point solution of the minimization problem. Different experimental results for synthetic and real shapes registration cases will be demonstrated to show the efficiency of the proposed techniques. Also, comparison with the state of the art approaches will be discussed in detail.

The treatment below on shape representation and registration is based on our work [10]. We keep similar notations as well. Later on in the chapter we expand on this work for simultaneous segmentation and registartion of objects using shape priors.

2 Shape Representation

A map, $\mathbf{C}(p) : [0, 1] \subset R \rightarrow R^2$ defines a planar smooth curve with parameter p. The cartesian coordinates of the point vector can be defined by $\mathbf{C}(p) = [x(p)y(p)]^T$ where $0 \leq p \leq 1, 0 \leq x, \leq X$ and $0 \leq y, \leq Y$. This is the explicit representation of the given shape or contour $\mathbf{C}$. Open shapes have the relation $\mathbf{C}(0) \neq \mathbf{C}(1)$. A closed contour will always have $\mathbf{C}(0) = \mathbf{C}(1)$. Parameterizing complicated topology shapes is a challenge which is considered a disadvantage of the explicit shape representation method. Thus, a parametrization-free representation is needed. The implicit shape representation satisfies this condition as shown.

Given a smooth curve $\mathbf{C}^\alpha$ (defined above), that represents boundaries of the shape of interest, the following implicit vector function is defined as $\Phi_\alpha(\mathbf{X}) : \Omega_\alpha \subset \mathbf{R^2} \rightarrow \mathbf{R^2}$ where

$$\Phi\alpha(\mathbf{X}) = \mathbf{X}_0 - \mathbf{X} = [\phi_1\phi_2]^T, \mathbf{X} \in \Omega_\alpha, \phi_1 \; and \; \phi_2 \in R, \tag{1}$$

where $\mathbf{X}_0$ is the point on $\mathbf{C}^\alpha$ with the minimum Euclidean distance to $\mathbf{X}$ where $\mathbf{X} \in \Omega_\alpha$ (Ω_α is the domain that includes the shape/contour). The surface or boundary points always satisfy the relation $||\Phi_\alpha(\mathbf{C}^\alpha)|| = 0$. Note, that the implicit representation is dependent only on the boundary position, not on any parameterizations, and hence, it is suitable to represent a cloud of points or even scattered edge boundaries.

If a global transformation is applied to the given shape represented by the designed vector map, one can predict the map of the new shape. We define a shape β that is obtained by applying a transformation $\mathbf{A}$ to a given shape α. Let us assume that the transformation has a scale matrix $\mathbf{S}$, a rotation matrix $\mathbf{R}$, and a translation vector $\mathbf{T}$. The transformation can be written for any point $\mathbf{X}$ in the space as $\mathbf{A} = \mathbf{SRX} + \mathbf{T}$.

Applying the transformation to the given points results in the pair of points $\hat{\mathbf{X}}, \hat{\mathbf{X}}_0 \in \Omega_\beta$ (Domain of the Target Shape where $\Omega_\beta \subset R^2$). It is straightforward to show that:

$$\Phi_\beta(\mathbf{A}) = \hat{\mathbf{X}}_0 - \hat{\mathbf{X}} = \mathbf{SR}(\mathbf{X}_0 - \mathbf{X}) \tag{2}$$

as such the following relation holds:

$$\Phi_\beta(\mathbf{A}) = \mathbf{SR}\Phi_\alpha(\mathbf{X}) \tag{3}$$

Which illustrates that this representation can give a vector similarity measure that includes inhomogeneous scales and rotations. Also, it is invariant to the translation parameters, while the effect of scales and rotations can be predicted. This measure overcomes the problem of using the conventional signed distance maps that leads to the use of homogeneous scales only. Note, that the VDF components are smooth and differentiable at the boundary points.

3 Global Registration of Shapes

Finding point-wise correspondences (between the two given source and target shapes defined respectively by $\mathbf{C}^{\alpha}$ and $\mathbf{C}^{\beta}$) is the objective of the registration problem. An energy function is built based on the vector dissimilarity measure. The VDF shape representation changes the problem from the shape boundary domain to the higher dimensional vector representation. A transformation, $\mathbf{A}$, that gives pixel-wise vector correspondences between the two shape representations Φ_{α} and Φ_{β}, is required to be estimated.The problem now can be considered as a global optimization that includes all points in the image domain. Sum of squared differences will be considered with energy optimized by the gradient descent approach.

According to the properties of the implicit vector representation shown, the following dissimilarity measure is used: $\mathbf{r} = \mathbf{SR}\Phi_{\alpha}(\mathbf{X}) - \Phi_{\beta}(\mathbf{A})$ and the optimization energy function is formulated by the sum of squared differences as: $E(\mathbf{S}, \mathbf{R}, \mathbf{T}) = \int_{\Omega\alpha} \mathbf{r}^T \mathbf{r} d\Omega_{\alpha}$. The complexity of the problem is reduced by considering only points around the zero level of the vector function and neglecting mapping of far away points. The matching space is limited to a small band around the surface that can be selected by introducing the following energy function:

$$E(\mathbf{S}, \mathbf{R}, \mathbf{T}) = \int_{\Omega\alpha} \delta_{\varepsilon}(\Phi_{\alpha}, \Phi_{\beta}) \mathbf{r}^T \mathbf{r} d\Omega_{\alpha}. \tag{4}$$

where $\delta_{\varepsilon}$is an indicator function defined in [1].
The optimization of the given criterion is handled using the gradient descent method:

$$\frac{d}{dt}\vartheta = 2 \int_{\Omega\alpha} \delta_{\varepsilon} \mathbf{r}^T [\nabla_{\vartheta}(\mathbf{SR}\Phi_{\alpha}(\mathbf{X})) - \nabla\Phi_{\beta}^T(\mathbf{A})\nabla_{\vartheta}\mathbf{A}] d\Omega_{\alpha} \tag{5}$$

where $\vartheta \in \{S_x, S_y, \theta, T_x, T_y\}$ represents the set of scale, rotation, and translation parameters respectively.

3.1 Evaluation of Global Registration

In [10] we reported results for an experiment that involved 100 registration cases, using the corpus callosum (simple shape) and the hippocampus (four separate parts). Each case considers a source and a target shape. The source is fixed and the target is generated by applying a transformation on the source. Parameters $(S_x, S_y, \theta, T_x, T_y)$ are created and selected randomly from the ranges [0.8, 1.2], [0.8, 1.2], $[-60°, 60°]$, $[-60, 60]$ respectively. These generated patterns are kept as the ground truth for each case. The gradient descent optimization is performed to obtain a steady state estimate for each parameter associated with each registration case. The algorithm shows successful results for the one hundred cases and the energy decreases smoothly with the increase of the iteration number until perfect alignment is achieved. The measurements show that the mean errors and standard deviations (Table 1) are very

Table 1 Mean error (μ) and its standard deviation (δ) for the transformation parameters of the corpus callosum (CC) and hippocampus (HC) cases ($\mu \pm \delta$)

Structure	S_x	S_y	θ°	T_x	T_y
CC	-0.005 ± 0.009	-0.003 ± 0.007	-0.002 ± 0.018	-0.5 ± 0.4	-0.3 ± 0.5
HC	0.009 ± 0.007	0.005 ± 0.004	0.001 ± 0.09	0.00 ± 0.02	-0.0 ± 0.02

Parameter ranges:[0.8, 1.2], [0.8, 1.2], [$-60°$, $60°$], [-60, 60], [-60, 60], are used

appropriate and satisfactorily small. The final registration emphasizes that for each experiment, the boundaries of the source and target shapes become very close to one another. The gradient descent successfully estimates the scales, rotations, and translations with proper initialization.

In addition, we formed three groups of different shapes (Fighter Jet, Fishes, Number Four). Each group includes 11 instances of its corresponding shape. Different global registration processes are conducted by randomly taking 11 pairs from each group. For each pair of shapes, the correlation coefficient is calculated to measure the similarity between the shape representations: $\gamma = \frac{E[(||\Phi_\alpha||-\mu_\alpha)(||\Phi_\beta||-\mu_\beta)]}{\sigma_\alpha \sigma_\beta}$ where μ, σ stand for mean and standard deviation of the shape vector representations magnitudes respectively. The global registration process successfully increases the coefficient dramatically. Before alignment, the mean correlation coefficients and their standard deviations for the groups are (0.836 ± 0.047), (0.834 ± 0.087), and (0.754 ± 0.092), respectively. After alignment, the coefficients become (0.969 ± 0.013), (0.953 ± 0.03), and (0.911 ± 0.039). Note, that the last group has the largest local shape variations and hence, has the smallest average coefficient 0.911, which is small compared to other groups of coefficients.

For comparison with other techniques, two synthetic shape images have been created. The second image results by stretching the first with large inhomogeneous scales ($S_x = 2.5$, $S_y = 3.3$). Mutual information is used to register these contours (images) according to the technique in [15]. Mutual information suffers in such a situation because the scale range will increase/decrease the energy in one direction, providing unacceptable results (minimum position does not provide the correct parameters as shown in Fig. 1 left image). The proposed approach aims to align the contours of the given images to each other to obtain a global minimum at these scales exactly as shown, which is considered to be an advantage over the mutual information.

3.2 Global Registartion for Segmentation of Lung Nodule Regions

We use the global alignment approach with the shape-based segmentation as an application. In our previous work [2], we formulated the problem as a global registration between a shape and an intensity model implicit representation. In this paper, we adopt the above alignment technique to segment lung nodule regions [16]. Nodule size is an important factor in volumetric analysis of lung nodules. It has been

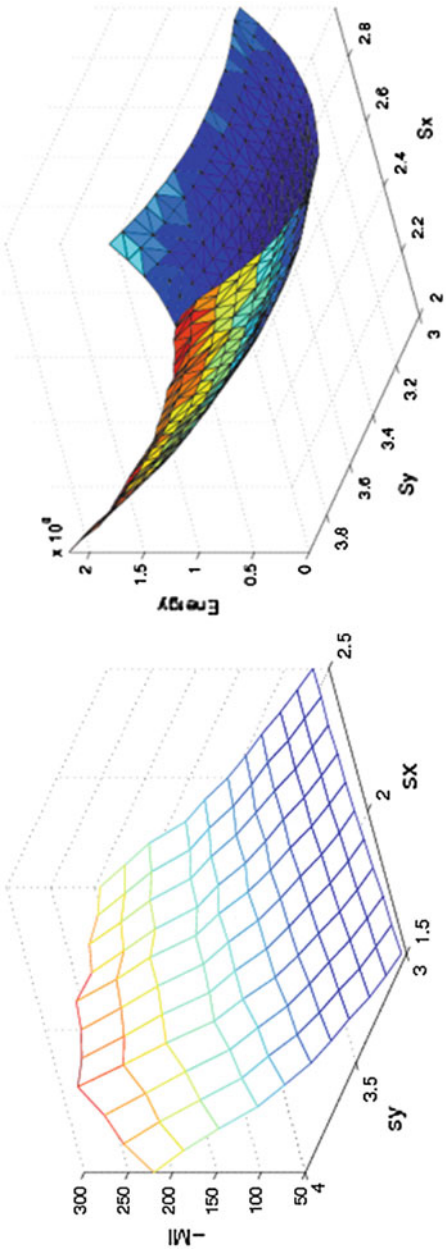


Fig. 1 Large scales registration case: the negative of mutual information is given to the *left*. The proposed energy is given to the *right*. The mutual information and the proposed energy are shown as functions of and the proposed energy are shown as functions of S_x and S_y

shown clinically that size is linked to nodule malignancy, with non-calcified nodules larger than in diameter having a higher rate of malignancy than smaller nodules. Size computation is usually performed by applying volumetric methods to a segmentation result. However, lung nodules segmentation in CT imaging is a complex and challenging process. One of the most important problems arises from possible attachments of the nodules to other anatomical objects. The lungs are a complex anatomical structure. Vessels, fissures, bronchi or pleura are structures that can be located close to lung nodules.

From our experience, we noticed that the main nodule regions considered for size computation are elliptic. A circle model is represented implicitly by the vector function Φ_p. A region of interest image (ROI) is taken from the whole lung CT scan to include the nodule. Intensity segmentation of the ROI is represented implicitly by the vector function Φ_g. Aligning the two models using the above approach will result in an ellipse that includes the nodule region. The model is initialized and then the alignment parameters are estimated using the gradient descent optimization. Different scales, rotation, and translation parameters are computed in each case to obtain an ellipse exactly around the nodule (see Fig. 2). The ellipse axis rotates while its size changes to include the boundaries of the nodule. A thresholding technique can be used later to remove the non-nodule parts from the elliptic areas.

4 The Elastic Registration Problem

Our objective is to find a function that gives the point correspondences between the two given domains (source and target). Let us define the $2D$ shape elastic registration as follows:A map $\mathbf{C}^{\hat{\alpha}}(\tau_s) : [0, 1] \in R \rightarrow R^2$ defines a planar source curve with a parameter τ_s (it is the source shape $\mathbf{C}^{\alpha}$ after applying the global transformation estimated by the methods above). The target is defined by $\mathbf{C}^{\beta}(\tau_t) : [0, 1] \in R \rightarrow R^2$. Assume that $\mathbf{C}^{\hat{\alpha}}(\tau_s)$ is the corresponding point of $\mathbf{C}^{\beta}(\tau_t)$ (the criteria for finding the correspondences can be found in the following sections). The output will be a C^0 function $f : R^2 \rightarrow R^2$ with $f(\mathbf{C}^{\hat{\alpha}}(\tau_s)) = \mathbf{C}^{\beta}(\tau_t)$. Different interpolation functions have been proposed to handle this problem [12]. We choose the free form deformation *FFD* model, based on B-splines [13, 14], which is a powerful tool for modeling deformable objects and has been previously applied to the tracking and motion analysis problems. The basic idea is to deform the shape by manipulating a mesh of control points. The resulting deformation controls the shape of the object and produces a smooth and continuous transformation.

Consider an $M \times N$ lattice of control points $\mathbf{P} = \mathbf{P}_{m,n}; m \in \{1, \ldots, M\}; n \in \{1, \ldots, N\}$, each point on the source shape will have the following form of deformation:

$$\mathbf{L}(\tau_s) = \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) \delta \mathbf{P}_{i+k, j+1} \tag{6}$$

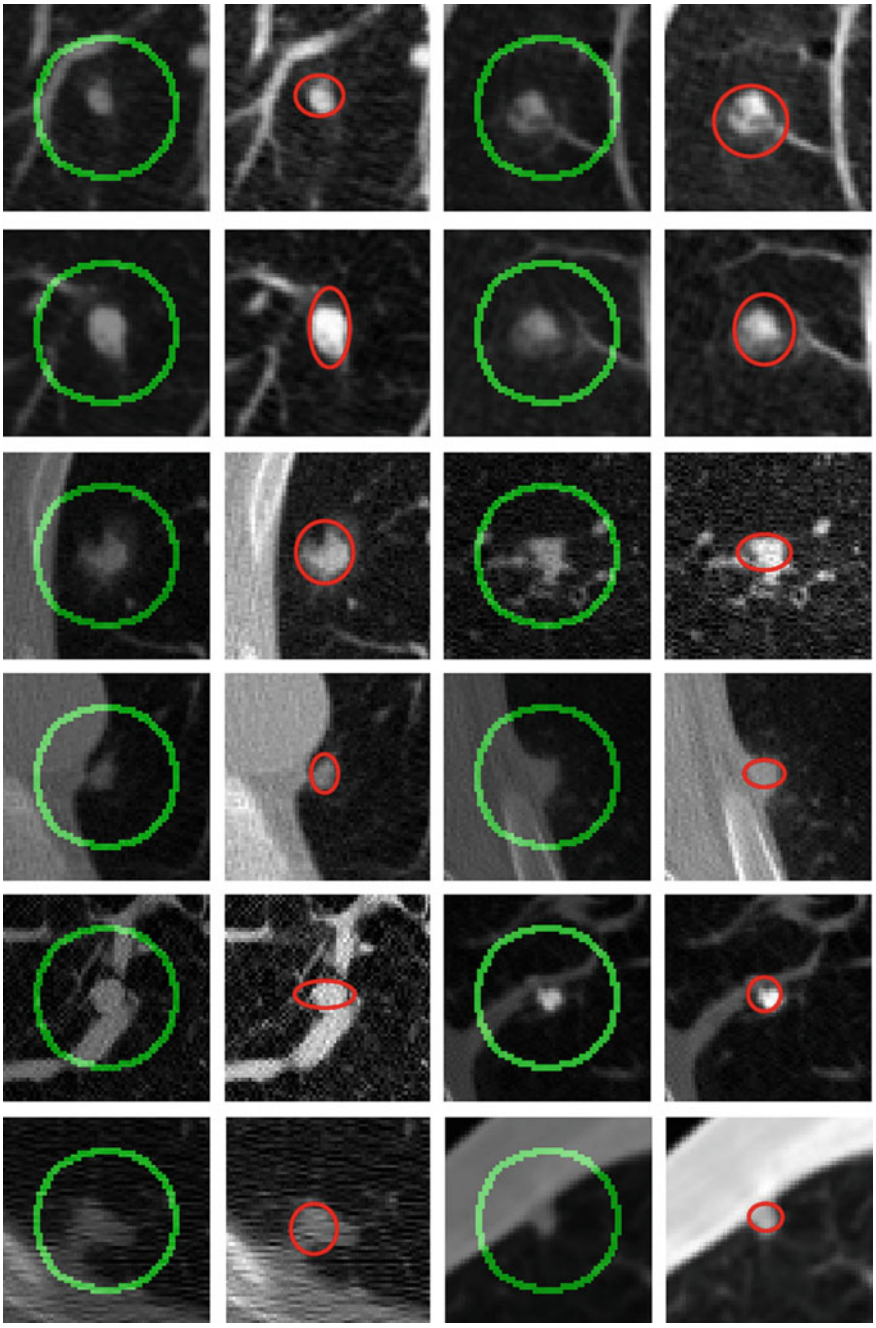


Fig. 2 Initial positions are shown in *green* at the first and third columns while final ellipses are demonstrated in *red* at the second and fourth columns

where $\delta\mathbf{P} = \delta\mathbf{P}_{m,n} \in \{[\delta\mathbf{P}^x_{1,1}\delta\mathbf{P}^y_{1,1}]^T, \ldots, [\delta\mathbf{P}^x_{M,N}\delta\mathbf{P}^y_{M,N}]^T\}$ is the control point deformation vector, $i = (x.(M-1)/X)+1$, $j = (y.(N-1)/Y)+1$, $u = x.M/X)$, $v = y.N/Y - (y.N/Y)$, and the spline basis functions (B) are defined in [14]. So the cubic B-spline is used as an approximation function for our interpolation problem. Below, we propose and discuss the problem solution in implicit and explicit spaces. In [10], we devised a closed form solution of the interpolation function parameters.

4.1 A Coarse to Fine Strategy with IFFD's

The control lattice points are required to move to correctly obtain correspondences over shape boundaries. A very small error can be achieved when using a high resolution control lattice since the number of degrees of freedom increases. However, this is not enough. Such sudden movement will result in unnecessary cross overs of the domain grid lines and the registration process will be meaningless. This will result in changing and corrupting the object topology. A better way is to move the grid step by step towards the target.

To avoid this, a coarse to fine strategy is used (equivalent to the incremental free form deformations used in [15]). We start with a resolution of 4×4 and solve for the deformation. Iteratively we increase the resolution to 5×5, 6×6, and so on and so forth. In each step, the positions of the control points are computed and the contour moved to the new position until a satisfactory error distance is obtained. The result is smooth and the correspondence is achieved accurately. This process handles the error extremely well and provides an impressive infinitesimal energy function and smooth grid deformations simultaneously.

4.2 Solution in Vector Implicit Spaces

Following the work in [11], a local deformation vector $\mathbf{L} = \mathbf{L}(\mathbf{X}) = \mathbf{L}(\delta\mathbf{P})$(described above) is applied to the globally transformed shape represented by $\hat{\alpha}$. The following dissimilarity measure is considered:

$$\mathbf{r}_n - \Phi_{\hat{\alpha}}(\mathbf{X}) - \Phi_{\beta}(\mathbf{X} + \mathbf{L}) \tag{7}$$

and hence the non rigid energy function will be defined as:

$$E^{\Phi}_n(\delta\mathbf{P}) = \int_{\Omega_{\hat{\alpha}}} \mathbf{r}^T_n \mathbf{r}_n d\Omega_{\hat{\alpha}} \tag{8}$$

The local deformations are smoothed by adding another term that includes their derivatives as follows:

$$E_n^{\Phi}(\delta\mathbf{P}) = \int_{\Omega_{\hat{\alpha}}} \mathbf{r}_n^T \mathbf{r}_n d\Omega_{\hat{\alpha}} + \lambda \int_{\Omega_{\hat{\alpha}}} (||\mathbf{L}_x||^2 + ||\mathbf{L}_y||^2 + ||\mathbf{L}_{xx}||^2 + ||\mathbf{L}_{yy}||^2) d\Omega_{\hat{\alpha}} \quad (9)$$

As an interpretation, the energy contains a term for covering the local deformations and another for penalizing large derivatives. To make the addition homogeneous, we weight the second term by $\lambda \in R^+$. Again, we take the derivative of the energy with respect to each of the unknown parameters as follows:

$$\frac{\partial E_n^{\Phi}}{\partial \delta \mathbf{P}} = -2 \int_{\Omega_{\hat{\alpha}}} \mathbf{r}_n^T (\nabla \Phi_\beta)^T \frac{\partial \mathbf{L}}{\partial \delta \mathbf{P}} d\Omega_{\hat{\alpha}} + 2\lambda \int_{\Omega_{\hat{\alpha}}} ((\mathbf{L}_x)^T \frac{\partial \mathbf{L}_x}{\partial \delta \mathbf{P}} + (\mathbf{L}_y)^T \frac{\partial \mathbf{L}_y}{\partial \delta \mathbf{P}} + (\mathbf{L}_{xx})^T \frac{\partial \mathbf{L}_{xx}}{\partial \delta \mathbf{P}} + (\mathbf{L}_{yy})^T \frac{\partial \mathbf{L}_{yy}}{\partial \delta \mathbf{P}}) d\Omega_{\hat{\alpha}} \quad (10)$$

We assume that the amount of pixel deformation is relatively small such that its vector representation can be approximated using Taylor series expansion as: $\Phi_\beta(\mathbf{X}+\mathbf{L}) \approx \Phi_\beta(\mathbf{X}) + (\nabla \Phi_\beta(\mathbf{X})^T)\mathbf{L}$. The control points are required to move and minimize the above objective function and hence satisfy the following zero condition: $\frac{\partial E_n^{\Phi}}{\partial \delta \mathbf{P}} = [0 \quad 0]^T$. By setting $\Phi(\mathbf{X}) = \Phi_{\hat{\alpha}}(\mathbf{X}) - \Phi_\beta(\mathbf{X})$, the above formulation will lead to:

$$\int_{\Omega_{\hat{\alpha}}} \Phi^T (\nabla \Phi_\beta)^T \frac{\partial \mathbf{L}}{\partial \delta \mathbf{P}} d\Omega_{\hat{a}} = \int_{\Omega_{\hat{\alpha}}} ((\nabla \Phi_\beta)^T \mathbf{L})^T (\nabla \Phi_\beta)^T \frac{\partial \mathbf{L}}{\partial \delta \mathbf{P}} d\Omega_{\hat{a}} + \lambda \int_{\Omega_{\hat{\alpha}}} ((\mathbf{L}_x)^T \frac{\partial \mathbf{L}_x}{\partial \delta \mathbf{P}} + (\mathbf{L}_y)^T \frac{\partial \mathbf{L}_y}{\partial \delta \mathbf{P}} + (\mathbf{L}_{xx})^T \frac{\partial \mathbf{L}_{xx}}{\partial \delta \mathbf{P}} + (\mathbf{L}_{yy})^T \frac{\partial \mathbf{L}_{xx}}{\partial \delta \mathbf{P}}) d\Omega_{\hat{a}} \quad (11)$$

Fortunately, the above equation is linear in terms of control points deformations. We can formulate the following linear system to give a closed form solution for the unknown deformations:

$$\bar{\mathbf{\Psi}} \mathbf{\Theta} = \bar{\mathbf{\Lambda}} \quad (12)$$

where:

$$\bar{\Psi}_{r,c} = \int_{\Omega_{\hat{\alpha}}} ((\nabla \Phi_\beta)^T \mathbf{L}^{r,c})^T (\nabla \Phi_\beta)^T \frac{\partial \mathbf{L}}{\partial \theta_r} d\Omega_{\hat{\alpha}} + \lambda \int_{\Omega_{\hat{\alpha}}} ((\mathbf{L}_x^{r,c})^T \frac{\partial \mathbf{L}_x}{\partial \theta_r} + (\mathbf{L}_y^{r,c})^T \frac{\partial \mathbf{L}^y}{\partial \theta_r} + (\mathbf{L}_{xx}^{r,c})^T \frac{\partial \mathbf{L}_{xx}}{\partial \theta_r} + (\mathbf{L}_{yy}^{r,c})^T \frac{\partial \mathbf{L}_{yy}}{\partial \theta_r}) d\Omega_{\hat{\alpha}}, \quad (13)$$

$$\bar{\Lambda}_r = \int_{\Omega_{\hat{\alpha}}} \Phi^T (\nabla \Phi_\beta)^T \frac{\partial \mathbf{L}}{\partial \theta_r} d\Omega_{\hat{\alpha}}. \quad (14)$$

Note, that this will lead to computing new positions of the control lattice points and hence, we can compute the entire domain deformation field. Other approaches use gradient descent to compute the position of each point in space. Unfortunately, the use of this form of local deformation does not guarantee proper handling of the registered shape since it cannot preserve topology. Also, it results in scattered front points leading to an open surface which is not the case. Another issue is that the gradient descent does not guarantee the desired solution especially when using a large number of deformation vectors.

Now we will illustrate the whole algorithm for elastic shape registration in vector implicit spaces. Assume that $N_x^i \times N_y^i$ is the resolution of the control lattice initially denoted by i. The resolution at any time will be $N_x \times N_y$. The basic algorithm steps are shown as follows:

1. Set $N_x = N_x^i$ and $N_y = N_y^i$ (initial grid size).
2. Compute the vector distance representation of the source and target shapes $\Phi_{\hat{\alpha}}$ and Φ_β respectively.
3. Construct a control lattice of size $N_x \times N_y$ and initialize its point deformation vectors to zeros.
4. Construct and Solve Eq. 12 to obtain the new deformation of each control point and hence, compute its new position.
5. Based on the new lattice, update the source points and its vector representation, $\Phi_{\hat{\alpha}}$, by computing the new deformation field using Eq. 6.
6. Set $N_x = N_x + 1$ and $N_y = N_y + 1$.
7. Check the stopping criteria. Either the objective function goes below a certain threshold or a number of maximum resolution levels is reached, otherwise go to step #3.

4.3 Evaluation

The point-based algorithm described in [8] looks simple and does not require huge space to store the shape representation. However, for many registration cases, we obtain unsatisfactory results. This is due to its use of the closest point criteria to decide the correspondences. Examples of shapes that show the failure of the algorithm are shown in Fig. 3. It is clear that these examples fail because the left end of the source shape arrives at the center line of the target. The decision to go left or right becomes very difficult since both directions have the same distance.The above algorithm works efficiently and handles the cases that the former algorithm fails to register as shown in last row of Fig. 3 (see also [10]).

The reason for the success is that the approach minimizes the differences between the two implicit representations and hence, makes the two contours very close to each other. The neighborhood vectors around the shape boundaries have small magnitudes. This property with the delta function described above helps in moving the contour smoothly in the proper direction. This creates a force that stretches the source to the

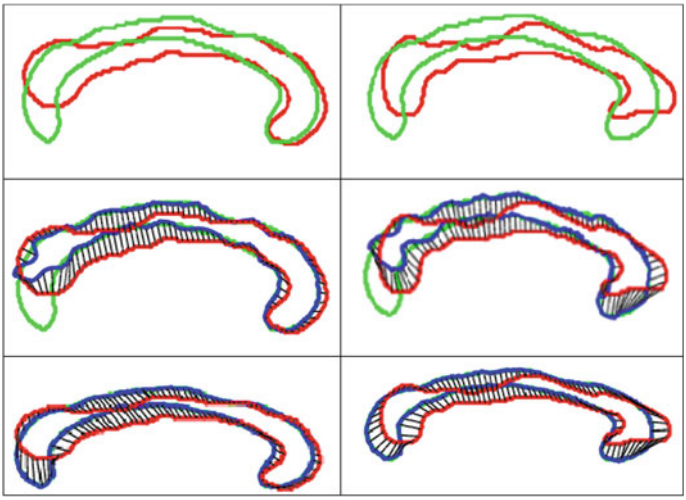


Fig. 3 Different elastic registration examples: source contour is given in *red*, target contour is drawn in *green*, and deformed contour is shown in *blue*. Initial contours are shown in the first row. The second row shows the failure of the ICP [16] algorithm with the IFFD given in [9]. The success of our approach is demonstrated in the last row

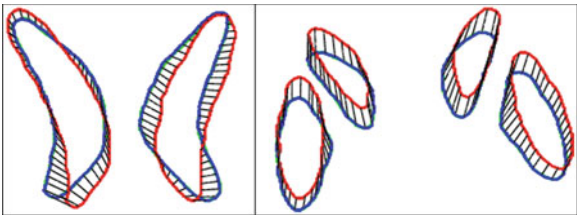


Fig. 4 Different elastic registration examples of shapes containing multiple parts using the implicit vector representation: source contour is given in *red* while target contour is drawn in *blue*. Correspondences for the ventricle is shown to the *left* while hippocampus results are illustrated to the *right*

target while the free form deformation preserves the topology of the shape. In all of the registration cases, we notice that the grid deformation is smooth and each grid line is kept in its order without crossovers or folding.

The algorithm works for multiple objects without any problem since it is not necessary to handle the parameterizations. Elastic registration cases of two-part and four-part shapes are illustrated in Fig. 4 for the brain ventricles and the hippocampus shapes.

For more validation of the above algorithm, a deformed shape is generated from a given tooth model. The target shape is generated by applying random deformations on the source image such that correspondences are stored as ground truth for validation. The model is from real data of a Cone-beamed Computed Tomography scan of resolution 0.2 mm per pixel (Fig. 5a).The approach is applied by increasing the

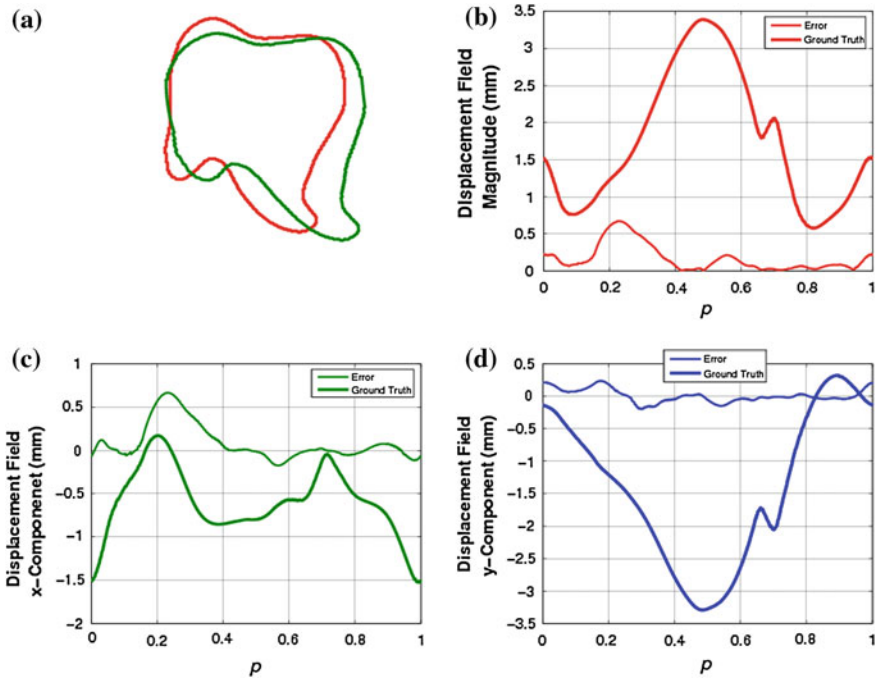


Fig. 5 A local registration example with displacement field measurements: **a** initial shape contours of the source and target models, **b** displacement field magnitude plot over the shape contour, **c** x-component of the displacement field error, and **d** y-component of the displacement field error

resolution of the control lattice one step in each direction at a time starting from a grid of 5×5. The contours come closer to each other iteratively until steady state is reached. The approach shows very high accuracy. The displacement field is achieved with an average error of 0.1677 mm. As shown in Fig. 5b–d, errors of the displacement fields are plotted versus the curve parameterizations allowing a follow up of the error distribution over the whole shape boundary.

Compared to the proposed in [15], the above algorithm is more complicated but the closed form solution for the control points positions possesses a great advantage. Also, if we do not use the closed form solution, the total execution time will be doubled. Assume that the registration problem needs N incremental levels of free form deformations, each level has, $N_{cp} = n_x \times n_y$ control points, and hence, $2 \times N_{cp}$ unknown variables x and y components for the gradient descent. If the average number of iterations of the gradient descent needed for each variable to reach the steady state is N_{Iter} with average time per iteration of Δt (with the method in [15]), the total time will be: $Time_1 = \sum_{i=1}^{N} (2 \times N_{CP}^i * N_{Iter} \times \Delta t)$. For the same IFFD setup with the gradient descent of Eq. 10 which does not use the closed form solution, the total time will be doubled; $Time_2 = \sum_{i=1}^{N} (2 \times N_{CP}^i \times N_{Iter} \times (2 \times \Delta t))$, since we use an implicit vector representation which has two components. The gradient descent execution time for an iteration will be roughly twice that of Δt. In the case of applying the proposed closed

form solution, the gradient descent iterations will be omitted. The new execution time can be estimated as: $OurTime = \sum_{i=1}^{N}(2 \times N_{CP}^{i} \times (2 \times \Delta t))$. The time to construct the linear system of the closed form (Eq. 12) is equal to the time of one gradient descent iteration for all variables. Our time holds the relation: $OurTime = 2 * Time_1/N_{Iter}$. A good steady state solution for the gradient descent needs a number of iterations greater than which guarantees that our execution time is less than that of the approach in [15].

5 Variational Shape-Based Segmentation

Variational approaches segment shapes through an energy minimization framework that controls the evolution of an implicit/explicit contour/surface. The active contour models proposed by Kass et al. [20] and level sets proposed by Osher and Sethian [19] are the most important variational methods in the literature. The active contour models minimize the energy formulation using the explicit shape representation, which requires parameterizations of the contour. Explicit shape representations suffer when applied to shape modeling since they do not allow the shape to undergo topological changes. The level sets method uses implicit shape representation, which does not need contour parameterizations, and handles the topological changes of shapes.

Tsai et al. [24] proposed a shape model using a signed distance function of the training data. The Eigenmodes of implicit shape representations are used to model the shape variability. They proposed a shape prior using a coefficient of each training shape. Cremers et al. [26] proposed a simultaneous kernel shape based segmentation algorithm with a dissimilarity measure and statistical shape priors. This method is validated using various image sets which objects are tracked successfully. In [15] the distance function is used to implicitly represent open/closed shapes (structures). The images of distance functions are registered using the mutual information approach. In addition to global registration, they used a b-spline based incremental Free Form Deformation (IFFD) to minimize a dissimilarity measure. Taron et al. [25] proposed an invariant representation of shapes, and computing uncertainties on the registration process. They proposed a dimensionality reduction technique to lower the cost of the density estimate computation of kernel based shape model. Mahmoodi [28] proposed a shape-based active contours for fast video segmentation. Their level sets implement is based on Mumford-Shah [29] and Chan-Vese [21] methods. They compared their method with only intensity based segmentation method.

5.1 Methods

The intensity (as the existing information) and shape (as the prior information) are modeled to obtain the optimum segmentation in this study. The intensity information is modeled using the histogram of gray levels of the image. This information is mod-

eled using the Gaussian distribution. The model estimates the marginal density for each class. Kendall [18], definesshape the geometrical information that remains when location, scale, and rotational effects are filtered out from an object. Hence, the shape information is modeled after the sample shapes are transformed into the reference space. The shape variability is modeled using the occurrences of the transformed shapes. To label the image into meaningful areas, the chosen information is modeled to fit progressively in each of the regions by an optimization process. Each pixel in the image will have two probabilities to be an object and a background class based on the intensity and shape models. These probabilistic values will guide the energy (cost) functionals in the optimization process. Next sections detail the proposed method

5.1.1 Generation of the Shape Prior

As described in [30], the shape model is required to capture the variations in the training set. This model is considered to be a weighted sum of the new projected SDFs's as follows:

$$\Phi_{\mathbf{P}} = \sum_{a=1}^{N} \omega_a \Phi_a^t \tag{15}$$

Let $\mathbf{W} = [\omega_1, \ldots, \omega_N]^t$ to be the weighting coefficient vector. By varying these weights, $\Phi_{\mathbf{P}}$ can cover all values of the training distance functions and, hence, the shape model changes according to all of the given images. A new probabilistic and dynamic shape model is synthesized using the first four principal components. Two shape probability density functions which represent the probability of i) the object (inside of a boundary) and ii) background regions (outside of a boundary) are obtained:

$$P_o^s(x) = \frac{\sum_{j=1}^{N} \omega_j |\Phi_j^t(x)| H(-\Phi_j^t(x))}{\sum_{j=1}^{N} \omega_j |\Phi_j^t(x)|}, P_b^s(x) \frac{\sum_{j=1}^{N} \omega_i |\Phi_j^t(x)| H(\Phi_j^t(x))}{\sum_{j=1}^{N} \omega_j |\Phi_j^t(x)|}, \tag{16}$$

where $H(.)$ is the Heaviside step function as a smoothed differentiable version of the unit step function. Also, we should note that $P_o^s(\mathbf{x}) + P_b^s(\mathbf{x}) = 1$. This step is integrated into the registration step which is described in section 5.1.2, hence the shape model is dynamically reconstructed in the registration process.

Figure 6 shows the detailed description of the shape model where the shape weighting coefficients are normalized, i.e. $\mathbf{w} = \{\omega_1, \ldots, \omega_N\} = \{1/N, \ldots, 1/N\}$. The green color shows the background region which does not have any intersection with any training shape. The blue color shows the object region which is the intersection of all projected training shapes.

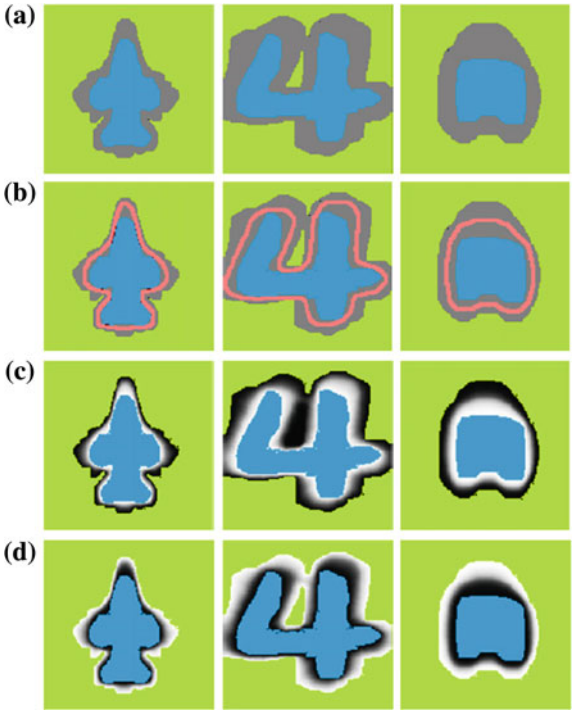


Fig. 6 **a** The *gray* color represents the variability region. **b** The *red* color shows the contour of the average shape (Φ_p). **c** The object (p_o^s) and **d** background (p_b^s) shapes are modeled in the variability region which the pixel values are defined in (0 : 1)

5.1.2 Level Sets Segmentation

The level sets formulation was first introduced by Osher and Sethian [19]. Topology changes like merging and splitting, are handled naturally without the need of parameterization. Given a curve C, it can be embedded into a higher dimension function Φ as $C = \{\mathbf{x} : \Phi(\mathbf{x}) = 0\}$. Then the curve is defined as the zero level of the implicit function. If the time t is added to the function, curve evolution function is changed to $\Phi = \Phi(\mathbf{x}, t)$. The surface function Φ evolves with the time and the evolution front is always represented as the zero level. In the literature, the final level sets formulation is defined as follows:

$$\Phi(t + \Delta t) = \Phi(t) - F|\nabla\Phi|\Delta t = 0 \tag{17}$$

There has been various methods to model the speed function, F. In this paper, a new method which is integrating the intensity and prior shape information is proposed. We use two energy functionals to be minimized. The first functional is to extract object regions using image intensities only with a statistical level set evolution as

described in [22]. We need this step to obtain the image feature to be used in the shape registration process. The second functional, which is slightly different than the formulation proposed in [26], depends on the dissimilarity measure between our shape model and the resulting contour which is obtained in the first phase.

The data is assumed to consist of two classes: object and background. Suppose that the intensity probability density function (*pdf*) within each of these two regions, denoted as p_o^I and p_b^I, can be modeled using a Gaussian distribution whose parameters are adaptively updated during the course of evolution ofthe level set function. The segmentation process starts by initializing the level set function as the signed distance function of a circle centered at a seed point(s) that is placed automatically using the Matched filter [31] or with manual annotation. Then, the statistical parameters corresponding to the *pdf* for the object and background are estimated as follows:

$$\mu_o = \frac{\int_\Omega I(\mathrm{X})H(-\Phi_{\mathbf{f}^*})d\Omega}{\int_\Omega H(-\Phi_{\mathbf{f}^*})d\Omega}, \quad \mu_b = \frac{\int_\Omega I(\mathrm{X})H(\Phi_{\mathbf{f}^*})d\Omega}{\int_\Omega H(\Phi_{\mathbf{f}^*})d\Omega},$$

$$\sigma_o^2 = \frac{\int_\Omega (I(\mathrm{X}) - \mu_o)^2 H(-\Phi_{\mathbf{f}^*})d\Omega}{\int_\Omega H(-\Phi_{\mathbf{f}^*})d\Omega}, \quad \sigma_b^2 = \frac{\int_\Omega (I(\mathrm{X}) - \mu_b)^2 H(\Phi_{\mathbf{f}^*})d\Omega}{\int_\Omega H(\Phi_{\mathbf{f}^*})d\Omega},$$

$$\pi_o = \frac{\int_\Omega H(-\Phi_{\mathbf{f}^*})d\Omega}{\int_\Omega d\Omega}, \textit{and} \quad \pi_b = \frac{\int_\Omega H(\Phi_{\mathbf{f}^*})d\Omega}{\int_\Omega d\Omega} \tag{18}$$

where μ, σ, and π are the mean, standard deviation, and prior probability of the corresponding *pdf* [22]. Object and background regions are represented by $H(-\Phi)H(\Phi)$, respectively. The pixel position, (x, y), is represented as $(\mathbf{x})$. The intensity based energy term is modeled to maximize posterior probability of each region as follows:

$$E_{intensity}(\Phi_{\mathbf{f}^*}) = -\int_\Omega P_o^I(I(x))H(-\Phi_{\mathbf{f}^*})d\Omega - \int_\Omega P_b^I(I(x))H(-\Phi_{\mathbf{f}^*})d\Omega + \in L, \tag{19}$$

where L is the front length of the surface area and $\in$ is a constant between 0 and 1. The change of the level set function with time is calculated by the Euler-Lagrange with the gradient descent given as:

$$\frac{\partial \Phi_{\mathbf{f}^*}}{\partial t} = -\frac{\partial E_{intensity}}{\partial \Phi_{\mathbf{f}}^*} = \delta(\Phi_{\mathbf{f}^*})[P_0^I(I(x)) - P_b^I(I(x))] + \in K \tag{20}$$

where k is the curvature of the evolving contour (or derivative of L) and δ is the derivative of the Heaviside step function. By solving this gradient descent formulation, the initial segmented region $(\Phi_{\mathbf{f}^*})$ is obtained. After this step, the shape energy (E_{shape}) is optimized using the shape based functions which are defined in Eqs. 15 and 16.

After the object region is initially segmented, the shape model is embedded into this domain by minimizing the new energy functional. It should be noted that the method is implemented in 2D dimension in this work. However, the extension of 3D

dimension is straightforward. A transformation matrix, **T**, that gives pixel-wise correspondences between the two shape representations Φ_{source} and Φ_{target} is required. The transformation has scaling, rotation, and translation components represented as follows:

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}, R = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}, Tr = [t_x, t_y]^t. \tag{21}$$

The transformation will be in the form of $\boldsymbol{T(x)} = \boldsymbol{X} = \boldsymbol{SRx} + Tr$ where $\boldsymbol{X} \in \Phi_{\mathbf{f}^*}$ and $\boldsymbol{x} \in \Phi_{\mathbf{P}}$ The proposed dissimilarity measure is

$$E_{shape}(\Phi) = \rho E_{Global} + E_{Local}, \tag{22}$$

where ρ is the normalization constant which controls the relationship between the first and second terms which can be described as follows:

$$E_{Global}(\Phi_{\mathbf{P}}, \Phi_{\mathbf{f}^*}|T) = \int_{\Omega} (\sqrt{s_x s_y}\Phi_{\mathbf{P}}(x) - \Phi_{\mathbf{f}^*}(X))^2 d\Omega, \tag{23}$$

$$E_{Local}(\Phi_{\mathbf{f}^*}, P_{o,b}^{S,I}|W) = -\int_{\Omega} P_o^S(x) P_o^I(x) H(-\Phi_{\mathbf{f}^*}(x)) d\Omega \; - \int_{\Omega} P_b^S(x) P_b^I(x) H(\Phi_{\mathbf{f}^*}(X)) d\Omega. \tag{24}$$

The first term of the proposed energy formulation is the (sum-of-squared distance) SSD of matched distances. It helps to estimate the registration parameters $(s_x, s_y, \theta, t_x, t_y)$, iteratively. Distance changes anisotropically in x-y directions. That's why the geometric mean between s_x and s_y as an approximation is proposed, since the SDF is not invariant to inhomogeneous scaling. After the registration parameters are estimated the shape model, Φ_p, and the projected training shapes, $\{\Phi_1^t, \ldots, \Phi_N^t\}$, are registered to the target domain using the affine transformation. However, this approximation still may not be enough to perfectly align the shapes. Hence, it is needed to add the other shape *pdf* term. A pixel inside the object of interest needs to have bigger object probability. At the same time, this pixel needs to have smaller background probability as well. So, the second term maximizes the probability for object pixels to be correctly classified as internal points. The same will happen for the background points. This step helps to estimate the shape weighting coefficients $(\mathbf{w} = \omega_1, \ldots, \omega_N)$ and to refine the result of the first component more accurately. Our proposed framework including the training step is shown in Fig. 7. The registration and weighting parameters $(s_x, s_y, \theta, t_x, t_y, \omega_1, \ldots, \omega_N)$ are computed to minimize E_{shape} using the Nelder-Mead simplex optimization method which was first proposed by Nelder and Mead and proved using theoretical results by Lagarias et al. [32]. The Nedler-Mead method aims to minimize a scalar-valued nonlinear function of n variables using function values, hence it is one of the direct search methods.

5.2 Evaluation

To assess the accuracy and robustness of our proposed framework, we tested it using clinical data sets as well as synthetic and phantom images. All algorithms are implemented on a PC with a 3 GHz AMD Athlon 64 × 2 Dual processor, with 3 GB RAM. First, we describe the experimental results on synthetic images. Second, validation on the European Spine Phantom (ESP) with various noise levels and clinical data sets will be shown. Effect of initialization will be evaluated.Shape based segmentation is useful when the target shape has some occlusions and missing information.

5.2.1 Shape-Based Segmentation of Synthetic Objects

Figures 8 and 9 show results on synthetic jet airplane and number four images with some missing information or occlusions. As seen in the results, the first component (of Eq. 22) is useful for an approximate transformation of the shape model. The second component enhances the segmentation with updated shape coefficients. Hence, the proposed dissimilarity measure is able to improve the global registration results. The results show that occlusions and missing information mislead those methods based only on intensity model. Using the shape prior information the desired shapes are recovered. Also, we observe that the proposed method slightly improves segmentation quality of our previous study [27]. As shown in Fig. 8. The proposed method is more able to capture the fine details and corners of the objects.

In [11, 24], the dissimilarity measures have limitations to capture the object-of-interest if the source and target shapes have inhomogeneous scale differences. Figure 10 shows the results when the target shapes have (i) homogeneous, and (ii–iv) inhomogeneous scale differences. Because dissimilarity measures of two alternative methods discard a possible scale difference in x or y directions, they fail when the target shapes are scaled inhomogeneities in x-y directions. The results prove that the proposed method overcomes the problems inhomogeneous scale differences. The computational costs of the two alternative methods [11, 24] and our method on 40 images (with 128 × 128 size) are approximately 220, 340, and 360 s, respectively. Since the method described in [11] does not estimate the shape coefficients in the optimization, it executes the experiment in faster time. Also, since the proposed method estimates two scaling parameters (s_x, s_y), the execution time may be expected to be slightly higher than other two alternatives.

5.2.2 Shape-Based Segmentation of Vertebral Body from CT

Our approach is also tested on clinical CT images to segment vertebral bodies (VBs) as well as the European spinal phantom (ESP). The vertebra consists of the VB and spinal processes. The red color shows the contour of the region of interest in Fig. 11. The objective is to segment the VB region correctly. Spinal processes and ribs should

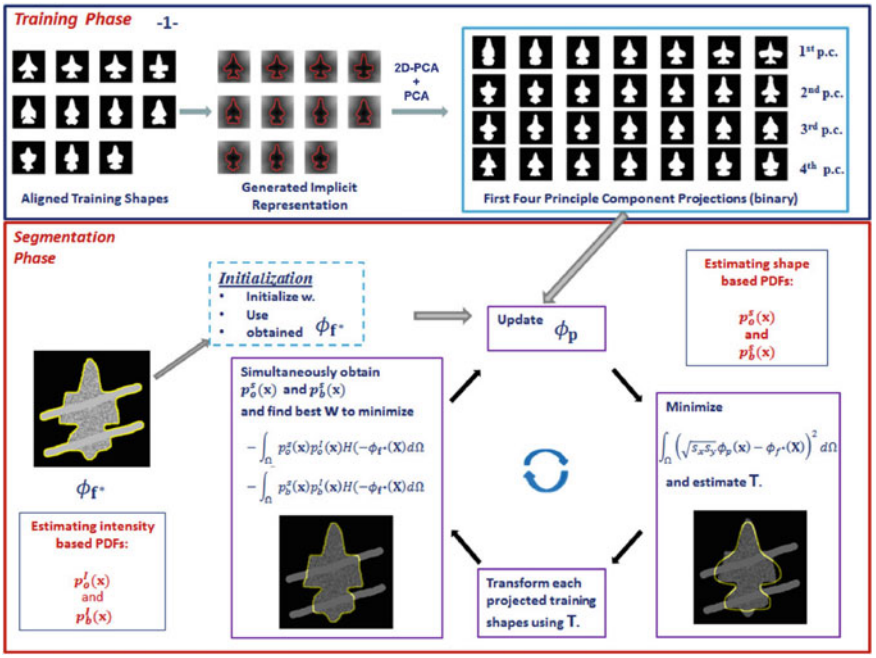


Fig. 7 The general framework is shown. The steps can listed as follows: *1* Obtain shape projections to define the shape variability. *2* Segmentation

not be included in the bone mineral density (BMD) measurements. The clinical data sets were scanned at 120 KV and 2.5 mm slice thickness. In this experiment, 260 testing CT slices (totals to 15 VBs) which are obtained from 13 different patients and different spine bone regions (i.e. lumbar, thoracic, etc.) are tested. To assess the proposed method under various challenges, a zero mean Gaussian noise was added to the CT images with different signal-to-noise ratios (SNR). To compare the proposed method with other alternatives, VBs are subsequently segmented using two other methods; (1) the active appearance model (AAM) [23], and (2) our earlier PCA-based approach which is described in [8].

Segmentation accuracy is measured for each method using the ground truths (expert segmentation). To evaluate the results, the percentage segmentation accuracy (A) is calculated as follows:

$$Dice's\ Coefficient(A\%) = \frac{100 * 2TP}{2TP + FP + FN}. \tag{25}$$

where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. The segmentation accuracy is shown in Table 2. It is clear that the noise immunity of our method are much higher than other alternatives. Figure 11 shows the segmentation results of the proposed framework with different scaling, translation, and rotation initializations.

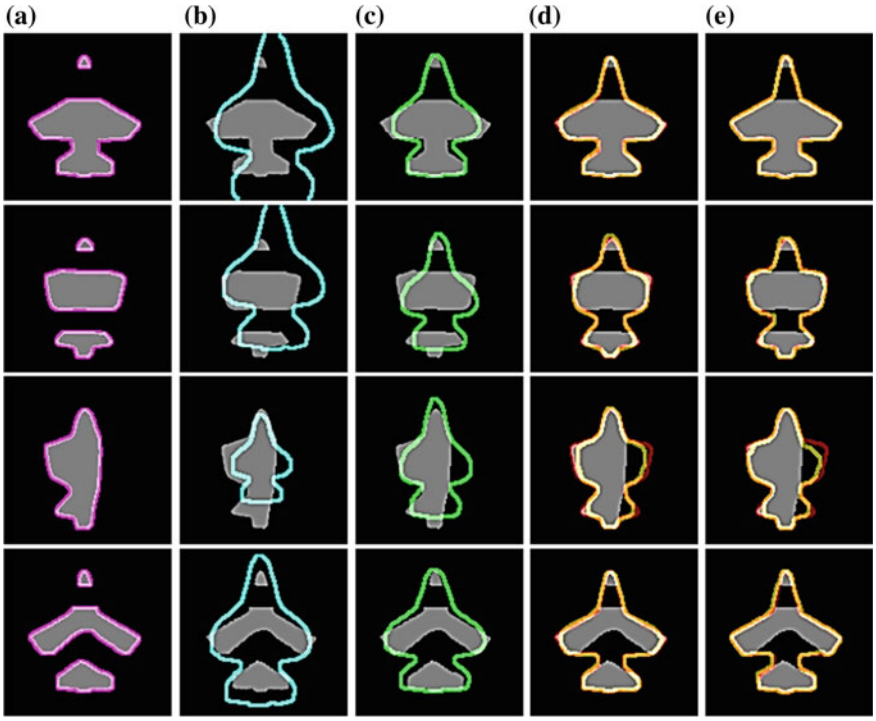


Fig. 8 Segmentation results of a synthetic jet airplane images with different missing information and initializations. **a** The intensity only based segmentation results. **b** Different shape model initialization. **c** The results using only the first term of Eq. 22. **d** [27]. **e** The segmentation of the proposed method (the *red* and *yellow* colors show the contour of the ground truth shape region, and the contour of the automatically segmented region, respectively)

Table 2 Average segmentation accuracy of the proposed vb segmentation on 272 ct images. The size of each image is 512×512

	$SNR = 100\ dB$	$SNR = 50\ dB$	$SNR = 10\ dB$	$SNR = 1\ dB$	s/slice
Intensity based, %	79.3	66.2	57.9	51.8	5.6
AAM [23], %	85.2	83.7	79.0	76.1	7.2
PCA-based [8], %	89.3	83.6	81.8	81.3	10.8
Proposed, %	94.3	92.9	89.3	86.8	11.3

Results indicate that the performance of our method is almost constant with different initialization parameters. To quantitatively demonstrate the accuracy of our approach, we calculate the average segmentation accuracy of our segmentation method on 272 CT images (including 12 ESP images) the under various signal-to-noise ratios and compare the results with the two other methods (Intensity-based and PCA-based). Our 2D-PCA based framework outperforms the conventional PCA described in [8] as shown in Fig. 12a.

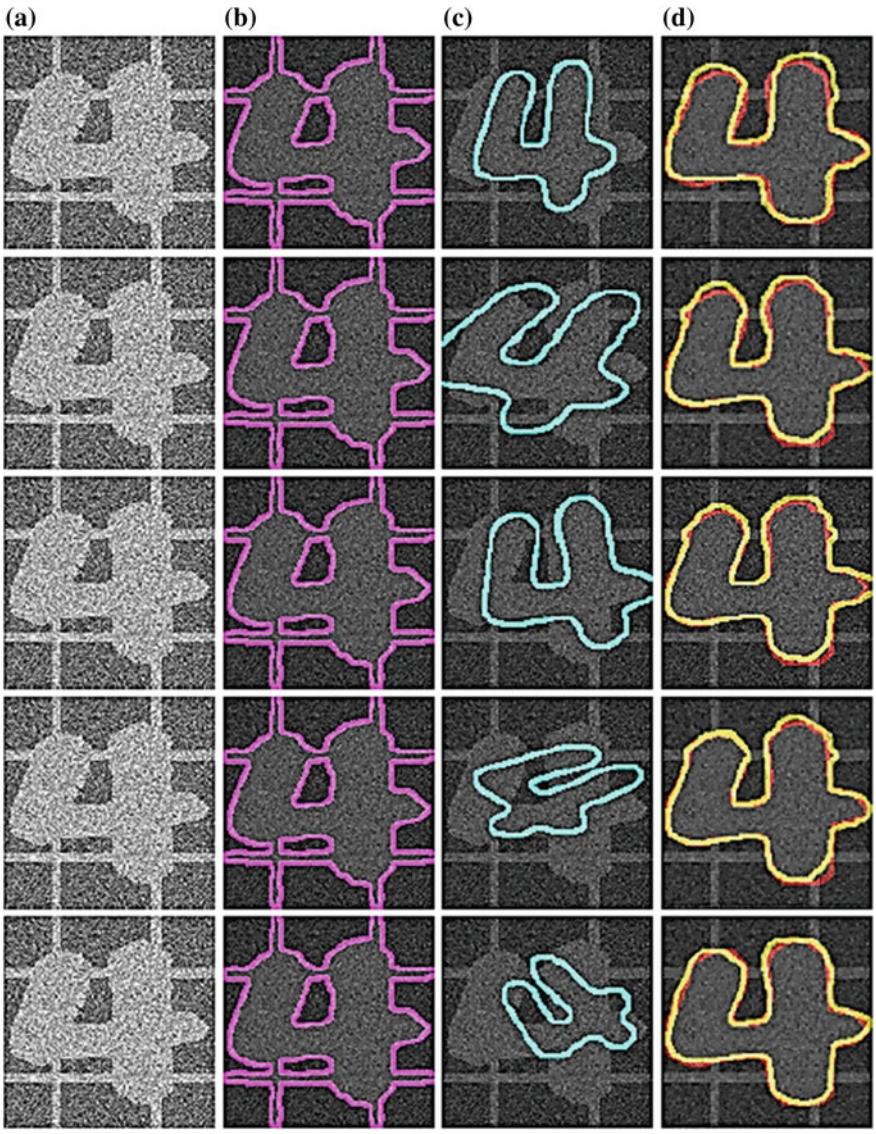


Fig. 9 Segmentation results on a synthetic number "4" with occlusions and different shape initializations. **a** the image with occlusions and noise. **b** the segmentation results using intensity only information. **c** different shape model initializations. **d** the result of the proposed method (the *red* and *yellow* colors show the contour of the ground truth shape region, and the contour of the automatically segmented region, respectively)

Additionally, Fig. 12b studies the effect of choosing the number of the projected training shapes N (by changing the chosen value of L) on the segmentation accuracy. From this figure, we can conclude that the performance of 2D-PCA is better than

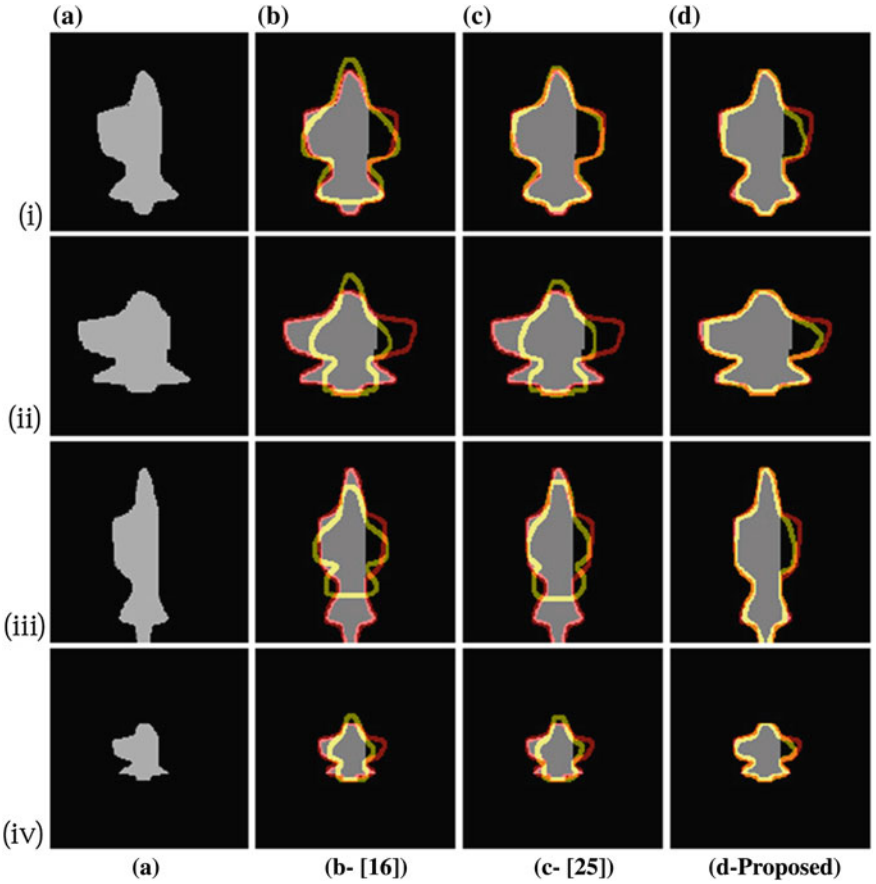


Fig. 10 Comparison with two closest works described in [11, 24]. Testing shapes with (**i**) homogeneous and (**ii**–**iv**) inhomogeneous scaling factors. (i) sx = 1.0, sy = 1.0, (ii) sx = 0.7 sy = 1.3, (iii) sx = 1.2, sy = 0.7, (iv) sx = 0.4, sy = 0.7 (the *red* and *yellow* colors show the contour of the ground truth shape region, and the contour of the automatically segmented region, respectively

the conventional PCA under the same number of training shapes. In other words, to get the same accuracy of PCA framework, the 2D-PCA needs fewer training shapes. Using the shape model, the spinal processes are eliminated automatically without any computational cost and execution time. This contribution is very important for the BMD measurements which are restricted to the VBs.

6 Summary and Possible Extensions

This chapter considered elastic registration of shapes and its applications in the segmentation problem. Shape representation was performed using the vector distance function (VDF). The energy function for global and local registrations was described.

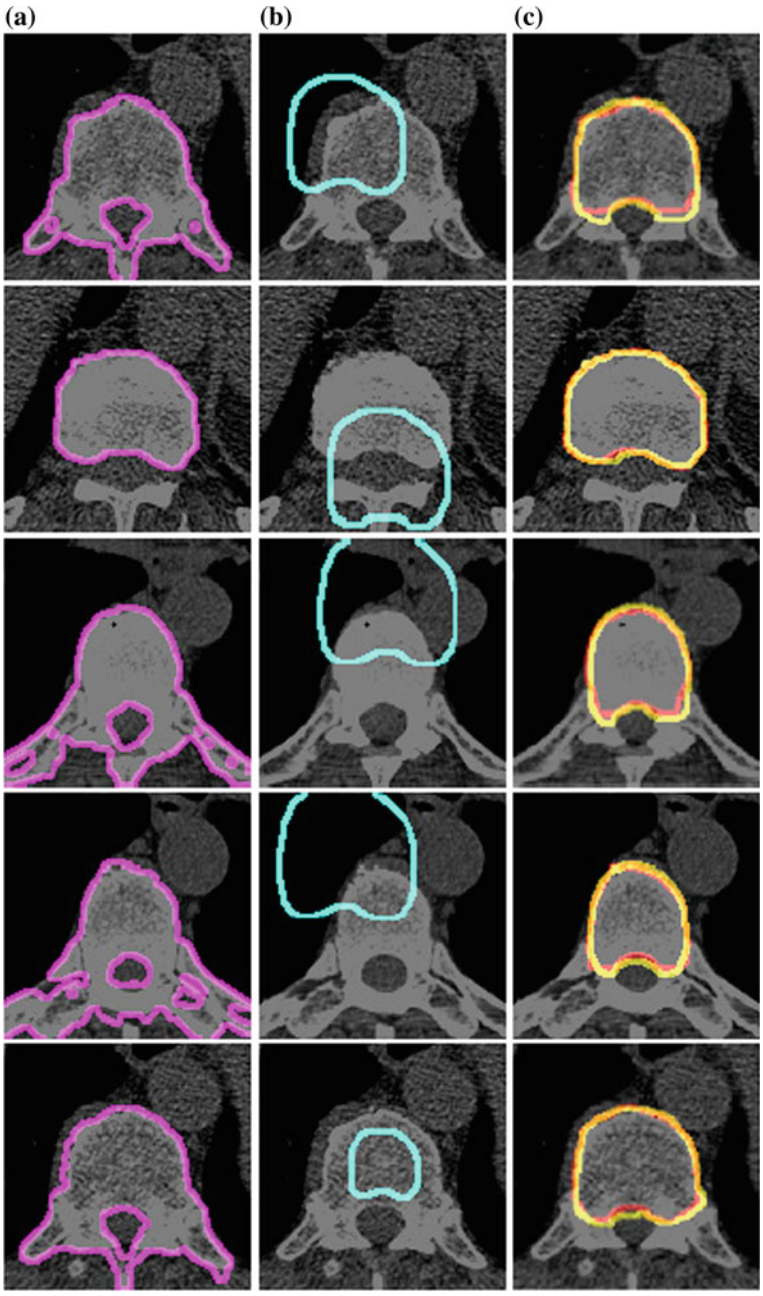


Fig. 11 Segmentation results of clinical CT images. **a** intensity only based segmentation results. **b** different initialization of the shape model. **c** the proposed segmentation results (the *red* color shows the contour of the ground truth shape region, the *yellow* color shows the contour of the automatically segmented region)

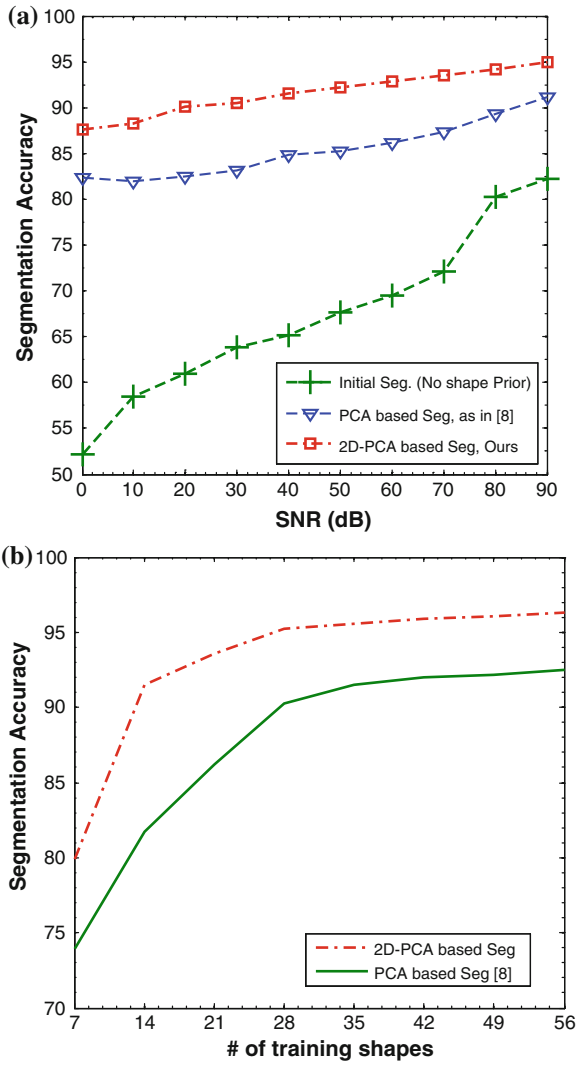


Fig. 12 **a** the average segmentation accuracy of different segmentation methods on 272 CT images under various signal-to-noise ratios. **b** the effect of choosing the number of the projected training shapes N on the segmentation accuracy

Creating a shape prior was studied for a number of examples. We have demonstrated the registration problem by matching vector implicit spaces representation of shapes. We formulated the process as an energy minimization problem. Gradient descent is used for optimizing the global registration energy with proper initialization of the transformation parameters.

The use of vector implicit representation helps generalize the global transformation and hence, results are improved. In local registration, the situation is different because the number of unknown parameters used to represent deformations is large. Gradient descent is an awkward step in this situation. We adopt a closed form solution for the elastic registration problem by formulating a quadratic function, which leads to a convex optimization system. The proposed approach avoids using large number of iterations required for the minimization by gradient descent optimization. We demonstrated several experimental results for synthetic and real shape registration cases. The proposed approach is competitive when compared to the state of the art techniques. Qualitative, quantitative, and comparative experimental results have been demonstrated for both global and local registration cases. An application for nodular region segmentation has been illustrated to assure that the proposed technique has a wide range of applications. Regarding future directions, the proposed approach can be implemented in 3D in a straightforward manner, which will help in applications like 3D face recognition.

References

1. Abd El Munim H, Farag AA (2007) Shape representation and registration using vector distance functions. In: Proceedings of IEEE conference on computer vision and pattern recognition (CVPR'07), Minneapolis, MN, June 18–23 2007
2. Abd EL Munim H, Farag AA (2007) Curve/surface representation and evolution using vector level sets with application to the shape-based segmentation problem. IEEE Trans Pattern Anal Mach Intell 29(6):945–958
3. Abd El Munim HE, Farag AA (2007) A new global registration approach of medical imaging using vector maps. In: Proceedings of international symposium on biomedical imaging, ISBI'07, Metro Washington DC, pp 584–587, April 12–15 2007
4. Abd El Munim HE, Farag AA (2007) A new variational approach for 3D shape registration. In: Proceedings of international symposium on biomedical imaging, ISBI'07, Metro Washington DC, pp 1324–1327, April 12–15 2007
5. Fahmi R, Farag AA (2008) A fast level set algorithm for shape-based segmentation with multiple selective priors. In: Proceedings of IEEE international conference on image processing (ICIP'08), San Diego, California, Oct 12–15 2008
6. Farag A, Abdelmunim H, Graham J, Farag AA, Elshazly S, Ali AM, Farag A, Al Mogy S, Al Mogy M, Falk R, Al Jafary S, Mahdi H, Milam R (2011) Variational approaches for segmentation of lung nodules. IEEE international conference on image processing (ICIP), pp 2157–2160, Sept 2011
7. Aslan MS, Ali A, Farag AA, Abdelmunim H, Arnold B, Xiang P (2011) A new segmentation and registration approach for vertebral body analysis. In: Proceedings of 2011 IEEE international symposium on biomedical imaging (ISBI), pp 2006–2009, March 2011
8. Aslan MS, Mustafa E, Abdelmunim H, Shalaby A, Farag AA, Arnold B (2011) A novel probabilistic simultaneous segmentation and registration using level set. In: Proceedings of 2011 IEEE international conference on image processing (ICIP), pp 2161–2164, Sept 2011
9. Abdelmunim H, Farag AA (2011) Elastic shape registration using an incremental free form deformation approach with the ICP algorithm. In: Eighth Canadian conference on computer and robot vision (CRV), pp 212–218, May 2011

10. Abd El Munim HE, Farag AA, Farag AA (2013) Shape representation and registration in vector implicit spaces: adopting a closed form solution in the optimization process. In: IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI'13, March 2013
11. Paragios N, Rousson M, Ramesh V (2002) Matching distance functions: a shape-to-area variational approach for global-to-local registration. In: European conference in computer vision. Denmark, Copenhagen, June 2002
12. Xie Z, Farin GE (2004) Image registration using hierarchical B-splines. IEEE Trans Visual Comput Graphics 10(1):85–94
13. Sederberg T, Parry S (1986) Free-form deformation of solid geometric models. In: ACM SIGGRAPH, pp 151–160
14. Rueckert D, Sonoda L, Hayes C, Hill D, Leach M, Hawkes D (1999) Nonrigid registration using free-form seformations: application to rreast MR images. IEEE Trans Med Imaging 8:712–721
15. Huang X, Paragios N, Metaxas DN (2006) Shape registration in implicit spaces using information theory and free form deformations. IEEE Trans PAMI 28(8):1303–1318
16. Diciotti S, Lombardo S, Falchini M, Picozzi G, Mascalchi M (2011) Automated segmentation refinement of small lung nodules in CT scans by local shape analysis. IEEE Trans Biomed Eng 58(12):3418–3428
17. Besl P, Mckay N (1992) A method for registration of 3-D shapes. IEEE Trans PAMI, 14(2):239–256
18. Kendall DG (1989) A survey of the statistical theory of shape. Stat Sci 4(2):87–120
19. Osher S, Sethian JA (1988) Fronts propogating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. J Comput Phys 79:12–49
20. Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. Int J Comput Vis 1(4):321331
21. Chan TF, Vese LA (2001) Active contours without edges. IEEE Trans Image Process 10(2):266277
22. Farag AA, Hassan H (2004) Adaptive segmentation of multi-modal 3D data using robust level set techniques. In: Proceedings of international conference on medical image computing and computer-assisted intervention (MICCAI'04), pp 143–150
23. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models—their training and application. Comput Vis Image Understanding 61:38–51
24. Tsai A, Yezzi A, Wells W, Tempany C, Tucker D, Fan A, Grimson WE, Willsky A (2003) A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans Med Imaging 22(2):137–154
25. Taron M, Paragios N, Jolly MP (2007) Registration with uncertainties and statistical modelling of shapes with variabile metric kernels. IEEE Trans Pattern Anal Mach Intell 1(8):1–14
26. Cremers D, Osher SJ, Soatto S (2006) Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. Int J Comput Vision 69(3):335–351
27. Aslan MS, Abdelmunim H, Farag AA (2011) A probabilistic shape-based segmentation using level sets. In: First IEEE workshop on information theory in computer vision and pattern recognition (ICCV workshop)
28. Mahmoodi S (2009) Shape-based active contours for fast video segmentation. IEEE Trans Signal Process Lett 16(10):857–860
29. Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. Xommun Pure Appl Math 42(4):577–688
30. Shalaby A, Aslan M, Abdelmunim H, Farag AA (2012) 2D-PCA based shape prior for level sets segmentation framework of the vertebral body. In: Proceedings of the 6th Cairo international conference on biomedical engineering (CIBEC'12), Cairo, Egypt, pp 134–137, Dec 22–25 2012
31. Kumar BVKV, Savvides M, Xie C (2006) Correlation pattern recognition for face recognition. Proc IEEE 94(11):1963–1976
32. Lagarias JC, Reeds JA, Wright MH, Wright PE (1998) Convergence properties of the nelder-mead simplex method in low dimensions. SIAM J Optim 9(1):112–147

Image Computing Based on Bayesian Models (BM)

Zhong Xue and Stephen Wong

Abstract Many medical image computing tasks apply the prior knowledge about the variability of shapes or deformations to improve the performance of shape analysis, segmentation, registration, as well as group comparison or computer-aided diagnosis. Statistical model-based algorithms play important roles in capturing such prior information and applying them for robust image segmentation and registration. Given the prior distribution of a high-dimensional data, which can be a shape description, a deformation field, or other feature vectors from the training images, the objectiveis to come up with the best estimation of the shape, the deformation, or feature vectors from an observed data/image. The traditional maximum a posteriori (MAP) framework is the most commonly used methodology to incorporate the prior information in the estimation. One example of MAP estimation is the active shape model (ASM), which encodes the prior information of object shapes using principal component analysis (PCA) and then extracts the shape of an object from the PCA model that matches the image the best. Such statistical model-based method is constrained by the prior distribution from sample data for improved robustness. However, ASM takes directly the reconstructed object shape as the matching result, and it may not be able to match a new image accurately if the variations of the shape are not presented in the sample data, or if the number of model modes has been truncated too severely. This chapter introducesa new Bayesian model (BM) for accurate and robust medical image computing. BM overcomes the limitation of MAP by incorporating an intermediate variable and jointly estimating the result and the intermediate variable simultaneously. In this way the BM framework allows for convenient incorporation of additional constraints, reduces the constraints of the prior distributions, and increases the flexibility of shape matching. In this chapter, after a brief literature review of the statistical model-based image

Z. Xue (✉) · S. Wong
Houston Methodist Research Institute, Houston Methodist Hospital, Houston, TX, USA
e-mail: zhong.xue@ieee.org

Z. Xue · S. Wong
Weill Cornell Medical College, New York, NY, USA

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_4,

analysis methods, we first introduce the BM framework in both segmentation and registration and then present our two works that apply the BM methodology. From the techniques presented we can see that image segmentation and registration can be uniformly formulated in the same BM framework, and such formulation can also easily facilitate other image computing tasks.

Abbreviations

ASM	Active shape model
BM	Bayesian model
BSM	Bayesian shape model
GM	Gray matter
MAP	Maximum a posteriori
MR	Magnetic resonance
PDE	Partial differential equation
PVA	Principal component analysis
ROI	Region of interest
SMD	Statistical models of deformations
SPM	Statistical parametric mapping
SVM	Support vector machine
TS	Temporal smoothness
WM	White matter
WPT	Wavelet packet transform

1 Introduction

Image segmentation, registration, group comparison, and classification are the most commonly used medical image computing tasks. In image segmentation, a group of voxels from the input image are highlighted (using curves or surfaces or region of interest (ROI)) to represent specific anatomical structures, e.g., shapes or volumes of tumors. Sometimes all the images voxels are partitioned into different tissue types such as MR brain image segmentation. In image registration, one image is aligned onto another globally and locally (using elastic deformations) in order to define correspondences between images. After defining the correspondences, automatic measurements or image normalization can be performed for either group comparison or automatic classification for computer-aided diagnosis. The template-based image segmentation deforms shape templates of organs or structures to match the shapes in the input images, and atlas-based image registration solves the deformation fields that match the atlas onto subject images. Therefore, they are essentially solving a similar image computing problem: deform a template (shape or image) onto input subject images, and the only difference is that they are dealing with different sizes of

voxels or feature points. ROI extraction normally involves in less number of points, while image registration considers each voxel in the image.

In the literature, many segmentation and registration algorithms had been proposed. In this chapter, we first give a brief review of them. As our goal is to introduce the uniform methodology by applying the BM in image computing, we will focus on summarizing the common methodologies used in medical image segmentation and registration, and discussing how the BM framework can be applied to them.

In image segmentation, for more than 25 years deformable models have been attracting much attention in the areas of object detection and matching because of their flexibility of adapting themselves to fit objects more closely than fixed template matching. Generally, deformable models can be classified into two classes [1]: the free-form models and the parametric or deformable template models. The free-form models, *e.g.*, active contours or snakes [2] and later level set methods [3–5], can be used to match any arbitrary shape provided some general regularization constraints, such as continuity and smoothness, are satisfied. On the other hand, the parametric models are more constrained because some prior information of the geometrical shape is incorporated [1, 6, 7]. Compared with the free-form deformable models, it has been demonstrated that the parametric models are more robust to irrelevant structures and occlusions when being applied to detect specific shapes of interest. The parametric models, such as deformable templates/models [1, 6, 8, 9], G-Snake [10] and Active Shape Model (ASM) [11, 12], encode specific characteristics of a shape and its variations using global shape model, which is formed by a set of feature parameters or well defined landmark/boundary points of that shape. A successful and versatile scheme in this field is statistics-based shape models in the maximum a posteriori (MAP) framework. In these models, the prior knowledge of the object as well as the observation statistics is utilized to define the optimal estimate. However, most of these existing parametric models encode the shape information in a "hard" manner in that the prototype contour is fixed during the matching process. As a result, only a small amount of local deformations can be tolerated.Statistical model-based segmentation such as ASM represents a typical MAP algorithm in applying prior distribution constrained methods. It encodes the prior information of object shapes using PCA and matches object by directly using the geometrical transformed version of the ASM model. Many other training algorithms had been proposed such as kernel-PCA and support vector machine (SVM) to capture nonlinear distributions.

Medical image registration is the process of finding a transformation that maps one individual image to another. It is used frequently for anatomical segmentation and labeling, for morphological analysis using shape transformations, and for spatial normalization of structural and functional data.In the literature, many registration methods have been proposed [13–22], and they aim at finding the deformation field between two images by not only maximizing the similarity between two images but also by using proper constraints on the deformation field. Therefore, the registration performance of an algorithm may be affected by the way it defines the image similarity measure and the form or the constraints it uses for the deformation field. As a result, even for the same set of data, the registration results of different algorithms are different, and they are biased toward different image similarity measures and differ-

ent deformation definitions and constraints. While the ground truth is not available, it is often difficult to evaluate which registration result is better or to provide more robust registration results.

In fact, deformable template-based segmentation and image registration found the same ground in terms of image processing. In image segmentation, the object shape is defined as a group of voxels, a shape, or a region, and the objective is to deform such a shape model so that it matches the object in the input image. Image registration can also be regarded as a special segmentation procedure, and the difference is that now the object shape is the target image or a shape that include all the image voxels. The goal is the same: deforming the shape (the template image) and find the best match between the deformed template image and the input subject image to define the voxel-wide image correspondences. Therefore, it is natural to formulate the segmentation and registration procedure using similar methodologies.

How to constrain the deformation of a shape model or a template image is one of the major tasks to define an appropriate method. Generally, there are two groups of methods. One is to use the general constraints such as smoothness constraints or topological constraints to make sure that the matching procedure is valid. Another is to apply prior knowledge in the matching procedure. Such prior knowledge about the possible variability of shapes or image deformations can be learned from a number of training samples and captured using statistical models, and the statistical model can then be used to constrain the matching procedure. Typical examples are the ASM and statistical model-based registration mentioned above.

Statistical models play important role in capturing such prior information [23]. Given the prior distribution of a high-dimensional data, which can be shape description, deformation fields, and other feature vectors from the training images, the objective of knowledge-based estimation is to come up with the best estimation of the shape, deformation, or feature vectors from an observed data/image. The MAP framework is the most commonly used methodology to incorporate the prior information. However, ASM takes the reconstructed object model as the matching results, and it may not be able to match a new image accurately if the variations of the shape are not presented in the sample data, or if the number of model modes has been truncated too severely.In addition, representing prior statistical knowledge of high-dimensional scalar or vector fields is of fundamental importance in a variety of scientific areas including computational anatomy, shape analysis, pattern recognition, and hypothesis testing applied to images or their deformations [8, 12, 24]. For instance, statistical study of deformations can be used to provide voxel-based morphological (VBM) characterization of different groups [25]; to incorporate prior knowledge of deformations from training samples into image segmentation and registration algorithms [17]; to provide an efficient way of synthesizing new deformation fields for validation of registration and segmentation methods [26]; to regularize deformations according to prior knowledge of sample deformations; and to estimate the missing parts of a deformation from parts that are observed. In fact, all these applications and the plethora of automated methods for deformable registration of brain images have necessitated the construction of a statistical model that effectively captures the

prior distribution of high-dimensional deformation fields, in order to represent the true and full range of anatomical variability.

We have proposed a new Bayesian shape model (BSM) for accurate object matching that overcomes the shortcoming of the ASM model that over-constrains the shapes using the shape prior distribution [27]. We refer the framework as the Bayesian model (BM) because in this chapter the segmentation and registration will be presented in a uniform way and we are not only dealing with shapes. BM overcomes the limitation of MAP by incorporating an intermediate variable and jointly estimating the results and the intermediate variable. The BM framework allows for convenient incorporation of additional constraints to the medical image computing tasks and can be applied to image segmentation, registration, and group comparison.

The organization of this chapter is as follows. First, we introduce the traditional MAP framework and then present the BM method that incorporates an intermediate parameter for estimation. Then, we introduce the recent works on how BM framework can be applied for medical image segmentation and registration.

2 The Maximum A Posteriori (MAP) Framework

In many image segmentation and registration works, given the input data D, which can be the input image or input template and subject images, the objective is to find a shape that matches the objects in the input image or a deformation field that matches the two images. Based on the MAP framework, such a problem can be described as maximizing the following posterior distribution [28]:

$$P(S|D) = \frac{P(D,S)}{P(D)} = \frac{P(D|S)P(S)}{P(D)} \tag{1}$$

The optimal shape can be estimated by:

$$S^* = \text{argmax}\{P(S|D)\} = \text{argmax}\{P(D|S)P(S)\} \tag{2}$$

where $P(D) = 1$ because the input data D is known already. $P(D|S)$ is the conditional probability of obtaining data D given the shape S, and $P(S)$is the prior distribution of shape S. So the solution of Eq. (2) is to find the best shape S^* so that it not only gives higher probability of the observed data D but also subject to a higher probability according to the prior distribution of the shape.

When the probabilities are described by the Gibbs distribution [29, 30], the MAP problem is equivalent to minimizing the energy function:

$$E_{\text{MAP}} = E_{\text{ext}}(D|S) + E_{\text{int}}(S), \tag{3}$$

where $E_{\text{ext}}(D|S)$ stands for the external energy term that reflects the matching degree between shape S and the input image D. One example in deformable shape

matching is that the external energy term is defined as the matching degree between the shape (curve) and the object boundary features (such as boundaries) extracted from the image. $E_{\text{int}}(S)$ is the internal energy term that regularizes the shape itself. If the shapes from different subjects vary largely, normally the constraints are defined as the smoothness or topological regularizations. If the shapes from different subjects are similar but are subject to local elastic changes, a statistical shape model can be used to capture the variability of the shapes from a number of training samples, and then such a shape distribution can be used as the regularization term (internal energy) in Eq. (3). In many medical image computing tasks, because the shapes of organs are similar cross different subjects unless there exist some pathological conditions, statistical model-based methods play an important role in regularizing such shape variability. Therefore, many deformable model methods and statistical atlas based deformable registration algorithms fall in this MAP formulation. In deformable models, different internal and external energy/force terms are defined for automatic object matching.

In image registration, we can use different notes to describe the MAP method, but the formulation remains the same. Given two images, the template I_T and the subject I_S, the goal is to find the deformation field that aligns the two images. According to MAP, the goal is to find that $\mathbf{f}$ maximizes the following posterior distribution:

$$P(\mathbf{f}|I_T, I_S) = \frac{P(\mathbf{f}, I_T, I_S)}{P(I_T, I_S)} = \frac{P(I_T, I_S|\mathbf{f})P(\mathbf{f})}{P(I_T, I_S)} \tag{4}$$

Similarly, using the Gibbs distribution, the MAP problem is equivalent to minimizing the energy function:

$$E_{MAP} = E_{\text{sim}}(I_T, I_S|\mathbf{f}) + E_{\text{reg}}(\mathbf{f}), \tag{5}$$

where $E_{\text{sim}}(I_T, I_S|\mathbf{f})$ stands for the image distance measure (the reverse of image similarity measure) between the two images under the current deformation field $\mathbf{f}$. $E_{\text{reg}}\mathbf{f}$ is the regularization term of the deformation field.

Comparing Eqs. (3) and (5), we can see that they are actually the same kind of formulation, and we use different names for the sake of different conventions in segmentation and registration. The difference is that now in image registration, the image similarity energy term calculates the differences between the two images, such as the sum of voxel-wise intensity differences or mutual information of the subject image and the deformed template image using $\mathbf{f}$. Similar to the internal energy, the deformation field regularization term here can use either smoothness and topological regularization or can be derived from a statistical model if the template image is fixed, and multiple images need to be registered onto the same template.

3 The Bayesian Model (BM) Framework

One drawback of the MAP is that it might over regularize the shapes during the matching procedure. The reason is that from the formulation we can see that the final

result must be exactly subject to the prior distribution. In another word, the MAP method requires that the resultant shape is subject to the prior distribution. This renders that it may not be able to match a new image accurately if the variations of the shape are not presented in the sample data. For example, if a PCA model is used, because of the limited numbers of principal components used, the shape reconstructed from PCA model cannot reflect the detailed variability of the high-dimensional data.

It is in this context that we proposed a Bayesian shape model (BSM), which is also referred to as Bayesian model (BM) in this chapter. In BM, an intermediate shape is introduced in the matching procedure. In this section, the BM is first presented in the context of shape-based segmentation, and then, its application to image registration is introduced.

Given the input image or data D, our goal is to jointly estimate the resultant shape S and an intermediate shape $\bar{S}$ by maximizing the following posterior probability [27, 31]:

$$P(S, \bar{S}|D) = \frac{P(D, S, \bar{S})}{P(D)} = \frac{P(D|S, \bar{S})P(S|\bar{S})P(\bar{S})}{P(D)} = \frac{P(D|S)P(S|\bar{S})P(\bar{S})}{P(D)} \tag{6}$$

It can be seen that by using an intermediate shape $\bar{S}$, we can decouple the two requirements in the matching procedure. Namely, we require that the resultant shape $\bar{S}$ match the image data and the intermediate shape match the requirement of prior distribution. So these two shapes do not need to be the same as in the ASM model. The middle term requires that the two shapes are close with each other. This is the advantage of BM, and on one hand, the prior shape constraint is applied and on the other hand, the formulation allows for flexibility to better match the object.

The final derivation in Eq. (6) is based on the assumption that the data D and the intermediate shape $\bar{S}$ are independent each other. This may not be exactly true but in practice if we are not requiring a match between the intermediate shape $\bar{S}$ and the input data D, it is a valid assumption in the formulation. Using the Gibbs distribution, the problem is equivalent to minimizing the energy function:

$$E_{BSM} = E_{\text{ext}}(D|S) + E_{\text{int}}(S|\bar{S}) + E_{\text{con}}(\bar{S}), \tag{7}$$

where $E_{\text{ext}}(D|S)$ stands for the external energy that matches the shape S with the image $E_{\text{int}}(S|\bar{S})$, represents the distance between the two shapes, and $E_{\text{con}}(\bar{S})$ is the constraint energy term that regularizes the intermediate shape $\bar{S}$. We can use the smoothness constraints or statistical model constraints to define $E_{\text{con}}(\bar{S})$.

The key feature for BM is that now we are solving a shape S that on one hand matches the image data D, and on the other hand, to be as close as the intermediate shape $\bar{S}$, which is regularized by the constraint energy term. In this way, the resultant shape S is not necessarily the exact regularized shape according to the shape priors. Such a formulation allows for accurate matching of the object shape, while still be regularized by the prior information.

We now translate the BM into image registration applications. According to the BM framework, the goal of image registration is to solve jointly the deformation

field $\mathbf{f}$ and an intermediate deformation $\bar{\mathbf{f}}$ field by maximizing the following posterior distribution [27]:

$$P(\mathbf{f}, \bar{\mathbf{f}}|I_T, I_S) = \frac{P(\mathbf{f}, \bar{\mathbf{f}} I_T, I_S)}{P(I_T, I_S)} = \frac{P(I_T, I_S|\mathbf{f}, \bar{\mathbf{f}})P(f, \bar{\mathbf{f}})P(\bar{\mathbf{f}})}{P(I_T, I_S)} = \frac{P(I_T, I_S|f)P(\mathbf{f}, |\bar{\mathbf{f}})P(\bar{\mathbf{f}})}{P(I_T, I_S)}, \tag{8}$$

Similarly, using the Gibbs distribution, the BM problem is equivalent to minimizing the energy function:

$$E_{BM} = E_{\text{sim}}(I_T, I_S|\mathbf{f}) + E_{\text{reg}}(\mathbf{f}|\bar{\mathbf{f}}) + E_{\text{con}}(\bar{\mathbf{f}}) \tag{9}$$

where $E_{\text{sim}}(I_T, I_S|\mathbf{f})$ stands for the image similarity measure, $E_{\text{reg}}(\mathbf{f}|\bar{\mathbf{f}})$ is the regularization of $\mathbf{f}$, and $E_{\text{con}}(\bar{\mathbf{f}})$is the constraint energy of the deformation field. $E_{\text{con}}(\bar{\mathbf{f}})$ can be modeled using statistical models of the training deformations defined on the same template image space.

In fact, by comparing Eq. (5) with (9), we can see that the use of intermediate deformation $\bar{\mathbf{f}}$ in the BM method relaxes the direct regularization of the deformation field and increases its flexibility. This formulation allows for more accurate registration of medical images. If no statistical models are used, e.g., only using smoothness constraints, such an intermediate matching procedure may not be necessary and the MAP method may perform similar to the BM. Because for high dimensional shape and deformation data, the generalization and specificity of statistical models may be limited, requiring the result to be exactly subject to the prior distribution could be loosed by using the BM method.

In the next two sections, we introduce the applications of the BM framework in medical image segmentation and registration. First, in Sect. 4, we introduce an atlas-based level-set segmentation for MR brain images, where the BM framework has been applied to use the statistical models of level-set functions in the atlas space for constraining the segmentation of anatomical structures in subject images. This comes up with a new joint parametric and nonparametric segmentation [32, 33]. Then, in Sect. 5, we introduce a statistical model of deformation (SMD) constrained image registration algorithm [19, 34]. Although with different applications, both algorithms apply the same basic methodology presented in this Section.

4 BM for Medical Image Segmentation

4.1 Introduction

In this chapter we introduce how to apply BM in level-set-based MR brain image segmentation. Image segmentation or automatic extraction of anatomical structures plays an important role in medical image analysis. In the literature, various

deformable models or partial differential equation (PDE)-based methods have been proposed to extract the anatomical structures of interest in medical images. These model scan also be regarded as model fitting problems in a probabilistic framework [7] based on the Bayes' theorem, where the input image is divided into different regions by matching the external and internal hidden features so that the posteriori probability of the segmentation is maximized.

The internal features reflect the geometric properties of the evolving curve such as curvature, curve-length, and smoothness measures. They are represented by either explicit or implicit curved descriptors. An unjustified property of explicit descriptors is that geometrical evolution of a curve is not intrinsic since it is dependent on the way the curve is parameterized. Implicit curve descriptors are generally parameterized by arc-length [33, 35, 36]. One of the typical methods using implicit curve descriptor, known as the level set method [37], has become very important in image segmentation due to its advantages such as parameter-free representation of curves and easy handling of topology changes.

The external features are traditionally represented by image specific information such as image gradient [2, 33, 35, 36, 38], or regional information [3, 39]. Particularly, in boundary-based approaches with an implicit curve representation [33, 36] curve evolution is achieved by defining a geodesic curve of minimal weighted length and a boundary-based metric from image data. However, these methods are sensitive to noise or images with poor contrast, and they often stuck in local minima since the objective functions rely on boundary-based external features and there is no global shape constraints about the object shape to be matched.

On the other hand, region-based external features are defined as regional image statistics inside and outside the contours and are more stable than boundary-based features. Mumford and Shah [40] proposed a new energy function in the variation framework. The aim is to find smooth regions as optimal piecewise approximation of input image and with sharp boundaries by minimizing the energy function. In order to simplify the optimization process, Mumford and Shah's energy function is used in greedy algorithm based on the region growing and merging [41] or in statistical framework [39].

Another solution was proposed by Chan and Vese [3] to provide an efficient variation formulation in level set representation for minimizing the Mumford–Shah energy function [42–44]. In curve evolution-based methods, regional image statistics are obtained based on the parametric statistics [39, 43, 44]. However,the performance of parametric methods can be depended on the assumed parametric models and the class of input images. Therefore,nonparametric methods such as Parzen windowing are used to estimate the underlying distribution of the pixel intensities within each region [23, 41, 45]. Another way to improve region-based methods is to integrate prior knowledge into image-based segmentation function to process globally high-level information of object shapes. This knowledge is trained as the object-oriented model based on using learning procedures, for example on space of implicit function [4, 46–48], Fourier-driven curve representation [8, 33], and parametric space of curve representation [11, 12, 15].

One important property of the model-based approach is that the variability of object shapes is captured from training samples and acts as the shape priors. According to level set representation,the shape prior information can be used on both linear subspace of parametric level set function spanned by principal components and nonlinear space of nonparametric level set function. In parametric curve evolution on linear subspace, Tsai et al. [5] proposed to parameterize the level set energy function with respect to shape and pose parameters. Curve evolution is performed parametrically via updating the shape and poses parameters constrained by the statistical parametric model obtained using Principal Component Analysis (PCA). In nonlinear space of evolving curve, Leventon et al. [49] used PCA directly on the space of implicit functions to estimate principal modes of shape variation of training samples. This prior information constraints and regularizes the evolving curve derived separately by image specific energy term in [36]. In order to couple segmentation and prior model, Rousson [48] proposed a statistical prior model in variational form and included an average shape and mode of variation defined from training samples. This statistical model is integrated with an existing data-driven variational method to extract the objects. One can instead estimate and impose the shape prior information on the zero-crossing rather than on the level set function using variational framework. For promising extensions, based on the level set representation in statistical shape prior models, Charpiat et al. [50] proposed nonlinear shape metrics and estimated the principal shape variations.

Cremers et al. [51] proposed nonlinear statistical shape model by performing kernel PCA along level set-based Mumford–Shah segmentation [40]. In [52] and [4], nonparametric density estimation forshape priors is assimilated in the level set framework to extract object. In summary, the parametric statistical model-based curve evolution enforces the constraints about the underlying shapes by the shape prior and therefore the algorithm is robust and less affected by noise or low image contrast. On the other hand, the nonparametric evolution energy term can handle infinite degree of curve variations and helps us to achieve better accuracy than the method that only uses the parametric statistical models.

Recent works by [53] and [46] also use similar statistical model in level set method. These approaches directly apply a PCA model to constrain the level set function while matching it with the image features. Thus, the level set function is more constrained by the prior distribution. The optimization procedure of the curve is achieved using the calculus of variation on the nonlinear space of nonparametric level set function.

As described in the previous section, BM facilitates a joint estimation of the resultant shape and an intermediate shape simultaneously. Thus, we can integrate the advantageous of both parametric and nonparametric curve evolution using in the BM framework. The core of such joint curve evolution is a unified energy function that drives the active contour toward the desired boundaries through two steps. In the first step, the curve is driven by the nonparametric level set function toward a homogeneous intensity regions and constraint by the parametric shape-based model. Therefore, minimization process of this step is achieved on nonlinear space of level set function. In the second step, the curve is optimized by parametric level set func-

tion and is driven toward the global shape prior and image regions parametrically and constraint by nonparametric model from previous step. Optimization of the second step is achieved on linear subspace spanned by principal components obtained in statistical shape-based model. Therefore,in the joint curve evolution, both local features such as statistical region properties of the image, curvature and length of the curve and global features of the object such as shape and position parameters are updating respectively through these two steps. In this way, two curves are used to match the object shape in an iterative manner. They are jointly constrained via a similarity measure between them in each iterative evolution. Finally, the nonparametric level set curve is regarded as the final matching result since it is more accurate as compared to the parametric curve.

In experiments, the proposed algorithm has been applied to segment the frontal horns of ventricles and putamen of MR brain images. The reason that we choose these two structures is that the boundaries of the ventricle frontal horn are very clear, while the white matter/gray matter contrast of putamen boundaries is lower compared to other anatomical structures of the brain. Thus we can fully evaluate the performance of the algorithm under different conditions. The comparative results have shown that the proposed joint curve evolution is as robust as Tsai et al. method and yields more accurate results as compared with the parametric only statistical model [5] and nonparametric model [3] by using manually marked curves as the gold standard.

4.2 Previous Methods

Before introducing the joint BM segmentation method, we briefly introduce the nonparametric and parametric curve evolution algorithms that the joint curve evolution algorithm is built on.

Nonparametric Curve Evolution

Chan and Vese proposed a nonparametric curve evolution method by solving the Mumford Shah problem, called minimum partition [3]. From the variational analysis point of view, the basic idea of the algorithm is to find the level set function ϕ a given image I by minimizing the energy function,

$$E = \int_W (I(\mathbf{X}) - C_W)d\mathbf{x} + \int_{\Omega\setminus W} (I(\mathbf{x}) - c\setminus W)d\mathbf{x} + \nu \int_\Omega |\nabla H(\phi)|d\mathbf{x}, \quad (10)$$

where υ is a constant. The shape ϕ divides the image domain Ω into two homogeneous regions W and $\Omega\setminus W$. W represents object region and $\Omega\setminus W$ is the region outside the object. C_W is the average intensity of image within region W, and $C\setminus W$ is the average intensity within region $\Omega\setminus W$. $H(\phi)$ is the level set function and ϕ is the shape represented by the zero level set. It can be seen that this nonparametric

curve evolution algorithm uses region-based image features and infinite degree of curve variation for image segmentation.

Parametric Model-Based Curve Evolution

In order to improve robustness, many statistical model-based methods have been proposed to constrain the evolution of level set functions. Tsai et al. [5] integrated parametric statistical shape modeling with the region-based curve evolution to drive the curve by using the statistical shape model as constraints. The shape variability is estimated by performing PCA on the globally aligned level set functions, and the major shape variation is then statistically reflected by the changes along the principal components. Subsequently, a parametric level set function Φ can be described as a function of the feature vector in the PCA space, $\mathbf{w}$, and the pose parameter $\mathbf{p}$, reflecting a global transformation from the shape space onto the image space (see Eq. (12) for details). Then, the associated energy functions (similar to Eq. (10)) can be defined as follows,

$$E = S_W^2/A_W + S_{\Omega\backslash W}^2/A_{\Omega\backslash W} \tag{11}$$

Notice that the parametric level set function ϕ is determined by $\mathbf{w}$ and $\mathbf{p}$, and it divides the image domain into two regions: inside $W = \{\mathbf{V}\epsilon\phi(\mathbf{V}) < 0\}$ and outside $\Omega\backslash W = \{\mathbf{V}\epsilon\phi(\mathbf{V}) > 0\}$ the zero level set curve ϕ, the areas of these two regions are A_W and $A_{\Omega\backslash W}$, and S_W and $S_{\Omega\backslash W}$ are the total intensities of these regions.

Notice that the energy functions in Eqs. (10) and (11) are similar, but the difference is that the zero level set curve in nonparametric curve evolution means the shape ϕ is defined by a free-form level set, while in parametric-model based curve evolution, the shape is defined by a parametric model $\Phi[\mathbf{W}, \mathbf{P}]$, and will be detailed in the following section.

Statistical Models of Level Set Functions

The PCA-based statistical model of the level set functions can be trained as follows. First, all the N sample images are globally aligned onto the standard shape space, and according to PCA, a parametric level set function can be represented by,

$$\Phi[\mathbf{W}, \mathbf{P}] = \bar{\Phi} + \Sigma_{k=1}^{K} W_k \varphi_k(\mathbf{p}) \tag{12}$$

where $\bar{\Phi}$ is the mean level set function, and φ_k is the kth eigenvector of the covariance matrix of the $\mathbf{N}$ samples. The corresponding K largest eigenvalues are denoted as $\psi_k, k = 1, \ldots N$.

It is worth noting that other statistical modeling methods can also be used to model the distributions of level set functions, shapes, or deformations. In this chapter, we use PCA as an example. In general, it has been proved that nonlinear modeling such as kernel-PCA or SVM performed better in capturing the detailed variability of shapes.

These statistical models are not introduced in this chapter, because our focus is to present a uniform BM framework for both segmentation and registration.

4.3 Joint Curve Evolution Algorithm

As stated in the introduction, nonparametric level set curve evolution can match image boundaries accurately but often suffer from local minima, and parametric level set curve evolution is more robust by using statistical shape models but might be less accurate in matching object boundaries. Using the proposed BM framework, we can estimate both the parametric and nonparametric shapes simultaneously, so that the level set evolution results can be not only accurate but also robust.

Joint Curve Evolution Using the BM Framework

If we denote the parametric shape parameter as $\mathbf{w}$ and the nonparametric shape as the level set function ϕ, according to Eq. (6), is S replaced by the nonparametric shape ϕ, and $\bar{S}$ is replaced by the shape parameter $\mathbf{w}$. Then, the joint estimation can be formulated by following Eq. (7), as follows:

$$(\hat{\phi}, \hat{\mathbf{w}}) = \text{argmax}\{P(I|\phi)P(\phi|\mathbf{w})P(\mathbf{w})/P(I)\}. \tag{13}$$

Using the Gibbs distribution, the energy function can be defined as:

$$E(\phi, \mathbf{w}) = E_{\text{ext}}(I|\phi) + \lambda_{\text{int}} E_{\text{int}}(\phi, \mathbf{w}) + \lambda_{\text{con}} E_{\text{con}}(\mathbf{w}) \tag{14}$$

where λ_{int} and λ_{con} are the weighting coefficients for the energy internal and the constraint energy terms. The first energy term reflects the matching degree between ϕ and the image data. , and it drives the nonparametric curve toward the object in the image. Therefore, we use Eq. (10) for calculating $E_{\text{ext}}(I, \phi)$.

The second energy term of Eq. (14) E_{int}, reflects the difference between the parametric and nonparametric curves. And it can be defined as

$$E_{\text{int}}(\phi|\mathbf{w}) = \int_{\Omega} (H(\phi(\mathbf{x})) - H(\Phi(\mathbf{w})))^2 d\mathbf{x}, \tag{15}$$

where $H()$ is the signed distance map for a given shape. The reason to use the differences of the signed distance map to define the shape differences is because that the prior of the parametric shape Φ is defined on the signed distance maps,and the nonparametric Φ shapes are represented by level set functions. The third energy term of Eq. (14), E_{con}, constrains the variations of the parametric shape according to the prior distribution trained from sample shapes using PCA according to Eq. (12).

Curve Evolution with Pose Invariance

The above formulation does not consider the global transformations among the training samples. In fact all the training samples can be globally aligned into a standard shape domain before training the statistical model so that no variations of the position of images have been included in the statistical model. Therefore, in this subsection, by considering the pose parameter **p** between the standard shape domain and the image domain, a parametric curve can be represented by (**w, p**) and thus we augment the proposed Bayesian formulation in Eq. (14) as,

$$E(\phi, \mathbf{w}, \mathbf{p}) = E_{\text{ext}}(I|\phi) + \lambda_{\text{int}} E_{\text{int}}(\phi|\mathbf{w}, \mathbf{p}) + \lambda_{\text{con}} E_{\text{con}}(\mathbf{w}, \mathbf{p}). \tag{16}$$

Minimization of the Energy Function

The energy function of Eq.(15) can be minimized by iteratively updating (**w, p**) and Φ using gradient descent method:

- Step 1: Calculating **w**. First the parametric vector **w** can be updated by assuming that **p** and ϕ are known. Since nonparametric contour ϕ is not related to **w**, the partial derivatives involve in optimization of Eqs. (12) and (15).
- Step 2: Updating **p**. The pose information **p** is updated by assuming that **w** and ϕ are known and fixed, and the gradient of E with respect to **p** can be calculated as from Eqs. (12) and (15).
- Step 3: Updating ϕ. The nonparametric shape ϕ can be updated by fixing **w** and **p**, and the gradient of E can be calculated from Eqs. (10) and (15).

4.4 Results

In order to evaluate the performance of the proposed joint curve evolution algorithm, in this section, we applied the algorithm to extract the ventricle frontal horns and putamen from MR brain images. The reason that we choose these two structures is that the boundaries of the ventricle frontal horn are very clear, while the white matter/gray matter contrast of putamen boundaries is lower compared to other anatomical structures of the brain. Thus we can fully evaluate the performance of the algorithm under different conditions. Prior to the experiments, corresponding slices from different images are selected and the ventricle frontal horn and putamen shapes are manually marked to act as both ground truth and training samples of the statistical models.

Altogether, we used 14 ventricle images and 13 putamen slices in the experiments, and the pixel spacing of the images is 0.9375 mm $\times$ 0.9375 mm. These 2D images were selected by first globally registering all the 3D images to one randomly selected image and then choosing the same slice position for all the images. Leave-one-out strategy is used for training and testing the algorithms. In each iteration of the cross-validation, the MR brain image of one subject is left out from the training data that are

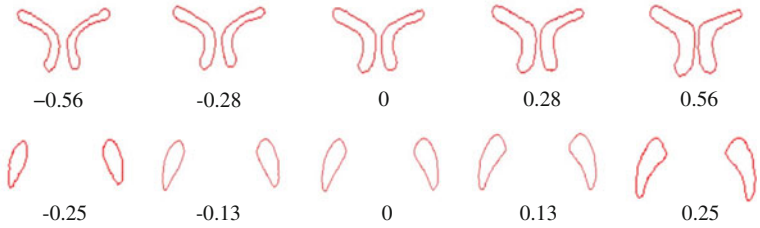


Fig. 1 Shapes by varying the weights for the first principal component. The values shown are the corresponding weights [32]

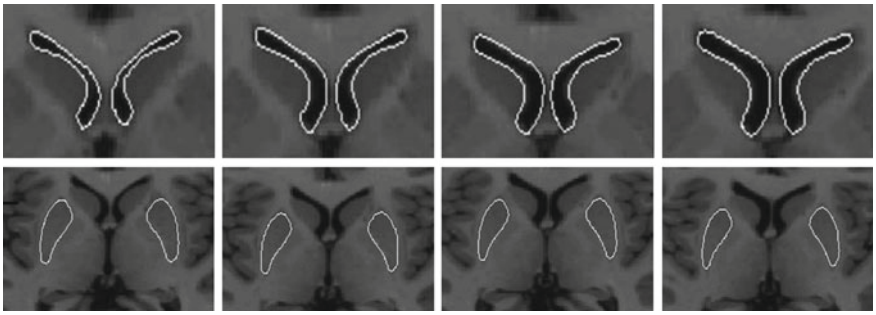


Fig. 2 Sample segmentation results for the putamen and the ventricle frontal horns of MR images by using the joint curve evolution algorithm with leave-one-out training strategy [32]

used for training the statistical model, and is tested by using the joint curve evolution algorithm trained on them. The validation iterates until all the images are left out once and only once for testing. The average distance between the resultant curve and the ground truth is then recorded, reflecting the accuracy of the segmentation algorithm with respected to that testing image.

Figure 1 shows the shape changes by varying the weights of the first principal component, trained using PCA model. Figure 2 illustrates the sample results for matching the putamen and the ventricle frontal horns using our methods. The initialization is done as follows. We manually indicate the location of the objects (i.e., the visual estimate of the object center), and then rigidly shift the mean shape to that location, acting as the initial curve. We performed more than 3 initializations for each image, and the errors for manual estimation of the object center are within the range of $[-7, 7\,\text{mm}]$, and the proposed algorithm obtained satisfactory results for all the tests with different initializations. The results also indicate that the proposed algorithm is not only effective for shapes with highly distinct boundaries such as the ventricles, but also satisfactory with low contrast shapes such as the putamen.

We also evaluated the object matching performance by quantitatively comparing the results of the proposed joint curve evolution and the parametric-only curve evolution with the manual ground truth. For comparison purposes, the same leave-one-out strategy was also used for Tsai method.

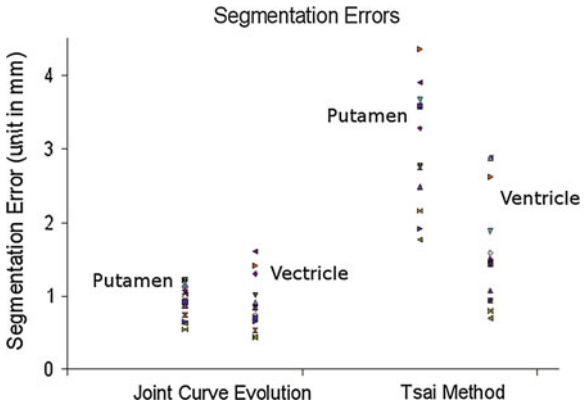


Fig. 3 Comparison results for extracting the shapes of the ventricle frontal horn and putamen [32]

Figure 3 plots the segmentation errors for matching both putamen and ventricles in all the testing images. It can be seen that the matching errors of the proposed joint evolution algorithm are much lower than those of Tsai method. For putamen, the average and standard deviation of our algorithm and Tsai method are 0.95 mm/0.24 mm and 3.06 mm/0.81 mm, respectively; for ventricle segmentation, they are 0.89 mm/0.37 mm and 1.59 mm/0.76 mm, respectively.

Obviously we obtained better segmentation accuracy for ventricles than putamen, which is because that the boundaries of ventricles are much clearer. In summary, both visual and quantitative results indicate that the proposed joint curve evolution algorithm is not only robust but also more accurate than the parametric curve evolution method for medical image segmentation.

4.5 Concluding Remarks

We have proposed a joint curve evolution algorithm for medical image segmentation. In this algorithm, both parametric and nonparametric curve evolution methods are employed based on the Bayesian framework to extract object shapes using level set functions. As a result, the algorithm has the advantages of both parametric and nonparametric curve evolution. It is as robust as the statistical model-based parametric curve evolution algorithm and at the same time, yields more accurate segmentation results. In the future work, we will extend our algorithm to match various anatomical structures in 3D medical images and improve the algorithm so that it can be used to segment multiple objects simultaneously.

5 BM for Group-Wise Registration

5.1 Introduction

As introduced in Sects. 2 and 3, atlas-based image registration uses one fixed template image for normalizing a large number of subject images for group analysis or labeling. In order to improve the robustness of the registration statistical models of deformations (SMD) that captures the variability of deformations from sample images can be used to constrain deformable registration. Although low-dimensional statistical models, such as active shape and appearance models have been successfully used in statistically-constrained deformable models, constraining of high-dimensional warping is a more challenging task, since conventional PCA-based statistics are limited to capture the full range of anatomical variability.

In this Section, we first introduce an SMD that is built upon the wavelet-PCA model and then uses it to constrain the deformable registration, wherein the template image is adaptively warped based on SMD during the registration procedure. Compared to the original template image, the adaptively deformed template image acts as an intermediate image that is formulated in the BM framework, and it is more similar to the subject image, e.g., the deformation is relatively small and local, and it is less likely to be stuck in undesired local minima.

Conventional deformable registration methods [14–16, 54] aim to find a deformation field between two images by maximizing the image-similarity measure and simultaneously constraining/regularizing the deformation field according to various deformable models. When training samples are available, statistical models that capture the variability of these deformations can be utilized to constrain the registration procedure in order to obtain more robust registration results [17]. Nevertheless, a good statistical model must effectively limit the searching space of deformations and, at the same time, accurately capture complex nature of deformation fields.

The popular principal component analysis (PCA)-based algorithms [12, 24, 55] can be used to capture the variability of deformations using the principal modes of shape variation. However, they fail when applied to deformation gelds, due to undertraining in practical settings, e.g., high dimensionality and small training samples. Different statistical algorithms have been proposed to deal with these problems, among which the wavelet-PCA model [32], applying PCA model in each wavelet band of deformation fields, has been approved to be more accurate and effective for estimating probability distribution functions (*pdf*s) of deformations. To alleviate possible discontinuity or some negative Jacobian values that could be generated by the wavelet-PCA model, we proposed the SMD [23], which uses additional constraints to regularize deformation fields, including wavelet-PCA models of both deformations and their Jacobian determinants coupled with a Markov random field (MRF).

We show that the proposed statistically-constrained deformable registration is more robust and accurate than the conventional registration. Because the HAMMER [15] registration program is used in this Section, we referred to the new registration method under the BM framework as SMD+HAMMER [19, 34].

This Section proposes a statistically-constrained deformable registration, using the Bayesian framework formulation. The basic idea is to adaptively deform the template image according to the statistics of deformations (i.e., SMD) and register the input subject image and the deformed template image by using a conventional algorithm. Since the intermediate deformed template image is more similar to the input subject image than the original template image, the registration between them is much easier, due to relatively small and local deformations between these two images.

5.2 Statistical Model of Deformations (SMD)

Denoting **f** as the deformation field defined over the template image domain, the goal of SMD is to estimate its *pdf*, *i.e.*, $p(\mathbf{f})$, from a set of training samples. Generally, compared to the number of dimensions the number of samples is way smaller. In order to capture finer and more localized variations of **f**, SMD follows and extends the wavelet-based PCA model. The wavelet-PCA model decomposes using the wavelet packet transform (WPT) and subsequently captures within-scale statistics via PCA in each wavelet band. The fundamental assumption in wavelet-PCA is that the wavelet based rotation renders the covariance matrix of deformation close to block-diagonal, thereby enabling a more accurate estimation of the statistical distribution in each block (wavelet band) from a limited set of examples, compared to the usual sample covariance estimation, due to both of lower dimensionality and relatively strong correlations among variables.

Ideally, if the wavelet-PCA model captures the statistics of deformation accurately, we can just use it as the statistical model; however, the assumption that the covariance matrix of is block-diagonal in the wavelet packet basis does not hold exactly. Although it is well-known that for broad classes of signals, correlations across scales diminish rapidly, they are nonetheless non-negligible for adjacent scales. In order to alleviate this problem, we observe that additional constraints imposed on the deformation fields can be used to define subspaces in which the deformation must belong to.

Therefore, SMD requires that a valid deformation field belongs to the intersection of the following three subspaces:

- The first subspace: The Wavelet-PCA model applied to the sample deformation fields.
- The second subspace: The wavelet-PCA model of the Jacobian determinants of the sample deformation fields. The reason to use Jacobian is that they reflect local volume changes of anatomical structures, which are important from the perspective of spatial distribution of the amount of brain tissue. It also makes sure that the deformation fields have valid Jacobian determinants and are topologically correct.
- The third subspace: A nested MRF regularization applied to eliminating the potential discontinuities of deformations.

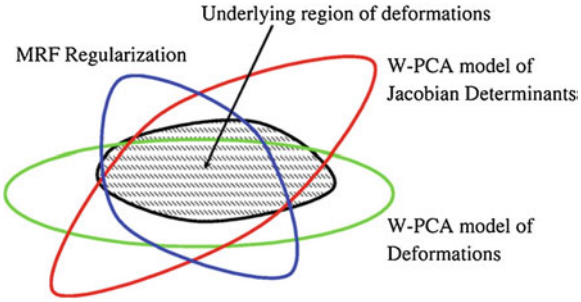


Fig. 4 The space of valid deformations is represented as the intersection of different subspaces reflecting different aspects of a warping field [23]

Figure 4 illustrates that the intersection of the three spaces represents the underlying region of deformation fields, and the details of the method was presented in [23].

Given an input deformation field, we can iteratively project it onto each of the three subspaces, and generate the SMD regularized deformation field according to these priors. This procedure is referred to as the SMD regularization algorithm, and it consists of the following five steps:

- Step 1. Project the deformation field onto the wavelet-PCA model of valid deformation fields;
- Step 2. Project the Jacobian of the deformation field onto the wavelet-PCA model of valid Jacobian determinants;
- Step 3. Find new deformation field with Jacobian matching the one generated in Step 2;
- Step 4. Apply the nested MRF regularization to imposing spatial smoothness on the deformation at all scales;
- Step 5. Go to step 1 and iterate until convergence, i.e., Until the MRF-regularized deformation field belongs to the subspaces of valid Jacobian and deformations.

SMD has been approved to be more accurate in capturing the statistics of deformation fields than the conventional PCA-based method and been used for simulating realistic images for evaluation of atlas-based registration and segmentation algorithms.

5.3 Statistically-Constrained Deformable Registration Using BM

After estimating the statistical model of deformation fields, we can use it to constrain a deformable registration. Let I_T and I_S be the template and the subject images respectively, according to the Bayesian model (BM) framework, we can estimate two deformation fields. One is denoted as **f**, and it is constrained by the prior distribution,

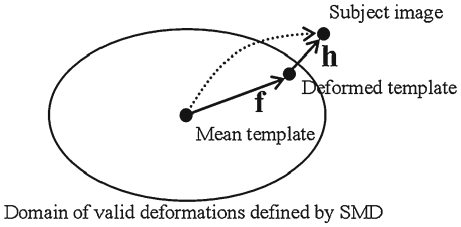


Fig. 5 Instead of directly registering the template image with the subject image (illustrated by the *dotted line*), in this work, the template image is first warped toward the subject image in the space of SMD (by deformation **f**), and a conventional registration is then performed for the deformed template and the subject images (using deformation **h**) [34]

and the other is denoted as $\bar{\mathbf{f}}$, and it is the final result. Then, similar to Eq. (8), the statistically-constrained deformable registration can be formulated as maximizing,

$$P(\bar{\mathbf{f}}, \mathbf{f}|I_T, I_S) = P(I_T, I_S|\bar{\mathbf{f}})P(\bar{\mathbf{f}}|\mathbf{f})P(\mathbf{f})/P(I_T, I_S), \tag{17}$$

where $P(\mathbf{f})$ is the prior distribution of deformation fields. Once the template image is fixed and this prior can be trained from a large number of images, and herein the prior can be defined by the SMD as described previously. $P(I_T, I_S|\bar{\mathbf{f}})$ is the likelihood of the images for a given deformation $\bar{\mathbf{f}}$, meaning we need to find a deformation field $\bar{\mathbf{f}}$ to match the two images. $P(\bar{\mathbf{f}}|\mathbf{f})$ indicates that these two deformation fields should be similar to each other.

Notice that once $\mathbf{f}$ is known, we can deform the template image I_T using $\mathbf{f}$, and generate a new deformed template image that is more similar with the subject image. Figure 5 shows such a case. The domain of the prior distribution of the deformation field $\mathbf{f}$ is represented by the ellipse, and the deformation field $\bar{\mathbf{f}}$ can be calculated by combining the deformations $\mathbf{f}$ and $\mathbf{h}$, $\bar{\mathbf{f}} = \mathbf{h} \circ \mathbf{f}$.

If we substitute $\bar{\mathbf{f}} = \mathbf{h} \circ \mathbf{f}$ in Eq. (17), the first term tells us to find a deformation field $\mathbf{h}$ to match the deformed template and the subject image, and the second term means the deformation field $\mathbf{h}$ should be small so $\mathbf{f}$ and $\bar{\mathbf{f}}$ are similar. The third term means that the deformation field $\mathbf{f}$ should be constrained by the prior distribution.

To solve the optimization problem, we can use the iterative two steps. The first step will fix the field $\mathbf{f}$, and try to find the best registration result $\mathbf{h}$; and the second step will fix the $\bar{\mathbf{f}} = \mathbf{h} \circ \mathbf{f}$, and try to constrain it to get a new deformation $\mathbf{f}$.

These two tasks are accomplished by the following two steps and also illustrated in Fig. 6:

- Step 1. Use the *SMD regularization algorithm* to regularize the combined deformation $\bar{\mathbf{f}} = \mathbf{h} \circ \mathbf{f}$ and generate a new deformation $\mathbf{f}$ and a new deformed template image;
- Step 2. Apply a conventional registration algorithm to register the deformed template image with the subject image and generate a new deformation $\mathbf{h}$. In this work,

the HAMMER [15] registration program was used, and therefore, the proposed framework is referred to as SMD+HAMMER.

Figure 7 gives an example of the statistically-constrained deformable registration. It can be seen that compared to the original template image in Fig. 7a, the intermediate deformed template image in Fig. 7c is more similar to the subject image in Fig. 7b, which renders the registration between Fig. 7b and c relatively easy than that between Fig. 7a and b, e.g., the deformation between them is relatively small and local.

Other registration algorithms could also be used instead. Notice that the deformation in this Section consists of two parts, one is a deformation of the template image defined according to SMD, and the other is a nonlinear deformation by registering the warped template with the subject image using a conventional registration; whereas in our previous work [23], only the former, i.e., the deformation regularized by SMD, is regarded as the registration result.

5.4 Results

Simulated deformations and images were used to compare the registration accuracy of HAMMER and the proposed SMD+HAMMER. First, we used two groups of T1-weighted MR brain images (79 for each group) as the training samples for two SMDs, respectively. Then, one SMD was used for simulating new deformations and images, and another was used for constraining the registration. We have simulated 9 such images and registered them onto the template image space, and then calculated the deformation errors between the registration results and the simulated ground truth. The histogram of these voxel-wise deformation errors of all the 9 images are shown in Fig. 8. It can be seen that SMD+HAMMER yields more accurate registration than HAMMER, with respective population means as 0.59 and 0.86 mm.

In order to test the robustness of registration, from a group of normal MR brain images (10 images), we simulated atrophy on the superior temporal gyrus and the precentral gyrus [20], and then registered all these images onto the template image. Using the tissue density maps, i.e., the RAVENS maps [56] of gray matter (GM), white matter (WM), and ventricular CSF (VN) calculated from the resultant deformation fields, we performed a paired t-test for group analysis using the statistical parametric mapping (SPM) software package. A smaller p-value or a larger t-value of this t-test indicates better separation. Table 1 shows the statistical measures for the two clusters detected in the locations of the precentral gyrus and the superior temporal gyrus, respectively. It can be seen that smaller p-values (both of $p_{FWE-corr}$ and $p_{FDR-corr}$) and larger t-values are obtained by SMD+HAMMER. Thus, according to these experiments, SMD+HAMMER generated more robust and stable deformation fields and is more powerful in detecting group differences.

In another test on simulated images, we simulated atrophy on corpus callosum on the T1-weighted images of 150 different subjects. All the original 300 original and simulated images were then registered on a randomly selected subject. The SMD

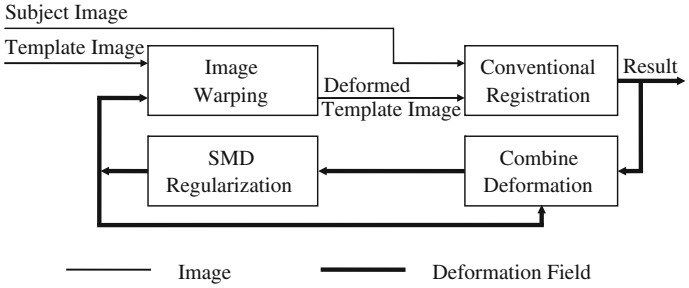


Fig. 6 The structure of the statistically-constrained deformable registration [34]

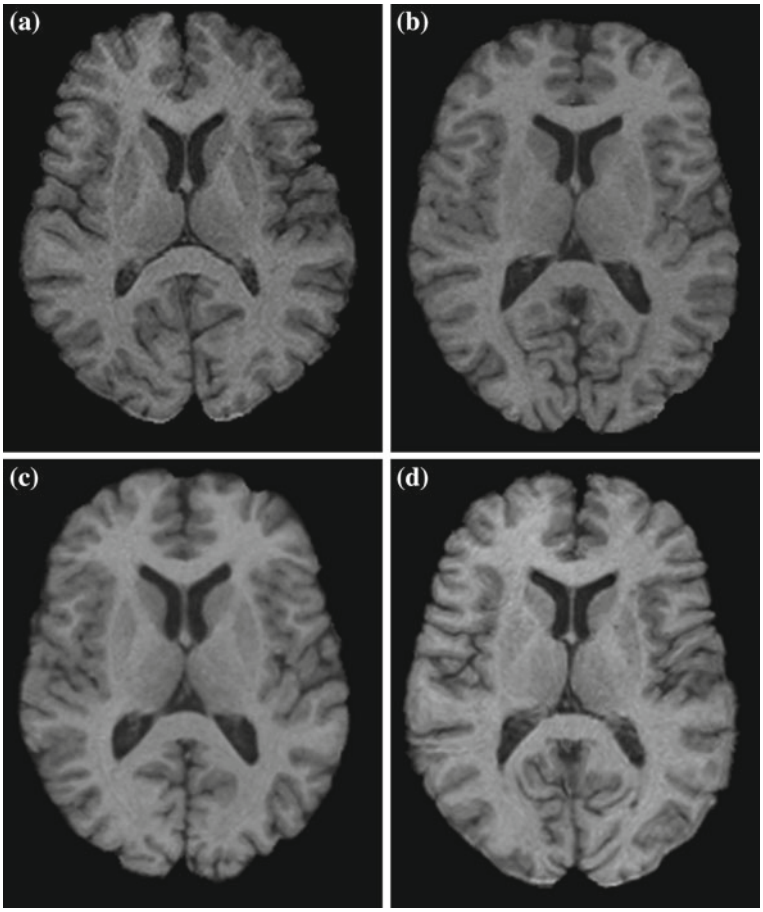


Fig. 7 An example of the registration results. **a** the template image; **b** the subject image; **c** the intermediate deformed template image; **d** final registration result: the warped template image

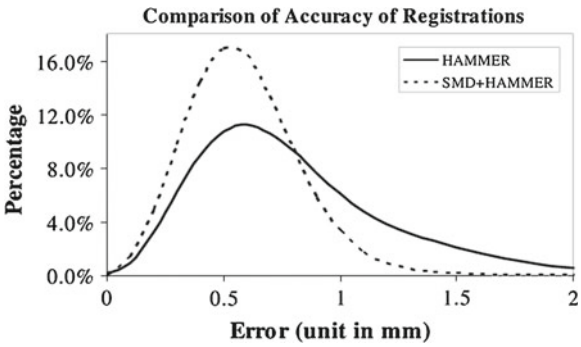


Fig. 8 Comparison of accuracy of registrations [34]

Table 1 Paired t-test for the two image groups

	HAMMER	SMD+HAMMER
Cluster 1 (in the superior temporal gyrus)	$p_{\text{FWE-corr}} = 0.342$ $p_{\text{FDR-corr}} = 0.057$ $T = 12.91$	$p_{\text{FWE-corr}} = 0.032$ $p_{\text{FDR-corr}} = 0.003$ $T = 16.70$
Cluster 2 (in the precentral gyrus)	$p_{\text{FWE-corr}} = 0.272$ $p_{\text{FDR-corr}} = 0.057$ $T = 13.26$	$p_{\text{FWE-corr}} = 0.003$ $p_{\text{FDR-corr}} = 0.003$ $T = 21.12$

was also trained on the deformation fields obtained from the 150 original images. We than performed a paired t-test for group analysis. Figure 9 shows some examples of the original images and the atrophy images in a flowchart. After registering all the images with the template, we calculated the Jacobian determinant map of all the deformation fields, and then performed t-test. The reverse of the p-value map $(1-p)$ is also shown in Fig. 9. It can be seen that the corpus callosum had been extracted indicating the group differences.

Robustness of registration can also be observed by warping serial images of the same subject. For the serial images captured from normal subjects, the longitudinal changes are relatively small, thus a registration algorithm should be able to accurately measure these subtle longitudinal changes or provide temporally consistent/smooth results, even without using any temporal smoothness constraints. Therefore, we tested the registration algorithm by registering the serial images of six subjects onto the template image then calculated the voxel-wise temporal smoothness (TS) of the serial deformation fields. TS is defined as the average of the absolute values of temporal gradients of deformations along longitudinally corresponding points, can be used to reflect such kind of longitudinal consistency.

A smaller TS value means the longitudinal deformation is temporally smooth, and viceversa. To illustrate the TS values of SMD+HAMMER relative to those of HAMMER, we calculated the histograms of the differences of TS maps between SMD+HAMMER and HAMMER, for the six series of images tested. Figure 10

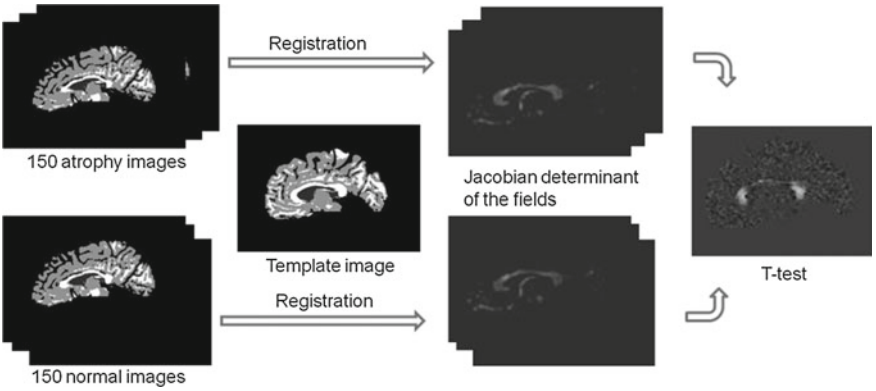


Fig. 9 Test of group comparison using simulated corpus callosum images

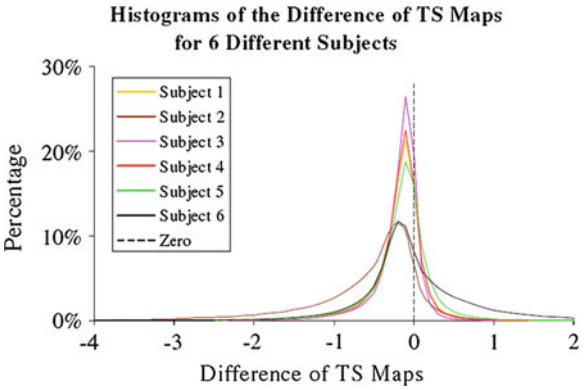


Fig. 10 Histograms of the difference of TS values between SMD+HAMMER and HAMMER [34]

shows the results of the six subjects. Notice the histogram now shows the distribution of the differences of TS, i.e., TS of SMD+HAMMER minus that of HAMMER. From the figure, we can see that most of the TS values of SMD+HAMMER are smaller than those of HAMMER, indicating that the new SMD+HAMMER method generates more longitudinally-consistent registration.

5.5 Concluding Remarks

This Section proposed a statistical model of deformation (SMD) and uses it to constrain deformable registration. The template image can be adaptively warped according to SMD during the registration procedure, and the conventional registration is performed by aligning the input subject image with the intermediate deformed tem-

plate image. More robust and accuracy registration is achieved compared to the conventional registration. The proposed framework can be easily applied to other conventional deformable registration methods.

6 Conclusion

We introduced the Bayesian Model (BM) that overcomes the limitation of MAP by incorporating an intermediate variable and jointly estimating the results and the intermediate variable. The BM framework allows for convenient incorporation of additional constraints to the medical image computing tasks and can be applied to image segmentation and registration. After a brief literature review of the statistical model-based image analysis methods, we first introduce the BM framework in terms of object matching. Then, we introduce how BM framework can be applied for medical image segmentation, registration, and group analysis. Using two examples for segmentation and registration, we showed that the BM framework can be used for formulation of image segmentation and registration in the same framework, and the results showed improved accuracy and robustness as compared with the traditional MAP framework.

References

1. Jain AK, Zhong Y, Lakshmanan S (1996) Object matching using deformable templates. IEEE Trans Pattern AnalMachine Intell 18:267–278
2. Kass M, Witkin A, Terzopoulos D (1987) Snakes - Active Contour Models. Int J Comput Vis 1:321–331
3. Chan TF, Vese LA (2001) Active contours without edges. IEEE Trans Image Process 10:266–277
4. Kim J, Cetin M, Willsky AS (2007) Nonparametric shape priors for active contour-based image segmentation. Sign Process 87:3021–3044
5. Tsai A, Yezzi A, Wells W, Tempany C, Tucker D, Fan A et al (2003) A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans Med Imaging 22:137–154
6. Lakshmanan S, Grimmer D (1996) A deformable template approach to detecting straight edges in radar images. IEEE Trans Pattern Anal Machine Intell 18:438–443
7. McInerney T, Terzopoulos D (1996) Deformable models in medical image analysis: a survey. Med Image Anal 1:91–108
8. Staib LH, Duncan JS (1992) Boundary finding with parametrically deformable models. IEEE Trans Pattern Anal Machine Intell 14:1061–1075
9. Yuille AL, Hallinan PW, Cohen DS (1992) Feature-extraction from faces using deformable templates. Int J Comput Vis 8:99–111
10. Lai KF, Chin RT (1995) Deformable contours-modeling and extraction. IEEE Trans Pattern Anal Machine Intell 17:1084–1090
11. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Machine Intell 23:681–685
12. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models–their training and application. Comput Vis Image Underst 61:38–59

13. Li H, Xue Z, Guo L, Wong ST (2009) Simultaneous consideration of spatial deformation and tensor orientation in diffusion tensor image registration using local fast marching patterns. Inf Process Med Imaging 21:63–75
14. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ (1999) Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imaging 18:712–721
15. Shen D, Davatzikos C (2002) HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans Med Imaging 21:1421–39
16. Toga AW, Thompson PM (2001) The role of image registration in brain mapping. Image Vis Comput 19:3–24
17. Twining CJ, Cootes T, Marsland S, Petrovic V, Schestowitz R, Taylor CJ (2005) A unified information-theoretic approach to groupwise non-rigid registration and model building. Inf Process Med Imaging 19:1–14
18. Xue Z, Li H, Guo L, Wong ST (2010a) A local fast marching-based diffusion tensor image registration algorithm by simultaneously considering spatial deformation and tensor orientation. Neuroimage 52:119–130
19. Xue Z, Shen D (2009) A new statistically-constrained deformable registration framework for MR brain images. Int J Med Eng Inform 1:357–367
20. Xue Z, Shen D, Karacali B, Stern J, Rottenberg D, Davatzikos C (2006b) Simulating deformations of MR brain images for validation of atlas-based segmentation and registration algorithms. Neuroimage 33:855–66
21. Xue Z, Wong K, Wong STC (2010b) Joint registration and segmentation of serial lung CT images for image-guided lung cancer diagnosis and therapy. Comput Medl Imaging Graph 34:55–60
22. Yeo BT, Sabuncu M, Vercauteren T, Ayache N, Fischl B, Golland P (2008) Spherical demons: fast surface registration. In: Medical image computing and computer-assisted intervention MICCAI 2008, 11 May 2008 Vol. 11. pp 745–53
23. Xue Z, Shen D, Davatzikos C (2006a) Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping. Med Image Anal 10:740–751
24. Miller M, Banerjee A, Christensen G, Joshi S, Khaneja N, Grenander U et al (1997) Statistical methods in computational anatomy. Stat Methods Med Res 6:267–299
25. Ashburner J, Friston KJ (2000) Voxel-based morphometry-the methods. Neuroimage 11:805–821
26. Xue Z, Shen D, Karacali B, Davatzikos C (2005) Statistical representation and simulation of high-dimensional deformations: application to synthesizing brain deformations. Med Image Comput Comput Assist Interv 8:500–508
27. Xue Z, Li SZ, Teoh EK (2003) Bayesian shape model for facial feature extraction and recognition. Pattern Recogn 36:2819–2833
28. Heimann T, Meinzer HP (2009) Statistical shape models for 3D medical image segmentation: a review. Med Image Anal 13:543–563
29. Chen T, Metaxas D (2003) Gibbs prior models, marching cubes, and deformable models: a hybrid framework for 3D medical image segmentation. In: Medical image computing and computer-Assisted intervention MICCAI 2003, vol 22879. pp 703–710
30. Chen T, Metaxas D (2000) Image segmentation based on the integration of Markov Random Fields and deformable models. In: Medical image computing and computer-assisted intervention MICCAI 2000, vol 1935. pp 256–265
31. Xue Z, Li SZ, Teoh EK (2002) Al-EigenSnake: an affine-invariant deformable contour model for object matching. Image Vis Comput 20:77–84
32. Farzinfar M, Xue Z, Teoh EK (2008) Joint parametric and non-parametric curve evolution for medical image segmentation. In: European conference on computer vision ECCV 2008, Springer, pp 167–78
33. Farzinfar M, Xue Z, Teoh EK (2010) A novel approach for curve evolution in segmentation of medical images. Comput Med Imaging Graph 34:354–361

34. Xue Z, Shen D (2007) Statistically-constrained deformable registration of MR brain images. In: 4th IEEE International Symposium on biomedical imaging from Nano to Macro, ISBI 2007, pp 25–8
35. Caselles V, Catte F, Coll T, Dibos F (1993) A geometric model for active contours in image-processing. Numer Math 66:1–31
36. Caselles V, Kimmel R, Sapiro G (1997) Geodesic active contours. Int J Comput Vis 22:61–79
37. Osher S, Sethian JA (1988) Fronts propagating with curvature-dependent speed: algorithms based on Hamilton–Jacobi formulations. J Comput Phys 79:12–49
38. Cohen LD, Cohen I (1993) Finite-element methods for active contour models and balloons for 2-D and 3-D images. IEEE Trans Pattern Anal Machine Intell 15:1131–1147
39. Zhu SC, Yuille A (1996) Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans Pattern Anal Machine Intell 18:884–900
40. Mumford D, Shah J (1988) Boundary detection by minimizing functionals. In: Image understanding, pp 19–43
41. Dal Maso G, Morel J-M, Solimini S (1992) A variational method in image segmentation: existence and approximation results. Acta Math 168:89–151
42. Samson C, Blanc-Feraud L, Aubert G, Zerubia J (2000) A variational model for image classification and restoration. IEEE Trans Pattern Anal Machine Intell 22:460–472
43. Yezzi Jr A, Tsai A, Willsky A (1999) A statistical approach to snakes for bimodal and trimodal imagery. In: Proceedings of the Seventh IEEE international conference on computer vision, 1999, Vol 2. pp 898–903
44. Paragios N, Deriche R (2002) Geodesic active regions: a new framework to deal with frame partition problems in computer vision. J Vis Commun Image Represent 13:249–268
45. Kim JM, Fisher JW, Yezzi A, Cetin M, Willsky AS (2005) A nonparametric statistical method for image segmentation using information theory and curve evolution. IEEE Trans Image Process 14:1486–1502
46. Bresson X, Vandergheynst P, Thiran JP (2006) A variational model for object segmentation using boundary information and shape prior driven by the Mumford-Shah functional. Int J Comput Vis 68:145–162
47. Chen YM, Tagare HD, Thiruvenkadam S, Huang F, Wilson D, Gopinath KS et al (2002) Using prior shapes in geometric active contours in a variational framework. Int J Comput Vis 50:315–328
48. Rousson M, Paragios N (2008) Prior knowledge, level set representations & visual grouping. Int J Comput Vis 76:231–243
49. Leventon ME, Grimson WEL, Faugeras O (2000) Statistical shape influence in geodesic active contours. In: Proceedings of IEEE Conference on computer vision and pattern recognition, Hilton Head Island, 13–15 June 2000, Vol 1. pp 316–323
50. Charpiat G, Faugeras O, Keriven R (2005) Approximations of shape metrics and application to shape warping and empirical shape statistics. Found Comput Math 5:1–58
51. Cremers D, Kohlberger T, Schnorr C (2002) Nonlinear shape statistics in Mumford-Shah based segmentation. In: European conference on computer vision ECCV 2002, Pt 2, vol 2351. pp. 93–108
52. Cremers D, Osher SJ, Soatto S (2004) Kernel density estimation and intrinsic alignment for knowledge-driven segmentation: teaching level sets to walk. Pattern Recogn 3175:36–44
53. Rousson M, Paragios N, Deriche R (2004) Implicit active shape models for 3D segmentation in MR imaging. In: Proceedings of medical image computing and computer-assisted intervention MICCAI 2004, Pt 1, vol 3216. pp 209–216
54. Duncan JS, Ayache N (2000) Medical image analysis: progress over two decades and the challenges ahead. IEEE Trans Pattern Anal Machine Intell 22:85–106
55. Davatzikos C, Tao X, Shen D (2003) Hierarchical active shape models, using the wavelet transform. IEEE Trans Med Imaging 22:414–423
56. Davatzikos C, Genc A, Xu D, Resnick SM (2001) Voxel-based morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. Neuroimage 14:1361–1369

Shape-Constrained Deformable Models and Applications in Medical Imaging

Jürgen Weese, Irina Wächter-Stehle, Lyubomir Zagorchev and Jochen Peters

Abstract The recognition and segmentation of organs and anatomical structures in medical images is the basis for an efficient workflow and quantitative measurements in diagnostic and interventional applications. Numerous methods have been developed in the past for specific applications, and many of them are based on variants and extensions of active contours or active shape methods. We present an overview over shape-constrained deformable models that have specifically been developed for organ segmentation in 3D medical images. They rely on a pre-defined shape space like active shape models, but preserve some flexibility of active contour approaches as they allow deviations from the shape space. In particular, we describe our approach to shape parametrization and the concept of "Simulated Search" that we use to train boundary detection. Fully automatic segmentation is achieved by a segmentation chain comprising a localization step based on the Generalized Hough Transformation and subsequent model adaptation with increasing degrees of freedom. Finally, we show how shape-constrained deformable models allow to address clinical applications in cardiology and neurology.

J. Weese (✉) · I. Wächter-Stehle · J. Peters
Philips Technologie GmbH, Innovative Technologies, Research Laboratories, Röntgenstrasse 24–26, 22335 Hamburg, Germany
e-mail: juergen.weese@philips.com

I. Wächter-Stehle
e-mail: irina.waechter@philips.com

J. Peters
e-mail: jochen.peters@philips.com

L. Zagorchev
Philips Research North America, 345 Scarborough Road, Briarcliff Manor, NY 10510, USA
e-mail: Lyubomir.Zagorchev@philips.com

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_5,

1 Introduction

Today, medical imaging systems like CT and MR scanners produce detailed high quality patient images that contain a huge amount of information. This information is, however, not readily available and recognition and segmentation techniques are key to extract and visualize the desired information efficiently in clinical practice. For that reason, numerous methods have been developed in the past to segment vessels, tumors and anatomical structures. In many cases, these techniques have been designed to address specific questions or clinical applications.

We present an overview of shape-constrained deformable models (SCDM) as introduced in [1] that have specifically been developed for organ segmentation in 3D medical images. A specific goal was to develop a framework that works for images of different modalities such as CT or MR as well as for different organs or anatomical structures. SCDMs rely on a pre-defined shape space like active shape models [2], but preserve some flexibility of active contour approaches [3] as they allow deviations from the shape space. Compared to atlas-based segmentation methods that are often used for brain segmentation [4], SCDMs share the concept of deforming a reference anatomy in order to segment the organ of interest. Adaptation relies, however, on triangle meshes and on boundary detection rather than on voxel-based non-rigid registration.

Starting with a brief description of active contours and active shape models, we introduce the basic concept of SCDMs in the following section. In Sect. 3, we discuss their application for establishing corresponding points on 3D organ surfaces in the context of 3D statistical shape model construction. As an alternative approach to parameterize shape variability, we introduce multi-linear models that are based on a subdivision of an anatomical structure into several parts. Next to a parameterization of the shape variability, robust and accurate model adaptation requires suitable techniques for boundary detection. Our boundary detection approach is presented in Sect. 4. Training is based on the definition of a large number of boundary detection function candidates and a mechanism that we call "Simulated Search" which selects the optimal boundary detection function for each triangle from the candidates.

The approaches described in Sects. 2, 3 and 4 have been integrated in a segmentation framework that is described in Sect. 5. Next to model building, fully automatic segmentation by localization using the Generalized Hough Transformation (GHT) and subsequent model adaptation with increasing degrees of freedom is described. In addition, segmentation results for different organs in CT and MR images are presented. Section 6 describes four clinical applications of the segmentation framework. Clinical applications comprise left ventricle (LV) segmentation for functional analysis in CT, CT-based measurements for planning transcatheter aortic valve implantations (TAVI), the generation of models of the left atrium and pulmonary veins for guiding atrial fibrillation (AF) ablation procedures and volumetry of brain structures using MR images for evaluating traumatic brain injury (TBI) patients.

2 Approaches

SCDMs combine properties of deformable models or active contour models and active shape models. We briefly describe therefore both approaches, before introducing the concept of SCDMs. For illustration, we also present results for vertebra segmentation in CT images.

2.1 Deformable Models and Active Shape Models

Active contour models have been proposed by Kass et al. [3] for the segmentation of objects in 2D images. The basic idea consists in the determination of the contour

$$\boldsymbol{x}(s) \quad s \in [0, 1] \tag{1}$$

of an object by minimization of an energy functional that consists of an external energy E_{ext} attracting the contour to image features and an internal energy E_{int} that enforces smoothness of the contour. The energy functional may, for instance, be given by

$$E[\boldsymbol{x}] = -\int_{s=0}^{1} \parallel \nabla I(\boldsymbol{x}(s)) \parallel^2 ds + \alpha \int_{s=0}^{1} \parallel \partial^2/\partial s^2 \boldsymbol{x}(s) \parallel^2 ds. \tag{2}$$

After proper initialization, the contour would be attracted towards borders in the image I with high gradients and the weight α controls smoothness of the contour. This approach has been extremely successful and a huge number of papers were published proposing variations of energy terms and applying it to numerous problems. The approach was also applied in medical image analysis and extended for the segmentation of surfaces in 3D images [5, 6].

Active shape models [2] are another very successful approach. The approach relies on establishing a number V of corresponding points $\boldsymbol{X}_i = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,V})^\top$ on a set of N reference objects ($i = 1, \ldots, N$) of the same kind. After aligning the objects, a point-distribution model (PDM)

$$\boldsymbol{X}^{\text{PDM}} = \bar{\boldsymbol{M}} + \sum_{k=1}^{P} p_k \boldsymbol{M}_k \tag{3}$$

is constructed via principal component analysis (PCA), i.e. the mean shape

$$\bar{\boldsymbol{M}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{X}_i \tag{4}$$

is computed and modes $\boldsymbol{M}_k = (\boldsymbol{m}_{k,1}, \ldots, \boldsymbol{m}_{k,V})^\top$ describing shape variations are derived from the P largest eigenvectors of the matrix

$$\boldsymbol{C} = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{X}_i - \bar{\boldsymbol{M}})(\boldsymbol{X}_i - \bar{\boldsymbol{M}})^\top. \tag{5}$$

Assuming that an initial model position and shape is given, adaptation of the PDM is done by detecting borders along normals of the object boundary and modifying the parameters p_k of the PDM to minimize the distance to the detected border points. Also for this approach numerous variants have been introduced and the basic approach was applied to a large number of problems. In particular, the approach was extended for 3D medical image segmentation [7].

When segmenting an anatomical object like a bone or the heart with a stable topology between different individuals, active shape models differ in two important aspects from active contour models. First, the segmentation accuracy of active shape models is limited by the number of eigenvectors and the training samples used to construct them while active contour models can in principle approximate any (smooth) shape. Of course, the flexibility of active shape models may be increased by adding more eigenvectors and artificially generating additional reference shapes, but the advantage of a compact model with a few parameters will be lost. Second, by establishing corresponding points, the points of an active shape model can, at least approximately, be associated with an anatomical location, while active contour models do not allow this. This property is very important as it allows to attach individual boundary detectors to each point and enables segmentation of organs with different boundary properties for different parts.

2.2 Shape Constrained Deformable Models

SCDMs [1] have been introduced to combine the advantages of active contour models with those of active shape models. In particular, they have been designed to allow for an accurate approximation of anatomical objects, while each point of the model can be associated with an anatomical region.

In the following we consider the segmentation of anatomical objects in 3D medical images. The deformable model is represented by a mesh with V vertices $\boldsymbol{x}_i$ and T triangles with centers $\boldsymbol{c}_i$. Adaptation of the mesh to an image is done by iterating a boundary detection step and a mesh deformation step. Boundary detection is done by detecting points $\boldsymbol{x}_i^{\text{target}}$ with high gradient along profiles parallel to the mesh normal (see also Sect. 4). Mesh deformation is done by minimizing an energy

$$E = E_{\text{ext}} + \alpha E_{\text{int}}. \tag{6}$$

The external energy E_{ext} drives the mesh towards the detected borders

$$E_{\text{ext}}^{p} = \sum_{i=1}^{T} \tilde{w}_i (\boldsymbol{x}_i^{\text{target}} - \boldsymbol{c}_i)^2 \tag{7}$$

or the associated surfaces

$$E_{\text{ext}}^{s} = \sum_{i=1}^{T} \tilde{w}_i \left(\frac{\nabla^\top I(\boldsymbol{x}_i^{\text{target}})}{\| \nabla I(\boldsymbol{x}_i^{\text{target}}) \|} (\boldsymbol{x}_i^{\text{target}} - \boldsymbol{c}_i) \right)^2 \tag{8}$$

with weights $\tilde{w}_i$ derived from the boundary detection step. The internal energy

$$E_{\text{int}} = \sum_{i=1}^{V} \sum_{j \in N(i)} \left(\boldsymbol{x}_i - \boldsymbol{x}_j - s\boldsymbol{R} \left(\bar{\boldsymbol{m}}_i - \bar{\boldsymbol{m}}_j + \sum_{k=1}^{P} p_k (\boldsymbol{m}_{k,i} - \boldsymbol{m}_{k,j}) \right) \right)^2 \tag{9}$$

compares the difference vectors between two neighboring mesh vertices of the deforming model and the shape model. $N(i)$ denotes the neighbors of vertex i. The scaling parameter s and the rotation matrix $\boldsymbol{R}$ allow to properly align both models. Finally, it should be mentioned that in addition to the mesh vertices, the modes p_k of the shape model, the scaling parameter s, and the rotation matrix $\boldsymbol{R}$ are optimized in the mesh deformation step.

The internal energy of a SCDM does essentially embed a shape model into a deformable model and introduces regularization by penalizing deviations from the shape model instead of imposing a smoothness constraint. In addition, the approach approximately maintains the distribution of mesh vertices as given by the shape model. This mechanism preserves the property of shape models that each point can be associated with an anatomical location.

2.3 Example Results for Vertebra Segmentation

For illustration, segmentation results for 18 vertebra in 6 CT images generated with a SCDM with $P = 10$ modes are briefly discussed [1]. For initialization, the mean vertebra model has been placed with an Euclidean surface-to-surface error of 2.5–3.5 mm compared to the reference segmentation. Model adaptation resulted in an average error of 1.3 mm when attracting the model directly to the detected boundaries (Eq. 7), and in 0.9 mm when attracting the model to the associated surfaces (Eq. 8). Fig. 1 shows an exemplary result.

The example shows that the basic idea of SCDMs works. In addition, it demonstrates the importance of the external energy and whether the model is directly attracted towards the detected boundary points or to the associated surfaces. Though the difference between the corresponding average segmentation errors of 1.3 and 0.9 mm is only 0.4 mm, the segmentation result looks unacceptable in the former case, while a good segmentation result is obtained in the latter case. This impression

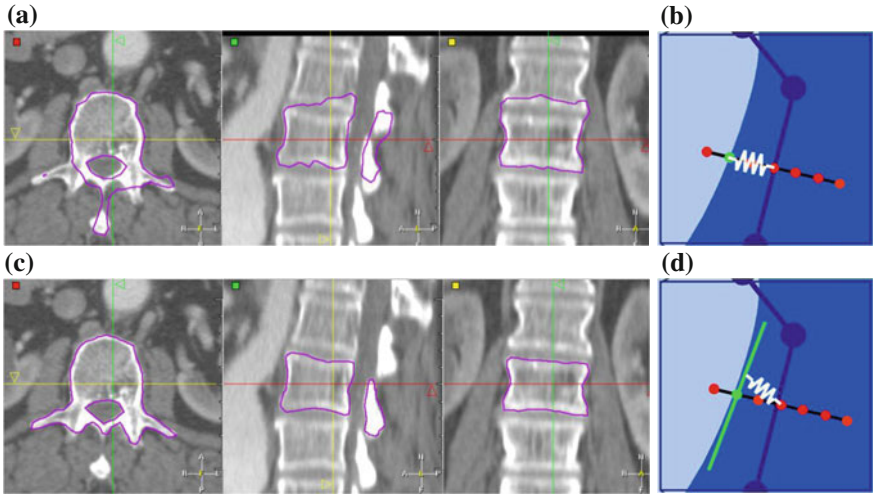


Fig. 1 Sample segmentation results starting from the identical initialization, when using the external energy of Eq. (7) (**a**, **b**) and of Eq. (8) (**c**, **d**) [1]

can potentially be explained by large segmentation errors in some regions while an accurate segmentation is obtained in other regions, and shows that comparatively small differences in surface-to-surface metrics can make a relevant difference.

3 Shape Variability

Robust and accurate model adaptation relies on a suitable shape parameterization. We first summarize results related to PCA-based shape models where SCDMs have been used to establish corresponding points. Afterwards we introduce multi-linear models as an alternative approach to parameterize shape variability. The latter approach is exploited in the SCDM framework as it offers an intuitive way to parameterize shape variability that leads to good shape approximations and can easily be generalized to complex models.

3.1 PCA-Based Shape Models

Construction of an active shape model requires to establish V corresponding points $\boldsymbol{X}_i = (\boldsymbol{x}_{i,1}, \ldots, \boldsymbol{x}_{i,V})^\top$ on a set of N reference objects of the same kind. Looking at the surface of a 3D organ, this is a non-trivial task, because there is usually no dense set of anatomical landmarks that can unambiguously be identified. A number of methods has been proposed to address this task. They can be classified into mesh-to-mesh, mesh-to-volume and volume-to-volume registration [7].

SCDMs have been designed in a way that each point can be associated with an anatomical location. They can therefore be used for establishing corresponding points. Starting with a set of N binary images representing the different instances of the anatomical structure-of-interest, an arbitrarily selected instance can be used as reference to derive a mesh model by triangulation. This mesh model is then subsequently adapted to the remaining $N-1$ binary images. This approach belongs to the mesh-to-volume category and has been used to construct shape models of vertebrae and femurs [8].

The results for 32 vertebrae (L1–L4) derived from CT scans with an in-plane resolution of $0.2 \times 0.2\,\text{mm}^2$ to $0.7 \times 0.7\,\text{mm}^2$ and a slice thickness between 2 and 3 mm show that the reference mesh can be adapted to the binary images representing the other vertebrae with an average surface-to-surface error of 0.8 mm. The accuracy of corresponding anatomical positions has been assessed using the tips of the processes as landmarks. They have been encoded in the reference mesh by marking the corresponding triangle. After adapting the reference mesh to the binary image of another instance, these triangle positions have been compared with the manually defined landmarks resulting in an average Euclidean distance of 2.1 mm. This result provides evidence for the statement that points of a SCDM can approximately be associated with an anatomical location, especially when keeping in mind that the result is also affected by the accuracy of manual landmark definition.

Generalizability of the PCA-based shape models has been addressed by leave-one-out experiments where a PDM was built from all vertebrae but one and adapted to the left out vertebra. Figure 2 shows the resulting surface-to-surface error in dependence of the number of modes P. The results show that the approximation accuracy increases with the number of modes P. The benefit of additional modes decreases, however, and the approximation accuracy is only marginally improved when using more modes than $P = 13$. The resulting approximation accuracy of 1.4 mm is clearly worse than the 0.9 mm obtained with SCDM segmentation [1]. Although care must be taken in the comparison since metric and test set are not identical, the results show

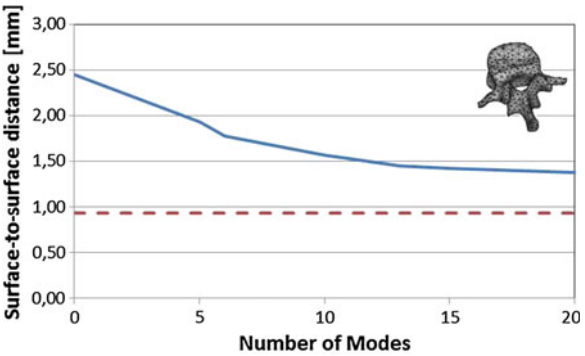


Fig. 2 Approximation of an unseen vertebra with a PCA-based shape model [8]. The *dashed line* indicates the segmentation accuracy of SCDMs [1]

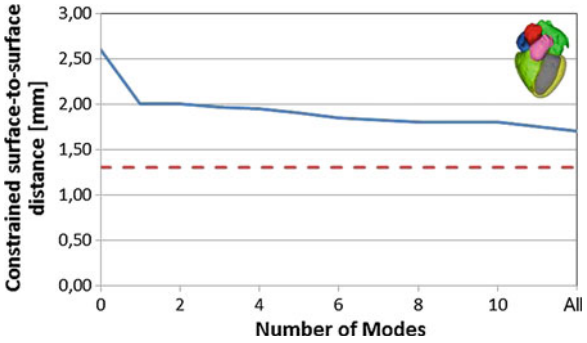


Fig. 3 Approximation of an unseen heart with a PCA-based shape model [9]. The *dashed line* corresponds to shape approximation with a multi-linear model

that SCDMs relax shape constraints and can, therefore, lead to more accurate results than active shape models.

A similar experiment was performed for a 4-chamber heart model [9]. In this work, 28 cardiac CTA images from 13 patients corresponding to different heart phases with an in-plane resolution of $0.48 \times 0.48\,\text{mm}^2$ and a slice thickness between 0.67 and 3 mm were used. Corresponding meshes and ground truth annotations have been generated simultaneously in a bootstrap approach using SCDMs, and the resulting meshes have been used to build PCA-based shape models. Generalizability has been addressed by leave-one-patient-out experiments to avoid that results are biased by the shape of a patient's heart. Figure 3 shows the mean constrained surface-to-surface error, which is close to a surface-to-surface error but avoids meaningless comparisons of non-corresponding mesh parts. The approximation accuracy increases with the number of modes, but the curve becomes flat and usage of more than $P = 8$ modes leads only to marginal improvements.

Both examples illustrate that shape models with a few modes derived by principal component analysis allow to approximate anatomical structures with a surprisingly high accuracy. Though other approaches for establishing correspondences than SCDMs may lead to better results, approaches that further increase the flexibility of shape models and allow for a more accurate approximation are desired.

3.2 Multi-Linear Models

One class of approaches to increase model flexibility is to subdivide an anatomical structure into several parts and model their shape variability independently [7, 10–12]. One particular approach [9, 13] consists in partitioning an anatomical object in K parts and modeling the variability of part k by a linear transformation $\mathcal{T}_k(\mathbf{q}_k)[.]$ where $\mathbf{q}_k$ describes the transformation parameters. With this approach, the mean model can be transformed according to

$$\mathscr{T}_{\text{multi-linear}}(\mathbf{q})[\bar{\boldsymbol{m}}_i] = \sum_{k=1}^{K} w_{i,k} \cdot \mathscr{T}_k(\mathbf{q}_k)[\bar{\boldsymbol{m}}_i] \tag{10}$$

If all parts would be independent, the subdivision can be described by weights

$$w_{i,k} = \begin{cases} 1 \text{ if vertex } i \text{ belongs to part } k \\ 0 \text{ otherwise} \end{cases}. \tag{11}$$

To avoid discontinuities between the individual parts, the transformations are interpolated in transition regions that depend on the geodesic distance to the border between the parts. Practically, this is done by assigning weights $0 < w_{i,k} < 1$ with $\sum_{k=1}^{K} w_{i,k} = 1$ in transition regions.

This approach has been used to parameterize the shape of a 4-chamber heart model [9]. In particular, affine transformations ($K = 5$) have been assigned to the left atrium, the right atrium, the left ventricular epicardium, the left ventricular endocardium (together with the aortic trunk) and the right ventricle (together with the pulmonary artery). The capability to approximate an unknown shape has been assessed in the same way and using the same data as it has been done for the PCA-based shape model of Fig. 3. Therein, the resulting mean constrained surface-to-surface error is shown by the dashed red line.

The result shows that this intuitive way of parameterizing shape leads to a better approximation. Though the multi-linear model has more parameters (60 resulting from 5 affine transformations with 12 parameters) than the PCA-based shape model (37 corresponding to 25 modes and an affine transformation describing pose), there is no indication that global PCA-analysis would lead to a more compact model that approximates unknown shapes with similar or better accuracy. In addition, local modeling of shape variability via partitioning allows to describe shape variability of complex anatomical structures and organs within a single framework. A multi-linear model has, for instance, been used to construct a shape parameterization of the heart with the major vessels [14]. Shape variability of the vessels was modeled by assigning similarity transformations to vessel segments (Fig. 4).

Modeling of shape variability by multi-linear transformations can easily be included in the SCDM framework. For that purpose, the internal energy of Eq. (9) is replaced by

$$E_{\text{int}} = \sum_{i=1}^{V} \sum_{j \in N(i)} \sum_{k=1}^{K} w_{i,k} \left(\boldsymbol{x}_i - \boldsymbol{x}_j - \left(\mathscr{T}_k(\mathbf{q}_k)[\bar{\boldsymbol{m}}_i] - \mathscr{T}_k(\mathbf{q}_k)[\bar{\boldsymbol{m}}_j] \right) \right)^2. \tag{12}$$

The parameters $\mathbf{q}$ of the multi-linear transformation are optimized in addition to the vertex positions $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_V$ during mesh deformation by energy minimization.

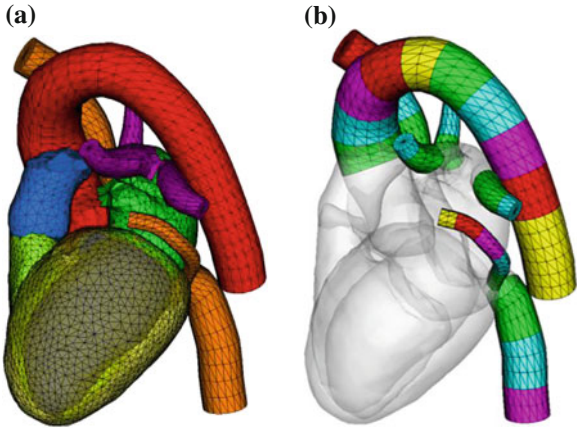


Fig. 4 Heart model [14] with large vessels (**a**). For multi-linear shape modeling, the vessels are subdivided into segments (**b**)

4 Boundary Detection Framework

A key problem when adapting a surface model to medical images is the abundance of misleading structures that need to be distinguished from the wanted organ surfaces. In addition, boundary characteristics may vary over the organ surface. The boundary detection framework is therefore of key importance for robust and accurate adaptation of a SCDM to images. We first describe how boundaries are detected and the type of features that we use. Specific boundary detection functions are generated via the analysis of training images. "Simulated Search" is used to select the best boundary detection function per triangle. Finally, we summarize some results that have been generated in the context of heart segmentation.

4.1 Boundary Detection

As explained in Sect. 2.2, adaptation of the mesh to an image is done by iterating a boundary detection step and a mesh deformation step that minimizes the energy $E = E^s_{\text{ext}} + \alpha E_{\text{int}}$ (see Eqs. (6)–(9)). Similar as for active shape models [2, 7], boundary detection is done by searching for each mesh triangle a characteristic point along profiles parallel to the mesh normals.

Formally, for triangle i, we establish a search profile $\{\boldsymbol{x}_i^{-\ell}, \ldots, \boldsymbol{x}_i^{+\ell}\}$ centered at the triangle center $\boldsymbol{c}_i$. The profile is oriented along the triangle normal with equidistantly spaced points $\boldsymbol{x}_i^k$. Among these profile points, a target point is detected:

$$\boldsymbol{x}_i^{\text{target}} = \underset{\{\boldsymbol{x}_i^k \mid k=-\ell,\ldots,+\ell\}}{\text{argmax}} \left(F_i(\boldsymbol{x}_i^k) - D \cdot (\boldsymbol{x}_i^k - \boldsymbol{c}_i)^2 \right). \tag{13}$$

Here, $F_i(\boldsymbol{x})$ is a triangle-specific function with a large positive response for surface points (e.g., the local gradient magnitude $\|\nabla I(\boldsymbol{x})\|$). The distance penalty with weighting factor D allows to bias the search to nearby target points. The feature response together with the distance penalty is also used to define the weight

$$\tilde{w}_i = \max \left\{ 0,\, F_i(\boldsymbol{x}_i^{\text{target}}) - D \cdot (\boldsymbol{x}_i^{\text{target}} - \boldsymbol{c}_i)^2 \right\}. \tag{14}$$

of the detected boundary point in the external energy of Eqs. (7) and (8).

4.2 Features

Organ boundaries in medical images are often associated with intensity transitions. The image gradient $\nabla I(\boldsymbol{x})$ builds, therefore, the basis of the features $F_i(\boldsymbol{x})$ that we use. The image gradient is projected onto the triangle normal $\boldsymbol{n}_i$ to suppress edges that deviate strongly from the local surface orientation. To control the maximum feature response for structures like metal implants in CT images, we use a heuristic damping of gradients exceeding a threshold $g_{\max}$ [1]:

$$G_{\text{proj}}^{\text{limit}}(\boldsymbol{x}) = \left(\boldsymbol{n}_i \cdot \nabla I(\boldsymbol{x}) \right) \cdot \frac{g_{\max} \cdot (g_{\max} + \|\nabla I(\boldsymbol{x})\|)}{g_{\max}^2 + \|\nabla I(\boldsymbol{x})\|^2}. \tag{15}$$

This quantity builds together with the sign $s_i = \pm 1$ the basis of our features

$$F_i(\boldsymbol{x}) = s_i \cdot G_{\text{proj}}^{\text{limit}}(\boldsymbol{x}). \tag{16}$$

The sign indicates per triangle i whether the boundary is expected to show a bright-to-dark transition or vice versa.

Such a simple feature leads to many positive responses, and a large number of unwanted boundary points are typically detected. Local gray values and gradient sizes [15–19] have been used to design more specific features and to improve discrimination from false boundaries. We use a general formulation and constrain the admissible edges by a set S of local image criteria $Q_j(\boldsymbol{x})$

$$F_i(\boldsymbol{x}) = \begin{cases} s_i \cdot G_{\text{proj}}^{\text{limit}}(\boldsymbol{x}) & \text{if } Q_j(\boldsymbol{x}) \in [\text{Min}_{i,j}, \text{Max}_{i,j}]\ \forall\, j \in S, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

If any of the criteria's value is not within its triangle-specific acceptance interval $[\text{Min}_{i,j}, \text{Max}_{i,j}]$, the associated boundary is rejected by setting its response to 0.

Typically, we use criteria $Q_j(\boldsymbol{x})$ such as local gray values on either side of the surface, averaged over several points perpendicular or parallel to the surface, and changes of gray values across the surface. The latter can be characterized by first or second order Taylor coefficients of a local gray value profile or simply by the difference of inside versus outside gray values. If needed, these criteria can be extended to better capture local characteristics and to improve discrimination. If, for instance, the gray-value transition (bright-to-dark or vice versa) is less predictable, the signed gradient $s_i \cdot G^{\text{limit}}_{\text{proj}}(\boldsymbol{x})$ may be replaced by its absolute value $\left\| G^{\text{limit}}_{\text{proj}}(\boldsymbol{x}) \right\|$. Furthermore, gray-value normalization may be introduced to allow application of the features for imaging modalities without calibrated gray-scale (see Sect. 5.3).

4.3 Feature Parameterization

The feature functions of Eq. (17) depend on several parameters such as the sign s_i of the expected gradient orientation and the acceptance intervals $[\text{Min}_{i,j}, \text{Max}_{i,j}]$ per triangle. In the simplest case, the same feature function may be used for all triangles. Its parameters can be derived from a statistical analysis of the parameter values on a collection of training images with reference meshes. This analysis takes the positions $\boldsymbol{c}_i$ of the triangles of a reference mesh per image and evaluates $Q_j(\boldsymbol{c}_i)$. Based on these values, we can define acceptance intervals $[\text{Min}_j, \text{Max}_j]$ that extend, for instance, from a lower to an upper percentile value of the observed values of the criteria $Q_j(\boldsymbol{x})$.

To account for variations of the image appearance across the organ surface, the global statistical analysis can be preceded by a clustering procedure that groups the mesh triangles into clusters of similar appearance (see, e.g., [18, 20–22]). This can be done using K-means clustering based on the vector of criteria $Q_j(\boldsymbol{c}_i)$. The statistical analysis is then performed per cluster of mesh triangles, resulting in acceptance intervals $[\text{Min}_{k,j}, \text{Max}_{k,j}]$ for each cluster k. The sign s_k can also be adjusted per triangle cluster. With this clustering-based approach, all triangles of a cluster share the same parameterized feature function.

Definition of the feature functions depends on various parameters and decisions. E.g., we can include or discard the gradient sign in Eq. (17). We can also use different sets of criteria $Q_j(\boldsymbol{x})$ to constrain boundary detection and change the width of the acceptance intervals. These modifications will improve boundary detection for some triangles and deteriorate boundary detection for other triangles. For that reason, adaptation of a SCDM to images can be further improved by an approach that selects the optimal feature function for a triangle from a large number of feature function candidates, while the feature functions of Eq. (17) together with the clustering approach are very suited for generating a large number of good parametrized feature function candidates.

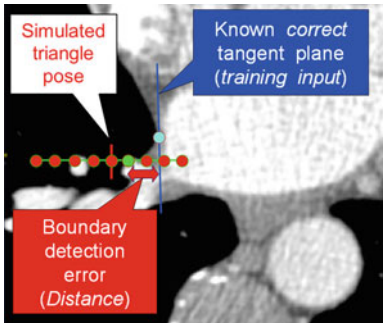


Fig. 5 Sketch illustrating the basic idea of "Simulated Search". Along the simulated search profile (*red points*), the *green* target point is detected and its distance to the reference plane is determined

4.4 "Simulated Search"

"Simulated Search" [23, 24] is an approach to rate the boundary detection performance of a feature function for a given triangle on the basis of training images and corresponding reference meshes. Starting with a set of feature function candidates, it is therefore possible to select the best performing feature function per mesh triangle and assign an optimal feature function $F_i(\boldsymbol{x})$ to each triangle i.

The key idea of "Simulated Search" is to simulate boundary detection for an individual triangle as illustrated in Fig. 5. By shifting the position $\boldsymbol{c}_i$ of a triangle as given by the reference mesh in a pre-defined neighborhood, we construct a pool of *simulated search profiles*. Also, the orientations of the search profiles may be varied to cover a certain range around the reference triangle's normal direction. We then perform boundary detection according to Eq. (13) for a feature function out of the set of candidates and record for each simulated search profile the geometric distance between the detected target point $\boldsymbol{x}_i^{\text{target}}$ and the reference plane as defined by the normal of the reference triangle. This procedure is performed for all reference meshes. Finally, we compute the root-mean-square (RMS) error from all geometric distances that have been recorded for a specific triangle and a specific feature function, and assign the feature function candidate with the smallest average RMS error as feature function $F_i(\boldsymbol{x})$ to triangle i.

"Simulated Search" reflects the goal to put the mesh triangles as close as possible to the desired boundary as defined by the reference meshes in the training images. It addresses the question to what extend a feature function will accurately detect the desired boundary during model adaptation, and is therefore complementary to other approaches that construct highly characteristic feature functions by statistical analysis of gray-value profiles [20, 25] or constraining gradient-based boundary detectors [15–19, 26].

It is important to mention that the concept of "Simulated Search" relies only on the boundary detection mechanism described in Sect. 4.1, but not on the specific type of feature functions described in Sect. 4.2. Essentially, any approach leading to some plausible form and parametrization of the candidate functions $F(\boldsymbol{x})$ can be tested and compared. As example, "Simulated Search" was used to demonstrate

improved boundary detection for prostate segmentation in MR images with modified scale-invariant feature transformation (SIFT) features [27].

4.5 Boundary Detection Training

For boundary detection training, we typically use the feature functions of Sect. 4.2 and feature parametrization of Sect. 4.3 to generate a set of feature function candidates. In particular, we start with a set of feature templates in form of predefined sets S of criteria $Q_j(\boldsymbol{x})$ and perform a statistical analysis of all training images in the vicinity of the reference surfaces. Per set S, the values of the criteria Q_j are used to subdivide the mesh into appearance regions using K-means clustering. Per cluster k, we create acceptance intervals $[\mathrm{Min}_{k,j}, \mathrm{Max}_{k,j}]$ reflecting the observed boundary appearance. The results lead to a list of feature function candidates with a form as defined by Eq. (17). In the final step, we assign to each triangle the feature function candidate that minimizes the boundary detection error resulting from "Simulated Search".

Boundary detection training using "Simulated Search" has been compared to the clustering-based approach for defining feature functions [24]. In particular, experiments have been performed using the 4-chamber heart model and the CT data that were also used for the experiments with the PCA-based shape model (see Fig. 3). Between 80 and 120 different initializations have been generated and either adaptation of global pose and scaling (similarity transformation), adaptation of multi-affine model parameters or SCDM adaptation has been performed. Similar to SCDM adaptation, model adaptation is done in the former two cases by iterating boundary detection and minimization of the external energy of Eq. (8) with respect to the parameters used to transform the mean shape model. The results were used to define a capture range, i.e., the maximum of the constrained surface-to-surface distance between the initially placed model and the reference model up to which an accurate segmentation is obtained. In addition, the segmentation accuracy has been defined as the mean of the constrained surface-to-surface distance between the adapted model and the reference model for cases with initial model placement within the capture range. Results are summarized in Table 1. The results show that "Simulated Search" increases the capture range and improves adaptation accuracy compared to a cluster-based definition

Table 1 Capture range and segmentation accuracy for 4-chamber heart segmentation in CT data when using a clustering-based approach for defining feature functions and boundary detection training using "Simulated Search" [24]

	Capture range (mm)		Segmentation accuracy (mm)	
	Clustering	Sim. search	Clustering	Sim. search
Similarity transformation	18.3	19.7	3.27	3.16
Multi-affine transformation	4.3	7.6	1.63	1.33
SCDM adaptation	4.0	6.4	1.02	0.76

of feature functions. This effect is more pronounced for adaptation of multi-linear model parameters or SCDM adaptation than for adaptation of global pose.

5 Segmentation Framework

Within the previous sections, the basic concepts of SCDMs, an approach to model shape variability and a method for boundary detection training have been presented. Using these concepts, we now describe how to generate and train a model. Model adaptation is done in several steps comprising localization and adaptation with increasing degrees of freedom. Finally, we show that the segmentation framework can be applied to different modalities and different anatomical structures.

5.1 Model Generation

The generation of a SCDM consists of several steps (Table 2). Initially, a suitable mesh model is generated. The starting point can be a single binary or multi-label image from which the mesh model may be generated by triangulation. More complex models like the 4-chamber heart model may be generated by fusing different mesh parts [26]. There is a multitude of different options and many tools developed in the field of computer graphics can be used in this context. In a second step, the mesh model can be subdivided in order to define the multi-linear transformation for parameterizing shape variations.

The next steps cover model training. Robust, reliable and accurate adaptation of a model to new images requires training on a representative set of images. As manual annotation from scratch is very time consuming—especially for complex models like a 4-chamber heart model with great vessels or a brain model—we follow a bootstrap approach that starts with manual annotation of only a few reference images. The mesh model is adapted to the resulting binary or multi-label images as it has been described in the context of constructing PDMs [8]. These images and the resulting corresponding meshes are used to train preliminary boundary detectors and to define an initial shape model. This results in a first model that can be adapted to images.

To augment our training database, we adapt this initially trained model to further training images, select those for which reasonable segmentation results are obtained, manually correct the results, and train a next model. This procedure is iterated until the desired set of training images has been annotated. In addition to reducing manual interaction time for annotations, this bootstrapping approach has the advantage of enforcing consistent annotations as defined in the first fully manual annotations.

Next to model training, we also use the reference annotations to evaluate the segmentation performance of a model in cross-validation experiments and to optimize the overall configuration of the adaptation chain. The bootstrap approach for generating annotations may lead to a bias and the segmentation error may be underestimated

Table 2 Different steps within the construction of a SCDM

Input	Processing	Output
Single label image	Triangulation/meshing	Mesh model
New label images	Adaptation of mesh model	Coarse segmentations
Corresp. gray-value images	Manual corrections	Reference segmentations
Gray-value images	Boundary detection training	Initial boundary detectors
+ reference segmentations	PCA analysis	Initial shape model
New gray-value images	Model adaptation	Coarse segmentations
+ manual model initialization	Selection and manual corrections	Reference segmentations
Augmented training set	Repeated training	Updated model

in such experiments, especially if manual corrections were not done thoroughly. It is, however, very difficult and extremely time-consuming to generate a sufficient number of consistent 3D annotations for complex models independently, and we consider the bootstrap approach as a useful and practical approach.

5.2 Adaption Chain

There are two challenges related to model adaptation. First, the organ to be segmented must be localized in the image. The result can be used for initial model placement. Especially for complex models, the initialization may be good for some parts, but less good for other parts. Robust adaptation of complex models after initial model placement is therefore the second challenge. Addressing both challenges in the context of SCDMs will be explained at the example of heart segmentation in CT images [9, 14].

The Generalized Hough Transformation (GHT) [28] is one approach to localize a shape in an image. A limitation of this approach is that the GHT requires a so-called accumulator array with a dimension given by the degrees of freedom of the pose of the shape to be localized. If, for instance, the shape may be at different positions and different orientation in a 2D image, the Hough accumulator has 3 dimensions (2 for translation and 1 for rotation). In the case of a 3D image, the Hough accumulator would need 6 dimensions (3 for translation and 3 for rotation), making its application unfeasible. Application of the GHT for localizing anatomical structures in 3D images is essentially enabled by the fact that patients are scanned in standardized poses and that the orientation of organs shows little variability. In many cases, it is therefore sufficient to consider translations and possibly scaling (see [29] for examples). In particular, the GHT has been used to detect the heart and initialize subsequent model adaptation [9]. Next to the GHT, Hough forests [30, 31] and classification approaches [32] have recently been used for the localization of anatomical structures in 3D images.

Table 3 Segmentation error after different adaptation steps [9]

Stage of the chain	Mean constrained surface-to-surface error (mm)		
Localization (GHT)	8.14	8.14	8.14
Similarity transformation	3.46	3.46	–
Multi-affine transformation	1.30	–	3.12
SCDM adaptation	0.82	0.96	2.27

Initial model placement is only a starting point. Typically, the final pose and shape of the complete model or some of its parts need considerable changes, and an immediate SCDM adaptation is not robust enough. E.g., some parts of complex models may be quite distant from their final destination, no matter how we place the mean model in the image. As a result, they would adapt to wrong image structures. To overcome these problems, we adapt the model in several steps, starting with very few degrees of freedom and adding more and more degrees of freedom later. For the first steps, a (low-)parametric model is defined by transforming the mean shape

$$\bar{\boldsymbol{m}}_i(\boldsymbol{q}) = \mathscr{T}(\boldsymbol{q})[\bar{\boldsymbol{m}}_i] \tag{18}$$

with a transformation $\mathscr{T}$ that depends on the parameters $\boldsymbol{q}$. Parametric model adaptation is done by iterating boundary detection and minimization of the external energy (8) with respect to the parameters $\boldsymbol{q}$. Several parametric adaptation steps that successively increase the degrees of freedom may be done before the result is finally refined using SCDM adaptation.

For heart segmentation in CT images [9], parametric model adaptation with one global similarity transformation was used to refine pose and size of the 4-chamber heart model after GHT-based heart localization. Then, parametric model adaptation using a multi-affine transformation with $K = 5$ (see Sect. 3.2) was used. For final refinement, SCDM adaptation was used. Table 3 shows the segmentation error after each step of the adaptation chain for 28 cardiac CTA images with different heart phase from 13 different patients. The table also shows that errors of subsequent adaptation steps become larger if one of the parametric adaptation steps is omitted.

Additional adaptation mechanisms have been introduced in [14] to allow for the efficient adaptation of even more complex models such as the heart model with the great vessels of Fig. 4. Most importantly, model parts can be inactive during initial phases of model adaptation and are activated after a pre-defined number of iterations has been performed. In that way, a model part that has already been adapted to the image can be used to initialize the pose of another model part. This mechanism has been used to segment the aorta, pulmonary veins, superior vena cava, inferior vena cava, and coronary sinus by successively activating and initializing vessel segments after adapting the heart chambers to an image. To make model adaptation efficient, mechanisms have been introduced to initialize a high resolution mesh model by first adapting a corresponding model with a lower resolution, and to freeze already well adapted model parts while model adaptation continues for other parts.

With these mechanisms, segmentation of the heart with the great vessels in CT images can be accomplished in 12s (PC with two 2.33 GHz Dual-Core Intel Xeon Processors and 4 GB RAM).

5.3 Application to Multiple Modalities

All previous examples refer to the segmentation of anatomical structures in CT images. The entire segmentation framework is, however, not tailored for a specific modality and can easily be adapted or extended to other modalities if suitable boundary detection functions can be defined and trained.

One particular property of the boundary detection functions of Eq. (17) is the use of characteristic acceptance intervals $[\mathrm{Min}_j, \mathrm{Max}_j]$. Using such intervals presumes the existence of a calibrated gray-scale such as the Hounsfield scale in CT that is well-suited to characterize several tissues. To enable use of these features for imaging modalities such as MRI that do have uncalibrated gray values, we use a simple histogram-based calibration scheme (see also [33]). In particular, we determine the gray-value histogram of the image and map the gray-value interval between the L- and $(100 - L)$-percentile linearly to a standardized interval. This calibration makes the appearance of images acquired with the same protocol from different individuals more similar. We do, however, not transform the entire image, but include this calibration step in the boundary detection functions.

The 4-chamber heart segmentation has been adjusted for a steady-state free-precision MR protocol used to inspect the coronary arteries for ischemic disease. The 42 images have been acquired on Philips Intera and Achieva 1.5 T scanners and have an in-plane resolution of $0.5 \times 0.5\,\mathrm{mm}^2$ to $0.7 \times 0.7\,\mathrm{mm}^2$ and a slice distance between 0.7 and 0.9 mm. Compared to the 4-chamber CT segmentation, several parameter settings have been modified (see [34] for more details). For instance, the damped gradient in direction of the mesh normal in Eq. (16) was replaced by its absolute value. Furthermore, an additional parametric model adaptation step with a global affine transformation was added in between the parametric adaptations with the global similarity and the multi-affine transformation. Table 4 shows the results for segmentation without gray-scale calibration and with gray-scale calibration. The table also includes the fraction of triangles with a mean error below 1.0 mm, between 1.0 and 2.0 mm, and larger than 2.0 mm. For comparison, respective values for 4-chamber CT segmentation are also included.

Table 4 shows that gray-scale calibration reduces the overall segmentation error by 0.57 mm. Even more important is that the percentage of triangles with a large error >2.0 mm is reduced from 19.9 to 0.7 %. This represents a really huge improvement from an application point-of-view, because the regions with a large segmentation error usually need manual corrections. Overall, the resulting segmentations are as good or even slightly more accurate than the results for CT heart segmentation.

Table 4 Mean constrained surface-to-surface error for 4-chamber heart segmentation in CT and MR images [9, 34]. In addition, the percentage of triangles with mean error in distinct ranges is given

	Mean error (mm)	Triangles within error range (mm)		
		<1.0	1.0–2.0	>2.0
CT	0.82	73.7 %	25.9 %	0.4 %
MR (without calibration)	1.33	42.7 %	37.4 %	19.9 %
MR (with calibration)	0.76	76.6 %	22.7 %	0.7 %

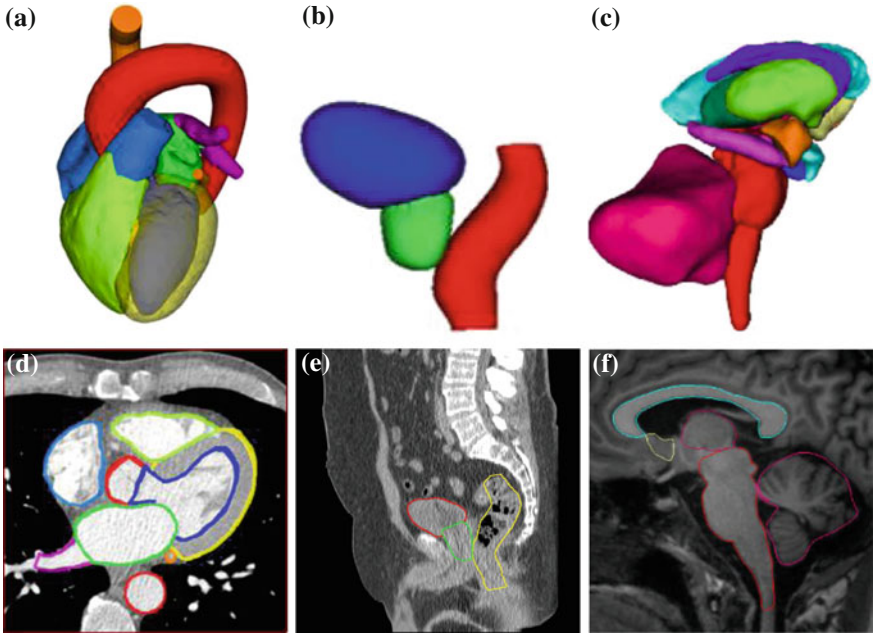


Fig. 6 Sample models (**a**–**c**) and segmentation results (**d**–**f**) obtained with the SCDM framework (see [14, 35])

5.4 Application to Different Organs

Implementation of the entire segmentation framework has been done in a way that model information (shape model, boundary detectors, adaptation parameters) is separated from the algorithms used for model initialization and adaptation [14]. The segmentation framework can, therefore, not only be used for different modalities, but also for a variety of anatomical structures. For illustration, Fig. 6 shows exemplarily models of the heart chambers, of male pelvic structures (bladder, prostate, rectum), and several brain structures together with sample segmentation results.

Since the mesh model preserves its topology during adaptation, the framework is particularly well suited for anatomical structures showing limited variability between individuals such as, but not limited to bones, the heart, and deep brain structures. Without extensions, the framework cannot handle anatomical changes or pathologies that imply topological changes of the shape. For instance, the heart model of Fig. 6 cannot be used to segment hearts with a complex defect such as the tetralogy of Fallot.

6 Applications

The SCDM framework has been applied to segment different organs (e.g. heart, brain, vertebrae, abdominal organs) in different imaging modalities (e.g. CT and MR). In this section, we describe four sample applications of the SCDM framework: accurate left ventricle segmentation for functional analysis in CT, aortic valve segmentation to support TAVI planning, left atrium segmentation to guide AF ablation procedures, and volumetry of brain structures using MR images for evaluating TBI patients. In the context of the sample applications, we outline extensions of the SCDM framework aiming at a further improvement of the segmentation accuracy, the derivation of measurements from the segmentation result and the handling of anatomical variants.

6.1 Left Ventricular Function from CT Images

6.1.1 Motivation

Left ventricular (LV) ejection fraction is a widespread biomarker to characterize heart function [36, 37] and has been used in numerous studies to select or characterize patient populations. In addition, numerous methods have been developed for the efficient segmentation of the LV in images of different modalities [38, 39] that are the basis for computing the ejection fraction. Still, accurate determination of the ejection fraction from clinical images requires often considerable user interaction and its accurate automatic determination remains a challenging task.

SCDMs have been used to segment the LV in cine MR images [18] and the four heart chambers in CT and MR images [9, 14, 34]. Results of the segmentation accuracy suggest that SCDMs as described in the previous sections are suited for the automatic determination of chamber volumes and related quantities from CT images. Indeed, several related studies have been carried out for specific patient groups. They suggest that this approach allows to obtain accurate, clinically relevant results for LV volumes and LV ejection fraction [40], that it results in better accuracy and time savings when compared to specific manual and semi-automatic methods [41], and that it enables reproducible global heart function to be obtained rapidly [42].

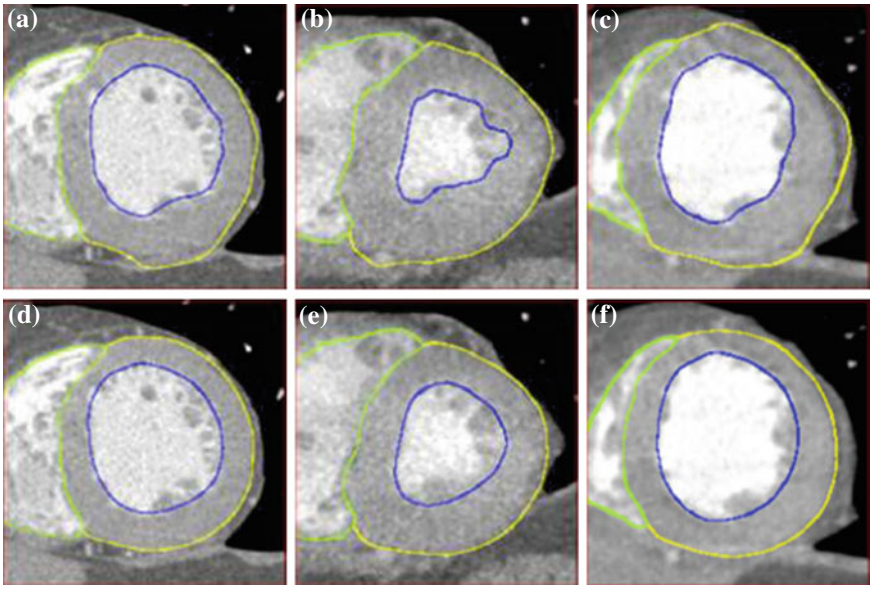


Fig. 7 LV segmentations using the algorithm of Ecabert et al. [9] (**a**–**c**). Papillary muscles are not consistently cut off and occasionally the pericardium is segmented instead of the epicardium. With penalized "Simulated Search" and the improved adaptation chain (**d**–**f**), more accurate and consistent LV segmentation results are obtained [43]

It is also desirable to use SCDMs for a more detailed analysis of the LV function and to determine quantities like regional wall motion or wall thickening that are very relevant in the context of myocardial ischemia. A closer inspection of some segmentation results (see Fig. 7) shows, however, that occasionally the pericardium is segmented instead of the epicardium and that papillary muscles and trabeculations may be cut off arbitrarily. For the assessment of regional LV function, a more accurate and highly consistent segmentation of the LV epi- and endocarium is therefore needed.

6.1.2 Accurate LV Segmentation and Penalized "Simulated Search"

As explained in Sect. 4.4, "Simulated Search" selects the feature function that detects a boundary close to the reference annotation with the smallest average boundary detection error. If there are two boundaries close to each other as in the case of the epi- and pericardium, "Simulated Search" may select a feature function that detects either of both boundaries reliably, but that does not reliably discriminate between them. Similarly, "Simulated Search" will select a feature function that detects the closest nearby boundary, even if the reference annotation does not mark a visible boundary as in the case of the papillary muscles. To obtain a segmentation that

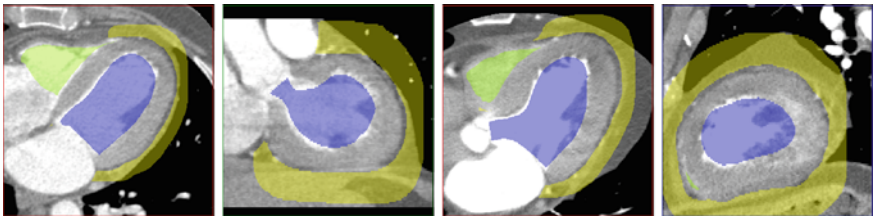


Fig. 8 Masks defining the regions where boundary detection is penalized [43]

(1) consistently segments the epicardium and rejects nearby competing boundaries and (2) avoids attraction towards the bloodpool in regions of papillary muscles and trabeculation, the concept of "Simulated Search" must be extended.

To take the requirements (1) and (2) into account, regions containing unwanted boundaries were explicitly annotated (see Fig. 8). With penalized "Simulated Search", it is counted for a given feature function, how often a boundary is detected in these regions, and for feature selection the average RMS error is penalized by adding a term proportional to the frequency of unwanted boundary detection. A weighting factor balances the tradeoff between avoiding unwanted boundaries and detecting boundaries geometrically accurate [43].

For accurate LV segmentation penalized "Simulated Search" is combined with a modified adaptation chain. After GHT-based model placement and parametric model adaptation, deformable adaptation is applied twice. In the first pass, a large search range (± 10 mm) is used for boundary detection according to Eq. (13) and detection of unwanted boundaries is strongly penalized. This will minimize the detection of boundaries inside the penalized regions. If no suitable boundaries can be detected for some model parts, adaptation is guided by the shape constraints. In the second pass, a small search range (± 2 mm) is used and detection of unwanted boundaries is weakly penalized. As many unwanted boundaries are outside the search range, the second pass improves local segmentation accuracy.

6.1.3 Results

To demonstrate the improvements resulting from penalized "Simulated Search" with the modified adaptation chain, reference segmentations and masks were defined for 67 CT images from 33 patients. Quantitative results from 3-fold-cross validation show that the surface-to-surface error decreased from 1.26 to 0.76 mm for the endocardium and from 0.96 to 0.68 mm for the epicardium [43]. The examples in Fig. 7d–f illustrate qualitatively that LV segmentation becomes more consistent and accurate.

In addition, the local wall thickness has been computed. For that purpose, the normal of the endocardial wall was estimated for endocardial triangles from a regression plane fitted through the triangle vertices and their neighbors. A ray was then cast from the triangle center outwards along the normal, and the distance to the epicardial wall was taken as wall thickness. Figure 9 shows the result for the sample segmen-

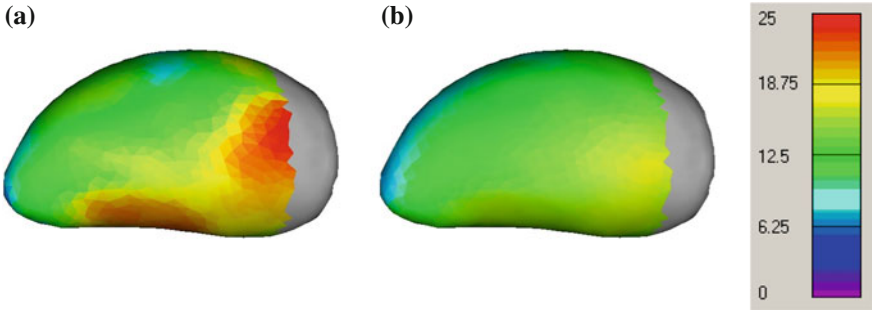


Fig. 9 Wall thickness encoded in color of the LV models shown in Fig. 7b (**a**) and e (**b**). The example illustrates the reduction of LV wall thickness variations by an accurate and consistent LV segmentation [43]

tation of Fig. 7b and e. Variations of the wall thickness resulting from inconsistent LV segmentation are reduced by penalized "Simulated Search" with the modified adaptation chain.

6.2 TAVI Measurements from CT Images

6.2.1 Motivation

Aortic valve stenosis is an abnormal narrowing of the aortic valve that impedes blood flow. In patients with age over 65, aortic valve stenosis is most often caused by calcification that restricts valve leaflet mobility and prevents proper valve opening and closing. Untreated, aortic valve stenosis can cause heart failure. Since a few years, minimally invasive percutaneous valve implantation is feasible [44, 45]. The artificial valve is mounted on a stent which is delivered through a catheter, the so-called transcatheter aortic valve implantation (TAVI).

Accurate pre-interventional assessment of the valve anatomy is essential for TAVI [46]: the proper stent size must be selected in dependence of the aortic annulus diameter, the risk of occluding the coronary ostia by the valve leaflets or the stent must be evaluated, and a projection direction for interventional X-ray imaging must be determined that allows to image the aortic bulbus without foreshortening during the TAVI procedure. While SCDMs allow to segment the aortic valve anatomy given a suitable model, functional extensions are needed to support the desired measurements.

6.2.2 Aortic Valve Segmentation and Information Encoding

As discussed in Sect. 2.2 and illustrated at the example of PCA models of vertebrae in Sect. 3.1, SCDMs approximately maintain the distribution of mesh vertices as given

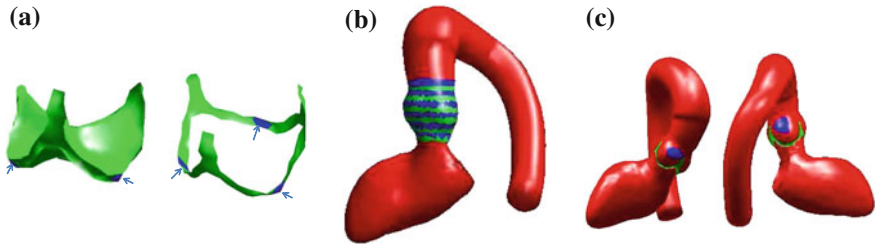


Fig. 10 Information encoded into the aortic valve model for aortic valve plane estimation (**a**), for construction of planes in the LV outflow tract and aortic bulbus (**b**), and for coronary ostia detection (**c**) [47]

by the shape model. A model point can, therefore, be associated with an anatomical location. This property of SCDMs allows to encode information about anatomical structures or landmarks in the shape model and to use this information after model adaptation for subsequent processing steps or measurements.

To support TAVI measurements, a detailed model of the aortic valve, aortic bulbus, and LV outflow tract was created, trained on CT data, and integrated into the heart model of Fig. 6a [47]. Three patches were encoded on the basal ring of the aortic annulus (see Fig. 10a) that allow to determine the aortic valve plane after model adaptation. In addition, rings were encoded in the model from which planes can be derived via regression analysis in the left-ventricular outflow tract and the aortic bulbus (see Fig. 10b). For diameter measurements, the adapted model is cut by the desired plane, and an ellipse or an inner and outer circle is fitted to the resulting contour.

While the planes for diameter measurements can be constructed directly from information encoded into the model, the precise position of the coronary ostia varies between patients. Marking of the vertex closest to a manually determined coronary ostium location in the training images leads to patches with diameters of 10–15 mm (see Fig. 10c). The basic idea of coronary ostia detection is to restrict detection to the well-defined area on the surface of the aortic bulbus defined by these patches. In particular, the candidate patch is determined after adapting the model to the actual image and a complete search is performed on the patch to find a structure defined by a bright half-sphere surrounded by a darker ring. The orientation of the half-sphere and the ring are given by the mesh normal at the patch center and the search is performed for varying radius. The location with the best match of the structure provides the estimated ostia position.

6.2.3 Results

CT data sets from 20 patients were annotated and used to build and train a detailed model of the aortic valve, aortic bulbus, and LV outflow tract [47]. Figure 11 shows exemplary segmentation results of patients with calcified aortic valve. The same

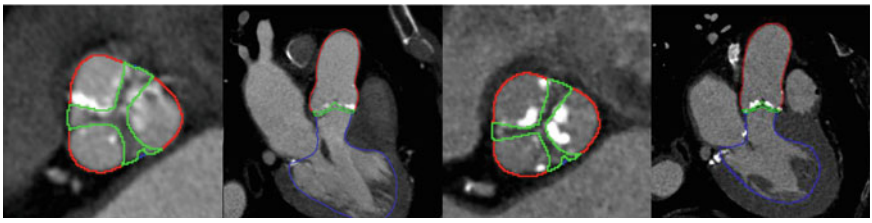


Fig. 11 Examples of aortic valve segmentation [47]

data sets have been used within leave-one-out experiments to assess the accuracy of SCDM-based segmentation and the accuracy of TAVI measurements. The mean surface-to-surface error was 0.5 mm for left-ventricular outflow tract, aortic valve and aortic bulbus. Diameter measurements had a RMS error of 0.8–1.0 mm (aortic valve annulus), 1.0–1.2 mm (LV outflow tract) and 0.6–0.7 mm (mid of aortic bulbus). The RMS error of the distance of the coronary ostia to the aortic valve plane was 0.9 mm for the left side and 0.6 mm for the right side.

SCDM-supported TAVI measurements on CT images have also been assessed in a clinical study [48]. In particular, manual measurements of two users, SCDM-supported measurements of two users, and automatic SCDM-based measurements were compared for 49 patients considered for TAVI and 17 patients without aortic stenosis. The study demonstrates "that a model-based segmentation of the aortic root can be used to objectively and consistently quantify the diameter of the aortic annulus and its distance to the coronary ostia on ECG-gated cardiac CT images" [48]. The study also showed that SCDM-based assessment of the aortic annulus instead of manual measurements on 2D-coronal CT slices would have modified the implantation strategy for 25 % of the patients with aortic stenosis.

6.3 AF Ablation Guidance Using LAPV Models

6.3.1 Motivation

Atrial fibrillation (AF) is a common cardiac arrhythmia characterized by a chaotic contraction of the atrium that is commonly treated in many major hospitals throughout the world by catheter ablation [49]. Within this procedure, tissue around the pulmonary veins (PVs) of the left atrium (LA) is ablated to achieve electrical isolation. X-ray fluoroscopy is commonly used to visualize the catheters during mapping of electrical potentials and ablation. The LA and the PVs are, however, only clearly visible in x-ray fluoroscopy images, if contrast agent is applied. To improve interventional guidance it has been proposed to overlay the anatomy of the LA and PVs onto the fluoroscopy images [50–52]. The anatomy of the LA and PVs can be determined from pre-interventional 3D CT or MR images or from interventional rotational

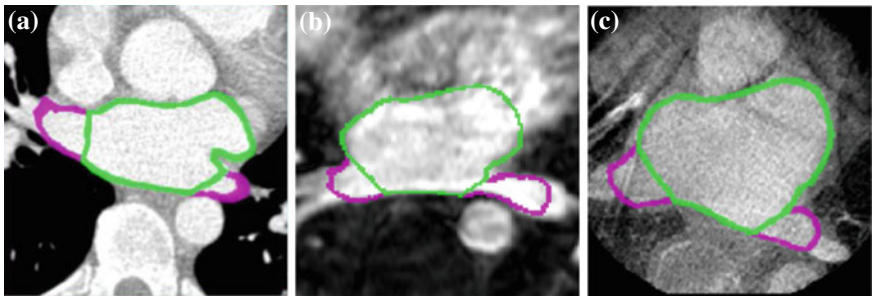


Fig. 12 Exemplary results for LAPV segmentation in a CT (**a**), MR (**b**) and RXA (**c**) image (see [54, 55])

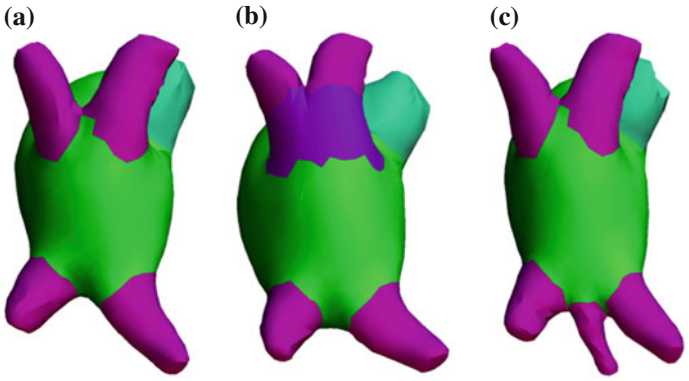


Fig. 13 Models of the left atrium with different pulmonary vein variants: **a** normal, **b** CLT, **c** RMPV [57]

X-ray angiography (RXA) acquisitions [53]. SCDMs allow to construct models of the standard LAPV anatomy from all three modalities [54]. Exemplary segmentation results are shown in Fig. 12.

The pulmonary veins can, however, have different configurations [56] and the actual configuration may have procedural implications. In the majority of patients, the LA is joined by two PVs on each side through individual ostia. The most frequent variation on the left side of the LA is a common trunk (CLT). The most frequent variation on the right side of the LA is the right middle pulmonary vein (RMPV), which is characterized by an accessory PV joining the LA body with a separate ostium and can be observed in 13–24 % of the patients. Figure 13 shows models of the normal LAPV anatomy and the two variants. Especially, the RMPV variant represents a topological change of the shape that SCDMs cannot automatically recognize and segment.

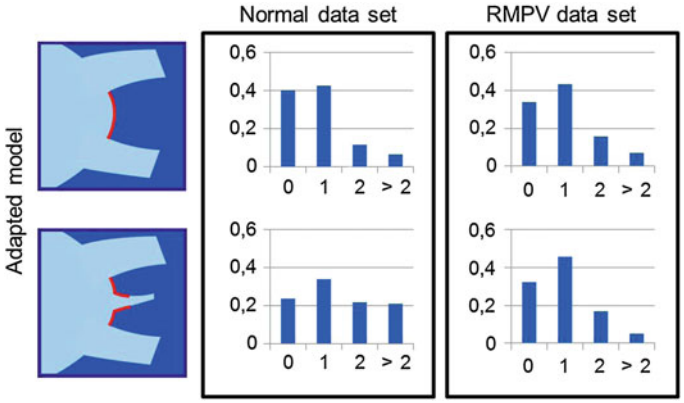


Fig. 14 Average feature vectors evaluated in the pre-encoded region (*red*) for data sets with normal LAPV anatomy and with RMPV anatomy when the normal or the RMPV model is adapted [57]

6.3.2 LAPV Variant Recognition and Segmentation Quality Assessment

SCDMs allow to make individual models for the variants of Fig. 13 and to segment the LA and PVs after selecting the proper model. To recognize the LAPV variant for a given patient image, all different model variants can be adapted, the segmentation quality can be assessed and the best fitting model can be selected [57].

The key element required for this approach is a feature that characterizes the segmentation quality after model adaptation. In particular, a histogram of the unsigned distances between the detected target points and corresponding triangle centers in the pre-encoded regions where the models differ was proposed. This feature vector exploits the fact that the model is only approximately adapted to the detected target points because of the internal energy of Eq. (9) or (12). Figure 14 shows the average feature vectors for data sets with normal LAPV anatomy and with RMPV anatomy when the normal or the RMPV model is adapted [57]. Respective feature was used as input for a support vector machine (SVM) selecting the anatomical variant on the right side of the left atrium. Recognition of the anatomical variant (normal vs. CLT) on the left side of the atrium was done accordingly.

6.3.3 Results

LAPV variant recognition was tested using 59 whole heart MR scans [57]. The anatomy was classified by a clinical expert (31 normal, 8 CLT, 15 RMPV, 5 other), and the images were used to build, train and assess the segmentation accuracy of the models shown in Fig. 13. The overall surface-to-surface error was 1.1 mm. The errors at the PVs were between 0.8 and 1.8 mm. The LAPV variant was correctly recognized in 45 out of the 59 cases. In 3 out of the 14 misclassified cases, a CLT was detected while the LAPV anatomy on the left side was normal, but the segmentation

result was (at least) equivalent. The results are not only of interest for generating personalized LAPV models to guide AF ablation procedures. The overall approach is also a promising extension of SCDMs for anatomical structures showing topological shape variants.

6.4 Brain Structure Volumentry of TBI Patients

6.4.1 Motivation

Traumatic brain injury (TBI) is a significant public health problem worldwide with an average of 1.4 million cases occurring each year in the United States [58]. TBI is caused, for instance, by falls, motor vehicle crashes, struck by or against events, and assaults. CT is used in the emergency department to identify skull fractures, hemorrhages, contusions or edemas. Mild TBI can cause long-term cognitive problems, but often appears normal in conventional CT and MR scans. Advanced MR neuroimaging techniques have, therefore, been applied to investigate the more subtle changes of the brain in mild TBI. Several studies indicate that TBI goes along with morphometric brain abnormalities or brain volume reduction (atrophy) [59]. SCDMs present in this context a unique opportunity for accurate and efficient segmentation of sub-cortical brain structures and atrophy assessment.

6.4.2 Model Generation Using Independent Reference Annotations

As described in Sect. 5.1, we generate reference annotations in the SCDM framework often via manually refining segmentation results obtained with a preliminarily trained model. This is an efficient way for generating consistent reference annotations, but the approach potentially introduces a bias. In the context of the presented TBI results, the reference annotations were generated independently with a different set of tools to avoid a potential bias when investigating subtle volume changes.

Using a software prototype for the interactive manipulation of parametric spline contours [60], the brainstem, caudate, cerebellum, corpus callosum, hippocampus, putamen, and thalamus were traced by experts in 42 high resolution MR volumes (Philips 3T, T1-weighted, sagittal TFE, TR = 6.77 ms, TE = 3.15 ms, flip angle 8°, $1 \times 1 \times 1.2\,\text{mm}^3$ voxel size, spacing between slices 1.2 mm, rows 256, columns 256, slices 140). Triangular meshes were constructed from the contour stack, the mean mesh geometry was obtained for each structure and all mean meshes were combined together, to define the geometry of the brain model shown in Fig. 6c. In addition, the model was trained and the adaptation chain configured.

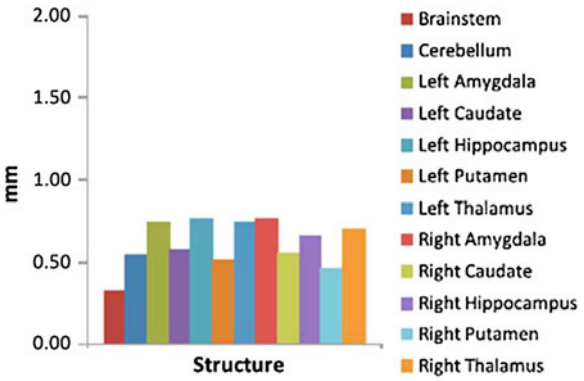


Fig. 15 Mean surface-to-surface error for the different brain structures [35]

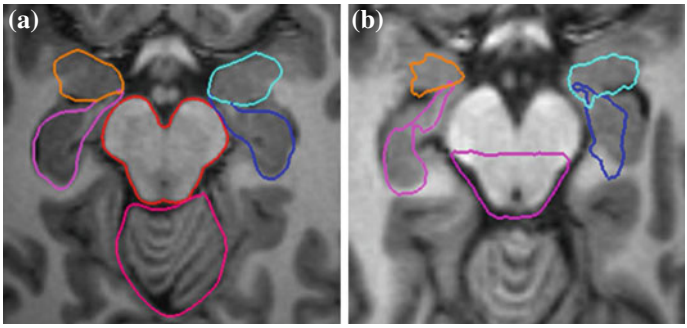


Fig. 16 Exemplary segmentation result are for SCDMs (**a**) and FSL (**b**) [35]

6.4.3 Results

The segmentation accuracy has been assessed quantitatively using 8 expert-traced ground truth data sets in a leave-one-out approach [35]. The mean surface-to-surface error was 0.61 mm. Errors differed between anatomical structures with 0.32 mm for the brainstem and 0.77 mm for the right amygdala. The results for the individual structures are displayed in Fig. 15. Figures 6f and 16a show exemplary segmentation results.

In addition, consistency of the SCDM-based brain structure segmentation has been assessed by segmenting images of 9 healthy control subjects acquired at two different time points. No changes of the volume of the brain structures are expected in this case, and SCDM-based segmentation resulted indeed in little variability (see Fig. 17a). For comparison, the same experiment was performed with FSL/FIRST 4.1.0 [61] showing that SCDM-based volume measurements are more consistent. Figure 16 shows a comparison of a segmentation result obtained with SCDMs and FSL.

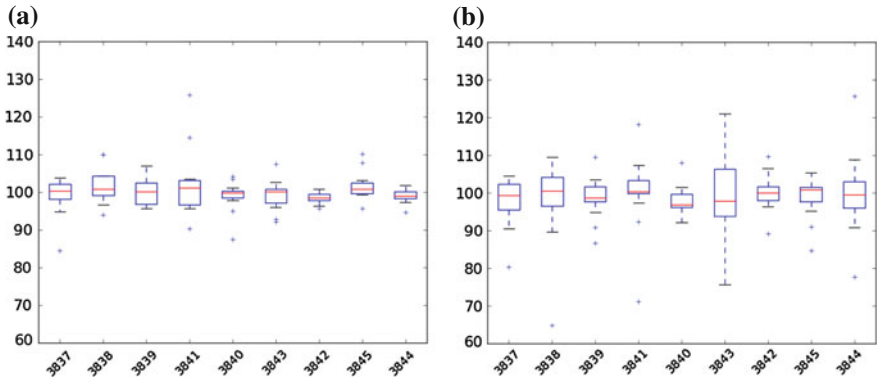


Fig. 17 Box-Whisker plot of the relative volume difference between the two time points for all segmented brain structures for SCDMs (**a**) and FSL (**b**) [35]

Finally, SCDM-based brain segmentation was applied to the retrospective evaluation of a cohort of age-matched male subjects: healthy controls and individuals diagnosed with moderate TBI. Statistical analysis of the resulting volumes detected a significant bilateral volumetric difference in the putamen, thalamus, caudate, and brainstem [35]. These results suggest that the SCDM-based brain segmentation has the necessary sensitivity to detect volumetric changes in individuals affected by (mild) TBI. This could allow for the development of novel imaging biomarkers for diagnosing and monitoring of TBI.

7 Summary

We presented an overview of the SCDM segmentation framework. The framework relies on GHT-based detection of the anatomical structure-of-interest, parametric model adaptation with increasing degrees of freedom and SCDM adaptation. A parametric description of the shape variability is generated by subdividing the structure-of-interest into different parts and assigning a parameterized linear transformation to each part. Optimal boundary detection functions are selected for each model triangle from a large set of boundary detection function candidates using "Simulated Search". Next to describing the basic mechanisms underlying SCDMs, we also presented results for four clinical applications.

Compared to other variants of deformable models, points of a SCDM can approximately be associated with an anatomical location. This property has important implications. SCDMs can be used to establish corresponding points as required for PCA-based shape models. Furthermore, individual boundary detection functions can be assigned to the triangles of the model mesh allowing to segment complex anatomical structures with varying boundary characteristics. This property can also

be used to encode information into the models that can be used for landmark detection or measurements after SCDM segmentation.

The SCDM framework can be adapted to different segmentation tasks and allows to adapt complex anatomical models fast, robustly and accurately to images. This has been shown for a heart model comprising the four chambers and the attached great vessels, and a brain model with amygdala, brainstem, caudate, cerebellum, corpus callosum, hippocampus, putamen, and thalamus. The framework is also suited for images of different modalities. While many specific approaches have been proposed for specific segmentation tasks, we believe that a consistent framework suitable for various anatomical structures and various modalities will contribute to a more consistent interpretation of images in the future.

In the context of the four sample applications, we described extensions of the SCDM framework aiming at a further improvement of the segmentation accuracy, the derivation of measurements from the segmentation result, and the handling of anatomical variants. These extensions facilitate left ventricle segmentation for functional analysis in CT, CT-based measurements for TAVI planning, the generation of models of the left atrium and pulmonary veins for guiding atrial fibrillation ablation procedures, and volumetry of brain structures using MR images for evaluating TBI patients. First related clinical studies have been published and give an indication of the clinical value of the SCDM framework.

References

1. Weese J, Kaus M, Lorenz C, Lobregt S, Truyen R, Pekar V (2001) Shape constrained deformable models for 3D medical image segmentation. In: Insana M, Leahy R (eds) Information processing in medical imaging—IPMI 2001. Lecture notes in computer science, vol 2082. Springer, Berlin, pp 380–387
2. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models—their training and application. Comput Vis Image Underst 61(1):38–59
3. Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. Int J Comput Vis 1(4):321–331
4. Cabezas M, Oliver A, Lladó X, Freixenet J, Cuadra MB (2011) A review of atlas-based segmentation for magnetic resonance brain images. Comput Methods Programs Biomed 104(3):e158–e177
5. McInerney T, Terzopoulos D (1996) Deformable models in medical image analysis. Med Image Anal 1(2):91–108
6. Montagnat J, Delingette H, Ayache N (2001) A review of deformable surfaces: topology, geometry and deformation. Image Vis Comput 19:1023–1040
7. Heimann T, Meinzer HP (2009) Statistical shape models for 3D medical image segmentation: a review. Med Image Anal 13(4):543–563
8. Kaus MR, Pekar V, Lorenz C, Truyen R, Lobregt S, Weese J (2003) Automated 3-D PDM construction from segmented images using deformable models. IEEE Trans Med Imaging 22(8):1005–1013
9. Ecabert O, Peters J, Schramm H, Lorenz C, von Berg J, Walker MJ, Vembar M, Olszewski ME, Subramanyan K, Lavi G, Weese J (2008) Automatic model-based segmentation of the heart in CT images. IEEE Trans Med Imaging 27(9):1189–1201

10. Bruijne M, Ginneken B, Viergever MA, Niessen WJ (2003) Adapting active shape models for 3d segmentation of tubular structures in medical images. In: Taylor C, Noble J (eds) Information processing in medical imaging—IPMI 2003. Lecture notes in computer science, vol 2732. Springer, Berlin, pp 136–147
11. Davatzikos C, Tao X, Shen D (2003) Hierarchical active shape models, using the wavelet transform. IEEE Trans Med Imaging 22(3):414–423
12. Zhao Z, Aylward SR, Teoh EK (2005) A novel 3D partitioned active shape model for segmentation of brain MR images. In: Duncan J, Gerig G (eds) Medical image computing and computer-assisted intervention MICCAI 2005. Lecture notes in computer science, vol 3749. Springer, Berlin, pp 221–228
13. Ecabert O, Peters J, Weese J (2006) Modeling shape variability for full heart segmentation in cardiac CT images. In Reinhardt JM, Pluim JPW (eds) Proceedings of SPIE medical imaging, vol 6144, 6144–3R
14. Ecabert O, Peters J, Walker M, Ivanc T, Lorenz C, von Berg J, Lessick J, Vembar M, Weese J (2011) Segmentation of the heart and great vessels in CT images using a model-based adaptation framework. Med Image Anal 15:863–876
15. Delingette H (1999) General object reconstruction based on simplex meshes. Int J Comput Vis 32:111–142
16. Montagnat J, Sermesant M, Delingette H, Malandain G, Ayache N (2003) Anisotropic filtering for model-based segmentation of 4D cylindrical echocardiographic images. Pattern Recogn Lett 24:815–828
17. Sermesant M, Forest C, Pennec X, Delingette H, Ayache N (2003) Deformable biomechanical models: application to 4D cardiac image analysis. Med Image Anal 7:475–488
18. Kaus MR, von Berg J, Weese J, Niessen W, Pekar V (2004) Automated segmentation of the left ventricle in cardiac MRI. Med Image Anal 8:245–254
19. Sermesant M, Moireau P, Camara O, Sainte-Marie J, Andriantsimiavona R, Cimrman R, Hill DLG, Chapelle D, Razavi R (2006) Cardiac function estimation from MRI using a heart model and data assimilation: advances and difficulties. Med Image Anal 10(4):642–656
20. Brejl M, Sonka M (2000) Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples. IEEE Trans Med Imaging 19(10):973–985
21. Heimann T (2008) Statistical shape models for 3D medical image segmentation. PhD thesis, Ruprecht-Karls-Universität, Heidelberg
22. Heimann T, Münzing S, Meinzer HP, Wolf I (2007) A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation. In: Karssemeijer N, Lelieveldt B (eds) Information processing in medical imaging—IPMI 2007. Lecture Notes in Computer Science, vol 4584. Springer, Berlin, pp 1–12
23. Peters J, Ecabert O, Weese J (2005) Feature optimization via simulated search for model-based heart segmentation. In Lemke H, Inamura K, Doi K, Vannier M, Farman A (eds) Computer assisted radiology and surgery—CARS. International Congress Series, vol 1281. Elsevier, Amsterdam, pp 33–38
24. Peters J, Ecabert O, Meyer C, Kneser R, Weese J (2010) Optimizing boundary detection via simulated search with applications to multi-modal heart segmentation. Med Image Anal 14(1):70–84
25. Cootes TF, Hill A, Taylor CJ, Haslam J (1994) The use of active shape models for locating structures in medical images. Image Vis Comput 12(7):355–366
26. Lorenz C, von Berg J (2006) A comprehensive shape model of the heart. Med Image Anal 10(4):657–670
27. Yang M, Li X, Turkbey B, Choyke P, Yan P (2013) Prostate segmentation in MR images using discriminant boundary features. IEEE Trans Biomed Eng 60(2):479–488
28. Ballard DH (1981) Generalizing the Hough transform to detect arbitrary shapes. Pattern Recogn 13(2):111–122
29. Schramm H, Ecabert O, Peters J, Philomin V, Weese J (2006) Towards fully automatic object detection and segmentation. In: Reinhardt JM, Pluim JPW (eds) Proceedings of SPIE medical imaging, vol 6144, pp 6144–02

30. Gall J, Lempitsky V (2009) Class-specific Hough forests for object detection. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1022–1029
31. Criminisi A, Shotton J, Robertson D, Konukoglu E (2011) Regression forests for efficient anatomy detection and localization in CT studies. In: Menze B, Langs G, Tu Z, Criminisi A (eds) Medical computer vision. Recognition techniques and applications in medical imaging. Lecture notes in computer science, vol 6533. Springer, Berlin, pp 106–117
32. Zheng Y, Barbu A, Georgescu B, Scheuering M, Comaniciu D (2008) Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans Med Imaging 27(11):1668–1681
33. Nyul LG, Udupa JK (1999) On standardizing the MR image intensity scale. Magn Reson Med 42:1072–1081
34. Peters J, Ecabert O, Meyer C, Schramm H, Kneser R, Groth A, Weese J (2007) Automatic whole heart segmentation in static magnetic resonance image volumes. In: Ayache N, Ourselin S, Maeder A (eds) Medical image computing and computer-assisted intervention MICCAI 2007. Lecture notes in computer science, vol 4792. Springer, Berlin, pp 402–410
35. Zagorchev L, Meyer C, Stehle T, Kneser R, Young S, Weese J (2011) Evaluation of traumatic brain injury patients using a shape-constrained deformable model. In: Liu T, Shen D, Ibanez L, Tao X (eds) Multimodal brain image analysis. Lecture notes in computer science, vol 7012. Springer, Berlin, pp 118–125
36. Alfakih K, Reid S, Jones T, Sivananthan M (2004) Assessment of ventricular function and mass by cardiac magnetic resonance imaging. Eur Radiol 14(10):1813–1822
37. Thomas JD, Popovic ZB (2006) Assessment of left ventricular function by cardiac ultrasound. J Am Coll Cardiol 48(10):2012–2025
38. Noble J, Boukerroui D (2006) Ultrasound image segmentation: a survey. IEEE Trans Med Imaging 25(8):987–1010
39. Petitjean C, Dacher JN (2011) A review of segmentation methods in short axis cardiac MR images. Med Image Anal 15(2):169–184
40. Ghersin E, Abadi S, Yalonetsky S, Engel A, Lessick J (2009) Clinical evaluation of a fully automated model-based algorithm to calculate left ventricular volumes and ejection fraction using multidetector computed tomography. Acute Cardiac Care 11:47–55
41. Plumhans C, Keil S, Ocklenburg C, Mühlenbruch G, Berendt FF, Günter RW, Mahnken AH (2009) Comparison of manual, semi- and fully automated heart segmentation for assessing global left ventricular function in multidetector computed tomography. Invest Radiol 44(8):476–482
42. Abadi S, Roguin A, Engel A, Lessick J (2010) Feasibility of automatic assessment of four-chamber cardiac function with MDCT: initial clinical application and validation. Eur J Radiol 74(1):175–181
43. Peters J, Lessick J, Kneser R, Waechter I, Vembar M, Ecabert O, Weese J (2010) Accurate segmentation of the left ventricle in computed tomography images for local wall thickness assessment. In: Jiang T, Navab N, Pluim JPW, Viergever MA (eds) Medical image computing and computer-assisted intervention MICCAI 2010. Lecture notes in computer science, vol 6361. Springer, Berlin, pp 400–408
44. Cribier A, Eltchaninoff H, Tron C, Bauer F, Agatiello C, Sebagh L, Bash A, Nusimovici D, Litzler PY, Bessou JP, Leon MB (2004) Early experience with percutaneous transcatheter implantation of heart valve prosthesis for the treatment of end-stage inoperable patients with calcific aortic stenosis. J Am Coll Cardiol 43(4):698–703
45. Leon MB, Smith CR, Mack M, Miller DC, Moses JW, Svensson LG, Tuzcu EM, Webb JG, Fontana GP, Makkar RR, Brown DL, Block PC, Guyton RA, Pichard AD, Bavaria JE, Herrmann HC, Douglas PS, Petersen JL, Akin JJ, Anderson WN, Wang D, Pocock S (2010) Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. N Engl J Med 363(17):1597–1607
46. Piazza N, de Jaegere P, Schultz C, Becker AE, Serruys PW, Anderson RH (2008) Anatomy of the aortic valvar complex and its implications for transcatheter implantation of the aortic valve tomography. Circ Cardiovasc Interv 1:74–81

47. Waechter I, Kneser R, Korosoglou G, Peters J, Bakker N, Boomen R, Weese J (2010) Patient specific models for planning and guidance of minimally invasive aortic valve implantation. In: Jiang T, Navab N, Pluim J, Viergever M (eds) Medical image computing and computer-assisted intervention MICCAI 2010. Lecture notes in computer science, vol 6361. Springer, Berlin, pp 526–533
48. Korosoglou G, Gitsioudis G, Waechter-Stehle I, Weese J, Krumsdorf U, Chorianopoulos E, Hosch W, Kauczor HU, Katus HA, Bekeredjian R (2013) Objective quantification of aortic valvular structures by cardiac computed tomography angiography in patients considered for transcatheter aortic valve implantation. Cathet Cardiovasc Interv 81(1):148–159
49. Calkins H, Kuck KH, Cappato R, Brugada J et al (2012) 2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design. Europace 14(4):528–606
50. Sra J, Krum D, Malloy A, Vass M, Belanger B, Soubelet E, Vaillant R, Akhtar M (2005) Registration of three-dimensional left atrial computed tomographic images with projection images obtained using fluoroscopy. Circulation 112(24):3763–3768
51. Rhode K, Sermesant M, Brogan D, Hegde S, Hipwell J, Lambiase P, Rosenthal E, Bucknall C, Qureshi S, Gill J, Razavi R, Hill DLG (2005) A system for real-time XMR guided cardiovascular intervention. IEEE Trans Med Imaging 24(11):1428–1440
52. Sra J, Narayan G, Krum D, Malloy A, COooley R, Bhatta A, Dhala A, Blanck Z, Nangia V, AkhtarM (2007) Computed tomography-fluoroscopy image integration-guided catheter ablation of atrial fibrillation. J Cardiovasc Electrophysiol 18(4):409–414
53. Manzke R, Reddy VY, Dalal S, Hanekamp A, Rasche V, Chan RC (2006) Intra-operative volume imaging of the left atrium and pulmonary veins with rotational x-ray angiography. In: Larsen R, Nielsen M, Sporring J (eds) Medical image computing and computer-assisted intervention MICCAI 2006. Lecture notes in computer science, vol 4190. Springer, Berlin, pp 604–611
54. Manzke R, Meyer C, Ecabert O, Peters J, Noordhoek NJ, Thiagalingam A, Reddy VY, Chan RC, Weese J (2010) Automatic segmentation of rotational X-ray images for anatomic intra-procedural surface generation in atrial fibrillation ablation procedures. IEEE Trans Med Imaging 29(2):260–272
55. Weese J, Peters J, Meyer C, Waechter I, Kneser R, Lehmann H, Ecabert O (2009) Patient-specific heart models for diagnosis and interventions. Medica Mundi 53(3):72–78
56. Marom EM, Herndon JE, Kim YH, McAdams HP (2004) Variations in pulmonary venous drainage to the left atrium: implications for radiofrequency ablation. Radiology 230(3):824–829
57. Kutra D, Saalbach A, Lehmann H, Groth A, Dries SP, Krueger MW, Doessel O, Weese J (2012) Automatic multi-model-based segmentation of the left atrium in cardiac MRI scans. In: Ayache N, Delingette H, Golland P, Mori K (eds) Medical image computing and computer-assisted intervention—MICCAI 2012. Lecture notes in computer science, vol 7511. Springer, Berlin, pp 1–8
58. Langlois JA, Rutland-Brown W, Wald MM (2006) The epidemiology and impact of traumatic brain injury: a brief overview. J Head Trauma Rehabil 21(5):375–378
59. Shenton M, Hamoda H, Schneiderman J, Bouix S, Pasternak O, Rathi Y, Vu MA, Purohit M, Helmer K, Koerte I, Lin A, Westin CF, Kikinis R, Kubicki M, Stern R, Zafonte R (2012) A review of magnetic resonance imaging and diffusion tensor imaging findings in mild traumatic brain injury. Brain Imaging Behav 6(2):137–192
60. Zagorchev L, Goshtasby A, Paulsen K, McAllister T, Young S, Weese J (2011) Manual annotation, 3-D shape reconstruction, and traumatic brain injury analysis. In: Liu T, Shen D, Ibanez L, Tao X (eds) Multimodal brain image analysis. Lecture notes in computer science, vol 7012. Springer, Berlin, pp 52–59
61. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33(3):341–355

Part II
Application Cases

Accurate Pathology Segmentation in FLAIR MRI for Robust Shape Characterization

April Khademi, Alan R. Moody and Anastasios Venetsanopoulos

Abstract Shape analysis of pathology requires an accurate initial segmentation. However, in magnetic resonance images (MRI) of the brain, an artifact known as partial volume averaging (PVA) pathology severely impedes segmentation accuracy. Traditional MRI brain segmentation techniques rely on Gaussianmixture models to handle noise and PVA, or high-dimensional feature sets that exploit redundancy in multispectral datasets. Unfortunately, model-based techniques have limited performance on images with non-Gaussian noise distributions and pathology, and multispectral techniques do not make efficient use of imaging resources. For robust segmentation, a generalized PVA modeling approach is developed for FLAIR MRI with white matter lesion (WML) pathology that does not depend on predetermined intensity distribution models or multispectral scans. Instead, PVA is estimated directly from each image using an adaptively defined global edge map constructed by exploiting a mathematical relationship between edge content and PVA. The final PVA map is used to segment WML with sub-voxel accuracy. Using the highly accurately segmented WML, shape analysis experiments were conducted to characterize the types of lesions in the brain. Currently, WML are divided into periventricular white matter lesions (PVWML) and deep white matter lesions (DWML) and radiologists differentiate between them manually. It is important classify these two types of WML since the pathogenic mechanisms between them provide clues regarding the pathophysiology of many diseases (such as MS, stroke, etc.). In this work, we used boundary-based

A. Khademi (✉)
Department of Biomedical Engineering, University of Guelph, 50 Stone Rd E, Guelph, ON N1G 2W1, Canada
e-mail: akhademi@ieee.org

A. R. Moody
Department of Medical Imaging, University of Toronto, 263 McCaul Street, Toronto, ON M5T 1W7, Canada

A. Venetsanopoulos
Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON M5S 3G4, Canada

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis,* Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_6,

and Fourier descriptors to automatically classify the WML into PVWML and DWML classes. A supervised, linear discriminant classifier was used, where a leave-one-out training and testing strategy was employed. It was found that circularity features alone provided the highest classification rate (90%).

A cerebral vascular accident (CVA), or stroke, is an acute neurological injury caused by an interruption of the vascular supply of blood to the brain. Since blood is a carrier of nutrients and oxygen, the affected neurons begin to die within minutes due to oxygen and/or nutrient starvation (ischemic stroke) [1]. Stroke can result in significant neurological deficits, leading to various physical impairments such as sensory motor paralysis, loss of sensation and motor control, as well as difficulties in interpreting spatial relationships [1]; stroke can also be fatal. According to the Canadian Heart and Stroke Society, about 50,000 Canadians suffer new or recurrent strokes each year, which on the average means a stroke occurs every 10 min. It is the third cause of death behind heart disease and cancer, and costs the Canadian economy roughly $3.6 billion a year in physician services, hospital costs, lost wages and decreased productivity [2].

To reduce the mortality rates and long-term disabilities associated with stroke, physicians are investigating Magnetic Resonance Images (MRI) of the brain to determine whether precursors to stroke exist. Identifying the early stages or features of the disease through MRI can lead to the development of intervention protocols and therapeutic strategies which ultimately may reduce the incidence of stroke.

Using Fluid Attenuation Inversion Recovery (FLAIR) MRI, researchers found that abnormal changes in the white matter, known as White Matter Lesions (WML), are a surrogate for future stroke [3]. In FLAIR images, WML appear as hyperintense objects scattered throughout the white matter and have enhanced discrimination of ischemic pathology [4]. The volumes of the lesions (lesion loads) were important markers used to correlate WML and stroke.

However, since the volumes of these lesions were obtained using manual methods, the volumetric measurements are observer-dependent (subjective), error-prone, and are time consuming and labourious to obtain, especially for large patient cohorts. Image analysis techniques offer a great alternative since they can automatically segment WML and compute the volume in a quantitative, efficient, reproducible and reliable manner. Any number of images can be included in the study and lesion loads can be obtained within minutes.

Current manual methods have mainly focused on finding the total volume of the lesions and so the value of automated WML segmentation schemes for volume computation is known. However, other features, that are not readily defined and measured using visual cues and human perception, may hold even more valuable information about the disease. For example, the shape characteristics of the lesions, which cannot be easily defined nor measured by a human observer, could be measured using image analysis techniques. For disease characterization, shape descriptors could possibly describe disease in a unique and novel manner. For example, robust shapes features could possibly differentiate between types of WML as well as neurological diseases

such as Alzheimers and Multiple Sclerosis (MS). These features could also be used to monitor disease progression or response to therapy.

Currently, white matter lesions are divided into periventricular white matter lesions (PVWML) and deep white matter lesions (DWML). DWML are identified as being close to the periphery of the brain ("deep" in the WM), and PVWML as being alongside the ventricles. The pathogenic mechanisms between PVWML and DWML are providing some clues for understanding pathophysiology of many diseases associated with white matter lesions [5]. Since the visual appearance of the lesions in these regions differ, this chapter investigates whether shape analysis techniques can be used to differentiate between DWML and PVWML.

To ensure that the extracted shape features are reliable, a major focus of this chapter is on the design and implementation of robust image processing algorithms for WML segmentation in FLAIR MRI. This is a challenging problem since MRI are degraded by many types of imaging artifacts, such as partial volume averaging (PVA) and acquisition noise. Modern acquisition systems generate even more interesting noise properties, such as spatial correlation, non-Gaussian intensity distributions and nonstationarity [6]. Traditional segmentation techniques that rely on model-based approaches are inaccurate in such modern acquisition systems since predetermining a model for the noise is challenging if not impossible. Moreover, WML, or any pathology cannot be modeled using a "nice" distribution such as those used in model-based approaches.

To combat these challenges, a model-free, efficient approach to segment WML accurately in the presence of noise and PVA is presented. Because the technique is PVA-based, the WML are segmented with subvoxel precision and produce accurate WML boundaries ideal for shape analysis. Both boundary-based and global shape metrics are used to classify the lesions.

1 Background

This section focuses on the background material, such as the motivation for the current work, the principle of magnetic resonance for imaging purposes as well as the FLAIR MRI modality, which is used for WML segmentation and characterization.

1.1 Magnetic Resonance Imaging (MRI): The Fundamentals

Unlike Computed Tomography (CT) and X-Ray imaging, MRI relies on nonionizing radiation to obtain anatomical images of the human body, and has the ability to localize soft-tissue structures. Consequently, MRI has become the leading imaging modality for the examination of brain disease and disorders [4].

Magnetic resonance imaging is based on the principle of nuclear magnetic resonance (NMR), which is the behaviour of nuclei under the influence of externally

applied RF (radiofrequency) and magnetic fields. A nucleus is NMR-Active if it has a nonzero magnetic dipole moment (MDM), created by an odd number of protons and neutrons, such as hydrogen (one proton and one neutron). The human body is comprised of approximately 63 % hydrogen [7] (fat and water) and MR imaging exploits the magnetic behaviour of these hydrogen atoms to generate anatomical images.

In the presence of an externally applied magnetic field, hydrogen protons begin to precess (like tops) around the axis parallel to the external magnetic field. The net effect of these MDMs (the bulk magnetization $\overrightarrow{M}$) is a vector that lines up with the external field $\overrightarrow{B}_0$ (the transverse components cancel each other out). Once the bulk magnetization has lined up with the external field, an RF field B_1 is applied to excite the nuclei. The frequency of the RF field is the same as the rate of precession of the nuclei (Larmor rate) so that energy from the excitation may be absorbed. Only the atoms that are precessing at the Larmor rate will be excited and reach a higher energy state. Once the RF field is removed, the nuclei begin to return to their original state, releasing energy in the process. The energy released in the transverse plane is known as the T1-signal and the energy released in the longitudinal plane is known as the T2-signal. Different tissues and pathologies have different T1-T2 relaxation times (which translates into different contrast in the output image). Because regions of the brain are composed of more or less water/fat, soft-tissue structures of the brain are imaged with good results.

The 3D position from which photons were released is learned by applying additional *gradient* fields during the scan. These strong magnetic field gradients cause the nuclei at different locations to rotate at different speeds and makes the magnetic field strength vary depending on the position within the patient. This in turn makes the frequency of released photons dependant on their original position in a predictable manner, and the original locations can be mathematically recovered from the resulting signal by the inverse Fourier transform. In other words, gradient fields selectively excite regions in the patient that are used to decode the corresponding pixels in the image.

The result is a volumetric data set which offers a 3D representation of anatomy and pathology. Each image (slice) is acquired in a specific direction, which is parallel to an imaging plane. There are three perpendicular imaging planes that are most commonly used: axial (horizontal), sagittal and coronal. An example sagittal T1-weighted MR dataset, acquired from a single-coil scanner, is shown in Fig. 1.

The area the gradient field excites determines the resolution of the pixels (voxels) in the image, i.e., if a smaller area is excited, the resolution of the pixel in the final image increases. Slice thickness works the same way, i.e., if a large region in the z-direction (thick slice) is excited, the resolution in this dimension is low. It is possible to generate images of very high resolution, by exciting smaller and smaller volumes, except that this would require many more excitations and gradients to be applied, which significantly increases scan time.

Long acquisition times are a natural side effect of MR imaging. Each slice needs to be spatially encoded, which requires a series of excitations and during each excitation,

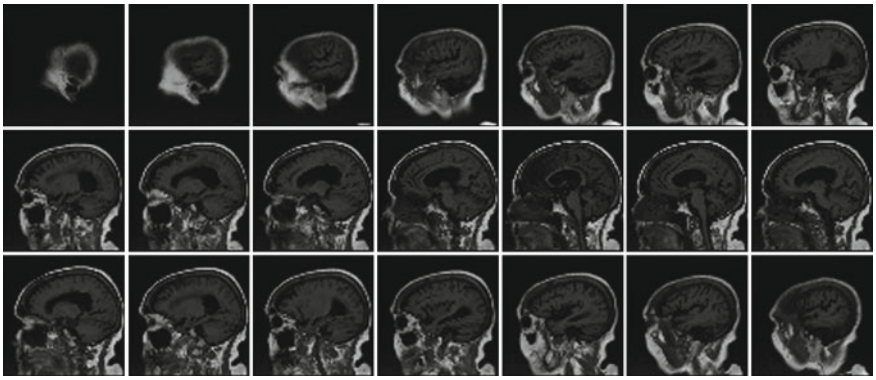


Fig. 1 MR volume of the cerebral region acquired in the sagittal plane (slice thickness is 5 mm, spacing between slices is 6 mm, pixel resolution is 0.4297 mm $\times$ 0.4297 mm)

a line in the Fourier domain (k-space) is acquired. Between excitations it is necessary to wait for the excited spins to return to equilibrium again (relaxation). Therefore, relaxation time and the number excitations have direct impact on the acquisition time.

In single-coil technologies, to speed up scanning, moderate values for the pixel height, width and slice thickness are usually chosen. Moreover, there is usually a gap left between adjacent slices, to facilitate even more efficient scanning times. Using larger volume elements creates faster scanning times at the expense of resolution. Resultantly, there have been many research efforts dedicated to overcoming the balance between resolution and scan time. For example, single-shot techniques such as Fast Low Angle SHot (FLASH) [8] or Echo Planar Imaging (EPI) [9]. The acquisition times of these techniques are significantly shorter, but the images are in general of lesser quality. They suffer from low signal-to-noise ratios (SNRs) and are very sensitive to the inhomogeneities of the main magnetic field [10]. Details on single-coil technologies and the way they affect noise are discussed in the following section.

The latest and most innovative methods for speeding up acquisition times are achieved by sparsely sampling the k-space data (which reduces the number of required excitations). However, reduction of the sampling rate in the frequency domain (k-space) violates the Nyquist theorem of perfect reconstruction, and consequently there is aliasing in the spatial domain images. To compensate for the data loss in the spatial domain, multiple receiver coils and Parallel MRI (pMRI) reconstruction techniques can be used to recover the missing data.

The principle of pMRI is focused around the use of multiple receiver coils (phased-array coils or multicoil) to capture the image data [11]. Each coil has a spatially varying sensitivity map that dictates the image reception profile of the coil. Moreover, each coil is positioned so that each has a different sensitivity over the the Field of View (FOV) [12]. The distinct sensitivity profile of each coil is modulated with the

pixel value of the image and since there are multiple sources capturing the same information, the missing k-space information can be calculated [12]. Since these coils are working simultaneously, the introduction of redundancy gives way to a significant reduction in acquisition times. Details on pMRI and the way it affects noise in MRI are discussed in the next section.

1.2 Fluid Attenuation Inversion Recovery (FLAIR) MRI

To automatically analyze WML, Fluid Attenuation Inversion Recovery (FLAIR)-weighted MRI are used because of their ability to localize ischemic brain pathology. FLAIR images have very similar properties to T2-weighted images, in terms of tissue contrast characteristics. In T2-weighted images, water-filled tissues are imaged as bright or high-signal regions, whereas fatty tissues are represented with low intensities. This translates to dark gray and light gray intensities for white matter (WM) and gray matter (GM), high-intensity values for cerebrospinal fluid (CSF).

Most pathology is associated with increased water content [4], so T2-weighted images highlight ischemic pathology such as WML. However, CSF also shows up as high-intensity signal in T2 images which reduces the discrimination of periventricular WML (lesions close to the ventricles). FLAIR MRI overcomes these challenges by removing the T2 mobility so that the CSF signal is nulled (i.e., the intensity of CSF is set to zero).[1] This produces better discrimination of pathology [4] since the CSF signal does not interfere with the signal of the hyperintense pathology (WML). Figure 2 contains the corresponding T2- and FLAIR weighted images for a patient with WML. The ventricular high signal material shown in the T2 images impede delineation of WML, whereas the CSF is dark in FLAIR MRI, providing much better visualization of white matter pathology.

The FLAIR modality has been used in many studies to detect the presence of white matter lesions and other neurodegenerative disorders [4, 13, 14] due to its ability to localize pathology (as shown in Fig. 2). Therefore, this modality is an excellent candidate for the automatic detection and characterization of WML.

Some additional sample FLAIR MRI with varying lesion loads (total WML volume) are shown in Fig. 3. In each image, there are three major tissue classes in the cerebral region (excluding outer head structures):

- CSF: darkest class in the image,
- Brain (both GM and WM): comprised of medium intensity values,
- WML: most intense objects (after brain extraction).

Normally, in MRI, WM and GM are clearly differentiable and compose two separate tissue classes. However, imaging parameters in FLAIR MRI cause GM and WM classes to have similar intensity values and thus are treated as a single class [15].

[1] Mobility refers to how mobile the protons are. Large mobility produces large T2 times (intense regions) and small mobility results in short T2 times (dark regions).

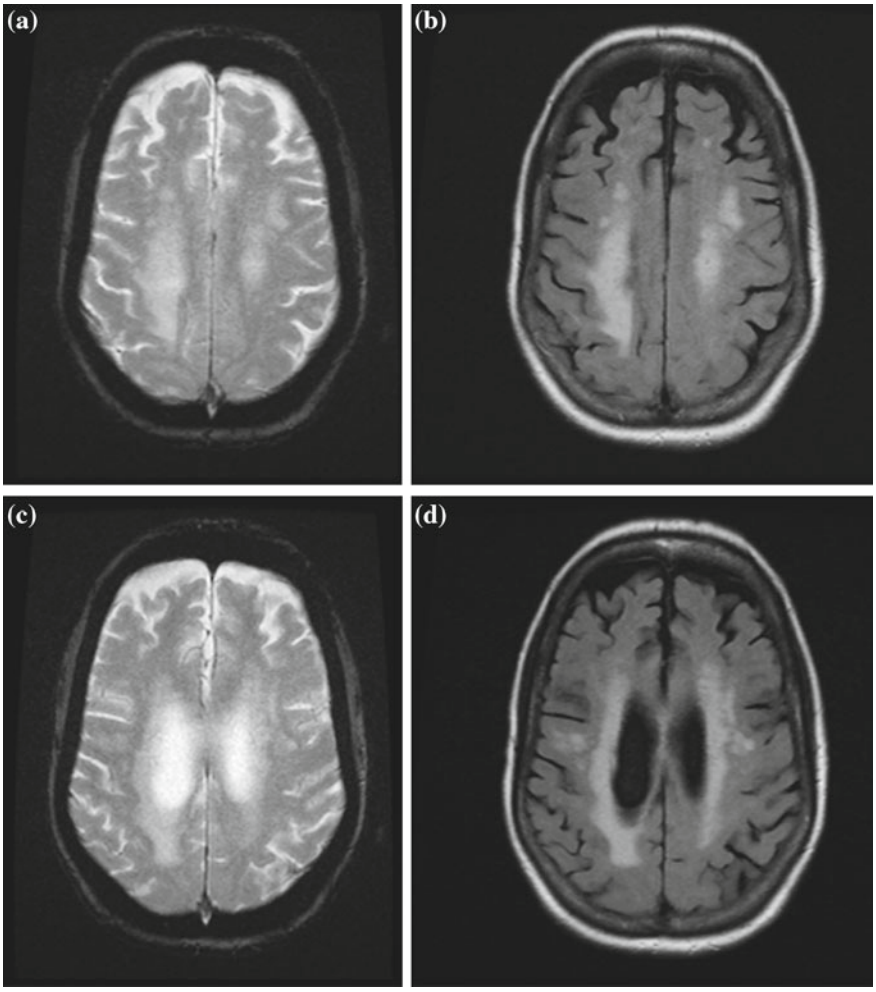


Fig. 2 **a** T2-weighted. **b** FLAIR MRI. **c** T2-weighted MRI. **d** FLAIR MRI. Two T2 images (**a**, **c**) and the two corresponding FLAIR images (**b**, **d**)

Therefore, a tissue segmentation scheme for FLAIR MRI would search for estimates of the CSF, brain and WML classes, individually.

2 Challenges of Segmenting FLAIR MRI

Images with inherent artifacts may cause slight difficulty for human interpretation, but they can cause significant challenges for algorithms. A computer can easily

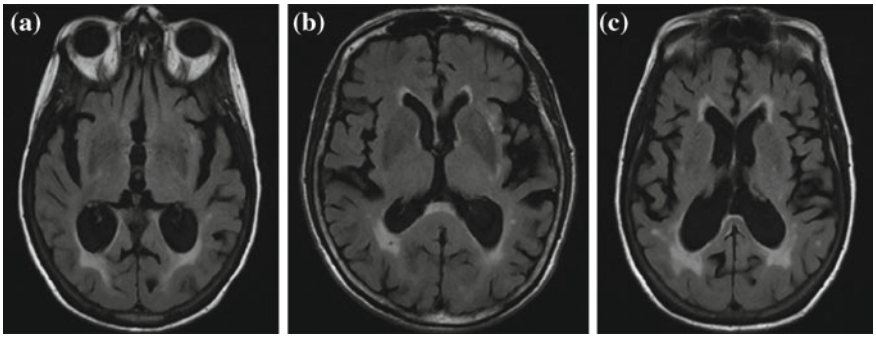


Fig. 3 Three examples of FLAIR MRI with WML. **a** Example 1. **b** Example 2. **c** Example 3

tell the difference between two pixels' intensity values, even if they only differ from one another by a few graylevel values. As a result, imaging artifacts can cause erroneous results in intensity-based segmentation schemes [15], misclassified pixels in automated tissue classification algorithms [16], inaccurate 3D reconstructions, etc. There are many noise sources in MRI that are produced by the imaging acquisition system.

Much research has gone into mathematically modeling the images' noise fields in MRI and there are three main issues: acquisition noise [17], the tissue nonuniformity issue [18] and the PVA effect [17]. Consider an undistorted, "clean" image $f(x_1, x_2)$ distorted by a multiplicative and additive noise source:

$$y(x_1, x_2) = f(x_1, x_2) \times \beta(x_1, x_2) + n(x_1, x_2), \tag{1}$$

where $y(x_1, x_2)$ is the distorted image. In MR images $\beta(x_1, x_2)$ is known as the bias field or tissue nonuniformity artifact, and $n(x_1, x_2)$ is the additive (acquisition) noise field. The bias field is a smoothly varying field caused by inhomogeneity of the magnetic field [19]. The bias field is not discussed in this work[2] so we set $\beta(x_1, x_2) = 1$. The additive noise source, $n(x_1, x_2)$, is generated by acquisition noise from the system [20] and must be considered in any automated MRI segmentation technique.

A third type of artifact is known as partial volume averaging (PVA) and it cannot be modeled with Eq. 1 as it has its own degradation model. In concerns the imaging of finite volume elements that have more than one tissue type/structure within the imaging section, pixel or voxel. The consequence is that the signals of the structure are averaged. As will be shown, PVA and acquisition noise are inherently related with one another, as well as the imaging system that is used to acquire the images. Handling these two artifacts is challenging but is required for accurate WML segmentations. These artifacts are discussed in this section to provide motivation for the proposed segmentation methodology.

[2] Experimental images do not possess significant bias field.

2.1 Acquisition Noise

In MR images, the signal-to-noise ratio (SNR) is intrinsically related to the resolution of the images [21]. As a result, the SNR of the output images may be modest or even low, making acquisition noise more apparent in the output image. Additive noise poses significant challenges for automated WML segmentation since it changes the intensity profile each tissue class; image classes are no longer clearly discernible in the intensity histogram and contrast between tissue classes is significantly reduced.

To combat the downfalls associated with acquisition noise, many researchers have investigated MRI noise characteristics and have tuned their processing technologies to these properties. This requires good knowledge of the acquisition process since noise in the output image is inherently related to the way the image is acquired. There are two dominant technologies used for MRI acquisition that depend on either single-coil technologies, or multicoil phased array systems. These two families create substantially different noise properties in the final image.

2.1.1 Single-Coil Technologies

Single-coil technologies use a single coil to transmit and receive NMR signals, which has direct bearing on the way noise is rendered in the image. It is well known that during acquisition the distribution of the noise within the single coil is Gaussian with zero mean and σ standard deviation [20–23]. Each MR image (slice) is collected from the coil in the Fourier domain (k-space), which is inverted through the inverse Fourier transform to get the spatial representation of the image. The spatial domain image retains the same noise distribution as k-space since the inverse Fourier transform is a linear and orthogonal transformation and does not change the characteristics of the noise.

The spatial domain representation $y(x_1, x_2)$ is formed by summing the imaginary and a real components found by the inverse Fourier transform as in

$$y(x_1, x_2) = y_R(x_1, x_2) + j \cdot y_I(x_1, x_2), \tag{2}$$

where x_1 and x_2 are the spatial coordinates $x_1, x_2 \in Z^2$ and $j = \sqrt{-1}$. Both $y_R(x_1, x_2)$ and $y_I(x_1, x_2)$ are individually corrupted by Gaussian noise, $N(0, \sigma)$. Because this expression contains complex values, it must be transformed in order to gain a visual representation of the data. Typically, the magnitude of Eq. 2 is taken for visualization purposes [22]:

$$|y(x_1, x_2)| = \sqrt{y_R^2(x_1, x_2) + y_I^2(x_1, x_2)}. \tag{3}$$

The modulus operator $|y|$ is nonlinear and transforms noise into a Rician distribution [20, 22, 23]. For high SNR images, the Rician distribution closely approximates a Gaussian. Background regions in MR images contain no signal (SNR of 0) and

consequently, the Rician distribution simplifies to a Rayleigh distribution in non-signal regions.

The underlying noise characteristics for single-coil technologies have been exploited in the design of image processing methodologies. Model-based approaches classify tissues in presence of noise by using predetermined intensity distributions for the tissues (which is known because of knowledge of the image acquisition process). Probability density functions (PDF) of the graylevel values, $p(y|\omega)$, are used to model the image intensity formation process, which is the probability of observing intensity y for some class ω. For normal brain tissues (i.e., no pathology) and single-coil systems, the tissue classes are typically modelled using Gaussian distributions [24–26]

$$p(y|\omega) = \frac{1}{\sigma_\omega \sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{(y-\mu_\omega)^2}{2\sigma_\omega^2}}, \tag{4}$$

where μ_ω is the mean intensity of class ω, σ_ω^2 is the variance of the distribution and $\theta_\omega = \{\mu_\omega, \sigma_\omega^2\}$ are the model parameters which need to be estimated. Each tissue class can have its own unique parameter set $\theta_\omega = \{\mu_\omega, \sigma_\omega^2\}$.

Parameter estimation can be completed by the Expectation Maximization (EM) algorithm [17, 27], which uses an iterative approach to update posterior and parameter estimates such that the log-likelihood is maximized. The joint representation, or normalized histogram, may be reconstructed from these class distributions by

$$p(\mathbf{y}) = \sum_{\omega \in \Omega} P(\omega) \cdot p(y|\omega), \tag{5}$$

where $\mathbf{y}$ is the vector representation of all intensity levels in MRI, $P(\omega)$ is the prior probability of the class ω and Ω is the set of all possible classes. This joint distribution is known as a Gaussian Mixture Model (GMM) [17]. Tissue classification is performed using the parameters and distributions of each tissue class. Understanding the characteristics of the data is of utmost importance in achieving good results, especially with model-based approaches. Newer multicoil phased array systems generate noise that is not as easily handled.

2.1.2 Multicoil Technology

Multicoil MR systems are fast becoming the norm for brain imaging studies. The image statistics in multicoil images depend strongly on the pMRI method used to combine the images from different receiver coils. In this section, we examine the effect that multicoil reconstruction algorithms have on noise.

Following the inverse fast Fourier transform (FFT) operation, each voxel in the image is represented by column vector, $\mathbf{p}$ (one complex value for each of the n coils). The elements of $\mathbf{p}$ are the product of the signal we are trying to measure, A, and the sensitivity profiles of the coils, represented by column vector $\mathbf{b}$

$$\mathbf{p} = A\mathbf{b}. \tag{6}$$

Coil images are corrupted by additive Gaussian noise with covariance matrix, Ψ. Diagonal elements represent the noise variance for each coil, while the off-diagonal elements describe the degree of correlation between coils. Noise correlation between coils is caused by inductive coupling combined with the tendency for coils with overlapping sensitivity profiles to be similarly contaminated by thermal radiation. It is possible to estimate the noise covariance matrix from a simple pre-scan that can be acquired in a few seconds [28].

There are many ways to reconstruct the image from the redundant information collected from each coil. Two of the most common spatial reconstruction algorithms are known as root sum-of-squares [29] and SENSE [30]. Root sum-of-squares is the simplest algorithm for combining images from multiple coils because it does not require any knowledge of the coil sensitivities. If all coils have equivalent noise and are uncorrelated (i.e., $\Psi = \mathbf{I}\sigma$, where $\mathbf{I}$ is the identity matrix), the average noise is equal to $\sigma/\sqrt{2}$ for all voxels in the image. However, if the coils are correlated or have variable noise (which is often the case), the noise in the resulting image will be nonstationary, and the signal will deviate from the noncentral chi distribution.

SENSE is a method for accelerating image acquisition by undersampling k-space data [30]. In its simplest form, regularly spaced lines are skipped along a single dimension. This allows scan time to be reduced in proportion to the number of lines skipped, referred to as the acceleration factor, R. The resulting images have a characteristic aliasing pattern, with voxels from outside of the reduced field of view wrapped around onto other parts of the image.

The SENSE algorithm utilizes coil sensitivity information to unfold these aliased images. SENSE reconstruction produces complex valued images but it is common practice to take the magnitude for display purposes. The SNR and noise resulting from SENSE reconstruction are

$$SNR_{SENSE} = \frac{\sqrt{2} \cdot \left|\mathbf{u}^T\mathbf{a}\right|}{\sqrt{R \cdot \mathbf{u}^T \Psi \mathbf{u}^*}}, \tag{7}$$

and

$$\sigma_{SENSE} = \sqrt{\frac{1}{2} \cdot \mathbf{u}^T \Psi \mathbf{u}^*}, \tag{8}$$

respectively. Although these techniques speed up the acquisition process, there are direct consequences on the noise in the image. In particular, the noise level in a SENSE image varies from pixel to pixel (nonstationary), there is noise correlation between pixels [30] and the noise varies dependent on the coil geometry [31]. As shown in [31], the histogram or distribution of the noise changes substantially depending on the reconstruction method used, the speed up factor, etc. Confounding issues further, there are several variations of the SENSE algorithm, including modified SENSE (mSENSE) and regularized SENSE [31], which all modify the noise statistics in different ways.

There are many other reconstruction techniques, such as those completed in the frequency (k-space) domain. Alternate lines in the k-space are collected and the missing intermediary k-space lines are calculated from the signals recorded by the different coil elements. Usually, this is completed by combining weighted signals from each coil, using methods such as SiMultaneous Acquisition of Spatial Harmonics (SMASH) [32] or Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA) [33]. As shown in [31], the way the missing k-space lines are estimated not only affects reconstruction, but also significantly changes the distribution and characteristics of the noise.

There are non-Cartesian k-space acquisitions such as PROPELLER or variable-density spirals that result in coloured noise properties [34], and application of a Fermi filter prior to reconstruction, which introduces spatial correlations across the reconstructed image as well [35].

As can be seen, many types of noise characteristics can be generated by mulit-coil MRI acquisitions. Resultantly, traditional noise modeling approaches cannot be applied to pMRI-based techniques, since the intensity distributions of the tissues are non-Guassian, nonstationary and could possess correlation as well. Authors are attempting to exploit some of these characteristics in their algorithms. For example, the authors in [36] propose a nonlinear anisotropic diffusion filter which adapts to the nonstationarity of the noise in SENSE images to reduce noise more robustly than traditional approaches [36].

Today, many pMRI methods available and the choice of an optimal method is not straightforward since each method has its own advantages and disadvantages. Moreover, imaging technology is rapidly changing and many of the reconstruction algorithms in MR scanners are proprietary. This impedes the performance of traditional model-based approaches, since they depend on the noise satisfying several "nice" properties.

2.2 Partial Volume Averaging (PVA) Effect

Imaging elements that contain a single tissue are known as "pure" tissue voxels, whereas volumes that contain more than one tissue within the extent of the voxel are called mixture tissues and they constitute the partial volume averaging (PVA) artifact [17]. The final intensity of a PVA voxel is proportional to the intensities of the tissues that exist within the 3D volume that is being imaged. These mixture voxels or mixels, create fuzzy boundaries around image objects and can lead to 30–60 % error in the volume measurement of complex brain structures [37, 38].

Partial volume averaging is most noticeable in regions where different tissues meet. It causes the graylevel transition between two pure tissues to be ill-defined, non-crisp and fuzzy [17, 39], in contrast to an ideal step-like edge, or crisp margin. This is visually demonstrated in Fig. 4a, by the fuzzy halo surrounding several WML. This ill-defined border makes it difficult to determine the location of the lesions' borders automatically. The corresponding one-dimensional scanlines taken from the center of

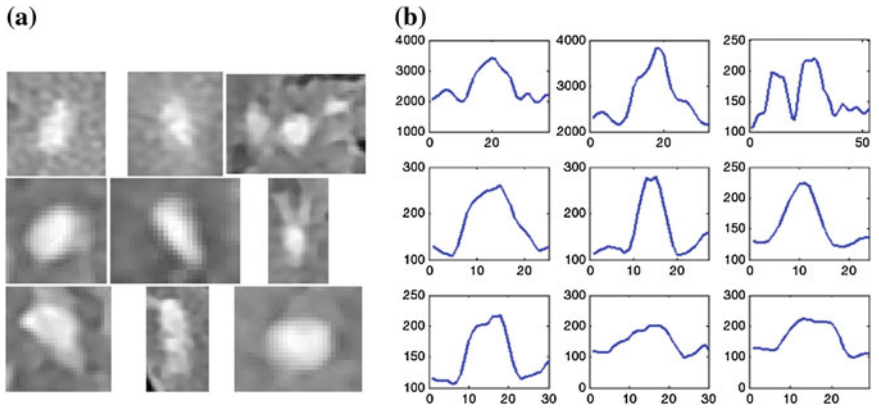


Fig. 4 Nine regions of interest extracted from FLAIR MRI containing WML. **a** WML. **b** 1D profile of (**a**)

each lesion are included in Fig. 4b, which demonstrate the gradual increase/decrease of graylevel values in regions surrounding the lesion (PVA).

According to MR physics, PVA generates an image intensity which is linearly dependant on the proportion of each tissue in the voxel [38, 40]. In neurological imaging studies, two tissue types are commonly present in mixture (PVA) voxels [38]. Therefore, a PVA voxel's intensity at a spatial coordinate $\mathbf{x} = (x_1, x_2) \in Z^2$, is determined by the proportion of tissue 1 present at $\mathbf{x}$, as in

$$Y_{12}(\mathbf{x}) = \alpha(\mathbf{x}) \cdot Y_1(\mathbf{x}) + (1 - \alpha(\mathbf{x})) \cdot Y_2(\mathbf{x}), \tag{9}$$

where $Y_{12}(\mathbf{x})$ is the resultant intensity of the PVA voxel, $Y_1(\mathbf{x})$ is the intensity of the first tissue where $Y_1 \sim p_1(y)$, $Y_2(\mathbf{x})$ is the intensity of the second tissue where $Y_2 \sim p_2(y)$ and α is the proportion of the first tissue present in the PVA voxel where $\alpha \in [0, 1]$. This parameter α is commonly referred to as the tissue fraction.

This equation strictly governs the transition of the graylevel values between two pure tissue classes. Consider Fig. 5, which is the 1D profile of a WML taken from Fig. 4a. This image shows the approximate regions of pure (non-PVA regions) and mixed tissues (PVA regions). The intensity values of PVA voxels are governed by Eq. 9 and pure tissues are drawn from some distribution that is related to the acquisition noise profile.

As stated, the intensity of PVA voxels is proportional to the intensities of the tissues present in the voxel. The "amount of PVA" is dependant on the slice thickness, i.e., thick slices will have a more noticeable PVA effect creating large, blended regions between pure tissues. Consequently, the histograms for each tissue class cannot be described by a single intensity value, but will be defined over a range of values that overlap neighbouring classes as a result of noise and PVA. Obviously this creates challenges for intensity-based segmentation techniques.

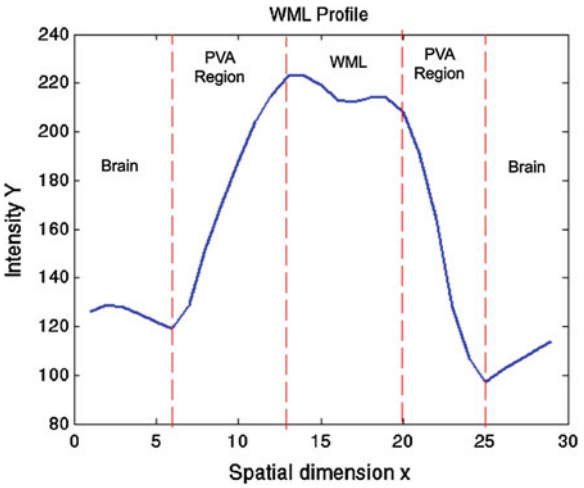


Fig. 5 1D scanline of WML

As shown in Sect. 2.1, model-based approaches use Gaussian distributions to model the intensity distributions of each tissue class. Together, these class distributions form a Gaussian Mixture Model, which describes the joint PDF of the image's intensity values. Some of the more recent model-based techniques account for PVA in the presence of Gaussian noise.

The classes which are most likely to mix are white matter (WM) and gray matter (GM) (denoted by GW) and cerebrospinal fluid (CSF) with GM (denoted by CG) [17]. Therefore, in addition to modeling the intensity distribution of the three pure tissue classes, i.e., $\Omega_{pure} = \{\text{WM, GM, CSF}\}$, a new set of classes are included to account for the PVA effect: $\Omega_{pure-mix} = \{\text{WM, GM, CSF, CG, GW}\}$.

Although several variations exist, two types of methods are most popular. In the first method, each class $\Omega_{pure-mix} = \{\text{WM, GM, CSF, CG, GW}\}$ is independently modeled by a single Gaussian distribution (as in Eq. 4), each with their own parameter set $\theta_\omega = \{\mu_\omega, \sigma_\omega^2\}$, which may be estimated by the EM algorithm.

In the second method, the pure tissue classes, $\Omega_{pure} = \{\text{WM, GM, CSF}\}$, are modeled by Gaussians, as in Eq. 4 and the mixture tissues, $\Omega_{mix} = \{\text{CG, GW}\}$, are approximated by estimating the mixing process [41]. The mixing process is modeled by mixture densities with the following PDF

$$P(y|\omega_{mix}, \alpha) = \frac{1}{\sigma_\omega(\alpha)\sqrt{2\pi}} \cdot \mathrm{e}^{-\frac{(y-\mu_\omega(\alpha))^2}{2\sigma_\omega^2(\alpha)}}, \tag{10}$$

where $y \in \Omega_{mix}$ is a mixture of two pure tissues $\omega_1, \omega_2 \in \Omega_{pure}$ and α is the fraction of ω_1 present in the mixture voxel y. In this work, the authors assume that the distribution of α is uniform [17] and therefore the likelihood density may be found by numerical computation of

$$P(y|\omega) = \int_0^1 P(y|\omega, \alpha) d\alpha. \tag{11}$$

The final intensity distribution of all tissue classes can be found by substituting Eq. 11 into the GMM computation of Eq. 5. Again, knowledge of the acquisition process is needed to accurately represent pure and PVA voxels in the presence of noise using these methods. As PVA and acquisition noise are inherently related to one another, they should be handled together for robust segmentation.

3 PVA Quantification and WML Segmentation

To segment neurological structures in MRI in the presence of PVA and noise in the past, most works focused on intensity-based tissue classification using the Expectation Maximization (EM) framework [17, 42, 43]. As previously discussed, with EM-based approaches, mixture models are constructed to represent the graylevel histogram using assumed *apriori* distributions [17, 24, 38]. Usually, Gaussian intensity models with equal variances are used to represent the intensity distributions of the pure tissue classes, and the PVA voxels are modeled by a separate Gaussian [40], as a combination of the neighbouring tissue classes' Gaussians [17] or as a uniform distribution [38]. Other extensions to this parametric method include Markov Random Fields (MRF), which impose spatial constraints [24, 37].

Although results from these techniques are promising, they are based on assumptions regarding the underlying distributions and require estimates for distributional parameters. This causes inaccurate segmentations for images acquired from multicoil MR machines, since the intensity distributions of tissues are often non-Gaussian and in many cases may not even be known [6]. Additionally, pathology, such as white matter lesions, do not follow known distributions and thus cannot be handled easily by a model-based approach. Model-selection issues of such techniques are compounded by the challenges of the EM algorithm: it is computationally complex, requiring hours of processing to find optimal parameters [17], and it is local, often requiring a "proper" initialization [17]. Some algorithms employ computationally intensive Monte carlo [38] or Metropolis methods [37] for parametric tuning, but this can take an entire day [44].

There are nonparametric classification techniques in the literature that attempt to overcome these issues by processing coregistered, multicomponent datasets (i.e. T1, T2, PD) [45–47] to segment WML robustly. The introduction of redundancy removes the dependency on distributional parameters, but increases image acquisition cost (several modalities per patient), computational complexity, memory requirements and registration error, ultimately reducing the appeal of such approaches.

Unfortunately, previous model-based methods are not directly applicable since pathology modifies the intensity distribution in a manner that is difficult to model. Moreover, neurological MRI often have non-Gaussian or unknown noise properties, causing techniques that rely on normality to be inaccurate [6]. To combat these chal-

lenges, the current work focuses on a model-free, adaptive PVA modeling approach for robust segmentation of WML in FLAIR. It is computationally efficient and operates on a single modality by exploiting a mathematical relationship between edge content and PVA for robust PVA quantification and tissue segmentation. The following subsection details this method, which was initially introduced in [48].

3.1 PVA Model

Recal that in neurological MRI, there are usually two tissue types mixing per PVA voxel [38]. The intensities of these mixels, $Y_{jk}(\mathbf{x})$, are determined by the proportion of the first tissue j, in comparison to that of the second tissue k, as in:

$$Y_{jk}(\mathbf{x}) = \alpha(\mathbf{x}) \cdot Y_j(\mathbf{x}) + (1 - \alpha(\mathbf{x})) \cdot Y_k(\mathbf{x}), \tag{12}$$

where $Y_j(\mathbf{x})$ is the intensity of the first tissue $\sim p_j(y)$ at spatial location $\mathbf{x} = (x_1, x_2) \in Z^2$, $Y_k(\mathbf{x})$ is the intensity of the second tissue $\sim p_k(y)$, and $\alpha \in [0, 1]$ is the tissue fraction which describes the proportion of tissue j present at $\mathbf{x}$ (the remainder of the voxel is a fraction of tissue k, i.e. $1 - \alpha$). Using this mathematical relationship we will quantify PVA in a new way based on the edge content of the image.

To show the motivation for an edge-based approach, an ideal signal model is used where no other artifact aside from PVA is present. Tissue intensities are simulated as constant quantities $Y_j = I_j$ and the tissue fraction α is deterministic. An ideal tissue model is useful as it highlights image features that are solely associated with PVA. In the context of WML segmentation in FLAIR MRI, there are three pure tissue classes in the ideal signal model:

1. Cerebrospinal fluid (CSF), $Y_3 = I_3$
2. Gray matter (GM) AND white matter (WM), $Y_2 = I_2$
3. White matter lesions (WML), $Y_1 = I_1$.

Applying these variables to Eq. 12 results in an idealized multiclass PVA model

$$Y_{12}(\mathbf{x}) = \alpha_{12}(\mathbf{x}) \cdot I_1 + (1 - \alpha_{12}(\mathbf{x})) \cdot I_2, \tag{13}$$

$$Y_{23}(\mathbf{x}) = \alpha_{23}(\mathbf{x}) \cdot I_2 + (1 - \alpha_{23}(\mathbf{x})) \cdot I_3, \tag{14}$$

where $Y_{12}(\mathbf{x})$ and $Y_{23}(\mathbf{x})$ are the intensities of PVA voxels in the WML-brain and brain-CSF boundaries, respectively, and the brightest tissue (WML) is denoted by I_1 where $I_1 > I_2 > I_3 \geq 0$. The parameter α_{12} describes the percentage of the voxel that is made up of WML and its quantification is required for robust WML volume computation. Finding an image-based estimate for α would be of great value, since it would not rely on distributional assumptions nor require multiparametric image sets.

The ideal tissue model is simulated and shown in Fig. 6 to highlight unique characteristics of MRI. Firstly, there are three pure classes (CSF, brain, WML) that

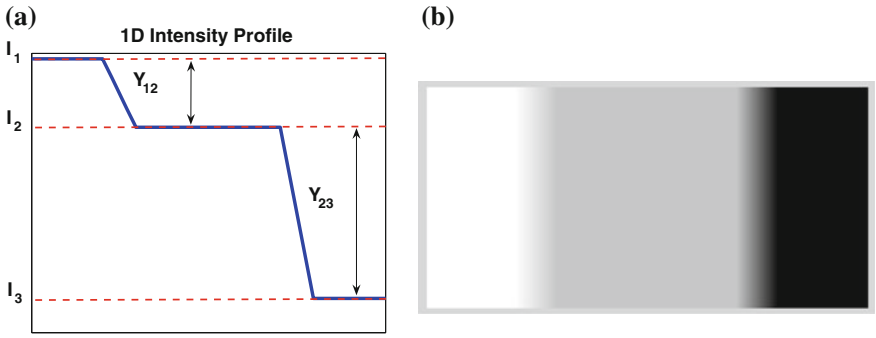


Fig. 6 Simulated mixels for ideal signal model. *Left* image representation (2D). *Right* 1D profile or scanline taken from a cross section of the 2D image. **a** Ideal $y(\mathbf{x})$. **b** $y(10, x_2)$

correspond to three unique intensity values I_1, I_2, I_3. These intensities occur in the *flat* regions of the image. Secondly, there are two mixel (PVA) classes—Y_{12} and Y_{23}, which occur over specific intensity ranges that are bounded, i.e. $I_k \leq Y_{jk}(\mathbf{x}) \leq I_j$. These graylevels occur in *edgy* regions. This ideal signal model shows that intensity and edge strength features can discriminate between pure and PVA regions and therefore, we examine how edge information can be used to model PVA (estimate $\alpha(y)$).

3.2 Edge-Based PVA Modeling

To examine the edge content in the PVA regions, the gradient of the ideal signal model is taken resulting in

$$Y'_{12} = \alpha'_{12} \cdot (I_1 - I_2)\,, \qquad Y'_{23} = \alpha'_{23} \cdot (I_2 - I_3)\,, \tag{15}$$

where α' is the *change* in the tissue fraction which dictates how the proportion of one tissue changes as a function of space. Solving for the change in the tissue fraction α' results in two PVA quantifiers

$$\alpha'_{12} = \frac{Y'_{12}}{I_1 - I_2}, \qquad \alpha'_{23} = \frac{Y'_{23}}{I_2 - I_3}. \tag{16}$$

Because edge content is quantified by the gradient, each PVA measure α'_{jk} is a class specific representation of the edge information in PVA voxels. It is a normalized representation because the largest possible value of Y'_{jk} is $I_j - I_k$ (maximum intensity change in one pixel step) and the minimum is 0 in a constant region, resulting in $0 \leq \alpha'_{jk} \leq 1$. Since these class specific variables describe PVA in terms of the

gradient, the current work focuses on an edge-based estimate for α' and uses it to decode the proportion of tissues parameter α. This is a great innovation that holds the potential to change the way PVA is handled in MR image analysis techniques. It does not depend on the intensity distribution models or multi-modalities, like traditional approaches, but is based on a mathematical relation that describes the PVA voxels in terms of the their respective edge content.

3.3 Fuzzy Edge Model

To estimate $\alpha'(\mathbf{x})$, a fuzzy technique based on the cumulative distribution function (CDF) of the gradient [39, 49], is employed. First, the traditional magnitude of the gradient, $g = \|\nabla y\|$ is estimated by

$$g = \|\nabla y\| = \sqrt{\left|\frac{\partial y}{\partial x_1}\right|^2 + \left|\frac{\partial y}{\partial x_2}\right|^2}, \tag{17}$$

where the Sobel operator is used. Based on the probability distribution function (PDF) of the gradient $p_G(g)$, the CDF of the gradient magnitude is found and used as an estimate for the edge information in the image

$$\alpha'(g) = \text{Prob}(G \leq g) = \sum_{i=0}^{g} p_G(i), \tag{18}$$

where $\alpha'(g) \in [0, 1]$. This nonlinear, fuzzyfication of the edge information quantifies the "certainty of edge presence". Note that this parameter is expressed as a function of the gradient and to be used to approximate $\alpha'(\mathbf{x})$, $\alpha'(g)$ is mapped back to the spatial domain: $\alpha'(g) \rightarrow \alpha'(\mathbf{x})$.

As shown in the works [39, 50], this fuzzy edge measure assigns large and similar values to significant edges, despite them occurring over a wide range of g and with few occurrences. It groups significant edges, while suppressing the irrelevant ones. As this edge measure is normalized, i.e. $0 \leq \alpha'(\mathbf{x}) \leq 1$, and representative of the edge information in the image, it is used to represent PVA.

Although such a nonlinear mapping function localizes PVA in ideal images and demonstrates the motivation for an edge-based approach, because of the local nature of the gradient operator, noise severely degrades its performance in noisy images. Since MRI are inherently noisy, a new estimate based on the global edge content is utilized instead to combat noise.

3.4 Global Edge Description

PVA occurs over specific intensity ranges, and moreover, with high edge values. These two features, intensity and edge strength, are coupled together in the following section to arrive at a new and denoised version of the fuzzy edge metric.

To arrive at the global PVA metric, first, edge and intensity information are coupled through the conditional PDF of $\alpha'(\mathbf{x})$, for a particular intensity y by

$$p_{\alpha'(\mathbf{x})|Y}(\alpha'(\mathbf{x}) = a|Y = y) = \frac{\#\text{ voxels with } \alpha'(\mathbf{x}) = a|Y = y}{\#\text{ voxels with } Y = y}, \tag{19}$$

where $0 \leq a \leq 1$, $0 \leq y \leq y_{max}$, a is the realization of $\alpha'(\mathbf{x})$ and y_{max} is the maximum graylevel in the image. This PDF quantifies the distribution of the edge information $\alpha'(\mathbf{x})$ for a specific graylevel y. Because it was computed on the entire image (or volume for 3D approaches), it describes the global clustering trend of edge information as a function of intensity. Generally, in flat regions - pure tissues, there is clustering in the PDF for low edge values at corresponding intensities. Across anatomical boundaries (PVA), high edge values dominate the PDF for the respective intensity ranges.

To robustly approximate $p_{\alpha'(\mathbf{x})|Y}(a|y)$, a kernel density estimator is used since it creates a smooth construction of the frequency distribution [51]. Given a sample of fuzzy edge values $\alpha'(\mathbf{x}_1), \alpha'(\mathbf{x}_2), \cdots \alpha'(\mathbf{x}_n)$, for a specific intensity y, the result is the summation of a series of kernels for all sample data

$$p_{\alpha'(\mathbf{x})|Y}(a|y) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{a - \alpha'(\mathbf{x}_i)|y}{h_n}\right), \tag{20}$$

where $K(\cdot)$ is a Gaussian kernel function and h_n is the standard deviation of the Gaussian. This technique can be applied in either 2D or 3D, where n would be the number of pixels or voxels with intensity y from the image or volume, respectively. Because the 3D representation offers an even richer description of the edge information (the distribution is generated from more samples and thus is more accurate), the final results will be obtained using a 3D approach.

To discriminate between pure tissue voxels and PVA voxels, two bin locations at $a = 0$ and $a = 1$ are used to solve Eq. 20. The result is a conditional PDF that describes edge presence for each graylevel y for the non-edge class (pure) and edge class (PVA), respectively. The probability distribution for the non-edgy regions (pure tissue class) is described by:

$$p_{\alpha'(\mathbf{x})|Y}(a=0|y) = \frac{1}{nh_n}\sum_{i=1}^{n} K'\left(0-\alpha'(\mathbf{x}_i)|y\right), \tag{21}$$
$$= \frac{1}{nh_n}\sum_{i=1}^{n} K'\left(\alpha'(\mathbf{x}_i)|y\right),$$

where the last line is valid for symmetric kernels, since $K'(x) = K'(-x)$ and $K'(x) = K(\frac{x}{h_n})$. The probability a voxel lies in a PVA region (significant edge) is found by

$$p_{\alpha'(\mathbf{x})|Y}(a=1|y) = \frac{1}{nh_n}\sum_{i=1}^{n} K'\left(1-\alpha'(\mathbf{x}_i)|y\right). \tag{22}$$

Estimation of the PDF this way automatically classifies the voxel y as belonging to either the pure tissue or PVA class. Pure regions correspond to maxima in $p\,(a=0|y)$ and minima in $p\,(a=1|y)$, whereas minima in $p\,(a=0|y)$ and maxima in $p\,(a=1|y)$ indicate with high likelihood that these voxels belong to a PVA region.

To determine the global estimate of $\alpha'(y)$, the conditional expectation operator is used. It offers the best prediction of α' given that the intensity is y in the mean square error (MSE) sense. The result is an enhanced edge map $\alpha'(y)$, which provides a global representation of the edge information in the image:

$$\alpha'(y) = \sum_{\forall a} a \cdot p_{\alpha'(\mathbf{x})|Y}(a|y), \tag{23}$$
$$= 0 \cdot p_{\alpha'(\mathbf{x})|Y}(a=0|y) + 1 \cdot p_{\alpha'(\mathbf{x})|Y}(a=1|y),$$
$$= p_{\alpha'(\mathbf{x})|Y}(a=1|y).$$

The last line of the equation above indicates that quantification of PVA content is directly proportional to the probability that a voxel is located on an edge.

3.5 Estimating α

To decode $\alpha(y)$, regions of $\alpha'(y)$ are retained, while others discarded. Recall that the maxima of $p(a=1|y)$ dictate which voxels y are most likely PVA (maximally edgy), while the minima are correlated to voxels y from pure tissue classes (minimally edgy or flat). Ideally, in pure tissues regions, there should be no edge information, but noise generates "artificial" edginess creating the minima of $\alpha'(y)$ to be non-zero in these regions. To account for the relative nature of $\alpha'(y)$, an adaptive threshold is applied.

An adaptive threshold that retains voxels most likely (in a probabilistic sense) to contain mixture components are computed for the left and right side of each PVA pulse:

$$t_{jk}^{L} = \min_k + \frac{\max_{jk} - \min_k}{2}, \tag{24}$$
$$t_{jk}^{R} = \min_j + \frac{\max_{jk} - \min_j}{2},$$

where t_L and t_R are the left and right thresholds, respectively, and min_k is the minimum of $\alpha'(y)$ corresponding to tissue k, min_j is the minimum of $\alpha'(y)$ corresponding to pure tissue j and max_{jk} is the maximum of the PVA pulse that describes the mixture of tissue j and k. The minima and maxima values are easily found with a peakfinder algorithm based on the derivative to locate optima automatically.

For two mixel classes, an array of thresholds are defined

$$T = [t_{23}^{L}, t_{23}^{R}, t_{12}^{L}, t_{12}^{R}], \tag{25}$$

where t_{23}^{L}, t_{23}^{R} are applied to the first PVA pulse (CSF-brain PVA), t_{12}^{L}, t_{12}^{R} to the second pulse (brain-WML PVA)

$$\alpha'_T(y) \rightarrow \alpha'(y) > T, \tag{26}$$
$$\alpha'_T(y) \rightarrow p\ (a = 1|y) > T. \tag{27}$$

The nonzero regions correspond to the adaptively found, class specific estimates for $\alpha'_{jk}(y)$, where $\alpha'_T(y) = \alpha'_{23}(y) \cup \alpha'_{12}(y)$.

To decode the tissue fraction $\alpha_{jk}(y)$, each refined PVA pulse $\alpha'_{jk}(y)$ is integrated over the corresponding intensity values

$$\alpha_{jk}(y) = \frac{\int_{y_k}^{y} \alpha'_{jk}(t)dt}{\int_{y_k}^{y_j} \alpha'_{jk}(t)dt}, \quad y_k < y \leq y_j, \tag{28}$$

where the denominator is a normalizing constant. The final estimate for the proportion of tissues parameter is the union of these two PVA regions: $\alpha(y) = \alpha_{23}(y) \cup \alpha_{12}(y)$. This technique automatically detects which graylevels correspond to PVA, and also how much of each tissue is present within these voxels.

3.6 WML Segmentation

As the PVA map $\alpha(y)$ dictates how tissue classes are mixing, it can be easily modified to get class membership function for the WML. Since a value of $\alpha_{12}(y) > 0.5$ indicates that tissue 1 (WML) is dominating at this graylevel y, this point is used to define where the class membership becomes more in favour of WML (tissue 1).

If y_2 and y_1 denote the starting and ending graylevel values for $\alpha_{12}(y) > 0$, respectively, the class membership $\xi_{WML}(y)$ for the WML class may be found by

$$\xi_{WML}(y) = \begin{cases} \alpha_{jk}(y), & y_2 \leq y < y_1, \\ 1 & y \geq y_1. \end{cases} \tag{29}$$

A value of '1' is assigned to voxels that are pure WML, a '0' for voxels that are not part of the lesion class, and $0 < \xi < 1$ for PVA voxels (mixture of brain and WML tissues). This class membership function leads to automatic segmentation of WML ($\xi_{WML}(y)$ is mapped to the spatial domain $\xi_{WML}(\mathbf{x})$, where the value represents the fraction of WML that is present at voxel ($\mathbf{x}$)). Other class membership functions for the brain and CSF tissue classes are similarly defined.

4 Shape Analysis

Shape analysis techniques are a set of image processing tools that are focused on characterizing segmented objects based on their shape. Typically, shape analysis has been used for object recognition and matching, boundary filtering as well as general shape characterization and there are many literature reviews on the subject [52]. In this chapter we are focused on characterizing the shape of lesions to explore shape signatures of different types of WML.

Currently, white matter lesions are divided into periventricular white matter lesions (PVWML) and deep white matter lesions (DWML). DWML are identified as being close to the periphery of the brain ("deep" in the WM), and PVWML as being alongside the ventricles. Although the meaning of these two terms vary by study and this dichotomization itself is still in debate, a possible dissimilarity in pathogenic mechanisms between PVWML and DWML are providing some clues for understanding pathophysiology of many diseases associated with white matter lesions [5]. This work investigates whether WML shape signatures can differentiate between DWML and PVWML.

Shape descriptors are computed using the segmented WML and therefore, the shape and boundary of these objects must accurately represent the underlying pathology. Since the proposed method is PVA-based, WML are segmented with subvoxel precision and produce WML objects with accurate boundaries that are ideal for shape analysis.

Very broadly, two families of shape analysis techniques are examined in this work: (1) external (boundary-based) and (2) global methods. External techniques focus on analyzing the boundary of the segmented object and global methods analyze the geometric shape of the object. Since the proposed WML segmentation technique segments WML boundaries robustly, both methods are employed. This section will discuss the shape analysis methods that will be used to describe the difference between DWML and PVWML.

4.1 Boundary-Based Techniques

In boundary-based methods, the contour of the object is usually represented by its spatial coordinates x_k, y_k, where $k = 0, 1, \ldots, N-1$ for a boundary with N pixels. The k^{th} pixel along the contour is described with two parametric equations:

$$x(k) = x_k, \tag{30}$$

$$y(k) = y_k, \tag{31}$$

which can be described jointly using complex notation: $s(k) = x(k) + iy(k)$.

A series of features are extracted from the boundary representation known as Fourier Descriptors (FD), and they are computed by taking the Fourier transform of $s(k)$

$$S(f) = \mathcal{F}\left(s(k)\right), \tag{32}$$

$$= \frac{1}{n}\sum_{k=1}^{N} s(i)e^{-2\pi j\frac{kf}{N}}, \tag{33}$$

where the length of the sequence is determined by the number of points used to computed the FFT. Since the Fourier transform is linear

$$S(f) = \mathcal{F}\left(x(k) + iy(k)\right), \tag{34}$$

$$= S_x(f) + iS_y(f), \tag{35}$$

where $S_x(f)$ and $S_y(f)$ are the spectral coefficients of the x and y coordinates. Usually, the magnitude spectrum is used to analyze the overall shape and defines the final set of FD as

$$z(f) = ||S(f)|| = \sqrt{S_x(f).^2 + S_y(f)^2}. \tag{36}$$

Fourier descriptors $z(f)$, are a series of spectral components that describe the way a contour changes as the boundary is traversed. For example, a bumpy contour, or one that has rapid changes in the x or y coordinate generate a spectrum that is dominated by high frequency content. Conversely, a smooth boundary, or one that has little change in the x or y coordinates of the object boundary would have little high-frequency content in the spectrum.

The descriptors $z(f)$ for values of f close to zero will describe an approximative shape (global shape characteristics) and the higher frequencies will describe the details. For $f = 0$, $z(f)$ represents the center of gravity of the shape and does not describe anything about the shape of the object. The first frequency component, $z(f)$ for $f = 1$, describes the overall size of the segmented shape. If all the other components are set to zero, the shape becomes a circle. The other frequency components will make alterations on the circle described by $z(1)$.

If only the low frequency descriptors are kept, the reconstructed curve is an approximation to the outline of a shape. By increasing the number of components in the description, high frequencies define sharp curves and details of the shape.

The (in)variance properties of the Fourier descriptors are related to the properties of the Fourier transform, in terms of translation, rotation, scale, and start point. To ensure that the shape descriptors of the WML are robust, these properties are examined and invariant Fourier descriptors are defined.

Translation Translation of an object is the same as adding a single constant to all of the values of $x(k)$ and $y(k)$, which effects the DC term of the boundary descriptor $s(k)$. This zero-frequency component describes its mean position (location) in the image and therefore, is the only coefficient that is translation variant. Without this term the FD are translation invariant and thus this term is usually dropped.

Rotation Rotation in the complex plane by some angle is the same as multiplication with $e^{i\theta}$. Therefore, rotation of an object about the origin of the coordinate system, results in FD that are scaled by $e^{i\theta}$. Computing the magnitude normalizes the scaling constant $e^{i\theta}$ which removes variance to rotation.

Start Point Changing the starting point of the contour (the first point in the boundary descriptor $s(k)$) is the same as translation in the spatial domain (in this case, k) which causes a phase-shift in the Fourier domain. Therefore, the magnitude of $S(f)$, $z(f)$ are invariant to the choice of the starting point.

Scaling Resizing the object is equivalent to multiplying $x(k)$ and $y(k)$ by a constant, which is the same as multiplying the corresponding FD by the same constant (ignoring first value). Since the first coefficient ($z(1)$) describes the shape, we can normalize the FD by this coefficient to render a scale invariant representation.

Invariant Descriptors The following operations are performed to make the FD invariant to translation, scale, rotation and start-point:

- $z(0) = 0 \rightarrow$ Translation Invariance,
- $z(f) = ||S(f)|| \rightarrow$ Rotation and Start Point Invariance,
- $z(f) = \frac{z(f)}{z(1)} \rightarrow$ Scale Invariance,
- Different sized lesions $\rightarrow$ Boundaries are extended and resampled so that there is fair comparison between boundaries of different lengths.

4.2 Global Shape Metrics

Global shape metrics quantify the global shape description, or the geometric shape of the WML. These metrics are attractive because they are not sensitive to boundary definitions and they also have a more intuitive meaning. For example, how "circular" an object is can be easily understood by a user, or clinician.

Each metric is computed based on the binary segmentation mask for each WML, called $B(x_1, x_2)$, where

$$B(x_1, x_2) = \begin{cases} 1, & (x_1, x_2) \in \text{WML}, \\ 0, & (x_1, x_2) \in \text{Background}. \end{cases} \tag{37}$$

Using the binary masks, several global shape metrics are considered for WML shape analysis and are briefly described below:

Circularity The simplest formula to quantify the circularity of an object (also called compactness or shape factor) is found by

$$Circ = \frac{P^2}{A}, \tag{38}$$

where P is the perimeter and A is the area of the WML in $B(x_1, x_2)$.

Haralick's Circularity Another roundness measure is known as Haralick's measure of circularity and it is computed as

$$Circ_H = \frac{\mu_R}{\sigma_R}, \tag{39}$$

where R is the Euclidean distance computed between every boundary point and the centroid, μ is the mean of these distances, and σ is the standard deviation. This metric is sometimes preferred over the traditional circularity metric since it is less sensitive to noise and digitization artifacts.

Elongatedness In addition to describing the roundness, or degree of circularity of the WML, another set of features that describe how elongated the WML are also explored. A convenient measure is dependant on central moments μ_{pq} of the segmented WML and is calculated as the ratio of the lengths of the axes that describe the best fit ellipse. The best fit ellipse is derived by performing eigenanalysis on the covariance matrix of the image constructed using the second order central moments. The two eigenvalues of the covariance matrix are

$$\lambda_1 = \frac{\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2}, \tag{40}$$

$$\lambda_2 = \frac{\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2}, \tag{41}$$

which correspond to the major and minor axes of the image intensity, respectively. The ratio of the eigenvalues define the elongatedness metric

$$elon = \frac{\lambda_1}{\lambda_2}, \tag{42}$$

which may be simplified and reformulated as

$$elon = \frac{\sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{\mu_{20} + \mu_{02}}, \tag{43}$$

in order to ensure $elon \in (0, 1)$.

Solidity A last global shape measure used is known as solidity, which describes the proportion of WML pixels that are contained in the corresponding convex hull (CH)

$$sol = \frac{n(WML \cap CH)}{n(CH)} \tag{44}$$

where $n(WML \cap CH)$ is the number of WML pixels inside the convex hull and $n(CH)$ is the total number of pixels that make up the convex hull. This may be useful in detecting PVWML since they would comprise a low number of $n(WML \cap CH)$.

5 Results

This section describes the experimental results of the proposed WML segmentation and shape characterization approaches. Both simulated and real images are used, and the performance is quantitatively measured using several metrics.

5.1 Databases

5.1.1 Simulated FLAIR with WML

To validate performance of the proposed PVA quantification scheme robustly, a series of brain images are simulated from the discrete templates provided by McGill's BrainWeb Simulated Database [53, 54]. The noise-free, inhomogeneity-free, 1.0mm slice thickness Brainweb templates were downloaded and the ground truth masks for three most commonly present tissues (GM, WM and CSF) as well as masks for pathology are retained for image simulation. Simulated images are particularly useful to quantify PVA estimation performance, since the underlying $\alpha(x_1, x_2)$ used to generate the images is known. This information is not available in real images and thus simulated images are the only real way to accurately judge the performance of a PVA quantifier.[3] The images are re-simulated so that we have strict control over the way artifacts are simulated, giving way to a flexible approach where noise and PVA can be freely modified according to a variety of parameters.

Each one of the discrete masks from BSD are denoted as $C_j(x_1, x_2)$ and each describes the membership of the voxel located at the spatial position (x_1, x_2), to the

[3] As noted in [55], there is a discrepancy with the partial volume model used in BrainWeb. Since we are particularly interested in validating the performance of the partial volume model, we simulate our own images.

j-th pure tissue type. These maps are binary and nonoverlapping. To generate partial volume averaging, we simulate the image formation process using these discrete tissue masks.

Image formation in the presence of partial volume can be considered as a convolution of the pure tissue class memberships, with a point spread function (PSF) [56]. Considering a single-class membership class $C_j(x_1, x_2)$, the partial volumed class membership function $\xi_j^{true}(x_1, x_2)$ is found by convolving the binary membership map with a PSF $h(x_1, x_2)$ as in

$$\xi_j^{true}(x_1, x_2) = C_j(x_1, x_2) * h(x_1, x_2), \tag{45}$$

where a rectangular PSF is used since it satisfies the requirements of a "nice" PSF, i.e., $\int h(x_1, x_2)dx = 1$ and $h(x_1, x_2) \geq 0, \forall(x_1, x_2)$ [56]. Using a symmetric, $M = m \times m$ rectangular kernel, if the mask lies entirely inside a pure tissue region, all M pixels are of the same class j

$$\xi_j^{true}(x_1, x_2) = \frac{1}{M} \sum_{i=1}^{M} C_j(\mathbf{x}_i) = 1. \tag{46}$$

A value of '1' indicates a pure voxel from tissue j. If there are s number of pixels inside the object, and t outside, this is a PVA voxel, and the membership is computed by

$$xi_j^{true}(x_1, x_2) = \frac{1}{M} \sum_{i=1}^{M} C_j(\mathbf{x}_i) = \frac{s}{M} * 1 + \frac{t}{M} * 0 = \frac{s}{M}, \tag{47}$$

which models PVA since $\alpha^{true}(x_1, x_2) = \frac{s}{M}$ and $1 - \alpha^{true}(x_1, x_2) = \frac{t}{M}$. The size M of the blurring function $h(x_1, x_2)$ determines the severity of the partial voluming.

Noise is added to each pure class: $Y_j \sim N(\mu_j, \sigma_j)$, where $j = \{\text{CSF, brain, WML}\}$ for FLAIR with WML. The values of μ_j were found by manually selecting voxels from real MRI and the noise variance $\sigma = \sigma_j$ is the same for all classes but is varied to determine algorithmic performance as a function of noise. In PVA regions, noise is added according to Eq. 12, using the ground truth PVA maps $\alpha^{true}(x_1, x_2)$.

The noise level in the images is measured by considering the contrast to noise ratio (CNR). Since the proposed technique is based on the gradient, and the quality of the gradient depends on the contrast between classes as well as the noise in each class, the CNR will be more descriptive for highlighting the advantages and limitations of the proposed methods. The CNR is defined as

$$CNR = \frac{\mu_j - \mu_k}{\sigma}, \tag{48}$$

where μ_j and μ_k are the class means of the two most similar (in intensity) classes, and σ is the standard deviation of the noise.

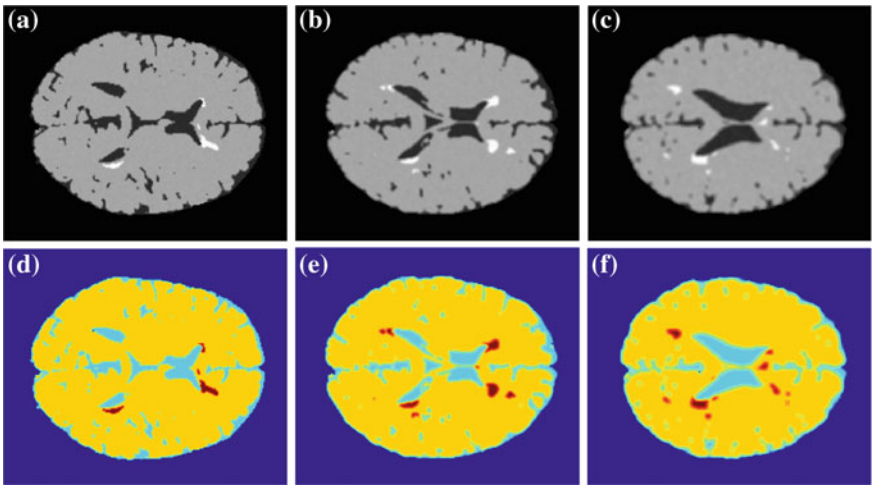


Fig. 7 Simulated FLAIR with WML (*first row*) using the corresponding class membership masks *below*. *Left* ST = 1, $\sigma = 15$. *Middle* ST = 2, $\sigma = 20$. *Right* ST = 3, $\sigma = 30$. **a** Slice 85. **b** Slice 90. **c** Slice 95. **d** $\xi^{true}(x_1, x_2)$. **e** $\xi^{true}(x_1, x_2)$. **f** $\xi^{true}(x_1, x_2)$

To test the effect of slice thickness on algorithm performance, the width of the PSF is varied. If the width/height of the PSF is denoted by m pixels, then the slice thickness (ST) is related by $m = 2 * ST + 1$ (the PSF is symmetric and thus must have an odd width/height). Several example images, along with the class membership maps used to generate the images are shown in Fig. 7.

5.1.2 Real FLAIR with WML

To verify the performance of the algorithm thoroughly, validation studies are also conducted on real images since the natural variability of patient data is not represented in simulated images. To use real FLAIR images with WML, a Research Ethics Protocol was submitted and received Institutional Ethics Approval in January 2008. A database of 24 patients were used to examine the WML segmentation performance and lesion load characteristics.

Images were acquired in the axial plane on a 1.5T GE Signa Excite, which uses a multi-phased coil array to reconstruct each image. As a result, the images are inflicted with *non-Gaussian noise*. The imaging parameters are as follows: pixel bandwidth of 97.65, 90 degree flip angle, 16bit, $0.5 \times 0.5 \times 5\,\text{mm}^3$ voxel dimensions, TR/TE/TI = 8000/128/2000 ms, FOV 180×240 mm. These images will highlight the benefits of the current approach over traditional model-based approaches since images have non-Gaussian noise characteristics and pathology which cannot be accurately accounted for with Gaussian intensity distribution models.

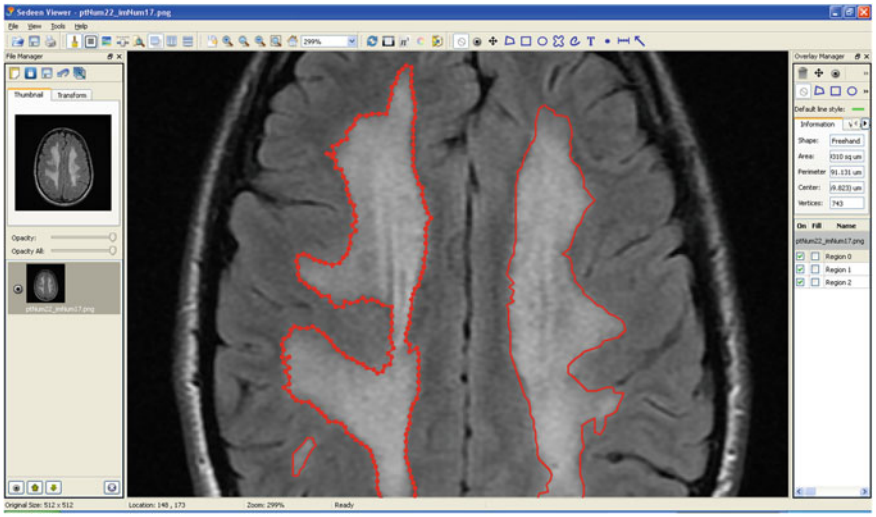


Fig. 8 Examples of manual segmentation outlines drawn in sedeen viewer

Validation of algorithms on real images requires manual segmentation of the objects that are being detected. To generate ground truth data, a Radiologist with many years experience in the interpretation of neurological MRI used a specialized software program called Sedeen[4] to outline WML objects on each image. An example of an image loaded into Sedeen is shown in Fig. 8. It is a user-friendly program that allows the user to load multiple images, zoom, pan, draw ROI (saved as an overlay), as well as manual editing of the drawn ROI. The ROIs are saved as a series of x-y coordinates.

An example image with the radiologist's outlines, the corresponding binary mask and the estimated (discrete) class membership map is shown in Fig. 9. The estimated and ground truth WML masks will be compared in the results section using various metrics.

5.2 PVA Quantification and WML Segmentation Performance

This section will detail the results of the PVA quantification and WML segmentation using real and simulated images. To objectively validate the performance of the proposed work on the experimental databases, the amount of overlap ("agreement") between a segmented object and the ground truth is measured by the Dice Similarity Coefficient (*DSC*)

$$\mathrm{DSC}(A,B) = \frac{2|A(\mathbf{x}) \cap B(\mathbf{x})|}{|A(\mathbf{x})| + |B(\mathbf{x})|}, \tag{49}$$

[4] http://www.pathcore.ca/sedeen/

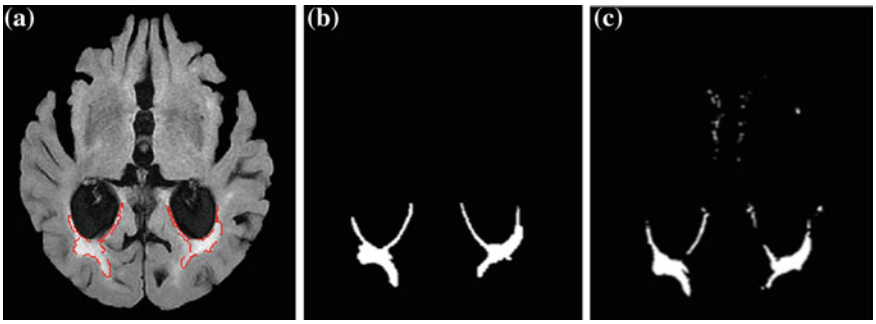


Fig. 9 Examples of manual segmentation (*outline*), corresponding binary map and automatic segmentation results. **a** Outline. **b** Manual (binary). **c** Automated (binary)

where $A(\mathbf{x})$ and $B(\mathbf{x})$ are binary masks for the segmentation and ground truth.

To further quantify the performance of the proposed method, specificity and sensitivity are also used. The sensitivity *sens* is defined as the true positive rate

$$sens = \frac{TP}{TP + FN}, \tag{50}$$

where TP and FN are the number of true positives and false negatives, respectively. Conversely, the specificity *spec* describes the true negative rate

$$spec = \frac{TN}{TN + FP}, \tag{51}$$

where TN and FP are the number of true negatives and false positives, respectively. The TP, FP, TN, FN are all defined on a voxel-by-voxel basis, using binary masks of the ground truth and segmentation result [57, 58]. These metrics describe the classification rate of our system.

Since there is usually a trade-off between these measures, sensitivity and specificity are often represented graphically by a Receiver Operating Characteristic (ROC) curve, which has 1-specificity on the x-axis and sensitivity on the y-axis. Perfect classification would yield a point in the upper left corner of the ROC curve (i.e. (0,1)), which represents 100 % sensitivity (no false negatives) and 100 % specificity (no false positives). A completely random guess would result in a value along the diagonal line from the bottom left to the top right corner. This is called the line of no discrimination and divides the ROC space; points above this line represent good classification, whereas points below indicate poor performance.

The results for one slice of the partial volumed noisy image $Y(\mathbf{x})$ is shown in Fig. 10a. The spatial domain images for the original local edge estimate ($\alpha'(g) \rightarrow \alpha'(\mathbf{x})$) and the new and improved global estimate ($\alpha'(y) = p(a = 1|y)$) are shown in Fig. 10b, c. The localized measure is extremely noisy in flat regions while noise is

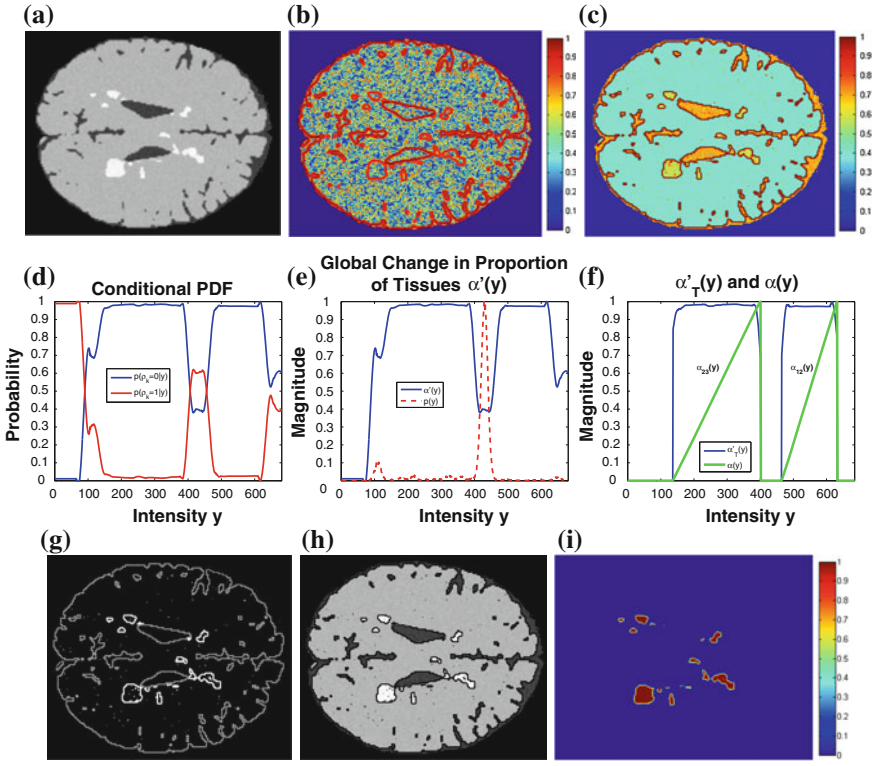


Fig. 10 Simulated images and algorithm inner workings ($\sigma = 30$). *Top row* (from *left* to *right*) simulated FLAIR MRI with WML $Y(\mathbf{x})$, localized edge metric mapped to spatial domain ($\alpha'(g) \rightarrow \alpha'(\mathbf{x})$), global edge measure mapped to spatial domain $\alpha'(y) \rightarrow \alpha'(\mathbf{x})$. *Middle row* (from *left* to *right*) conditional probability estimates $p(0|y)$, $p(1|y)$, estimated global edge function $\alpha'(y)$ and thresholded PVA map $\alpha'_T(y)$ with corresponding tissue fraction $\alpha(y)$. *Bottom row* (*left* to *right*) segmented PVA voxels, segmented pure voxels, and WML class membership mask. **a** $Y(\mathbf{x})$. **b** $\alpha'(g) \rightarrow \alpha'(\mathbf{x})$. **c** $\alpha'(y) \rightarrow \alpha'(\mathbf{x})$. **d** $p(0|y)$, $p(1|y)$. **e** $\alpha'(y)$, $p(y)$. **f** $\alpha'_T(y)$, $\alpha(y)$. **g** PVA voxels. **h** Pure voxels. **i** $\xi_{WML}(\mathbf{x})$

suppressed in the globalized version. Moreover, PVA is localized with high values and classes are clearly discernible in the new measure.

Figure 10d displays the initial estimate for $p(a|y)$ and Fig. 10e contains the estimated change in tissue fraction parameter $\alpha'(y)$ (Eq. 23), plotted on top of the graylevel PDF $p(y)$. The local minimums of $\alpha'(y)$ line up with peaks in $p(y)$ (purest voxels) and the value is nonzero in these flattest regions caused by noise. The refined estimate, α'_T is shown in Fig. 10f along with the decoded proportion of tissues $\alpha(y)$ found with Eq. 28.

Using $\alpha(y)$, the estimated intensities of PVA voxels are segmented (all y, $\alpha(y) > 0$) alongside the pure voxels (all y, $\alpha(y) = 0$) in Fig. 10g–h, showing good localization of both PVA and pure tissues. The class membership ξ_{WML} is shown in Fig. 10i,

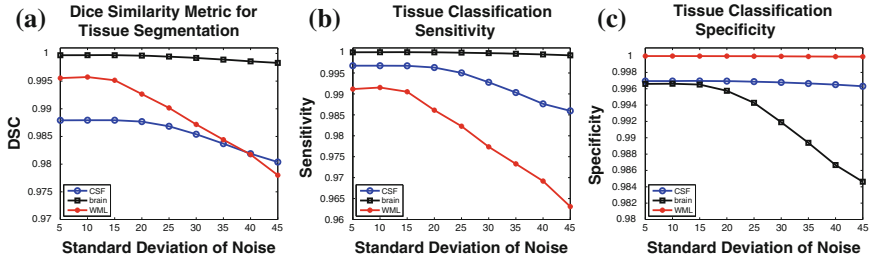


Fig. 11 Dice similarity coefficient, sensitivity and specificity of the tissue classification performance of the proposed method on synthetic data. **a** DSC. **b** Sensitivity. **c** Specificity

demonstrating that the center of the lesion (purest) receives the highest value, while the boundaries (PVA) receive proportionally less values.

The *DSC* between the true and estimated class membership masks,[5] averaged over all slices is shown in Fig. 11a. The *DSC* is very high indicating excellent detection and quantification performance. The CSF and brain tissue classes were detected with almost perfect results (average *DSC* of 0.9855 and 0.9992) and so were the WML (0.9890 on average). The sensitivity and specificity also confirm these results and are shown in Fig. 11.

For the validation studies on real FLAIR MRI with WML, a total of 195 FLAIR MR images were used to verify performance. These images were collected from 24 patients and they possess varying lesion load characteristics. To prepare the images, brain extraction [39] and bilateral filtering are employed. Manual segmentation was performed on 25 randomly chosen images. The manual segmentation result and corresponding PVA quantification and binary segmentation for the WML class are shown in Fig. 12 for several images. The interior of WML (pure) are classified with high values and the boundaries (PVA) receive a decreasing value. Note that although only the WML segmentation performance is shown, the other tissue classes are also robustly detected.

The estimated class membership $\xi_{WML}(\mathbf{x})$ were converted to binary by thresholding at various values $\tau \in (0, 1)$ and were compared to the ground truth masks via the *DSC* (segmentation "overlap") and ROC analysis (classification performance). Each point on the ROC corresponds to a sensitivity and specificity for a particular threshold value τ.

The *DSC* over all thresholds, and the average *DSC*, for all images, are shown in Fig. 13a, b, respectively. The *DSC* is high for lower thresholds as the graphs are skewed to the left. The optimal threshold τ^* (best performance over all images) is found to be 0.15, suggesting that expert WML detection includes majority of the PVA voxels in WML.

Similar results are found for the corresponding ROC curves shown in Fig. 13c, d. The ROC has a steep increase for low thresholds indicating that for high sensitivities,

[5] $\xi_{WML} \geq 0.5$.

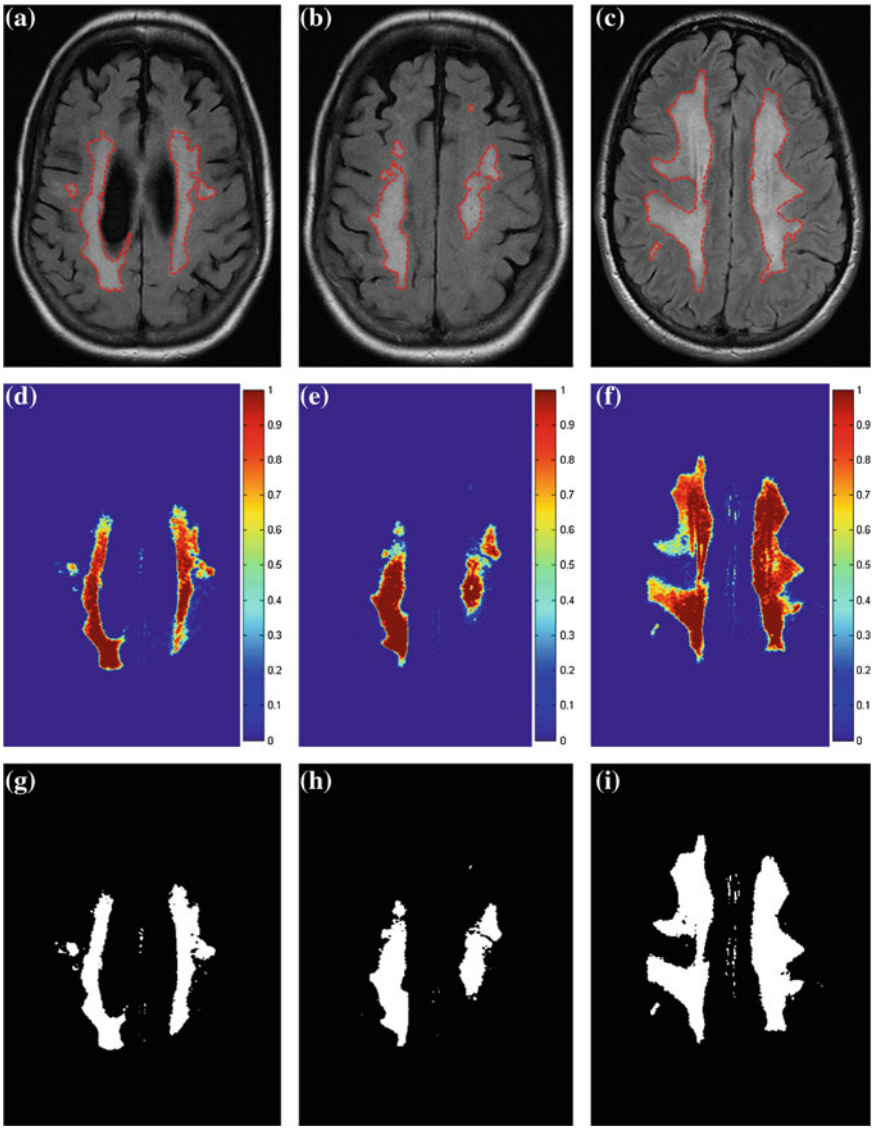


Fig. 12 Segmentation of WML in FLAIR MRI. *Top* original images with manual segmentations. *Middle* WML class membership. *Top*: *Bottom* Thresholded ($\tau^* = 0.15$) and binarized class membership of WML. **a** Im2. **b** Im4. **c** Im22. **d** ξ_{WML}. **e** ξ_{WML}. **f** ξ_{WML}. **g** $\xi_{WML} > \tau^*$. **h** $\xi_{WML} > \tau^*$. **i** $\xi_{WML} > \tau^*$

high specificities are achieved. The average *DSC*, sensitivity, and specificity at the optimal threshold value $\tau^* = 0.15$ for all images is $\overline{DSC} = 0.83$, $\overline{Sens} = 0.82$ and $\overline{Spec} = 0.99$ and the individual results are tabulated in Table 1. These results demonstrate good segmentation performance over a wide variety of patients and lesion loads.

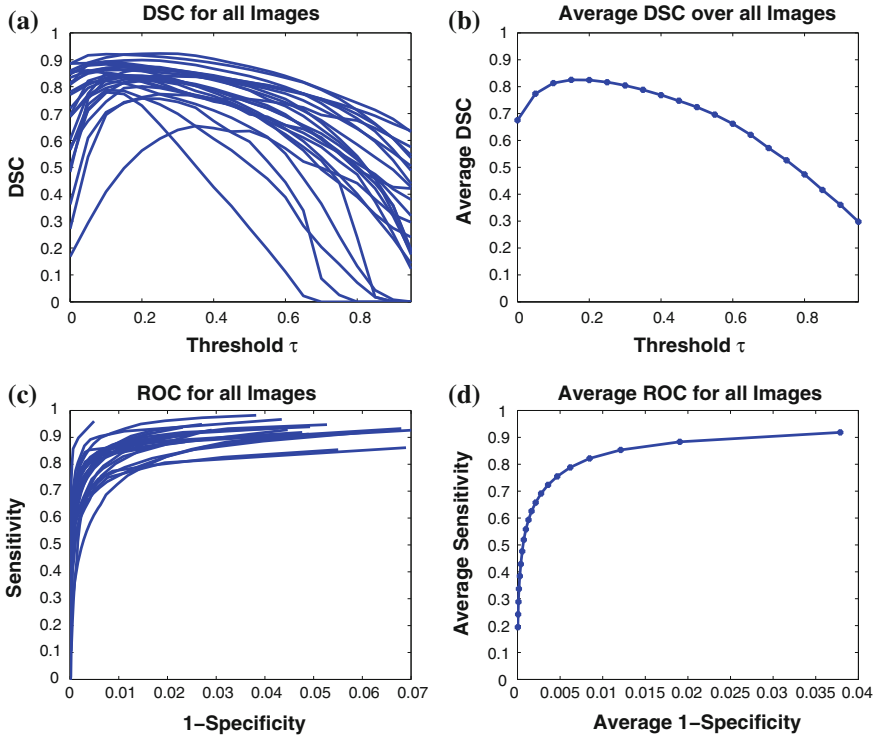


Fig. 13 **a** DSC as a function of threshold τ for all patients, **b** Average DSC curve, which shows optimal performance is achieved with $\tau = 0.15$. **c** ROC plots for all patients, **d** Average ROC graph

The simulated and real FLAIR MRI are corrupted by PVA and Gaussian as well as non-Gaussian noise sources, and regardless, algorithm performance is high in both scenarios. The algorithms use information calculated directly from a single modality so in addition to accuracy, the proposed work is also efficient. These characteristics are advantageous over traditional model- or multimodality-based approaches, since it does not depend on pretermitted intensity distribution models that inaccurately model noise in modern MRI systems and pathology, nor does it use multiple modalities to segment the lesions.

Results on simulated and real FLAIR MRI with WML show the algorithm is segmenting WML robustly. Due to the fractional nature of the class membership maps, WML segmentation is completed with subvoxel accuracy. Therefore, this algorithm presents a great opportunity for automatic shape analysis of WML (discussed next) since accurate segmentation is absolutely crucial for shape analysis.

Table 1 Results for real FLAIR MRI with WML using $\tau* = 0.15$

	IM1	IM2	IM3	IM4	IM5	IM6	IM7	IM8	IM9
DSC	0.84	0.90	0.82	0.92	0.83	0.81	0.74	0.74	0.79
Sensitivity	0.82	0.89	0.84	0.95	0.89	0.88	0.84	0.74	0.77
Specificity	0.98	0.99	0.99	0.99	0.98	0.99	1.00	1.00	0.99
Volume (mL)	8.89	8.76	2.51	6.14	4.87	2.24	0.38	0.28	2.34
	IM10	IM11	IM12	IM13	IM14	IM15	IM16	IM17	IM18
DSC	0.78	0.84	0.84	0.81	0.85	0.86	0.87	0.87	0.88
Sensitivity	0.73	0.80	0.81	0.76	0.89	0.86	0.87	0.85	0.83
Specificity	1.00	1.00	1.00	0.99	1.00	0.98	0.99	0.99	0.99
Volume (mL)	1.55	0.21	1.81	3.41	1.35	11.29	4.62	5.20	7.42
	IM19	IM20	IM21	IM22	IM23	IM24	IM25		
DSC	0.84	0.77	0.89	0.92	0.87	0.83	0.50		
Sensitivity	0.75	0.68	0.83	0.89	0.83	0.77	0.78		
Specificity	1.00	1.00	0.99	0.99	0.98	1.00	0.99		
Volume (mL)	3.89	2.71	6.84	18.78	20.33	2.02	0.37		

5.3 Shape Characterization

The previous subsections validated the PVA quantification and WML segmentation methodology on simulated and real FLAIR MRI. As was shown, WML are extracted with sub-voxel accuracy, making them great candidates for shape analysis techniques. The binary masks $B(x_1, x_2)$ and the corresponding detected boundaries (x_k, y_k) for several WML are shown in Fig. 14. These WML masks and boundaries are individually fed into the shape analysis engine.

From the real FLAIR MRI database, five images that had the WML segmentation performance validated were selected for shape analysis, resulting in a total of 48 lesions for experimentation. Segmentation postprocessing was performed to tidy up the binary masks to analyze only relevant lesions. Operations included removing false positives from the periphery of the brain (left over from brain extraction) as well as removing lesions with the largest diameter $<3\,mm$ and lesions comprised of 9 pixels or less. This removed irrelevant and spurious noise from the WML segmentation results.

The lesions were labeled as PVWML if they are touching the ventricles, and DWML if they are located away from the ventricles. These class labels, in conjunction with the shape analysis features extracted from each lesion, are used for automatic classification of lesion type.

Boundary-based and global shape metrics described in Sect. 4 were extracted from the lesions. For the lesions in Fig. 14, boundary-based features (Fourier Descriptors), as well the global shape features are shown in Figs. 15, 16.

The eight most relevant Fourier descriptors $z(2), z(3), \ldots, z(10)$, which are the most commonly used for shape analysis and description are shown for each lesion

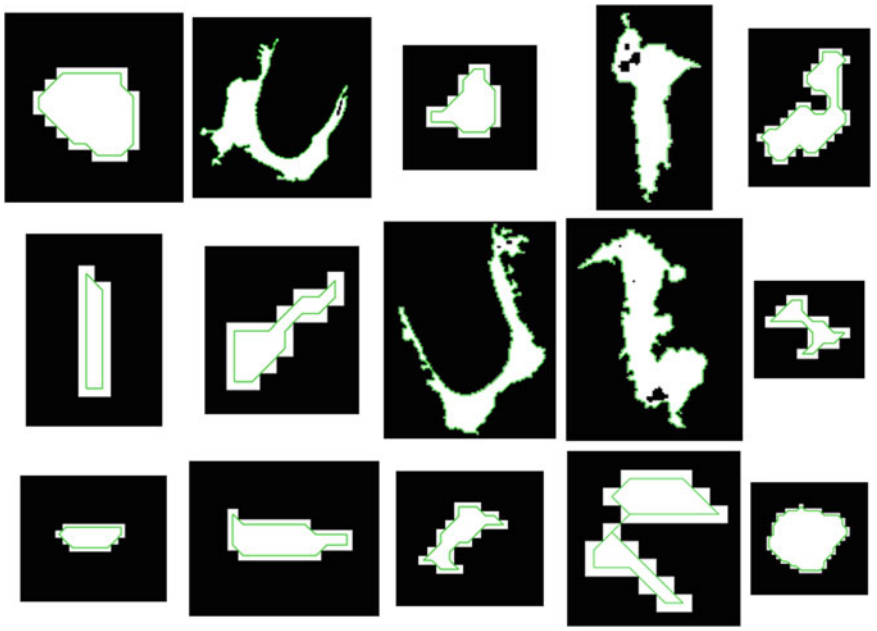


Fig. 14 WML segmentation masks and detected boundaries for a single patient. Lesions are numbered 1 through 15, starting at the *top left corner* traversing each row

(a single line corresponds to the spectrum for a single lesion). As can be seen, it is hard to differentiate between the DWML and PVWML spectra, and moreover, the meaning of these features are hard to interpret.

The global shape metrics are shown per lesion and are coded differently to highlight the discriminatory behaviour of each feature in classifying between DWML and PVWML. The circularity metric has the most separation between PVWML and DWML, whereas Haralick's Circularity measure does not differentiate between these lesion types. However, it is noted that two lesions have high Haralick's Circularity measures - these lesions correspond to the round lesions in Fig. 14 (lesion 1 and lesion 15). The solidity also discriminates between the lesion types with good results. That is due to the fact that the shape of the lesions take up less pixels inside the convex hull. The elongation metric does not appear to differentiate between the DWML and PVWML classes, but some discrimination is noted. A classification scheme will be used on these features to determine their ability to classify between lesion classes.

A supervised, linear discriminant classifier was used, where a leave-one-out training and testing strategy was employed. This allows us to determine which features robustly discriminate between PVWML and DWML. The performance of the classification system is judged by the misclassification rate (*MCR*), which is the percentage of WML labeled incorrectly. Several experiments were conducted to test the individual discrimination power of the different shape features. First, the performance of

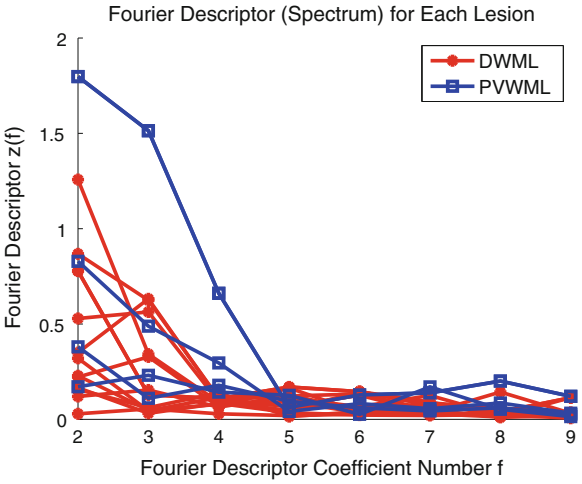


Fig. 15 Fourier descriptor (spectrum) for each lession

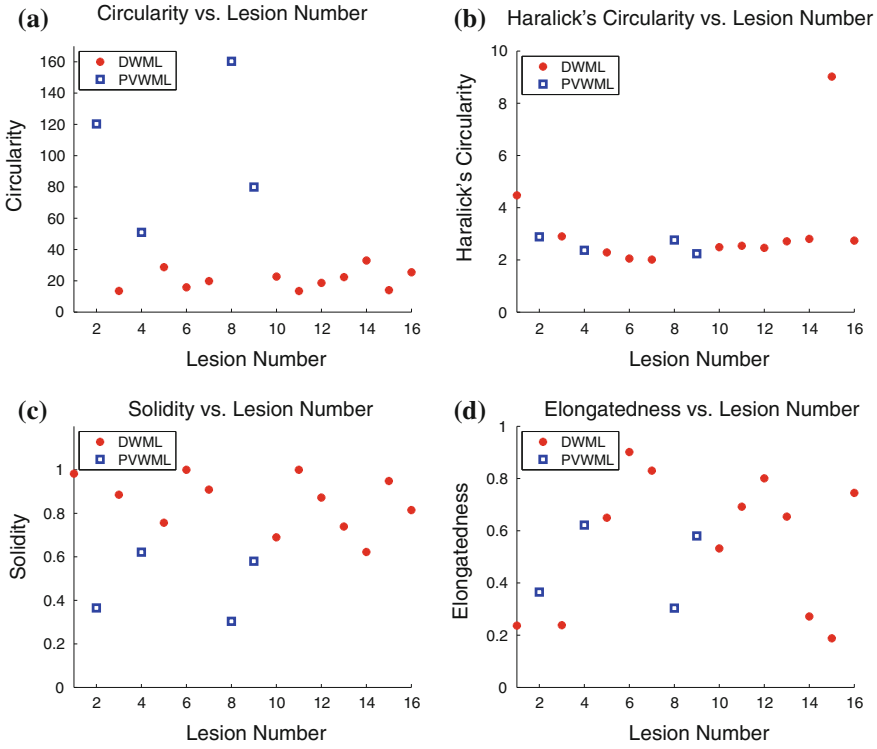


Fig. 16 Global shape metrics for each lesion. **a** Circularity. **b** Haralick's circularity. **c** Solidity. **d** Elongatedness

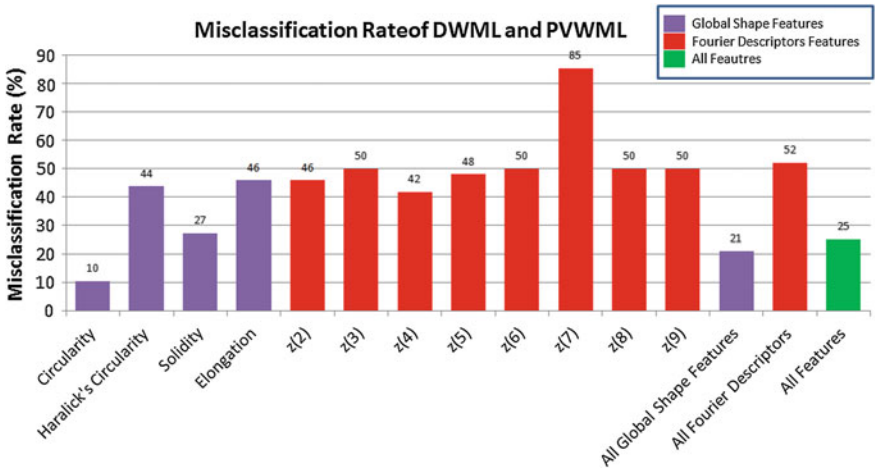


Fig. 17 WML segmentation masks and detected boundaries

each global shape feature, and each Fourier descriptor were analyzed individually. Next, the global and boundary-based features are grouped creating a global shape feature superset and boundary-based feature superset. The performance of the system to these two feature sets were tested as well. The last experiments uses all proposed shape analysis features together for classification between DWML and PVWML. These results are displayed in Fig. 17.

As can be seen, the circular and elongation metrics clearly outperform all the other features with $MCR = 10\ \%$, $MCR = 27\ \%$, respectively, including the FD as well. Using all global features and all features also results in good classification performance, with $MCR = 10\ \%$, $MCR = 27\ \%$, respectively. Interestingly, using a single feature (circularity) results in the best performance showing that very few lesions were misclassified. This shows the great utility and promise that shape analysis techniques hold for neurological disease quantification and classification. We plan to apply these methods on a larger database of lesions to examine the performance further and for other applications.

References

1. Martini FH (2001) Fundamentals of anatomy and physiology, 5th edn. Prentice Hall, New Jersey
2. Canadian Hearth and Stroke Association (2011) Statistics. World Wide Web, 2011. http://www.heartandstroke.com/site/c.ikIQLcMWJtE/b.3483991/k.34A8/Statistic.htm.
3. Fox AJ, Hachinski VC, Barnett HJM, Streifler JY, Eliasziw M, Benavente OR, Alamowitch S (2002) Prognostic importance of leukoaraiosis in patients with symptomatic internal carotid artery stenosis. Stroke 33:1651–1655

4. Malloy P, Correia S, Stebbins G, Laidlaw DH (2007) Neuroimaging of white matter in aging and dementia. Clin Neuropsychol 21:73–109
5. Kim KW, MacFall JR, Payne ME (2008) Classification of white matter lesions on magnetic resonance imaging in the elderly. Biol Psychiatr 64(4):273–280
6. Khademi A, Hosseinzadeh D, Venetsanopoulos A, Moody AR (2009) Nonparametric statistical tests for exploration of correlation and nonstationarity in images. In: International conference on digital signal processing (DSP), 2009, pp 1–6
7. Chang R (2007) Chemistry, 9th edn. McGraw-Hill, New York
8. Haase A, Frahm J, Matthaei D, Hanicke W, Merboldt KD (1986) Flash imaging: rapid NMR imaging using low flip angle pulses. J Magn Reson 67(2):258–266
9. Mansfield P (1977) Multi-planar image formation using NMR spin echoes. J Phys C 10:L55–L58
10. Salvolini U, Scarabino T (2006) High field brain MRI: use in clinical practice. Springer Berlin Heidelberg, Printed in Germany
11. Dietrich O (2011) Parallel imaging in clinical MR applications, Part I, chapter MRI from k-Space to parallel imaging. Medical Radiology, Springer Berlin Heidelberg, Printed in Germany, pp 3–17
12. Griswold MA (2011) Parallel imaging in clinical MR applications, Part I, chapter basic reconstruction algorithms for parallel imaging, Medical Radiology, Springer Berlin Heidelberg, Printed in Germany, pp 19–36
13. Murphy RE, Moody AR, Morgan PS, Martel AL (2008) Brain white matter hyperintensities are associated with carotid intraplaque hemorrhage. Radiology 248(1):202–209
14. Altaf N, Daniels L, Morgan PS, Lowe J, Gladman J, MacSweeney ST, Moody A, Auer DP (2006) Cerebral white matter hyperintense lesions are associated with unstable carotid plaques. Eur J Vasc Endovasc Surg 31:8–13
15. Jack CR, O'Brien PC, Rettman DW, Shiung MM, Xu YC, Muthupillai R, Manduca A, Avula R, Erickson BJ (2001) FLAIR histogram segmentation for measurement of leukoaraiosis volume. J Magn Reson Imaging 14(6):668–676
16. Cuadra MB, Platel B, Solanas E, Butz T, Thiran J-P (2002) Validation of tissue modelization and classification techniques in T1-weighted MR brain images, Lecture notes in computer science (LCNS). In: Medical image computing and computer-assisted intervention (MICCAI) Conference, vol 2488, pp 290–297
17. Cuadra MB, Cammoun L, Butz T, Cuisenaire O, Thiran J-P (2005) Comparison and validation of tissue modelization and statistical classification methods in T1-weighted MR brain images. IEEE Trans Med Imaging 24(12):1548–1565
18. Suri J, Singh S, Reden L (2002) Computer vision and pattern recognition techniques for 2-D and 3-D MR cerebral cortical segmentation (part I): a state-of-the-art review. Pattern Anal Appl 5(1):46–76
19. Sled JG, Zijdenbos AP, Evans AC (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17(1):87–97
20. Sijbers J, Poot D, den Dekker AJ, Pintjens W (2007) Automatic estimation of the noise variance from the histogram of a magnetic resonance image. Phys Med Biol 52:1335–1348
21. Nowak RD (1999) Wavelet-based Rician noise removal for magnetic resonance imaging. IEEE Trans Image Process 8(10):1408–1419
22. Chang L-C, Rohde GK, Pierpaoli C (2005) An automatic method for estimating noise-induced signal variance in magnitude-reconstructed magnetic resonance images. In: SPIE, medical imaging conference, vol 5747, pp 1136–1142
23. Wood JC, Johnson KM (1999) Wavelet packet denoising of magnetic resonance images: importance of Rician noise at low SNR. Magn Reson Med 41:631–635
24. Lin P, Yang Y, Zheng C-X, Gu J-W (2004) An efficient automatic framework for segmentation of MRI brain image. In: International conference on computer and IT, 2004, vol 00, pp 896–900
25. Yang J, Huang S-C (1999) Method for evaluation of different MRI segmentation approaches. IEEE Trans Nucl Sci 46(6):2259–2265

26. Clarke LP, Velthuizen RP, Phuphanich S, Schellenberg JD, Arrington JA, Silbiger M (1993) MRI: stability of three supervised semgentation algorithms. Magn Reson Imaging 11(1):95–106
27. Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20(1):45–57
28. Kellman P, McVeigh ER (2005) Image reconstruction in SNR units: a general method for SNR measurement. Magn Reson Med 54(6):1439–1447
29. Roemer PB, Edelstein WA, Hayes CE, Souza SP, Mueller OM (1990) The NMR phased array. Magn Reson Med 16(2):192–225
30. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P (1999) Sense: sensitivity encoding for fast MRI. Magn Reson Med 42:952–962
31. Thunberg P, Zetterberg P (2007) Noise distribution in SENSE-and GRAPPA-reconstructed images: a computer simulation study. Magn Reson Imaging 25(7):1089–1094
32. Sodickson DK, Manning WJ (1997) Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. Magn Reson Med 38:591–603
33. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47:1202–1210
34. Newbould R, Liu C, Bammer R (2006) Colored noise and effective resolution: data considerations for non-uniform k-space sampling reconstructions. In: Proceedings of the international society of magnetic resonance in medicine, vol 14, pp 2939
35. Lowe M, Sorenson J (1997) Spatially filtering functional magnetic resonance imaging data. Magn Reson Med 37(5):723–729
36. Samsonov A, Johnson C (2004) Noise-adaptive nonlinear diffusion filtering of MR images with spatially varying noise levels. Magn Reson Med 52:798–806
37. Van Leemput K, Maes F, Vandermeulen D, Suetens P (2003) A unifying framework for partial volume segmentation of brain MR images. IEEE Trans Med Imaging 22(1):105–119
38. Ballester MAG, Zisserman AP,Brady M (2002) Estimation of the partial volume effect in MRI. Med Image Anal 6(4):389–405
39. Khademi A, Venetsanopoulos A, Moody AR (2009) Automatic contrast enhancement of white matter lesions in FLAIR MRI. In: IEEE international symposium on biomedical imaging, 2009, pp 322–325, © 2009 IEEE. Reprinted with permission
40. Ruan S, Jaggi C, Xue J, Fadili J, Bloyet D (2000) Brain tissue classification of magnetic resonance images using partial volume modeling. IEEE Trans Med Imaging 19(12):1179–1187
41. Santago P, Gage HD (1993) Quantification of MR brain images by mixture density and partial volume modeling. IEEE Trans Med Imaging 12(3):566–574
42. Santago P, Gage H (2003) Statistical models of partial volume effect. IEEE Trans Image Process 4(11):1531–1540
43. Dugas-Phocion G, Ballester MAG,Malandain G, Lebrun C, Ayache N (2003) Improved em-based tissue segmentation and partial volume effect quantification in multi-sequence brain MRI. In: Lecture notes in computer science (LCNS): medical image computing and computer-assisted intervention (MICCAI) Conference, 2003, vol 3216, pp 26–33
44. Chiverton JP, Wells K (2003) Adaptive partial volume classification of MRI data. Phys Med Biol 53(20):5577–5594
45. Anbeek P, Vincken K, van Osch M, Bisschops R, van der Grond J (2004) Probabilistic segmentation of white matter lesions in MR imaging. NeuroImage 21(3):1037–1044
46. Lao Z, Shen D, Jawad A, Karacali B, Liu D, Melhem ER, Bryan RN, Davatzikos C (2006) Automated segmentation of white matter lesions in 3D brain MRI, using multivariate pattern classification. In: IEEE international symposium on biomedical imaging (ISBI), 2006, pp 307–310
47. de Boer R, van der Lijn F, Vrooman HA, Vernooij MW, Ikram MA, Breteler MMB, Niessen WJ (2007) Automatic segmentation of brain tissue and white matter lesions in MRI. In: IEEE international symposium on biomedical, imaging, 2007, pp 652–655

48. Khademi A, Venetsanopoulos A, Moody AR (2012) Robust white matter lesion segmentation in flair mri. IEEE Trans Biomed Eng 59(3):860–871
49. Meer P, Georgescu B (2001) Edge detection with embedded confidence. IEEE Trans Pattern Anal Mach Intell 23(12):1351–1365
50. Khademi A, Venetsanopoulos A, Moody AR (2010) Edge-based partial volume averaging estimation in FLAIR MRI with white matter lesions. In: IEEE engineering in medicine and biology conference, 2010, pp 6114–6117
51. Kvam PH, Vidakovic B (2007) Nonparametric statistics with applications to science and engineering. Whiley-Interscience, USA
52. Loncarica S (1988) A survey of shape analysis techniques. Pattern Recognit 31(8):983–1001
53. Kwan RK-S, Evans AC, Pike GB (1996) An extensible MRI simulator for post-processing evaluation. In: Lecture notes in computer science (LCNS): visualization in biomedical computing (VBC), 1996, vol 1131, pp 135–140
54. Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, Evans AC (1998) Design and construction of a realistic digital brain phantom. IEEE Trans Med Imaging 17(3):463–468
55. Bromiley PA (2008) Problems with the brainweb MRI simulator in the evaluation of medical image segmentation algorithms, and an alternative methodology. Technical report, Imaging Science and Biomedical Engineering Division, University of Manchester Medical School, Tina Memo No. 2008–002, Internal Memo
56. Pham DL, Prince JL (2000) Unsupervised partial volume estimation in single-channel image data. In: IEEE workshop on mathematical methods in biomedical image analysis (MMBIA), 2000, pp 170–177
57. Anbeek P, Vincken KL, van Osch MJ, Bisschops RH, van der Grond J (2004) Automatic segmentation of different-sized white matter lesions by voxel probability estimation. Med Image Anal 8(3):205–215
58. Popovic A, de la Fuente M, Engelhardt M, Radermacher K (2007) Statistical validation metric for accuracy assessment in medical image segmentation. Int J Comput Assist Radiol Surg 2:169–181

Groupwise Registration of Brain Images for Establishing Accurate Spatial Correspondence of Brain Structures

Zhenyu Tang and Yong Fan

Abstract For establishing accurate spatial correspondence of brain structures among different subjects, many groupwise image registration methods have been proposed to register brain images taken from different subjects onto a common space. Except the congealing method, most groupwise image registration methods achieve the image registration by registering images to a template image using pairwise image registration algorithms. For these groupwise image registration methods built upon pairwise image registration, the key points are template determination, registration path identification, and pairwise image registration. Focusing on the graph-based groupwise image registration methods due to their high computation efficiency and accuracy, this chapter introduces briefly the congealing method and groupwise image registration methods with different strategies for template determination and registration path identification. To demonstrate the strength of state-of-the-art groupwise image registration methods, a quantitative comparison study has also been presented for representative graph-based groupwise image registration methods based on two publicly available 3D MR brain image datasets.

1 Introduction

Following the rapid advancement of medical imaging techniques, medical image analysis plays an increasingly important role in the field of medical science. In neuroimaging studies, since brain structures of different subjects have variant shapes, brain images of different subjects are often transformed into a common space to

Z. Tang · Y. Fan (✉)
Brainnetome Center, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing100190, China
e-mail: yfan@nlpr.ia.ac.cn; yong.fan@ieee.org

Z. Tang
e-mail: zytang@nlpr.ia.ac.cn

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*, Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_7,

facilitate statistical image analysis. Such a common space is typically defined by a brain image atlas, e.g., the MNI brain and the Talairach atlas [1].

Recently, groupwise image registration methods for registering multiple images have attracted interests due to their efficacy [2–11]. Instead of registering each individual image to a specific atlas image, the groupwise image registration methods register images to a common space identified automatically from the images to be registered. Such an image registration strategy can effectively reduce registration bias introduced by a specific image atlas, and therefore might be able to improve the subsequent image analysis. In this chapter, different kinds of groupwise image registration methods are introduced, and their performance in the registration of 3D MR brain images is evaluated.

Many methods for the groupwise image registration have been proposed in the last decade [2–10]. Except the congealing method that registers images by optimizing a groupwise image similarity, i.e., sum of entropies of pixel stacks [10, 12], most of the existing groupwise image registration methods are built upon pairwise image registration algorithms and achieve the final registration in two different ways: (1) registering all images directly to a template image that can be a representative image selected from the images to be registered (typically a group center image) or iteratively to their evolving average image [2, 5, 6], referred to as direct pairwise image registration based groupwise image registration methods; and (2) registering images to their similar images and then composing the resulting deformation fields to achieve the registration of all images to the representative image [3, 4, 7–9, 13, 14], referred to as intermediate template based groupwise image registration methods.

This chapter is organized as follows. Pairwise image registration methods are firstly briefly introduced, followed by introductions of groupwise image registration methods, including the congealing method, the direct pairwise image registration based, and the intermediate templates based groupwise image registration methods. Finally, the chapter is concluded with a brief discussion.

2 Pairwise Image Registration

The pairwise image registration methods find a spatial transformation for registering one image, often referred to as floating image, to another image, referred to as target image. A survey of image registration methods has classified image registration methods into feature based and area based methods [15]. To register two images (target and floating images), a feature based image registration method typically follows three steps. First, each image's feature points are identified and characterized by simple image features such as line intersections [16] and corners [17] or more sophisticated image features, e.g., histogram of local image patch and SIFT features [18]. Second, matched feature points of the images to be registered are obtained by matching their image features. Finally, parameters of a transformation model, e.g., rigid transformation or projection mapping, are estimated to establish the correspondence between the matched feature points. The resulting parameters of the

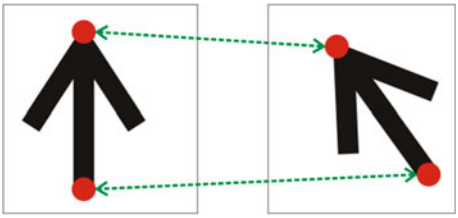


Fig. 1 Images contain an arrow shape with different orientations. By finding the matched feature points (marked as *red points*), one is able to estimate the rotation angle and translation of a rigid transformation for registering the images

transformation model can be used to spatially transform the floating image to the target image's space. Figure 1 shows two 2D images, each of them containing an arrow with different orientations. Once one gets at least two pairs of matched feature points, parameters of a rigid transformation (the rotation angle and translation) can be estimated to establish their spatial correspondence for achieving image registration.

In the area based image registration methods, instead of feature points, image regions are considered. In other words, all image pixels within a region of interest are regarded as one feature "point". Typically, parameters of certain transformation model in the area based image registration are determined based on a predefined cost function. The most commonly adopted cost functions include Pearson correlation coefficient for modeling linear relationship between intensities of images to be registered and mutual information for modeling their non-linear relationship [15].

The image registration methods can also be classified as either parametric image registration or non-parametric image registration based on the transformation model used in the registration. The advantage of parametric image registration methods is their high computational efficiency. In particular, linear parametric transformation models used in image registration include rigid transformation, affine transformation, and projection mapping, while Thin Plate Spline (TPS) and Free-Form Deformations (FFD) are typically adopted as non-linear parametric transformation models in image registration [19, 20].

Since the deformation modeled by a parametric model typically has low degrees of freedom, non-parametric image registration methods are often used for registering images with diffuse, local differences, e.g., brain images [21]. Among the non-parametric image registration methods, Thirion's Demons algorithm [22] is one well-known method. The Demons algorithm is derived from the optical flow equation [23], and the deformation vector for each pixel is formulated as

$$\boldsymbol{v}(x) = \begin{cases} \frac{(I_f(x)-I_t(x))\cdot\nabla I_t(x)}{\|\nabla I_t(x)\|^2+(I_f(x)-I_t(x))^2}, & \text{if } \| \nabla I_t(x) \|^2 + (I_f(x) - I_t(x))^2 > \varepsilon, \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $\boldsymbol{v}$ is the deformation vector at each pixel x, I_f and I_t are floating and target images, and ε is a small tolerance to improve the solution's numerical stability. The Demons algorithm achieves image registration in an iterative way. In particular, at

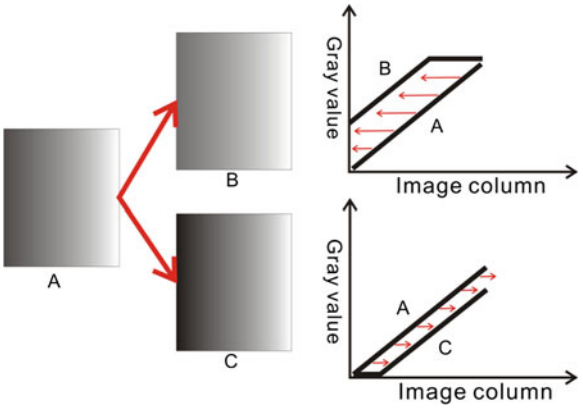


Fig. 2 Illustration of the Demons. *Left* images with a ramp structure of gray values. *Right top* intensity profile and the deformation vectors of pixels at each column for registering image *A* to image *B*. *Right bottom* intensity profile and the deformation vectors of pixels at each column for registering image *A* to image *C*

each iteration step, deformation vectors for all pixels are calculated and regularized by a Gaussian filter, and then the resulting deformation vectors are added to an accumulated deformation field. At the end, the accumulated deformation field is used to warp I_f to I_t.

Figure 2 illustrates how the deformation vector is calculated by Eq. (1) for registering one image (denoted by A) to two different images (denoted by B and C). Each image contains a different ramp structure of gray values. For registering image A to image B, since the gray value of each column of A is smaller than the corresponding column in B, $I_f(x) - I_t(x)$ is negative according to the Eq. (1). Therefore, the directions of deformation vectors for each column of A is opposite to the directions of image gradient of B as shown in the right top of Fig. 2. On the contrary, for registering image A to image C, the deformation vectors and the image gradient of C have the same direction since the gray value of each column of A is bigger than the corresponding column in C, as shown in the right bottom of Fig. 2.

The Demons algorithm has been improved for achieving better accuracy and higher efficiency in different ways and validated in different applications [22, 24–70]. Among its improved versions, the Diffeomorphic Demons algorithm has been widely adopted in the registration of brain images [65]. Since deformation vectors are projected onto the Lie groups by exponential maps, the Diffeomorphic Demons algorithm is able to obtain a diffeomorphic deformation field, i.e., no folding and invertible. This algorithm has also been used in groupwise image registration methods that are built upon pairwise image registration [8, 11, 71].

The pairwise image registration has been extended for registering multiple images using different strategies, referred to as groupwise image registration. The remainder of this chapter reviews representative groupwise image registration methods, including the congealing method, the direct pairwise image registration based and

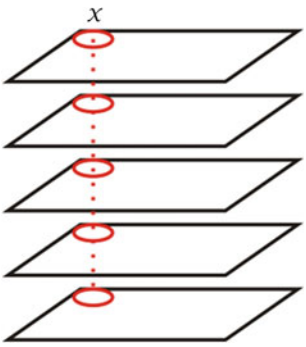


Fig. 3 Image stack used in the congealing method. The *rectangles* stand for input images and the *red circles* constitute a pixel stack at one pixel location x

the intermediate templates based groupwise image registration. In particular, graph based groupwise image registration methods are introduced as special cases of the intermediate template based groupwise image registration.

3 Groupwise Image Registration Using Congealing

The congealing method was first proposed by Miller [72] for the task of image classification, and then was adopted for groupwise image registration [73]. Parametric deformation models are typically used in the congealing based groupwise image registration, including affine transformation and B-spline deformation [12, 73]. The congealing method in conjunction with the Diffeomorphic Demons algorithm has also been adopted in fMRI data registration [74].

In the congealing method, no template is required. The similarity among images to be registered is measured by the sum of pixel wise (voxelwise) entropies across the whole image space

$$E = \sum_{x=1}^{n} H(s(x)), \tag{2}$$

where x is the spatial location of a pixel, n is the number of pixels in the image space, $s(x)$ is a variable defined by the intensities of a pixel x across all of the images to be registered, and $H(\cdot)$ is the entropy function. The variable defined by the intensities of a pixel x across all of the images is often referred to as an image stack, as illustrated by Fig. 3. The congealing method achieves groupwise image registration by iteratively deforming images one by one for minimizing the cost function defined by Eq. (2) until convergence. Its computational cost is very high since it requires one evaluation of the cost function defined by Eq. (2) for every update of the deformation parameters for each image.

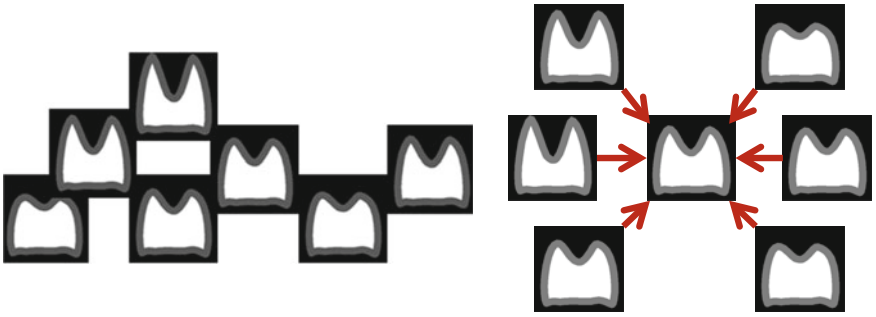


Fig. 4 Illustration of groupwise registration using direct pairwise image registration. *Left* input images; *Right* one input image is selected as a template and other images are registered to the template for achieving the final groupwise registration

To achieve efficient groupwise image registration, many groupwise image registration methods are built upon pairwise image registration. Basically, pairwise image registration is adopted to achieve groupwise image registration in two ways: (1) registering all images directly to a template image that can be a representative image selected from the images to be registered (typically a group center image) or iteratively to their evolving average image [2, 5, 6]; and (2) registering images to their similar images, known as intermediate templates, and then composing the resulting deformation fields to achieve the registration of all images to the representative image [3, 4, 7–9, 13, 14].

4 Groupwise Image Registration Using Direct Pairwise Image Registration

It is an intuitive way to achieving the groupwise image registration by directly registering all images to a template image [2, 5, 6]. The basic idea is to choose or to produce an image serving as a template from a given group of images. Then all the images in the given image group are registered to the template to achieve the final alignment (Fig. 4). The key of this kind of methods is how to determine an optimal template.

The template image can be defined as the geometric mean of the input images based on their pairwise registration results [5]. In particular, each image is first registered to all the other images, then the geometric mean is estimated using the Multi-Dimensional Scaling (MDS) based on the registration result of all image pairs, and finally the closest image to the geometric mean is chosen as the template. In [75], the template image was defined as the average image of deformed input images, and each deformed input image was produced by warping the input image with the average deformation field resulted from averaging all deformation fields determined from pairwise image registrations between this image and all the other images. However,

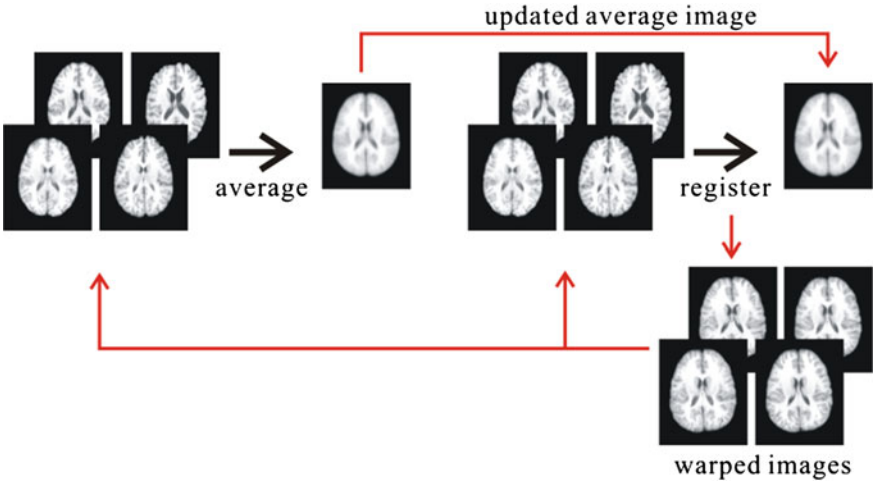


Fig. 5 Illustration of the iterative process of the groupwise registration method using the group average image as the template

the template selection methods have relatively high computational cost since the pairwise image registration is required for every possible pair of images. As mentioned in [75], it would take 4096 h, i.e., 170 days, to get the template for an image group containing 64 images using one processor.

Another efficient way to obtain the template is to generate the average image from the input images iteratively [6]. Particularly, an average image of registered input images using affine transformation is first generated as the template image and then all images are registered to the template image using a nonlinear registration method. The registered images can be averaged again to generate a new template image so that the groupwise image registration is improved iteratively. The iteration process stops when all images have been well aligned or the maximum number of iterations exceeded. The flowchart of such a groupwise image registration algorithm is shown in Fig. 5.

At the very beginning of the iterative groupwise image registration, images are typically not well aligned and the resulting average image is blurred. Such a blurred template image could hamper the groupwise image registration. To overcome this problem, a sharp average image can be generated using an adaptive weighting strategy at each iteration step [7]. The weight for each image's pixel/voxel at each iteration step is inverse proportional to the squared intensity difference between image patches centered at the pixel/voxel in the image considered and the group mean image produced in the previous iteration step, so that the pixel/voxel in an image more similar (i.e., similar image intensity) to that in the group mean image contributes more to the pixel/voxel intensity in the new group mean image. The resulting group mean image is often with sharp contrast.

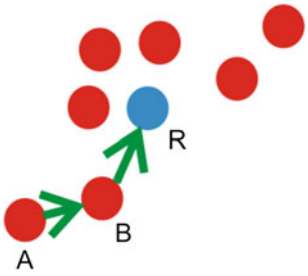


Fig. 6 Illustration of groupwise registration using intermediate templates. *Red points* indicate images. *A* is the image to be registered, *B* is the intermediate template, and *R* is the final template

5 Groupwise Image Registration Using Intermediate Templates

Since the registration of similar images is typically easier than the registration of images that are much different from each other, several groupwise image registration algorithms have been proposed to register images to their similar images, referred to as intermediate templates, and then composing the resulting deformation fields for achieving the registration of all images to a representative image of input images [3, 4, 7–9, 13, 14]. The basic idea of this kind of registration methods for registering image A to final template R via intermediate template B is illustrated by Fig. 6. To determine the optimal intermediate template for each image, different strategies have been proposed [14, 71].

A statistical deformation model has been used to generate the intermediate templates in a brain image registration method called RABBIT [14]. In particular, the method first gets a set of deformation fields that register a template brain image to individual brain images, and then applies Principal Component Analysis (PCA) to the deformation fields for generating a statistical deformation model that is able to characterize a brain image deformation field with a small number of parameters. To register the template brain image to a brain image, the statistical deformation model is used to generate a deformation field that warp the template image to produce an intermediate template close to the brain image considered, then a deformation field that registers the intermediate template and the brain image considered is estimated by a pairwise image registration algorithm, and finally the obtained deformation field is composed with that generated by the statistical deformation model to register the template brain image to the brain image considered. Since the difference between the intermediate template image and the image considered is relatively small, a better registration result can be obtained.

In most of groupwise image registration methods, the intermediate templates for each input image are its neighboring images chosen from the rest of input images. For example, in a method called ABSORB [71], each image's neighboring images serve as its intermediate templates if the neighboring images are closer than itself to a group center image. The center image is determined by choosing the image

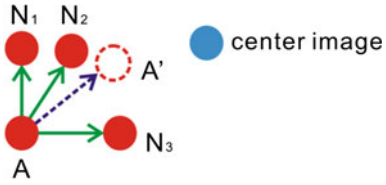


Fig. 7 Illustration of the registration strategy proposed in ABSORB. Given a group of image, represented by *red points*, to be registered, each image (e.g., A) is warped toward the group center with a deformation field obtained by averaging deformation fields warping A to its neighboring images (N_1, N_2 and N_3) that are closer to the group center than A itself. The procedure is iterated to warp all images to the group center image that is updated after each iteration step

near the group geodesic mean on an estimated manifold of input images using graph theoretic techniques (discussed in the Sect. 6). The groupwise image registration is then achieved by iteratively warping each input image with a mean deformation field obtained by averaging deformation fields that map the input image to its intermediate templates, as illustrated by Fig. 7.

Recently, graph theoretic techniques have been adopted in the groupwise image registration for learning a manifold of input images so that a large deformation between images can be decomposed into a series of small and anatomically meaningful deformations between similar images [3, 4, 8, 9]. The graph based groupwise image registration methods are intermediate template based registration methods too since some input images are used as intermediate templates of other images for achieving the final groupwise image registration. The representative groupwise image registration methods using graph theoretic techniques are introduced in Sect. 6.

6 Groupwise Image Registration Using Graphs

In the graph based methods, a graph of input images is constructed. In such a graph, each node corresponds to an input image and similar nodes are connected with edges. The weight of each edge connecting two images is determined by their similarity measure. The graph of images helps estimate the manifold of input images and the geodesic distance between two images can be approximated by the their shortest path in the graph identified using graphical algorithms, such as Dijkstra method [76]. Images on the shortest path between a pair of images can serve as intermediate templates and the registration between the pair of images can be achieved by subsequently composing deformations between all adjacent image pairs along their shortest path. Typically, a root image, which serves as the template, is first determined in the graph based methods, and then all images are registered to the root image along their corresponding shortest paths to the root. The template (or root) in the graph based methods is usually identified as a pseudo-geodesic mean of the images [8] by

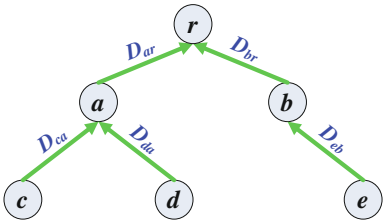


Fig. 8 Illustration of the graph based groupwise image registration. Each *circle* denotes one image, and $D_{xy}, x, y = a, b, c, d, e, r, x \neq y$, denotes a deformation field that registers image x to image y

$$I_{root} = \underset{I_i}{argmin} \sum_{j=1}^{n} g(I_i, I_j), i = 1, \cdots, n, \tag{3}$$

where $g(I_i, I_j)$ is the length of the shortest path between images I_i and I_j. Once the root image is determined and the shortest paths from non-root images to the root are fixed, a tree of images is obtained and the root of the tree is the group center of images. An example of such a tree with 6 images is illustrated in Fig. 8.

As shown in Fig. 8, there are 5 non-root images (denoted as a, b, c, d, e) and each of them has a corresponding shortest path, following which the non-root image can be registered to the root image by subsequently composing pairwise deformation fields along the shortest path, i.e.,

$$D_{xr} = D_{xk_1} \circ D_{k_1k_2} \circ \cdots \circ D_{k_nr}, x, k_1, k_2, \cdots, k_n, r \in P_{xr}, \tag{4}$$

where $P_{xr} = (x, k_1, k_2, \cdots, k_n, r)$ is the shortest path from image x to the root image r, and D_{xr} denotes the deformation field mapping x to r. For instance, in Fig. 8 the shortest path of image c to the root image r is $P_{cr} = (c, a, r)$ and the registration of image c to the root image can be achieved by subsequently composing deformation fields D_{ca} and D_{ar} , i.e., $D_{cr} = D_{ca} \circ D_{ar}$. The graph based registration methods help decompose a large deformation into small ones, and it is relatively easier to achieve accurate results for the registration of similar images than the registration of images with larger differences.

In the rest of this section, typical graph based methods are discussed in detail, and their performance in the registration of MR brain images is evaluated.

6.1 kNN Graph Based Groupwise Image Registration

Most of the graph based groupwise image registration methods adopt kNN graph to build the graph of images [4, 8]. In a kNN graph, each node is connected to its k nearest neighboring nodes that are determined based on an image similarity measurement. The typically used image similarity is the Euclidean distance between image intensities defined as

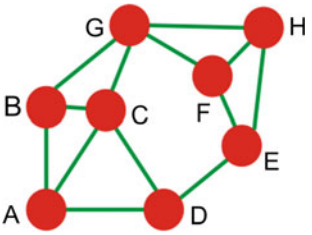


Fig. 9 An example of the graph construction using kNN algorithm. Each node is connected to its k nearest neighbors ($k = 3$). For instance, node A is connected to B, C and D

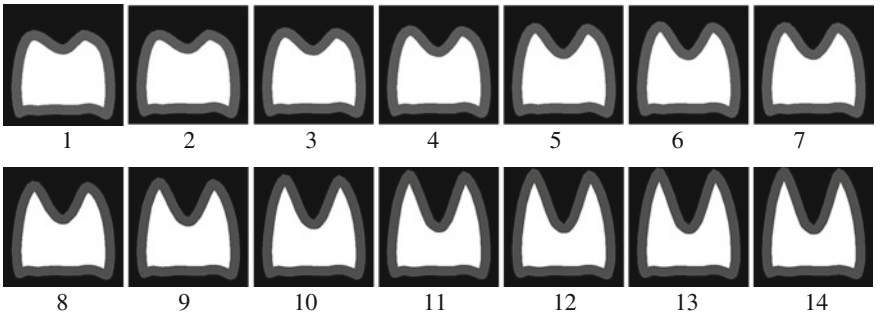


Fig. 10 A set of synthetic images that have two peaks of different heights

$$d_{Euclidean}(I_i, I_j) = \left(\sum_{x=1}^{n} (I_i(x) - I_j(x))^2 \right)^{1/2}, \tag{5}$$

where n is the total number of pixels/voxels in each image, $I_i(x)$ and $I_j(x)$ are the image intensities at location x in images I_i and I_j and respectively. An example kNN graph is shown in Fig. 9.

The kNN graph has been demonstrated to be able to estimate the manifold of high-dimensional data such as images [77]. However, the performance of kNN graph based method is sensitive to the selection of k. Therefore, the resulting shortest paths identified in a kNN graph, i.e., geodesic distances measured in the estimated manifold, are unstable. The sensitivity of kNN graph based method is illustrated by the estimation of the manifold of synthetic images shown in Fig. 10. These images lie on a one dimensional manifold. The kNN graph based estimation result varies very much with the parameter k, as indicated by the identified shortest paths of images to a group center image shown in Fig. 11. In particular, kNN graphs with different settings of k are constructed based on similarity measures defined by Eq. (5), and then Dijkstra method is used to identify the shortest path from each image to the group center determined using Eq. (3). It is clear that the shortest paths for some images are changed dramatically with different settings of k. It is worth noting that the number in each circle shown in Fig. 11 corresponds to the image with the same number in Fig. 10.

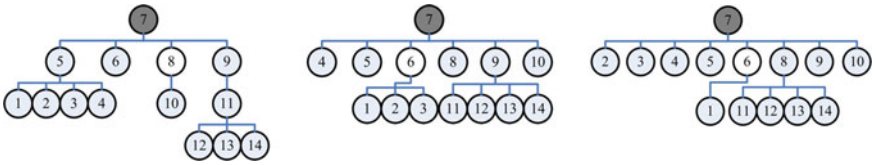


Fig. 11 The shortest paths derived from kNN graphs with different values of k. The root image is marked as a *gray circle* and the ID in each *circle* corresponds to the image with the same ID in Fig. 10

The graph of images can also be constructed using a so-called ϵ-ball method in a similar way to the kNN graph method. Different from the kNN graph method, the ϵ-ball method determines the neighboring images with connections to a given image in the graph using a distance threshold. In particular, given a group of images, each image's neighboring images are those within a ball centered at the image considered with radius ϵ. Unfortunately, the ϵ-ball method has the same problem as the kNN graph method.

6.2 kNN+MST Based Groupwise Image Registration

Built upon the kNN graph method, the Minimum Spanning Tree (MST) method has also been adopted to identify the shortest paths for images in groupwise image registration [4, 7]. The MST is a tree with minimal sum of edge weights. Two commonly used algorithms for constructing a MST are Prim's algorithm [78] and Kruskal's algorithm [79], and the latter can also be used to construct a minimum spanning forest if the graph is not connected, i.e., a MST for each connected component. The pseudo-code of Kruskal's algorithm for connected graphs is summarized in **Algorithm 9.1**.

Algorithm 9.1: *Kruskal's algorithm*

Input: nodes (i.e. images) $\{I_i, i = 1, \cdots, n\}$ and edges with weights of each pair of connected nodes $\{e_{ij} = (I_i, I_j), i, j = 1, ...,n, i \neq j\}$

1. $G = \{V, E\} = \emptyset$
2. Initialize n sub-graphs (e.g., U_1,…,U_n each of which contains a node only (no overlap)
3. Rank edges according to their weights in an increased order
4. **While** G is not a connected graph
5. Choose the edge e_{ab} which has the minimal weight
6. **If** $I_a \in U_x$ and $I_b \in U_y$ && $x \neq y$
7. $G.V = G.V \cup \{U_x, U_y\}$
8. $G.E = G.E \cup U_x.E \cup U_y.E \cup e_{ab}$
9. Merge U_x and U_y (including e_{ab})
10. **Next**

Output: G

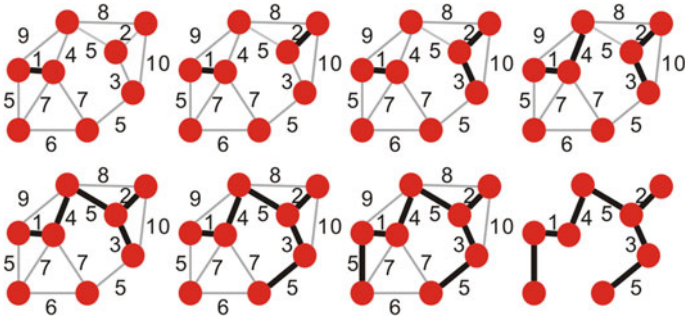


Fig. 12 Illustration of the construction of MST in a graph where each edge has a distance measure as indicated by the number next to itself. Kruskal's algorithm is applied to find the MST by adding one edge (in *black*) with the minimal distance step by step as illustrated by the figures shown from *top left* to *top right*, then *bottom left* to *bottom right*. The resulting MST is shown in the *bottom right*

Figure 12 shows an example how to find the MST in a graph where each edge is weighted by a distance measure as indicted by the number next to itself. Starting from a void set, the Kruskal's algorithm finds the MST in the graph by adding one edge with the minimal distance step by step.

It is worth noting that the MST is usually generated from a graph obtained by the kNN graph method in the groupwise image registration methods [4]. Therefore, it suffers from the same problem as the kNN graph based groupwise image registration methods.

6.3 Sparse Graph Based Groupwise Image Registration

A sparse coding based algorithm was proposed recently for constructing a sparse graph of images, referred to as ℓ^1-graph [80]. Particularly, given a group of images $I_i, i = 1, \cdots, m$, each with n voxels, the algorithm measures similarities among images using sparse coding [81]. Mathematically, the similarity measures for one image, e.g., I_i, to all the other images can be derived from the solution of the sparse coding problem

$$\min_{\alpha} \parallel I_i - D\alpha \parallel + \lambda \parallel \alpha \parallel_1, \tag{6}$$

where I_i is an n dimensional column vector of image intensities, $D = [I_1, \ldots, I_{i-1}, \ldots, I_{i+1}, \ldots, I_m] \in \mathbb{R}^{n \times m-1}$ is a matrix containing all the other images, α is a (m-1)-dimensional vector containing the coefficients corresponding to images in D. Since each element of α is proportional to the similarity measure between I_i and its corresponding image in D, its inverse or a monotonically increasing function of its inverse can be used to measure the image distance. According to the sparse coding based image distance measures, one can build a directed graph of images in which

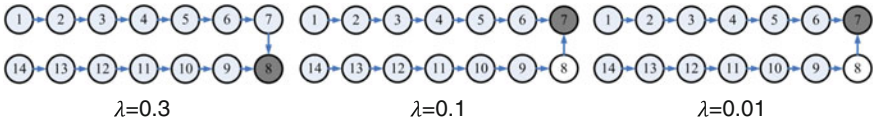


Fig. 13 The shortest paths derived from ℓ^1-graph with different values of λ. The root image is marked as a *gray circle* and the number in each *circle* correspond to the image with the same number in Fig. 10

each node's outgoing edges to other images are weighted by its distance measures to other images calculated based on the sparse coding. The advantage of the ℓ^1-graph method over the *k*NN graph is that each image's neighbors are determined adaptively to the image itself and the resulting graph is sparse due to the L1 norm constraint. Furthermore, this method is more robust to its parameter λ . As shown in Fig. 13, the shortest paths (using Dijkstra method) and the geodesic group center derive from the ℓ^1-graph with different settings of λ for the images shown in Fig. 10 are stable and the shortest paths are consistent with the underlying manifold of the images.

A groupwise image registration method built upon the sparse graph of images, referred to as dynamic sparse graph based method in this chapter, has been proposed in [11]. In particular, a sparse graph of images is constructed based on an adaptively weighted sparse coding method for estimating image similarity measures among different images. Different from the sparse coding problem formulated by Eq. (6), a weight vector is introduced in the adaptively weighted sparse coding

$$\min_{\boldsymbol{\alpha}} \|I_i - D\boldsymbol{\alpha}\| + \lambda \|diag(\boldsymbol{w})\boldsymbol{\alpha}\|_1 . \tag{7}$$

where I_i is an n-dimensional column vector of image intensities, $D = [I_1, \ldots, I_i, \ldots, I_m] \in \mathbb{R}^{n \times m}$ is a matrix containing all the input images, α is an m-dimensional vector containing the coefficients corresponding to images in D, and $\boldsymbol{w} = (w_1, \ldots, w_m) \in \mathbb{R}^m$ is a weight vector for defining the participation of images in D. To avoid a trivial similarity measurement of I_i, its corresponding weight w_i in $\boldsymbol{w}$ is set large enough to prevent I_i from representing itself.

Different from the existing graph based groupwise image registration methods [4, 8], this method then achieves the groupwise image registration in an iterative fashion. Once a graph of images is obtained, each image is registered to its direct parent image, and the graph of images is updated based on the registration results. Then, according to the updated graph of images, each image is again registered to its direct parent image in the current graph. This procedure is repeated until all images are registered to the root image. At every iteration step, images with the same directed parent image become close to each other and form a cluster after they are registered to the same parent image. If the sparse coding as formulated in the Eq. (7) with the same setting, i.e., fixed weight $\boldsymbol{w}$, is used to estimate image similarity measures for the registered images, the images in the same cluster might represent one another, resulting in larger values of similarity measures among themselves and

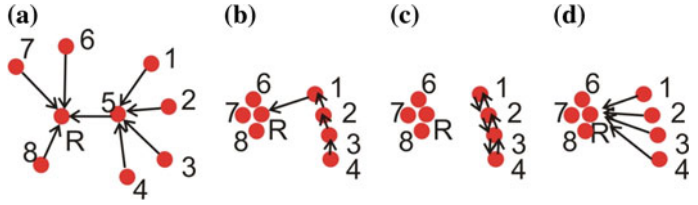


Fig. 14 Illustration of the shortest path of each image to the root identified in the updated graph of images with the image similarity measures estimated using the sparse coding with fixed and adaptive weight $\boldsymbol{w}$. *Red dots* numbered 1–8 and R denote different images, and R is the root image identified in the graph of images. **a** The shortest paths in the directed graph at one of the iteration steps. Images 1–4 have the same parent image 5, and form a cluster. Images 5–8 have the same parent image R and form another cluster. **b** The shortest path for image 4 derived from updated graph using the sparse coding with fixed weight $\boldsymbol{w}$. **c** The updated graph becomes disconnected if the image similarity measures are estimated using the sparse coding with fixed weight $\boldsymbol{w}$. **d** The shortest paths of images derived from the updated graph with image similarity measures estimated using the sparse coding with adaptive weight $\boldsymbol{w}$

small values of similarity measures (mostly 0) with the other images outside of the cluster. Therefore, some of images in the same cluster may have the shortest paths traversing others in the same cluster as illustrated by Fig. 14b, or all the images in one cluster are isolated from images in other different clusters, thus generating a disconnected graph as illustrated by Fig. 14c. The registration of one image with the shortest path traversing multiple images highly similar to it requires multiple pairwise image registration between similar images, which increases the computation cost with little improvement of the overall registration accuracy. In a graph with disconnected components, one is not able to find a root image that is reachable from all the other images. To solve the problems illustrated in Fig. 14, the weight $\boldsymbol{w}$ is adaptively updated along with the image registration. Particularly, for calculating image similarity of image , large values are set to weight elements in $\boldsymbol{w}$ corresponding to the images that have the same direct parent image as I_i in the previous iteration step. Due to the constraint of the L1 norm in Eq. (7), images having the large weights will not participate in the representing of I_i. Therefore, images in the same cluster are isolated from each other and the problems illustrated in Fig. 14b and c can be solved, as illustrated by Fig. 14d.

7 Groupwise Registration of Brain Images

In this section, applications of groupwise image registration methods to neuroimaging studies are presented. In particular, three graph based groupwise image registration methods, including kNN, kNN+MST and sparse graph, are evaluated based on two public 3D MR brain image datasets, including LPBA40 [82] and NIREP-NA0 [83]. The same pairwise image registration, namely Diffeomorphic Demons [65], is used in all of the three methods.

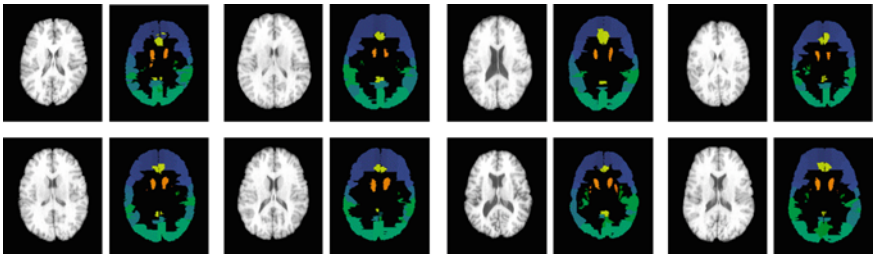


Fig. 15 2D slices of 8 randomly selected 3D MR brain images from LPBA40 and their corresponding label images

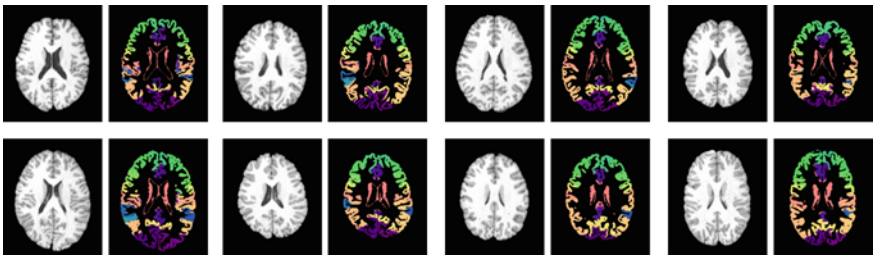


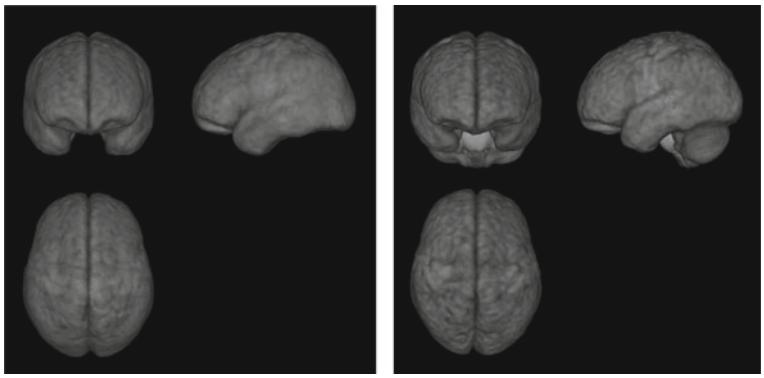
Fig. 16 2D slices of 8 randomly selected 3D MR brain images from NIREP-NA0 and their corresponding label images

The LPBA40 dataset consists of 3D MR brain images obtained from 40 normal subjects and each of them has a manually labeled image with 54 brain regions. The NIREP-NA0 dataset contains 3D MR brain images obtained from 16 normal subjects and corresponding manually labeled images with 32 regions. The overlap for the labeled regions across different images can be used to evaluate the performance of image registration algorithms. Some images and their corresponding label images from both datasets are shown in Figs. 15 and 16, respectively.

All the T1-weighted images are affine-registered to MNI152 space using FLIRT after preprocessed using histogram matching, and the transforms are then applied to their corresponding individual label images [84]. For each dataset, an average image is obtained and its 3D surface rendering is shown in Fig. 17. As the 3D surface rendering results shown, the average images are blurred, especially in gyri, indicating that the images cannot be well aligned using an affine image registration algorithm due to large inter-subject differences. The registered images using affine transformation are used as input images in the graph based groupwise image registration methods.

7.1 kNN Graph Based Groupwise Image Registration

Images in the datasets of LPBA40 and NIREP-NA0 are registered separately using the *k*NN graph based groupwise image registration method [8]. In particular, the *k*NN



Average image of LPBA40 Average image of NIREP-NA0

Fig. 17 3D surface rendering results of average images of LPBA40 (*left*) and NIREP-NA0 (*right*)

graph was adopted to estimate the underlying manifold of input images and Dijkstra method was used to find the shortest path between each pair of images. After the determination of the root image using Eq. (3), deformation fields mapping images to their direct parent images along the corresponding shortest paths were calculated. The deformation field mapping each image to the root image was produced by composing deformation fields along its shortest path as formulated in Eq. (4). It is worth noting that the image distance in the construction of *k*NN is the Euclidean distance defined in Eq. (5).

Figure 18 left shows the tree of images formed by the shortest paths derived from the *k*NN graph of LPBA40 dataset with k set to 3. These 2D Points in Fig. 18 are projection results of 3D MR brain images in LPBA40 dataset by applying Principal Component Analysis (PCA) to reduce the high dimensional original images to 2D points. The surface rendering results of average of the registered images and the root image are shown in Fig. 18 right. Compared with the surface rendering result of the average image shown in Fig. 17, the one shown in Fig. 18 has sharper contrast, indicating better registration results have been achieved by the *k*NN graph based groupwise image registration algorithm. Registration results of NIREP-NA0 dataset are shown in Fig. 19 with the same layout as Fig. 18.

7.2 kNN+MST Based Groupwise Image Registration

Images in the datasets of LPBA40 and NIREP-NA0 are registered separately using the *k*NN+MST based groupwise image registration method [4]. Figure 20 left shows the resulting *k*NN+MST of LPBA40 dataset and surface rendering results of the average image of registered images and the root image. The results of NIREP-NA0 dataset are shown in Fig. 21 with the same layout as Fig. 20. Visually, these results are similar to those obtained by the*k*NN graph based groupwise image registration method.

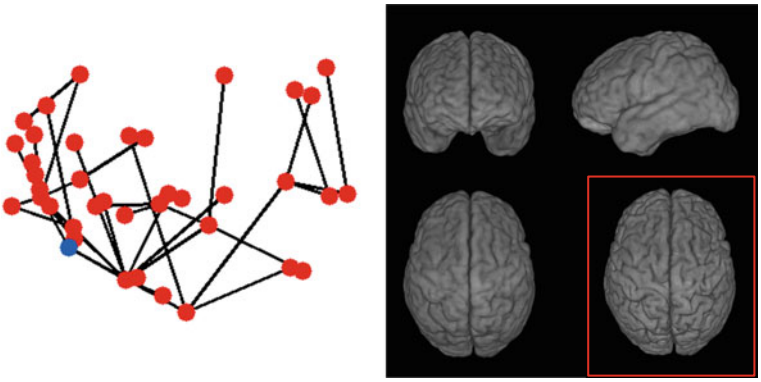


Fig. 18 *Left* Tree of images formed by the shortest paths from each image (*red point*) to the root image (*blue point*) derived from the *k*NN graph of LPBA40 dataset. *Right* Surface rendering results of average image of registered images and the root image (within the *red rectangle* on the *right bottom*)

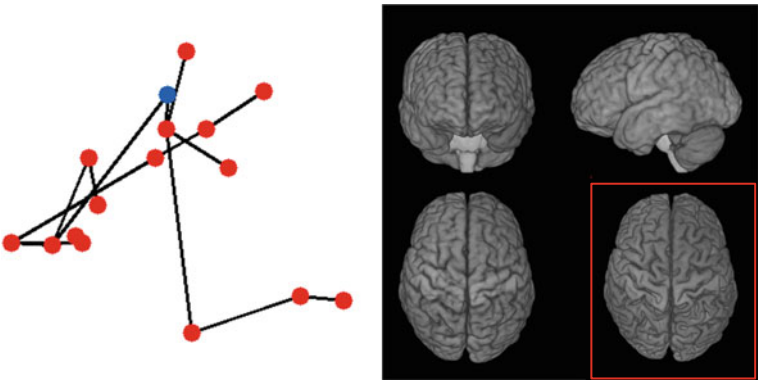


Fig. 19 *Left* Tree of images formed by the shortest paths from each image (*red point*) to the root image (*blue point*) derived from the kNN graph of NIREP-NA0 dataset. *Right* Surface rendering results of average image of registered images and the root image (within the *red rectangle* on the *right bottom*)

7.3 Sparse Graph Based Groupwise Image Registration

Images in the datasets of LPBA40 and NIREP-NA0 are registered separately using the dynamic sparse graph based groupwise image registration method [11]. Figure 22 shows the tree formed by the shortest paths derived from the graph at each iteration step for the registration of images in LPBA40 dataset.

Registration results of NIREP-NA0 dataset can be found in Fig. 23 with the same layout as Fig. 22. These results are not visually different from those obtained by the *k*NN graph based groupwise image registration method and *k*NN+MST graph based

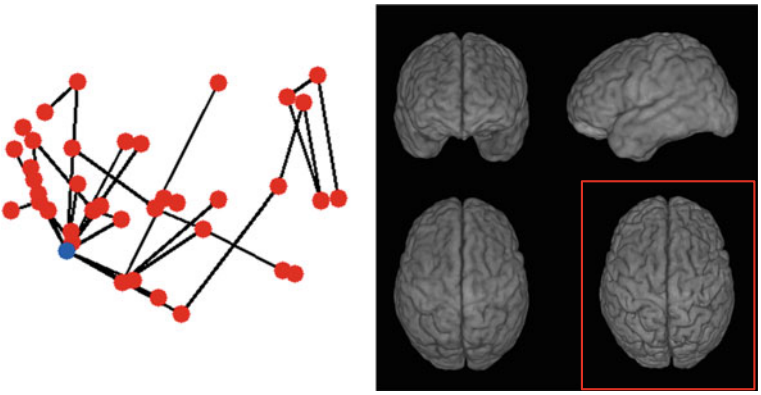


Fig. 20 *Left* *k*NN+MST of LPBA40 dataset. *Right* Surface rendering results of average image of registered images and the root image (within the *red rectangle* on the *right bottom*)

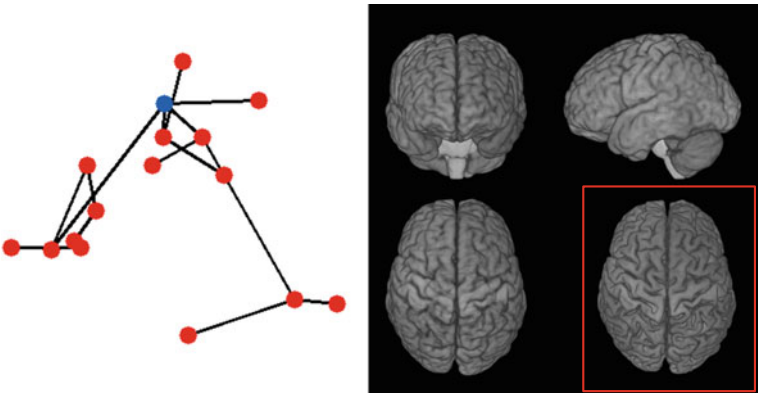


Fig. 21 *Left* *k*NN+MST of NIREP-NA0 dataset. *Right* Surface rendering results of average image of registered images and the root image (within the *red rectangle* on the *right bottom*)

groupwise image registration method. However, it seems that the group average images obtained by the sparse graph based groupwise image registration method are closer to their corresponding root images for both LPBA40 and NIREP-NA0 datasets. Such the observation is supported by the quantitative evaluation results, presented in the following section.

7.4 Evaluation of Graph Based Groupwise Image Registration

The registration performance of different graph based groupwise image registration methods has been evaluated quantitatively using overlap measures among different

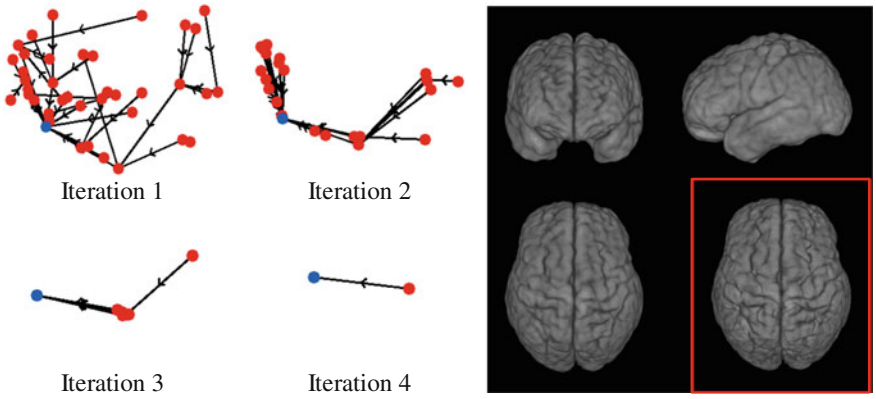


Fig. 22 *Left* Trees formed by the shortest paths derived from corresponding graphs of LPBA40 dataset at four different iteration steps. *Right* Surface rendering results of average image of registered images and the root image (within the *red rectangle* on the *right bottom*)

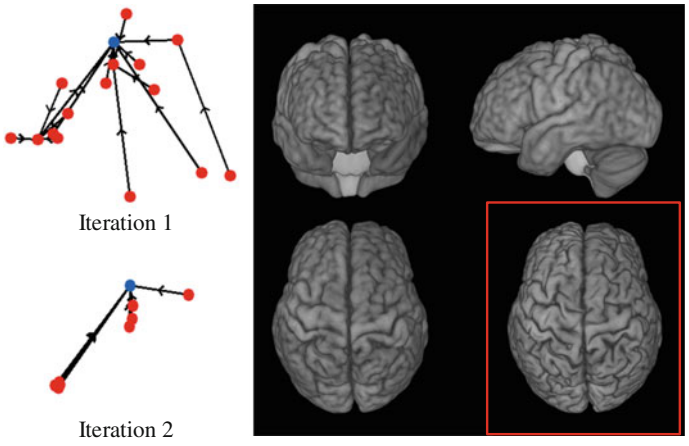


Fig. 23 *Left* Trees formed by shortest paths derived from corresponding graphs of NIREP-NA0 dataset at two different iteration steps. *Right* Surface rendering results of average image of registered images and the root image (within the *red rectangle* on the *right bottom*)

labeled regions of registered images. Particularly, the overlap between corresponding regions of registered images is evaluated using Dice index [85]. Given two regions X and Y, their overlap can be measured by $DI = \frac{2|X \cap Y|}{|X|+|Y|}$. To eavluate the registation performance of a given method, we obtain the overlap measure for each labeled region between every pair of images. So, given registration results obtained by one registration method, we obtain Dice index values for $\binom{40}{2} = 780$ image pairs and $\binom{16}{2} = 120$ image pairs for the LPBA40 and NIREP-NA0 datasets, respectively.

Besides the overlap measures for different regions, we also calculated the weighted average Dice index value across of all the labeled regions for each pair of registered images for the LPBA40 and NIREP-NA0 datasets, which is formulated as

$$DI_w = \sum_{i=1}^{n} \frac{Vol_i}{Vol_w} DI_i, \tag{8}$$

where Vol_w is the volume of all labeled regions, Vol_i is the volume of the i-th region, n is the total number of labeled regions, and DI_i is Dice index value of the i-th region.

Figure 24 shows the evaluation results based on the LPBA40 dataset for *k*NN, *k*NN+MST, and dynamic sparse graph based groupwise image registration methods. The evaluation results include Dice index values for 54 brain regions (ID: 1–54) and the weighted average Dice index value of all the brain regions (ID: 55).

Figure 25 shows the evaluation results based on the NIREP-NA0 dataset for methods including *k*NN, *k*NN+MST, and dynamic sparse graph based groupwise image registration methods, including Dice index values for all 32 brain regions (ID: 1–32) and the weighted average Dice index value of all the brain regions (ID: 33).

In summary, the average value of the weighted average Dice index values of all the brain regions of LPBA40 dataset are 0.692 (*k*NN), 0.691 (*k*NN+MST), and 0.709 (dynamic sparse graph). The paired t-tests revealed that the dynamic sparse graph performed better than *k*NN and *k*NN+MST with p values $< 1.0 \times 10^{-20}$. For the NIREP-NA0 dataset the values are 0.628 (*k*NN), 0.622 (*k*NN+MST), and 0.639 (dynamic sparse graph). Again, the paired t-tests revealed that the dynamic sparse graph performed better than *k*NN and *k*NN+MST with p values $< 1.0 \times 10^{-20}$. It is evident that the registration accuracy of the dynamic sparse graph based method is higher than the *k*NN and *k*NN+MST based methods. Since Diffeomorphic Demons [65] is used in all of the three methods, the applied image similarity measures and graph construction strategies contribute to the performance difference of these methods.

8 Discussion and Conclusion

In this chapter, several groupwise image registration methods have been discussed. These groupwise image registration methods are typically used to generate a common template or to model the shape variation of objects of interest. Except the congealing method, most groupwise image registration methods achieve the image registration by registering images to a template image using pairwise image registration algorithms. The key points in the groupwise image registrations methods are template determination, registration path identification, and pairwise image registration. For the template determination, the ultimate objective is to choose an unbiased image as the template to which all images can be registered with almost equal effort. Several strategies for the template determination have been reviewed in this chapter, includ-

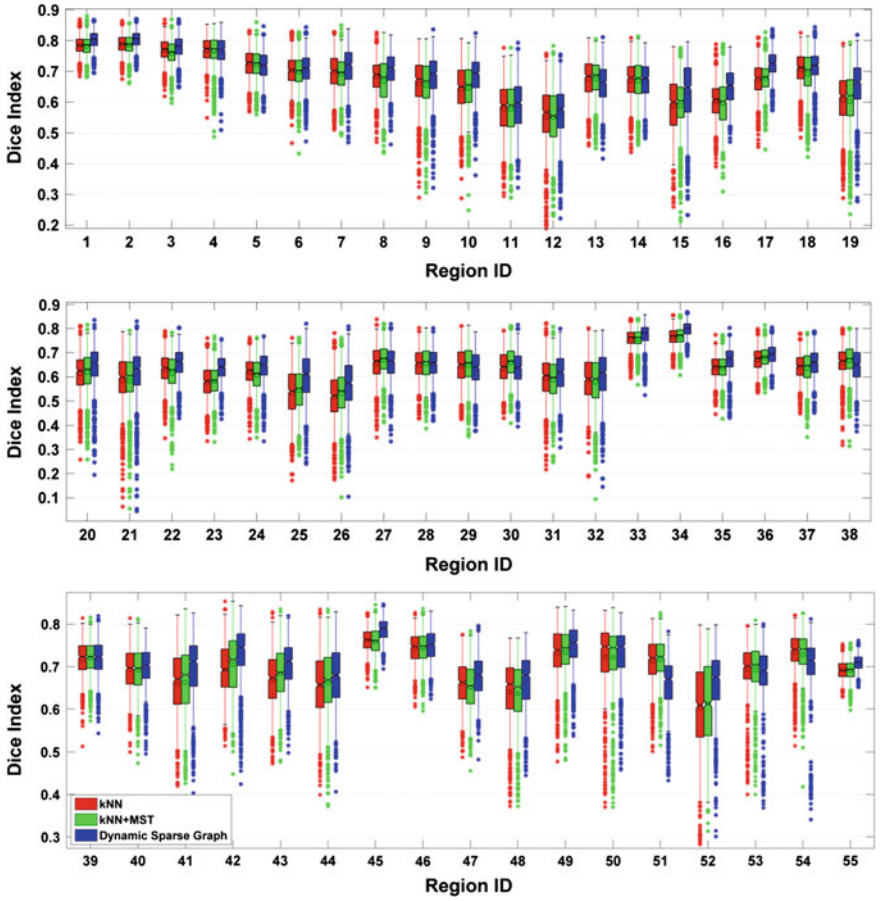


Fig. 24 Dice indexes of registered brain regions of LPBA40 dataset using *k*NN, *k*NN+MST and dynamic sparse graph based groupwise image registration methods. The region ID's corresponding brain region name is following. 1/2: L/R superior frontal gyrus; 3/4: L/R middle frontal gyrus; 5/6: L/R inferior frontal gyrus; 7/8: L/R precentral gyrus; 9/10: L/R middle orbitofrontal gyrus; 11/12: L/R lateral orbitofrontal gyrus; 13/14: L/R gyrus rectus; 15/16: L/R postcentral gyrus; 17/18: L/R superior parietal gyrus; 19/20: L/R supramarginal gyrus; 21/22: L/R angular gyrus; 23/24: L/R precuneus; 25/26: L/R superior occipital gyrus; 27/28: L/R middle occipital gyrus; 29/30: L/R inferior occipital gyrus; 31/32: L/R cuneus; 33/34: L/R superior temporal gyrus; 35/36: L/R middle temporal gyrus; 37/38: L/R inferior temporal gyrus; 39/40: L/R parahippocampal gyrus; 41/42: L/R lingual gyrus; 43/44: L/R fusiform gyrus; 45/46: L/R insular cortex; 47/48: L/R cingulate gyrus; 49/50: L/R caudate; 51/52: L/R putamen; 53/54: L/R hippocampus; and 55: weighted average Dice index of all the brain regions

ing the geometric mean, the geodesic mean, and the iterative average group image. For the registration path identification, the simplest way is to directly register image to the template. However, it is difficult to achieve accurate image registration when images to be registered have large difference. It typically leads to better performance

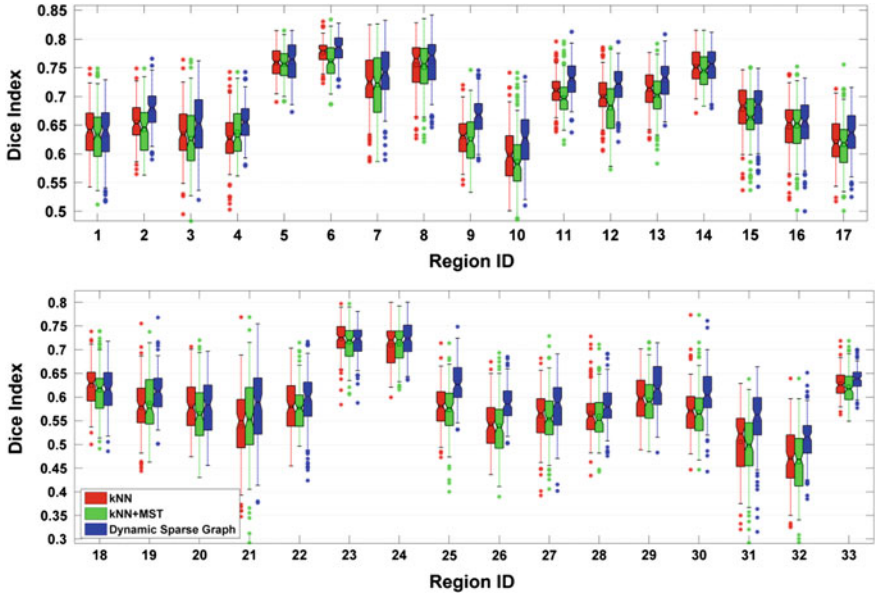


Fig. 25 Dice indexes of registered brain regions of NIREP-NA0 dataset using *k*NN, *k*NN+MST and dynamic sparse graph based groupwise image registration methods. The region ID's corresponding brain region name is following. 1/2: L/R occipital lobe; 3/4: L/R cingulate gyrus; 5/6: L/R insular gyrus; 7/8: L/R temporal pole; 9/10: L/R superior temporal gyrus; 11/12: L/R inferior temporal gyrus; 13/14: L/R parahippocampal gyrus; 15/16: L/R frontal pole; 17/18: L/R superior frontal gyrus; 19/20: L/R middle frontal gyrus; 21/22: L/R inferior gyrus; 23/24: L/R orbital frontal gyrus; 25/26: L/R precentral gyrus; 27/28: L/R superior parietal gyrus; 29/30: L/R inferior parietal lobule; 31/32: L/R postcentral gyrus; and 33: weighted average Dice index of all the brain regions

if intermediate template images are adopted to reach to the final template since large difference can be decomposed into many small ones. For the pairwise image registration, which is used to register images to the intermediate template or the final template in the groupwise registration methods, many algorithms could be adopted, not limited to Diffeomorphic Demons adopted in this study.

We have been focusing on the graph based groupwise image registration methods in this chapter, including *k*NN, MST and sparse graph (including dynamic sparse graph), due to their high computation efficiency and accuracy. The basic idea of these methods is to estimate the underlying manifold of input images so that the shortest paths that approximate the geodesic distance between each pair of images can be derived. Based on all of the geodesic distances between images, the geodesic center of image can be chosen as a template image. Final deformation field mapping each image to the template can be obtained by composing deformation fields along its corresponding shortest paths. For such methods, image similarity measurement plays an important role. Therefore, in this chapter, the pairwise (e.g., Euclidean distance) and groupwise (e.g., sparse coding and adaptively weighted sparse coding) image similarity measurements have been reviewed. Evaluation of both kinds of similarity

measurements using the LPBA40 and NIREP-NA0 datasets has demonstrated that the groupwise image similarity measurement can lead to better groupwise image registration performance than the pairwise image similarity measurements due to that the groupwise image similarity measures are estimated from a global point of view, while the pairwise image similarity measures considers only two images to be measured and no information of the other images is utilized.

Graph based manifold learning techniques have been increasingly adopted in the groupwise image registration methods to capture the distribution of images to be registered so that a group center image can be identified and the deformation between the group center and other images can be decomposed into small ones that register similar images pairwisely. For the graph based manifold learning techniques, a critical parameter is the neighborhood size for constructing a graph of images. Since the distribution of images to be registered is not necessarily uniform, especially in applications of medical image analysis with a limited number of images, the manifold of images estimated using kNN graph might be sensitive to its parameter, which inevitably renders the groupwise image registration unstable. Although the sparse coding based graph construction could be robust for estimating manifold of images with a non-uniform distribution, it merits further investigation how the image registration performance is hinged on its sparsity parameter.

Groupwise image registration algorithms have been successful in many applications of brain image analysis [4, 7, 10–12, 74]. However, all these algorithms have been focusing on imaging data of single image modality, either structural imaging data or functional imaging data. Groupwise registration of both structural and functional data might be helpful to better align the brain structure and function units across subjects [74, 86]. How to effectively integrate both structural and functional information of multimodality images for groupwise image registration merits investigation.

Acknowledgments This study was partially supported by the National Basic Research Program of China (973 Program) 2011CB707801, the National High Technology Research and Development Program of China (863 Program) 2012AA011603, the National Science Foundation of China (Grant No. 30970770, 91132707, 81271514, and 81261120419), and the Hundred Talents Program of the Chinese Academy of Sciences.

References

1. Chau W, McIntosh AR (2005) The Talairach coordinate of a point in the MNI space: how to interpret it. Neuroimage 25:408–416
2. Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113
3. Jia H, Wu G, Wang Q, Wang Y, Kim M, Shen D (2012) Directed graph based image registration. Comput Med Imaging Graph 36:139–151
4. Jia H, Yap PT, Wu G, Wang Q, Shen D (2011) Intermediate templates guided groupwise registration of diffusion tensor images. Neuroimage 54:928–939
5. Park H, Bland PH, Hero AO 3rd, Meyer CR (2005) Least biased target selection in probabilistic atlas construction. MICCAI 8:419–426

6. Joshi S, Davis B, Jomier M, Gerig G (2004) Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage 23:S151–S160
7. Wu GR, Jia HJ, Wang Q, Shen DG (2011) SharpMean: groupwise registration guided by sharp mean image and tree-based registration. Neuroimage 56:1968–1981
8. Hamm J, Davatzikos C, Verma R (2009) Efficient large deformation registration via geodesics on a learned manifold of images. MICCAI 12:680–687
9. Ye D, Hamm J, Kwon D, Davatzikos C, Pohl K (2012) Regional manifold learning for deformable registration of brain MR images. In: MICCAI 2012, vol. 7512, Springer, Berlin Heidelberg, pp 131–138
10. Wang Q, Chen L, Yap PT, Wu G, Shen D (2010) Groupwise registration based on hierarchical image clustering and atlas synthesis. Hum Brain Mapp 31:1128–1140
11. Tang Z, Jiang D, Fan Y (2013) Image registration based on dynamic directed graphs with group-wise image similarity. In: 2013 international symposium on biomedical imaging: from nano to macro, San Francisco, CA, USA
12. Balci SK, Golland P, Shenton M, Wells WM (2007) Free-form B-spline deformation model for groupwise registration. MICCAI 10:23–30
13. Donoghue CR, Rao A, Pizarro L, Bull AMJ, Rueckert D (2012) Fast and accurate global geodesic registrations using knee MRI from the Osteoarthritis Initiative. In: 2012 IEEE computer society conference on computer vision and pattern recognition workshops, Providence, RI, USA, pp 50–57
14. Tang SY, Fan Y, Wu GR, Kim M, Shen DG (2009) RABBIT: rapid alignment of brains by building intermediate templates. Neuroimage 47:1277–1287
15. Zitova B, Flusser J (2003) Image registration methods: a survey. Image Vis Comput 21:977–1000
16. Stockman G, Kopstein S, Benett S (1982) Matching images to models for registration and object detection via clustering. IEEE Trans Pattern Anal Mach Intell 4:229–241
17. Bhattacharya D, Sinha S (1997) Invariance of stereo images via the theory of complex moments. Pattern Recogn 30:1373–1386
18. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60:91–110
19. Zhuang XH, Arridge S, Hawkes DJ, Ourselin S (2011) A nonrigid registration framework using spatially encoded mutual information and free-form deformations. IEEE Trans Med Imaging 30:1819–1828
20. Rueckert D, Sonoda LI, Hayes C, Hill DLG, Leach MO, Hawkes DJ (1999) Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imaging 18:712–721
21. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46:786–802
22. Thirion JP (1998) Image matching as a diffusion process: an analogy with Maxwell's demons. Med Image Anal 2:243–260
23. Barron JL, Fleet DJ, Beauchemin SS (1994) Performance of optical-flow techniques. Int J Comput Vision 12:43–77
24. Bloy L, Verma R (2010) Demons registration of high angular resolution diffusion images. In: 2010 7th IEEE international symposium on biomedical imaging: from nano to macro, pp 1013–1016
25. Cahill ND (2012) Motion coherent image registration and demons: practical handling of deformation boundaries. Medical imaging 2012: image processing, vol 8314
26. Cahill ND, Noble JA, Hawkes DJ (2009) Demons algorithms for fluid and curvature registration. In: 2009 IEEE international symposium on biomedical imaging: from nano to macro, vols 1 and 2, pp 730–733
27. Cahill ND, Noble JA, Hawkes DJ (2009) A demons algorithm for image registration with locally adaptive regularization. In: Proceedings of medical image computing and computer-assisted intervention—MICCAI 2009, Pt I, vol 5761, pp 574–581

28. Cifor A, Risser L, Chung D, Anderson EM, Schnabel JA (2012) Hybrid feature-based log-demons registration for tumour tracking in 2-D liver ultrasound images. In: 2012 9th IEEE international symposium on biomedical imaging (Isbi), pp 724–727
29. Ebrahimi M, Martel AL (2009) Image registration under varying illumination: hyper-demons algorithm. In: Proceedings of energy minimization methods in computer vision and pattern recognition, vol 5681, pp 303–316
30. Forsberg D, Rathi Y, Bouix S, Wassermann D, Knutsson H, Westin CF (2011) Improving registration using multi-channel diffeomorphic demons combined with certainty maps. Multimodal Brain Image Anal 7012:19–26
31. Freiman M, Voss SD, Warfield SK (2011) Demons registration with local affine adaptive regularization: application to registration of abdominal structures. In: 2011 8th IEEE international symposium on biomedical imaging: from nano to macro, pp 1219–1222
32. Gu X, Jia X, Dong B, Gautier Q, Jiang S (2011) A contour-guided demons deformable image registration algorithm for adaptive radiotherapy. Int J Radiat Oncol Biol Phys 81:S803–S804
33. Gu XJ, Pan H, Liang Y, Castillo R, Yang DS, Choi DJ, Castillo E, Majumdar A, Guerrero T, Jiang SB (2010) Implementation and evaluation of various demons deformable image registration algorithms on a GPU. Phys Med Biol 55:207–219
34. Guo YJ, Cheng WH, Lu CC (2007) Non-rigid mammogram registration using demons algorithm: preliminary results. In: Proceedings of the ninth iasted international conference on signal and image processing, pp 437–442
35. Hub M, Kessler ML, Karger CP (2010) B-spline registration versus demons algorithm—a quantitative comparison of accuracy and invertibility based on artificially created test cases for the lung. World congress on medical physics and biomedical engineering, vol 25, Pt 4: image processing, biosignal processing, modelling and simulation. Biomechanics 25:790–792
36. Jin S, Li DW, Wang HJ, Yin Y (2013) Registration of PET and CT images based on multiresolution gradient of mutual information demons algorithm for positioning esophageal cancer patients. J Appl Clin Med Phys 14:50–61
37. Li DW, Yin Y (2012) Deformable registration using multi-resolution demons algorithm for 4DCT. Med Phys 39:3672–3673
38. Li W, Ibanez L, Andreasen NC, Magnotta VA (2011) The effectiveness of geometry features on multi-resolution diffeomorphic demons registration in the implementation of human cortex surface parcellation. In: 2011 8th IEEE international symposium on biomedical imaging: from nano to macro, pp 586–589
39. Lin XB, Qiu TS, Nicolier F, Ruan S (2008) An improved method of "demons" non-rigid image registration algorithm. In: Proceedings of Icsp, (2008) 9th international conference on signal processing, vols 1–5, pp 1091–1094
40. Lu C, Mandal M (2010) Improved demons technique with orthogonal gradient information for medical image registration. Ieice Trans Inf Syst E93d:3414–3417
41. Lu C, Mandal M (2010) Improved image registration technique based on demons and symmetric orthogonal gradient information. In: 2010 international conference on signal processing and communications (Spcom)
42. Lu H, Reyes M, Serifovic A, Weber S, Sakurai Y, Yamagata H, Cattin PC (2010) multi-modal diffeomorphic demons registration based on point-wise mutual information. In: 2010 7th IEEE international symposium on biomedical imaging: from nano to macro, pp 372–375
43. Lu HX, Cattin PC, Reyes M (2010) A hybrid multimodal non-rigid registration of MR images based on diffeomorphic demons. In: 2010 annual international conference of the IEEE engineering in medicine and biology society (Embc), pp 5951–5954
44. Mansi T, Pennec X, Sermesant M, Delingette H, Ayache N (2011) iLogDemons: a demons-based registration algorithm for tracking incompressible elastic biological tissues. Int J Comput Vis 92:92–111
45. Muyan-Ozcelik P, Owens JD, Xia JY, Samant SS (2008) Fast deformable registration on the GPU: a CUDA implementation of demons. Proceedings of the International Conference on Computational Sciences and Its Applications, pp 223–233

46. Nithiananthan S, Brock KK, Daly MJ, Chan H, Irish JC, Siewerdsen JH (2009) Demons deformable registration for CBCT-guided procedures in the head and neck: convergence and accuracy. Med Phys 36:4755–4764
47. Nithiananthan S, Brock KK, Daly MJ, Chan H, Irish JC, Siewerdsen JH (2010) Demons deformable registration for cone-beam CT guidance: registration of pre- and intra-operative images. Medical imaging 2010: visualization, image-guided procedures, and modeling, vol 7625
48. Nithiananthan S, Mirota D, Uneri A, Schafer S, Otake Y, Stayman JW, Siewerdsen JH (2011) Incorporating tissue excision in deformable image registration: a modified demons algorithm for cone-beam CT-guided surgery. Medical imaging 2011: visualization, image-guided procedures, and modeling, vol 7964
49. Nithiananthan S, Schafer S, Mirota DJ, Stayman JW, Zbijewski W, Reh DD, Gallia GL, Siewerdsen JH (2012) Extra-dimensional demons: a method for incorporating missing tissue in deformable image registration. Med Phys 39:5718–5731
50. Nithiananthan S, Schafer S, Uneri A, Mirota DJ, Stayman JW, Zbijewski W, Brock KK, Daly MJ, Chan H, Irish JC, Siewerdsen JH (2011) Demons deformable registration of CT and cone-beam CT using an iterative intensity matching approach. Med Phys 38:1785–1798
51. Peyrat JM, Delingette H, Sermesant M, Pennec X, Xu CY, Ayache N (2008) Registration of 4D time-series of cardiac images with multichannel diffeomorphic demons. In: Proceedings of medical image computing and computer-assisted intervention—MICCAI 2008, Pt Ii, vol 5242, pp 972–979
52. Peyrat JM, Delingette H, Sermesant M, Xu CY, Ayache N (2010) Registration of 4D cardiac CT sequences under trajectory constraints with multichannel diffeomorphic demons. IEEE Trans Med Imaging 29:1351–1368
53. Pheiffer TS, Ou JJ, Miga MI (2010) Automatic generation of boundary conditions using Demons non-rigid image registration for use in 3D modality-independent elastography. Medical imaging 2010: visualization, image-guided procedures, and modeling, vol 7625
54. Pheiffer TS, Ou JJ, Ong RE, Miga MI (2011) Automatic generation of boundary conditions using demons nonrigid image registration for use in 3-D modality-independent elastography. IEEE Trans Biomed Eng 58:2607–2616
55. Seiler, C., Pennec, X., Reyes, M.: Geometry-Aware Multiscale Image Registration via OBBTree-Based Polyaffine Log-Demons. Medical Image Computing and Computer-Assisted Intervention (Miccai 2011), Pt Ii 6892, 631–638 (2011).
56. Seiler, C., Pennec, X., Ritacco, L., Reyes, M.: Femur Specific Polyaffine Model to Regularize the Log-Domain Demons Registration. Medical Imaging 2011: Image Processing 7962, (2011).
57. Sharp GC, Kandasamy N, Singh H, Folkert M (2007) GPU-based streaming architectures for fast cone-beam CT image reconstruction and demons deformable registration. Phys Med Biol 52:5771–5783
58. Shen JK, Matuszewski BJ, Shark LK, Skalski A, Zielinski T, Moore CJ (2008) Deformable image registration—a critical evaluation: demons, B-spline FFD and spring mass system. In: Proceedings medivis, (2008) fifth international conference biomedical visualization—information visualization in medical and biomedical informatics, pp 77–82
59. Siless V, Guevara P, Pennec X, Fillard P (2011) Joint T1 and brain fiber diffeomorphic registration using the demons. Multimodal Brain Image Anal 7012:10–18
60. Sosa-Cabrera D, Tristan-Vega A, Vegas-Sanchez-Ferrero G, Gonzalez-Fernandez J, Gomez-Deniz L, Alberla-Lopez C, Ruiz-Alzola J (2008) A new approach to elastography using a modified demons registration algorithm—art. no. 69200X. Medical imaging 2008: ultrasonic imaging and signal processing, vol 6920, pp X9200–X9200
61. Suh JW, Kwon OK, Scheinost D, Sinusas AJ, Cline GW, Papademetris X (2011) Whole body nonrigid Ct-Pet registration using weighted demons. In: 2011 8th IEEE international symposium on biomedical imaging: from nano to macro, pp 1223–1226
62. Tristan-Vega A, Vegas-Sanchez-Ferrero G, Aja-Fernandez S (2008) Local similarity measures for demons-like registration algorithms. In: 2008 IEEE international symposium on biomedical imaging: from nano to macro, vols 1–4, pp 1087–1090

63. Vercauteren T, Pennec X, Malis E, Perchant A, Ayache N (2007) Insight into efficient image registration techniques and the demons algorithm. In: Proceedings information processing in medical imaging, vol 4584, pp 495–506
64. Vercauteren T, Pennec X, Perchant A, Ayache N (2008) Symmetric log-domain diffeomorphic registration: a demons-based approach. In: Proceedings medical image computing and computer-assisted intervention—MICCAI 2008, Pt I, vol 5241, pp 754–761
65. Vercauteren T, Pennec X, Perchant A, Ayache N (2009) Diffeomorphic demons: efficient non-parametric image registration. Neuroimage 45:61–72
66. Vereauteren T, Pennec X, Perchant A, Ayache N (2007) Non-parametric diffeomorphic image registration with the demons algorithm. In: Proceedings medical image computing and computer-assisted intervention—MICCAI 2007, Pt 2, vol 4792, pp 319–326
67. Wang H, Dong L, O'Daniel J, Mohan R, Garden AS, Ang KK, Kuban DA, Bonnen M, Chang JY, Cheung R (2005) Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. Phys Med Biol 50:2887–2905
68. Yang D, Li HD (2009) A probabilistic demons algorithm for texture-rich image registration. In: 2009 16th IEEE international conference on image processing, vols 1–6, pp 161–164
69. Yeo BTT, Sabuncu M, Vercauteren T, Ayache N, Fischl B, Golland P (2008) Spherical demons: fast surface registration. In: Proceedings of medical image computing and computer-assisted intervention—MICCAI 2008, Pt I, vol 5241, pp 745–753
70. Yeo BTT, Sabuncu MR, Vercauteren T, Ayache N, Fischl B, Golland P (2010) Spherical demons: fast diffeomorphic landmark-free surface registration. IEEE Trans Med Imaging 29:650–668
71. Jia H, Wu G, Wang Q, Shen D (2010) ABSORB: atlas building by self-organized registration and bundling. Neuroimage 51:1057–1070
72. Miller EG, Matsakis NE, Viola PA (2000) Learning from one example through shared densities on transforms. In: Proceedings of IEEE conference on computer vision and pattern recognition, vol 1, pp 464–471
73. Zollei L, Learned-Miller E, Grimson E, Wells W (2005) Efficient population registration of 3D data. Lect Notes Comput Sc 3765:291–301
74. Jiang D, Du Y, Cheng H, Jiang T, Fan Y (2013) Groupwise spatial normalization of fMRI data based on multi-range functional connectivity patterns. Neuroimage 82C:355–372
75. Seghers D, D'Agostino E, Maes F, Vandermeulen D, Suetens P (2004) Construction of a brain template from MR images using state-of-the-art registration and segmentation techniques. In: Proceedings od medical image computing and computer-assisted intervention—MICCAI 2004, Pt 1, vol 3216, pp 696–703
76. Dijkstra EW (1959) A note on two problems in connexion with graphs. Numerische Mathematik 1:269–271
77. Tenenbaum JB, de Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323
78. Prim RC (1957) Shortest connection networks and some generalizations. Bell Syst Tech J 36:1389–1401
79. Kruskal JB (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. Proc Am Math Soc 7:3
80. Cheng B, Yang JC, Yan SC, Fu Y, Huang TS (2010) Learning with l(1)-graph for image analysis. IEEE Trans Image Process 19:858–866
81. Donoho DL (2006) For most large underdetermined systems of equations, the minimal l(1)-norm near-solution approximates the sparsest near-solution. Commun Pure Appl Math 59:907–934
82. Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW (2008) Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 39:1064–1080
83. Christensen GE, Geng X, Kuhl JG, Bruss J, Grabowski TJ, Pirwani IA, Vannier MW, Allen JS, Damasio H (2006) Introduction to the non-rigid image registration evaluation project (NIREP). In: Proceedings of biomedical image registration, vol 4057, pp 128–135

84. Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5:143–156
85. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302
86. Conroy BR, Singer BD, Guntupalli JS, Ramadge PJ, Haxby JV (2013) Inter-subject alignment of human cortical anatomy using functional connectivity. Neuroimage 81:400–411

Modeling of Lung Nodules from LDCT of the Human Chest: Algorithms and Evaluation for CAD Systems

Amal A. Farag, Mostafa Farag, James Graham, Salwa Elshazly, Mohamed al Mogy and Aly Farag

Abstract This chapter provides a complete model-based approach for analysis of lung nodules visibly observed in clinical low dose CT (LDCT) scans of the human chest. The purpose is to highlight elements of computer-assisted diagnosis (CAD) software that can be validated using multiple radiologists using modern computing and information technology. The front-end components of the proposed approach are the following: lung nodule modeling, nodule detection, nodule segmentation, and CAD system design and evaluation. The implicit steps involved in developing these components, include filtering of the LDCT scans to reduce noise artifacts and other uncertainties associated with the imaging protocol; segmentation of the lung tissue from the rest of organs appearing in the LDCT of the chest; and creating an ensemble of nodules by human experts. As nodules take various shapes, sizes and pathologies, we limit our treatment to small size nodule ≤ 1 cm in diameter. Our ultimate goal is to create a robust system for early detection and classification, as well as tracking, of small-size nodules before they turn into cancerous. The entire development in the chapter is model-based and data-driven, allowing design, calibration and testing for the CAD system, based on archived data as well as data accrued from new patients. We provide standard development using two clinical datasets that are already available from the ELCAP and LIDC studies.

A. A. Farag
Imaging Biomarkers and Computer-Aided Diagnosis Laboratory,
NIH, Bethesda, MD, USA

M. Farag, J. Graham, S. Elshazly and A. Farag (✉)
Department of Electrical and Computer Engineering, University of Louisville,
Louisville, KY, USA
e-mail: aly.farag@louisville.edu
url: http://www.cvip.uofl.edu

S. Elshazly
Kentucky Imaging Technologies, LLC, Louisville, KY, USA

M. al Mogy
Al Mogy Scan Center, Mansoura, Egypt

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_8,

1 Introduction

In this chapter, we highlight a state-of-the-art analytic approach to lung nodule analysis using low dose CT (LDCT) of the human chest. Our focus is on small-size nodules ($\leq$1 cm in diameter) that appear randomly in the lung tissue. Radiologists diagnose these nodules by visible inspection of the LDCT scan. Despite the wide range of nodule classifications among radiologists, the nodule classification of Kostis et al. [1] is found to be particularly useful in the algorithmic evaluation presented in this work. Nodules in Kostis's work are grouped into four categories:(i) well-circumscribed where the nodule is located centrally in the lung without being connected to vasculature; (ii) vascularized where the nodule has significant connection(s) to the neighboring vessels while located centrally in the lung; (iii) pleural tail where the nodule is near the pleural surface, connected by a thin structure; and (iv) juxta-pleural where a significant portion of the nodule is connected to the pleural surface.

Figure 1 shows examples of small size nodules ($\leq$1 cm in diameter) from the four categories. The upper and lower rows show zoomed images of these nodules. Notice the ambiguities associated with shape definition, location in the lung tissues, and lack of crisp discriminatory features.

Modeling aims at representing the objects with mathematical formulation that captures their characteristics such as shape, texture and other salient features. The histogram of the object's image provides some information about its texture—the modes of the histogram indicate the complexity of the texture of the object. Figure 2 shows sample of nodules and their histograms. These histograms are essentially bimodal, for the nodule and background regions, and may be sharpened if the region of interest (ROI) is limited to be around the spatial support of the nodules.

Another difficulty of small-size nodules lies with inabilities of exact boundary definition. For example, radiologists may differ in outlining the lung nodules spatial support as shown in Fig. 3. Difference in manual annotation is common of small objects that have not well-defined description. This adds another dimension of difficulty for automatic approaches, as they are supposed to provide outputs that mimic human experts. In other words, human experts differ among themselves, how would they judge a computer output? Validation of automatic approaches for lung nodule detection, segmentation and classification - using only the visible information in an image - is an order of magnitude more difficult than that of automatic face recognition, for example.

Farag [2] studied the behavior of the intensity versus the radial distance of the nodule centroids [2]. The intensity versus radial distance distribution for small nodules was shown to decay almost exponentially. An empirical measure of the region of support of the nodules was derived based on this distribution. This approach has been tested further on three additional clinical studies in this work and has shown to hold true. The summation of the intensities of Hounsfield Units (HU) in concentric circles (or ellipses) beginning from the centroid of the nodule, decays in a nearly exponential manner with the distance from the centroid.

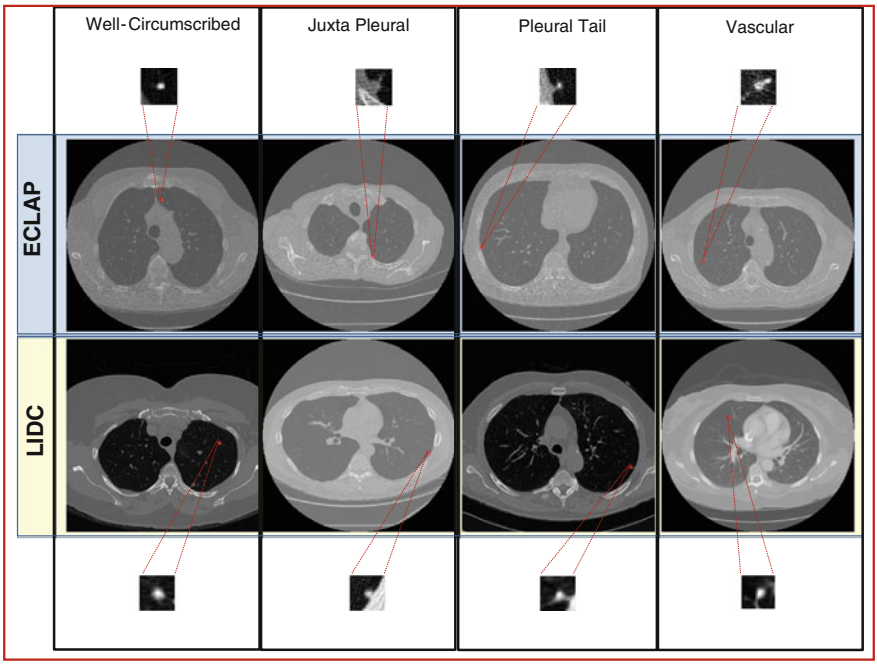


Fig. 1 Examples of lung nodules of size below 10 mm from two clinical studies. The *upper* and *lower rows* show zoomed pictures of the nodules

Figure 4 shows the radial distances for four nodule types from the LIDC clinical studies [4]. This behavior provided a clue for empirically deciding the spatial support (ROI) of the nodules—which is used for auto cropping of the detected nodules. Of course a refinement step is needed to precisely define the exact ROI of the nodule—this is carried out in nodule segmentation. This behavior is similar with the ELCAP study as well.

Object segmentation is a traditional task in image analysis. Real world objects are hard to model precisely; hence the segmentation process is never an easy task. It is more difficult with the lung nodules due to the size constraints.

Figure 5 shows the average intensity (HU) histograms of the manually cropped nodules in the ELCAP and LIDC screening studies. The histograms are distinctly bimodal and a binary classifier (thresholding) may be used for separating the nodules and non-nodules regions. The decision boundary (threshold) may be selected by various techniques, including fitting one-dimensional Gaussian density for the nodule and non-nodule regions and using the expectation-maximization approach (EM) to estimate the parameters (e.g., [2]). Unfortunately, this approach does not work well due to the uncertainties associated with the physical nodules as previously described.

There is a vast literature on object modeling and considerably larger literature on the subsequent steps of modeling; e.g., synthesis, enhancements, detection, segmen-

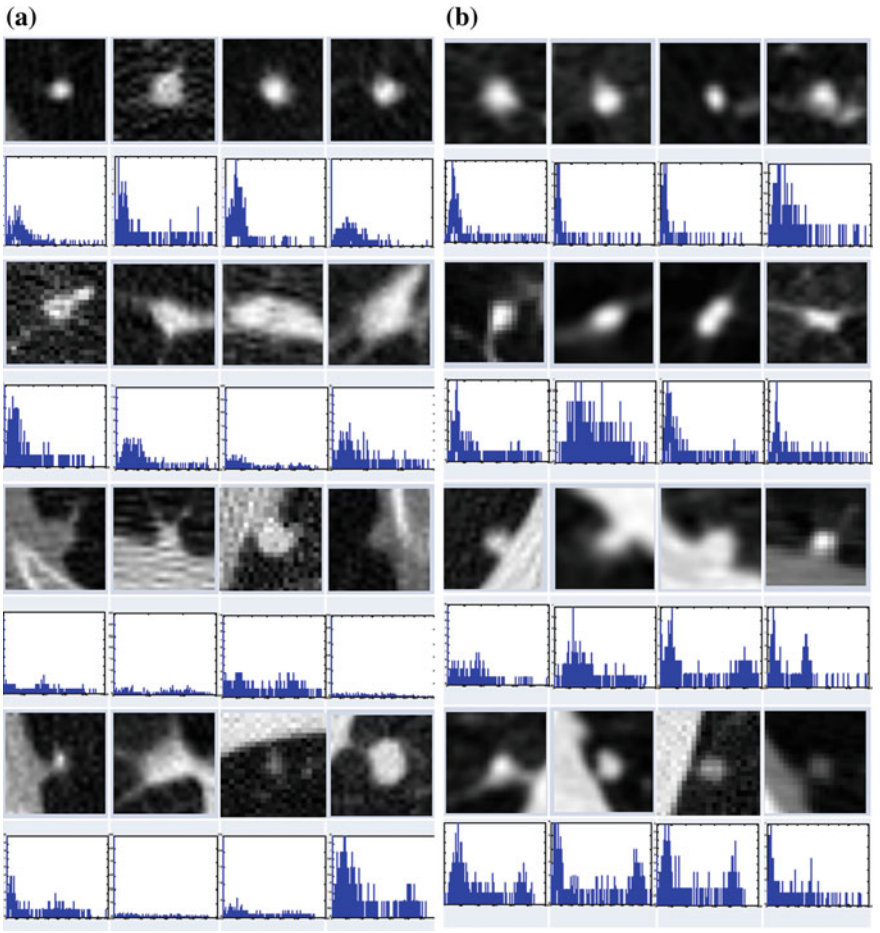


Fig. 2 Sample of nodules and their gray level (Hounsfield Units) histograms. Nodules in left are from ELCAP [3] study and those in the right table from LIDC [4] study. On *top row*, from *left* to *right*: well-circumscribed, vascular, juxta-pleural and pleural-tail nodules, respectively. **a** Nodules and histograms from the ELCAP study. **b** Nodules and histograms from the LIDC study

tation, recognition, and categorization. Farag [2] considered a five-step system for modeling of small lung nodules: (i) Acquisition and Enhancement; (ii) Parametric Modeling; (iii) Detection; (iv) Segmentation; and (v) Categorization (Classification) [2]. By constructing a front-end system of image analysis (CAD system) for lung nodule screening, all of these steps must be considered. Activities in the past few years have led to the following discoveries: (1) Feature definitions on small size objects are hard to pin point, and correspondences, among populations, is very tough to obtain automatically; (2) Classical approaches for image segmentation based on statistical maximum a posteriori (MAP) estimation and the variational level sets

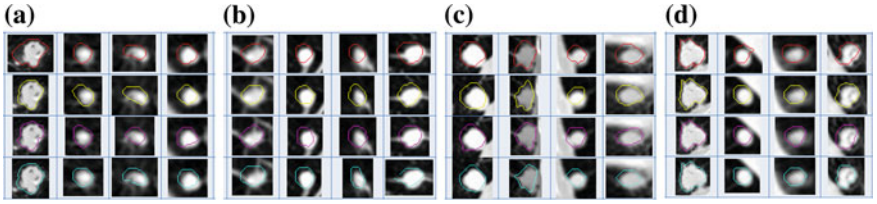


Fig. 3 Manual annotation of the main portion of the spatial support of lung nodules by *four* radiologists. Note the difference in size and shape of the annotations. **a** Outlines of fouur well-circumscribed nodules. **b** Outlines of four vascular nodules. **c** Outlines of four juxta-pleural nodules **d** Outlines of four pleural-tail nodules

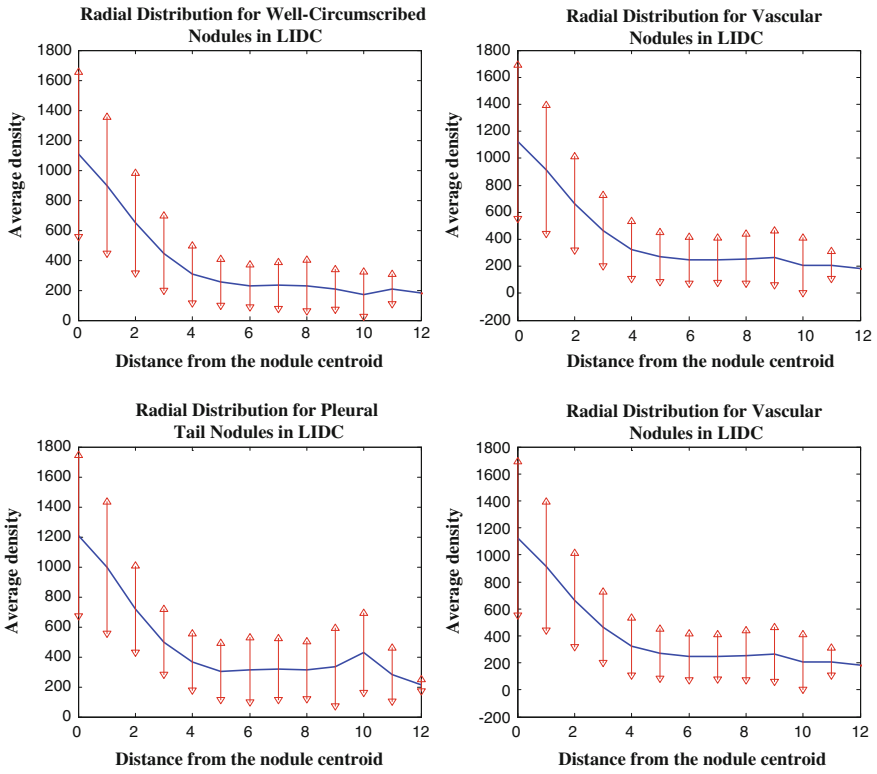


Fig. 4 Distribution of the nodule intensity (HU) for four nodule types manually cropped from the LIDC (over 2000 nodules). For nodules less than 10 mm in diameter, an ROI of size 21×21 pixels may be used

approaches do not perform well on small size objects due to unspecific object characteristics; (3) Prior information is essential to guide the segmentation and object detection algorithms—the more inclusive the a-priori knowledge, the better the performance of the automated algorithms; (4) An integration of attributes is essential for

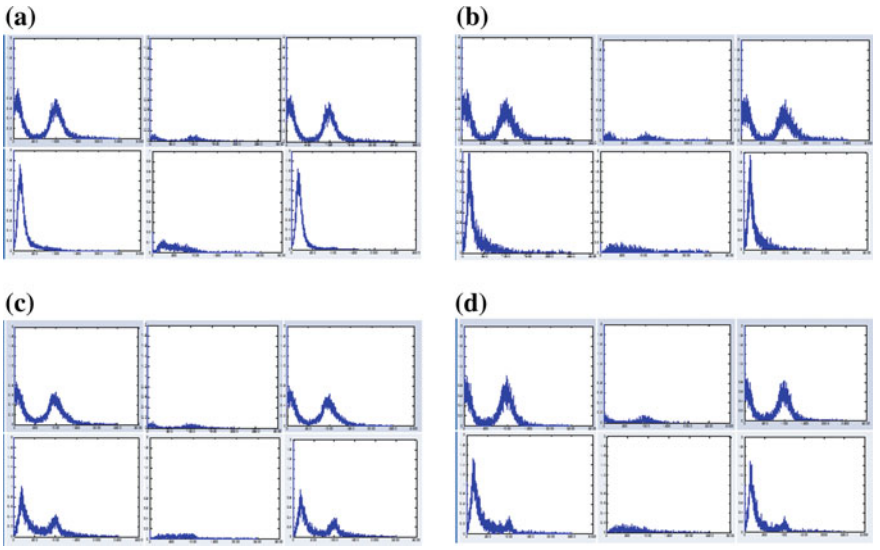


Fig. 5 The intensity (HU) histograms of the manually cropped nodules from the ELCAP and LIDC screening studies. These histograms are bio-modal showing the nodule and non-nodule regions in the ROI. These histograms are used as estimates of the probability density functions in the nodule segmentation process. **a** Intensity of well-circumscribed nodules for ELCAP (*upper*) and LIDC (*lower*) **b** Intensity of vascular nodules for ELCAP (*upper*) and LIDC (*lower*) **c** Intensity of juxta-pleural nodules for ELCAP (*upper*) and LIDC (*lower*) **d** Intensity of pleural-tail nodules for ELCAP (*upper*) and LIDC (*lower*)

robust algorithmic performance; in particular shape, texture, and approximate size of desired objects are needed for proper definition of the energy functions outlining the MAP or the level sets approaches. These factors play a major motivational role of this work.

The rest of the material in this chapter will focus on four steps related to an analytic system for lung nodule analysis: lung nodule modeling by active appearance; lung nodule detection; lung nodule segmentation; and lung nodule categorization.

2 Modeling of Lung Nodules by Deformable Models

Deformable models are common in image modeling and analysis. Random objects provide major challenges as shapes and appearances are hard to quantify; hence, formulation of deformable models are much harder to construct and validate. In this work, we devise an approach for annotation, which lends a standard mechanism for building traditional active appearance (AAM), active shape (ASM) and active tensor models (ATM). We illustrate the effectiveness of AAM for nodule detection.

Automatic approaches for image analysis require precise quantification of object attributes such as shape and texture. These concepts have precise definitions, but their descriptors vary so much from one application to another. A shape is defined to be the information attributed to an object that is invariant to scale, origin and orientation [5]. A texture may be defined as the prevalence pattern of the interior of an object [6]. Geometric descriptors identify "features" that are "unique" about an object. Shape, texture and geometric descriptors are major concepts in this work; they will be defined and used in the context of modeling small size objects under uncertainties [7]. The theoretical development in this work falls under the modern approaches of shape and appearance modeling. These models assume the availability of an ensemble of objects annotated by experts—the ensemble includes variations in the imaging conditions and objects attributes to enable building a meaningful statistical database.

Active shape models (ASM) and active appearance models (AAM) have been powerful tools of statistical analysis of objects (e.g., [8, 9]). This section highlights some of the authors' work on *data-driven* lung nodule modeling and analysis (e.g., [10, 11]), with focus on active appearance models (AAM).

2.1 Lung Nodule Modeling

Real world objects may take various forms of details, and may be linear, planar or three-dimensional. In [7], Dryden and Marida, define *anatomical landmarks* as points assigned by an expert that corresponds between organisms in some biologically meaningful way; *mathematical landmarks* as points located on an object according to some mathematical or geometrical property, i.e. high curvature or an extremum point; and *pseudo-landmarks* as constructed points on an object either on the outline or between landmarks. Figure 6 is a sample of small-size nodules smaller than 1 cm in diameter from the LIDC [4] clinical study, showing the variations that can be captured by shape and appearance models.

From a computer vision prospective, AAM and ASM modeling have been used with great successes in objects having distinct landmarks (e.g., [8, 9]). A shape is considered to be a set of $n-$vertices $x \in R^k$; for the two-dimensional case:

$$\boldsymbol{x} = [x_1; x_2; \cdots; x_n; y_1; y_2; \cdots; y_n]^T \tag{1}$$

The shape ensemble (realizations of the *shape process* of a certain object) is to be adjusted (aligned) on the same reference to enable filtering of scale, orientation and translation among the ensemble, per the shape definition. This alignment generates the so-called *shape space*, which is the set of all possible shapes of the object in question. To align the shapes in an ensemble, various procedures may be used. The *Procrustes* procedure is common for *rigid* shape alignments. The alignment process removes the redundancies of *scale*, *translation* and *rotation* using a similarity measure that provides the minimum *Procrustes distance*.

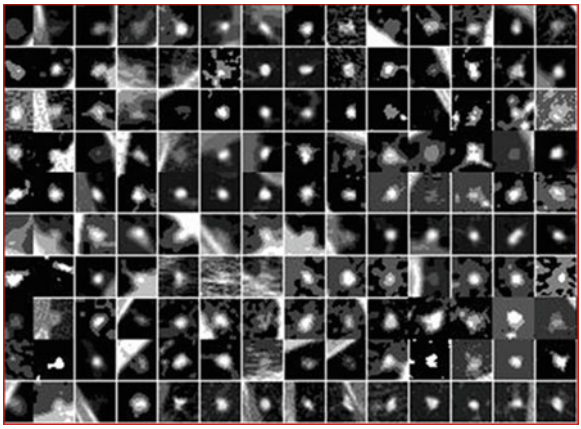


Fig. 6 An ensemble of 140 nodules manually cropped from the LIDC study

Suppose an ensemble of shapes is available with one-to-one point (feature) correspondence is provided. The Procrustes distance between two shapes s_1 and s_2 is the sum of squared distance (SSD)

$$P_d^2 = \sum_{j=1}^{n} \left(x_{j_1} - x_{j_2}\right)^2 + (y_{j_1} - y_{j_2})^2 \tag{2}$$

Annotated data of an ensemble of shapes of a certain object carries redundancies due to imprecise definitions of landmarks and due to errors in the annotations. Principal Component Analysis (PCA) may *be used for reducing these* redundancies. In PCA, the original shape vector is linearly transformed by a mapping such that has $\mathbf{z} = \mathbf{M}\mathbf{x}$ less correlated and highly separable features. The mapping $\mathbf{M}$ is derived for an ensemble of N shapes as follows**:**

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i;\ \sum_x = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})^T \tag{3}$$

are the mean and covariance of $\mathbf{X}$. Therefore, the mean and covariance of $\mathbf{z}$ would be**:**

$$\bar{z} = \frac{1}{N}\Sigma_{j=1}^{N}\mathbf{z}_j \tag{4a}$$

$$\Sigma_{\mathbf{z}} = \frac{1}{N}\Sigma_{i=1}^{N}(\mathbf{z}_i - \bar{z})(\mathbf{z}_i - \bar{z})^T = \mathbf{M}\Sigma_x\mathbf{M}^T \tag{4b}$$

If the linear transformation $\mathbf{M}$ is chosen to be orthogonal; i.e., $\mathbf{M}^{-1} = \mathbf{M}^T$, and selecting it as the eigenvectors of the symmetric matrix $\sum_x$, this would make $\sum_z$ to be a diagonal matrix of the eigenvalues of $\sum_x$. The eigenvectors corresponding to the small eigenvalues can be eliminated, which provides the desired reduction.

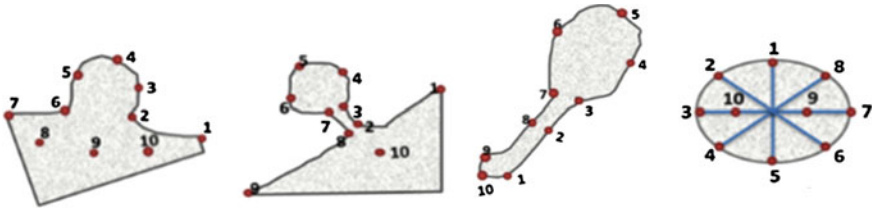


Fig. 7 Definition of Control points (landmarks) for nodules. *Right*-to- *left*: juxta-pleural, pleural tail, vascular, a well-circumscribed nodule models

Therefore , $\mathbf{x}$ may be expressed as:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\,\mathbf{b} \tag{5}$$

where $\mathbf{P} = (\boldsymbol{p}_1|\boldsymbol{p}_2|\ldots|\boldsymbol{p}_m)$ matrix of m largest eigen vectors of $\sum_x$ and $\mathbf{b} = \mathbf{P}^{\mathbf{T}}(\mathbf{x} - \bar{\mathbf{x}})$ **is an** $m \times 1$ vector. Equation (5) is the statistical shape model, which is derived using PCA. By varying the elements of $\mathbf{b}$ one can vary the synthesized shape $\mathbf{x}$ in Eq. (5). The variance of the ith parameter $b_i \in \mathbf{b}$ can be shown across the training set to be equal to the eigenvalue λ_i [8].

2.2 Nodule Annotation

In order to construct the active appearance or active tensor models, we need an annotated ensemble of objects. In case of random objects, the annotation process becomes extremely difficult; it takes yet another level of difficulty with small-size. Yet, the major goal of this work is to address such objects, specifically, small size lung nodules, which are used for early detection screening of possible lung cancer. We used the fuzzy description of lung nodules from Kostis et al. [1] to devise a feature definition approach for four categories of nodules; *well-circumscribed*, *vascularized, juxta-pleural* and *pleural-tail* nodules. Figure 7 illustrates the landmarks that correspond to the clinical definition of these four nodule categories.

Using the above definitions, we created a manual approach to annotate the nodules. First, we take the experts' annotation, zoom it and manually register it to a template defining the nodule type/category, and then we select the control points on the actual nodule using the help of the template. This annotation enabled creation of active appearance models, which mimics largely the physical characteristics of lung nodules that cannot be modeled otherwise.

Figure 8 shows examples for the nodule models generated by ensembles from the ELCAP and LIDC clinical lung screening studies**.** The average nodules (shown in Fig. 8) capture the main features of real nodules. Incorporation of other basis has been studied in Farag et al. [11]. Figure 9 shows examples of AAM nodule models with additional "Eigen nodules".

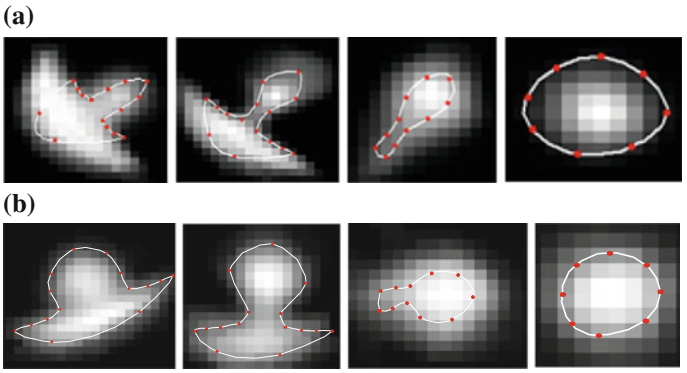


Fig. 8 AAM Models for lung nodules from clinical CT scans. *Right*-to-*left*: juxta-pleural, pleural tail, vascular, a well-circumscribed nodule models. **a** Average nodules from ELCAP study. **b** Average nodules from LIDC study

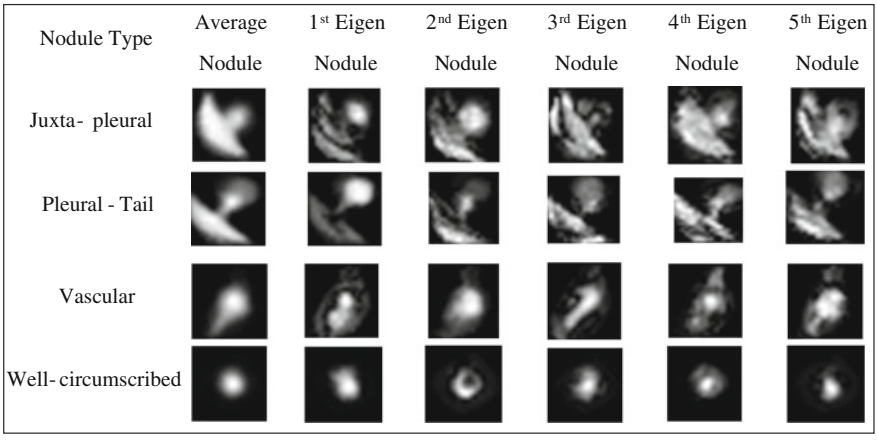


Fig. 9 Average and 1st five eigen nodules on ELCAP study

3 Lung Nodule Detection

The above modeling approach has provided tremendous promise in three subsequent steps of lung nodule analysis: detection, segmentation, and categorization. Due to space limitations, we only consider lung nodule detection using the AAM nodule models. Further, we use only a basic detection approach that is based on template matching with normalized cross-correlation (NCC) as similarity measure. Other measures have been examined in our related work (e.g., [11]). We report the detection performance by constructing the ROC of both the ELCAP and LIDC clinical studies. We chose to limit the ensemble size for modeling to be 24 *per nodule type* for the two studies, to provide a comparison with our earlier work [10]. The ROCs are

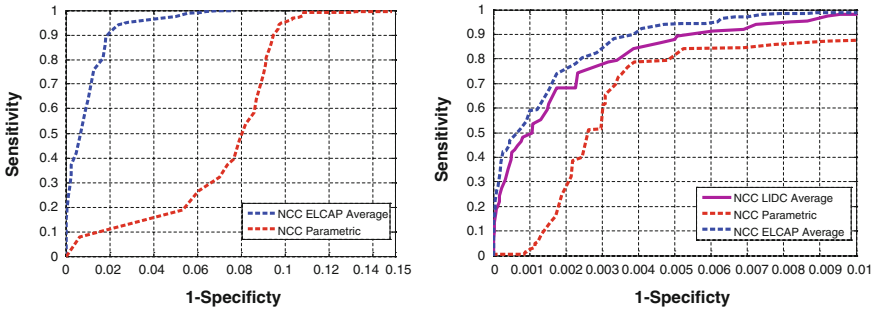


Fig. 10 ROC curves for template matching detection on the ELCAP and LIDC database versus the circular and semi-circular models. **a** ROC for the ELCAP study. **b** ROC for the LIDC study

built to show the overall *sensitivity* and of the detection process. The textures of the parametric nodules were generated by the analytical formulation in our earlier work (e.g., [10]).

3.1 Clinical Evaluation

ELCAP Data: The ELCAP database [3] contains 397 nodules, 291 identified and categorized nodules are used in the detection process. Results using only the average (mean) template models generated from the AAM approach is examined against parametric nodule models, (i.e. circular and semi-circular) of radius 10, templates in this first set of experiments.
LIDC Data: The Lung Imaging Data Consortium (LIDC) [4] contains 1018 helical thoracic CT scans from 1010 different patients. We used ensembles of 24 nodules per nodule type to design the nodule models (templates) and the rest to test the detection performance.

Figure 10 shows the ROC of 1—specificity versus sensitivity. The results show the superior performance of the AAM-models over the parametric models. In generating these ROC curves, we used the mean in the AAM models as the nodule template (note: in [11] we used other *eigen-nodules* besides the mean).

We note from Figure 10 that the templates from the ELCAP ensemble provided better performance than those from the LIDC ensemble. This because the wide range of variations in texture information found in the LIDC database, which affects the appearance of the resulting nodule model (template). We used 24 nodules, per nodule type, in both ELCAP and LIDC in order to have even comparison. It is expected that the better AAM models result with larger ensemble size; which is possible with the LIDC study as it contains over 2000 nodules vs. ELCAP which is only few hundreds.

3.2 Extensions

We note that in the ELCAP database, the data acquisition protocol was the same throughout; very low resolution. That was reflected in the AAM model, showing a texture that is relatively more homogenous than that in the LIDC case, which uses data from various imaging centers and various imaging scanners, with somewhat variable range of Hounsfield Units (HU). In general, if we include more nodules in the design, we expect a better appearance modeling; the LIDC database allows such choice.

This section dealt with modeling of small-size lung nodules using two clinical studies, the ELCAP and LIDC. We discussed the process of nodule annotation and the steps to create AAM nodule models. These models resemble the real nodules, thus using them as templates for nodule detection is more logical than the non-realistic parametric models. These types of models add two additional distinctions over the parametric approaches; it can automate the processes of nodule segmentation and categorization. Tensor modeling may also be used to generate the nodule models. From the algorithmic point of view, an *adaboost* strategy for carrying out the detection may lend speed advantage over the typical cross-correlation implementation used in this work.

4 Nodule Segmentation

This section describes a variational approach for segmentation of small-size lung nodules which may be detected in low dose CT (LDCT) scans. These nodules do not possess distinct shape or appearance characteristics; hence, their segmentation is enormously difficult, especially at small size ($\leq$1 cm). *Variational methods hold promise in these scenarios despite the difficulties in estimation of the energy function parameters and the convergence. The proposed method is analytic and has a clear implementation strategy for LDCT scans.*

The lungs are a complex organ which includes several structures, such as vessels, fissures, bronchi or pleura that can be located close to lung nodules. Also, the main "head" of the nodule is what radiologists consider when computing the size. In the case of detached nodules (i.e. well-circumscribed nodules) the whole segmented nodule is considered in size computations and growth analysis, while in detached nodules (i.e. juxta-pleural, vascularized and pleural-tail) the "head" is required to be extracted from the anatomical surrounds. Intensity-based segmentation [13, 14] has been applied to nodule segmentation using local density maximum and thresholding algorithms. These classes of algorithms are primarily effective for solitary nodules (well-circumscribed), however, fail in separating nodules from juxtaposed surrounding structures, such as the pleural wall (i.e., juxta-pleural and pleural-tail nodules) and vessels (vascular nodules), due to their similar intensities.

More sophisticated approaches have been proposed to incorporate nodule-specific geometrical and morphological constraints (e.g., [1, 15–17]). However, juxta-pleural, or wall-attached, nodules still remain a challenge because they can violate geometrical assumptions and appear frequently. Robust segmentation of the juxta-pleural cases can be addressed in two approaches: a) global lung or rib segmentation (e.g., [18]), and b) local non-target removal or avoidance [14]. The first can be effective but also computationally complex and dependent on the accuracy of the whole-lung segmentation. The second is more efficient than the former but more difficult to achieve high performance due to the limited amount of information available for the non-target structures. Other approaches have been proposed in the literature (e.g., [19]), but require excessive user interaction. In addition, some approaches assumed predefined lung walls before segmenting the juxta-pleural nodules (e.g., [20]).

4.1 Variational Approach for Nodule Segmentation

The level set function as a signed distance map is able to capture complicated topological deformations. A level set function $\emptyset : \Omega \subset R^2 \rightarrow R$ can be defined as the minimum Euclidean distance between the point $\mathbf{X} \in \Omega$ and the shape boundary points. A curve can be initialized inside an object, and then evolves to cover the region guided by image information. The evolving curve within the level set formulation is a propagating front embedded as the zero level of a 3D scalar function $\emptyset(\boldsymbol{X}, t)$, where $\boldsymbol{X}$ represents a location in space. In order to formulate the intensity segmentation problem, it is necessary to involve the contour representation. Given an image $I : \Omega \subset R^2 \rightarrow R$, the segmentation process aims to partition the image into two regions: object (inside the contour denoted by **o**) and background (outside the contour denoted by **b**). An error term can be computed by counting the number of correctly classified pixels and then measuring the difference with respect to the total number of pixels. This can be done by summing up the probabilities of the internal pixels to be *object* and the external pixels probabilities to be classified as *background*. This is measured by the term:

$$Error = 1 - \pi_o \int_{\Omega_o} P_o(I(\boldsymbol{X}))\, d\Omega - \pi_b \int_{\Omega_b} p_b(I(\boldsymbol{X}))\, d\Omega \tag{6}$$

where p_o and p_b are the probabilities of the object and background according to the intensity values (Gaussian distributions are used to model these regions). Prior probabilities of regions (π_o and π_b) are involved in the formulation as well. Minimizing this error term is equivalent to minimizing the energy functional:

$$E(\emptyset) = -\pi_o \int_{\Omega_o} p_o H_\epsilon(\emptyset)\, d\Omega - \pi_b \int_{\Omega_b} p_b H_\epsilon(-\emptyset)\, d\Omega \tag{7}$$

where H is the Heaviside step function and $\epsilon \in R^+$ represents the narrow band region width. An extra term is added to the energy function to represent the contour arc-length (L) which also needs to be minimal to guarantee a smooth evolution. The new energy will be:

$$E(\emptyset) = -\pi_o \int_{\Omega_o} p_o H_\epsilon(\emptyset)\, d\Omega - \pi_b \int_{\Omega_b} p_b H_\epsilon(-\emptyset)\, d\Omega + \lambda L \tag{8}$$

where $\lambda \in R^+$. The level set function evolves to minimize such a functional using the Euler-Lagrange formulation with the gradient descent optimization:

$$\frac{\partial \emptyset}{\partial t} = \delta_\epsilon(\emptyset)(\pi_o p_o - \pi_b p_b) + \lambda \mathrm{k} \tag{9}$$

where δ is the derivative of the Heaviside function and k is the curvature. Thus, the evolution depends on the local geometric properties (local curvature) of the front and the external parameters related to the input data I. The function $\emptyset(\cdot, \cdot)$ deforms iteratively according to the above equation, while solving $\emptyset(\boldsymbol{X}, t = 0$ gives the position of the 2D front iteratively. Let $\emptyset_g$ denote the intensity segmented region function representation The Gaussian distribution and prior probabilistic parameters are computed according to the method in [21].

4.2 Shape Alignment

This process aims to compute a transformation $\mathbf{A}$ that moves a source shape (α) to its target (β). The in-homogeneous scaling matching criteria from [21] is adopted, where the source and target shapes are represented by the signed distance functions $\emptyset_\alpha$ and $\emptyset_\beta$ respectively. The transformation function is assumed to have scaling components: $\boldsymbol{S} = diag(s_x, s_y)$, rotation angle, θ (associated with a rotation matrix $\mathbf{R}$) and translations: $\boldsymbol{T} = [T_x, T_y]^T$ A dissimilarity measure to overcome the scale variance issue is formulated by assuming that the signed distance function can be expressed in terms of its projections in the coordinate directions as: $\mathbf{d}_\alpha = [d_x, d_y]^T$ at any point in the domain of the shape α . Applying a global transformation $\mathbf{A}$ on $\emptyset_\alpha$ results in a change of the distance projections to $\mathbf{d}'_\alpha = \mathbf{RSd}_\alpha$ which allows the magnitude to be defined as: $\emptyset'_\alpha = ||\mathbf{Sd}_\alpha||$ which implies that $\emptyset'_\alpha \leq \max(s_x, s_y)\emptyset$ Thus, a dissimilarity measure to compute the difference between the transformed shape and its target representation can be directly formulated as:

$$r(\mathbf{X}) = ||\mathbf{RSd}_\alpha(\mathbf{X})|| - \emptyset_\beta(\mathbf{A}). \tag{10}$$

By summing-up the squared difference between the two representations, an energy function can be formulated as:

$$E_1 = \int_{\Omega} \delta'_{\epsilon}(\emptyset_{\alpha}, \emptyset_{\beta})\, r^2\, d\Omega \tag{11}$$

where δ'_{ϵ} reduces the complexity of the problem and ϵ is the width parameter of the band around the shape contour. The given measure r, from the shown derivations, satisfy the relation $r \leq s\emptyset_{\alpha}(\mathbf{X}) - \emptyset_{\beta}(\mathbf{A})$, where $s = \max(s_x, s_y)$. Thus, an energy function can be obtained where $E \leq E_1$;

$$E = \int_{\Omega} \delta'_{\epsilon}(\emptyset_{\alpha}, \emptyset_{\beta})\, \left(s\emptyset_{\alpha}(\mathbf{X}) - \emptyset_{\beta}(\mathbf{A})\right)^2 d\Omega \tag{12}$$

The above functional better describes the registration since it incorporates a scaled version of the source shape representation. In this work, the gradient descent optimization is used to solve the problem, which requires the involved functions to be differentiable. A smeared version of $s(s_x, s_y) = \max(s_x, s_y)$, is used at the line since, $(s_x = s_y)$ the function is not differentiable there, which is based on its original definition:

$$s(s_x, s_y) = \max(s_x, s_y) = s_x H_{\epsilon}(s_x - s_y) + s_y(1 - H_{\epsilon}(s_x - s_y)) \tag{13}$$

which will return s_x if $s_x - s_y \geq 0$, otherwise s_y. The smeared Heaviside step function H is used to obtain a smooth transition around the line $s_x = s_y$ allowing the function to be differentiable everywhere. The function derivatives will be calculated as

$$\frac{\partial s}{\partial s_x} = H_{\epsilon}(s_x - s_y) + (s_x - s_y)\delta_{\epsilon}(s_x - s_y) \tag{14}$$

$$\frac{\partial s}{\partial s_y} = H_{\epsilon}(s_y - s_x) + (s_y - s_x)\delta_{\epsilon}(s_y - s_x) \tag{15}$$

The parameters $\{s_x, s_y, \theta, T_x, T_y\}$ are required to minimize the energy functional E.

4.3 Level Set Segmentation Algorithm with Shape Prior

The above steps have resulted in an algorithm whose input is LDCT scans and output is segmented lung nodules. The algorithm can be summarized as follows:

***Lung Nodule Segmentation Algorithm*:**

1. Segment the Lungs from their surroundings—*Lung tissue segmentation* (e.g., [22]).
2. Train the lung nodule modeling step on a portion of the data at hand—*Lung Nodule Modeling*
3. Apply the lung nodule detection approach to compute the positions of the candidate nodules and hence crop them for classification. Cropping here means setting a

box around the nodule center and extracts its neighbor area from the surroundings; i.e., a region of interest, ROI, is cropped around the detected nodules—*Nodule detection and ROI determination*
4. Based on the input image size, construct the initial prior shape circle and its shape model representation $\emptyset_p$.
5. Solve Eq. 8 to compute the intensity segmentation region representation $\emptyset_g$. Solution is iterative until the function converges—reaches a certain state. Note the function keeps the sign distance property by following the approach in [22].
6. Initialize the transformation parameters to $s_x = 1, s_y = 1$ and . $\theta = 0$ At this moment the nodule center location is manually selected which initializes the translation parameters t_x and ty
7. Solve the gradient descent approach to minimize the energy in Eq. 11. Parameters converge to their steady state values and hence the final boundaries of the ellipse are computed.
8. Threshold the region inside the ellipse to accurately mark the nodule pixels. The resulting region may undergo a median filter smoothing step to remove noisy pixels.

4.4 Some Results

This work is validated using four different databases. The first is the ELCAP [3] public database, DB1.This database has nodules of diameter ranging from 2 to 5 mm. The second database (DB2) contains 108 nodules from LDCT scans of slice thickness 2.5 mm and a pixel-spacing of 0.72461×0.72461 mm (diameter from 2.9 to 6 mm). The third database (DB3) has 28 nodules, 1.25 and 2.5 mm slice thickness, and nodules diameter ranging from 7 to 20 mm. The fourth dataset is the LIDC (DB4) which contains nodules ranging in sizes. The slices are both low-dose and high-dose CT images [4].

Figure 11 demonstrates the performance of a number of model-based methods for nodule segmentation. Nodules are cropped by four different radiologists, and the approaches are applied to these cropped nodules. Overall the variational shape-based level set method provided the best segmentation results for obtaining the "head" of the nodule region. The results show that the intensity-based approaches can be used as an initial or post segmentation process to the variational shaped-based level sets. Also, approaches where a shape model can be embedded into the formulation of the segmentation method are necessary for such cases as nodule segmentation.

The developed approach uses a region of interest (ROI) image that contains the lung nodule as input. Image intensity segmentation using level sets (as described above) is used to extract the non-lung regions from the lung tissue regions and represents the slices by a level set function ($\emptyset_g$). Different scales, rotation, and translation parameters are computed in each case to obtain an ellipse exactly around the nodule head (see Fig. 12). Changes of the shape model can be noticed until the

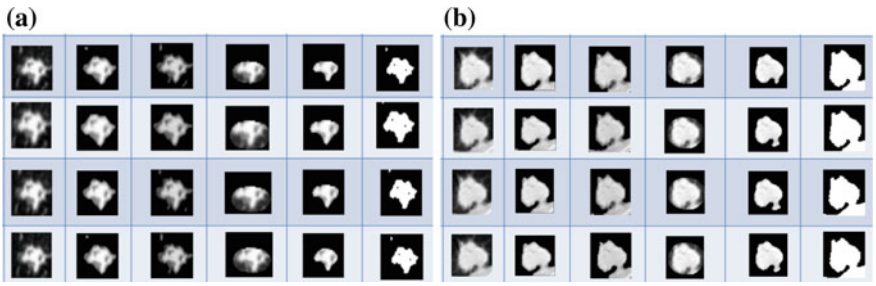


Fig. 11 Nodule segmentation by a number of approaches. Columns 5 and 6 show results of the variational approach with and without shape alignments. First column is a nodule segmented by four radiologists. Second column is the EM segmentation. Third column is the level set method. Fourth column is level sets plus shape priors. Fifth column is EM plus shape priors. Last column is graph cuts. **a** Nodule centrally located in the lung tissue. **b** Nodules connected to the pleural surface

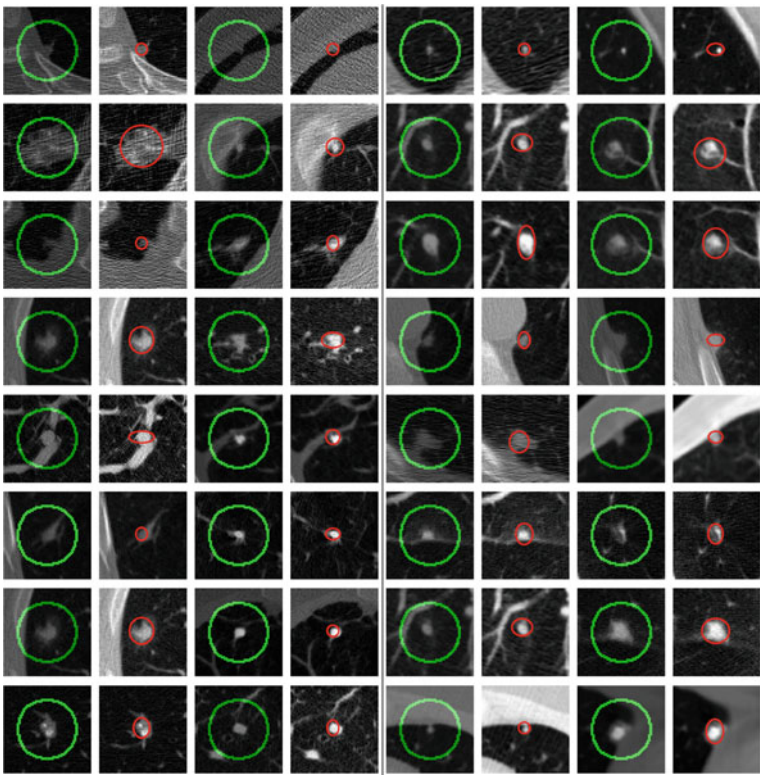


Fig. 12 Nodule segmentation results from DB1 (*left* block-first four columns) and DB2 (*right* block-last four columns). Initialization is given in green while final nodule boundaries are shown in *red*

steady state around the nodule boundaries is reached. Also, the axis of the ellipse rotates and varies in size to include the most boundary information of the nodule. The approach is robust for various nodule sizes from larger nodules (<1 cm) to nodules that occupy smaller spatial support regions (i.e. >1 cm). Similar results are obtained from other databases (e.g., [12, 23]).

4.5 Extensions

Among the possible extensions of the above algorithm are the following:

a. Proper modeling of the shape shape priors in the statistical segmentation approach.
b. Generlizing the transformation paramters that embed the shape model into the image domain, thus, avoiding the post EM step;
c. Incorporatation of the shape priors into the energy function, of general topological cliques in the MGRF models, and evaluation of the segmentation algorithm with respect to variational shape-based techniques such as level sets.

The nodule segmentation is a componenet of the CAD system for analysis of lung nodules; it requires exhaustive validation by large scale clinical studies and various radiologists.

5 Nodule Categorization

In the computer vision and biomedical imaging literature the terms categorization, classification, identification, and recognition share a lot of commonality of methods and purpose. In the lung nodule example, one may also denote the classification step as recognition. However, classification may indeed entail two aspects: assigning segmented objects into types (classes, such as the four nodule types that we have been considering in this chapter), or assigning them into a definitive group (e.g., pathology in the lung nodule case). Our focus is on descriptors that adhere to shape and appearance contexts

5.1 Object Feature Descriptors

In the past decade, several object descriptors have been introduced in the computer vision literature, including the local binary pattern (LBP) [24] and the scale-invariant feature transform (SIFT) [25]. A comprehensive evaluation of the geometric feature descriptors may be found elsewhere, in particular Mikolajczyk and Schmid 2005 [26]. Below we describe the LBP and SIFT descriptors and their performance for small-size nodule categorization (see [23] for detailed evaluation).

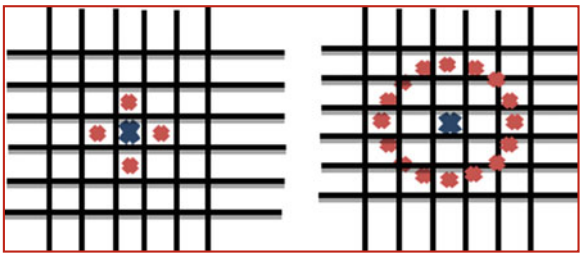


Fig. 13 Circularly symmetric neighbor sets for different values of (P , R); *left* (a) P = 4, R = 1.0; *right* P = 16, R = 2.0

5.1.1 Multi-Resolution Local Binary Pattern (LBP)

The Local Binary Pattern is an operator invariant to monotonic changes in grayscale and can resist illumination variations as long as the absolute gray-level value differences are not badly affected (e.g., [24]). The original operator labeled the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value and considered the result as a binary number. At a given pixel position (x_c, y_c), the decimal form of the resulting 8-bit word is

$$\mathrm{LBP}(\mathrm{x_c}, \mathrm{y_c}) = \sum_{\mathrm{i}=0}^{7} \mathrm{s}(\mathrm{I_i} - \mathrm{I_c})2^{\mathrm{i}} \tag{16}$$

where, $\mathrm{I_c}$ corresponds to the center pixel $(\mathrm{x_c}, \mathrm{y_c})$, $\mathrm{I_i}$ to gray level values of the *eight* surrounding pixels and function $\mathrm{s}(\cdot)$ is a unit-step function.

The LBP operator was extended to a circular neighborhood of different radius size to overcome the limitation of the small original 3×3 neighborhood size failing to capture large-scale structures. Each instance is denoted as (P, R), where P refers to the equally spaced pixels on a circle of radius R. The parameter P controls the quantization of the angular space and R determines the spatial resolution of the operator. An LBP pattern is considered uniform if it contains at most two bitwise transitions from 0 to 1 and vice-versa, when the binary string is circular. The reason for using uniform patterns is that they contain most of the texture information and mainly represent texture primitives. The operator is derived on a circularly symmetric neighbor set of P members on a circle of radius R denoting the operator as LBP^{u2}_{PR}.

Figure 13 illustrates examples of circularly symmetric neighbor sets for various (P, R). The LBP operator was further enhanced by combining it with a rotation invariant measure $VAR_{P,R}$, which characterizes the contrast of local image texture. The combination of the LBP^{u2}_{PR} operator and the variance measure produces a powerful operator that is rotation and gray-scale invariant.

In the multi-resolution analysis the responses of multiple operators realized with different (P , R) are combined together and an aggregate dissimilarity is defined as the sum of individual log-likelihoods computed from the responses of individual operators [24]. The notation LBP^{u2}_{PR} used in this chapter refers to the extended LBP

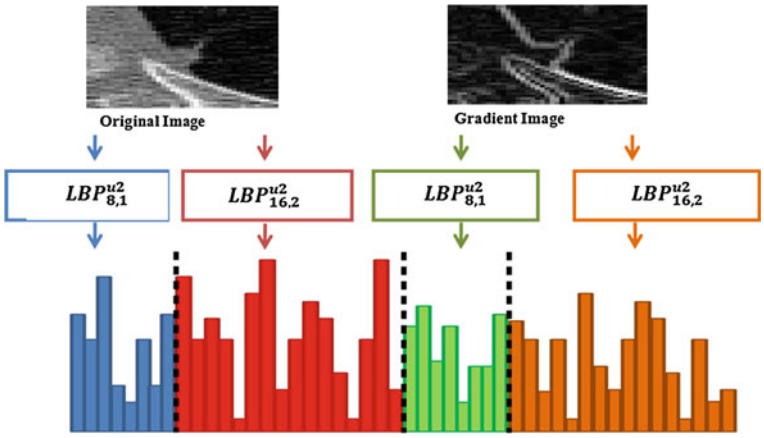


Fig. 14 Block diagram of generating the LBP for a juxta-pleural nodule. The equation for the above picture is: $\mathbf{LBP}^{u2}_{8,1} + \mathbf{LBP}^{u2}_{16,2} + \mathbf{LBP}^{u2}_{8,1} + \mathbf{LBP}^{u2}_{16,2}$, where the first two terms represent the original image and the last two terms represent the gradient image

operator in a neighborhood, with only uniform patterns considered. The LBP is used to generate a feature vector which describes the nodule region of interest in a LCDT slice. The LBP is applied to one of three scenarios on: (i) the original nodule images; (ii) the gradient of the nodule image or, (iii) an addition of the original and gradient nodule images. The gradient image was computed by first obtaining each individual image in the x- and y-spaces by filtering the corresponding directional-space original image with the corresponding parameter vector identified in the author's work (e.g., [23, 27]); the overall gradient nodule image is:

$$\nabla_{\text{nodule}} = \sqrt{\nabla_{\text{x}}^2 + \nabla_{\text{y}}^2} \tag{17}$$

A similarity measure is then used to classify these nodules to one of the four classes: juxta, well-circumscribed, pleural tail and vascularized. Principle component analysis (PCA) and linear discriminant analysis (LDA) are used to project the extracted LBP descriptors to a low-dimensional subspace where noise is filtered out. Figure 14 illustrates the formation of the LBP descriptors on lung nodules.

5.1.2 The Signed Distance Transform

The distance transform is a shape-based feature descriptor that represents each pixel of the binary edge map image with a distance to the nearest obstacle pixel (i.e., binary pixel). The extracted Signed Distance transform images were projected to a lower-dimensional subspace using PCA and LDA. The LBP of the signed distance image results were also obtained, thus, resulting in a combinational shape and texture feature descriptor representation of the nodules and non-nodules. The relevance of

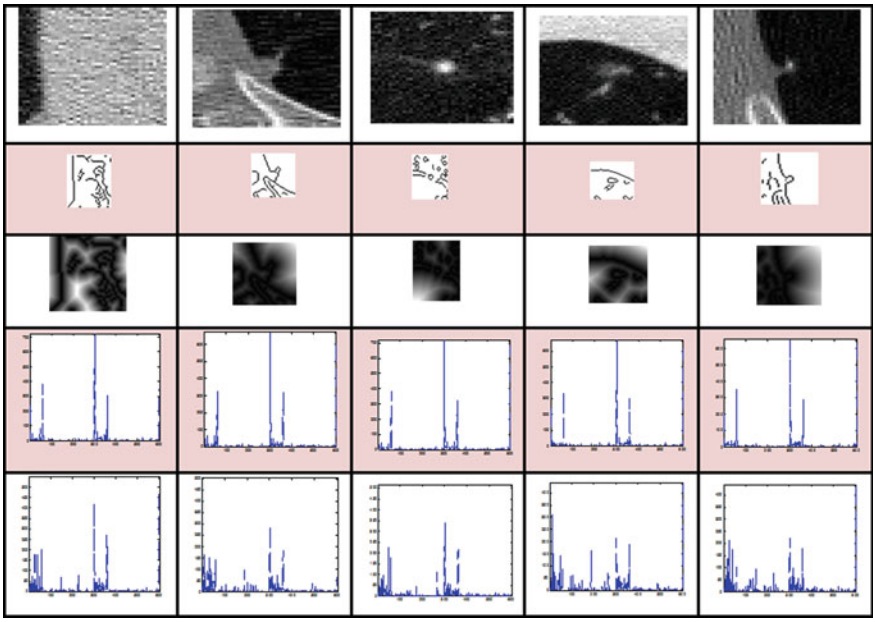


Fig. 15 *First* row shows typical non-nodule (*first*column) and nodule textures (juxta-pleural, well-circumscribed, vascularized and pleural tail, respectively). *Second* row shows edge maps (using the Canny Operator). *Third* row is the signed distance. *Fourth* row is the LBP of the nodules. Final results depict the LBP + Signed distance features

combining shape and texture feature vectors is described in the recognition stage. Figure 15 illustrates the approach which combines the LBP and Signed Distance Transform.

5.1.3 The Scale-Invariant Feature Transform (SIFT)

The SIFT is a combinational detector and descriptor approach introduced by Lowe [25] that allows extraction of distinctive scale and rotation invariant features from images. The SIFT is a combination of a scale invariant region detector known as the difference of Gaussian (DoG) detector and a proper descriptor referred to as SIFT-key. The approach consists of four major steps of computation to generate the set of image features: Scale Space extrema detection, Key-point Localization, Orientation assignment and Key-point descriptor. In the first stage of computation all scales and image locations are searched over using a DoG function to identify potential interest points that are invariant to orientation and scale. Once the potential interest points are found at each candidate location a detailed model is fitted to determine scale and location.

The keypoints selected are based on the stability measures. To each keypoint location one or more orientations are assigned based on the local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned scale, location and orientation for each feature. At the selected scale in the region around each keypoint the local image gradients are measured and transformed into a representation that allows for significant levels of change illumination and local shape distortion. The scale-space of an image defined as a function, $L(x, y, \sigma)$, was shown by Koenderink [28] and Lindeberg [29] as follows: The only possible scale-space kernel, under reasonable assumptions, is the Gaussian function, thus the scale-space of an image $L(x, y, \sigma)$ is produced from convolving a variable-scale Gaussian, $G(x, y, \sigma)$, with an input image, $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{18}$$

Lowe proposed using scale-space extrema in the difference-of-Gaussian function, to accurately detect stable keypoint locations in scale-space, convolved with the image, $D(x, y, \sigma)$ which from the difference of two nearby scales separated by a constant multiplicative k factor can be computed:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{19}$$

In order to detect the local maxima and minima of $D(x, y, \sigma)$, each sample point is compared to its eight neighbors in the current image and nine neighbors in the scale above and below. The keypoint is selected if it larger or smaller than all of these neighbors. Once the keypoint candidate is obtained a detailed fit to the nearby data for location, ratio of principal curvatures and scale is performed to reject points with low contrast or poorly localized along an edge. Consistent orientation assignment to each keypoint based on local image properties allows the keypoint descriptor to be represented relative to this orientation and thus achieve invariance to image rotation. The scale of the keypoint is used to select the Gaussian smoothed image, L, with the closest scale. Each image sample, $L(x, y)$, at this scale, the gradient magnitude, $m(x, y)$ and orientation $\theta(x, y)$ is pre-computed using pixel differences:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \tag{20}$$

$$\theta(x, y) = \tan^{-1}(L(x, y+1) - L(x, y-1) / (L(x+1, y) - L(x-1, y))) \tag{21}$$

An orientation histogram of 36 bins covering the 360° range of orientations is formed from the gradient orientation of sample points within a region around the keypoint. Additional samples added to the histogram is weighted by its gradient magnitude and by a Gaussian-weighted circular window with a σ that is 1.5 times that of the scale of the keypoint. All the weighted gradients for the descriptor are normalized to the main orientation of the circular region around the keypoint which is divided into 4×4 non-overlapping patches. The histogram gradient orientations

within the patches are computed and then histogram smoothing is performed to avoid sudden orientation changes and bin size reduction to eight bins to limit the descriptor's size results into a $4 \times 4 \times 8 = 128$ dimensional feature vector for each key-point. The feature vector is finally normalized to unit length and thresholded to reduce the effects of linear and non-linear illumination changes.

In nodule analysis framework, it is assumed that nodules have been already detected which correspond to interest/key points in Lowe's algorithm; hence, this step can be bypassed. In order to obtain a nodule SIFT descriptor which is invariant to orientation, a consistent orientation should be assigned to the detected nodule which is represented by its centroid, x_o. This orientation is based on the gradient of the nodule's local image patch. Considering a small window surrounding x_o, the gradient magnitude and orientation can be computed using finite differences. Local image patch orientation is then weighted by the corresponding magnitude and Gaussian window. Eventually the orientation is selected to be the peak of the weighted orientation histogram.

Building a nodule SIFT descriptor is similar to orientation assignment, for example a 16×16 image window surrounding the nodule centroid point is divided into sixteen 4×4 sub-windows, then an 8-bin weighted orientation histogram is computed for each sub-window, hence, $16 \times 8 = 128$ descriptors for each nodule is obtained. Thus, each detected nodule can now be defined at location (x_0, y_0), specific scale σ, explicit orientation θ and descriptor vector, $\mathrm{x}_o = \{\mathrm{x}_0, y_0, \sigma, \theta, \mathrm{d}\}$. Thus the SIFT operator $\mathcal{S} : \mathrm{I}(\mathrm{x}) \rightarrow \mathrm{X}$ can be viewed as mapping a CT slice I(x) to the nodule space with n-nodules, $\mathrm{X} = \{\mathrm{x_i}\}_{\mathrm{i}=1}^{\mathrm{n}}$ detected from I(x), where $\mathrm{x_i} = \left\{\mathrm{x}_0^{\mathrm{i}}, y_0^{\mathrm{i}}, \sigma_\mathrm{i}, \theta_\mathrm{i}, \mathrm{d_i}\right\}$. Principle component analysis (PCA) and linear discriminant analysis (LDA) are used to project the extracted SIFT descriptors to a low-dimensional subspace where noise is filtered out.

Example: Figure 16 shows four slices containing nodules <1 cm in size. The SIFT algorithm was applied to the four nodule types and the resulting discriptors were used to classify the nodules after a detection step, in order to reduce false positives. Small-size nodules lack textural distinction, but the shapes are distinct. Figure 17 shows the construction and values of the SIFT algorithm for the four nodule types. The values of the SIFT descriptor shows decent discrimination among the nodules.

5.2 Feature Distance Measures

The feature distance measurement is a numerical description of how far apart the feature vectors are from one another. Numerous methods found in the literature can be used; below are *described three different distance measurements.*

The Euclidean Distance: The Euclidian distance (ED) between feature point vectors ***p*** and ***t*** in the Euclidean n-space

$$\mathrm{ED} = \sqrt{\sum\nolimits_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{p_i} - \mathrm{t_i})^2} \tag{22}$$

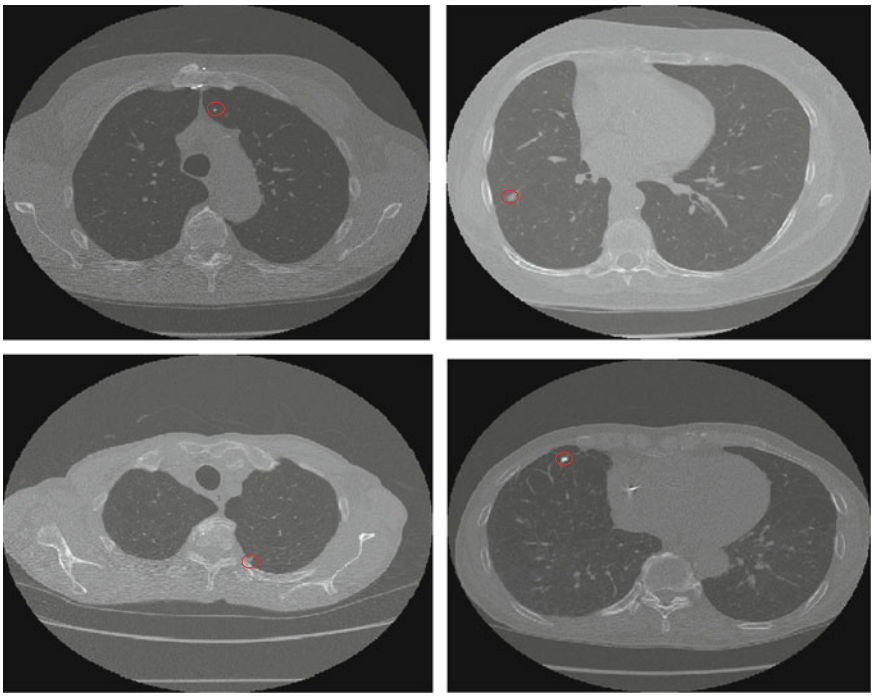


Fig. 16 Small-size lung nodules from LDCT scans. *Upper left* (well-circumscribed); *upper right* (vascular); *lower left* (juxta-pleural); *lower right* (pleural-tail). Nodules are marked by a *circle*

Note that the Euclidean distance is rotation invariant but not scale-invariant.

The Mahalanobis Distance: The Mahalanobis Distance is a scale-invariant distance measure based on correlations between variables by which variations can be identified for analysis. A multivariate vector $X = [x_1, x_2, \ldots, x_N]^T$ from a group of values with mean $\mu = [\mu_1, \mu_2, \ldots, \mu_N]^T$ and covariance matrix, S, is defined as:

$$D_M(\mathbf{X}) = \sqrt{(\mathbf{X} - \mu)^{\mathbf{T}} \mathbf{S}^{-\mathbf{1}} (\mathbf{X} - \mu)} \tag{23}$$

The Chebyshev Distance: This distance is a metric defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension. The distance between two vector points P and t with standard coordinates p_i and t_i is defined as:

$$D_{\text{Chebyshev}}(\mathbf{p}, \mathbf{t}) = max\ (|p_i - t_i|) \tag{24}$$

Evaluation of these distance measures for shape analysis exist elsewhere (e.g., [30, 31]).

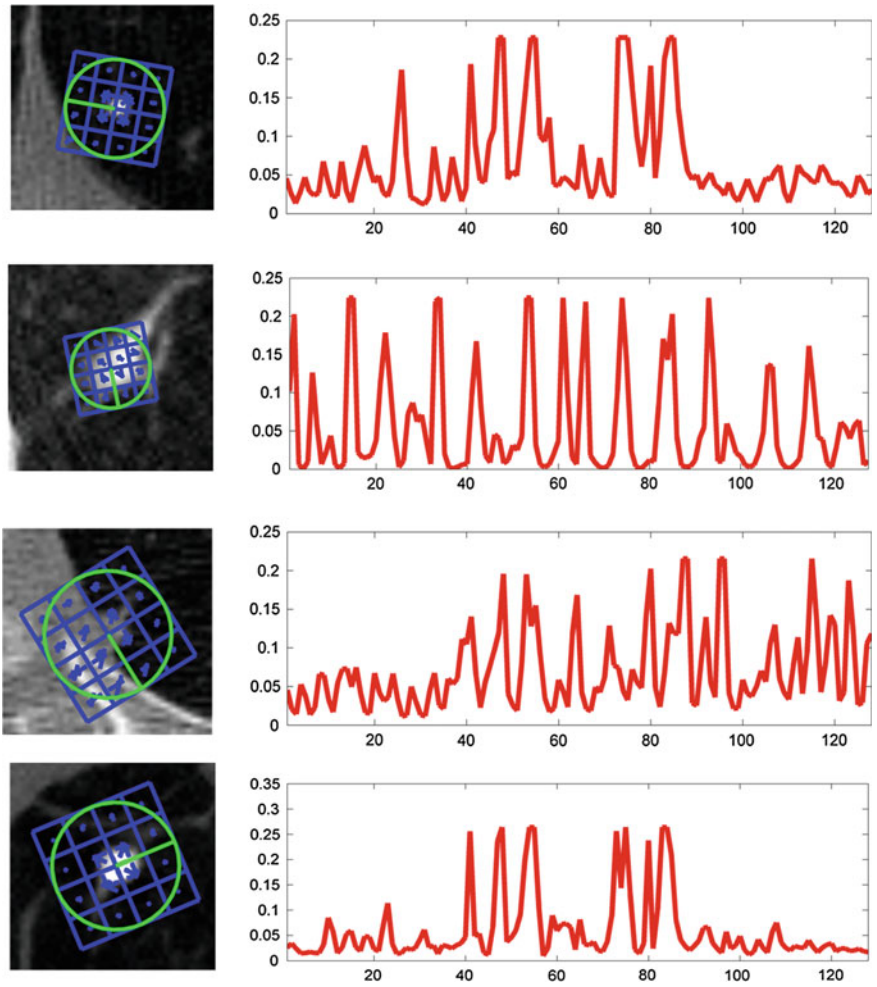


Fig. 17 SIFT desciptor applied to small-size nodule types in LDCT of the human chest. From top to bottom: well-cricumscribed, vascular, juxta-pleural and pleural-tails nodule types

5.3 Lung Nodule Classification

The above descriptors form the basis for the classification process to be examined in the following section.

5.3.1 General Approach

The general approach for nodule classification may be summarized by the following algorithm.

1. Construct a statistically sufficient database of pathological nodules;
2. Co-Register members of the nodule database to create the templates used for nodule detection, as described before;
3. Generate the feature vectors using the geometric descriptors (e.g., SIFT, ASFIT, SURF, LBP and Gabor Wavelet) for all members of the nodule database and store offline. Machine learning algorithms may be used such as PCA, RANSAC and Adaboost for optimal selections of the feature vector in terms of discrimination as well as execution time;
4. Perform the nodule detection using template matching;
5. Crop ROIs of sizes N $\times$ N over detected nodules (e.g., N=21)— these will be used for categorization;
6. Segment the nodule regions with the ROIs using the variational approach described in Chapter IV, enhanced with a priori information about shape and intensity, using the nodule database;
7. Repeat step #3 on the candidate nodules after segmentation; and
8. Calculate the distance between the feature vectors of candidate nodules and those in the pathological database, and assign the nodule category based on minimum distance.

The above algorithm may be carried out by various ways, depending the features available. Below feature-based and registartion-based nodule classifcation implemented in the author's recent work [23, 27] are discussed.

5.3.2 Feature-Based Classification

The most significant classification results were obtained when the shape based signed distance transform was combined to the texture based LBP approach. The results in Tables 1, 2, 3 illustrate the classification results of the signed distance transform versus the multi-resolution local binary pattern (LBP). A third feature descriptor using the combination of the methods is also shown.

Higher true-positive rates can be seen from the LDA projection in Tables 1 and 2 when more training is conducted using either the LBP or distance descriptors separately. When comparing the PCA results less training data resulted in better true-positive classification of nodules. In the non-nodule distance transform experimentations more training data was needed to obtain in some instances perfect results. This is understandable since the non-nodules do not have specific shape characteristics that can be defined or manipulated as in the nodules case.

Overall, the PCA combinational shape and feature description of nodules resulted in a drastic true-positive rate increase in classification. All of the results depicted in Tables 1 and 2 allow the conclusion to be made that non-nodules do in-fact contain

Table 1 Classification results for various nodules using raw LBP, LDA LBP and PCA LBP with variable training percentages

Nodule type	Raw LBP				LDA LBP				PCA LBP			
	100 %	75 %	50 %	25 %	100 %	75 %	50 %	25 %	100 %	75 %	50 %	25 %
Juxta pleural	52	50	47	38	100	86	65	50	64	64	59	67
Well-circumscribed	40	41	40	26	65	80	63	36	64	60	66	82
Vascular	22	29	32	10	32	76	56	32	20	22	37	56
Pleural tail	22	20	17	11	100	76	52	39	33	17	33	46
Non nodule	78	77	74	68	100	88	60	44	86	87	83	96

Table 2 Classification results for various nodules using raw distance transform, LDA LBP and PCA distance transform with variable training percentages

Nodule type	Raw distance transform				LDA distance transform				PCA distance transform			
	100 %	75 %	50 %	25 %	100 %	75 %	50 %	25 %	100 %	75 %	50 %	25 %
Juxta pleural	38	39	35	34	100	88	61	45	62	54	60	68
Well-circumscribed	33	33	36	34	74	83	63	45	46	59	48	55
Vascular	12	12	15	15	29	76	54	29	37	22	61	63
Pleural tail	17	17	17	15	100	85	54	33	17	24	35	52
Non nodule	63	68	68	49	100	87	65	49	83	89	85	79

Table 3 Classification results obtained from raw combinational feature transform and PCA combinational feature transform with variable training percentages

Nodule Type	Raw combinational feature descriptor				PCA on combinational feature descriptor			
	100 %	75 %	50 %	25 %	100 %	75 %	50 %	25 %
Juxta Pleural	40	41	39	37	78	76	76	79
Well-circumscribed	40	37	36	34	73	68	71	68
Vascular	24	20	22	12	51	54	44	76
Pleural tail	22	26	22	20	33	35	41	54
Non nodule	63	57	58	49	100	99	100	98

descriptor variations that allow them to be correctly classified. Also, combination of shape and texture feature information allows for better object representation to be obtained, thus improved results in classification.

Table 3 depicts impressive results when the LBP was obtained from the distance transform images. A 20 % true-positive rate increase was found, in the PCA 25 % training combinational vascular nodule case when comparing it to the PCA LBP results obtained when only the texture information was used for classification, and a 13 % increase over the distance transform results alone. Variations of percentage increases were seen for each nodule category.

5.3.3 Registration-Based Classification

The idea of the registartion-based classification is to compare the segmented nodules with nodule models, using a registration algorithm. Since the AAM approach generated impressive nodule models, which resembled both the shape and appreance of real nodules, it is plausible to use the normalized nodule models as templates to compare with candidate nodules for classifcation. From the face recognition analogy, a probe (test face) is compared to a gallary using either direct matching (by registration) or through the use of features.

The following terminologies are relavant to the categorization process:

(i) *Target* set $\mathcal{T}$: a set of textured regions containing the nodule models generated by the deformable model approach for all nodule types.
(ii) *Gallery* set $\mathcal{G}$: a subset of $\mathcal{T}$ containing template(s) to be matched in a certain matching setup.
(iii) *Query* set $\mathcal{Q}$: a set of textured regions of unknown nodule type, where nodule type identification is performed by matching all elements in the query set to the target set.
(iv) *Probe* set $\mathcal{P}_{\mathcal{G}}$:a subset of $\mathcal{Q}$, where each element has a match in the gallery set.
(v) *Imposter* set $\mathcal{P}_{\mathcal{N}}$: a subset of $\mathcal{Q}$, which contains elements that don't have a match in the gallery set.

As an example, again, using the face recognition terminology, a region centered at a well circumscribed nodule is considered an imposter to a gallery containing only juxta pleural nodules. Also a non-nodule region is always considered as an imposter. Comparing the feature vector for all nodule models in the gallery set with the feature vector for all regions in the probe set results in a similarity matrix $\mathcal{S}$, where the ijth element is the similarity between the ith element of the gallery and the jth element of the probe. The following metrics can be defined according to a similarity score: Normalized cross-corerlation (NCC), the mutual information (MI) or the output of descriptors such as SFIT, LBP, etc.

Identification Rate/Probability: It is calculated as the proportion of testing nodules correctly matched to its own type, i.e. probe $\mathrm{P_j}$ is identified correctly in the top N gallery nodule types, where $\mathrm{N} = \mathrm{renk}(\mathrm{P_j})$, such that:

$$\mathrm{rank}\left(\mathrm{p_j}\right) = \left|\left\{\mathrm{g_k} : \mathrm{s_{kj}} \geq \mathrm{s_{ij}},\ \mathrm{id}\left(\mathrm{g_i}\right) = \mathrm{id}\left(\mathrm{p_j}\right)\right\}\right| \qquad \forall \mathrm{g_k}, \mathrm{g_i} \in \mathcal{G} \tag{25}$$

For each probe in the probe set , the similarity measures are sorted against the gallery, and obtain the rank of the match. Identification performance is then stated as the fraction of probes whose gallery match is at rank or lower. Thus the probability of identification at specific rank is defined as:

$$\mathrm{P_I}\left(\mathrm{r}\right) = \frac{\left|\left\{\mathrm{p_j} : \mathrm{rank}\left(\mathrm{p_j}\right) \leq \mathrm{r}\right\}\right|}{\left|\mathcal{P}_{\mathcal{G}}\right|} \qquad \forall \mathrm{p_j} \in \mathcal{P}_{\mathcal{G}} \tag{26}$$

Table 4 Results of the nodule categorization using registration/matching nodule candidates to nodule models

Nodule model	Nodule and background				Nodule region segmented			
	Rank 1	Rank 2	Rank 3	Rank 4	Rank 1	Rank 2	Rank 3	Rank 4
Juxtal pleural	0.4606	0.9217	0.9826	1.0	0.4261	0.9217	0.9826	1.0
Well-circumscribed	0.764	0.7978	0.8427	1.0	0.8876	0.9663	0.9775	1.0
Vascularized	0.4146	1.0	1.0	1.0	0.5122	1.0	1.0	1.0
Pleural tail	0.3261	0.7609	0.8261	1.0	0.3913	0.5217	0.5435	1.0

These quantities have been calculated for all the nodules in the ELCAP data. Table 4 shows the results for the four nodule categories.

In measuring the ranking, the cropped nodules are used in two fashion; without segmentation (i.e., no extraction of the nodule part in the cropped region) and with segmentation. The segmentation of nodules were conducted by various homegrown methods (including use of shape and intensity priors in an energy model optimized by graph cuts; also experimented with were basic segmentation using adaptive thresholding of the cropped regions by median filtering and anisotropic diffusion filtering, etc.).

Figure 18 is the ROC for 291 nodules specified in the ELCAP dataset. Both the well-circumscribed and the viscularized nodules provide the best performance. This is because both nodule types possess the best texture and shape information that enhances the correlation between the nodules and the models.

In general, the results of the ranking (i.e., matching models with nodules) improved by segmentation of the nodule portion in the cropped region. Model-based approaches such level sets and combinations of Gibbs-Markov models enhance the segmentation at severe computation cost. Nodule segmentation is a work in progress issue. A code or signature for the models and the nodules will provide better matching than using the classic image registration methods on regions with small spatial support. The conclusion, however is that the cropped regions have always been correctly categorized within second ranks by a simple computational approach such the normalized cross-correlation. This indeed is very encouraging for moving into using context based image processing and the ability to invoke advanced machine learning approaches to perform the matching process.

The extensive analysis using the approaches described in this chapter has allowed several conclusions to be made:

(1) Texture and shape feature information separately are not sufficient for lung nodule categorization, since the combination of the approaches yielded great improvements.
(2) In all of the approaches used, the non-nodule features generated and projected by PCA or LDA provided excellent classification results; thus, non-nodules contain descriptor variations that allow them to be correctly classified and not confused with nodules.

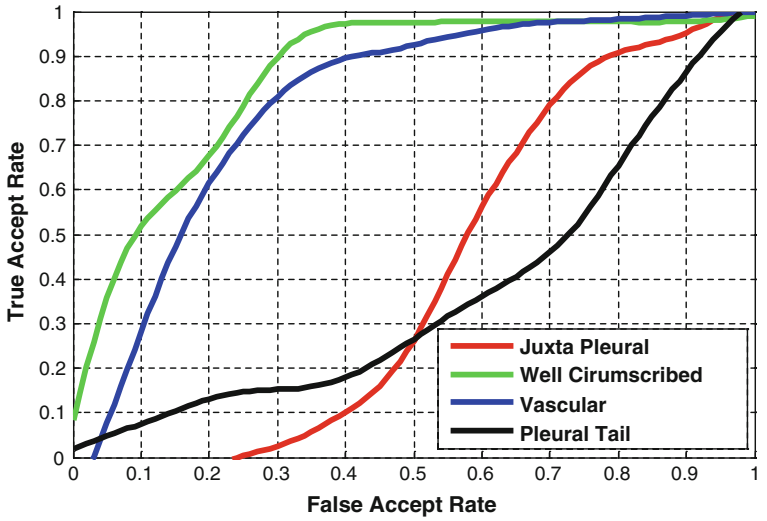


Fig. 18 ROC of automatic categorization on the ELCAP data. Well-circumscribed and vascular nodule types possess the best ranking for automatic categorization

(3) Intensity-based registration methods did not provide accurate categorization of small objects; a more appropriate similarity measures may be needed for these types of objects.
(4) Signatures of nodules—based on multiple approaches—may be generated and used for categorization; similar to face recognition methods. However, more extensive annotated databases of nodules are needed.

5.4 Summary

In this chapter, a system for nodule candidate detection and classification was created to show the robustness and accuracy of the produced models. Detection using a template matching method with normalized cross-correlation similarity measure without false positive reduction was implemented to show the robustness of the data-driven templates formulated from the AAM and ASM approaches over the known parametric template generation. Detection using the data driven template matching approach, after false positive reduction via SIFT and LBP feature extraction, was also implemented, further enhancing the detection process.

Classification of the nodules and non-nodules were examined using a k-NN leave-one-out algorithm with the Euclidean distance as the similarity measure, in order to test whether or not significant distinctions between the nodule classes exist. An overall 12% true-positive rate increase was found in the PCA combinational classification results over using the PCA LBP or the PCA distance transform separately. Various extensions and detailed analysis of biomedical imaging can be found in Farag [31].

References

1. Kostis WJ et al (2004) Small pulmonary nodules: reproducibility of three-dimensional volumetric measurement and estimation of time to follow-up. Radiology 231:446–52
2. Farag AA (2009) Lung nodule modeling and detection for computerized image analysis of low dose CT imaging of the chest. Master of Engineering, University of Louisville
3. ELCAP public lung image database. www.via.cornell.edu/databases/lungdb.html
4. Armato G, McLennan G, McNitt-Gray MF, Meyer CR, Yankelevitz D, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, Reeves AP, Croft BY, Clarke LP (2004) Lung image database consortium: developing a resource for the medical imaging research community. Radiology 232(3):739–748
5. Kendall DG (1984) Shape manifolds, procrustean metrics, and complex projective spaces. Bull Lond Math Soc 16(2):81–121
6. Julesz B et al (1973) Inability of humans to discriminate between visual textures that agree in second-order statistics-revisited. Perception 2:391–405
7. Dryden IL, Mardia KV (1998) Statistical shape analysis. John Wiley & Sons, New York
8. Cootes TF, Taylor CJ (2004) Anatomical Statistical Models and their role in feature extraction. Br J Radiol 77:S133–S139
9. Mathews I, Baker S (2004) Active appearance models revisited. Int J Comput Vision 60(2):135–164
10. Farag AA, Graham J, Farag AA, Elshazly S, Falk R (2010) Parametric and non-parametric nodule models: design and Evaluation. Workshop on pulmonary image processing in conjunction with MICCAI-10, Beijing, pp 151–162
11. Farag AA, Abdelmunim H, Graham J, Farag AA, Carter C, Elshazly S, El-Mogy S, El-Mogy MS, Falk R (2012) An AAM-based approach for lung nodule detection from low dose CT (LDCT) scans, International symposium on biomedical imaging (ISBI 12), Barcelona, Spain, pp 1040–1043
12. Farag AA (2013) A variational approach for small-size lung Nodule segmentation. International symposium on biomedical imaging (ISBI 13), San Francisco, CA, pp 81–84
13. Zhao B, Yankelevitz D, Reeves A, Henschke C (1999) Two-dimensional multi-criterion segmentation of pulmonary nodules on helical CT images. Med Phys 26(6):889–95
14. Ko JP, Rusinek H, Jacobs EL, Babb JS, Betke M, McGuinness G, Naidich DP (2003) Small pulmonary nodules: volume measurement at chest CT-phantom study. Radiology 228(3):864–870
15. Kuhnigk JM, Dicken V, Bornemann L, Bakai A, Wormanns D, Krass S, Peitgen HO (2006) Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans. IEEE Trans Med Imaging 25(4):417434
16. Kubota T, Jerebko AK, Dewan M, Salganicoff M, Krishnan A (2011) Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models. Med Image Anal 15(1):133–154
17. Dijia Wu, Le Lu, Jinbo Bi, Shinagawa Y, Boyer KL, Krishnan A, Salganicoff1 M (2010) Stratified learning of local anatomical context for lung nodules in ct images. The twenty-third IEEE conference on computer vision and pattern recognition, CVPR 2010, San Francisco, California, USA
18. Armato SG 3rd, Giger ML, Moran CJ, Blackburn JT, Doi K (1999) MacMahon H (1999) Computerized detection of pulmonary nodules on CT scans. J Radio Graph 19:1303–1311
19. Diciotti Stefano, Lombardo Simone, Falchini Massimo, Picozzi Giulia, Mascalchi Mario (2011) Automated segmentation refinement of small lung nodules in CT scans by local shape analysis. IEEE Trans Biomed Eng 58(12):3418–3428
20. Diciotti Stefano, Picozzi Giulia, Falchini Massimo, Mascalchi Mario, Villari Natale, Valli Guido (2008) 3-D segmentation algorithm of small lung nodules in spiral CT images. IEEE Trans Inf Technol Biomed 12(1):7–19

21. Munim HAE, Farag AA (2007) Curve/surface representation and evolution using vector level sets with application to the shape-based segmentation problem. IEEE Trans Pattern Anal Mach Intell 29(6):945–958
22. Farag A, Graham J, Farag A (2010) Robust segmentation of lung tissue in chest CT scanning. Proceedings of the 2010 IEEE international conference on image processing (ICIP), pp 2249–2252
23. Farag AA (2012) Modeling small objects under uncertainties: novel algorithms and applications. CVIP Lab, Ph.D. Dissertation, University of Louisville
24. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mac Intell 24:971–987
25. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
26. Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mac Intell 27(10):1615–1630
27. Farag AA, Elhabian S, Graham J, Farag AA, Falk R (2010) Toward precise pulmonary nodule descriptors for nodule type classification. Proceedings of the 13th international conference on medical image computing and computer assisted intervention (MICCAI), pp 626–633
28. Koenderink JJ (1984) The structure of images. Biol Cybern 50:363–396
29. Lindeberg T (1993) Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. Int J Comput Vis 11(3):283–318
30. Duda RO, Hart PE, Stork DG (2001) Pattern classification, 2nd edn. Wiley, New York
31. Farag Aly (2013) Biomedical image analysis: statistical and variational approaches. Cambridge University Press, Cambridge

Analyzing the Shape and Motion of the Lungs and Heart in Dynamic Pulmonary Imaging

Jianming Liang, Tim McInerney and Demetri Terzopoulos

Abstract The "Dynamic Chest Image Analysis" project aims to show focal and general abnormalities of lung ventilation and perfusion based on a sequence of digital chest fluoroscopy frames. An indispensable source of information recorded in the image sequence is the shapes and motions of the lungs and heart. This chapter employs this shape and motion information of lung and heart to detect abnormalities in both lung ventilation and perfusion. To extract the shape and motion information of lung and heart, we utilize a technique, called United Snakes, in which both the shape and motion of the lung and heart can be modeled using a single consistent theoretical and implementational framework. Along with case studies, we demonstrate the capability of United Snakes through four applications: lung registration, diaphragm motion analysis, cardiac motion analysis, and cardiac shape analysis, in revealing both lung ventilation and perfusion abnormalities.

1 Introduction

The respiratory system facilitates the exchange of gases (O_2 and CO_2) between the blood and ambient air; therefore, adequate pulmonary ventilation (air flow) and perfusion (blood flow) are essential for the lungs to function properly. Inadequate pulmonary function may be due to failure in ventilation and perfusion, among other

J. Liang (✉)
Department of Biomedical Informatics, Arizona State University, Scottsdale , AZ 85254, USA
e-mail: jianming.liang@computer.org

T. McInerney
Department of Computer Science, Ryerson University, Toronto, ON, Canada
e-mail: tmcinern@scs.ryerson.ca

D. Terzopoulos
Computer Science Department, University of California, Los Angeles, CA 90095, USA
e-mail: dt@cs.ucla.edu

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_9,

factors. To detect abnormalities of lung ventilation and perfusion, ventilation and perfusion isotope scans are conventionally used, but they can only provide a static, coarse 2D distribution of air and blood in the lungs, and also have a disadvantage of using radioactive isotopes. The primary imaging modality for diagnosing pulmonary disorders is chest X-ray, but the information about pulmonary function (ventilation and perfusion) that may be gleaned from a single chest X-ray is rather limited. To overcome this limitation, this chapter utilizes sequences of digital chest fluoroscopy frames to reveal focal and general pulmonary functional abnormalities by analyzing shape and motion of the lungs and heart.

2 Dynamic Pulmonary Imaging

2.1 Patient Examination

With Dynamic Pulmonary Imaging [1], we can collect a sequence of chest X-ray images of up to 512 × 512 pixels at a sampling frequency of 25 Hz with a copper filter of 3 mm. The reason of using a copper filter is to reduce the radiation dose to patients. Two separate examination procedures are used for ventilation and perfusion studies. In the ventilation study, the patient is asked to breathe naturally and normally in a supine position with posteroanterior projection. An image sequence of 55 frames with 192 × 144 pixels is collected in 4.32 s with a sampling frequency of 12.5 Hz, because in most cases the lungs can complete a full ventilation cycle in 4 s. Based on our experiments, a spatial resolution of 192 × 144 is sufficient for ventilation analysis. In the perfusion study, the patient is also in a supine position with posteroanterior projection, but with breath held to effectively remove the ventilation effects. An intravenous bolus of X-ray contrast medium may be further used to enhance the perfusion signal strength. Comparing with ventilation, perfusion has a higher frequency, thus requiring a higher temporal sampling rate but a shorter examination time. Furthermore, pulmonary perfusion is asynchronous,[1] demanding a higher spatial resolution. As a result, for perfusion analysis we acquire an image sequence of 52 frames with 384 × 288 pixels at a sampling frequency of 25 Hz in 2.04 s. The imaging parameters are summarized in Table 1.

The acquired image sequences may be represented with intensity function $I(x, y, t)$, where $0 \leq I \leq 255, 1 \leq x \leq$ width (192 for ventilation and 384 for perfusion), $1 \leq y \leq$ height (144 for ventilation and 288 for perfusion), and t is a discrete time point in $0 \leq t \leq$ *examtime* (4.32 s for ventilation and 2.04 s for perfusion). We may also represent it as $I(x, y, i)$, with i the frame index, such that $t = (i-1)/f$, where f is the sampling frequency of 12.5 Hz for ventilation analysis and 25 Hz for perfusion analysis.

[1] The speed of blood flow is roughly 10 cm/s. When the blood flows in the lungs, the phase (i.e., timeshifts) of a pulse signal at one location may be different from that at another location, although they have the same pulse frequency.

Table 1 Dynamic pulmonary imaging parameters used for ventilation and perfusion examinations

Examination	Image size	Temporal sampling frequency (Hz)	Number of frames	Examination time (s)
Ventilation	$\geq 192 \times 144$	12.5	55	4.32
Perfusion	$\geq 384 \times 288$	25	52	2.04

Because of the very short examination time and the use of a copper filter, the radiation dose to the patient is low. The entrance skin dose of a patient is about 0.1 to 0.2 mGy [1]. For comparison, the radiation dose of a normal chest X-ray image varies between 0.1 and 0.2 mGy, and the radiation dose of fluoroscopy is about 2 mGy per minute [1].

2.2 Ventilation and Perfusion Analysis

The 2D image sequence obtained from the patient examination carries valuable information for ventilation and perfusion studies thanks to the physical properties of X-rays: The attenuation of X-rays in air is much lower than in blood and soft tissue. As a result, the average pixel intensity of an area in the lung field varies over time due to the respiratory and cardiac cycles; this variation, called a lung functional signal, reflects the air and blood volume change in the corresponding 2D projectional area of the lung when the patient breathes naturally. When the patient is asked to hold their breath, we observe the perfusion signal disturbed by noise. The ventilation intensity variation depends on the depth of the tidal volume ventilation and on lung area. It is usually between 5 and 15 units in the 8-bit grey scale. The image intensity variation for perfusion is about 3 to 4 units without contrast media. The ventilation signal to noise ratio is about 10:1 and the perfusion signal to noise ratio is about 3:1. This phenomenon is illustrated in Figs. 1, 2 and 3.

We detect ventilation and perfusion abnormalities by extracting meaningful ventilation and perfusion parameters from the lung functional signals. To do so, it is necessary to accurately locate the "turning points" from the signal, but it is challenging due to the existence of both ventilation and perfusion components, in addition to noise. Furthermore, a phase (exhalation, inhalation, diastole, or systole) might not be complete in a signal. For instance, the signal in Fig. 1b does not have a complete exhaling phase. To this end, we introduce a mathematical function (Fig. 4):

$$M(A, D, U, S, L, t) = \begin{cases} A\cos(\pi t'/D) + L & \text{if } 0 \leq t' < D \\ A\cos(\pi(t' - D)/U + \pi) + L & \text{if } D \leq t' < (D + U) \end{cases} \quad (1)$$

where

$$t' = (t - S) \mod (D + U), \quad (2)$$

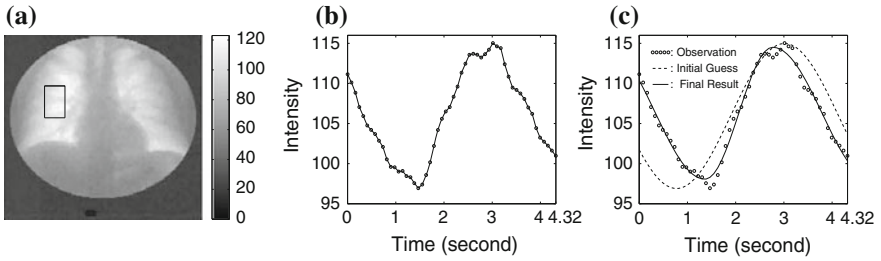


Fig. 1 A case in quiet breath. **a** A region of interest (ROI) in the right lung field and **b** its corresponding lung functional signal (observation), which reflects the air and blood change in the corresponding lung area over time during the examination due to the physical properties of X-rays. **c** A set of ventilation parameters can be extracted from the observation **b** with a ventilation model (see Fig. 4) via optimization, where the observation is indicated with an "o", the initial guess is plotted as a dashed curve and the final solution as a solid curve. The image gets whiter (higher intensity) during inhalation (more air in the lungs). The ROI shown here is a rectangle, but it may be of arbitrary shape. The ROI may be as large as a whole lung or may be as small as a single pixel

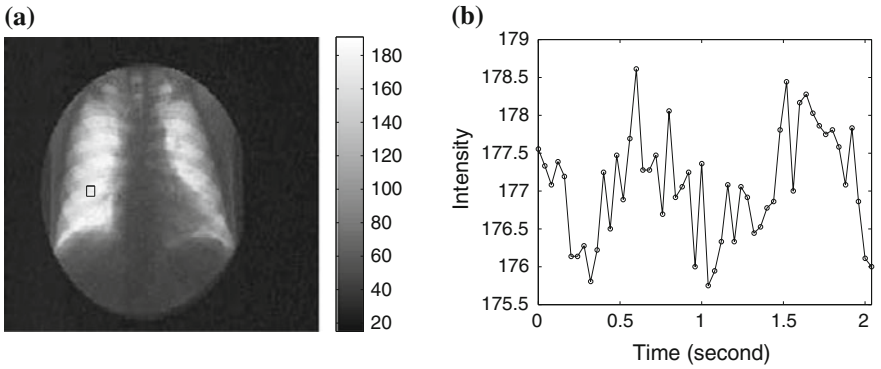


Fig. 2 A case with the breath held and an intravenous bolus of X-ray contrast media. **a** An ROI (region of interest) in the right lung field and **b** its corresponding observation—an enhanced lung perfusion signal which (due to the physical properties of X-rays) reflects the blood flow in the corresponding lung area with contrast media. The image gets darker (lower intensity) during the systolic phase (more blood in the lungs). Comparing to ventilation in Fig. 1, the perfusion signal is very noisy and weak (only about 3 intensity-unit variation)

with t for $time$, and $t' \in [0, D + U)$, so that ventilation and perfusion parameters can be extracted automatically from lung functional signals via optimization [2–4]. From these five extracted parameters, more parameters can be derived. In case of ventilation, we can compute:

- *Ventilation Frequency* ($\hat{F}_v$) (expressed as the number of breaths per minute):

$$\hat{F}_v = 60/(\hat{D}_v + \hat{U}_v), \tag{3}$$

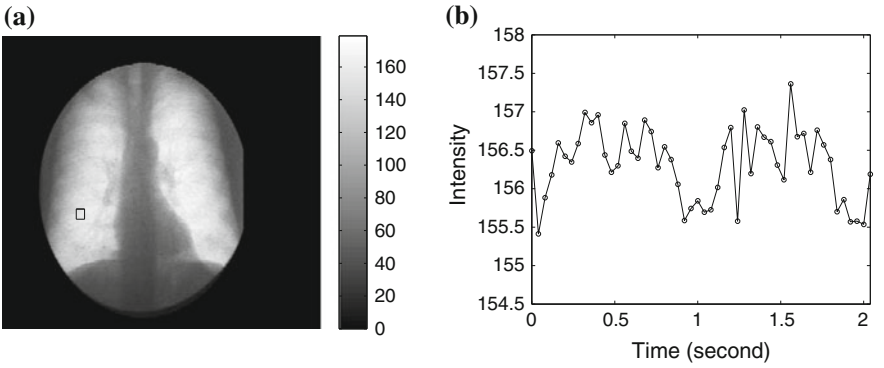


Fig. 3 A case with the breath held but using no X-ray contrast media. **a** An ROI (region of interest) in the right lung and **b** its corresponding observation—a perfusion signal reflecting the blood flow in the lung area due to the physical properties of X-rays. It is plotted in the same scale as in Fig. 2 for comparison

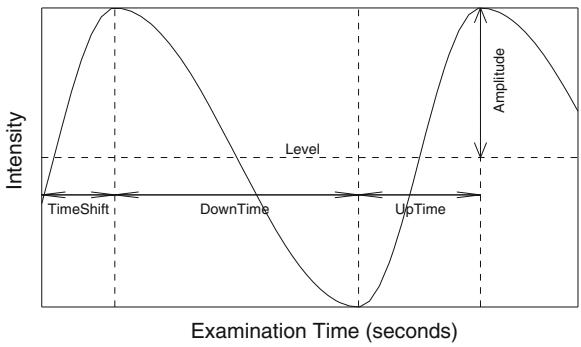


Fig. 4 A mathematical function can be used for both ventolation and perfusion analysis. In case of ventilation, it models the volume change of air during lung ventilation (increasing during inhalation and decreasing during exhalation) with five primitive parameters: *amplitude A* (ventilation strength), *downtime D* (time for exhalation), *uptime U* (time for inhalation), *timeshift S* (time from the starting of examination to the completion of the first inhalation) and *level L* (mean intensity; its value depending on various factors, medical meaning un-defined). In case of perfusion, it models the blood volume change, increasing during the diastolic phase and decreasing during the systolic phase, but its five free primitive parameters have completely different medical meanings: *amplitude A* (perfusion strength in the lung area), *downtime D* (time for the systolic phase in the lung area), *uptime U* (time for the diastolic phase in the lung area), *timeshift S* (time from the first image to the completion of the first diastolic phase) and *level L* (the mean intensity but with no well-defined medical meaning)

- *Inhaling Rate* ($\hat{I}_v$):

$$\hat{I}_v = \hat{A}_v / \hat{U}_v, \tag{4}$$

- *Exhaling Rate* ($\hat{E}_v$):

$$\hat{E}_v = \hat{A}_v / \hat{D}_v, \tag{5}$$

- *Normalized timeshift* ($\hat{H}_v$):

$$\hat{H}_v = \hat{S}_v / (\hat{U}_v + \hat{D}_v), \tag{6}$$

- *Updown Ratio* ($\hat{R}_v$):

$$\hat{R}_v = \log_{10}(\hat{U}_v / \hat{D}_v). \tag{7}$$

We take the logarithm of the "updown ratio" to make it symmetric. All these parameters may be used as ventilation abnormality indicators, but our experiments show that in most cases three parameters, $\hat{A}_v$ (amplitude), $\hat{H}_v$ (normalized timeshift) and $\hat{R}_v$ (updown ratio) are sufficient in revealing ventilation abnormalities.

2.3 Shape and Motion Analysis

An additional, indispensable source of information recorded in the image sequence is the shape and motion of the lungs and heart. This chapter employs this shape and motion information to detect abnormalities of the lungs and heart with the United Snakes technique, which is to be reviewed in Sect. 3. It should be noted that the tasks of lung registration and cardiac motion analysis are challenging, because of the reduced image contrast by the copper filter used in image acquisition to reduce the radiation dose to patients.

3 United Snakes

A snake [5] is a flexible, elastic contour whose behavior is governed by an energy functional, where an internal energy controls the degree of stretchiness and flexibility of the contour while an external energy couples the contour to an image, attracting the snake to features of interest (e.g., intensity edges). The active research in Snakes has resulted in a large family of Snakes algorithms [6–8], including finite element Snakes, B-Snakes, and Fourier Snakes, and related algorithms, such as "live-wire" (also known as "intelligent scissors") [9–15]. Each of these variants has its strengths and weaknesses.

To extract and model the shape and motion of both the lung and the heart in an accurate and robust manner, the differences between these two organs must be taken into consideration and the most appropriate Snake algorithm selected. For example, the lung boundary is smooth with readily identifiable curved corner regions. Consequently, a Hermite finite-element Snake, which can be constructed directly from user-defined lung boundary points and which can easily control the relative position of its nodal points, is most suitable for lung registration and motion analysis. On the other hand, much of the heart boundary is not visible in the image sequence

and there are no readily identifiable landmark points directly on the boundary. In this case, the reduced number of degrees of freedom, high degree of smoothness and control polygon of a B-spline Snake make it an ideal choice. Furthermore, the live-wire algorithm is effective for providing a quick delineation of the lungs and heart and this delineation can then be used by a Snakes algorithm for subsequent segmentation and motion tracking of an image sequence. As a result, a common framework combining the best features of the various Snakes algorithms and the live-wire algorithm is highly desirable.

To this end, United Snakes unifies various Snakes algorithms in a finite element framework, where a particular type of Snake can be derived simply by changing the shape functions at the user level. This unification expands the range of object modeling capabilities within a uniform Snake construction process and provides a uniform Snakes motion tracking mechanism. Consequently, both the shape and motion of the lung and heart can be modeled using a single consistent theoretical and implementational framework. United Snakes is also advantageously combined with live-wire by introducing an effective hard constraint mechanism. The United Snakes framework amplifies the efficiency and reproducibility of the component techniques, and it offers more flexible interactive control while further minimizing user interactions. The reader is referred to [16] for the mathematical details.

In the following sections, along with case studies, we present four applications of United Snakes: lung registration, diaphragm motion analysis, cardiac motion analysis, and cardiac shape analysis.

4 Lung Registration

Through our clinical studies, our expert radiologists have found it convenient and effective to use four rectangles (regions of interests, ROIs) covering the apex, upper, middle and lower lung field in each lung for a quick ventilation examination by inspecting the behaviors of the four corresponding lung functional signals. To facilitate this inspection, we propose an ROI-based analysis with lung registration and division.

4.1 Quick ROI-Based Analysis with Lung Registration and Division

The rapid ROI-based analysis is performed by first interactively delineating the lungs in the first frame with United Snakes (see Fig. 5), and then using the tracking capabilities of Snakes to automatically follow the motion throughout the entire image sequence (Fig. 6). The result is a lung delineation in each frame of the sequence. This step is followed by an automatic division of each lung field into four rectangular regions in each frame (Fig. 7) and an automatic calculation of the average intensity for each region. This process forms four lung functional signals in each lung field, from which ventilation parameters can be automatically extracted as shown in Fig. 8.

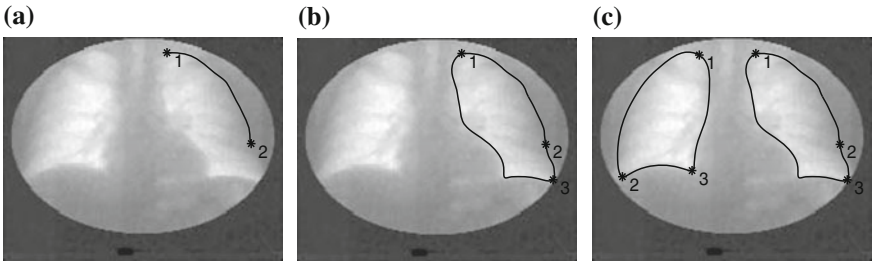


Fig. 5 Lung delineation with United Snakes. Once the first seed (point *1* in (**a**)) is placed on the lung boundary, an interactive snake is automatically constructed from the first seed to the current mouse position (position *2* in (**a**)). With three seeds, the left lung can be delineated (**b**), and similarly for the right lung (**c**)

4.2 Clinical Case Studies

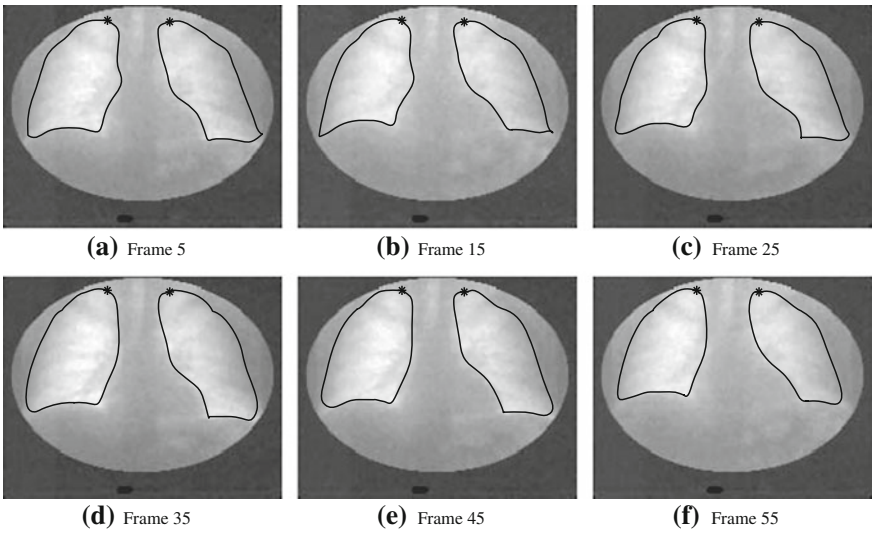
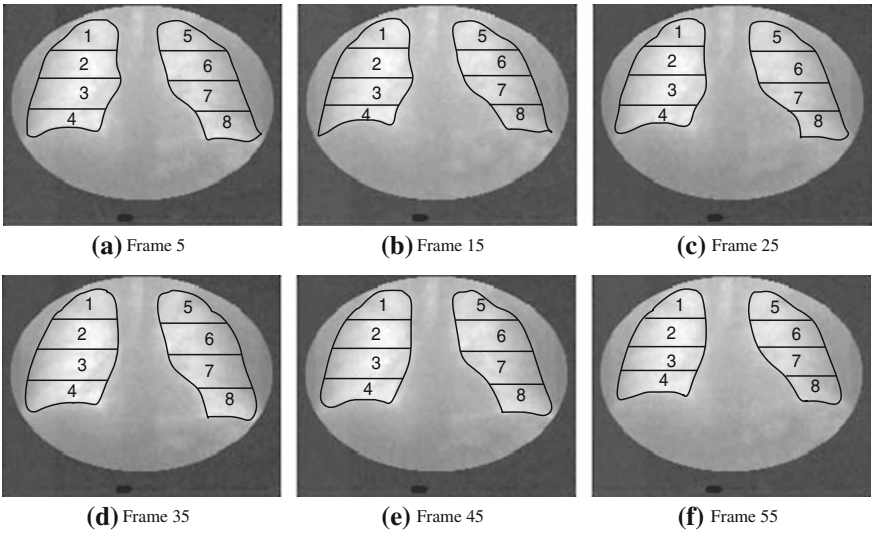


Fig. 6 The tracking result (every tenth image shown). The edge information at the lung apex is rather weak. Furthermore, there is no observable lung motion during quiet breathing. In order to make the snake firmly stick to the apex, it is desirable to maintain a hard constraint point there. Therefore, in the lung delineation as illustrated in Fig. 5, the first seed is usually placed at the lung apex so that it can be utilized as a hard constraint in the tracking process

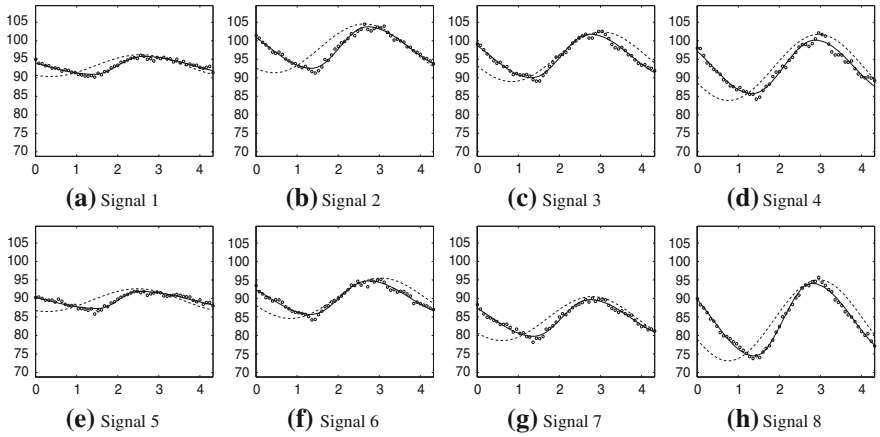
Here we present two representative cases—one normal and one abnormal—to illustrate typical abnormalities:



(a) Frame 5 **(b)** Frame 15 **(c)** Frame 25

(d) Frame 35 **(e)** Frame 45 **(f)** Frame 55

Fig. 7 Each lung field is divided into four regions with equal heights in each frame so that four lung functional signals in each lung field can be formed. Ventilation parameters are automatically extracted from these signals (see Fig. 8 and Table 2)



(a) Signal 1 **(b)** Signal 2 **(c)** Signal 3 **(d)** Signal 4

(e) Signal 5 **(f)** Signal 6 **(g)** Signal 7 **(h)** Signal 8

Fig. 8 Extracting the ventilation parameters from the eight lung functional signals resulting from the lung division (the observations are indicated with "o", the initial guesses with dashed curves and the final results with solid curves). The numerical results are given in Table 2

- Case 1: This case has been clinically classified as normal, but we have found some slightly asynchronous ventilation and non-symmetrical ventilation at both apices (see Table 2).

Table 2 Clinical case I: Extracted ventilation parameters

Signal	A	H	R
1	2.43	0.72	−0.325
2	5.61	0.79	−0.186
3	5.87	0.79	−0.160
4	7.23	0.83	−0.099
5	2.32	0.70	−0.421
6	4.54	0.80	−0.174
7	4.95	0.81	−0.181
8	9.79	0.81	−0.169

Clinically normal but slightly asynchronous ventilation (ASV) and non-symmetrical ventilation (NSV) seen in regions 1 and 5

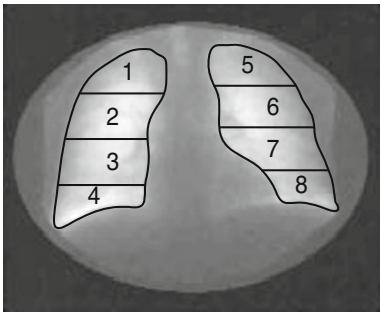


Fig. 9 Clinical case II with lung division

Table 3 Clinical case II: Parameter extraction. Reduced ventilation (RV) in regions 2, 3 and 7, and compensatory ventilation (CV) in regions 4 and 8

Signal	A	H	R
1	2.44	0.81	−0.129
2	2.98	0.74	−0.098
3	2.63	0.75	−0.133
4	15.94	0.76	−0.082
5	3.18	0.86	0.013
6	10.58	0.76	0.037
7	8.81	0.77	0.034
8	22.76	0.75	−0.023

It is abnormal that the updown ratios of the right lung are negative, while those of the left lung are almost all positive

- Case 2: A pathological case (see Fig. 9 and Table 3). By comparing this case with the previous normal case, it is clear that ventilation in regions 2 and 3 is poor. Ventilation amplitudes are expected to increase from region 1 to region 4 and from region 5 to region 8. Therefore, it is abnormal to observe an amplitude in region 7 smaller than region 6. It is also abnormal to observe negative updown ratios in the right lung and predominantly positive ratios in the left lung. Furthermore, regions

Table 4 Statistics of ventilation abnormalities with ROI-based analysis

Right/Left	NV	RV	ASV	NSV	CV
Apex(1/5)	0/0	4/5	25/19	23/17	0/1
Upper(2/6)	0/1	3/1	1/1	3/4	1/2
Mid(3/7)	2/0	9/4	5/3	11/3	0/1
Lower(4/8)	0/0	7/7	2/3	3/3	0/0

Table 5 Number of user interactions required to guarantee segmentation accuracy and robustness

	Seeds in first frame	User interactions in other frames
Max	5	2
Min	3	0
Average	3.3	0.3

4 and 8 are "hard" at work to compensate for the abnormal areas—the phenomena of "compensatory ventilation": an area with excessive ventilation in order to compensate for the abnormal areas in other parts of the lungs. This generally goes along with non-ventilation and reduced ventilation. The area itself should be considered as normal, but it is suggestive for abnormalities in other parts of the lungs. The HRCT report confirmed our findings in the right lung: bronchiectatic changes in the right middle and lower lobes, and scar changes in the right middle lobe medially. However, the report provided no explanation for the smaller amplitude of region 7 with respect to region 6.

The statistics of abnormal findings with our 53 ventilation patients are summarized in Tabel 4. Through the clinical cases studies, we have found that ROI-based ventilation analysis is efficient and effective with respect to the size of lung divisions. As the lung field is divided into finer regions, smaller ventilation irregularities are detected.

In the clinical studies, we also have found that United Snakes are not only efficient, providing real-time performance, but also accurate and require little user intervention. This performance is due, in part, to the use of the hard constraint mechanism offered by the United Snakes technique. The hard constraint at the lung apex plays a critical role by pinning the Snake—without it the snake would slide away from the lung apex. Furthermore, due to the interaction mechanisms provided by United Snakes, the segmentation and motion tracking can be made as accurate as desired by the user. The number of user interactions are given in Table 5, showing little user intervention is required.

5 Diaphragm Motion Analysis

The diaphragm is a dome-shaped sheet of muscle that separates the chest from the abdomen. As the diaphragm contracts and flattens, the volume of the chest increases and air is drawn into the lungs. As it relaxes, the dome pushes upward, forcing air out of the lungs. The diaphragm contracts without any voluntary control. During quiet breathing, there is no contraction of intercostal muscles. Therefore, analyzing the diaphragm motion gives first-hand information concerning pulmonary ventilation.

5.1 Quantifying Diaphragm Motion

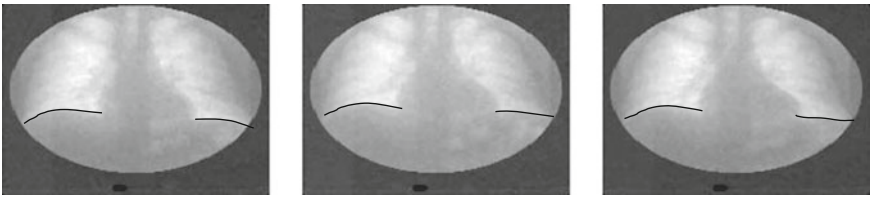
Although we can extract diaphragm motion from the lung delineation result in Sect. 4, a more efficient quantitative analysis can be obtained using United Snakes. In the current patient orientation, the diaphragm is restricted to an up-and-down motion. Consequently, we use an open Snake and restrict its motion along the y-axis, leading to efficient tracking as only a few Snake nodes are required and only one deformation direction is needed for each node. The diaphragm motion can be characterized using the average position of the Snake over time.

5.2 Clinical Case Studies

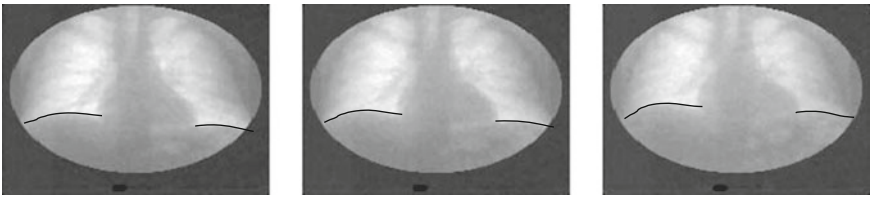
Three clinical cases are presented:

- Case 1 (Fig. 10): A clinically normal case. The diaphragm moves freely and continuously.
- Case 2 (Fig. 11): Local abnormal diaphragm motion. The right diaphragm is expected to contract even further during inhalation, but it exhibits no observable motion for about half a second. Air is not effectively drawn into the right lung.
- Case 3 (Fig. 12): Global diaphragm motion abnormality. The left diaphragm cannot move freely and exhibits further irregularities during exhalation.

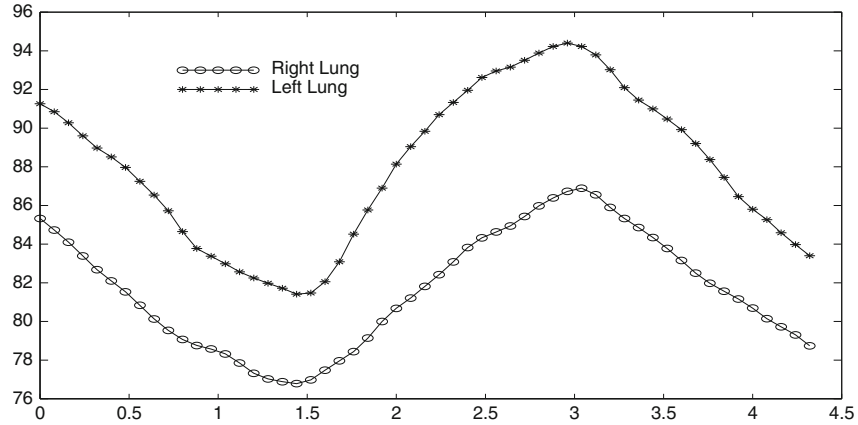
The number of user interactions needed for diaphragm motion analysis is minimal—only two seeds are required in the first frame to initialize a Snake, and no further user intervention is needed for any other frames due to the strong image contrast at the diaphragm. The abnormal diaphragm motion findings are summarized in Table 6.



(a) Frame 5 (t=0.32) **(b)** Frame 15 (t=1.12) **(c)** Frame 25 (t=1.92)



(d) Frame 35 (t=2.72) **(e)** Frame 45 (t=3.52) **(f)** Frame 55 (t=4.32)

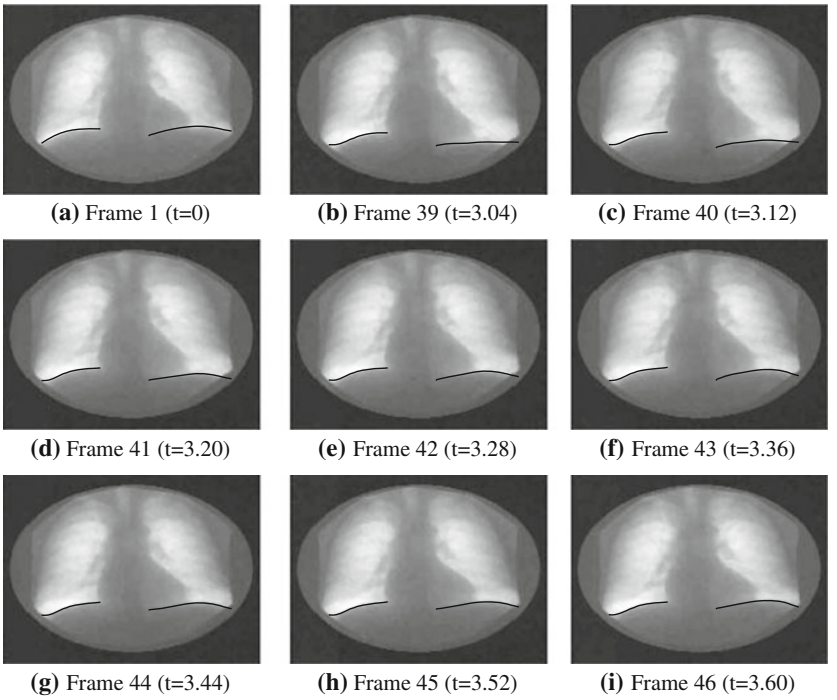


(g) Diaphragm Motion: Average y-positions of the two open snakes

Fig. 10 Clinical case I: Normal diaphragm motion. The diaphragm moves freely and continuously on both sides

6 Cardiac Motion Analysis

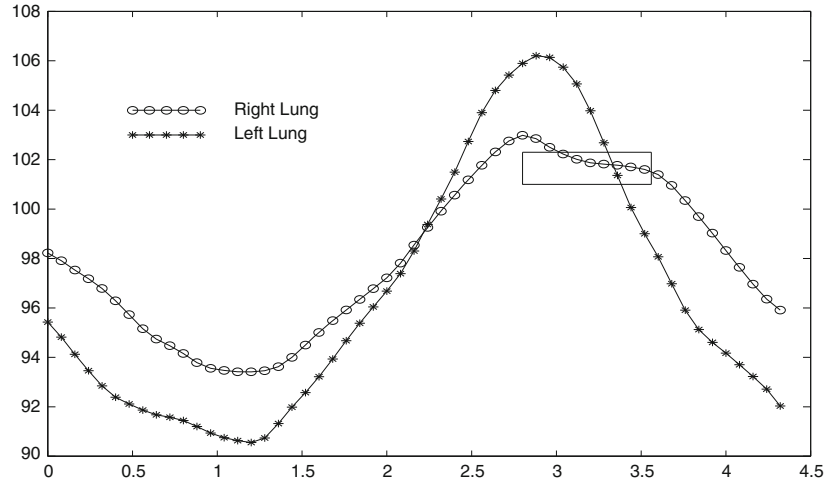
For effective perfusion analysis, it is essential to understand the cardiac function. Analyzing the cardiac motion gives the first-hand information concerning its function.



(a) Frame 1 (t=0) **(b)** Frame 39 (t=3.04) **(c)** Frame 40 (t=3.12)

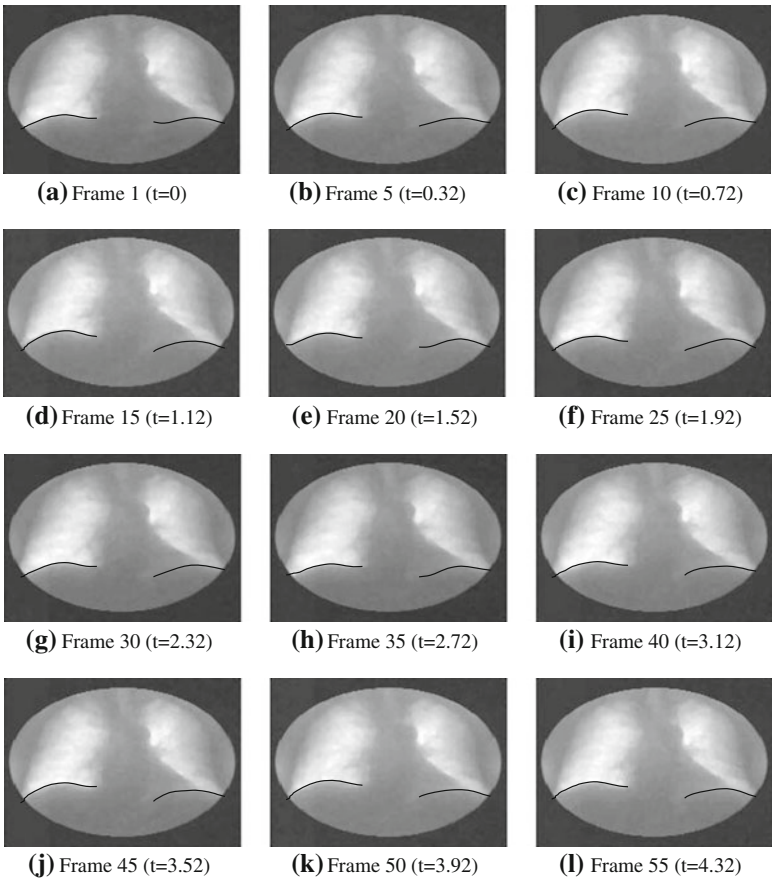
(d) Frame 41 (t=3.20) **(e)** Frame 42 (t=3.28) **(f)** Frame 43 (t=3.36)

(g) Frame 44 (t=3.44) **(h)** Frame 45 (t=3.52) **(i)** Frame 46 (t=3.60)



(j) Diaphragm Motion: Average y-positions of the two open snakes

Fig. 11 Clinical case II: Local abnormal diaphragm motion. **a** The two open snakes on the first frame. (**b**)–(**i**) From Frame 39 (t = 3.04) to Frame 46 (t = 3.60), the right diaphragm is expected to contract even further during inhalation, but it exhibits no observable motion (in term of pixels) for about half a second. **j** The average positions of the two open snakes. The local motion abnormality is indicated with a rectangle. The right diaphragm motion amplitude is also small



(a) Frame 1 (t=0) **(b)** Frame 5 (t=0.32) **(c)** Frame 10 (t=0.72)

(d) Frame 15 (t=1.12) **(e)** Frame 20 (t=1.52) **(f)** Frame 25 (t=1.92)

(g) Frame 30 (t=2.32) **(h)** Frame 35 (t=2.72) **(i)** Frame 40 (t=3.12)

(j) Frame 45 (t=3.52) **(k)** Frame 50 (t=3.92) **(l)** Frame 55 (t=4.32)

(m) Diaphragm Motion: Average y-positions of the two open snakes

Fig. 12 Clinical case III: Global abnormal diaphragm motion. The left diaphragm cannot move freely and exhibits irregularities during exhalation

Table 6 Statistics of diaphragm motion abnormalities

	Global	Local
Right	4	11
Left	3	7

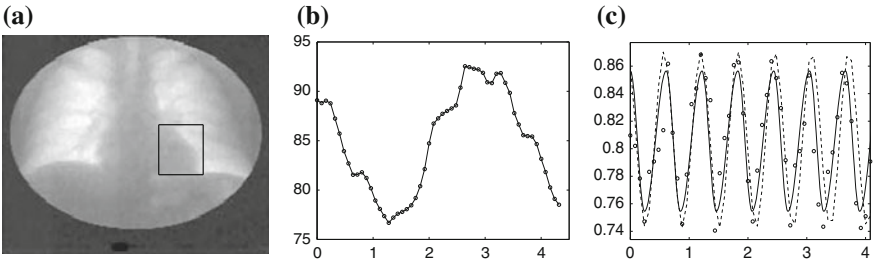


Fig. 13 **a** An ROI on the heart and **b** its corresponding observation. The observation is dominated by ventilation when the patient breathes. Our simple approach cannot extract the systolic and diastolic phases from such a signal. **c** However, with cardiac motion tracking (Fig. 15), the heart signal (indicated with "o") can be obtained with Eq. 8, and cardiac systolic and diastolic phases from the heart signal can be extracted with our perfusion model even when the patient breathes (the dashed curve indicates the initial guess and the final optimized result is shown with the solid curve)—the uptime corresponds to the diastolic phase (the heart proportion increases) and the downtime to the systolic phase of the heart (the heart proportion decreases). The extracted parameters for this patient are included in Table 7 as clinical case I, indicating that the heart is working properly

6.1 Characterizing Cardiac Motion

In order to use cardiac information to accelerate the pulmonary perfusion analysis, we proposed a simple approach to extract the cardiac systolic and diastolic phases from the heart [17]. We justify in Sect. 6.2 that this simple approach is sufficiently accurate for this purpose. However, it cannot be used to characterize complete cardiac motion for the following reasons:

- Its amplitude cannot be fully trusted in measuring the effectiveness of cardiac function. During perfusion examination, the patient is asked to hold their breath. The amount of air held in the lungs may differ from patient to patient and may differ from examination to examination, even for the same patient. As a result, when there is more air kept in the right lung, even if the heart doesn't pump effectively we may still have a higher amplitude due to the higher contrast along the cardiac boundary.
- This simple approach does not work when ventilation is present. When the patient breathes, the signal will be dominated by ventilation as shown in Fig. 13. Therefore, it cannot extract the systolic and diastolic phases from such a signal.

United Snakes offers a general solution for characterizing cardiac motion. By tracking cardiac motion (see Fig. 14), the heart proportion in an ROI can be computed

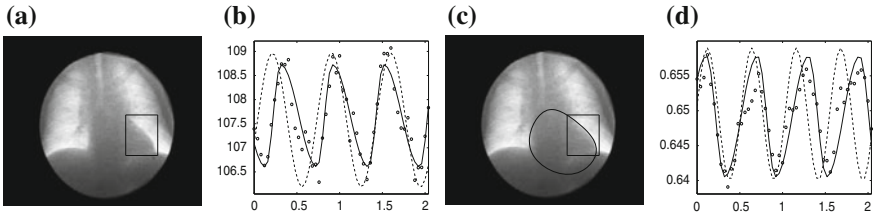


Fig. 14 Clinical case IV: Extracting cardiac motion parameters (with the same notation as in Fig. 16)

Table 7 Cardiac motion parameters given by our general approach for the four clinical cases

Clinical cases	Our new approach				
	A (amplitude %)	D (systolic)	U (diastolic)	S (timeshift)	F (frequency)
I	5.10	0.2398	0.3698	0.6106	98
II	2.40	0.3869	0.7931	0.9252	51
III	2.93	0.2065	0.9277	0.2144	53
IV	0.86	0.2106	0.3869	0.1105	100

The downtime (D) corresponds to the systolic phase, while the uptime (U) corresponds to the diastolic phase. The amplitude parameters show that the heart of Case I (a clinically normal case) is pumping effectively, while the heart pumping of Case IV (with advanced pulmonary embolism) is extremely weak. Cases II and III (both athletes) represent the typical "ineffective" phenomena of athletic hearts during rest: Low amplitude and low frequency

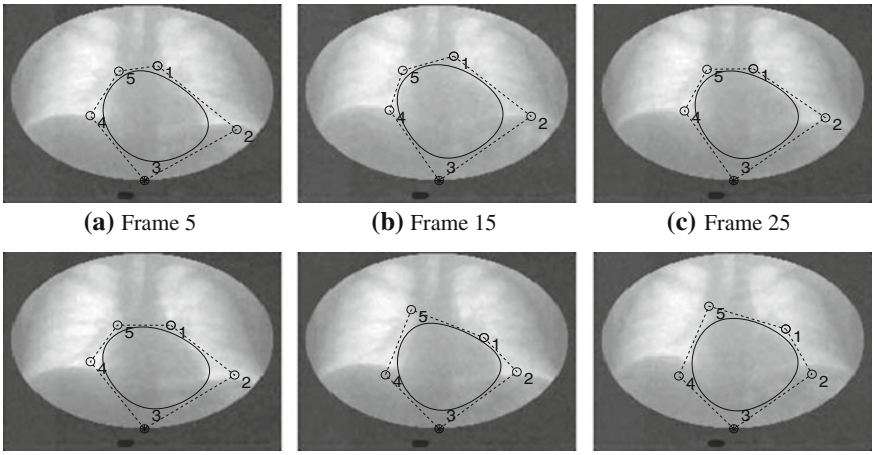
over time throughout the entire sequence resulting in a heart signal:

$$s_h = \frac{|\text{intersection of heart and ROI}|}{|\text{ROI}|}. \tag{8}$$

Analyzing the heart signal reveals the condition of cardiac function.

6.2 Clinical Case Studies

Clinical studies show that the new approach is effective in detecting abnormal cardiac function. We have applied this new approach to all of our clinical cases (53 ventilation patients and the 52 perfusion cases). We found weak cardiac function in 16 patients. Included here are three representative cases. Their results are listed in Table 7 and shown in Figs. 15, 16 and 17 as clinical cases II, III and IV. The amplitude parameters show the effectiveness of heart pumping action. According to our experiments, we expect that the amplitude of a healthy person is greater than 5.0 % (with the exception of athletes). Case I is a clinically normal case and the heart is pumping effectively. The heart pumping action of Case IV (with advanced pulmonary embolism) is extremely



(a) Frame 5 **(b)** Frame 15 **(c)** Frame 25

(d) Frame 35 **(e)** Frame 45 **(f)** Frame 55

Fig. 15 A united snake with B-spline shape functions (noise insensitive) is used to track cardiac motion (every tenth image shown). The dashed polygon is the B-spline control polygon. To effectively bridge the gap along the heart boundary, a hard constraint is further imposed on control polygon node 3. The robust tracking performance is largely due to the hard constraint

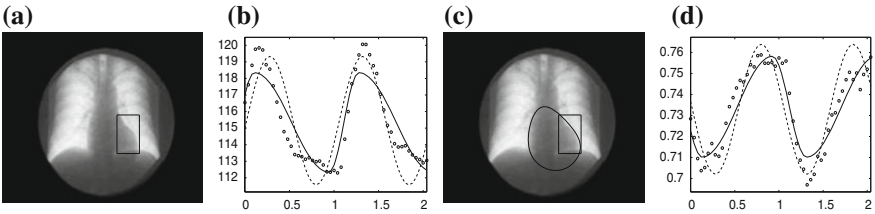


Fig. 16 Clinical case II: Extracting cardiac motion parameters with our previous, simple approach used for perfusion analysis in [17] (**a**, **b**) and the more general approach (**c**,**d**). The observations are indicated with "o", the initial guesses with the dashed curves and the final optimized results with the solid curves. The numerical results are given in Tables 8 and 7, respectively

ineffective. Cases II and III are both athletes representing the typical "ineffective" phenomena of athletic hearts during rest (low amplitude and slow cardiac rate).

The cardiac motion tracking approach can be used to justify the accuracy of our previous approach in [17]. By comparing the results given by our previous approach (see Table 8) and by the motion tracking approach (see Table 7), the absolute differences of the systolic/diastolic phases are less than 0.08 s. This means that our previous approach is a simple, fast, and accurate technique for extracting the systolic and diastolic phases from the heart.

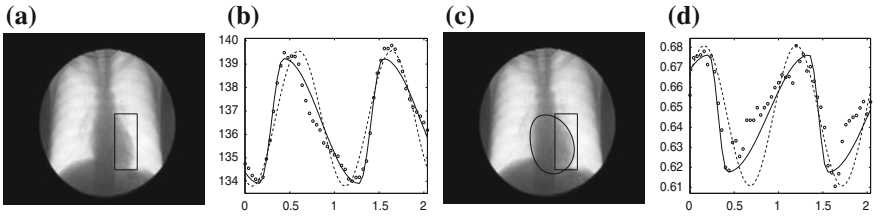


Fig. 17 Clinical case III: Extracting cardiac motion parameters (with the same notation as in Fig. 16)

Table 8 Cardiac motion parameters given by our previous approach

Clinical cases	Our previous approach			
	D (diastolic)	U (systolic)	S (timeshift)	F (frequency)
I	N/A	N/A	N/A	N/A
II	0.8361	0.3366	0.1067	51
III	0.8473	0.2658	0.4301	53
IV	0.4250	0.1770	0.3112	100

The method does not work when ventilation is present (Case I). The amplitude is not included, since it does not have a well-defined medical meaning. The uptime (U) for the systolic phase and the downtime (D) for the diastolic phase. The uptime and downtime in this table have completely different medical meanings from those in Table 7. Comparing with Table 7, the maximal absolute difference in systolic/diastolic phases is 0.08 s. Therefore, the simple approach is sufficiently accurate for extracting the systolic and diastolic phases and safe for perfusion analysis in [17]

7 Cardiac Shape Analysis

7.1 Revised Cardiothoracic Ratio (RCTR)

Cardiac size is an important and useful diagnostic parameter in chest radiographs. The conventional measurement for assessing cardiac enlargement is the cardiothoracic ratio (CTR) [18, 19]. Referring to Fig. 18, the CTR is defined as the ratio of the transverse diameter of the cardiac shadow to the greatest transverse diameter of the thorax [20] or the transverse diameter of the thorax at the highest level of the diaphragm [21]. Many researchers have reported the relationship between cardiac disease and heart size, and demonstrated the usefulness of the CTR and estimated cardiac size in clinical applications (e.g., [22–29]). Several research groups have attempted the automatic calculation of cardiac parameters for diagnosis [30–36].

However, the CTR is generally calculated from a single chest radiograph in which the heart may be in any phase of motion and, consequently, it is subject to measurement errors. In Dynamic Pulmonary Imaging, a whole cardiac motion cycle is available and the minimal CTR and maximal CTR may be computed to form an interval as a revised measurement of cardiac size. That is,

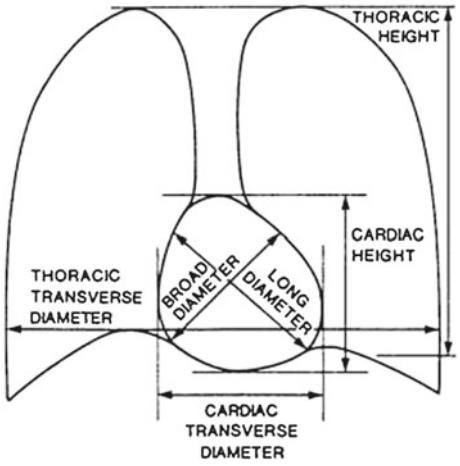


Fig. 18 Various parameters of the heart and lungs in chest radiographs given by Nakamori et. al. in [36]. "Cardiac broad diameter" may also be referred as "cardiac short diameter"

$$CTR_{min} = \frac{\min_t C(t)}{\max_t T(t)} \tag{9}$$

$$CTR_{max} = \frac{\max_t C(t)}{\max_t T(t)} \tag{10}$$

where $C(t)$ is the transverse diameter of the cardiac shadow at time t and $T(t)$ is the transverse diameter of the thorax at time t. The difference between CTR_{min} and CTR_{max} indicates the strength (effectiveness) of the heart pumping action. For a normal case, we expect that $CTR_{min} < CTR_{max} \leq 50\,\%$ and $CTR_{max} - CTR_{min} > 5\,\%$ with the exception of athletes.

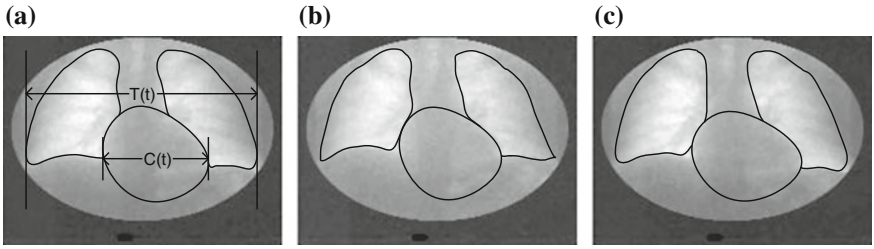


Fig. 19 Case I (with ventilation): Calculating RCTR. **a** The maximal thoracic transverse diameter (i.e., $max_t(T(t))$). **b** The minimal cardiac transverse diameter (i.e., $min_t(C(t))$). **c** The maximal cardiac transverse diameter (i.e., $max_t(C(t))$). A clinically normal case with normal heart size and effective heart pumping

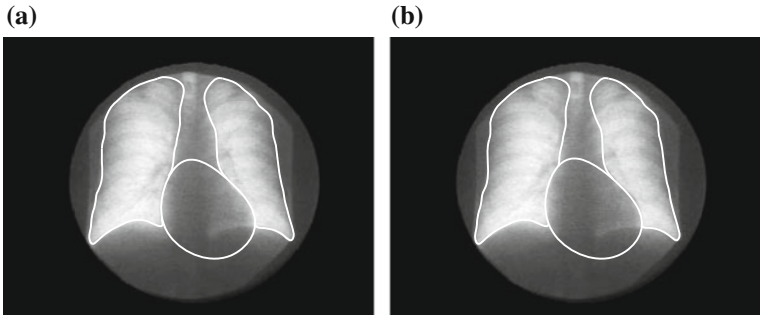


Fig. 20 Case II (with the breath held): **a** The minimal cardiac transverse diameter. **b** The maximal cardiac transverse diameter. The thoracic transverse diameter is a constant over time due to the held breath. An athlete. Heart size is normal, but heart pumping is weak. This is the typical "ineffective" phenomena of athletic hearts during rest

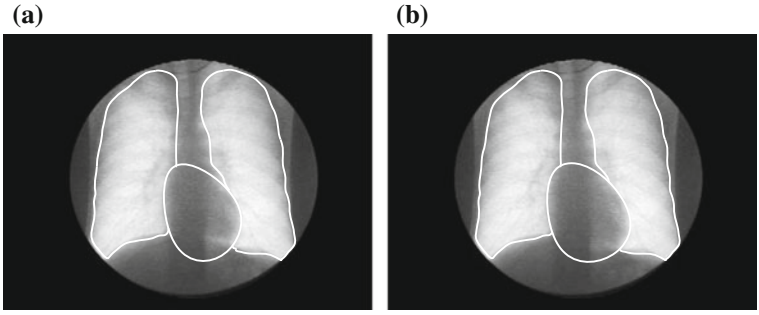


Fig. 21 Case III (with the breath held): Another athlete with the typical phenomena of athletic hearts during rest. The same notation is used as in Fig. 20

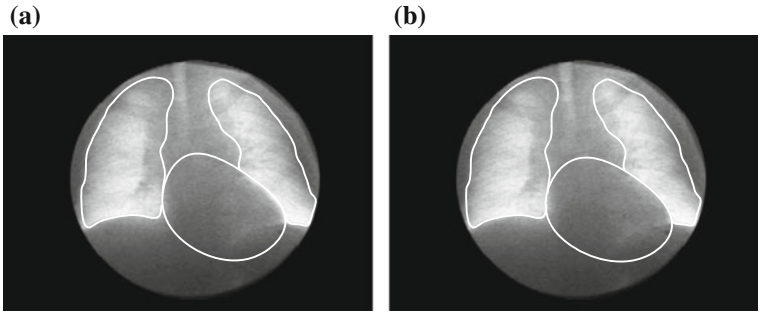


Fig. 22 Case IV (with the breath held): Calculating RCTR. **a** The minimal cardiac transverse diameter. **b** The maximal cardiac transverse diameter. The thoracic transverse diameter is a constant over time due to the breath held. A pathological case with advanced pulmonary embolism. Large heart and extremely ineffective heart pumping action

7.2 Clinical Case Studies

We have applied this new measurement to our 53 ventilation patients and 52 perfusion patients. We have found that 7 of them have large hearts and heart pumping is not effective in 16 cases. Here we list four representative cases:

- **Case I (with ventilation)** (see Fig. 19): A clinically normal case. Normal heart size and effective heart pumping ($CTR_{min} = 40.22\,\%$ and $CTR_{max} = 46.53\,\%$).
- **Case II** (see Fig. 20): An athlete. Heart size is normal, but heart pumping is weaker ($CTR_{min} = 45.97\,\%$ and $CTR_{max} = 47.29\,\%$) than Case I. This is the typical "ineffective" phenomena of athletic hearts
- **Case III** (see Fig. 21): Another athlete. Heart size is good but it does not pump as effectively ($CTR_{min} = 37.99\,\%$ and $CTR_{max} = 40.92\,\%$) as Case I. Again, the typical phenomena of athletic hearts during rest.
- **Case IV** (see Fig. 22): This is a pathological case with advanced pulmonary embolism. Large heart and extremely ineffective heart pumping ($CTR_{min} = 53.49\,\%$ and $CTR_{max} = 54.38\,\%$).

These patients are the same as in Sect. 6.2. Referring to Table 7, we can see that the findings there are corroborated by the new findings here.

8 Conclusions

This chapter has utilized our United Snakes technique to reveal pulmonary ventilation and perfusion abnormalities through motion and shape analysis of the lungs and heart. This fluoroscopical examination takes only about 4 s for ventilation studies and 2 s for perfusion studies with low radiation dose to the patient and with no preparation, radioactive isotopes, and contrast media.

References

1. Kiuru A, Svedstrom E (2003) Method to measure the relative perfusion of the lungs. U.S. Patent US 6,522,720 B1, 28 Feb 2003
2. Levenberg K (1944) A method for the solution of certain problems in least squares. Quart Appl Math 2:164–168
3. Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. SIAM J Appl Math 11:431–441
4. Scales LE (1985) Introduction to non-linear optimization. Macmillan Publishers Ltd, New York
5. Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. Int J Comput Vision 1(4):321–331
6. McInerney T, Terzopoulos D (1996) Deformable models in medical image analysis: a survey. Med Image Anal 1(2):91–108

7. Singh A, Goldgof D, Terzopoulos D (eds) (1998) Deformable models in medical image analysis. IEEE Computer Society Press, Los Alamitos
8. Blake A, Isard M (1998) Active contours. Springer-Verlag, New York
9. Mortensen EN, Barrett WA (1995) Intelligent scissors for image composition. In: Proceedings of computer graphics (SIGGRAPH'95), Los Angeles, CA, Aug 1995, pp 191–198
10. Barrett W, Mortensen E (1997) Interactive live-wire boundary extraction. Med Image Anal 1(4):331–341
11. Mortensen EN, Barrett WA (1998) Interactive segmentation with intelligent scissors. Grap Models Image Process 60:349–384
12. Falcão AX, Udupa JK, Samarasekera S, Hirsch BE (1996) User-steered image boundary segmentation. In: Proceedings of SPIE on medical imaging, vol. 2710, Newport Beach, CA, 1996, pp 278–288
13. Falcão AX, Udupa JK (1997) Segmentation of 3D objects using livewire. In: SPIE on medical imaging 1997, vol 3034, Newport Beach, CA, 1997, pp 228–239
14. Falcão AX, Udupa JK, Samarasekera S, Sharma S (1998) User-steered image segmentation paradigms: live wire and live lane. Grap Models Image Process 60:233–260
15. Falcão AX, Udupa JK, Miyazawa FK, An ultra-fast user-steered segmentation paradigm: Live-wire-on-the-fly. IEEE Trans Med Imaging (to appear)
16. Liang J, McInerney T, Terzopoulos D (2006) United snakes. Med Image Anal 10(2):33–215
17. Liang J, Järvi T, Kiuru A, Kormano M, Svedström E (2003) Dynamic chest image analysis: model-based perfusion analysis in dynamic pulmonary imaging. EURASIP J Appl Sig Process 2003(5):437–448 (Special Issue on Advances in Modality-Oriented Medical Image Processing)
18. Burgener FA, Kormano M (1991) Differential diagnosis in conventional radiology. Thieme, Stuttgart
19. Sutton D (1987) A textbook of radiology and imaging. Churchill Livingston, New York
20. Shanks SC, Kerley P (1962) A text-book X-ray diagnosis. Saunders, Philadelphia
21. Cooley RN, Schreiber MH (1978) Radiology of the heart and great vessels. William and Wilkins, Baltimore
22. Fuster V, Gersh BJ, Giuliani ER, Tajik AJ, Brandenburg RO, Frye RL (1981) The natural history of idiopathic dilated cardiomyopathy. Am J Cardio 47:525–531
23. Hutsebaut J, Scano G, Garcia-Herreros P, Degre S, DeCoster A, Sergysels R (1981) Hemodynamic characteristics in chronic obstructive lung diseases as related to cardiac size. Respiration 41:25–32
24. Gomez GA, Park JJ, Panahon AM, Pathasarathy KL, Pearce J, Reese P, Bakshi S, Henderson ES (1983) Heart size and function after radiation therapy to the mediastinum in patients with Hoggkin's disease. Cancer Threat Rep 67:1099–1103
25. Edwards DK, Higgins CB, Gilpin EA (1981) The cardiothoracic ratio in newborn infants. Am J Radiol 136:907–913
26. Kortman KE, Edwards DK, Deutsch AL, Higgins CB (1984) Heart size in newborn infants with birth asphyxia. Am J Radiol 143:533–535
27. Lauder IJ, Milns JS (1976) Longitudinal study of heart size in older people. Br Heart J 38:1286–1290
28. Nickol K, Wade AJ (1982) Radiographic heart size and cardiothoracic ratio in three ethnic groups: a basis for a simple screening test for cardiac enlargement in men. Br Heart J 55:399–403
29. Kabala JE, Wilde P (1987) The measurement of heart size in the anteroposterior chest radiograph. Br J Radiol 60:981–986
30. Myers PH, Nice CM, Becker HC, Nettleton WJ, Sweeney JW, Mechstroth GR (1964) Automated computer analysis of radiographic images. Radiology 83:1029–1033
31. Becker HC, Nettleton WJ, Myers PH, Sweeney JW, Mechstroth GR, Nice CM (1964) Digital computer determination of a medical diagnostic index directly from chest X-ray images. IEEE Trans Biomed Eng 11:67–72
32. Hall DH, Lodwick GS, Kruger RP, Dwyer SJ, Townes JR (1971) Direct computer diagnosis of rheumatic heart disease. Radiology 101:497–509

33. Kruger RP, Townes JR, Hall DL, Dwyer SJ, Lodwick GS (1972) Automated radiographic diagnosis via feature extraction and classification of cardiac size and shape desecrators. IEEE Trans Biomed Eng 19:174–186
34. Sezaki N, Ukena K (1973) Automatic computation of the cardiothoracic ratio with application to mass screening. IEEE Trans Biomed Eng 20:248–253
35. Paul JL, Levine MD, Fraser RG, Laszlo CA (1973) Automatic. IEEE Trans Biomed Eng 20:248–253
36. Nakamori N, Doi K, Sabeti V, MacMahon H (1990) Image feature analysis and computer-aided diagnosis in digital radiography: Automated analysis of sizes of heart and lung in chest images. Medl Phys 17(3):342–350

Epithelial Cell Segmentation via Shape Ranking

Alberto Santamaria-Pang, Yuchi Huang, Zhengyu Pang, Li Qing and Jens Rittscher

Abstract We present a robust and high-throughput computational method for cell segmentation using multiplexed immunohistopathology images. The major challenges in obtaining an accurate cell segmentation from tissue samples are due to (i) complex cell and tissue morphology, (ii) different sources of variability including non-homogeneous staining and microscope specific noise, and (iii) tissue quality. Here we present a fast method that uses cell shape and scale information via unsupervised machine learning to enhance and improve general purpose segmentation methods. The proposed method is well suited for tissue cytology because it captures the the morphological and shape heterogeneity in different cell populations. We discuss our segmentation framework for analysing approximately one hundred images of lung and colon cancer and we restrict our analysis to epithelial cells.

A. Santamaria-Pang (✉) · Y. Huang · Z. Pang · L. Qing · J. Rittscher
GE Global Research, Niskayuna, NY, USA
e-mail: santamar@ge.com

Y. Huang
e-mail: yuchi.huang@gmail.cn

Z. Pang
e-mail: pang@ge.com

L. Qing
e-mail: liq@ge.com

J. Rittscher
e-mail: rittsche@ge.com

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_10,

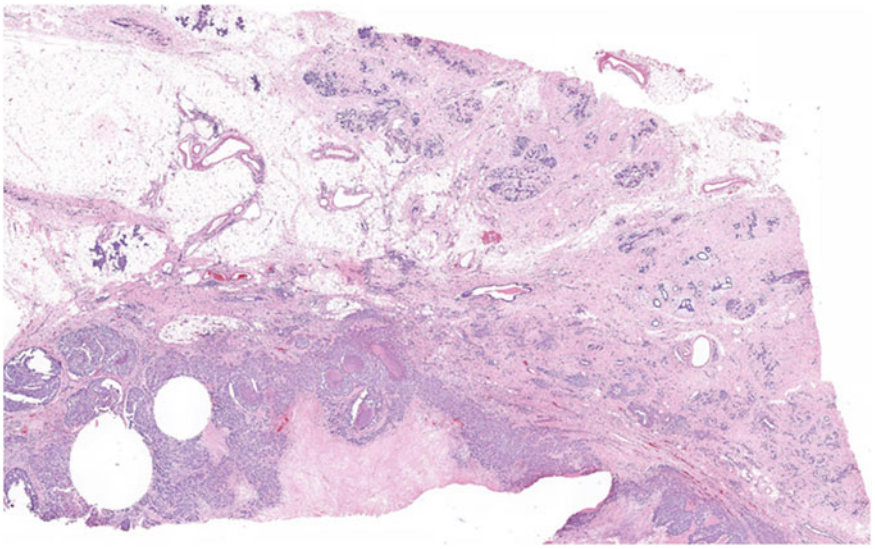


Fig. 1 Example of a whole slide image of breast tissue. Digital images of whole tissue sections are very large. This particular tissue section has been acquired at 20X resolution and the resulting image size is $71,300 \times 44,400$ pixels. While pathologists are trained on what areas of the tissue they need to focus, it is very difficult to construct analysis methods that mimic this process. As it is necessary to assess morphological properties of individual cells, it is, clear that the analysis of such tissue samples needs to take several different resolutions into account

1 Introduction

According to the American Cancer Society, in 2013 nearly 8 million of people will die due to cancer worldwide; cancer will account for approximately 22 % out of all the non-communicable diseases [1]. Even more concerning, the current mortality trend is projected to increase by 60 %, with an estimated 13.1 million deaths in 2030 [2]. Histology plays a critical role in both clinical practice and basic research. To date most pathologists make use of very traditional tools and techniques. Light microscopy, and in particular, Hematoxylin-Eosin (H&E) and Diaminobenzidine (DAB) staining, have been essential technologies that pathologist have used for a long time. Traditionally, pathologists make clinical assessments using H&E or DAB stained tissue sections by interrogating the cellular properties such as nuclei shape, texture, size, and the spatial relationship of the cells within the tumor.

Low cost, short acquisition time, consistent coloration, and the ability to highlight the overall tissue architecture are the advantages of this process. Figure 1 depicts a whole slide image of a breast cancer sample from a Hematoxylin-Eosin (H&E) stained tissue section acquired at 20X resolution. As it is shown, the tissue section is highly heterogeneous for both morphology and tissue components.

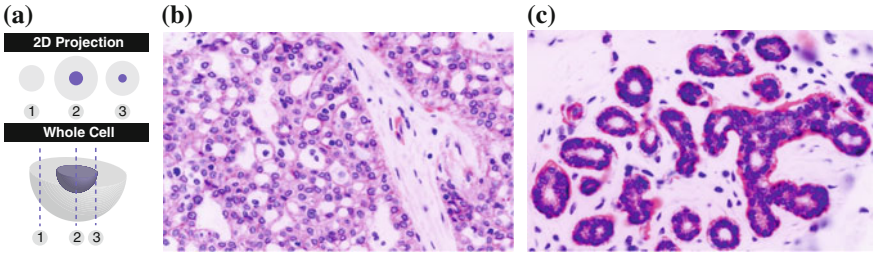


Fig. 2 Tissue analysis challenges. Most traditional histology methods are based on thin tissue sections. In that sense the histology slide is only a sample of the overall three dimensional tissue block. The graphic on the *left* (**a**) illustrates how this sampling process affects the appearance of individual cells. It is, for example, clear that not every cellular region needs to contain a nucleus. The tissue sections on the right give an example how disease can effect tissue architecture. Morphology of (**b**) breast cancer and (**c**) breast normal tissue

The field of digital microscopy is relatively new and it can be seen as a one of the key technologies that researchers use to increase our understanding of cancer. It promises to bring precision medicine to regular clinical practice with the goal of improving early detection, prognosis, and treatment [3, 4]. It is now possible to capture very rich information from tissue samples and cancer biopsies in form of high resolution images that can be reviewed on any computer rather than a dedicated microscope. This has already redefined clinical workflows and enabled methods of collaboration that have not been possible before [5, 6]. The impact of this digitization process will go beyond the way in how histology and cytology images are stored, viewed, analyzed and shared. Like Computer Assisted Diagnosis (CAD) enabled by advanced algorithms revolutionized certain radiology tasks, the evolving field of computational pathology will enable new methods of fully automated and user assisted diagnosis [4, 7, 8].

Guidelines like the Gleason Score [9] or the Nottingham Prognostic Index [10] include measurements such as the shape and size of nuclei. For example, in breast cancer nuclei size and mitotic features are diagnostic indicators. While anatomical atlantes [11] can greatly enhance our understanding of disease at the organ level, such prior knowledge is not available for analyzing the complex tumor heterogeneity at the microscopic scale. One of the major challenges in current pathology is human subjectivity in clinical assessment. Among the major benefits of integrating automated image analysis in clinical workflows are: objective quantification, study reproducibility, and consistency. The structure and appearance of the tumor microenvironment varies greatly according to the function of each specific organ [5, 12–14]. The major challenges towards a comprehensive framework in computational pathology and histopathology can be summarized as:

- Complex cell morphology. Cells are three-dimensional objects, and the corresponding microscopy image captures are two-dimensional projections which correspond to the slice of the tissue. Figure 2a illustrates examples of partial cell volumes at different focal planes.

- Tissue architecture. Tissue morphology varies greatly from organ to organ. Different organs and tissue types exemplify different tissue parenchymal architecture and stromal components. Cancer tissue shows very different morphological features from normal tissue even in their original organs. Figure 2b and c show different tissue architectures in breast cancer and normal tissue respectively.
- Process variability. There are a number of sources of variability including pre-analytical variables such as how tissues are fixed, processed, stored, and sectioned and analytical variables such as staining protocol, image acquisition and instrumentation.

As part of a novel fluorescent multiplexing method, referred to as MultiOmyx^{TM} and described in detail by Gerdes et al. [13], we have developed a single cell segmentation framework for tissue images. The hallmark of the MultiOmyx^{TM} process is the ability to image a large number of protein targets in a single tissue section. The overall goal of this segmentation framework is the detection and precise delineation of individual cells. This information is then compiled into a map of the tissue. Subsequently this tissue map is used to quantify the expression of the set of protein markers at a single cell level.

As part of different preclinical studies, this new imaging process is being applied to a large morphological variations of different tissue types and disease states. Hence our segmentation framework needs to handle a broad range of morphological variations. The concrete example of a lung and colon cancer dataset will be presented. In order to achieve more robust and reliable segmentation results, we have explored the utility of a shape ranking algorithm. Our focus here is the review of algorithm development, the discussion of the merits of the proposed method in context of a concrete study, and the presentation of some recent performance improvements.

The organization of this Chapter is the following. In Sect. 2, we will review relevant methods for data tissue analysis based on pattern recognition approaches and tissue morphological reconstruction. In Sect. 3, we will outline the GE MultiOmyx^{TM} imaging for acquiring immunohistochemistry (IHC) multiplexing imaging, in Sect. 4 we will present our cell shape segmentation algorithm. Results and discussion will be presented in Sects. 5 and 6 respectively. Finally in Sect. 7 we will present our future work.

2 Related Literature

As part of this Chapter we review a set of methods that enhance the ability of extracting the local tissue architecture in order to construct a tissue map. The limited space available does not permit a comprehensive review of relevant literature as many of the established image segmentation approaches have been adopted to cell and tissue image analysis. For a detailed review of image analysis methods in tissue histopathology we direct the reader to Gurcan et al. [8]. Here we would like to differentiate between approaches that focus on the extraction of disease specific pat-

terns or signatures and methods that apply model-based algorithms to reconstruct the architecture of the tissue. In Sect. 2.1 we summarize recent research [7, 15–22] that makes extensive use of sophisticated feature extraction techniques and machine learning algorithms. The second group, related to the model-based approaches for reconstructing the tissue morphology at the individual cell level [23–30] are reviewed in Sect. 2.2.

2.1 Pattern Recognition Methods

Doyle et al. [18, 19] presented a Bayesian multi-resolution classification method for prostate cancer tissue from whole slide histopathology images. The approach is based on extracting a number of intensity and texture based statistics at different resolution levels and deriving a Bayesian classification method. In Basavanhally et al. [16] a learning-based method for detection and grading lymphocytic infiltration in breast cancer tissue using histopathology images was proposed. The method automatically segments lymphocytes using region growing and Markov random field algorithms. A graph representation then is constructed from the segmented lymphocytes, defining fifty features used for lymphocytic classification. The method was applied to approximately one hundred images obtained from fifty eight patients. Monaco et al. [15] used a Markov Random Field to train a model for identifying glands from training data. While the model effectively deals with the wide variety of shapes, it focuses on segmenting and labeling regions and does not highlight individual cells.

Kaynig et al. [17] introduced an energy function that incorporates the probability output from a random forest classifier. The goal of this approach is to improve the segmentation of elongated structures (e.g., membranes) in electron microscopy images. Their model combines a discriminative model for membrane appearance learned from training data with perceptual grouping. The effectiveness of incorporating additional shape priors [31] that are identified through supervised training has been demonstrated recently. A graph-based method for mitosis identification in breast cancer in whole tissue slides is proposed by Roullier et al. [22]. The approach consisted of decomposing the whole image at multiple resolution levels and performing a graph-based clustering approach using a number of image features at different resolution levels.

2.2 Model-Based Methods

Typically, cell detection relies on structural markers that include nuclei and membrane markers [32, 33]. A variation of the watershed segmentation was presented by Na and Heru [33] to segment milk somatic cell images. The proposed method addresses the typical over-segmentation obtained from the watershed algorithm. Srinivasa et al. [34] proposed an active mask algorithm for cell segmentation in flu-

orescence microscope images based on punctate patterns. Their method is inspired by active-contour methods and multiresolution methods. The authors demonstrated their method in HeLa cells.

In Xiao et al. [35], a method for segmenting stem cells is proposed. The algorithm uses a morphological descriptor for cellular shapes in terms of a "symmetry axis transformation". The method accounts for morphological changes that are induced by cell growth. A method for segmenting cells for in-situ microscopy is presented in Martinez et al. [36]. Here the instrument is positioned inside a bioreactor in order to monitor cell culture processes. The method relies on a bubble segmentation algorithm which is based on shape from shading. Cells are segmented based on closed boundaries that are extracted from thresholding a depth map by applying Bichsel and Pentland's original shape from shading algorithm. A framework for supervised cell-image segmentation and a touching-cell splitting method is proposed by Kong et al. [37]. In this work cells are segmented by classifying the image pixels into either cell or extra-cellular category. The classification algorithm uses the color-texture extracted at the local neighborhood of each pixel. Local features utilized by the classification algorithm rely on a local Fourier transform from a color space.

Recently, a number of learning-based techniques have been suggested. Xiaong [38] proposed the use of computer generated models to provide synthesized images of healthy red blood cell populations. In order to develop cell segmentation and counting algorithms, learning-based techniques were used. The estimation of average cell shape and deformation was inferred from the synthetic models. Park [39] presented a watershed-based algorithm for the segmentation of clustered cells. The method incorporates specific image color-knowledge using the watershed transform with iterative shape alignment to segment the cell shape. Extensions to 3D segmentation include segmenting the spots in cDNA microarray images [40]. The segmentation is represented in a three-dimensional (3-D) space by a 3-D spot model posed via an optimization problem, which is solved by a genetic algorithm. The 3D segmentation is provided by contours of the 3-D spot models.

Given a set of training images Lempitsky and Zisserman [41] proposed to learn a linear classifier that allows cell counting. This approach is very general and has been applied to live cell data. The concept of learning from dot annotations has been applied by Arteta et al. [42] to detect cell like structures from a broad number of candidate regions that have been scored with a learning based measure such that the solution is globally optimal. Their approach has been successfully applied to different microscopy techniques. The major limitation of both approaches is that the segmentation does not use morphological assessment of individual cells. In a similar path, hierarchical segmentation schemes have been suggested to splitand merge cells [24, 26]. These models use supervised learning and enforce spatialconsistency through Markov random fields [26].

Here we are building on both of these developments to formulate an unsupervised cell segmentation method for epithelial cells that incorporates shape clues to dynamically adapt the segmentation to the cell shape and morphology. Based on statistical shape analysis we propose a framework for simultaneous cell classification and detection in tumor tissue with very heterogeneous morphology. The key idea is

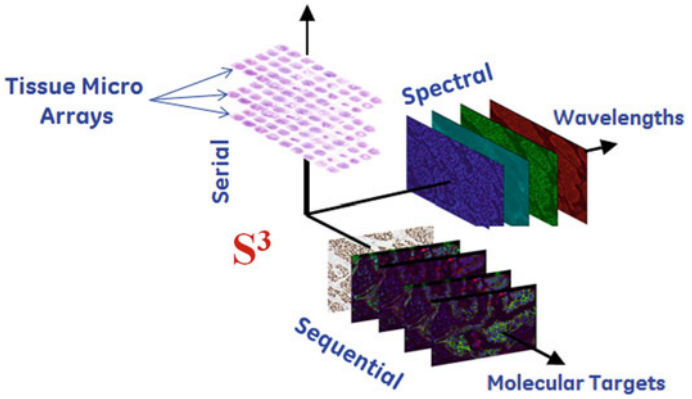


Fig. 3 Schematic of MultiOmyx™ IHC. The three main axes illustrate the key components of the technology: (i) serial information (ii) different excitation wavelengths such as DAPI, Cy3, Cy5, and (iii) repeated stain-image-bleach and de-stain sequence using direct antibody-fluorophor conjugation [43–46]

to optimize the cell segmentation according to the tumor tissue morphology via cell shape descriptors. This process is equivalent to solving in non-polynomial time an optimization problem, and here we propose an efficient solution via sparse random sampling.

3 Multiplexed Fluorescence Microscopy

Sequential multiplexed immunofluorescence is a powerful technique for understanding the complex interaction of different proteins in tissues while keeping the tissue morphology in context. We describe a novel image analysis method for an extended panel of biomarkers on a single, formalin-fixed paraffin embedded tissue section. We will provide a brief overview of the GE MultiOmyx™ methodology, for more details, we refer the reader to a more detail presentation in Gerdes et al [13]. We utilize a novel technology for simultaneous co-localization of multiple protein biomarkers on a single formalin-fixed paraffin embedded tissue section or core biopsy. The core of this technology is denoted as MultiOmyx™ and it has been developed at GE Global Research (Niskayuna, NY; USA). It comprises of a repeated stain-image-bleach and de-stain sequence using direct antibody-fluorophor conjugation [43–46]. Figure 3 illustrates the concept of applying MultiOmyx™ in Tissue Micro-Arrays (TMAs).[1]

[1] Tissue micro-arrays (TMAs) is a collection of tissue samples (biopsies) organized in a two-dimensional array. Typically, they contain hundred of samples organized in one or multiple two-dimensional arrays, where each sample has a diameter of approximately of 0.6 mm. The samples are collected using standardized tissue fixation protocols and each sample is embedded in paraffin and can be used as biomarker discovery tool [12–14].

This technology allows the simultaneous detection of multiple protein expressions at the individual cell level, revealing not only the complex morphology of the cancer tissue but the expression patterns of different molecules in each individual cell.

We have applied the MultiOmyx^{TM} to two TMAs using two different tissue types: lung and colon with approximately 100 images in each tissue type. Images were acquired at 20X magnification, a numerical aperture of 0.75, a pixel size of 0.37 μm. We used a customized Olympus^{TM} microscope with filters tuned to the emission wavelength of three dyes: cyanine3 (Cy3, emits at 570 nm); cyanine5 (Cy5, 670 nm); and 46-diamidino-2-phenylindole (DAPI, 460 nm). Image size was set to 2048×2048 pixels, using 12-bit digital camera. Tissue sections were 4 μm thick and they were fixed in paraffin. Modeling the inherent variability of the biological specimen combined with the image-to-image variation is the major challenges through robust tissue processing and analysis. Biomarker abundance, fixation methods, sectioning, and storage could affect the staining process. Different tissue types and, more importantly, different disease states exhibit vastly different morphology and architecture. In order to design computational tools that can, for example, effectively support the quantification of cell phenotypes, algorithms need to capture minute details and tissue specific morphological variations. Figure 4a presents a conceptual representation of the information content in TMA's. On the left we show a tissue micro-array, in the center a single tumor biopsy, and on the right a selected region of interest corresponding to epithelial tissue. The color images in the back, represent different excitation wavelengths corresponding to DAPI, Cy3, Cy5 channels and they target different molecular protein targets. Figure 4b–d present representative examples of epithelial cells in lung tissue exhibiting very different cell morphology. Cells in Fig. 4b, c show relatively uniformity in cell size and shape, whereas cells in Fig. 4d the show heterogeneous shape and size. Here we are combining general purpose segmentation with unsupervised machine learning techniques to construct algorithms that are robust while capturing phenotypically relevant information.

4 Cell Segmentation via Shape Ranking

The image analysis workflow of the MultiOmyx^{TM} process has been presented in more detail in [47]. In this chapter we focus on improving the robustness and fidelity of the cell segmentation algorithm. The current version of the single cell segmentation framework is based on a variant of a hierarchical watershed segmentation. For the purpose of this discussion, the reader can assume that any standard implementation the watershed segmentation can be utilized.

This approach is inspired by the idea of capturing the shape distribution for a given image class. Assuming that the available segmentation results are of reasonable quality, (i.e., the majority is correct) regions that correspond segmentation errors will be outliers of this shape distribution. Here we use a shape descriptor to capture the tissue's shape distribution. Based on this ranking function it is then possible to

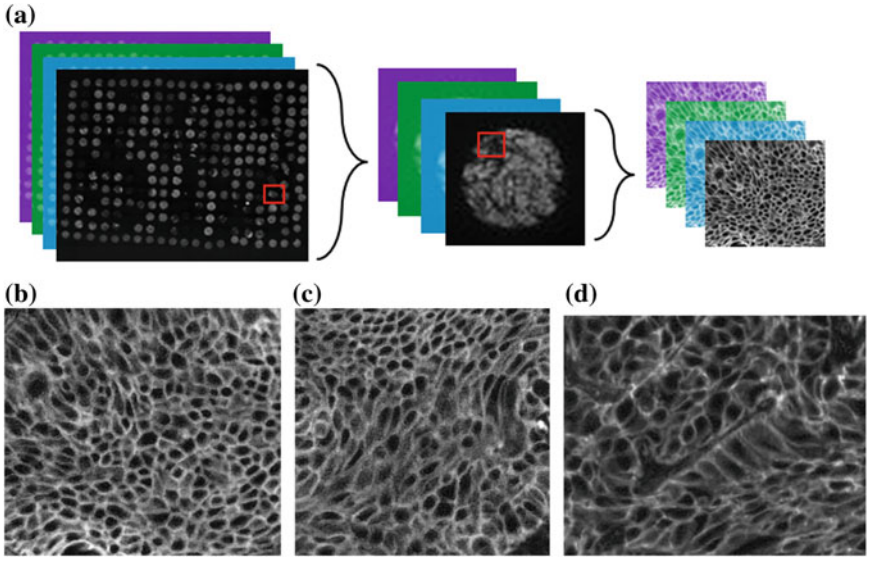


Fig. 4 Morphological heterogeneity in lung cancer tissue. Illustration of tissue micro-arrays and example of different size and morphology of epithelial cells from lung cancer tissue. **a** Regions of interest in tissue micro-arrays. Color images correspond to different excitation wavelengths corresponding to different molecular protein targets. Cells exhibiting: **b** smaller size, **c** medium size, and **d** larger size. Note the homogeneity of cell morphology and shape in **b** and **c** but the cell hetero- geneity in **d**

identify the regions in the hierarchical segmentation that rank high with respect to the given shape model.

4.1 Shape Ranking

The main idea of cell ranking is that abnormal objects can be consider to be outliers and they are likely the outcome of an error during the segmentation process. Assuming that the cell shape distribution is Gaussian-like, we aim to optimize the segmentation from a family of known parameters by minimizing the number of outliers during multiple segmentations. We presented a ranking method that maximizes the shape similarity using the k-nearest neighbors in [48]. While computing object similarity using k-nearest neighbor is computationally efficient, estimating a similarity matrix for a very large number of objects can be computationally very expensive. In this work, we propose to increase the efficiency of the cell ranking algorithm while preserving the fidelity of the shape ranking metric. Rather than estimating the object-to-object shape similarity from k-nearest neighbors, we propose to select an uniformed distributed sub-sample of the population.

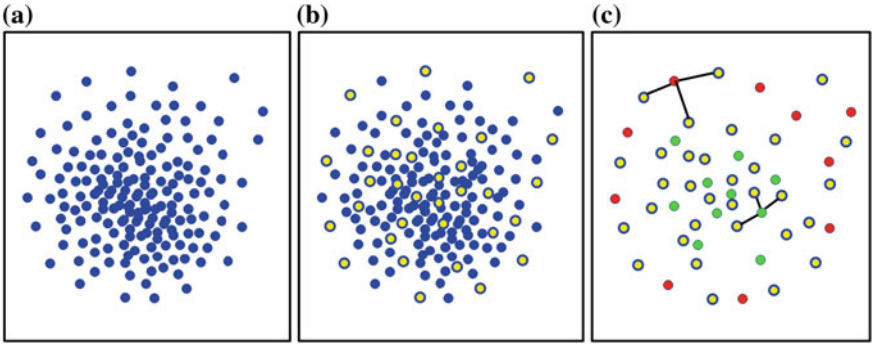


Fig. 5 Sub-sampling and ranking. Schematic illustrating the key idea of the similarity metric: outliers (abnormal points) should have a higher distance (lower ranking) as compared to points close to the center of the distribution (highest ranking). **a** Distribution of all points in X. **b** Distribution of a sample $A \subset X$ (*yellow*), **c** Outliers (*red*) and inliers (*green*) using k-nearest neighbors in A

Let $X = \{\mathbf{X}_1, ..., \mathbf{X}_n\}$ be a set of points, and $A \subset X$ a uniformly distributed sample, for $\mathbf{X}_i \in X$, let $N_i^A(k) \subset A$ represents the k-nearest neighbors within A. We then define the cost of every element of X within A as:

$$\mathscr{C}(\mathbf{X}_i) = \sum_{\mathbf{X}_j \in N_i^A(k)} d(\mathbf{X}_i, \mathbf{X}_j), \tag{1}$$

where $d(\mathbf{X}_i, \mathbf{X}_j)$ is a distance metric. Figure 5 illustrates the idea of the of the sub-sampling and ranking. Figure 5a and b shows all points in X (blue) and A (yellow). Figure 5c shows outliers (red) and normal (green) points. Note that according to our definition of similarity, the ranking of $\mathbf{X}_i$ is obtained from the cost $\mathscr{C}(\mathbf{X}_i)$ and indicates the degree of "abnormality"; that is, top candidates in the ranking have a high probability of being segmented correctly.

4.2 Shape Descriptors

Shape descriptors such as 'shape context' [49] and 'Histogram of Oriented Gradients' (HOG) [50] have been successfully applied to various problems in shape matching, recognition and object recognition. However, computing cell-to-cell similarity based on shape context involves a shape matching phase, which would be computationally very expensive. In the presence of thousands of different cells this would be unpractical. Histogram of Oriented Gradients is defined in the image coordinates, which is not translation, rotation and scale, invariant.

We represent the shape as a parameterized curve in polar-coordinates. The cell shape descriptors are rotational and scale invariant [48], and the provide an approx-

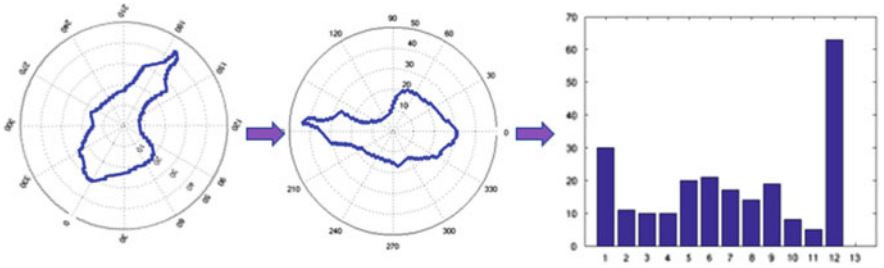


Fig. 6 Computation of the shape descriptors. The *left image* shows points corresponding to the cell border, the *center image* shows rotation along the major axis, and the *right image* illustrates a log-polar histogram

imation to the cell shape. We compute the shape descriptors in the following steps. First, given a segmented object, we select points that are two pixels away from the border and we translate the points with respect to its center of gravity, defining a new origin. Secondly, we rotate the border points with an angle equal to the axis of least inertia which is defined as the line for which the integral of the square of the distances to points on the shape boundary is a minimum. Thirdly, each boundary point is represented in the polar-coordinate system by a two-tuples (θ, ρ), where θ denotes the angular coordinate, ρ is the distance between the pole and the boundary point. The histogram from all boundary points is defined as:

$$x_i = \frac{1}{\sum\limits_{\forall j} B(j)} \sum_{\forall \theta \in B(j)} \log\left(1 + \frac{1}{\rho}\right), \tag{2}$$

where $B(j)$ denotes the degree interval of the i-th bin; $\frac{1}{\sum\limits_{\forall j} B(j)}$ is a normalization term. By this definition, our shape descriptor are translation, rotation and scale variant; the logarithm function makes them more sensitive to the boundary points near to the pole than those of points far away. The steps of computing the shape descriptor are illustrated in Fig. 6.

4.3 Shape Similarity Metric via Kernel Distance

We define a shape similarity measure by finding a suitable transformation K such that for any $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{\mathbf{n}}$, $K(\mathbf{X}, \mathbf{X}) = \mathbf{1}$ and as the distance $\mathbf{X}, \mathbf{Y}$ increases, $K(\mathbf{X}, \mathbf{Y})$ decreases. Then, K can be defined as a kernel transformation that induces a similarity measure. We encode shape and scale information as: $\mathbf{X} = [\mathbf{x}, \mathbf{a}] = [\mathbf{x_1}, ..., \mathbf{x_{n-1}}, \mathbf{a}]$, where x_i represents the shape descriptors and a is a scalar representing the object area.

We define a shape kernel distance measure that combines shape and scale information as:

$$K(\mathbf{X}_i, \mathbf{X}_j) = \alpha K_{shape}(\mathbf{x}_i, \mathbf{x}_j) + (1-\alpha) K_{scale}(a_i, a_j), \tag{3}$$

where K_{shape} and K_{scale} are kernels for the shape and scale terms, $0 \leq \alpha \leq 1$ regulates the shape-scale terms.

To measure shape-to-shape similarities, we use a Gaussian kernel:

$$K_{shape}(\mathbf{x}_i, \mathbf{x}_j) = e^{\left(-\frac{Dis_{shape}(\mathbf{x}_i, \mathbf{x}_j)}{\overline{Dis}_{shape}}\right)}, \tag{4}$$

where Dis_{shape} is the Chi-Squared histogram distance χ^2 to measure the distance between two shape descriptors:

$$Dis_{shape}(\mathbf{x}_i, \mathbf{x}_j) = \chi^2(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{2} \sum_{\forall k} \frac{|\mathbf{x}_i(k) - \mathbf{x}_j(k)|^2}{\mathbf{x}_i(k) + \mathbf{x}_j(k)}, \tag{5}$$

and $\overline{Dis}_{shape}$ is the mean value of the shape distances: $\overline{Dis}_{shape} = \frac{\sum \chi^2(\mathbf{x}_i, \mathbf{x}_j)}{N}$.

The scale information $K_{scale}(a_i, a_j)$ is defined as a Gaussian Kernel in terms of the cell area as:

$$K_{scale}(a_i, a_j) = e^{\left(-\frac{|a_i - a_j|)}{\mu_a}\right)}, \tag{6}$$

μ_a is the mean value of all scale differences. Note that $K(\mathbf{X}_i, \mathbf{X}_j)$ is a kernel since it is expressed as a linear combination of two kernel functions.

Figure 7 illustrates different types of ranking based on: size, shape and size plus shape, the ranking results are color coded from red (lowest ranking) to green (highest ranking). Figure 7a shows a synthetic image with a limited number of shape primitives at different sizes, Fig. 7b–d shows ranking results based on size, shape and size plus shape respectively. Figure 7e shows an instance of cell segmentation, and Fig. 7f–h represent the ranking results using the criteria previously described (note that green objects are the most similar).

4.4 Hierarchical Top-Down Segmentation

We define the cost (Eq. 1) of the object $\mathbf{X}_i \in X$, in terms of the its similarity function at the level $l \leq L$ as:

$$\mathscr{C}(\mathbf{X}_i) = S_i^l(\mathbf{X}_i) = \sum_{\mathbf{X}_j \in N_i^A(k)} d(\mathbf{X}_i, \mathbf{X}_j) = \sum_{j \in N_i^A(k)} -K(\mathbf{X}_i, \mathbf{X}_j), \tag{7}$$

where $N_i^A(k)$ is the set of k-nearest neighbors, $A \subset X$. We define the cumulative cost function $\mathscr{C}(\mathbf{X_i})$ in terms of the similarity S_i^l in the hierarchical segmentations as:

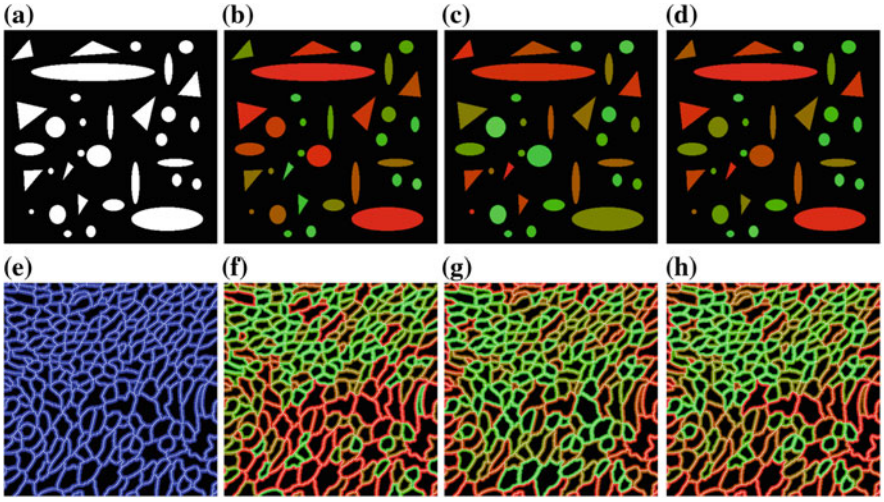


Fig. 7 Schematic illustrating different ranking results in a synthetic image and a lung tissue section. **a** Binary image with synthetic objects, **b** ranking by size, **c** ranking by shape and **d** ranking by size and shape. **e** Cell segmentation in lung tissue. **f** ranking by size, **g** ranking by shape, **h** ranking by size and shape

$$\min \sum_{i} \mathscr{C}(\mathbf{X}_i) = \min \sum_{i,l} S_i^l(\mathbf{X}_i) = \min \sum_{i,l} \sum_{j \in N_i^A(k)} -K(\mathbf{X}_i, \mathbf{X}_j). \tag{8}$$

The minimization of the previous function may not converge to a global minima and the convergence depends of the selection of $A \subset X$. However, if A is uniformly distributed sample, the solution of the minimization problem can be very close to the global minimum.

In general, minimizing Eq. 8 is an NP-hard optimization problem, given that the number of different object-to-object combinations grows exponentially as the number of scale levels increases. To overcome this difficulty, we efficiently approximate Eq. 8 by introducing topological constrains in the number of object-to-object combinations (Fig. 8). Let c_i^l be a partition of the image at the scale level l. We introduce a pyramidal-hierarchical relation in the multilevel-partition as: $c_i^l = \bigcup c_j^{l-1}$, where c_j^{l-1} are mutually exclusive.

The optimization process utilizes the ranking information to dynamically penalize the c_i^l objects corresponding to the top ranking (outliers) at the level l. We introduce a parameter $p \in [0, 1]$ that regulates the portion of objects to further subdivide if they meet a quality criteria. Intuitively, a cell c_i^l should be subdivided if and only if all their ancestors c_j^{l-1} have a lower average cost, that is, given $c_i^l = \bigcup c_j^{l-1}$, then:

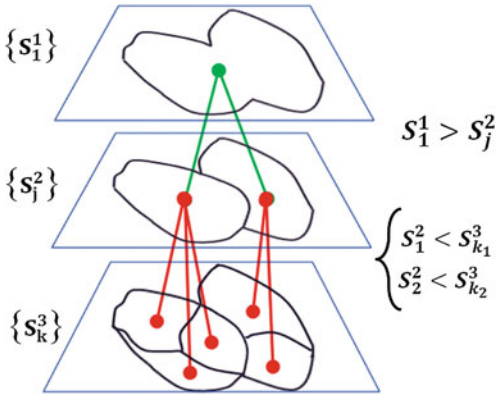


Fig. 8 Schematic of a hierarchical top-down segmentation at three different levels. The object at the highest level one the result of under segmentation, the objects in the level two correctly segmented and the objects at level three are the result of over-segmentation. Objects in level two have the lowest cost across all the levels

$$\left(\frac{\sum_{j=1}^{J} S_j^l}{J}\right) - S_i^{l-1} \leq 0. \tag{9}$$

We present our Cell Segmentation Via Shape Ranking algorithm.

Algorithm 1 *Cell Segmentation Via Shape Ranking*

1: Given a initial segmentation at a level l, select a sample (reference) $A \subset X$.
2: Using the reference set A, compute the cell similarity matrix $K(\mathbf{X}_i, \mathbf{X}_j)$, $\mathbf{X}_i \in X$, $\mathbf{X}_j \in A$ (Eq. 3).
3: Rank the cells according to N_i^A k-nearest neighbors, $A \subset X$.
4: Split cells with the lowest ranking if condition from Eq. 9 is met.
5: Go to Step 1 and iterate for level $l - 1$.

The parameter $p \in [0, 1]$ regulates the number of candidates of objects that may split across scales. If $p = 1$, all c_i^l candidates will split. If $p = 0$, the algorithm produces a shape ranking function at level l.

5 Results and Discussions

We applied our algorithm in immunohistopathology images obtained from lung and colon tissue. Dataset and image acquisition is as described in Sect. 3. Illumination correction [51] has been applied to all images. Since we are interested in segmenting epithelium cells, we utilized a pan cytokeratin marker to segment the epithelium tissue by removing the background via a minimum error thresholding algorithm [52]. The membrane protein marker used was Na+/K+-ATPase. For membrane detection, we applied the multiscale vessel enhancement filter proposed by Frangi [53] at dif-

ferent scale levels. Then the image was normalized to [0, 255]. We enhanced the membrane image by multiplying the membrane detection image with the original membrane image. To construct hierarchical partitions, the Morphological Watershed algorithm [54] was utilized at different flooding levels: 32, 16, 12, and 9. To implement our proposed shape descriptors, we used 12 bins, each one of 30°. We tested different values for the parameter p and we experimentally found that $p = 1/3$ provided good results. In our experiments the number of k-neighbors was 15.

Figure 9 shows an overlay of the membrane enhanced image with the segmentation results after applying the morphological watershed algorithm [54]. Figure 9a shows under-segmentation (flooding level 32), while in Fig. 9b under-segmentation is less evident (flooding level 16), and over-segmentation can be observed in Fig. 9c (flooding level 9). Figure 9d, shows results in a typical tissue core acquired at 20X. To visualize the results, cells are color-coded according to the ranking function from red (abnormal) to green (normal). Figure 9a–h show segmentation details.

We observed that the most homogeneous cluster corresponds to the "normal" (green) cells, which is expected since each cell minimizes the distance from its k-neighbors. Cells between red and green are not very homogenous (in overall they don't have smooth contours as cells in green) and they are relatively bigger or smaller than the green objects. The most heterogeneous cluster is the "abnormal" (red) cells. In general, this cluster contains the largest and smallest cells and the most irregular cell shapes.

Figure 10 presents the evolution of the ranking at different iterations. Figure 10a–c show the ranking evolution (best segmentations) by combining multiple scales, Fig. 10a shows the initial ranking, derived from the Watershed algorithm [54] at level 32, Fig. 10a shows the segmentation from levels 32, 16, and Fig. 10a from 32, 16, 12 and 9. Figure 10d presents the original image. Figure 10e–g present selected details from evolution of the segmentation via ranking.

Figure 11a presents a comparison of the algorithm running time as function of the percentage of the random sample size for a typical image. We compare four sampling rates: 100, 75, 50, and 25 %, all using four scale levels for the watershed algorithm: 32, 16, 12, and 9. As we can observe, the algorithm running time is reduced as the sampling rate decreases. We noted, that the execution time execution time is reduced from approximately 45 s to about 15 s, almost three times faster as compared with the full sampling. Figure 11b shows the number of detected cells in function of the sampling rate. When comparing the maximum vs. minimum number of detected cells, the difference was 1.24 %. Also, the maximum number of cells 4991, corresponded to 25 %, while the minimum number of cells 2929, corresponded to 100 %. The system used for these experiments was a machine with a 2.7 GHz processor in images with size of 2048 $\times$ 2048 and the algorithm was implemented in ITK. Figures 12 and 13 show results in colon tissue. Figures 12a and 13a depict the enhanced membrane image, while Figs. 12b and 13b depict the cell segmentation results in colon.

When viewed in context with related work [42] our results indicate the promise of modeling cell shape for extracting biologically relevant information. In future research we plan to extend the scope of shape modeling and integrate appearance models. In addition, we believe that there may be different Kernel transformations

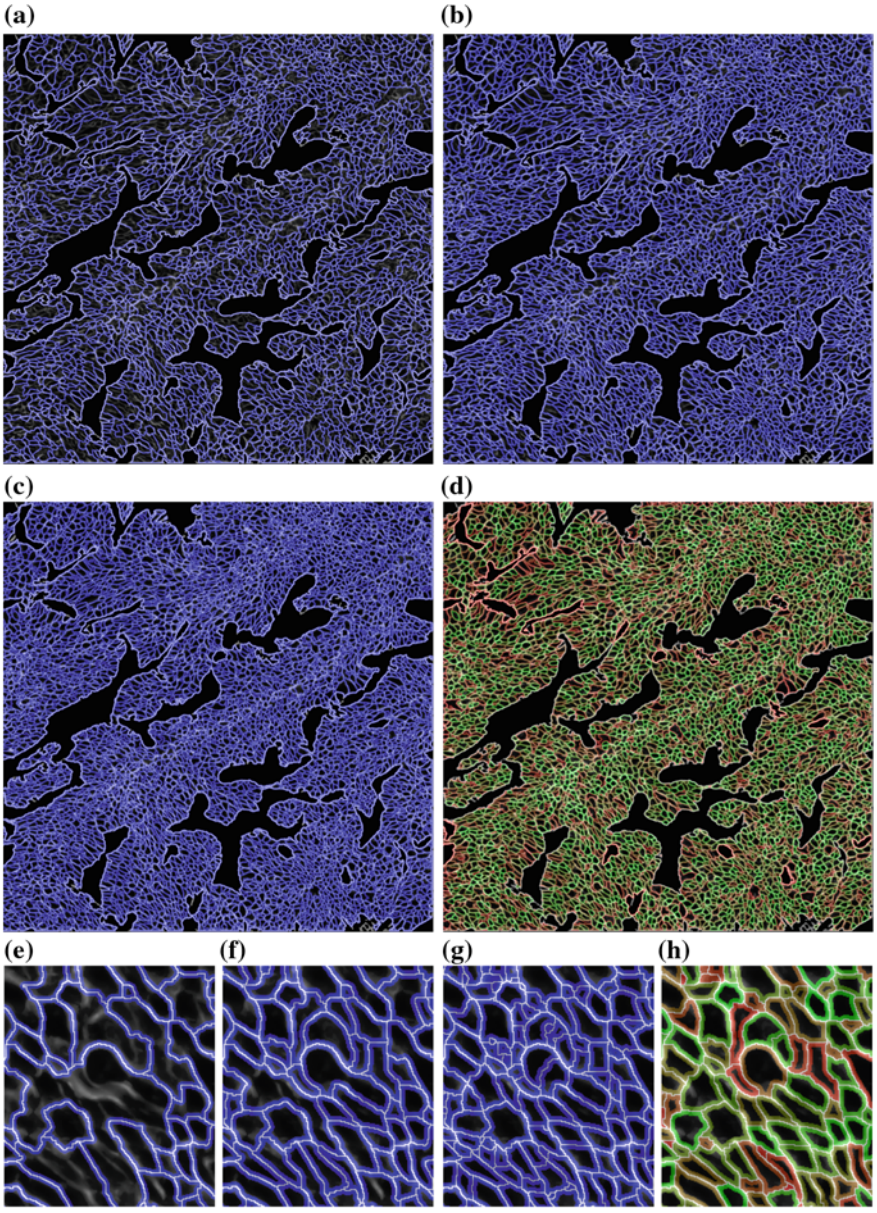


Fig. 9 Overlay of the enhanced membrane image with segmentations obtained with the morphological watershed implemented in ITK [54] at three different flooding levels: **a** high (level 32), **b** medium (level 16) and **c**, low (level 9). **d** Results from our algorithm. **e**–**h** detail of the segmentation results

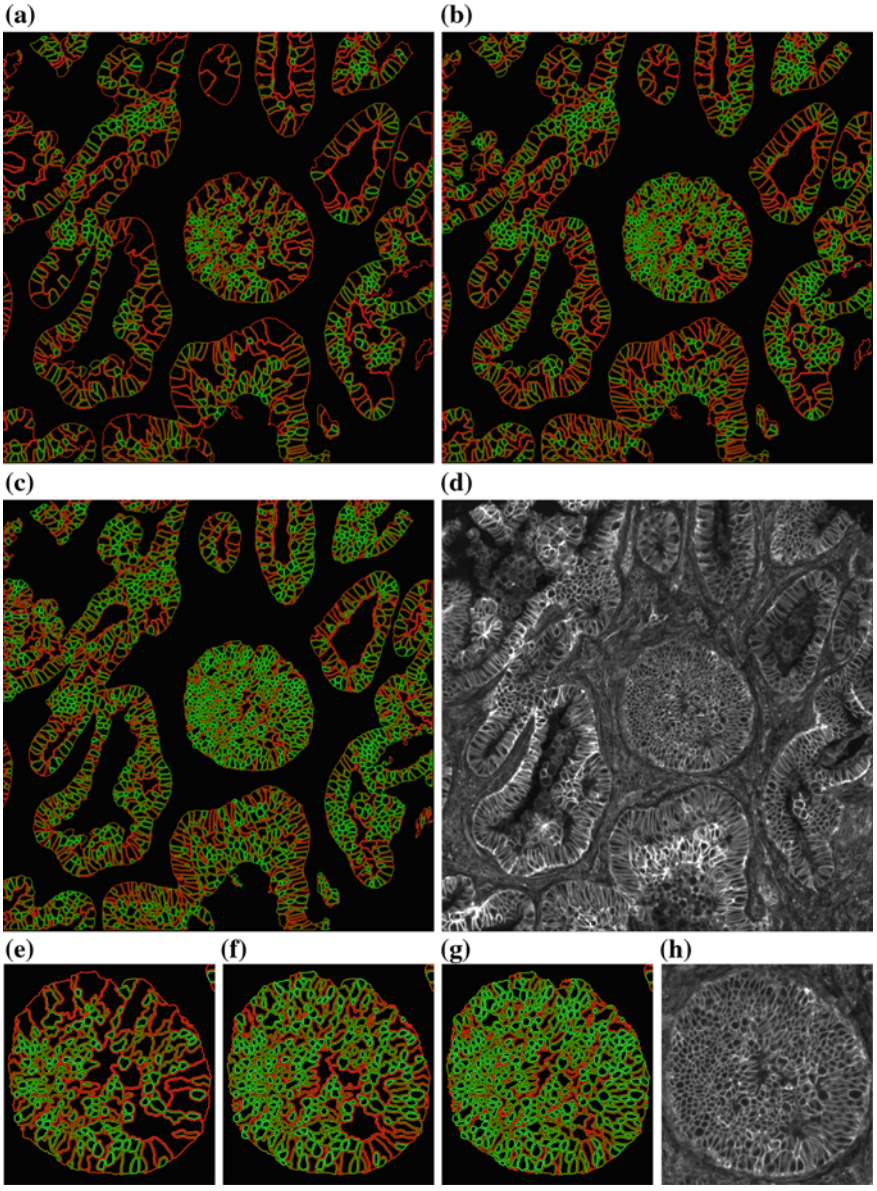


Fig. 10 Evolution of the ranking at different scale levels, from a coarser to detail scale levels derived from the ITK watershed algorithm [54]. **a** Ranking at level 32 (initialization), **b** ranking after two iterations with levels 32, 16, **c** ranking after four iterations with levels 32, 16, 12, 9, **d** original membrane image. **e**–**h** details of the previous images

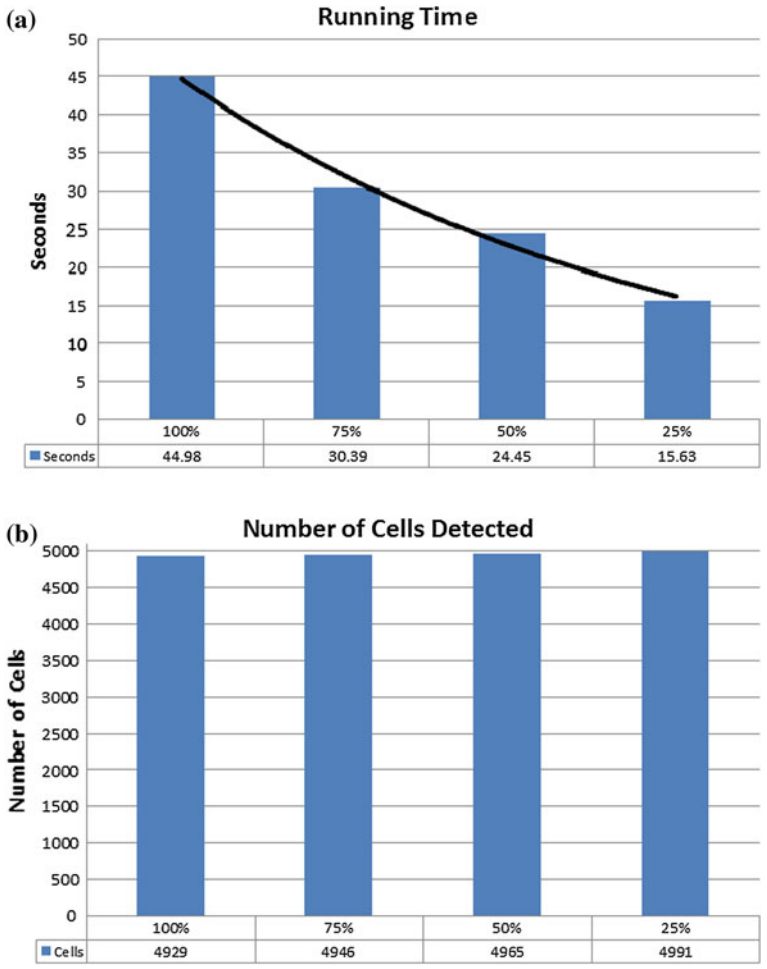


Fig. 11 Graphs corresponding to the running time and number of cells detected in a typical image. **a** Running time in seconds in function of the percentage of the percentage random sample with 100, 75, 50 and 25 %. **b** Number of cells detected in function of the percentage random sample with 100, 75, 50 and 25 %

that could be suitable for effectively modeling of the cell shape ranking (as opposed to Gaussian Kernel). Finally, the statistical analysis of sub-populations will need to be extended before the framework can be successfully applied to larger scale studies.

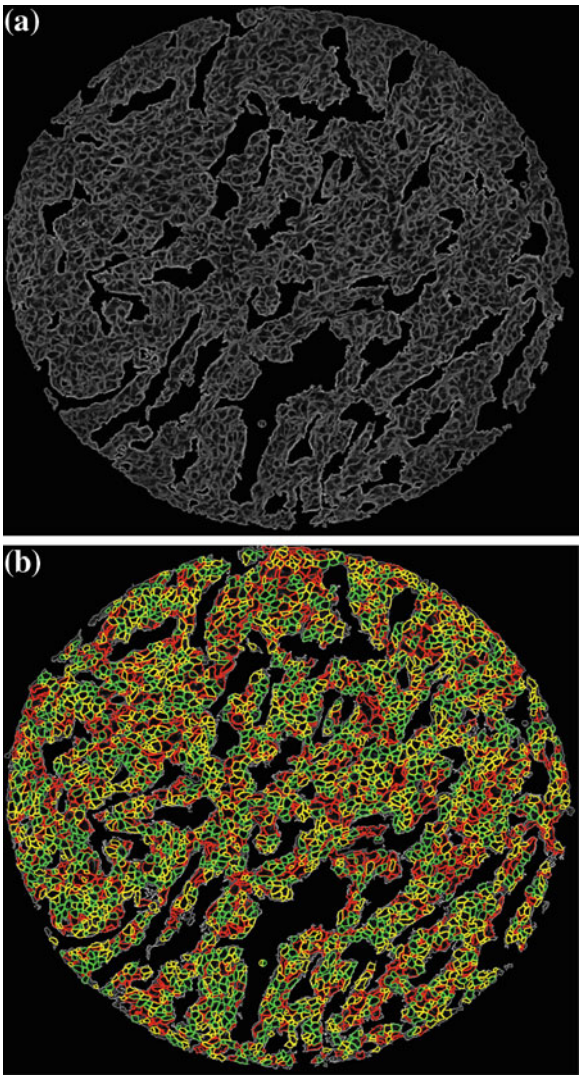


Fig. 12 Segmentation results in colon tissue. **a** Enhanced membrane image, **b** results of our segmentation algorithm. *Green*, *red* and *yellow* correspond to the most "normal", "abnormal" and "close to normal (or abnormal)" cell shape

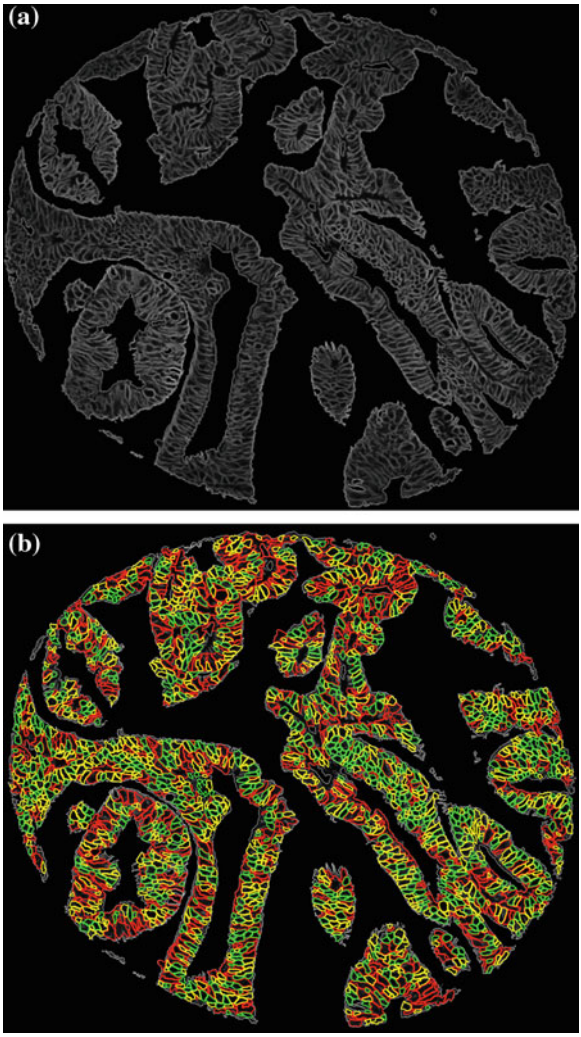


Fig. 13 Segmentation results in colon tissue. **a** Enhanced membrane image, **b** results of our segmentation algorithm. *Green*, *red* and *yellow* correspond to the most "normal", "abnormal", and "close to normal (or abnormal)" cell shape

6 Conclusion

We have developed this approach with the goal of supporting the concept of tissue based cytometry applications. This chapter presents a framework for enhancing general segmentation methods by incorporating a statistical shape model. The presented segmentation method is an extension of our previous work [48] and is approximately three times faster while keeping the fidelity of the segmentation results. Given a spe-

cific set of tissue samples it is necessary to extract as much high quality information as possible. Although MultiOmyx^{TM} methodology developed by GE is the primary use case for this technology, the overall concept is of much broader applicability [55]. The results that are being presented here clearly indicate the promise of modeling cell shape for extracting biologically relevant information.

This study focuses on a very localized analysis. In fact only the epithelial regions of the tissue are being analyzed. In the context of the lung study presented in Sect. 3 our results demonstrate that the proposed framework enables to capture a greater morphological variation of shapes. Currently we are in the process of extending this framework so that it will be possible to identify and analyze sub-populations of cells. Here it will be necessary to utilize advanced classification methods. We believe that user interactive annotation methods will be needed to capture the information for learning shape models that can deal with morphological variation in the tissue.

7 Future Work

Ultimately it will be necessary to construct methods and algorithms that can analyze the overall tissue architecture. Existing approaches for identifying certain tissue regions that rely on molecular markers are only a crude attempt to address this problem. Rather than merely extracting information on a single cell level, this information would need to be set in context with the spatial organization of the tissue itself. Although most tissue processing methods restrict ourselves to 2D processing, it might be possible to obtain full volumetric information routinely.

As this technology matures, the developed tools will be of increasing value to pathologists and life science researchers. While this is a dream at this point in time, we envision that future systems will enable practitioners to apply grading standards, as for example the Nottingham grading guideline for breast cancer, in a fully interactive fashion. There is the promise that such systems will reduce the high level of inter-operator variability. In turn they would also enable completely new applications in the area of data retrieval and data mining. Based on the extracted information the system could present a user with similar or related cases. At the same time it might be possible to highlight certain regions of the tissue for a more detailed review. This way such systems could insure that no critical areas have been omitted from the analysis.

We firmly believe that advanced feature extraction methods, sophisticated machine learning algorithms, and novel image understanding approaches will need to be developed to achieve this goal. Achieving the exceedingly high level of system performance that is necessary to enable real clinical applications is a significant challenge.

Acknowledgments This work has been developed as part of a larger interdisciplinary research program led by Fiona Ginty. In particular we would like to thank Michael Gerdes, Anup Sood, Christopher Sevinsky, and Brian Sarachan for valuable feedback. Without their collaboration it would have not been possible to evaluate the cell segmentation framework on such vast array of

tissue samples. Throughout these studies they have guided our thinking on how more robust and reliable methods could be developed. This work was performed while Yuchi Huang was in GE Global Research.

References

1. Society AC (2013) Cancer facts and figures 2013. Technical report, Atlanta, GA
2. Boyle P, Levin B (eds) (2008) World cancer report 2008. IARC Nonserial, Geneva
3. Fuller C, Straight A (2010) Image analysis benchmarking methods for high-content screen design. J Microsc 238(2):145–161
4. Hipp J, Flotte T, Monaco J, Cheng J, Madabhushi A, Yagi Y, Rodriguez-Canales J, Emmert-Buck M, Dugan M, Hewitt S, Toner M, Tompkins RG, Lucas D, Gilbertson JR, Balis U (2011) Computer aided diagnostic tools aim to empower rather than replace pathologists: lessons learned from computational chess. J Pathol Inf 2(1):25
5. Mccullough B, Ying X, Monticello T, Bonnefoi M (2004) Digital microscopy imaging and new approaches in toxicologic pathology. Toxicol Pathol 32(2 suppl):49–58
6. Meijering E (2012) Cell segmentation: 50 years down the road. IEEE Signal Process Mag 29(5):140–145
7. Fuchs TJ, Buhmann JM (2011) Computational pathology: challenges and promises for tissue analysis. Comput Med Imaging Graph 35(78): 515–530
8. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B (2009) Histopathological image analysis: a review. IEEE Rev Biomed Eng 2:147–171
9. Gleason DF, Mellinger GT (1974) Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol 111(1):58
10. Galea MH, Blamey RW, Elston CE, Ellis IO (1992) The nottingham prognostic index in primary breast cancer. Breast Cancer Res Treat 22(3):207–219
11. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, West RB, van de Rijn M, Koller D (2011) Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. Science Transl Med 3(108): 108ra113
12. Ginty F, Adak S, Can A, Gerdes M, Larsen M, Cline H, Filkins R, Pang Z, Li Q, Montalto MC (2008) The relative distribution of membranous and cytoplasmic met is a prognostic indicator in stage i and ii colon cancer. Clin Cancer Res 14(12):3814–3822
13. Gerdes MJ, Sevinsky CJ, Sood A, Adak S, Bello MO, Bordwell A, Can A, Corwin A, Dinn S, Filkins RJ, Hollman D, Kamath V, Kaanumalle S, Kenny K, Larsen M, Lazare M, Li Q, Lowes C, McCulloch CC, McDonough E, Montalto MC, Pang Z, Rittscher J, Santamaria-Pang A, Sarachan BD, Seel ML, Seppo A, Shaikh K, Sui Y, Zhang J, Ginty F (2013) Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. In: Proc Natl Acad Sci, 110(29):11982–11987, 2013
14. Nelson DA, Manhardt C, Kamath V, Sui Y, Santamaria-Pang A, Can A, Bello M, Corwin A, Dinn SR, Lazare M, Gervais EM, Sequeira SJ, Peters SB, Ginty F, Gerdes MJ, Larsen M (2013) Quantitative single cell analysis of cell population dynamics during submandibular salivary gland development and differentiation. Biol Open 2(5):439–447
15. Monaco JP, Tomaszewski JE, Feldman MD, Hagemann I, Moradi M, Mousavi P, Boag A, Davidson C, Abolmaesumi P, Madabhushi A (2010) High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models. Med Image Anal 14(4):617–629
16. Basavanhally AN, Ganesan S, Agner S, Monaco JP, Feldman MD, Tomaszewski JE, Bhanot G, Madabhushi A (2010) Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology. IEEE Trans Biomed Eng 57(3):642–653
17. Kaynig V, Fuchs T, Buhmann JM (2010) Neuron geometry extraction by perceptual grouping in ssTEM images. In: 2010 IEEE computer society conference on computer vision and pattern recognition, June 2010, pp 2902–2909

18. Doyle S, Feldman M, Tomaszewski J, Madabhushi A (2012) A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. IEEE Trans Biomed Eng 59(5):1205–1218
19. Doyle, S., Feldman, M., Tomaszewski, J., Shih, N., Madabhushi, A (2011) Cascaded multiclass pairwise classifier (cascampa) for normal, cancerous, and cancer confounder classes in prostate histology. In: 2011 IEEE international symposium on biomedical imaging: from nano to macro, pp 715–718
20. Chang J, Arbelez PA, Switz N, Reber C, Tapley A, Davis JL, Cattamanchi A, Fletcher D, Malik J (2012) Automated tuberculosis diagnosis using fluorescence images from a mobile microscope. In: Ayache N, Delingette H, Golland P, Mori K (eds) MICCAI (3), Volume 7512 of Lecture notes in computer science. Springer, pp 345–352
21. Nandy K, Gudla PR, Amundsen R, Meaburn KJ, Misteli T, Lockett SJ (2012) Automatic segmentation and supervised learning-based selection of nuclei in cancer tissue images. Cytometry Part A J Int Soc Anal Cytol 81(9):743–754
22. Roullier V, Lzoray O, Ta VT, Elmoataz A (2011) Multi-resolution graph-based analysis of histopathological whole slide images: application to mitotic cell extraction and visualization. Comput Med Imaging Graph 35(78): 603–615 ¡ce:title¿Whole Slide Image Process¡/ce:title¿
23. Nath SK, Palaniappan K, Bunyak F (2006) Cell segmentation using coupled level sets and graph-vertex coloring. Med Image Comput Computer-Assist Interv 9(Pt 1):101–108
24. Lin G, Chawla MK, Olson K, Barnes CA, Guzowski JF, Bjornsson C, Shain W, Roysam B (2007) A multi-model approach to simultaneous segmentation and classification of heterogeneous populations of cell nuclei in 3D confocal microscope images. Cytometry Part A 71A:724–736
25. Ta VT, Lzoray O, Elmoataz A, Schpp S (2009) Graph-based tools for microscopic cellular image segmentation. Pattern Recognit 42(6):1113–1125
26. Keuper M, Schmidt T, Rodriguez-Franco M, Schamel W, Brox T, Burkhardt H, Ronneberger O (2011) Hierarchical Markov random fields for mast cell segmentation in electron microscopic recordings. In: IEEE international symposium on biomedical, imaging, pp 973–978
27. Can A, Bello M, Gerdes M (2010) Quantification of subcellular molecules in tissue microarray. In: 2010 20th international conference on pattern recognition (ICPR), pp 2548–2551
28. Can A, Bello M, Cline HE, Tao X, Mendonca P, Gerdes M (2009) A unified segmentation method for detecting subcellular compartments in immuno-fluorescently labeled tissue images. In: Proceedings of the 2009 international workshop in microscopy image analysis with applications in biology, Bethesda, MD, Sept 2009
29. Cukierski WJ, Nandy K, Gudla PR, Meaburn KJ, Misteli T, Foran DJ, Lockett SJ (2012) Ranked retrieval of segmented nuclei for objective assessment of cancer gene repositioning. BMC Bioinform 13:232
30. Arslan S, Ersahin T, Cetin-Atalay R, Gunduz-Demir C (2013) Attributed relational graphs for cell nucleus segmentation in fluorescence microscopy images. IEEE Trans Med Imaging 32(6):1121–1131
31. Lou X, Koethe U, Wittbrodt J, Hamprecht FA (2012) Learning to segment dense cell nuclei with shape prior. In: 2012 IEEE conference on computer vision and, pattern recognition, June 2012, pp 1012–1018
32. Bamford P (2003) Empirical comparison of cell segmentation algorithms using an annotated dataset. In: Proceedings of the 2003 international conference on image processing, 2003 (ICIP 2003), vol 2, Sept 2003, II - 1073–6 vol. 3
33. Na S, Heru X (2009) The segmentation of overlapping milk somatic cells based on improved watershed algorithm. In: International conference on artificial intelligence and computational intelligence, AICI '09, vol 3, Nov 2009, pp 563–566
34. Srinivasa G, Fickus M, Guo Y, Linstedt A, Kovacevic J (2009) Active mask segmentation of fluorescence microscope images. IEEE Trans Image Process 18(8):1817–1829
35. Xiao Y, Pham T, Chang J, Zhou X (2011) Symmetry-based presentation for stem-cell image segmentation. In: 2011 IEEE 1st international conference on computational advances in bio and medical sciences (ICCABS), Feb 2011, pp 196–201

36. Martinez G, Frerichs JG, Scheper T (2011) Bubble segmentation based on shape from shading for in-situ microscopy. In: 2011 21st international conference on electrical communications and computers (CONIELECOMP), 28 Feb 2011–2 March 2011, p 1–4
37. Kong H, Gurcan M, Boussaid K (2011) Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting. IEEE Trans Med Imaging PP(99):1
38. Xiong W, Wang Y, Ong S, Lim JH, Jiang L (2010) Learning cell geometry models for cell image simulation: an unbiased approach. In: 2010 17th IEEE international conference on image processing (ICIP), Sept 2010, pp 1897–1900
39. Park M, Jin J, Peng Y, Summons P, Yu D, Cui Y, Luo S, Wang F, Santos L, Xu M (2010) Automatic cell segmentation in microscopic color images using ellipse fitting and watershed. In: 2010 IEEE/ICME international conference on complex medical engineering (CME), July 2010, pp 69–74
40. Zacharia E, Maroulis D (2010) 3-D spot modeling for automatic segmentation of cDNA microarray images. IEEE Trans Nanobiosci 9(3):181–192
41. Lempitsky V, Zisserman A (2010) Learning to count objects in images. Adv Neural Inf Process Syst 23:1324–1332
42. Arteta C, Lempitsky V, Noble J, Zisserman A (2012) Learning to detect cells using non-overlapping extremal regions. In: Medical image computing and computer assisted intervention, pp 348–356
43. Sood A, Montalto M, Gerdes M (2009) Sequential analysis of biological samples. USPTO Application #11/560,599. General Electric Company, United States of America
44. Gerdes MJ, Ginty F, Larsen M, Montalto M, Pang Z, Sood A (2010) Sequential analysis of biological samples. USPTO Patent #7741045. General Electric Company, United States of America
45. Larsen M, Sood A, Gerdes M, Montalto M, Pang Z, Ginty F (2010) Sequential analysis of biological samples. USPTO Application #11/864093. General Electric Company, United States of America
46. Treynor T, Sood A, Gerdes M, Pang Z (2012) Sequential analysis of biological samples. General Electric Company, United States of America
47. Rittscher J (2010) Characterization of biological processes through automated image analysis. Annu Rev Biomed Eng 12:315–44
48. Santamaria-Pang A, Huang Y, Rittscher J (2013) Cell segmentation and classification via unsupervised shape ranking. In: 10th IEEE international symposium on biomedical imaging: from nano to macro, 2013, (ISBI 2013). San Francisco, CA, April 2013
49. Mori G, Belongie S, Malik J (2005) Efficient shape matching using shape contexts. IEEE Trans Pattern Anal Mach Intell 27(11):1832–1837
50. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, CVPR'2005
51. Can A, Bello M, Cline H, Tao X, Mendonca P, Gerdes M (2009) A unified segmentation method for detecting subcellular compartments in immunofluorescently labeled tissue images. In: Microscopic image analysis with applications in biology
52. Kittler J, Illingworth J (1986) Minimum error thresholding. Pattern Recogn 19(1):41–47
53. Frangi AF, Niessen WJ, Vincken KL, Viergever MA (1998) Multiscale vessel enhancement filtering. In: MICCAI, Springer, pp 130–137
54. Ibanez L, Schroeder W, Ng L, Cates J (2005) The ITK software guide, 2nd edn. Kitware, Inc. ISBN 1-930934-15-7, http://www.itk.org/ItkSoftwareGuide.pdf.
55. Stewart E, Rittscher J, Sebastian TB (2007) Automatic characterization of cellular motion. US Patent #20080304732. General Electric Company, United States of America

Computational Modeling of the Spine

Tobias Klinder, Samuel Kadoury and Cristian Lorenz

Abstract Modeling of the human spine requires the extension from single object modeling to object ensembles. The spine consists of a constellation of vertebrae where the individual vertebrae show a complex shape. While most neighbouring vertebrae look very similar, their shape changes significantly along the spine. Due to these challenges, more sophisticated model formulations are needed that go beyond shape modeling of vertebrae. In this article, we combine several high-level models of the spine into one common framework. The individual vertebrae are represented as a set of models covering shape, gradient and appearance information as well as relative location and orientation. By encoding further anatomical information into the shape representation of the individual vertebrae, e.g., important anatomical regions or significant landmarks, clinically relevant parameters can be easily derived from the shape models. The spine is expressed as a sequence of rigid transformations between vertebrae and different statistical methods can be used to cover the variability of spinal curvatures. For selected applications that are vertebra labelling in limited field of view scans and segmentation in both CT and MRI, we show how this comprehensive framework can be used for an automatic image interpretation of medical images of the spine. Furthermore, the problem of change assessment for osteoporotic fracture detection is tackled with this framework as an example for CAD.

T. Klinder (✉) · C. Lorenz
Philips Research Europe-Hamburg, Röntgenstrasse 24-26, 22335 Hamburg, Germany
e-mail: tobias.klinder@philips.com

C. Lorenz
e-mail: cristian.lorenz@philips.com

S. Kadoury
Ecole Polytechnique de Montreal, 6079, Succursale Centre-Ville, Montreal, QCH3C 3A7, Canada
e-mail: samuel.kadoury@polymtl.ca

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_11,

1 Introduction

The spine represents both a vital central axis for the musculoskeletal system and a flexible protective shell surrounding the most important neural pathway in the body, the spinal cord. Spine related diseases or conditions are common and cause a huge burden of morbidity and cost to society. As an example, back pain is the second most common neurological ailment in the United States after headache. For many spine related diseases, imaging is required for diagnosis ranging from applications that include plain radiographs, CT, MR, ultrasound but also nuclear medicine. To assist the diagnosis, automatic image interpretation is desired.

Usually, the spine consists of 24 vertebrae grouped as cervical (C1–C7), thoracic (T1–T12) and lumbar (L1–L5) vertebrae. Thus, compared to other anatomical structures, the spine poses additional challenges as it consists of a constellation of vertebrae where the individual vertebrae show a complex shape. Although the shape of the individual vertebrae changes significantly along the spine, most neighbouring vertebrae look very similar and are difficult to distinguish.

In recent years, a variety of approaches has been presented in the context of spine related image analysis for various applications, such as, degenerative disc disease [1], osteoporosis [2, 29], scoliosis [8], fracture [40, 41], spine metastasis [15, 38, 39], or spondyloarthropathy [37]. Depending on the clinical application and target modality, different methods have been proposed. For most applications based on CT, segmentation of vertebrae is crucial and region-growing [3], graph-cuts [4, 5] and level-sets [2] were applied. In order to guide the segmentation prior knowledge was included, e.g., vertebra shape representations consisting of geometrical primitives [42]. Initialisation of the segmentation method was either not addressed, e.g., [2–4] or automatically determined, e.g., by first extracting the spine and then detecting the individual vertebrae [42]. In the context of scoliosis analysis, X-ray or more specifically bi-planar X-ray is usually the modality of choice. Particular efforts were made for the extraction of the three-dimensional spine curve instead of analyzing vertebrae individually. For the modeling of the global spine curve and its variability, a representation of the spine as an articulated model build from rigid transformations between neighboring vertebrae is often used [8, 9, 27, 43]. Another common approach for modeling the spine is by means of graph representations as done in [1, 11, 36].

While a lot of work has been done over the last decade, most solutions were very problem specific. In many cases, only dedicated parts of the spine were considered, e.g., lumbar spine as in [11, 31], and it is thus unclear how such methods can then be applied to other regions of the spine. In this article, we introduce a comprehensive framework for computational modeling of the whole spine integrating several high-level models for the individual vertebrae as well as the global spine curve. For the individual vertebrae, we propose a set of models covering shape, gradient and appearance information as well as relative location and orientation. By encoding further anatomical information into the detailed shape representation of the individual vertebrae, e.g., important anatomical regions or significant landmarks, clinically relevant

parameters can be easily derived from the shape models. The spine is expressed as a sequence of rigid transformations between vertebrae and different statistical methods can be applied to model the variability of spinal curvatures. These methods can be used to perform model adaptation to a new instance. Compared to other work, we are able to consider global spine curvature and local vertebra shape simultaneously. This article combines and unifies our recent work done seperately in two different groups meaning [16–21] and [23, 24, 29]. The proposed framework can be applied to a variety of applications. Examplarily, we selected three different clinical challenges: (i) vertebra labelling in limited field of view scans, (ii) segmentation in both CT and MRI and (iii) change assessment for osteoporotic fracture detection.

The reminder of this article is structured as follows. In Sect. 2, we introduce different models for the individual vertebrae, while Sect. 3 focusses on the articulated model to express the global spine curve. For the selected applications, results are given in Sect. 4. Finally, we conclude in Sect. 5.

2 Vertebra Modelling

Anatomical modeling by means of shape modeling has proven to be extremely beneficial for many problems in medical image analysis. However, modeling of the human spine requires the extension from single object modeling to object ensembles. Since the shape of the vertebrae changes along spine, we first introduce detailed shape representations for each of the 24 pre-sacral vertebrae in Sects. 2.1 and 2.2. In addition to the shape representation, we define for each vertebrae a coordinate system to express position and orientation in Sect. 2.3. This information can then be used to later model the whole spine shape. Finally, each vertebra model is further extended by adding gradient as explained in Sect. 2.4 enabling vertebra detection and appearance information as introduced in Sect. 2.5 to allow for vertebra identification.

2.1 Model Geometry

A first surface representation of the vertebrae was obtained by scanning of cadavers [17] or commercially available plastic phantoms [24] with a CT scanner. Out of the scanned data set, the individual vertebrae can be firstly segmented using simple thresholding and afterwards triangulated using the marching cubes algorithm [28]. Finally, mesh operations can be applied like surface smoothing, vertex insertion in the case of too large triangles, or edge deletion in the case of too small edges resulting in a smooth surface with a roughly uniform distribution of vertices. Thus, each of the 24 surface mesh representation $\mathbf{m}_i$ consists of V_i vertices $\{\mathbf{v}_j^i | j = 1, \ldots, V_i\}$ and T_i triangles.

Further prior knowledge can be encoded into the surface meshes by labelling the individual triangles. An example is given in Fig. 1a. This labelling allows to easily

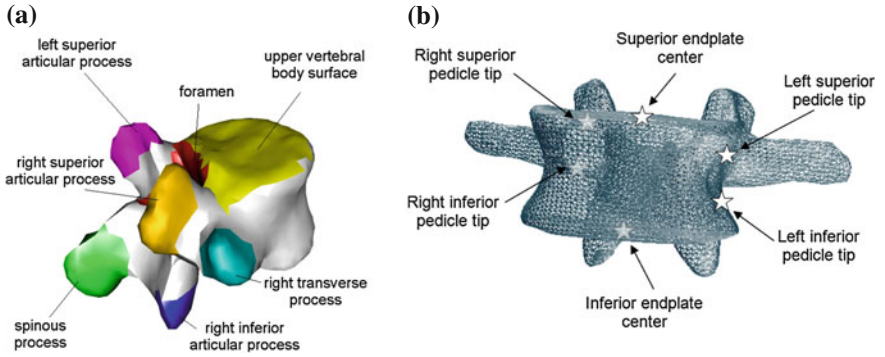


Fig. 1 Illustration of additional prior knowledge encoded into the shape models, **a** Labelling of anatomical regions. **b** Annotated anatomical landmarks

derive certain vertebra characteristics directly from the mesh, e.g., distance between vertebral body surfaces. Finally, important anatomical landmarks can be annotated in the mesh as shown in Fig. 1b. As both surface labelling and landmark setting only need to be done once during model generation, manual annotation is typically done.

The surface descriptions can be embedded into a multi-resolution representation with different levels of complexity. In [20], three levels of polygonal mesh catalogues were created to speed up processing and allow for optimal convergence using a coarse-to-fine strategy during registration.

2.2 Statistical Shape Modelling

From the initial vertebra surface representations, we can build statistical vertebra shape models by adapting the surfaces to a set of images. During patient individualization, we carefully preserve the established vertex distribution of the generated vertebra surface representation. By using the adaptation algorithm from Chapter 5, the topology remains unchanged and the vertex distribution is constrained to be similar to the respective reference mesh, so that the jth landmark of the ith mesh lies at approximately the same anatomical position for all cases. After segmentation of all vertebrae in a set of training images, a Procrustes co-registration is performed and finally a point distribution model can be obtained. In the following, we keep however only the mean surface models $\mathbf{s}_i$ for each vertebra. Thus, the overall geometrical spine model can be expressed as $\mathbf{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_L\}$ with $L = 24$ if all vertebrae are considered.

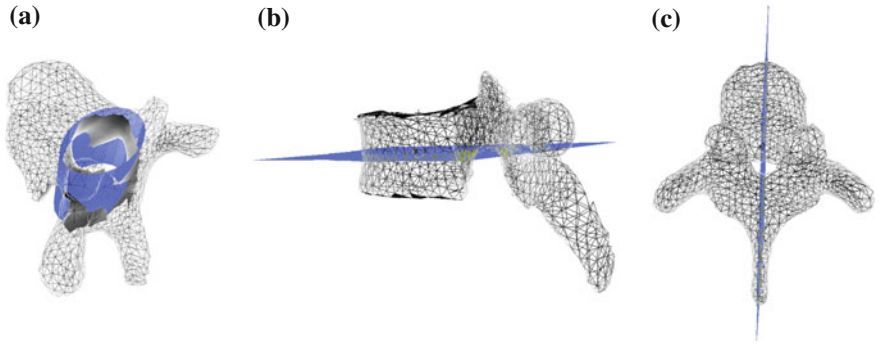


Fig. 2 Invariant features of vertebrae used to define a VCS: Cylinder fit to vertebral foramen (**a**), middle plane of upper and lower vertebral body surface (**b**) as well as sagittal symmetry plane (**c**)

2.3 Vertebra Coordinate System

In addition to the surface representation, each vertebra model carries its own local vertebra coordinate system (VCS). This allows to express relative location and orientation of the individual vertebrae. In order to have similar definition of the VCS for the individual vertebrae along the spine, a VCS is typically defined by invariant object characteristics and different definitions have been proposed. In [23], we derived the VCS of all vertebra from the surface mesh representation including the labelling of anatomical regions (see Fig. 1a). For the vertebrae C3 to L5, the VCS was defined in the same way since these vertebrae show similar shape characteristics. The definition of the VCS was based on three object-related simplified representations: a cylinder fit to the vertebral foramen, the middle plane of the upper and lower vertebral body surfaces, and the sagittal symmetry plane of the vertebra (see Fig. 2). The definition of the VCS was as follows: The origin of the VCS was located in the mid-vertebral plane at the centre of the vertebral foramen. The normal vector of the middle plane defined the $\mathbf{z}_{\mathrm{VCS}}$-axis, the $\mathbf{x}_{\mathrm{VCS}}$-axis was defined by the orthogonal component of the normal vector of the symmetry plane and the $\mathbf{z}_{\mathrm{VCS}}$-axis. The cross product of $\mathbf{z}_{\mathrm{VCS}}$-axis and $\mathbf{x}_{\mathrm{VCS}}$-axis resulted in the $\mathbf{y}_{\mathrm{VCS}}$-axis [23]. For the unique definition of the VCS for C1 and C2, other characteristics were chosen. A vertex subset was chosen containing the vertices belonging to the inferior articular faces and the anterior tubercle. The axes of the specific VCSs are given by the covariance analysis of the subset. The eigenvector corresponding to the largest eigenvalue of the covariance matrix points to the lateral direction defining the $\mathbf{x}_{\mathrm{VCS}}$-axis, the second eigenvector pointing towards the anterior tubercle gives the $\mathbf{y}_{\mathrm{VCS}}$-axis, and finally, the third eigenvector points towards the direction of the spinal canal denoting the $\mathbf{z}_{\mathrm{VCS}}$-axis. The barycentre of the subset defines the origin of the VCS.

Alternatively, the VCS for the thoracic and lumbar vertebrae was derived in [21] from six anatomical landmarks defined in the mesh (4 pedicle tips and 2 on the vertebral body, as depicted in Fig. 1b).

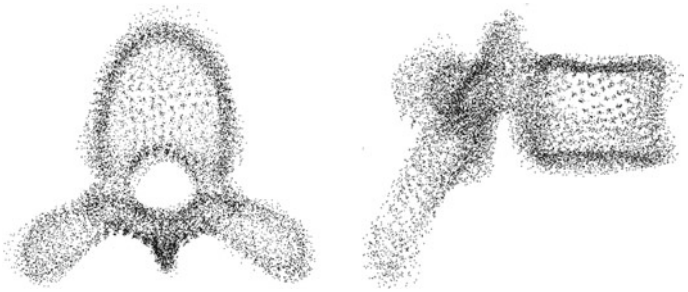


Fig. 3 Gradient model representation of sixth thoracic vertebra in axial and sagittal view showing the model points. The model has been created from ten samples

2.4 Gradient Model

For vertebra detection, we further created generalised Hough transform models of each vertebra. The generalised Hough transform (GHT) [6] is a robust and powerful method to detect arbitrary shapes in an image undergoing geometric transformations. During GHT learning, description of the shape is encoded into a reference table also called R-table. Its entries are vectors pointing from the shape boundary to a reference point being commonly the gravity centre of the shape. During detection, the gradient orientation is measured at each edge voxel of the new image yielding an index for an entry of the R-table. Then, the positions pointed by all vectors under this entry are incremented in an accumulator array. Finally, the shape is given by the highest peak in the accumulator array. Usually, the GHT is trained for one single reference shape of an object class. In addition to specific GHT models for C1 to L5 we build general cervical, thoracic and lumbar GHT models. The general models can be applied if the field of view of the available image is unknown while the more specific models can be used if some indication for a certain area of the spine is given.

For model generation, shape models adapted to a set of training images are superimposed based on the VCSs and the obtained aligned shapes are used to build up the R-table. Note that the orientation is not considered during GHT learning since all shapes were aligned via the respective VCS. In contrast to the above mentioned shape models, the GHT models do not only contain shape information but also gradient information. A gradient model for the sixth thoracic vertebra can be seen in Fig. 3. For more details about the gradient models, we refer to [24].

2.5 Appearance Model

Vertebra appearance models in this case are represented as average intensity volumes inside a bounding box around each vertebra without explicitly modeling the vertebra's shape. The information covered in the model about expected appearance in the

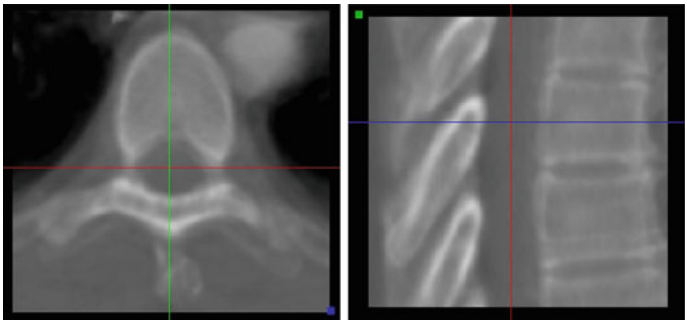


Fig. 4 Appearance model of sixth thoracic vertebra in axial and sagittal view generated from ten training images. The volume has a sample distance of 0.5 mm in each direction. The cross gives the origin of the VCS

direct neighbourhood of each vertebra can be used to identify vertebrae in an image by comparing the similarity between model and image content. For model generation, shape models need to be adapted to a set of training images. Mean intensity volumes are then generated by aligning volumes around corresponding vertebrae and averaging the intensity values. Alignment is performed by registration of the vertebra shape models to their respective model space. The obtained transformations can then be applied on the respective voxel positions. Location and pose of the object in the volumes are expressed by the VCSs of the mean vertebra models that were used for alignment. Exemplarily, the appearance model of the sixth thoracic vertebra is shown in Fig. 4. Inter-patient variability becomes obvious by blurry object boundaries whereas the particular objects are still clearly delimited. More details can be found in [24].

3 Articulated Models

In the previous section, each vertebra has been represented seperately by a set of detailed models. However, in order to model the overall constellation of the spine, the articulation between the individual vertebrae has to be considered. In this section, we first introduce a general representation of the articulated spine model in Sect. 3.1. Based on this representation, statistical modeling of the spine can be done and two different methods are proposed in Sects. 3.2 and 3.3. In each case, it is furthermore shown how the articulated model can be individualized to a new patient.

3.1 Model Definition

As defined in Sect. 2, the geometric model of the spine is defined as $\mathbf{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_L\}$ consisting of an interconnection of L objects. Each local shape $\mathbf{s}_i$ is represented as a triangulated mesh including a VCS to rigidly register each object to its upper neighbor.

The resulting rigid transforms are stored for each inter-object link. Hence, an articulated deformable model (ADM) is represented by a series of local inter-object rigid transformations T_i (translation and rotation) between neighbouring vertebrae resulting in

$$\mathbf{A} = [T_1, T_2, \ldots, T_{L-1}]\,. \tag{1}$$

While $\mathbf{A}$ covers the individual inter-object transformations, the global shape of the spine shape can be expressed by converting $\mathbf{A}$ into an absolute representation

$$\mathbf{A}_{\text{abs}} = [T_1, T_1 \circ T_2, \ldots, T_1 \circ T_2 \circ \ldots \circ T_{L-1}] \tag{2}$$

using recursive compositions. The respective transformations are expressed in the VCS of the lower object.

The relationship between the shape model $\mathbf{S}$ and the ADM is that the articulation vector $\mathbf{A}$ controls the position and orientation of the object constellation $\mathbf{S}$. The ADM can achieve deformation by modifying the vector of rigid transformations, which taken in its entirety, performs a global deformation of the spine [22].

The rigid transformations are the combination of a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$, while scaling is not considered. We formulate the rigid transformation $T = \{\mathbf{R}, \mathbf{t}\}$ of a triangular mesh model as $\mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{t}$ where $\mathbf{x}, \mathbf{y}, \mathbf{t} \in \Re^3$. Composition is given by $T_1 \circ T_2 = \{\mathbf{R}_1\mathbf{R}_2, \mathbf{R}_1\mathbf{t}_2 + \mathbf{t}_1\}$.

Alternatively, a rigid transformation can be fully described by an axis of rotation supported by a unit vector $\mathbf{u}$ and an angle of rotation θ. The rotation vector $\mathbf{r}$ is defined as the product of $\mathbf{u}$ and θ. Thus, a rigid transformation can be expressed as $T = \{\theta\mathbf{n}, \mathbf{t}\} = \{\mathbf{r}, \mathbf{t}\}$. The conversion between the two representations is simple, since the rotation vector can be converted into a rotation matrix using Rodrigues' formula (see Appendix for more details).

3.2 Linear Model

3.2.1 Model Generation

In order to capture variations of the spine curve, statistics need to be applied on the articulated model. As defined in the previous section, the spine is expressed as a vector of rigid transformation and there is no addition or scalar multiplication defined between them. Thus, as the representation no longer belongs to a vector space,

conventional statistics can not be applied. However, rigid transformations belong to a Riemannian manifold and Riemannian geometry concepts can be efficiently applied to generalize statistical notions to articulated shape models of the spine. The mathematical framework described in the following has been introduced by Boisvert et al. and explained, e.g., in [8–10].

Following the statistics for Riemannian manifolds, the generalization of the usual mean called the Fréchet mean can be defined for a given distance measure as the element μ that minimizes the sum of the distances with a set of elements $x_0, \ldots, x_N$ of the same manifold $\mathcal{M}$:

$$\mu = \arg\min_{x \in \mathcal{M}} \sum_{i=0}^{N} d(x, x_i)^2 \,. \tag{3}$$

Since the mean is given as a minimization, the gradient descent method can be performed on the summation obtaining

$$\mu_{n+1} = \mathrm{Exp}_{\mu_n}\left(\frac{1}{N}\sum_{i=0}^{N} \mathrm{Log}_{\mu_n}(x_i)\right). \tag{4}$$

The functions Exp and Log are respectively the exponential map and the log map associated with the distance $d(x, y)$. The exponential map projects an element of the tangent plane on the manifold $\mathcal{M}$ and the log map is the inverse function.

The intuitive generalization of the variance is the expectation of the squared distance between the elements x_i:

$$\sigma^2 = E[d(\mu, x)^2] = \frac{1}{N}\sum_{i=0}^{N} d(\mu, x_i)^2. \tag{5}$$

The covariance is usually defined as the expectation of the matricial product of the vectors from the mean to the elements on which the covariance is computed. The expression for Riemannian manifolds can thus be given by using again the Log map [8]:

$$\Sigma_{xx} = E[\mathrm{Log}(\mu^{-1}x)^T \mathrm{Log}(\mu^{-1}x)] = \frac{1}{N}\sum_{i=0}^{N} \mathrm{Log}_{\mu}(x)\mathrm{Log}_{\mu}(x)^T. \tag{6}$$

To use the Riemannian framework defined by Eqs. 3–6, a suitable distance needs to be defined first to find the structure of the geodesics on the manifold. For this purpose, we rely on the representation of transformation as a rotation vector as defined in Sect. 3.1. With this representation, the left-invariant distance definition between two rigid transformations can be defined [9]:

$$d(T_1, T_2) = N_\omega(T_2^{-1} \circ T_1) \quad \text{with} \quad N_\omega(T)^2 = N_\omega(\{\mathbf{r}, \mathbf{t}\})^2 = \|\mathbf{r}\|^2 + \|\omega\mathbf{t}\|^2 \tag{7}$$

where ω is used to weight the relative effect of rotation and translation, $\mathbf{r}$ is the rotation vector, and $\mathbf{t}$ the translation vector. An appropriate value for the weight factor ω in the case of the spine is found to be 0.05 [8].

Finally, exponential and logarithmic map associated with the defined distance are the conversions between the rotation vector and the rotation matrix combined with a scaled version of the translation vector [8]

$$\mathrm{Exp}_{Id}(T) = \begin{vmatrix} \mathbf{R}(r) \\ \omega^{-1}\mathbf{t} \end{vmatrix} \quad \text{and} \quad \mathrm{Log}_{Id}(T) = \begin{vmatrix} r(\mathbf{R}) \\ \omega\mathbf{t} \end{vmatrix} . \tag{8}$$

Given the articulated model of the spine is expressed as the multivariate vector $\mathbf{A}^T = [T_1, T_2, \ldots, T_L]^T$ from Eq. 1 of L rigid transformations, we obtain for the mean and the covariance

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_L \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_{T_1T_1} & \Sigma_{T_1T_2} & \ldots & \Sigma_{T_1T_L} \\ \Sigma_{T_2T_1} & \Sigma_{T_2T_2} & \ldots & \Sigma_{T_2T_L} \\ \vdots & \vdots & & \vdots \\ \Sigma_{T_LT_1} & \Sigma_{T_LT_2} & \ldots & \Sigma_{T_LT_L} \end{pmatrix} . \tag{9}$$

With Eq. 9, the formulation for the linear statistical spine model is given. In order to reduce the dimensionality of the model, principal component analysis (PCA) can be performed on the covariance matrix. As an example, Fig. 5 shows the first three eigenmodes after applying PCA to the statistics of transformations between the 12 thoracic vertebrae obtained from 18 patients.

3.2.2 Model Adaption

The statistical spine model can be used to support image segmentation. It allows to consider the object constellation during segmentation instead of separate segmentation of the individual vertebrae, thus going from single object to multi-object segmentation. Such a framework has been presented in [25] for the segmentation of the spine in CT. Its general idea is to follow a two-scale adaptation approach. During a global adaptation step, the individual vertebrae are at first roughly positioned in the image. Afterwards, exact segmentation of the vertebrae is achieved by adaptation of the respective vertebra shape models. In this section, we focus on the adaptation of the global model, while details for deformable surface mesh adaptation are basically explained earlier in Chap. 5.

The idea of global adaptation is to determine the inter-object transformations T_k by minimizing an energy term $E(T_k)$ where an external energy $E_{\mathrm{ext}}(T_k)$ drives the vertebrae towards image features, while an internal energy $E_{\mathrm{int}}(T_k)$ restricts attraction to a former learned constellation model

$$E(T_k) = E_{\mathrm{ext}}(T_k) + \alpha E_{\mathrm{int}}(T_k) . \tag{10}$$

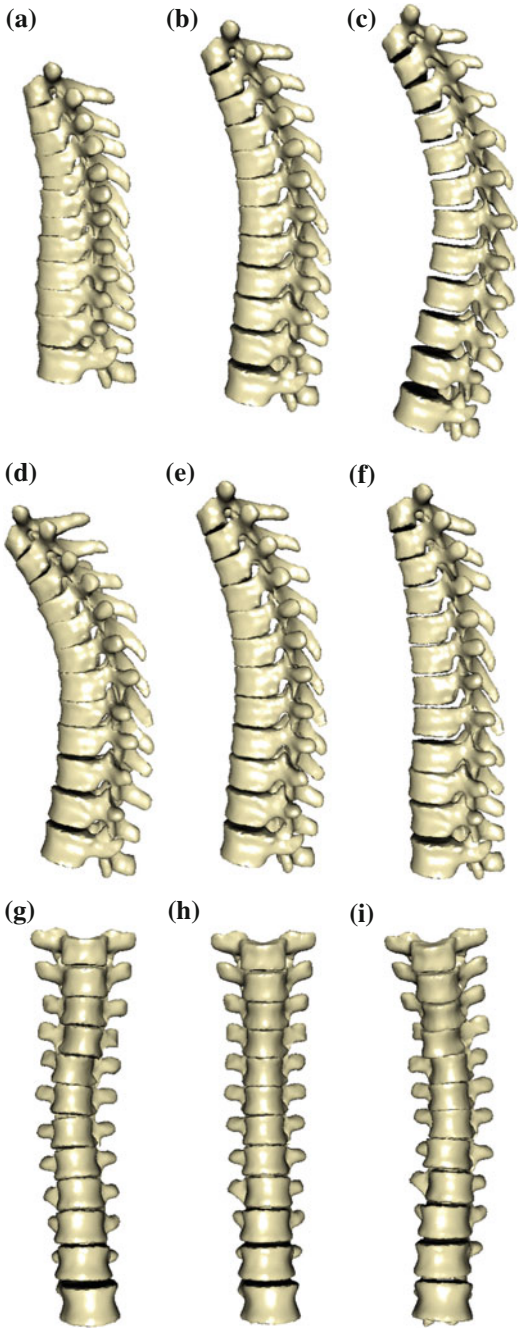


Fig. 5 First, second and third eigenmode of the statistical model on transformation between the local coordinate systems defined on the 12 thoracic vertebrae. Each row shows the mean transformation model μ and the mean plus a particular eigenmode weighted by $\pm 3\sqrt{\lambda_i}$ with $i = 1, ..., 3$. The resulting transformations are applied to mean shape models based on 18 patient data sets, **a** $\mu - 3\sqrt{\lambda_1}\phi_1$, **b** Mean μ, **c** $\mu + 3\sqrt{\lambda_1}\phi_1$, **d** $\mu - 3\sqrt{\lambda_2}\phi_2$, **e** Mean μ, **f** $\mu + 3\sqrt{\lambda_2}\phi_2$, **g** $\mu - 3\sqrt{\lambda_3}\phi_3$, **h** Mean μ, **i** $\mu + 3\sqrt{\lambda_3}\phi_3$

The parameter α controls the trade-off between both energy terms which are explained in the following.

The external energy attracts each vertebra towards its corresponding image structures. Thus, we first need to define a feature function such as

$$F_i(\mathbf{x}_i) = -\mathbf{n}_i{}^T \nabla I(\mathbf{x}_i) \frac{g_{\max}(g_{\max} + \|\nabla I(\mathbf{x}_i)\|)}{g_{\max}^2 + \|\nabla I(\mathbf{x}_i)\|^2} \tag{11}$$

that evaluates for each vertebra independently the match between the vertebra surface mesh and the underlying image structure. The feature function is evaluated for each triangle of the surface mesh at the position of the triangle barycentre $\mathbf{x}_i$. The image gradient $\nabla I(\mathbf{x}_i)$ is projected onto the face normal $\mathbf{n}_i$ of each triangle and damped by $g_{\max}$. The feature values of all triangles are summed up providing one value per vertebra for the current position.

The search for a new vertebra position is performed by testing various discrete locations inside a local neighborhood around a given position. For that purpose, a cartesian grid inside a bounding box around the current estimated vertebra location is defined. In order to not only cope with translations but also rotations, the original object is rotated in discrete steps around all axes obtaining M rotation matrices $\mathbf{R}_m$. At each of the N grid positions, the feature function from Eq. 11 is evaluated for the corresponding surface mesh translated by $\mathbf{t}_n$ as well as its M rotated versions. Due to the frequent presence of local minima, the exhaustive search of transformation parameters is preferred to other optimization strategies. It has to be noted again that this search aims at finding a new vertebra position while the original surface model is not deformed yet.

The transformation resulting in the highest feature strength determines the new position of the object:

$$\arg\max_{\mathbf{R}_m, \mathbf{t}_n} \sum_{i \in T} F_i(\mathbf{R}_m \mathbf{x}_i + \mathbf{t}_n)\,, \tag{12}$$

where T is the number of triangles per mesh.

After finding the most promising positions for a pair of neighboring objects m and n, we can define the global external energy for the corresponding kth transformation using the distance measure between transformations as introduced in Eq. 7:

$$E_{\text{ext}_k} = d(\tilde{T}_{\text{VCS}_m}^{-1} \cdot \tilde{T}_{\text{VCS}_n}, T_k)^2 \tag{13}$$

where $\tilde{T}_{\text{VCS}_m}$ and $\tilde{T}_{\text{VCS}_n}$ are the transformations of the corresponding VCSs at the new positions to the world coordinate system. Thus, $\tilde{T}_{\text{VCS}_m}^{-1} \cdot \tilde{T}_{\text{VCS}_n}$ gives the transformation between the corresponding VCSs.

Driving the model towards high image features as performed by the external energy is restricted by the internal energy to prevent false attraction. The internal energy preserves similarity of the ensemble towards an earlier learned constellation model. This is expressed by defining for each transformation T_k

$$E_{\text{int}_k} = d\left(\text{Exp}_{\mu_k}\left(\sum_{i=1}^{t} b_i \mathbf{c}_{i,k}\right), T_k\right)^2 \quad \text{with} \quad \mathbf{b} = \mathbf{C}^T(\text{Log}_{\mu}\mathbf{A}^T) \tag{14}$$

where the matrix of the individual eigenvectors $\mathbf{c}_i$ is denoted as $\mathbf{C}$ and b_i is the coordinate of the weight vector $\mathbf{b}$ associated with the ith principal component. The internal energy penalizes differences between the model and the current constellation. The closest constellation to the model is determined by projecting the given constellation $\mathbf{A}^T$ in the sub-space defined by the principal components.

After calculating the respective energy terms, the final transformations T_k between the individual objects are determined by minimizing

$$E(T_k) = d(T_{\text{ext}_k}, T_k)^2 + \alpha \cdot d(T_{\text{int}_k}, T_k)^2 \tag{15}$$

for each object separately. The transformations T_{ext_k} and T_{int_k} are obtained from the external and internal energy, respectively. The final transformation $T_{k_{\text{opt}}}$ is found using a Downhill-Simplex optimizer. As the representation of the ensemble as a consecution of rigid transformations requires a reference VCS, the vertebra with the highest feature strength is taken as the reference in each iteration.

After positioning the vertebra models using the global model adaptation, a local non-rigid free-form deformation similar to Chap. 5 of the individual surface meshes can be carried out. In order to improve segmentation, all vertebrae are adapted simultaneously with the individual shapes interacting on each other to prevent misadaptations. This has been realized in [25] by introducing a collision detection into feature search and decreasing the feature strength the deeper a target point lies inside another mesh.

3.3 Manifold Embedding

3.3.1 Model Generation

As an alternative to the linear statistical model proposed in the previous section, low-dimensional manifold embedding of the articulated model has been proposed in [20, 21].

For non-linear embedding, we rely on the absolute vector representation $\mathbf{A}_{\text{abs}} = [T_1, T_1 \circ T_2, \ldots, T_1 \circ T_2 \circ \ldots \circ T_{L-1}]$ as given in Eq. 2. Let us now consider N articulated shape models expressed by the feature vectors $\mathbf{A}^i_{\text{abs}}$, of dimensionality D. The aim is to create a low-dimensional manifold consisting of N points $\mathbf{Y}_i$, $\mathbf{Y}_i \in \Re^d$, $i \in [1, N]$ where $d \ll D$ based on [35]. In such a framework, if an adequate number of data points is available, then the underlying manifold $\mathcal{M}$ is considered to be "well-sampled". Therefore, it can represent the underlying population structure. In the sub-cluster corresponding to a pathological population, each point of the training set and its neighbours would lie within a locally linear patch as illustrated in Fig. 6.

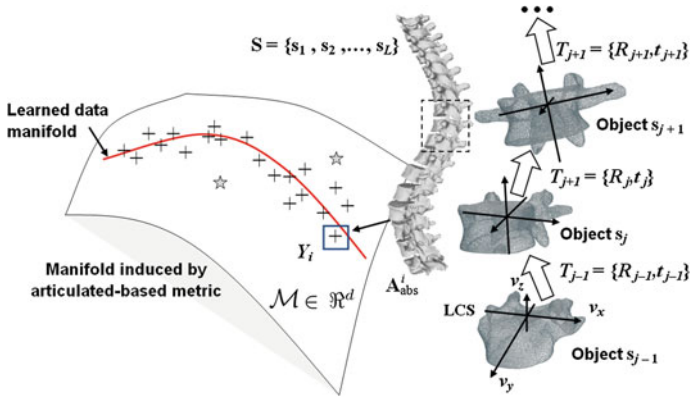


Fig. 6 Representation of inter-vertebral transformations in manifold space

The main limitation of embedding algorithms is the assumption of Euclidean metrics in the ambient space to evaluate similarity between sample points. Thus, similar to Sect. 3.2, a metric in the space of articulated structures is defined that accommodates for anatomical spine variability and adopts the intrinsic nature of the Riemannian manifold geometry allowing us to discern between articulated shape deformations in a topological invariant framework. For each point, the K closest neighbours are selected using a distortion metric which is particularly suited for geodesics. The metric $d_{\text{M}}(\mathbf{A}^i_{\text{abs}}, \mathbf{A}^j_{\text{abs}})$ estimates the distance of articulated models i, j. The distance measure for absolute representations can therefore be expressed as a sum of articulation deviations

$$d_M(\mathbf{A}^i_{\text{abs}}, \mathbf{A}^j_{\text{abs}}) = \sum_{k=1}^{L} d_M(T^i_k, T^j_k) = \sum_{k=1}^{L} \|\mathbf{t}^i_k - \mathbf{t}^j_k\| + \sum_{k=1}^{L} d_G(\mathbf{R}^i_k, \mathbf{R}^j_k). \quad (16)$$

While for the translation, the L_2 norm is chosen, geodesical distances are used between rotation neighbourhoods. This is expressed as $d_G(\mathbf{R}^i_k, \mathbf{R}^j_k) = \| \log((\mathbf{R}^i_k)^{-1} \mathbf{R}^j_k) \|_F$ where the log map is used to map a point in the manifold to the tangent plane.

Afterwards, the manifold reconstruction weights are estimated by assuming the local geometry of the patches can be described by linear coefficients that permit the reconstruction of every model point from its neighbours. In order to determine the value of the weights, the reconstruction errors are measured using the following objective function:

$$\varepsilon(\mathbf{W}) = \sum_{i=1}^{N} \left\| \mathbf{A}^i_{\text{abs}} - \sum_{j=1}^{K} W_{ij} \mathbf{A}^j_{\text{abs}} \right\|^2 \quad (17)$$

$$\text{subject to } \begin{cases} W_{ij} = 0 \text{ if } \mathbf{A}^i_{\text{abs}} \text{ not neighbor } \mathbf{A}^j_{\text{abs}} \\ \sum_j \mathbf{W}_{ij} = 1 \text{ for every } i. \end{cases} \tag{18}$$

Thus, $\varepsilon(\mathbf{W})$ sums the squared distances between all data points and their corresponding reconstructed points. The weights $\mathbf{W}_{ij}$ represent the importance of the jth data point to the reconstruction of the ith element.

The algorithm maps each high-dimensional $\mathbf{A}^i_{\text{abs}}$ to a low-dimensional $\mathbf{Y}_i$. These internal coordinates are found with a cost function Φ minimizing the reconstruction error:

$$\begin{aligned} \Phi(\mathbf{Y}) &= \sum_{i=1}^{N} \left\| \mathbf{Y}_i - \sum_{j=1}^{K} \mathbf{W}_{ij} \mathbf{Y}_j \right\|^2 \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{M}_{ij} \mathbf{Y}_i^T \mathbf{Y}_j \end{aligned} \tag{19}$$

with $\mathbf{M}$ as a sparse and symmetric $N \times N$ matrix enclosing the reconstruction weights $\mathbf{W}_{ij}$ such that $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})$, and $\mathbf{Y}$ spanning the $\mathbf{Y}_i$'s. The optimal embedding, up to a global rotation, is obtained from the bottom $d + 1$ eigenvectors of $\mathbf{M}$ and helps to minimize the cost function $\Phi(\mathbf{Y})$ as a simple eigenvalue problem. The d eigenvectors form the d embedding coordinates. The coordinates $\mathbf{Y}_i$ can be translated by a constant displacement without affecting the overall cost $\Phi(\mathbf{Y})$. The eigenvector corresponding to the smallest eigenvalue corresponds to the mean value of the embedded data $\mathbf{Y}^0 = \{\mathbf{y}_1, \ldots, \mathbf{y}_d\}$, $\mathbf{y}_i = 0, \forall i$. This can be discarded with $\sum \mathbf{Y}_i = 0$ to obtain an embedding centered at the origin. Hence, a new ADM can be inferred in the embedded d-space as a low-dimensional point $\mathbf{Y}^{\text{new}}$ by finding its optimal manifold coordinates $\mathbf{y}_i$.

To obtain the articulation vector for a new embedded point in the ambient space (image domain), one has to determine the representation in high-dimensional space based on its intrinsic coordinates. We first assume an explicit mapping $f : \mathcal{M} \to \Re^D$ from manifold space $\mathcal{M}$ to the ambient space $\Re^D$. The inverse mapping of $\mathbf{Y}_i$ is then performed by estimating the relationship between $\Re^D$ and $\mathcal{M}$ as a joint distribution, such there exists a smooth functional which belongs to a local neighborhood. Theoretically the manifold should follow the conditional expectation:

$$f(Y_i) \equiv E(\mathbf{A}^i_{\text{abs}} | \mathcal{M}(A_i) = \mathbf{Y}_i) = \int A_i \frac{p(\mathbf{Y}_i, A_i)}{p_{\mathcal{M}(A_i)}(\mathbf{Y}_i)} dD \tag{20}$$

which captures the overall trend of the data in D-space. Here, both $p_{\mathcal{M}(A_i)}(\mathbf{Y}_i)$ (marginal density of $\mathcal{M}(A_i)$) and $p(\mathbf{Y}_i, A_i)$ (joint density) are unknown. Based on the *Nadaraya-Watson* kernel regression [33], we replace densities by kernel functions as $p_{\mathcal{M}(A_i)}(\mathbf{Y}_i) = \frac{1}{K} \sum_{j \in \mathcal{N}(i)} G_h(\mathbf{Y}_i, \mathbf{Y}_j)$ and $p(\mathbf{Y}_i, A_i) = \frac{1}{K} \sum_{j \in \mathcal{N}(i)} G_h(\mathbf{Y}_i, \mathbf{Y}_j)$ $G_g(A_i, A_j)$ [12]. The Gaussian regression kernels G require the neighbors $\mathbf{A}^j_{\text{abs}}$ of

$j \in \mathcal{N}(i)$ to determine the bandwidths h, g so it includes all K data points ($\mathcal{N}(i)$ representing the neighborhood of i). Plugging these estimates in Eq. (20) gives:

$$f_{\mathrm{NW}}(\mathbf{Y}_i) = \int A_i \frac{\frac{1}{K}\sum_{j\in\mathcal{N}(i)} G_h(\mathbf{Y}_i, \mathbf{Y}_j) G_g(A_i, A_j)}{\frac{1}{K}\sum_{j\in\mathcal{N}(i)} G_h(\mathbf{Y}_i, \mathbf{Y}_j)} dD. \tag{21}$$

By assuming G is symmetric about the origin, we propose to integrate in the kernel regression estimator, the manifold-based distortion metric d_M which is particularly suited for geodesic metrics and articulated diffeomorphisms. This generalizes the expectation such that the observations $\mathbf{Y}$ are defined in manifold space $\mathcal{M}$:

$$f_{\mathrm{NW}}(\mathbf{Y}_i) = \arg\min_{\mathbf{A}^i_{\mathrm{abs}}} \frac{\sum_{j\in\mathcal{N}(i)} G(\mathbf{Y}_i, \mathbf{Y}_j) d_M(\mathbf{A}^i_{\mathrm{abs}}, \mathbf{A}^j_{\mathrm{abs}})}{\sum_{j\in\mathcal{N}(i)} G(\mathbf{Y}_i, \mathbf{Y}_j)} \tag{22}$$

which integrates the distance metric $d_M(\mathbf{A}^i_{\mathrm{abs}}, \mathbf{A}^j_{\mathrm{abs}})$ defined in Eq.(16) and updates $f_{\mathrm{NW}}(\mathbf{Y}_i)$ using the closest neighbors of point $\mathbf{Y}_i$ in the manifold space. This constrains the regression to be valid for similar data points in its vicinity since locality around $\mathbf{Y}_i$ preserves locality in $\mathbf{A}^i_{\mathrm{abs}}$.

To capture local vertebra shape variation, we rely on the assumption that global deformations, represented in a local neighborhood of $\mathcal{M}$, will also manifest similar local geometries due to the same type of pathological deviation affecting shape morphology. We assume that local appearances follow a linear distribution within the low-dimensional manifold. Hence, given a data point $\mathbf{Y}_j$ and its K neighbors, the local shape model $\mathbf{s}_i$, representing the ith element of the ADM, is obtained by building a particular class of shapes given the set of examples $\{\mathbf{s}^1_i, ..., \mathbf{s}^K_i\}$. We approximate the distribution of the shape using a point distribution model so that a new vertebra $\mathbf{s}^{\mathrm{new}}_i = \bar{\mathbf{s}_i} + [\mathbf{e}_1 \dots \mathbf{e}_n][\boldsymbol{\omega}_1 \dots \boldsymbol{\omega}_n]$ can be instantiated, where n are the eigenvalues with corresponding eigenvectors $\mathbf{e}$ and $\bar{\mathbf{s}_i}$ is the mean shape of the K neighboring local objects as well as weight vector $\mathbf{w} = [\boldsymbol{\omega}_1 \dots \boldsymbol{\omega}_n]$.

3.3.2 Model Adaptation

Once an appropriate modeling of spine shape variations is determined with a manifold, a successful inference between the image and manifold must be accomplished. We describe here how a new model is deformed. We search the optimal embedded manifold point $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_d)$ of the global spine model. Such a strategy offers an ideal compromise between the prior constraints, as well as the individual shape variation described by the weight vector $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)$ in a localized sub-patch. The energy E of inferring the model $\mathbf{S}$ in the image $\mathcal{I}$ is a function of the set of displacement vectors Δ in the manifold space for global shape representation. This involves: (a) a data-related term expressing the image cost and (b) a global prior term measuring deformation between low-dimensional vectors with shape models.

The third term represents (c) a higher-order term which is expressed by the reconstruction weights Ω for local vertebra modeling. The energy E can be expressed as the following combination of a global and local optimization:

$$E\left(\mathbf{S}^0, \mathcal{I}, \Delta, \Omega\right) = V\left(\mathbf{Y}^0 + \Delta, \mathcal{I}\right) + \alpha\ V\left(\mathbf{N}, \Delta\right) + \beta\ V\left(\mathbf{H}, \Delta, \Omega\right). \tag{23}$$

The global alignment of the model with the target image primarily drives the deformation of the model. The purpose is to estimate the set of articulations describing the global spine model by determining its optimal representation $\mathbf{Y}^0$ in the embedded space. This is performed by obtaining the global representation using the mapping in Eq. 22 so that: $f_{\mathrm{NW}}(Y_i + \Delta) = f_{\mathrm{NW}}(\{y_1 + \boldsymbol{\delta}_1, \ldots, y_d + \boldsymbol{\delta}_d\})$. This allows to optimize the model in manifold space coordinates while retrieving the articulations in $\mathcal{I}$. The global cost can be expressed as:

$$V\left(\mathbf{Y}^0 + \Delta, \mathcal{I}\right) = V\left(f_{\mathrm{NW}}(\{y_1 + \boldsymbol{\delta}_1, \ldots, y_d + \delta_d\}), \mathcal{I}\right). \tag{24}$$

The inverse transform allows to obtain $\mathbf{A}^i_{\mathrm{abs}} + \boldsymbol{D}$, with $\boldsymbol{D}$ as deformations in the image space. Since the transformations T_i are implicitly modeled in the absolute representation $\mathbf{A}^0_{\mathrm{abs}}$, we can formally consider the singleton image-related term as a summation of costs associated with each L vertebra of the model:

$$V\left(\mathbf{A}^0_{\mathrm{abs}} + \boldsymbol{D}, \mathcal{I}\right) = \sum_{i=1}^{L} V_i\left(\mathbf{s}_i * (T_i^0 + \boldsymbol{D}_i), \mathcal{I}\right) \tag{25}$$

where $V_i(\mathbf{s}, \mathcal{I}) = \sum_{\mathbf{v}_i \in \mathbf{s}} \mathbf{n}_i^T(\mathbf{v}_i) \nabla \mathcal{I}(\mathbf{v}_i)$ minimizes the distance between mesh vertices of the inferred shape and gradient image $\mathcal{I}$ by a rigid transformation. Here, $\mathbf{n}_i$ is the normal pointing outwards and $\nabla \mathcal{I}(\mathbf{v}_i)$ the image gradient at $\mathbf{v}_i$.

The prior constraint for the rigid alignment are pairwise potentials between neighboring models y_i such that the difference in manifold coordinates is minimal with regards to a prior distribution of neighboring distances P:

$$\alpha\ V\left(\mathbf{N}, \Delta\right) = \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y_i^0 + \boldsymbol{\delta}_i, y_j^0 + \boldsymbol{\delta}_j, P). \tag{26}$$

This term represents the smoothness term of the global cost function to ensure that the deformation $\boldsymbol{\delta}_i$ applied to point coordinates are regular, with $V_{ij} = (0, 1)$ a distance assigning function based on the distances to P.

Local shape geometry for each vertebrae of the articulated model is obtained by varying the weight parameters of the principal variations at each level. We parameterize these potentials with a set $\mathcal{C}$ of clique variables c, controlled by high-order potential V_c [34] which assigns a cost to a configuration of c. Each clique is assigned to weight vectors ω_c. Hence the third term of Eq. 23 is described as a high-order functional:

$$\beta\ V\Big(\mathbf{H}, \Delta, \Omega\Big) = \beta \sum_{c \in \mathcal{C}} V_c(\mathbf{w}_c^0 + \omega_c) \tag{27}$$

where independent clique variables c are treated as a graph minimization problem. The prior term is represented by higher-order potentials of degree n, based on the eigenvalues of the L local vertebrae from our model $\mathbf{S}$. Our work is inspired from a mesh reconfiguration Chap. 5 where costs are associated to cliques c based on the positions of the morphed mesh vertices $\mathbf{v}_i$. A search is performed along the normal $\mathbf{n}_i$ from $\mathbf{v}_i$ to find the optimal compromise between boundary detection and the distance to the mean eigenvalue shape. We therefore penalize deformations which deviates from the local distribution.

One can integrate the global data and prior terms along with local shape terms parameterized as the higher-order cliques, by combining Eqs. 24, 26 and 27:

$$\begin{aligned} E\Big(\mathbf{S}^0, \mathcal{I}, \Delta, \Omega\Big) = V\Big(f_{\mathrm{NW}}(\{y_1 + \boldsymbol{\delta}_1, \ldots, y_d + \boldsymbol{\delta}_d\}), \mathcal{I}\Big) \\ + \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y_i^0 + \boldsymbol{\delta}_i, y_j^0 + \boldsymbol{\delta}_j) + \beta \sum_{c \in \mathcal{C}} V_c(\mathbf{w}_c^0 + \omega_c). \end{aligned} \tag{28}$$

The optimization strategy of the resulting MRF (28) in the continuous domain is not a straightforward problem. The convexity of the solution domain is not guaranteed, while gradient-descent optimization approaches are prone to non-linearity and local minimums. We seek to assign the optimal labels $\mathcal{L}^{\Delta} = \{l_1, \ldots, l_d\}$ and $\mathcal{L}^{\Omega} = \{l_1, \ldots, l_n\}$ which are associated to the quantized space Δ of displacements and local weight parameters Ω respectively. We consider that displacing the coordinates of point y_i^0 by $\boldsymbol{\delta}^{l_i}$ is equivalent to assigning label l_i to y_i^0. An incremental approach is adopted where in each iteration t we look for the set of labels that improves the current solution s.t. $y_i^t = y_i^0 + \sum_t \boldsymbol{\delta}^{l_i t}$, which is a temporal minimization problem. Then (28) can be re-written as:

$$\begin{aligned} E^t(\mathcal{L}^{\Delta}, \mathcal{L}^{\Omega}) = V\Big(f_{\mathrm{NW}}(\{y_1^{t-1}, l_1^{\Delta}, \ldots, y_d^{t-1}, l_d^{\Delta}\}), \mathcal{I}\Big) \\ + \alpha \sum_{i \in G} \sum_{j \in \mathcal{N}(i)} V_{ij}(y_i^{t-1}, y_j^{t-1}, l_i^{\Delta}, l_j^{\Delta}) + \beta \sum_{c \in \mathcal{C}} V_c(\mathbf{w}_c^{t-1}, l_c^{\Omega}). \end{aligned} \tag{29}$$

We solve the minimization of the higher-order cliques in (29) by transforming them into quadratic functions [34]. We apply the FastPD method [26] which solves the problem by formulating the duality theory in linear programming.

4 Results

4.1 Spine Segmentation Using Articulated Models

4.1.1 Linear Model

Model building was based on vertebra shape models adapted to 18 thoracic CT volumes. From the adapted meshes, mean shape models of all vertebrae as well as a statical model of rigid transformations between the VCSs was created.

The articulated model was adapted to ten test thoracic CT data sets including pathologies like strong classifications between vertebrae or scoliosis. In-plane voxel-resolution varied between 0.85 and 0.97 mm and slice thickness between 2.5 and 3.5 mm. Model initialization was given as described in [23].

Automatic segmentation was compared to reference segmentations by calculating for each vertex of the adapted mesh the closest point on the reference surface. If segmentation for each vertebra is done seperately, misadaptation to neighbouring can occur. In the ten test cases, this scenario was observed for four patients. However, when considering the overall constellation by means of the statistical articulated model as described in Sect. 3.2.1, this problem could be circumvented resulting in a significant improvement. After global model adaptation, a simultaneous segmentation of the individual vertebrae was performed using collision detection to prevent overlapping between neighbouring vertebrae. Overall, a segmentation accuracy of 1.0 ± 0.3 mm with the proposed approach, which is a clear improvement to 2.7 ± 2.8 when adapting the vertebrae individually. Figure 7 illustrates the improvements when using the articulated model and simultaneous vertebra segmentation.

4.1.2 Manifold Embedding

The manifold was built from a database containing 711 scoliotic spines demonstrating several types of deformities. Each spine model in the database was obtained from biplanar radiographic stereo-reconstructions [16]. It is modelled with 12 thoracic and 5 lumbar vertebrae (17 in total), represented by 6 landmarks on each vertebra (4 pedicle extremities and 2 endplate center points) which were manually identified by an expert on the radiographic images. Once a 3D point-based vertebra model is obtained, each vertebra was fitted with a triangulated mesh surface using generic vertebra priors obtained from a serial CT-scan reconstruction of a cadaver specimen. The same six precise anatomical landmarks (4 pedicle tips and 2 on the vertebral body) were annotated on each triangulated model as shown in Fig. 1b. The coordinates for each anatomical points were used to determine the VCS of each vertebral body. Figure 8 displays the resulting embedding from the training data of 711 spine models in $\mathcal{M}$.

Adaptation of the articulated model was done on two different data sets. The first consisted of volumetric CT scans ($512 \times 512 \times 251$, resolution: 0.8×0.8 mm,

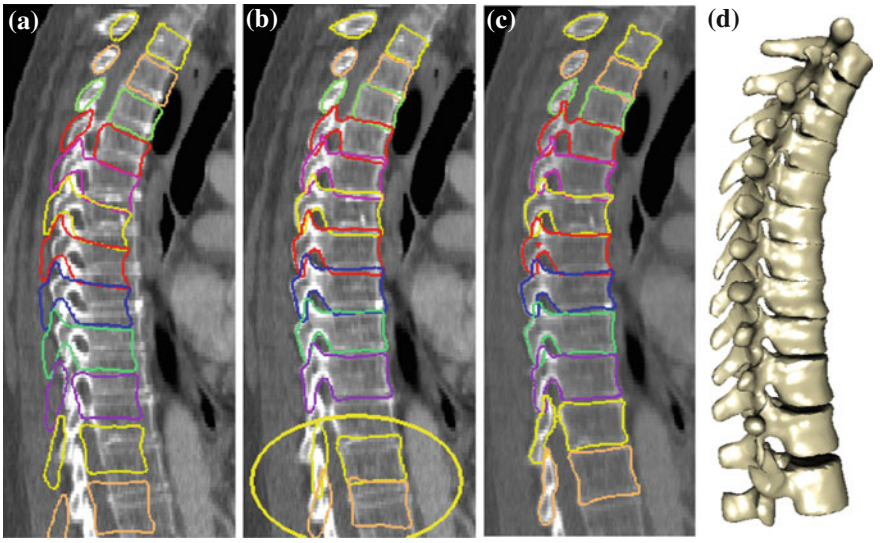


Fig. 7 Automatically positioned models **a** are individually adapted **b** and using the articulated model including simultaneous vertebra adaptation (**c**). Due to the consideration of overall constellation, misadaptation of T11 and T12 could be overcome and overlapping between neighbored models could be decreased. Surface rendering of adapted meshes (**d**)

thickness: 1–2 mm) of the lumbar and main thoracic regions obtained from 21 different patients acquired for operative planing purposes. The MR dataset comprised multi-parametric volumetric data ($256 \times 256 \times 160$, resolution: 1.3×0.9 mm, thickness: 1 mm) of 8 patients acquired for diagnostic purposes. For this study, only the T1 sequence was selected for the experiments. All patients on both datasets (29 in total) had 12 thoracic and 5 lumbar vertebrae. Both CT and MR data were manually annotated with 3D landmarks by an expert in radiology, corresponding to left and right pedicle tips as well as midpoints of the vertebral body. Segmentation of the vertebrae from the CT and MR slices were also made by the same operator. Quantitative assessment consisted of measuring landmark accuracy, as well as inferred surface distance errors.

CT imaging experiments. We first evaluated the model accuracy in CT images by computing the correspondence of the inferred vertebral mesh models to the segmented target structures. As a pre-processing step, a rough thresholding was performed on the whole volume to filter out noise artifacts. The overall surface-to-surface comparison results between the inferred 3D vertebral models issued from the articulated model and from known segmentations were first calculated. The mean errors are 2.2 ± 1.5 mm (range: 0.6–5.4 mm) for thoracic vertebra and 2.8 ± 1.9 mm (range: 0.7–8.1 mm) for lumbar vertebra. A qualitative assessment of the 3D model from the CT images is presented for the thoraco-lumbar region in Fig. 9a, with the outline of the vertebral body extracted from the inferred mesh model. One could observe accurate delineation of geometrical models on selected multi-planar views.

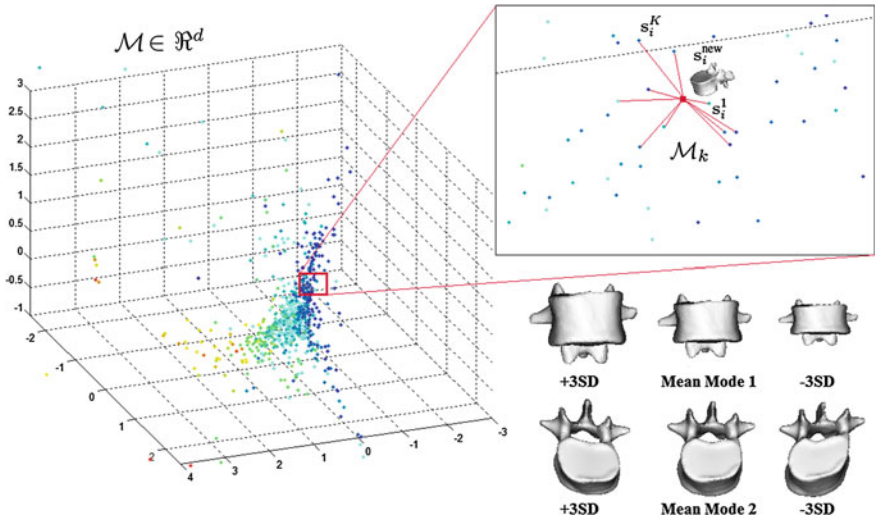


Fig. 8 Low-dimensional manifold embedding of the spine dataset comprising 711 models exhibiting various types of deformities. The sub-domain was used to estimate both the global shape pose costs and individual shape instances based on local neighborhoods

In fact, it is still very challenging to precisely capture the exact vertebra geometry (with transverse and spinous processes) given limited visibility and varying patient morphology.

Furthermore, we evaluated 3D landmark mean and S.D. differences to annotations made by an expert in radiology. While these types of data cannot be considered as ground-truth, as the annotated landmarks are also prone to human variability, this gives a good indication on the degree of convergence from the method. The overall mean difference (method vs. expert) for the selected cases was of 1.6 ± 0.6 mm for the pedicle and endplate landmarks. The errors also yields statistically significantly lower standard deviations compared to a manual technique (0.64 vs. 1.76 mm).

MR imaging experiments. For the experiments involving the segmentation of 3D spine models from MR images, an anisotropic filtering combining diffusion and shock filters was applied to the images in order to reduce inhomogeneity due to the magnetic and motion artifacts which hinders the bony boundaries. The surface-to-surface comparison showed encouraging results (thoracic: 2.9 ± 1.8 mm, lumbar: 3.0 ± 1.9 mm) based on differences to ground-truth. As in the previous experiments with CT imaging, ground-truth data was generated by manually segmenting the structures models which were validated by an expert in radiology. An illustrative result of the 3D model from the MR images is shown in Fig. 9b. As difficult as the CT inference is, the MR problem represent an even greater challenge as the image resolution is more limited and inter-slice spacing is increased compared to CT. Modeling of the statistical properties of the shape variations and global pose becomes even more important in this case, as it relies heavily in the non-linear distribution of

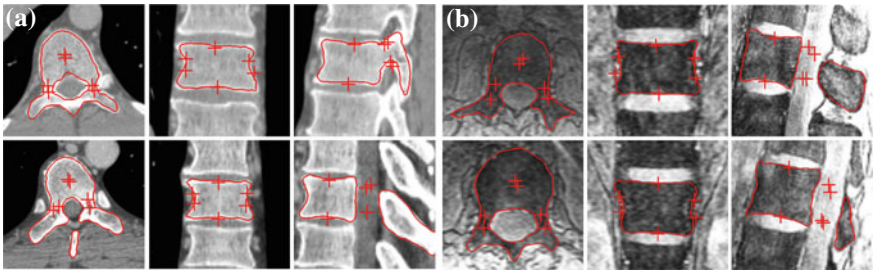


Fig. 9 Qualitative assessment of the spine model segmentation in CT **a** and MR **b** images for two levels with projected anatomical landmarks

the patient morphology. The accuracy is still comparable to ground-truth but not as reliable as in the case for CT imaging.

Comparison of 3D vertebral landmarks to those identified by an expert in radiology gave an overall mean squared distances (method vs. expert) of 2.0 ± 0.8 mm for the pedicle body landmarks and of 2.1 ± 0.8 mm for vertebral body landmarks. The global 3D point landmark difference was 2.0 ± 0.8 mm. When we separate the point-to-point differences in specific anatomical regions of the 3D models issued from the proposed technique, the mean difference was of 2.4 ± 1.0 mm for lumbar and 1.9 ± 0.7 mm for thoracic vertebrae.

4.1.3 Discussion

Both linear and non-linear articulated model representations could be adapted successfully to the respective image data with a segmentation accuracy of a few milimeters. In the case of the non-linear model, multi-model adaptation could also been shown. Nevertheless, in each case, the image data used for evaluation was still comparably small and it still needs to be shown that both are robust in a clinical environment. Unfortunately, in both cases different image data sets have been used limiting the comparison of both methods.

4.2 Vertebra Identification

A reliable identification of vertebrae in CT images containing parts of the spine column is essential for numerous orthopedic, neurological, and oncological applications. For example, in the case of spine surgery it is important that the correct vertebra is treated.

However, identification of individual vertebrae can be extremely challenging as illustrated in Fig. 10. Adjacent vertebrae can hardly be distinguished from each other, particularly if they belong to the same spinal region, like cervical, thoracic, or lumbar

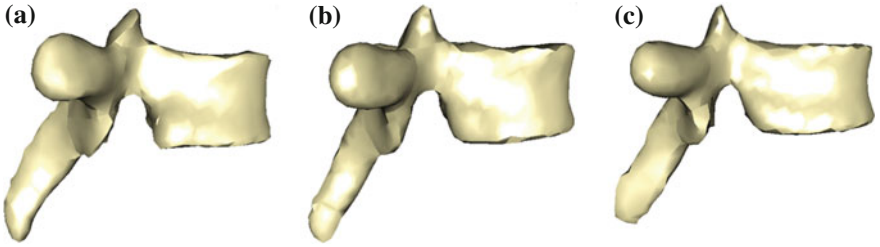


Fig. 10 Adjacent vertebra show very similar characteristics, especially if they belong to the same region as shown by means of T5–T7. There are only marginal differences like the shape of the processes or the size of the vertebral bodies

vertebrae. Marginal differences can be determined at the vertebral processes or for the size of the vertebral bodies. Even physicians usually solve the identification problem by searching for one characteristic vertebra, e.g., the first cervical or first thoracic vertebra, and subsequently label the neighbored vertebrae iteratively. In many cases, the labeling can be facilitated by the use of reference structures. However, spine CT scans may sometimes show a small part of the spine containing only a few vertebrae which additionally impedes distinguishing characteristics. Furthermore, surrounding structures may appear only sparsely, hence they cannot be used as reference structures for the identification of vertebrae.

However, with the prior knowledge captured in the appearance models, we are able to tackle vertebra identification. In the following, we shortly explain how the identification of vertebrae using appearance models can be performed in Sect. 4.2.1. Then, two key experiments are carried out. At first, the identification rate depending on the number of vertebrae shown in the image are evaluated in Sect. 4.2.2. The identification of individual vertebrae is analyzed in Sect. 4.2.3. More information can be found in [24].

4.2.1 Identification Using Appearance Models

All experiments were carried out on 64 patient images with a voxel size in-plane ranging from 0.316 to 0.976 mm and a slice thickness of 0.8–2.79 mm. Each image plane had 512×512 voxels with varying voxel numbers in z-direction of about 20–300. 18 head neck scans, 33 thorax scans including 3 scans showing also the lumbar spine, 9 abdominal scans showing a small part of the lower thoracic and upper lumbar spine, 4 scans that exclusively show the lumbar spine.

For identification, we first detected the vertebrae in the images using the gradient models from Sect. 2.4. Afterwards, the appearance models were registered to the image content around each candidate and the similarity to the model was evaluated. In this way, we can measure the similarity between all candidates and the appearance models as illustrated in Table 1.

Table 1 Evaluation of similarities of vertebra models to detected candidates. The identification is carried out by averaging each diagonal which represents a configuration

	C1	C2	C3	…	L5
Cand 1	$s_{1,1}$	$s_{1,2}$	$s_{1,3}$	…	$s_{1,24}$
Cand 2	$s_{2,1}$	$s_{2,2}$	$s_{2,3}$	…	$s_{2,24}$
Cand 3	$s_{3,1}$	$s_{3,2}$	$s_{3,3}$	…	$s_{3,24}$
⋮	⋮	⋮	⋮	⋱	⋮
Cand N_C	$s_{N_C,1}$	$s_{N_C,2}$	$s_{N_C,3}$	…	$s_{N_C,24}$

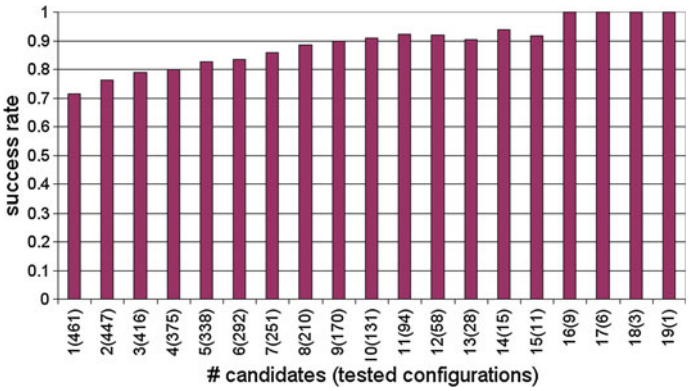


Fig. 11 Identification success rates for different amounts of vertebrae. Success rates increase as more candidates were detected. The values in parentheses on the **x**-axis denote the number of configurations for each case

For each diagonal of this table, denoted as a configuration, the average similarity measure was calculated and the diagonal with the highest mean similarity value was supposed to correspond to be the true configuration. Thus, so far we were always assuming that the detection provides consecutive candidates without missing or false detections. By evaluating the whole table, which means that, e.g., also the fifth lumbar vertebra was compared to the atlas, not only false detections beyond the vertebral column could be detected, but also the reliability of the identification of vertebra candidates was evaluated. As a similarity measure for the registration, we applied local correlation which turned out to provide the best results compared to cross correlation and sum of squared differences.

4.2.2 Dependence on the Amount of Candidates

By analysing each row of every table of similarities (see Table 1), the success rate for an arbitrary amount of detected candidates could be evaluated. Figure 11 shows the identification success rate for one up to 19 candidates (largest amount of visible vertebrae in image). Even if only one single vertebra is given, the registration successfully identified the object in more than 70 % of the cases. Increasing the number

of available vertebrae lead to an increase in identification rate reaching 100 % if 16 or more vertebrae were shown in the image since this reduced the number of possible configurations. The slight decrease for the cases of 13 and 15 shown vertebrae was probably caused by the fact that the more vertebrae were shown, the smaller the sample size was.

4.2.3 Identification Rate for Individual Vertebrae

Another key question is how reliable certain vertebrae can be identified. From a visual inspection, there are some vertebrae that can be identified more easily than others. In this experiment, we determined the identification rates for individual vertebrae by registering to each vertebra candidate in the training data all vertebrae and evaluating each row in the table of similarities (see Table 1). The results are shown in Fig. 12a. As one would expect, the identification rate was quite low in the middle of the thoracic spine since in this area the neighbouring vertebrae are very similar and can also be hardly distinguished by human observers. However, the large differences in success rates between the neighbouring vertebrae, e.g., fourth cervical vertebra (C4), were surprising and probably caused by the fact that the appearance models were created only from a small sample size of ten images. Another reason could be that neighbouring objects—besides of the cervical vertebrae—were extracted in parts from different samples. The shift in vertebrae in the case of a failure in identification is given in Fig. 12b. In 84 %, a shift by one vertebra, and in 13 % by two vertebrae occurred. It has to be noted that each row always showed a global maximum in the table of similarities.

4.3 Change Assessment

As an example for CAD, we show how the introduced models can be used for spine segmentation on follow-up exams for osteoporotic fracture detection. For more details, we refer to [29].

The image data base used in this study consisted of 16 patients with a base-line and a follow-up examination, including 11 patients with at least one new fracture in the follow-up scan. Patient age ranged from 44 to 79 years. In total, the study comprises 246 vertebra pairs in thoracic and lumber spine with no fracture in both scans and 20 vertebrae pairs with a new fracture in the follow up scan (Fractures: 1 × T5, 1 × T6, 1 × T9, 2 × T10, 2 × T11, 3 × T12, 4 × L1, 3 × L2, 2 × L3, 1 × L4). The fractures were diagnosed by two radiologists in consensus by using the spinal fracture index (SFI [13]). The time distance between base-line and follow-up scan ranged from 7 to 180 weeks. The 3D images were reconstructed in sagittal slices, (slice distance 3 mm, in-plane resolution between 0.3 × 0.3 and 1.3 × 1.3 mm), for which previous studies [7, 32] had indicated improved fracture detection.

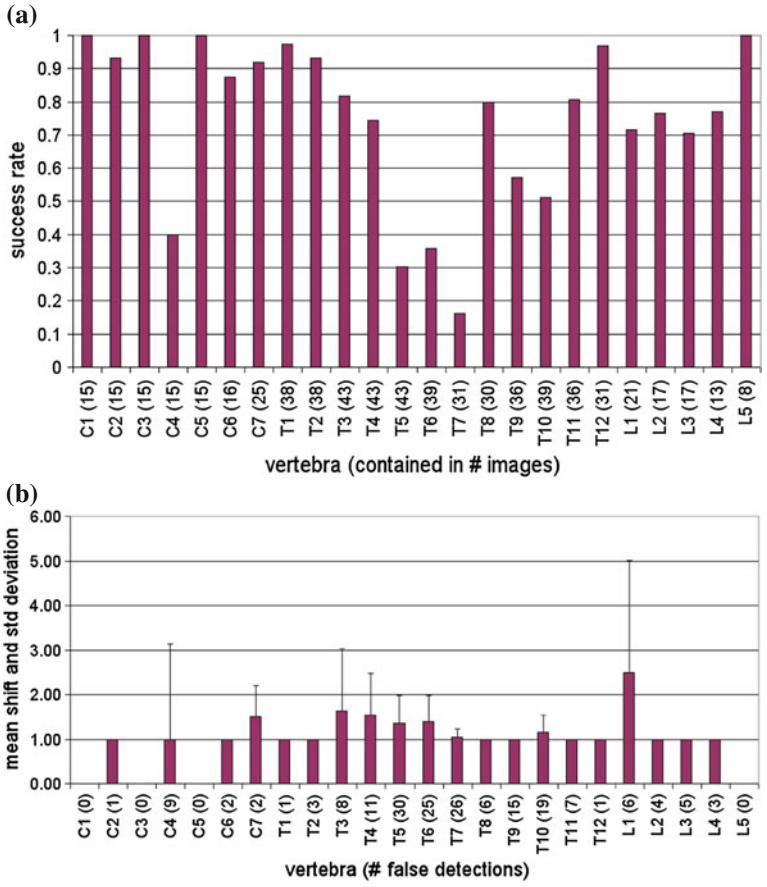


Fig. 12 Vertebra identification rates (**a**), mean transposition and standard deviation in the case of failed identification (**b**)

Segmentation of the vertebrae was done automatically using the approach described in [24]. The vertebrae were at first detected in curved planar reformatted images of the spine using the gradient models from Sect. 2.4 and then identified using the appearance models from Sect. 2.5. Finally, segmentation was done by adapting the surface models from Sect. 2.1. Since point correspondences were preserved during adaptation, a point-based rigid registration T_{bf} from the space of the base-line image I_b into the follow-up image I_f could be calculated between the meshes. Furthermore, in the neighborhood of the chosen vertebra, it allowed to define voxel correspondence between the images. In the same way, the transformation between model space and image space could be estimated, allowing geometric entities defined in the model space to be transformed into the image space. An example case is shown in Fig. 13 illustrating the result of the vertebra segmentation in base line and follow-up.

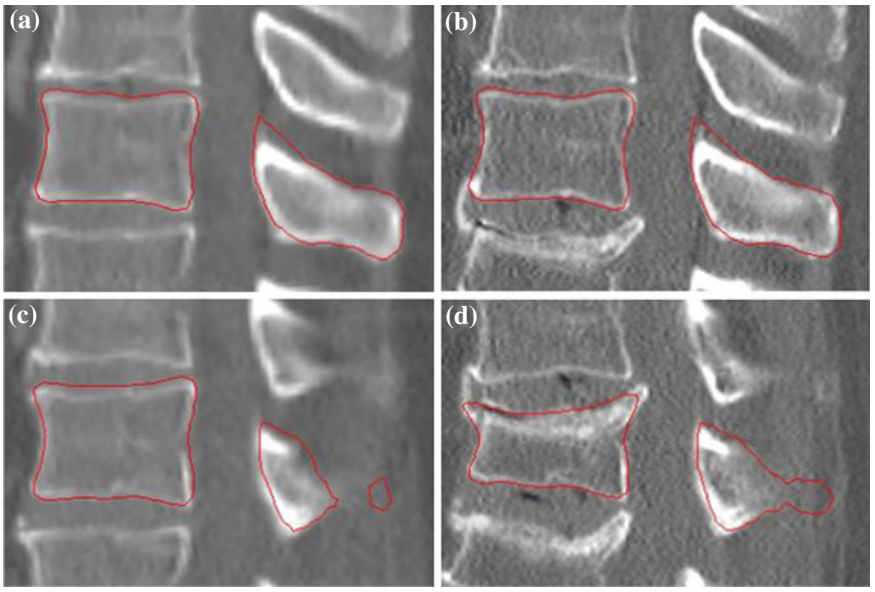


Fig. 13 Example case illustrating osteoporotic fracture detection. *Top row* T11 with no obvious change between base line and follow-up exam. *Bottom row* T12 where the vertebra body is significantly collapsed in the follow-up compared to the base line scan. In each case, adapted vertebra shape models are given by contours. The adapted meshes can be used for automatic change assessment

For change assessment, we define the mean vertex-to-vertex distance between base-line mesh $\mathbf{s}_b$ and follow-up mesh $\mathbf{s}_f$ of a given vertebra as

$$D_{vv}^{S} = \frac{1}{|S|} \sum_{i \in S} \left| T_{bf}(\mathbf{v}_i^b) - \mathbf{v}_i^f \right| \tag{30}$$

with S being the vertex index set, $\mathbf{v}_i^b$ and $\mathbf{v}_i^f$ being the vertices of base-line and follow-up mesh, respectively.
The vertex index set S may include either all mesh vertices (S_t), or a sub-set, e.g., all vertices that are associated to the lower and upper end plates of the vertebral body (S_e). While vertex-to-vertex distances are a measure of anatomical vertex correspondence, a vertex-to-surface measure is more appropriate for evaluating segmentation reproducibility. We therefore defined for a vertex $\mathbf{v}_i^f$ on mesh $\mathbf{s}_f$ the projection operator P_i^f that yields the point on the surface which is closest to the transformed vertex $T_{bf}(\mathbf{v}_i^b)$ of mesh $\mathbf{s}_b$ and obtain the following vertex-to-surface mesh distance:

$$D_{vs}^{S} = \frac{1}{|S|} \sum_{i \in S} \left| T_{bf}(\mathbf{v}_i^b) - P_i^f(T_{bf}(\mathbf{v}_i^b)) \right| \tag{31}$$

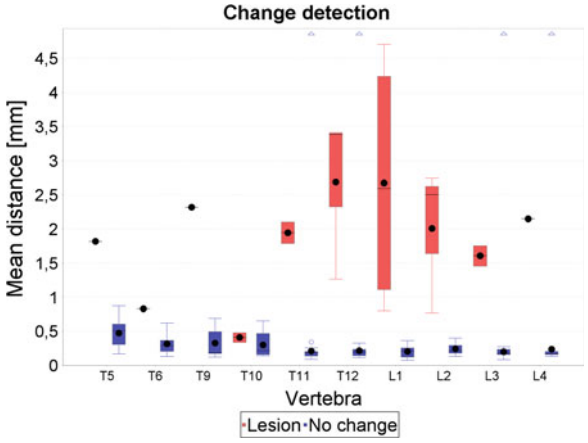
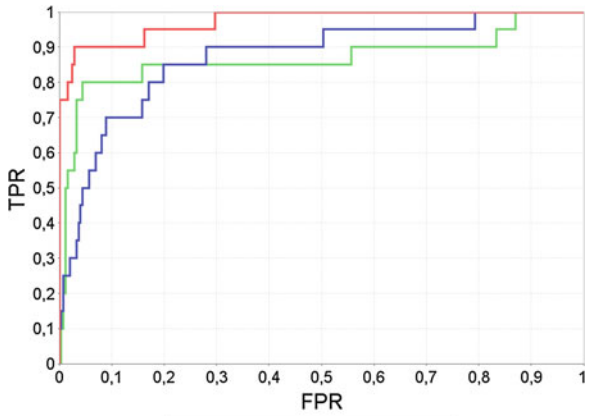


Fig. 14 Vertex-to-surface ($D_{vs}^{S_e}$) statistics for distances in mm between vertebra segmentations from base-line and follow-up scans. Only the vertebral body end plates were evaluated (see Fig. 1a). Distances in case of no change are displayed on the right side of the vertebra label (*blue*), and distances in case of a new lesion on the left side (*red*). Only vertebrae with lesion samples in the follow-up scans are shown

Furthermore, the difference of volume enclosed by base-line and follow-up mesh, D_{Vol}, was used as a further measure of change. To evaluate the possibility of an automated change detection, we selected all vertebrae with a lesion in the follow-up scan, but no lesion in the base-line scan. For the patient data used in this study, the lesion of interest is a compression fracture of the vertebral body, caused by osteoporotic fracture detection. Therefore, the distance measure was only evaluated for vertices in S_e, being part of to the end plates of the vertebral body. For each vertebra scan pair, the mean vertex-to-surface distance $D_{vs}^{S_e}$ was calculated.

Figure 14 depicts the resulting distribution of mean mesh distances for all vertebra types with lesion samples in at least one of the follow-up scans.

The results show, that the mesh distances for pairs with a lesion in the follow-up scan are typically well separated from the no-change cases. The distance ranges, however, overlap to some extend, making it impossible to define a distance threshold with 100 % classification accuracy. To show the threshold dependent classification potential, we calculated receiver operating characteristic (ROC) curves for the distance measures $D_{vs}^{S_e}$, $D_{vs}^{S_t}$, and D_{Vol}. The curves as depicted in Fig. 15, show the rate of true positive results (TPR) as function of the rate of false positive results (FPR), when varying the distance threshold. While the best operating point on the curve can be debated depending on application and work flow, the curve integral (Area Under Curve, $0 \leq AUC \leq 1$) is often used to compare classification approaches. We achieved AUC values between 0.97 and 0.86 depending on the differences measure. A max Youden Index of 0.75 and a max F1 score of 0.86 is both reached at a threshold of 1.26 mm when using the $D_{vs}^{S_e}$ distance measure. The finding that $D_{vs}^{S_e}$ leads to the highest AUC was consistent with the clinical approach to detect vertebral fractures.



	D_{vs}^{Se}	D_{vs}^{St}	D_{Vol}
AUC	0.97	0.87	0.86

Fig. 15 ROC *curves* for change detection based on three mesh distance measures. The area under *curve* (AUC) shows that a good discrimination can be achieved, especially with the end-plate based distance measure

The spinal fracture index (SFI) was based on visual inspection of the end plates of each vertebral body estimating the reduction in height without direct vertebral measurement [13].

5 Conclusion

In this article, we presented a comprehensive framework for computational modeling of the spine. The individual vertebrae are represented using a set of different models covering shape, relative location, gradient as well as appearance information. In order to model the global spine curve, the spine is expressed as a sequence of rigid transformations between the individual vertebrae.

Given the representation of the spine as a set of rigid transformations, statistical modeling is no longer straightforward as this representation does not belong to a vector space and statistics suited for Riemannian geometry have to be considered. Two different methods for statistical modelling of the articulated spine model was proposed using linear and non-linear embeddings. Linear statistical methods are usually more intuitive and thus easier to interpret the results. Furthermore, mapping a new instance to the linear space as well as the calculation of the inverse mapping are well defined. However, non-linear manifold embedding algorithms map high-dimensional observation data that are presumed to lie on a non-linear manifold, onto a single global coordinate system of lower dimensionality. Manifold embedding preserves neighborhood relationships of similar spine geometries and thereby reveals

the underlying structure of the data. Thus, dimensionality reduction using manifold embedding allows to recover the underlying manifold, whereas linear embedding methods map various data points to nearby points in the plane, creating distortions both in the local and global geometry. While we presented the mathematical concepts and also the possibility to individualize both models to an unseen patient in this article, thorough evaluation of the individual methods needs to be addressed in the future. So far, comparison was difficult as different data sets were used both for model building as well as for model adaptation.

A key task for many applications but also very challenging is vertebrae identification. Neighbouring vertebrae are usually so similar that even for a human observer it is difficult to identify the individual vertebrae without using additional context information or simply counting them. Based on the detailed modeling of each individual vertebra, automated identification was tackled using appearance models that captured statistical image information in the local neighbourhood around each vertebra. Identification was done by matching each model to the vertebra shown in the image and evaluating the similarity to the respective model. In this case, it could be shown that even if only one single vertebra is given, it could be identified in more than 70 % of the cases. However, in case of a misidentification in 84 % a shift by only one vertebra occured, and in 13 % by two vertebrae. Success rates were different for the individual vertebra which is intuitive as some vertebrae show quite characteristic features. It is worth noting that inspired by our pioneering work others have further investigated and improved vertebra identification, e.g., [14, 30].

Finally, automatic change assessment in the context of osteoporotic fracture detection was addressed as an example for CAD. The framework provides automated labeling, segmentation, and registration of vertebra pairs with high reproducibility and change sensitivity. Explicit encoding of anatomical information into the shape models, allows for efficient quantification once the model is individualized to the patient anatomy. It could be shown when using an evaluation based on the adapted shape models, pairs with a lesion in the follow-up scan can be typically well separated from the no-change cases.

Future efforts should be made to bring the framework closer to clinical practice. Currently, most results are promising but were still mostly done on a limited set of data. Thus, more work is needed to meet the requirements in a clinical workflow, e.g., in terms of robustness or computation time.

Appendix

The conversion between the two representations of rotation is described by Rodrigues' formula:

$$R = I + \sin(\theta)C(\mathbf{n}) + (1 - \cos(\theta))C(\mathbf{n})^2 \tag{32}$$

$$\text{with } C(\mathbf{n}) = \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{bmatrix}.$$

The inverse map (from a rotation matrix to a rotation vector) is given by the following equations [8]:

$$\theta = \arccos\left(\frac{tr(R) - 1}{2}\right) \quad \text{and} \quad C(\mathbf{n}) = \frac{R - R^T}{2\sin(\theta)}. \tag{33}$$

References

1. Alomari R, Corso J, Chaudhary V, Dhillon G (2010) Computer-aided diagnosis of lumbar disc pathology from clinical lower spine MRI. Int J Comput Assist Radiol Surg 5(3):287–293
2. Aslan M, Abdelmunim H, Farag A, Arnold B, Mostafa E, Xiang P (2011) A new shape based segmentation framework using statistical and variational methods. In: Proceedings of ICIP
3. Aslan M, Ali A, Arnold B, Fahmi R, Farag A, Xiang P (2009) Segmentation of trabecular bones from vertebral bodies in volumetric CT spine images. In: Proceedings of ICIP, pp 3385–3388
4. Aslan M, Ali A, Chen D, Arnold B, Farag A, Xiang P (2010) 3D vertebrae segmentation using graph cuts with shape prior constraints. In: Proceedings of ICIP, pp 2193–2196
5. Aslan M, Ali A, Farag A, Rara H, Arnold B, Xiang P (2010) 3D vertebral body segmentation using shape based graph cuts. In: Proceedings of ICPR, pp 3951–3954
6. Ballard D (1981) Generalizing the hough transform to detect arbitrary shapes. Pattern Recognition 13 (2): 111–122
7. Bauer JS et al (2006) Detection of osteoporotic vertebral fractures using multidetector CT. Osteoporos Int 17(4):608–615. doi:10.1007/s00198-005-0023-8
8. Boisvert J, Cheriet F, Pennec X, Labelle H, Ayache N (2008) Geometric variability of the scoliotic spine using statistics on articulated shape models. IEEE Trans Med Imaging 27(4):557–568
9. Boisvert J, Pennec X, Ayache N, Labelle H, Cheriet F (2006) 3D anatomical variability assessment of the scoliotic spine using statistics on lie groups. In: Proceedings of ISBI, pp 750–753
10. Boisvert J, Pennec X, Labelle H, Cheriet F, Ayache N (2006) Principal spine shape deformation modes using Riemannian geometry and articulated modes. In: Proceedings of AMDO, pp 346–355
11. Corso J, Alomari R, Chaudhary V (2008) Lumbar disc localization and labeling with a probabilistic model on both pixel and object features. In: Proceedings of MICCAI, vol LNCS 5241, pp 202–210
12. Davis B, Fletcher P, Bullitt E, Joshi S (2007) Population shape regression from random design data. In: ICCV, pp 1–7
13. Genant H et al (1993) Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res 8(9):1137–1148
14. Glocker B, Feulner J, Criminisi A, Haynor DR, Konukoglu E (2012) Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans. In: Proceedings of MICCAI
15. Jerebko A, Schmidt G, Zhou X, Bi J, Anand V, Liu J, Schoenberg S, Schmuecking I, Kiefer B, Krishnan A (2007) Robust parametric modeling approach based on domain knowledge for computer aided detection of vertebrae column metastases in MRI. In: Proceedings of IPMI, pp 713–725
16. Kadoury S, Cheriet F, Labelle H (2008) A statistical image-based approach for the 3D reconstruction of the scoliotic spine from biplanar radiographs. In: Proceedings of ISBI, pp 660–663

17. Kadoury S, Cheriet F, Labelle H (2009) Personalized X-ray 3-D reconstruction of the scoliotic spine from hybrid statistical and image-based models. IEEE Trans Med Imaging 28(9):1422–1435
18. Kadoury S, Cheriet F, Labelle H (2009) Segmentation of scoliotic spine silhouettes from enhanced biplanar X-rays using a prior knowledge bayesian framework. In: Proceedings of ISBI, pp 478–481
19. Kadoury S, Cheriet F, Labelle H (2010) Self-calibration of biplanar radiographic images through geometric spine shape descriptors. IEEE Trans Biomed Eng 57(7):1663–1675
20. Kadoury S, Labelle H, Paragios N (2011) Automatic inference of articulated spine models in CT images using high-order markov random fields. Med Image Anal 15(4):426–437
21. Kadoury S, Labelle H, Paragios N (2013) Spine segmentation in medical images using manifold embeddings and higher-order MRFs. IEEE Trans Med Imaging 32(7):1227–1238
22. Kadoury S, Paragios N (2009) Surface/volume-based articulated 3D spine inference through markov random fields. In: Proceedings of MICCAI, vol LNCS 5762, pp 92–99
23. Klinder T, Lorenz C, von Berg J, Dries S, Bülow T, Ostermann J (2007) Automated model-based rib cage segmentation and labeling in CT images. In: Proceedings of MICCAI, vol LNCS 4792, pp 195–203
24. Klinder T, Ostermann J, Ehm M, Franz A, Kneser R, Lorenz C (2009) Automated model-based vertebra detection, identification, and segmentation in CT images. Med Image Anal 13(3):471–482
25. Klinder T, Wolz R, Lorenz C, Franz A, Ostermann J (2008) Spine segmentation using articulated shape models. In: Proceedings of MICCAI, vol LNCS 5241, pp 227–234
26. Komodakis N, Tziritas G, Paragios N (2008) Performance vs computational efficiency for optimizing single and dynamic mrfs: setting the state of the art with primal dual strategies. CVIU 112(1):14–29
27. Lecron F, Boisvert J, Mahmoudi S, Labelle H, Benjelloun M (2012) Fast 3d spine reconstruction of postoperative patients using a multilevel statistical model. In: Proceedings of MICCAI, vol LNCS 7511, pp 446–453
28. Lorensen W, Cline H (1987) Marching cubes: a high resolution 3D surface construction algorithm. Comput Graphics 21:163–169
29. Lorenz C, Netsch T, Klinder T, Müller D, Baum T, Bauer J, Noel P (2013) Change assessment for CT spine imaging. In: Proceedings of ISBI, pp 97–100
30. Ma J, Lu L, Zhan Y, Zhou X, Salganicoff M, Krishnan A (2010) Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. In: Proceedings of MICCAI, vol 6361, pp 19–27
31. Mastmeyer A, Engelke K, Fuchs C, Kalender W (2006) A hierarchical 3d segmentation method and the definition of vertebral body coordinate systems for QCT of the lumbar spine. Med Image Anal 10(4):560–577
32. Müller D et al (2008) Significance of sagittal reformations in routine thoracic and abdominal multislice CT studies for detecting osteoporotic fractures and other spine abnormalities. Eur Radiol 18(8):1696–1702
33. Nadaraya EA (1964) On estimating regression. Theory Probab Appl 10:186–190
34. Rother C, Kohli P, Feng W, Jia J (2009) Minimizing sparse higher order energy functions of discrete variables. In: CVPR, pp 1382–1389
35. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326
36. Schmidt S, Kappes J, Bergtholdt M, Pekar V, Dries S, Bystrov D, Schnörr C (2007) Spine detection and labeling using a parts-based graphical model. In: Proceedings of Image Processing in Medical Imaging, vol LNCS 4584, pp 122–133
37. Tan S, Yao J, Ward M, Yao L, Summers R (2008) Computer aided evaluation of ankylosing spondylitis using high-resolution CT. IEEE Trans Med Imaging 27(9):1252–1267
38. Wiese T, Burns J, Yao J, Summers R (2011) Computer-aided detection of sclerotic bone metastases in the spine using watershed algorithm and support vector machines. In: Proceedings of ISBI, pp 152–155

39. Wiese T, Yao J, Burns J, Summers R (2012) Detection of sclerotic bone metastases in the spine using watershed algorithm and graph cut. In: Proceedings of SPIE Medical, Imaging, vol 8315, pp 8315121–8315128
40. Yao J, Burns J, Munoz H, Summers R (2012) Detection of vertebral body fractures based on cortical shell unwrapping. In: Proceedings of MICCAI, vol LNCS 7512, pp 509–516
41. Yao J, Burns J, Wiese T, Summers R (2012) Quantitative vertebral compression fracture evaluation using a height compass. In: Proceedings of SPIE, vol 8315, pp 83151X1–83151X8
42. Yao J, O'Connor S, Summers R (2006) Automated spinal column extraction and partitioning. In: Proceedings of ISBI, pp 390–393
43. Zhan Y, Maneesh D, Harder M, Zhou X (2012) Robust MR spine detection using hierarchical learning and local articulated model. In: Proceedings of MICCAI, vol LNCS 7510, pp 141–148

Shape Constraints for the Left Ventricle Segmentation from Cardiac Cine MRI Based on Snake Models

Yuanquan Wang, Yuwei Wu and Yunde Jia

Abstract Segmentation of the left ventricle (LV) is a hot topic in cardiac magnetic resonance (MR) images analysis. To make thorough use of the anatomical and functional information, it is necessary to segment the endocardium and epicardium of the left ventricle. However, automatic and accurate segmentation of the left ventricle remains a challenging problem because of papillary muscles, lack of edge information, image inhomogeneity and low contrast at the epicardium. In this chapter, we address the shape constraints for extracting the endocardium and epicardium from cardiac cine MRI based on snake models. For endocardium segmentation, a circle/ellipse-based shape energy term is incorporated into the snake model. With this prior constraint, the snake contour can conquer the unexpected local minimum stemming from artifacts and papillary muscle. After extracting the endocardium, the edge map is modified to yield a new external force field for active contours, which automatically pushes the snake contour directly to the epicardium by employing the endocardium result as initialization. However, the circle constraint does not work very well for the epicardium, and the ellipse constraint needs the troublesome calculation of the ellipse orientation during snake evolution. Assuming that the epicardium resembles the endocardium in shape, we further propose a novel shape similarity energy for epicardium segmentation. With this energy, the snake model can avoid being trapped into artifacts and leaking out at weak boundaries. Based on the circle constraint and shape similarity energy, we present an automatic algorithm

Y. Wang (✉)
School of Computer Science, Tianjin University of Technology, Tianjin 300384, People's Republic of China
e-mail: yqwang@bit.edu.cn

Y. Wu · Y. Jia
Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology (BIT), Beijing 100081, People's Republic of China
e-mail: wuyuwei@bit.edu.cn

Y. Jia
e-mail: jiayunde@bit.edu.cn

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*, Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_12,

to extract the endocardium and epicardium of the LV simultaneously. Both qualitative and quantitative evaluations on our dataset and the publicly available database (e.g., *MICCAI 2009*) demonstrate the good performance of our algorithm.

1 Introduction

Cardiovascular diseases (CVD) are the leading cause of death in most countries [1]. Nowadays, extensive techniques available for cardiac imaging provide qualitative and quantitative information about the morphology and function of the heart and great vessels [2]. These imaging modalities include angiocardiography, cardiac ultrasound (US), isotope imaging, computed tomography (CT) and magnetic resonance imaging (MRI). Cardiac magnetic resonance imaging has proven to be a versatile and non-invasive imaging modality. It can acquire the anatomical and functional information of a heart within a short period of time, and thus be widely used in clinical diagnosis due to its specific advantages:

- It owns a wide topographical field of view with visualization of the heart and its internal morphology and surrounding mediastinal structures.
- It is more favorable for the analysis of heart movements owning to the high soft-tissue contrast discrimination between the flowing blood and myocardium.

The segmentation of cardiac magnetic resonance images (MRIs) is one of the most critical prerequisites for quantitative study of the left ventricle (LV). Many clinically established diagnosis indices such as wall thickness, myocardial motion, ejection fraction, and circumferential shortening of myocardial fibers are evaluated by the segmentation results of MRIs.

In clinical practice, the LV segmentation task is often performed manually by an experienced clinician. Manual segmentation of the LV, however, is tedious, time consuming, subjective and irreproducible. This issue has motivated the development of automatic extraction of the contours of the LV. Although an impressive research effort has been devoted to automatic LV segmentation, it remains a challenging problem, mainly because of the difficulties inherent to MR cardiac images [3]. Figure 1 illustrates a typical MR image, where the blood pool appears bright and myocardium and surrounding structures appear dark. Still segmentation difficulties mainly include the following aspect:

- Since papillary muscles and the myocardium are connected, it is prone to take the papillary muscles as a part of the myocardium. According to clinical standards, papillary muscles should not be taken into account for endocardial wall segmentation.
- The myocardium and surrounding tissues such as the liver have almost the same intensity profile, leading to low contrast between them.
- A major difficulty in segmentation of the cardiac MR images is the intensity inhomogeneity due to the radio-frequency coils or acquisition sequences.

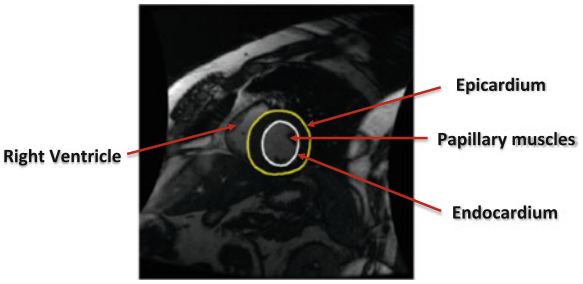


Fig. 1 A typical short axis cardiac MR image containing LV, RV and papillary muscles

There have been extensive researches such as graph cuts [4, 5], morphological operations [6, 7], dynamic weights fuzzy connectedness framework [8, 9], active contours or snake model [3, 10–14] and supervised learning methods [15–18], to overcome challenges of the LV segmentation. Petitjean and Dacher [19] presented a comprehensive review of LV segmentation algorithms. Among approaches mentioned above, the snake model is one of the most successful methods, which deforms a closed curve using both the intrinsic properties of the curve and the image information to capture the boundaries of the region of interest (ROI). However, the information (e.g. intensity, texture) only deriving from the image itself is not sufficient to get satisfactory segmentation results. The prior knowledge concerning the LV, therefore, is necessary to be incorporated into the snake model.

In this work, we address the segmentation of the LV focusing on the following challenges: (1) image inhomogeneity; (2) effect of papillary muscle; and (3) weak edges. Our strategy is the employment of the anatomical shape of the LV, rather than the constraints derived from a finite training set. The proposed method is based on the parametric snake model, in which the external forces include the gradient vector flow (GVF) and the proposed gradient vector convolution (GVC). The GVC model can be implemented in real time due to its convolutional nature and possesses similar properties of the GVF model. Considering the LV is roughly a circle, a circle-shape based energy is introduced into the snake model to extract the endocardium of the LV. The circle shape energy is also employed for epicardium segmentation. The ellipse constraint is employed for segmenting the LV as well. The drawback of the ellipse constraint is that one has to estimate the orientation of the ellipse during snake evolution. Assuming the epicardium resembles the endocardium in shape, we further develop a shape-similarity energy functional to prevent the snake contour from leaking out from weak boundaries. With the circle constraint and shape similarity energy, we can extract the endocardium and epicardium of the LV robustly and accurately. This work is based on our approaches presented in [13, 20–23].

2 Related Work

In this study, we address the shape constraints for segmentation of the LV from cine MRI based on snake models. We define intensity, texture or color statistics as *weak prior*. The prior shape obtained from a training set or from anatomical information (e.g., circle, ellipse) is regarded as *strong prior*. Based on these principles, in this section, we classify the relevant literature as *no prior*, *weak prior* and *strong prior* to review.

2.1 LV Segmentation Without Prior

The studies falling into this category have mainly concentrated on the design of the external energy. Ranganath [24] tracked the LV endocardium in cardiac MRI sequences by propagating the conventional snake from one frame to another. Makowski et al. [25] employed the balloon snake [26] to extract the LV endocardium and introduced an antitangling strategy to exclude the papillary muscles. The balloon force is defined as

$$\boldsymbol{F} = k_1 \overrightarrow{\boldsymbol{n}(s)} - k_2 \frac{\nabla E_{image}}{\|\nabla E_{image}\|}, \tag{1}$$

where $\overrightarrow{\boldsymbol{n}(s)}$ is the normal unit vector to the curve at point $C(s)$ and k_1 is the amplitude of the force. If we change the sign of k_1 or the orientation of the curve, it will have an effect of deflation instead of inflation. The curve expands and it is attracted and stopped by edges as before, but since there is a pressure force, if the edge is too weak the curve can pass through this edge. Therefore, it is prone to lead to weak edge leakage during LV segmentation. Based on the discrete contour model, Hautvast et al. [27] developed a method that attempts to maintain a constant contour environment in the vicinity of the cavity boundary. Due to the high performance at capture range enlarging, the gradient vector flow (GVF) snake [28] has been employed for the LV segmentation [29], but Santarelli et al. [29] did not consider the effect of weak boundaries, papillary muscle and artifacts stemming from swirling blood. It is also not clear how the GVF snake model captured the epicardium soon after the endocardium was extracted. The GVF field $\boldsymbol{V}(x, y) = [u(x, y), v(x, y)]$ is defined as

$$E(u, v) = \iint \left[\mu \left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) + |\nabla f|^2 |\boldsymbol{V} - \nabla f|^2\right] dxdy. \tag{2}$$

The behavior of the GVF will be discussed in Sect. 3.2. Lee et al. [30] presented the iterative thresholding method to extract the endocardium, which effectively alleviates the interference of papillary muscle. However, the endocardial contour is not smooth enough and the movement constraint based on image intensity for the snake

is too empirical. Nguyen et al. [31] compared the conventional snake, balloon snake and GVF snake on extracting the LV endocardium and concluded that the GVF snake has the best performance.

However, the information (e.g. intensity, texture) only derived from the image itself is not sufficient to get satisfactory segmentation results. It is difficult to deal with the noisy and incomplete data. The prior knowledge concerning the anatomy, appearance and motion of the LV, therefore, is necessary to be incorporated into the snake model. The prior information may be the statistical shape from a training set [11, 32, 33], be anatomical information such as ellipse in [34–36], or be intensity, texture or color statistics [37, 38].

2.2 LV Segmentation with Weak Priors

Lynch et al. [39] presented a novel and intuitive approach to combine 3-D spatial and temporal MRI data in an integrated segmentation algorithm to extract the myocardium of the left ventricle. By encoding prior knowledge about cardiac temporal evolution, an EM algorithm optimally tracks the myocardial deformation over the cardiac cycle. Punithakumar et al. [38] presented an original information theoretic measure of heart motion based on the Shannon's differential entropy (SDE), which allows heart wall motion abnormality detection. Folkesson et al. [33] extended the geodesic active region method by incorporating the statistical classifier for segmentation of cardiac MRI. Paragios [32] integrated visual information with anatomical constraint into the variational level set approach [40].

Ayed et al. [12] proposed to derive the curve evolution equations by minimizing two functionals each containing an original overlap prior constraint between the intensity distributions of the cavity and myocardium. This overlap prior constraint largely improves the segmentation accuracy and enhances the robustness of the algorithm. For each region $\boldsymbol{R} \in \{C^n, M^n, B^n, n = 1, 2, \ldots, N\}$, define $P_{\boldsymbol{R},\boldsymbol{I}}$ as the nonparametric (kernel-based) estimate of intensity distribution within region R in frame $I \in \{I^n, n = 1, 2, \ldots, N\}$

$$\forall z \in \mathbb{R}^+, P_{\boldsymbol{R},\boldsymbol{I}} = \frac{\int_{\boldsymbol{R}} K\left(z - I(\boldsymbol{X})d\boldsymbol{X}\right)}{a_{(R)}}, \tag{3}$$

where $a_{(R)}$ is the area of region $\boldsymbol{R}$

$$a_{(R)} = \int_{\boldsymbol{R}} d\boldsymbol{X}. \tag{4}$$

Ayed et al. [12] assume that a segmentation of the first frame I^1, i.e., a partition $\{C^1, M^1, B^1\}$ is given. The amount of overlap between the intensity distribution within the heart cavity region in I^n and the myocardium model learned from the first frame, is formulated as

$$\boldsymbol{B}_{in}^{n} = \mathcal{B}\left(P_{C^n,I^n}/P_{M^1,I^1}\right). \tag{5}$$

Similarly, the background model learned from the first frame is given by

$$\boldsymbol{B}_{out}^{n} = \mathcal{B}\left(P_{M^n,I^n}/P_{B^1,I^1}\right). \tag{6}$$

Mean-matching terms measures the conformity of intensity means within the cavity and the myocardium in the current frame to mean priors learned form the first frame

$$\begin{cases} \mathcal{M}_{in}^{n} &= \left(\mu_{in}^{n} - \mu_{in}^{1}\right)^2 \\ \mathcal{M}_{out}^{n} &= \left(\mu_{out}^{n} - \mu_{out}^{1}\right)^2, \end{cases} \tag{7}$$

where

$$\begin{cases} \mu_{in}^{n} &= \frac{\int_{C^n} I^n d\boldsymbol{X}}{a_{C^n}} \\ \mu_{out}^{n} &= \frac{\int_{M^n} I^n d\boldsymbol{X}}{a_{M^n}}. \end{cases} \tag{8}$$

The gradient terms is defined as

$$\begin{cases} \mathcal{G}_{in}^{n} &= \oint_{\overrightarrow{\Gamma}_{in}^{n}} (g_n + c) ds \\ \mathcal{G}_{out}^{n} &= \oint_{\overrightarrow{\Gamma}_{out}^{n}} (g_n + c) ds, \end{cases} \tag{9}$$

where c is a positive constant and g_n is an edge indicator. The functionals to minimize are a weighted sum of the three characteristic terms (i.e., overlap prior terms, mean-matching terms and gradient terms) given by

$$\begin{cases} \mathcal{F}_{in}^{n} &= \alpha \mathcal{O}_{in}^{n} + \beta \mathcal{M}_{in}^{n} + \lambda \mathcal{G}_{in}^{n} \\ &= \alpha\left(\boldsymbol{B}_{in}^{n} - \boldsymbol{B}_{in}^{1}\right)^2 + \beta\left(\mu_{in}^{n} - \mu_{in}^{1}\right)^2 + \lambda \oint_{\overrightarrow{\Gamma}_{in}^{n}} (g_n + c) ds \\ \mathcal{F}_{out}^{n} &= \alpha \mathcal{O}_{out}^{n} + \beta \mathcal{M}_{out}^{n} + \lambda \mathcal{G}_{out}^{n} \\ &= \alpha\left(\boldsymbol{B}_{out}^{n} - \boldsymbol{B}_{out}^{1}\right)^2 + \beta\left(\mu_{out}^{n} - \mu_{out}^{1}\right)^2 + \lambda \oint_{\overrightarrow{\Gamma}_{out}^{n}} (g_n + c) ds. \end{cases} \tag{10}$$

To relax the dependence on the choice of a training set, Zhu et al. [41] build subject-specific dynamic model from a user-provided segmentation of one frame in the current cardiac sequence, which is able to simultaneously handle temporal dynamics (intrasubject variability) and intersubject variability. The Bayesian framework combining the forward and backward segmentation is expressed as

$$\begin{aligned} \widehat{\boldsymbol{s}_t} &= \arg\max_{\boldsymbol{s}_t} P\left(\boldsymbol{s}_t | I_{1:N}\right) \\ &= \arg\max_{\boldsymbol{s}_t} P\left(I_t | \boldsymbol{s}_t\right) P\left(\boldsymbol{s}_t | \widehat{\boldsymbol{s}}_{1:t-1}^{+}\right) P\left(\boldsymbol{s}_t | \widehat{\boldsymbol{s}}_{t+1:N}^{-}\right) \\ &= \arg\max_{\boldsymbol{s}_t} \left\{ \underbrace{\log P\left(I_t | \boldsymbol{s}_t\right)}_{data\ adherence} + \underbrace{\log P\left(\boldsymbol{s}_t | \widehat{\boldsymbol{s}}_{1:t-1}^{+}\right)}_{forward\ dynamics} + \underbrace{\log P\left(\boldsymbol{s}_t | \widehat{\boldsymbol{s}}_{t+1:N}^{-}\right)}_{backward\ dynamics} \right\}. \end{aligned} \tag{11}$$

Ayed et al. [3] introduced a novel max-flow segmentation of the LV by recovering subject-specific distributions learned from the first frame via a bound of the Bhattacharyya measure. The total cost function is given by

$$\arg \min_{L:P\to 0,1} \Bigg[-\sum_{i\in I} \sqrt{\boldsymbol{P}_{L,I^n}(i)\mathcal{M}_{\boldsymbol{c},I}(i)} - \sum_{d\in D} \sqrt{\boldsymbol{P}_{L,D^c}(d)\mathcal{M}_{\boldsymbol{c},D}(d)} + \sum_{\{p,q\}\in\mathcal{N}} \frac{1}{||p-q||}\delta_{L(p)\neq L(q)} \Bigg], \tag{12}$$

where $\mathcal{M}_{\boldsymbol{c},I}$ is the learned model distribution of intensity and $\mathcal{M}_{\boldsymbol{c},D}$ is the model distribution of distances within the cavity in the first frame. Jolly et al. [42] combined the edge, region and shape information to extract the LV endocardium, the approximate shape of the LV is obtained based on the maximum discrimination method. They used the shape alignment method proposed by Duta et al. [43] to establish the correspondence between a subset $A^{'}$ of the template points and a subset $B^{'}$ of the candidate points. The goal of shape constraint in [42] is to find the coefficients (a, b, c, d) of the similarity transform which minimize the distance $f(N_c)$.

$$f(N_c) = \frac{1}{N_c^2} \sum_{j=1}^{N_c} w(B_j) \Big[\big(x_{A_j} - a x_{B_j} - c y_{B_j} - b\big)^2 + \big(y_{A_j} - a y_{B_j} - c x_{B_j} - d\big)^2 \Big] + \frac{2}{N_c}, \tag{13}$$

where N_c is the number of established correspondence.

2.3 LV Segmentation with Strong Priors

As shown by the growing literature on the LV segmentation, it benefit from the use of an anatomical constraint (e.g., shape model) on active contour model to enhance the robustness and accuracy of the segmentation. Pluempitiwiriyawej et al. [35] incorporated the ellipse constraint into the segmentation scheme. However, there should be an isolated step to estimate the five parameters of the ellipse, which does not comply with the evolution of the snake contour.

The atlas warping technique introduced by Lorenzo-Valdes et al. [44] is a typical training-based method, the atlas is constructed from manually segmented and temporally aligned data and is registered on the data for automatic segmentation. This algorithm is based on the 4D extension of expectation maximization (EM) algorithm. The active shape model (ASM) first proposed by Cootes et al. [15] is constructed from a training set of segmented objects using the principal component analysis (PCA) algorithm. The ASM minimizes the difference between the synthesized image from the model and an unseen image by tuning the model parameters, when it is applied to image interpretation or segmentation. Assen et al. [45] applied 3D-ASM to image

data sets with an increasing sparsity comprising images with different orientations and stemming from different MRI acquisition protocols. Later, Assen et al. [46] investigated the effect of initialization on segmentation results in [45] and proposed a semiautomatic segmentation method of cardiac CT and MR volumes, without the requirement of retraining the underlying statistical shape model.

The active appearance model (AAM) is extended from the ASM by taking into account the intensity distribution in the training set [16, 47]. AAM is able to simultaneously describe the shape and texture variation of objects. The AAM is more robust than the ASM when the intensity contrast is low and the object boundary is weak. There has been a great diversity of works devoted to the construction and application of the ASM/AAM models, particularly for the extraction of the LV from cardiac MRI. Gopal et al. [48] applied AAM and deformable superquadric models to automate the segmentation of the LV in cardiac MR cine images for the end-diastole and end-systole phases, in which the texture model in the AAM is modified by radially sampling gradient magnitude values. Ghose et al. [49] applied 2D AAM with a 3D shape restriction imposed by rigidly registering the obtained volume to a 3D average model of the prostate. Mitchell et al. [50] devised a multistage hybrid active appearance model (AAM) by combining ASM and AAM. Pfeifer et al. [51] used a 2D AAM for the blood mass of left and right ventricle and an additional three division 2D AAM to cope with the shape variation of the blood mass of the left and right ventricle. Zambal et al. [52] combined AAM and ASM for the LV segmentation, in which the global model construction interconnects a set of 2D AAM by a 3D shape model. Recently, Zhang et al. [17] also combined the ASM and AAM but to construct a biventricular model to segment the left and right ventricles simultaneously. Although excellent results have been achieved in [45, 48, 51, 52] where the LV shape is learned from an annotated training data set, the segmentation performance depends heavily on the size and richness of images in the training set.

3 Background: Active Contours

3.1 Tranditional Active Contours

Active contour models, or snakes [10], have been proven to be very effective tools for image segmentation. It integrates an initial estimate, geometrical properties of the contour, image data and knowledge-based constraints into a single process, and provides a good solution to shape recovery of objects of interest in visual data. A traditional active contour model is represented by a curve $C(s) = (x(s), y(s))$, $s \in [0, 1]$. It moves through the spatial domain of an image to minimize the energy functional

$$E(C) = \int_0^1 \underbrace{\frac{1}{2}\left(\alpha \mid C'(s) \mid^2 + \beta \mid C''(s) \mid^2\right)}_{Internal\ \ energy} + \underbrace{g\big(C(s)\big)}_{External\ \ energy} + \underbrace{E_{Con}\big(C(s)\big)}_{Constraint\ \ energy} ds, \tag{14}$$

where $C'(s)$ and $C''(s)$ denote the first and second derivatives of C' with respect to s, respectively. The first term of the integral stands for the internal force that keeps the contour continuous and smooth during deformation, the second term is the external force that drives the contour toward an object boundary or the other desired features within an image. The constraint energy term $E_{Con}\big(C(s)\big)$ is derived from the prior information (e.g., in this chapter, we concentrate on designing a novel constraint energy by considering the anatomical information of the LV to enhance the robustness and accuracy of the LV segmentation). The external energy function $g\big(C(s)\big)$ is derived from the image so that it takes on its smaller values at the features of interest, such as boundaries. Given a gray-level image $I(x, y)$, typical external energies are defined as follows,

$$\begin{aligned} g(x, y) &= -|\nabla I(x, y)|^2 \\ g(x, y) &= -|\nabla (G_\sigma(x, y) \otimes I(x, y))|^2, \end{aligned} \tag{15}$$

where $G_\sigma(x, y)$ is a two-dimensional Gaussian function with standard deviation σ, ∇ is the gradient operator, and $\otimes$ is convolution operation. If the image is a line drawing (black and white), then appropriate external energies include

$$\begin{aligned} g(x, y) &= I(x, y) \\ g(x, y) &= G_\sigma(x, y) \otimes I(x, y). \end{aligned} \tag{16}$$

By using the calculus of variation, the Euler equation to minimize $E(C)$ is

$$\alpha C''(s) - \beta C''''(s) - \nabla g\big(C(s)\big) - \nabla E_{con}\big(C(s)\big) = 0. \tag{17}$$

According to the representation and implementation, active contour models are classified into two categories: the parametric active contour models [26, 28, 53] and the geometric active contour models [54–57]. In this paper, we focus on the parametric active contour models, and our approach can be also integrated into geometric active contour models. Since the external force plays a leading role in driving the active contours to approach objects boundaries in the parametric active contour models, designing a novel external force field has been extensively studied [26, 28, 58–61]. Among all these external forces, gradient vector flow (GVF) proposed by Xu and Prince [28], has been one of the most successful external forces, which is computed as a diffusion of the gradient vectors of a gray-level or binary edge map derived from a given image to increase the capture range. Due to the outstanding properties of GVF, a large number of modified versions have been presented [58, 60, 62, 63] to improve the performance of active contour models. However, researchers found

that the GVF suffers from several challenges including narrow and deep concavity convergence as well as weak edge leakage. In the next section, we elaborate the behavior of the GVF snake model.

3.2 GVF Snake Model

Notwithstanding the marvelous ability in representing object shapes, the traditional active contour model is limited to capture range and poor convergence to boundary concavities. Gradient vector flow (GVF) was proposed by Xu and Prince [28] as a new external force for active contour model to overcome these issues. It is a dense vector field, generated by diffusing the gradient vectors of a gray-level or binary edge map derived from an image. The GVF field is defined as a vector field $\boldsymbol{V}(x, y) = [u(x, y), v(x, y)]$ that minimizes the following energy functional:

$$E(u, v) = \iint \left[\mu \left(u_x^2 + u_y^2 + v_x^2 + v_y^2\right) + |\nabla f|^2 |\boldsymbol{V} - \nabla f|^2\right] dxdy, \tag{18}$$

where f is the edge map defined as

$$\begin{aligned} f(x, y) &= |\nabla I(x, y)| \\ f(x, y) &= |\nabla (G_\sigma(x, y) \otimes I(x, y))|. \end{aligned} \tag{19}$$

$|\nabla f|$ is high near the edges and nearly zero in homogeneous regions and μ is a positive weight to control the balance between smoothness energy and edge energy. We see that when $|\nabla f|$ is small, the energy is dominated by sum of the squares of the partial derivatives of the vector field, yielding a slowly-varying field. On the other hand, when $|\nabla f|$ is large, the second term dominates the integrand and minimized by setting $\boldsymbol{V} = |\nabla f|$. This produces the desired effect of keeping $\boldsymbol{V}$ nearly equal to the gradient of the edge map when it is large, but forcing the field to be slow-varying when in homogeneous regions.

By the calculus of variation, the minimization of Eq. (18) reduces to solving the following Euler-Lagrange equation:

$$\mu \nabla^2 \boldsymbol{V} - (\boldsymbol{V} - \nabla f)\left(f_x^2 + f_y^2\right) = 0. \tag{20}$$

The Euler-Lagrange equations evolving Eq. (20), embedded into a dynamic scheme by treating $\boldsymbol{V}(x, y)$ as the function of t, x and y, formally are

$$\begin{cases} \frac{\partial u}{\partial t} = \mu \cdot \underbrace{\nabla^2 u}_{diffusion\ \ term} - \underbrace{(u - f_x)\left(f_x^2 + f_y^2\right)}_{data\ \ attraction\ \ term} = 0 \\ \frac{\partial v}{\partial t} = \mu \cdot \underbrace{\nabla^2 v}_{diffusion\ \ term} - \underbrace{(v - f_y)\left(f_x^2 + f_y^2\right)}_{data\ \ attraction\ \ term} = 0\,, \end{cases} \tag{21}$$

where ∇^2 is the Laplacian operator. The active contour model with $\boldsymbol{V}(x, y)$ as external force is called GVF active contour model.

GVF has successfully addressed the issues of building a satisfactory capture range and approaching boundary concavities, e.g., U-shape concavity convergence. However the GVF active contour model still fails to converge to narrow and deep concavity and would leak out around weak edges, especially neighbored by strong ones. More importantly, the computational cost of the GVF model is very high. In the next Section, we will elaborate the calculation of the GVF.

4 A New External Force: Gradient Vector Convolution

4.1 Computational Analysis of the GVF

The computation of a GVF field consists mainly of solving a huge discretized system of partial differential equations, which has restricted its potential applications to images with large sizes. Among the general numerical implementation of the GVF filed, diffusion is an iterative technique that depends upon a termination time. This involves how to choose an optimal iteration number in diffusion process. As discussed in [28] ,"*the steady-state solution of these linear parabolic equations is the desired solution of the Euler equations ...* ," this statement gives rise to the following question: dose "*the desired solution of the Euler equations*" be the desired external force for Snake model, i.e., the desired GVF? We answer this question and demonstrate the influence of iteration number in Fig. 2. GVF fields at 100, 200 and 2000 iterations of diffusion are given in Fig. 2a–c respectively. Visibly, the result in Fig. 2c is far from available in that the GVF flows into the ventricle from right and out from left-bottom depicted in the white rectangle. Surely, the result in Fig. 2c approximates the steady state solution, but it cannot serve as the external force for snake model. The reason behind this situation is that Eq. (21) is a biased version of $\boldsymbol{V}_t = \mu\nabla^2\boldsymbol{V}$ by $(\boldsymbol{V} - \nabla f)\left(f_x^2 + f_y^2\right)$, where $\boldsymbol{V}_t = \mu\nabla^2\boldsymbol{V}$ is an isotropic diffusion. As t increases, the isotropic smoothing effect will dominate the diffusion of Eq. (21) and converge to the average of the initial value. Small enough μ could depress this oversmoothing efficacy, but, at the same time, preserves excessive noise. Alternatively, an optimal iteration number, say, 200 for this example, would be an effective solution for this issue, however, it is still computationally expensive.

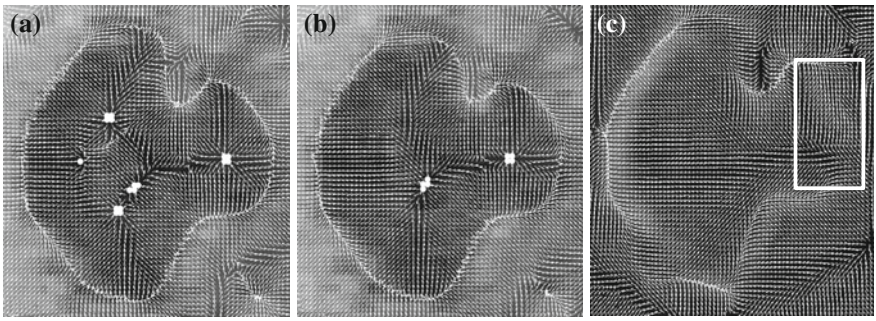


Fig. 2 Gradient vector flow fields at different iteration: **a** GVF at 100 iteration; **b** GVF at 200 iteration; **c** GVF at 2000 iteration. The white dots represent critical points of the flow field [22]. The parameters of GVF are $\mu = 0.15$, time step $\tau = 0.5$

One possible solution is to develop alternative external forces which can be efficiently computed. For example, Park and Chung [64] and Yuan and Lu [65] simultaneously proposed the virtual electric field (VEF) based external force, in which each pixel is considered as a static charge. The VEF possesses the advantage of being implemented in real time over the GVF while maintaining other desirable properties such as large capture range and concavity convergence. Jalba et al. [66] recently proposed the charged particle model, where each pixel is also considered as static charge. The authors demonstrated that the particles could be initialized randomly across the image and did not suffer from convergence issues related to GVF/GGVF. Li and Acton et al. [67] proposed convolving the image edge map with a vector field kernel, which comprises radial symmetric vectors pointing towards the center of the kernel.

Another solution is to design efficient numerical schemes for fast GVF computation. Vidholm et al. [68] employed the preconditioned conjugate gradient methods and multigrid methods to accelerate the computation of 3D GVF fields. Han et al. [69] proposed a multigrid GVF/GGVF algorithm, which can significantly improve the computational efficiency. Boukerroui [70] compared several efficient numerical schemes for GVF computation, and showed that the alternating direction explicit scheme (ADES) may be a suitable alternative to the multigrid method. Ren et al. [71] presented methods for fast GVF and GGVF computation by combining the augmented Lagrangian method and multiresolution scheme.

To overcome issues mentioned above, we argue that one can extend the gradient vector and suppress noise by convolving the gradient vector with a certain kernel. To this end, a convolution-based external force called gradient vector convolution (GVC) will be introduced in Sect. 4.2. Due to the fast fourier transform, the convolution operation would be implemented in real time and the snake model would benefit much from this convolution operation in computation time. The GVC method possesses some advantages of the GVF such as enlarged capture range, initializa-

tion insensitivity, concavity convergence, but its computational cost is low owing to its convolution mechanism. Some experiments are presented to demonstrate these advantages in Sect. 4.2.

4.2 Gradient Vector Convolution External Force

For any bounded $g \in \mathbb{R}^2$, the linear diffusion process $u_t = \nabla^2 u$, $u(x, 0) = g(x)$ possesses the unique solution $u(x, t) = \left(G_{\sqrt{2t}} \otimes g\right)(x)$, $t > 0$, where $\otimes$ denotes convolution, $G_{\sqrt{2t}}$ is the Gaussian kernel of standard deviation $\sqrt{2t}$. We argue that the solution of Eq. (21) can be approximated by convolving the $\nabla f = [f_x, f_y]$ with a kernel. This convolution-based external force is referred to as gradient vector convolution (GVC). Followed by fast Fourier transform, this convolution operation can be implemented in real time and the snake model would benefit much from this property in computation time. Denote the convolution kernel by K_{con}, the GVC takes the following form:

$$\begin{cases} u(x, y) = K_{con} \otimes f_x \\ v(x, y) = K_{con} \otimes f_y \ . \end{cases} \tag{22}$$

The GVC model possesses the desirable properties of the well-known GVF model such as enlarged capture range and insensitivity to initialization, whilst it can be implemented in real time owing to its convolution mechanism.

In practice, we take $K_{con} = \frac{1}{(r_h)^n}$, where $r_h = \sqrt{x^2 + y^2 + h}$, $h \in R^+$, $n \in R^+$. K_{con} always works well in terms of extending and smoothing gradient vector. Generally, the factor h plays a role analogous to scale space filtering, the greater the value of h, the greater the smoothing effect on the results. This property suggests that GVC would be robust to noise. In addition, large n makes the potential to degrade fast with distance and vice versa. Thereby it allows the GVC snake to preserve edges and to drive into C-shape concavities.

In order to well understand the behavior of h and n, we plot the proposed kernel K_{con} in 1D case with different h and n in Fig. 3. It can be seen from Fig. 3a that, the larger the value of h, the smaller the value of K_{con} at points nearby $x = 0$, but almost unchanged at points far from $x = 0$. Note that K_{con} is not defined at $x = 0$ when $h = 0$, we set $K_{con}(0) = K_{con}(1)$ for the sake of exhibition. Similar strategy is employed in Fig. 3b. From Fig. 3b, we can observe that the larger the value of n, the faster K_{con} degrades with distance. For example, although point A is 3 while B is 9 far from $x = 0$, due to varying n, the values of K_{con} at point A and B are almost identical. It seems as if the point B is as near as A to $x = 0$ in terms of the value of K_{con}. Similar results can be observed for points C and D and it seems as if the point C is as far as D from $x = 0$. As a result, if one wants to separate two closely-neighbored objects or preserve edges, one can use large n such that nearby points are less weighed as if they are far away. On the other hand, if concavity is too

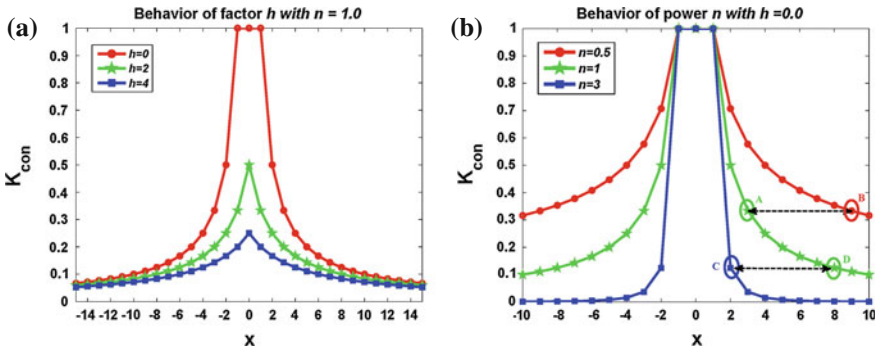


Fig. 3 Analysis of the behavior of h and n in 1D case

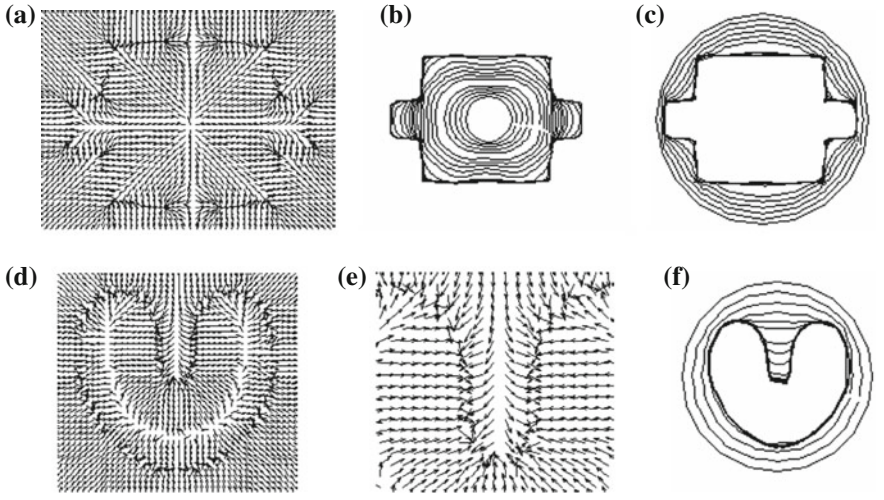


Fig. 4 The performance of GVC snake on U-shape and room images: **a** is the GVC field of room image; **b** and **c** are the convergence of the GVC snakes with the initial contours inside and outside the room, respectively; **d** is the GVC field of U-shape image; **e** is the close-up of GVC field within the concavity; **f** is the convergence of the GVC snake on the U-shape image

deep, small n can be employed to weigh relatively more on faraway points as if they are nearby.

Figure 4 shows two examples of the GVC snake. These experiments are implemented in MATLAB on an Intel Core2 2.66 GHz processor with 2GB RAM. The room and U-shape images are coined in [28] to demonstrate capture range enlarging and concavity convergence. The size of both images is 64×64. The parameters for GVC are: $h = 0, n = 2.0$, the kernel size is the same as that of the image. The GVC is able to obtain similar results as the GVF (see [28]). It is worth noting that the execution time of GVC for both images is 0.027 s while that of GVF is 2.36 s with 50 iterations.

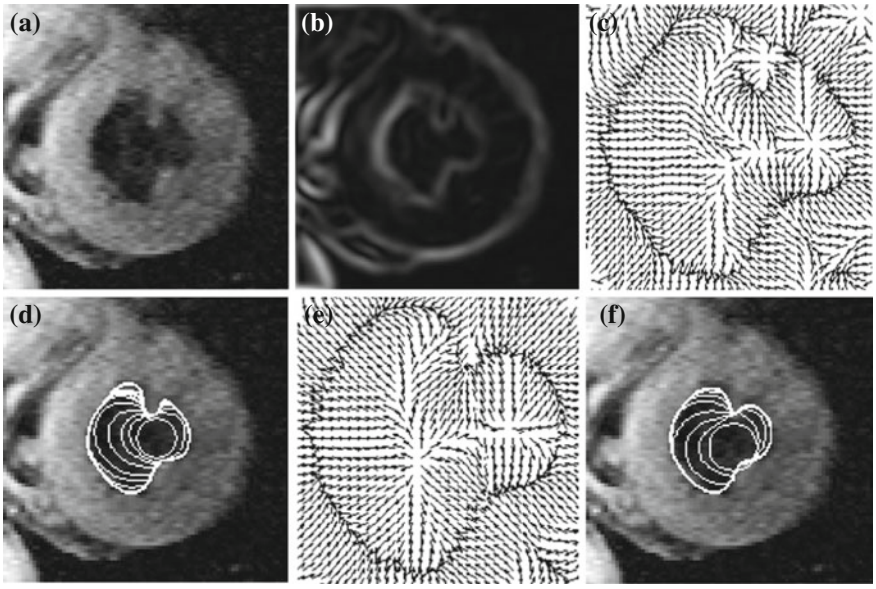


Fig. 5 Comparison of GVC and GVF on the heart image

As a second example, a comparative study of GVC and GVF is conducted on the heart image used in [28], shown in Fig. 5. The heart image is first blurred with a Gaussian kernel with standard deviation 3.0, and the edge map is shown in Fig. 5b. The GVC is calculated with $n = 2.0$, $h = 2.0$, and execution time is 0.531 s, as shown in Fig. 5c. A circle is used as initial contour, the snake contour can correctly locate the left ventricle and the result is in Fig. 5d. The GVF is also calculated, for all GVF in this experiment, $\mu = 0.15$ and time step is 0.5. Figure 5e presents the GVF after 200 iterations, and execution time is $6.407s$. However, due to the critical points in GVF [22], a larger circle including the critical points is employed as initialization and the segmentation result is fairly satisfactory, see Fig. 5f. This result implies that the critical point issue of GVC snake is less serious than that of the GVF snake. In addition, comparing Fig. 5d and f, although the initial contour in Fig. 5f is larger, but the segmentation result of the GVC snake is much better than that of the GVF snake. We also observe this phenomenon in the room image that GVC snake can detect the corner more accurately than that of the GVF snake. Let us keep in mind: *the computation time of the GVC is much shorter than that of the GVF.*

Furthermore, we use the C-shape image of 256 × 256 pixels to verify the performance of the GVC snake on concavity convergence. We apply the GVF [28], VEF [64] , VFC [67] and GVC snakes to a C-shape image, as shown in Fig. 6. The difference between C-shape concavity and U-shape concavity is that the C-shape is semi-close, while the U-shape is open. The results show that the GVC snake evolves into the concave region progressively, steadily and correctly. In contrast, others fail.

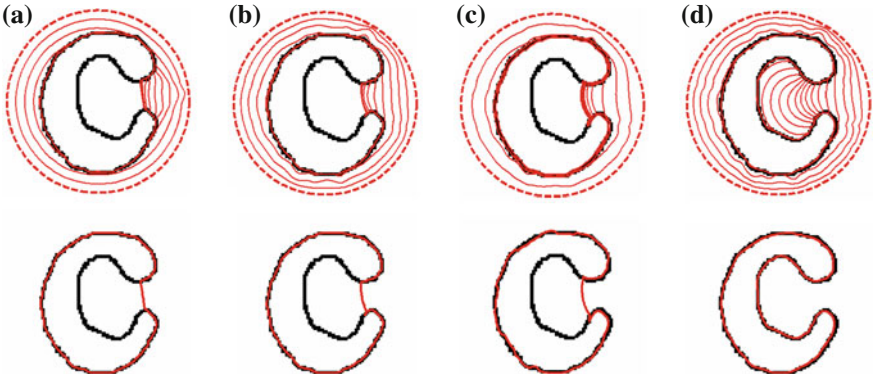


Fig. 6 Comparisons of C-shape image convergence on GVF, VEF, VFC and GVC snakes. Note that the *dashed red lines* represent the initial curves, and the *solid red lines* denote the final active contours

The success of GVC snake is ascribed by the larger weight on faraway points with a small n. The parameters of GVC snake are: $\alpha = 0.5$, $\beta = 0.5$, time step $\tau = 1$, $h = 1$ and $n = 2.6$.

5 Left Ventricle Segmentation with Shape Constraints

Though MRI provides quite good contrast between the myocardium and the blood pool, the difficulties in segmenting the endocardium originate primarily from artifacts and papillary muscles. In the classical internal energy of snake model (see Eq. (14)), the first and second derivatives control the continuity and smoothness of the curve, respectively. However, continuity and smoothness are only local geometrical properties. If the object is partially occluded or there are weak boundaries, the snake contour would not know how to bridge such gaps because there is no prior information about the overall shape of the object. If there are local minima caused by imperfectness of external force, the snake contour would get trapped. A solution to these issues is to incorporate the object overall shape into snake energies since the overall shape is a global property. The total energy of the snake model with shape constraints can be written as follows:

$$E(C) = \int_0^1 \underbrace{\frac{1}{2}\left(\alpha \mid C'(s) \mid^2 + \beta \mid C''(s) \mid^2\right)}_{Internal\ \ energy} + \underbrace{g\big(C(s)\big)}_{External\ \ energy} + \underbrace{E_{Con}\big(C(s)\big)}_{shape\ constraints} ds. \tag{23}$$

In this section, we focus on designing an effective shape constrains for LV segmentation based on GVF snake and GVC snake.

5.1 A Preliminary Study: GVF Snake Based LV Segmentation with Circle Constraint

5.1.1 Endocardium Segmentation

Observed that the endocardium of the LV is roughly a circle, a circle-shape constraint [72] is adopted for the endocardium segmentation. It is formulated as

$$E_{cir}(x, y) = \frac{\lambda}{2} \int_0^1 \left(R(s) - \overline{R}\right)^2 ds, \tag{24}$$

where

$$\begin{cases} R(s) = \sqrt{\left(x(s) - x_c\right)^2 + \left(y(s) - y_c\right)^2} \\ x_c = \int_0^1 x(s)ds \\ y_c = \int_0^1 y(s)ds \\ \overline{R} = \int_0^1 R(s)ds. \end{cases} \tag{25}$$

(x_c, y_c) is the centroid of the snake contour. The energy Eq. (24) measures the deviation of the snake contour from a circle with radius $\overline{R}$ and center (x_c, y_c). Both $\overline{R}$ and (x_c, y_c) are dynamic with the evolution of the snake contour. If the snake contour is attracted by artifacts or papillary muscle, this constraint would penalize the snake contour to be a circle, thus, the global shape of the LV would be maintained.

Suppose that there are n discrete points on the snake contour, the center (x_c, y_c) can be estimated by the following equations:

$$\begin{cases} x_c = \frac{1}{n} \sum_{i=1}^n x_i \\ y_c = \frac{1}{n} \sum_{i=1}^n y_i \\ R(i) = \sqrt{\left(x_i - x_c\right)^2 + \left(y_i - y_c\right)^2} \\ \overline{R} = \frac{1}{n} \sum_{i=1}^n R_i, \end{cases} \tag{26}$$

where $i = 1, 2, \ldots, n$. Since

$$\begin{cases} \left(R(s) - \overline{R}\right)^2 = \left(R(s) - \overline{R}\right)^2 \left(\cos^2(2\pi s) + \sin^2(2\pi s)\right) \\ \qquad\qquad = \left(R(s)\cos(2\pi s) - \overline{R}\cos(2\pi s)\right)^2 + \left(R(s)\sin(2\pi s) - \overline{R}\sin(2\pi s)\right)^2 \\ R(s)\cos(2\pi s) = x(s) - x_c \\ R(s)\sin(2\pi s) = y(s) - y_c, \end{cases}$$

by the calculus of variation, the discrete Euler equation of Eq. (24) is given by

$$\begin{cases} \lambda\left(x_i - x_c - \overline{R}\cos(2\pi i/n)\right) = 0 \\ \lambda\left(y_i - y_c - \overline{R}\sin(2\pi i/n)\right) = 0\,. \end{cases} \tag{27}$$

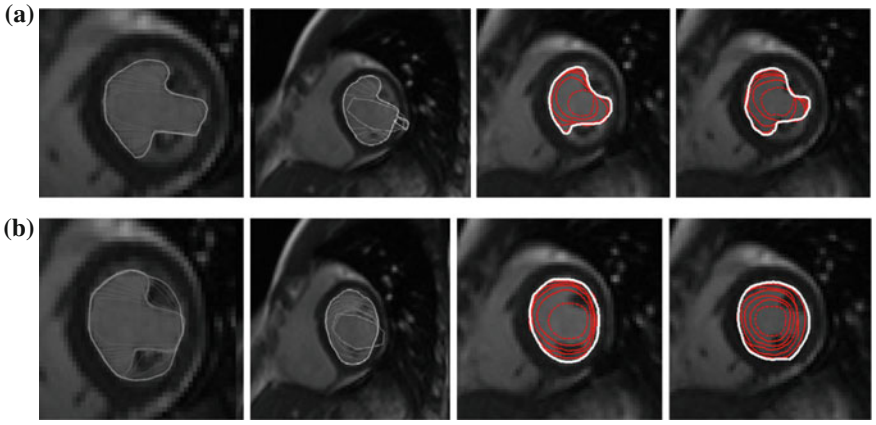


Fig. 7 Effectiveness of the *circle-shape* energy for endocardium segmentation. **a** Failed segmentations without the *circle-shape* constraint. **b** Succeeded segmentations with the *circle-shape* constraint

The solution of Eq. (27) obtained by treating x and y as the functions of time t is expressed as

$$\begin{cases} \frac{x_i^{t+1}-x_i^t}{\Delta t} + \lambda x_i^{t+1} - \lambda\left(x_c^t + \overline{R}^t \cos(2\pi i/n)\right) = 0 \\ \frac{y_i^{t+1}-y_i^t}{\Delta t} + \lambda y_i^{t+1} - \lambda\left(y_c^t + \overline{R}^t \sin(2\pi i/n)\right) = 0\,. \end{cases} \tag{28}$$

Equation (28) is incorporated into the shape constraints term in Eq. (23) to extract the endocardium, and the force vector $-\nabla g\big(C(s)\big)$ in Eq. (17) will be replaced by $[u(x, y), v(x, y)]$ in Eq. (21). The total energy of the GVF snake model with shape constraints can be written as follows:

$$E(C) = \int_0^1 \underbrace{\frac{1}{2}\left(\alpha \mid C'(s) \mid^2 + \beta \mid C''(s) \mid^2\right)}_{Internal\ \ energy} + \underbrace{g\big(C(s)\big)}_{GVF} + \underbrace{E_{Cir}\big(C(s)\big)}_{circle\ constraint} \; ds. \tag{29}$$

We demonstrate the effectiveness of the proposed circle shape constraint for segmenting the LV. Figure 7 illustrates segmentation results of the endocardium by using the circle-shape energy to conquer the papillary muscle and artifacts. The images are taken from mid-ventricle slice, where the papillary muscles are obstacles for the GVF snake model. When the initial contour excludes the papillary muscles, the snake contour halts at the papillary muscles and artifacts (see Fig. 7a) unless the initial contour is close enough to the endocardium. In contrast, when the global shape constraint is incorporated into the snake model, the snake contour conquers the papillary muscles successfully and sticks to the endocardium (see Fig. 7b).

We can extend the circle-shape constraint to the ellipse-shape constraint [72] which is a more versatile shape model. The ellipse-based energy function is defined as

$$E_{elli}(x,y) = \frac{\lambda}{2}\int_0^1 \Big[\Big(\big(x(s)-c_x\big)\cos(\theta) + \big(y(s)-c_y\big)\sin(\theta) - r_1\cos(2\pi s-\theta)\Big)^2 + \Big(-\big(x(s)-c_x\big)\sin(\theta) + \big(y(s)-c_y\big)\cos(\theta) - r_2\sin(2\pi s-\theta)\Big)^2\Big] ds, \tag{30}$$

where θ is the orientation of the semimajor axis of the ellipse with the x-axis, r_1 and r_2 represent the two radius of the ellipse, x_c and y_c denote the center coordinates, respectively. Here, we do not allow the snake shape to deviate significantly from an ellipse. It is clear from the formulation that we penalize the snake if it deviates from the best-fitted ellipse with the method of direct least squares method in the process of snake contour evolution every time. If there is no external force, the snake contour would just be an ellipse. If $r_1 = r_2$, i.e., the ellipse is a circle, then Eq. (30) will reduce to already mentioned circle-shape based constraint in Eq. (24). Once we obtain the values of parameters $[\theta, c_x, c_y, r_1, r_2]$, the Euler equations of the functional (30) is given by

$$\begin{cases} \lambda\Big(x(s) - c_x - r_1\cos(2\pi s-\theta)\cos(\theta) + r_2\sin(2\pi s-\theta)\sin(\theta)\Big) = 0 \\ \lambda\Big(y(s) - c_y - r_1\cos(2\pi s-\theta)\sin(\theta) - r_2\sin(2\pi s-\theta)\cos(\theta)\Big) = 0\,. \end{cases} \tag{31}$$

The solution of Eq. (31) obtained by treating x and y as the functions of time t is expressed as

$$\begin{cases} \frac{x_i^{t+1}-x_i^t}{\Delta t} + \lambda x_i^{t+1} \\ -\lambda\Big(c_x^t + r_1^t\cos(2\pi i/n+\theta^t)\cos(\theta^t) - r_2^t\sin(2\pi i/n-\theta^t)\sin(\theta^t)\Big) = 0 \\ \frac{y_i^{t+1}-y_i^t}{\Delta t} + \lambda y_i^{t+1} \\ -\lambda\Big(c_y^t + r_1^t\cos(2\pi i/n-\theta^t)\sin(\theta^t) - r_2^t\sin(2\pi i/n-\theta^t)\cos(\theta^t)\Big) = 0\,. \end{cases} \tag{32}$$

Similarly, Eq. (32) is incorporated into the shape constraints term in Eq. (23) to extract the endocardium, and the force vector $-\nabla g\big(C(s)\big)$ in Eq. (17) will be replaced by $[u(x,y), v(x,y)]$ in Eq. (21).

5.1.2 Epicardium Segmentation

The contrast between the myocardium and surrounding tissues (e.g. fat, lung and liver) is poor, and thus it would be more difficult to segment the epicardium [19]. The energy Eq. (24) regularize the snake to a canonical circle, but for the epicardium, the uniformly weighted circle shape constraint would cause poor performance because the left ventricle is not an exact circle but more like an ellipse and there are gaps on the

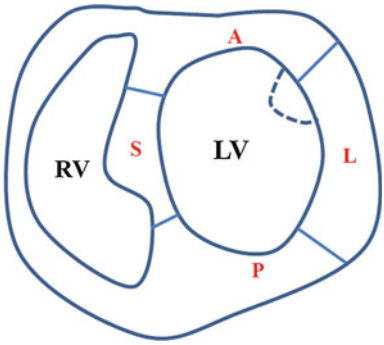


Fig. 8 Illustration of cardiac anatomy

epicardium boundary as well. As illustrated in Fig. 8, the LV can be generally divided into four partitions: septum, anterior, lateral and posterior, denoted as "S", "A", "L", and "P", respectively. The boundary at septum interfacing the RV is salient; the anterior and posterior parts are in junction with the RV, and therefore the boundaries are fraudulent; the lateral part interfaces the lung and the boundaries are usually not as salient as those at septum, possibly very weak. When the circle shape energy is applied uniformly, it would be too strong for some parts but too weak for others, so, we employ different weights for different partitions as $\lambda_S, \lambda_A, \lambda_L, \lambda_P$. In practice, more weights can be adopted according to the image quality, for instance, the posterior also partially neighbors other organs, therefore, one more weight can be associated to it.

Because the endocardium is a local minimum of the GVF field, the snake contour would become stationary there and it is impossible to extract the epicardium successively after the endocardium is located. But as far as we know, among the existent studies based on GVF snake, there is one exception in [29]; however, it is not clear how their method sequentially "*search for the first local minimum that is expected to correspond to the epicardium surface*". In this study, we develop a novel strategy to erase the endocardium, and then generate the external force for epicardium locating.

We take the center of the endocardium as reference and generate two edge maps by selectively choosing the gradient in direction of x and y according to the characteristics of the image. Taking the image in Fig. 9a as an example, on the left of the center, the septum interfaces the blood pools of RV and LV, the gradient in direction of x is taken as $f_l^x = |\min(I_x, 0)|$ such that the edge between septum and the LV blood pool is neglected; but for the right part, the gradient is $f_r^x = |\max(I_x, 0)|$ while the lateral is surrounded by fat (if no fat, choosing $f_r^x = |\min(I_x, 0)|$), thus, the edge between lateral and the LV blood pool and the one between the fat and the lung are neglected. Likewise, the gradient in direction of y up and down the center are $f_u^y = |\min(I_y, 0)|$ and $f_d^y = |\min(I_y, 0)|$. Then f_x and f_y are utilized as edge maps to calculate GVF external force.

Figure 9 shows the edge maps derived from this strategy, Fig. 9b is the usual edge map, and Fig. 9c, d are f_x and f_y, respectively. Generally speaking, this method is robust to noise because the min-max operation can weaken the side effects of noise; meanwhile, it depends on the characteristics of the image which should be known in advance.

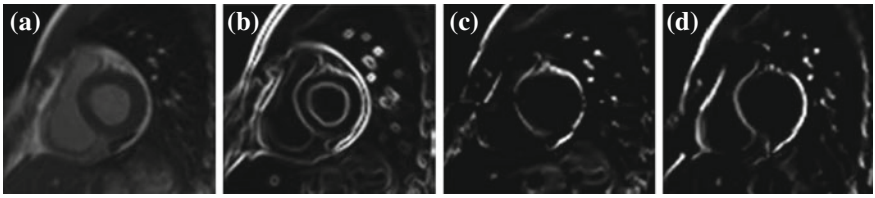


Fig. 9 Edge maps for epicardium locating

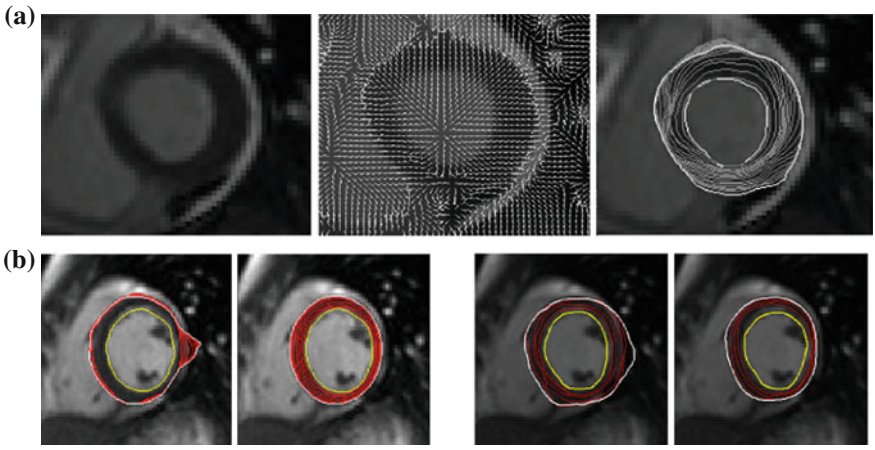


Fig. 10 Segmentation of the epicardium. **a** Epicardium extraction using new external force. **b** Comparison of segmentation results with and without shape energy

Figure 10a illustrates the epicardium segmentation; the left image is gray level, the external force by the proposed method is overlaid on the middle and the right is the segmentation result. Although the external force is not perfect, it can push the snake contour directly to the epicardium, with the shape energy, its imperfectness is conquered. Figure 10b shows segmentation results with and without shape energy. There are two groups and for each group, the left is the results without shape constraint and the right is with shape constraint. We can see that the snake contour with shape energy can prevent the snake contour from being trapped into local minimum and leaking from weak boundaries.

5.2 GVC Snake Based LV Segmentation with Shape Similarity Constraint

Although the segmentation results of LV are promising using the method described in Sect. 5.1, still drawbacks mainly include the following points:

- The GVF is computationally expensive.
- Since the epicardium is not an exact circle, we employ different weights for different partitions as λ_S, λ_A, λ_L, λ_P. However, the segmentation results are sensitive to these four parameters. In other words, the circle-based shape constraint is not suitable to segment the epicardium.
- The ellipse-based shape constraint needs to estimate the shape parameters explicitly during evolution of the curve $C(s)$.

To tackle these problems, we apply the GVC snake model to segment the endocardium and epicardium of the left ventricle. For endocardium segmentation, the proposed method pays particular attention to papillary muscle and artifacts by adopting a shape energy described in Sect. 5.1, with this energy, the snake contour can overcome the spurious edges stemming from artifacts and the final results could depend much less on the initial contour. In practice, the segmented endocardium would be beneficial significantly to segment the epicardium. We exploit the relationship between the endocardium and the epicardium in shape and position for the epicardium segmentation in the following aspects: (1) using the endocardium result as initialization to automatically segment the epicardium; (2) the endocardium is always encircled by and nearby the epicardium, as a result, one can build new external force according to this position relationship. (3) the endocardium usually resembles the epicardium in shape, therefore, the endocardium may serve as a priori shape to guide the segmentation of the epicardium. With these strategies, the epicardium is automatically extracted after the endocardium is segmented.

Since the strategy of segmenting the endocardium is similar to that in Sect. 5, we focus on a novel energy based on shape similarity to segment the epicardium. The total energy of the GVC snake with similarity constraint can be written as follows:

$$E(C) = \int_0^1 \underbrace{\frac{1}{2}\left(\alpha \mid C'(s) \mid^2 + \beta \mid C''(s) \mid^2\right)}_{Internal\ \ energy} + \underbrace{g\big(C(s)\big)}_{GVC} + \underbrace{E_{simi}\big(C(s)\big)}_{similarity\ constraint} \quad ds, \tag{33}$$

5.2.1 The Shape Similarity Constraint

Generally speaking, there would be spurious edges on the myocardium, and the contrast between myocardium and surrounding structures would be low. Even though the endocardium edge is removed, the new external force would not be good enough to prevent the snake contour from leaking out from weak boundaries. In order to get more accurate segmentation results of the epicardium, we employ the endocardium result as a *priori* shape and construct a new shape-similarity based constraint given by

$$E_{simi} = \frac{\rho}{2}\int_0^1 \big((R(s) - \overline{R}) - (r(s) - \overline{r})\big)^2 ds. \tag{34}$$

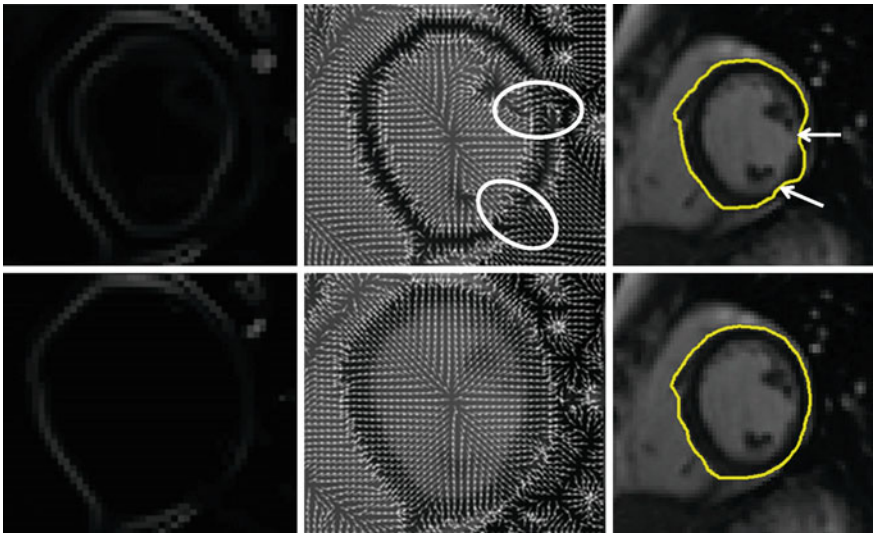


Fig. 11 Comparisons of external force for epicardium segmentation. *The upper row* the original edge map. *The lower row* the modified edge map

The variables in Eq. (34) have similar meanings as in Eq. (24), but R and $\overline{R}$ are for epicardium while r and $\overline{r}$ are for segmented endocardium. The snake contour for epicardium is supposed to be identically centered with th endocardium. It is clear in Sect. 5 that $r(s_i) - \overline{r}$ measures the deviation of the endocardium contour from a circle with radius $\overline{r}$ at snaxel s_i, thus $R(s_i) - \overline{R}$ measures the deviation of the snake contour for epicardium from a circle with radius $\overline{R}$ at snaxel s_i. Minimizing the energy E_{simi} will make the two deviations close, finally the snake contour for epicardium will resemble the endocardium in shape although their scales, i.e., $\overline{R}$ and $\overline{r}$ are different.

Similar to Eq. (24), by the calculus of variation, we obtain the Euler equation for Eq. (34) as follows:

$$\begin{cases} \rho\big(x_s - x^{endo}(s) - (\overline{R} - \overline{r})\cos(2\pi s)\big) = 0 \\ \rho\big(y_s - y^{endo}(s) - (\overline{R} - \overline{r})\sin(2\pi s)\big) = 0\,. \end{cases} \tag{35}$$

It is discretized as

$$\begin{cases} \rho\big(x_i - x_i^{endo} - (\overline{R} - \overline{r})\cos(2\pi i/n)\big) = 0 \\ \rho\big(y_i - y_i^{endo} - (\overline{R} - \overline{r})\sin(2\pi i/n)\big) = 0\,. \end{cases} \tag{36}$$

Similar to Eq. (28), this equation can be solved by taking x and y as the function of time t. Eq. (36) is incorporated into the similarity constraints term in Eq. (33) to extract the epicardium, and the force vector $-\nabla g\big(C(s)\big)$ in Eq. (17) will be replaced by $[u(x, y), v(x, y)]$ in Eq. (22) using the modified edge map depicted in Fig. 11.

Since the endocardium is a local minimum, this prevents the GVC active contour moving into the boundary of epicardium. In order to extract automatically the epicardium by taking the endocardium as initialization, the local minimum stemming from the endocardium edge should be filtered out. To this end, we directly set the original edge map around and within the endocardium to zero. This modified edge map is used to generate a new GVC force field, which leads to the fact that the endocardium is no longer a local minimum of the new GVC force. This new GVC force, therefore, can push final endocardium contour forward to the epicardium directly. In addition, since the endocardium and epicardium are adjacent, the capture range of the new GVC force needs not to be very large, so it takes even shorter time to calculate.

Figure 11 illustrates the effectiveness of the modified edge map for the epicardium segmentation. From left to right, the upper row in Fig. 11 shows the original edge map, the GVC field and the segmentation result, respectively. Moreover, the original force field flows into the myocardium at the weak epicardium boundaries (see the white ellipse on the upper row in Fig. 11). In contrast, when the edge map is modified using the proposed strategy, as shown on the lower row in Fig. 11, the associated force field can characterize the epicardium very well.

6 Segmentation Framework and Results

Our segmentation methods consist of the following steps as shown in Fig. 12:

1. Automatic localization of the LV. Hough transform is applied to intensity difference image to locate the LV centroid and the ROI.
2. Designing the external force for snake model. The external force field plays a leading role in driving the active contours to approach objects boundaries in the snake model, and thus significantly influences the segmentation performance. The GVF and (GVC) are utilized as the external forces of the snake model.
3. The endocardium segmentation. Considering that the LV is roughly a circle, the circle-shape or ellipse-shape based energy functional is integrated into the snake model to extract the endocardium.
4. The epicardium segmentation. After extracting the endocardium, the edge map is modified to yield a new external force field for active contours, which automatically pushes the snake contour directly to the epicardium by employing the endocardium result as initialization. We also adopt the segmentation result of the endocardium as a *priori* shape to get an accurate estimate of the epicardium.
5. Assessment of segmentation accuracy. The segmentation results are compared with the state-of-the-art methods using the mean absolute distance (MAD) and Dice metric (DM) [3].

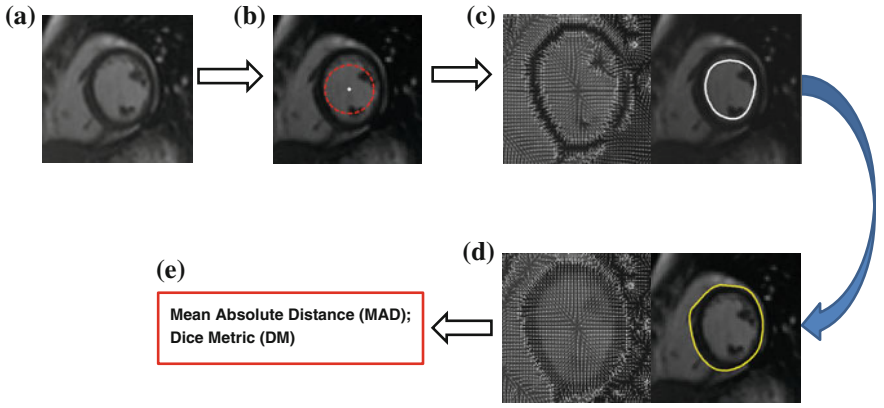


Fig. 12 Framework for segmenting the cardiac cine MRI. **a** The cardiac cine MRI input; **b** Automatic localization of the LV; **c** The external force of snake model for segmenting endocardium and the segmentation result; **d** The external force of snake model for segmenting epicardium and the segmentation result; **e** Evaluation of segmentation results

6.1 Automatic Localization of the LV

In a short-axis view of cardiac MR images, the myocardium is a dark area between two concentric circles enclosing a bright area corresponding to the blood in LV. The left side of the myocardium is a bright region corresponding to the blood in RV. The right side of the myocardium is a very dark area corresponding to the lungs. Under breath-hold condition, LV moves more obviously than its surrounding structures that are almost static during the cardiac cycle. This trait encourages intensity difference algorithm upon two consecutive frames in temporal image sequences to remove stationary background structures, and thereby locating the moving region of the left ventricle.

Suppose a cardiac MR images sequence $I_t(x, y)$, where (x, y) denotes the spatial coordinates of an image and $t \in T$ is the time instant. The nearly non-moving background pixels in two consecutive frames are excluded by the difference or subtraction operation defined as

$$D(x, y) = \left[I_{t+1}(x, y) - I_t(x, y)\right] > Th. \tag{37}$$

Here $D(x, y)$ is the intensity difference image, Th is a threshold value which we consider as non-moving background. Th is estimated by the OTSU method [73]. Figure 13 shows one of adjacent frames from a cardiac MR images sequence. In this sequence, myocardium moves along with heart beating; the chest cavity and lung change little due to the breath-hold condition; and there exists the nearly non-moving image background. The intensity difference image represents the most dynamic region. Observing the Fig. 13, the intensity values near the myocardial boundaries

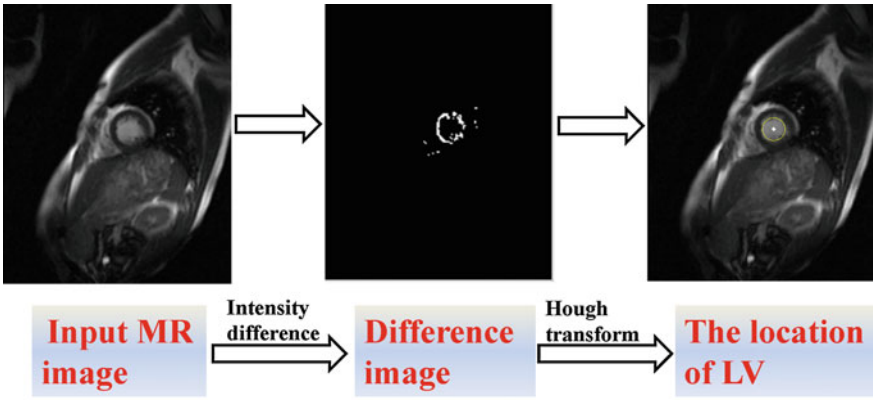


Fig. 13 Automatic localization of the LV

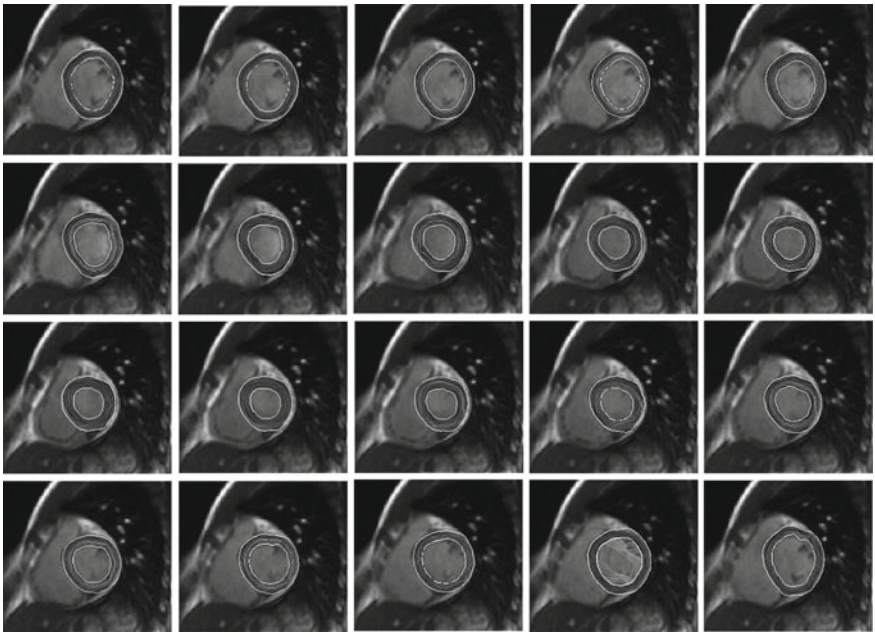


Fig. 14 LV segmentation using GVF snakes with circle constraint

are different from other region because of larger movement of the LV. The dense highlight circle-like region implies that the endocardium moves faster than the epicardium. Applying Hough transform to the difference image, we can obtain the LV centroid and the region of interest represented by a yellow solid circle.

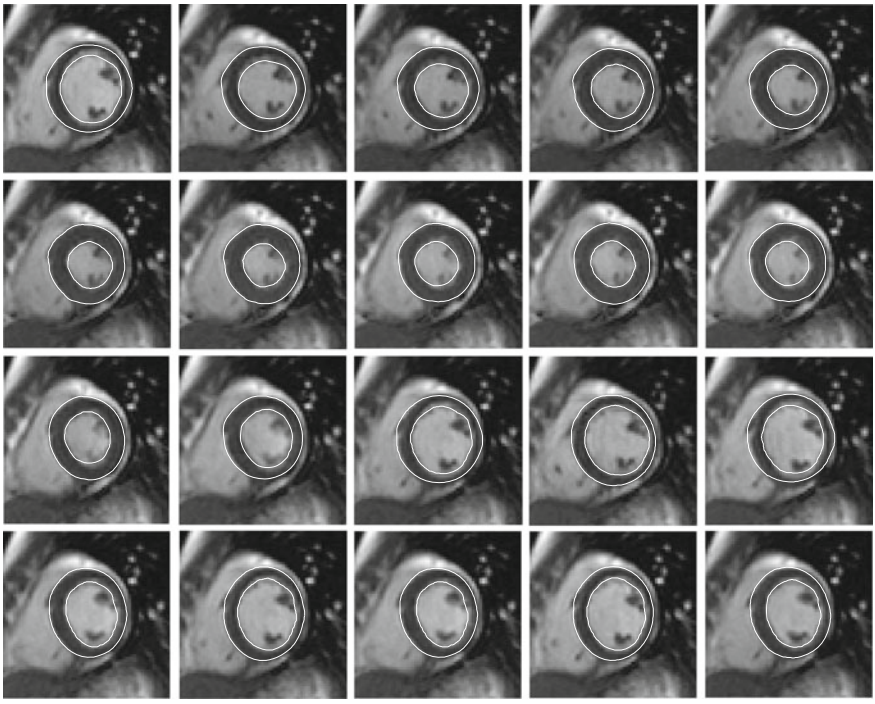


Fig. 15 LV segmentation using GVF snakes with ellipse constraint

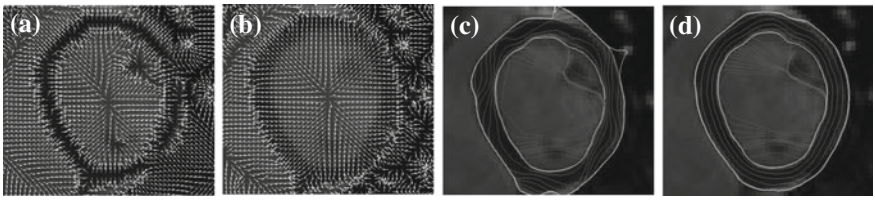


Fig. 16 Segmentation of the epicardium

6.2 Evaluation Criterions

For quantitative evaluation, we use two measurements to assess the segmentation performance. Suppose there are n points on the snake contour denoted by $S = \{s_1, s_2, \ldots, s_n\}$, k points on the ground truth expressed as $M = \{m_1, m_2, \ldots, m_k\}$, the mean absolute distance (MAD) [74] is defined as

$$MAD(S, M) = \frac{1}{2}\Big(\frac{1}{n}\sum\nolimits_{i=1}^{n} d(s_i, M) + \frac{1}{k}\sum\nolimits_{j=1}^{k} d(m_j, S)\Big), \tag{38}$$

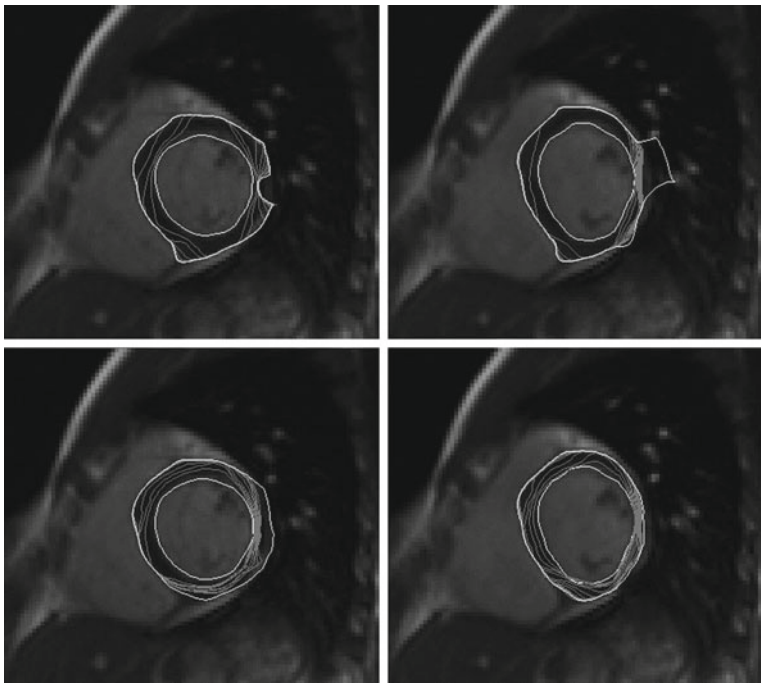


Fig. 17 Effectiveness of the shape-similarity based constraint for epicardium segmentation. *The upper row* without the shape-similarity based constraint. *The lower row* with the shape-similarity based constraint

where $d(s_i, M) = \min_j \|s_i - s_j\|$ is the distance from point s_i to the closet point on contour M.

The dice metric (DM) [3] is a measure of contour overlap utilizing the contour areas automatically segmented $\boldsymbol{A_a}$, manually segmented $\boldsymbol{A_m}$, and their intersection $\boldsymbol{A_{am}}$. It is expressed as

$$DM = 2\boldsymbol{A_{am}}\left(\boldsymbol{A_a} + \boldsymbol{A_m}\right)^{-1}. \tag{39}$$

DM is always between 0 and 1, with higher DM indicating better match between automatic and manual segmentations.

6.3 Experiment Results

6.3.1 GVF Snake Based Results

We applied the proposed strategies described in Sect. 5.1 to our own data set. Images of this dataset (147 images) were acquired using a $1.5T$ Siemens MRI scanner

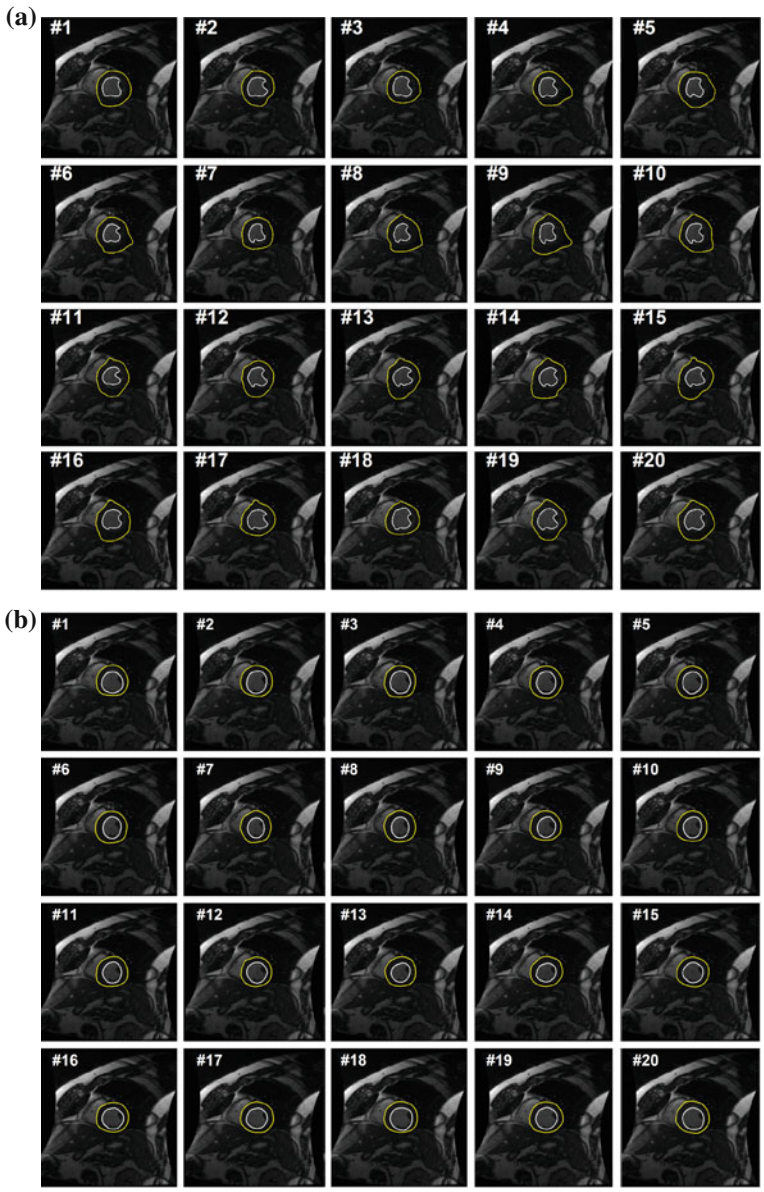


Fig. 18 Effectiveness of the shape based constraints for the LV segmentation. **a** The segmentation results of the LV without shape constraints. **b** The segmentation results of the LV with shape constraints

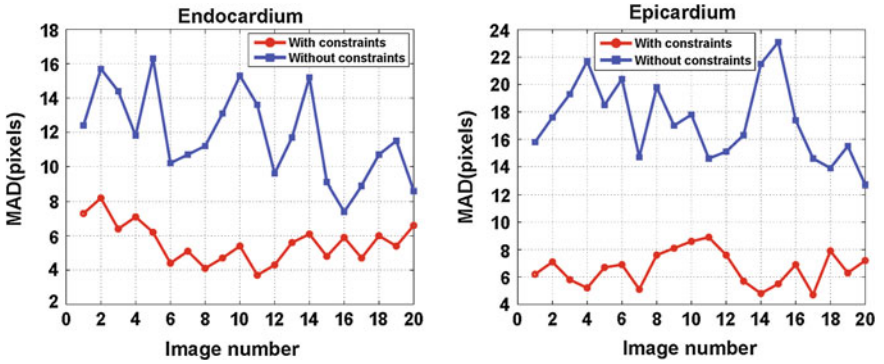


Fig. 19 The MAD errors corresponding to Fig. 18 of one subject on *MICCAI 2009* dataset with and without shape based constraints

from a healthy volunteer. Typical parameters were: TR: 29.16 ms; TE: 1.08 ms; flip angle: 50°, image dimension: 192 × 156; typical spatial resolution: 1.82 × 1.82 mm^2 in-plane; and slice thickness: 8 mm. There are 7 slices covering the entire LV from the apex to the base, and 21 cardiac phases in each slice. Figure 14 shows the segmentation results of the images from one slice using GVF snakes with the circle constraint in our dataset. The parameters of the snake model for endocardium are $\alpha = 1, \beta = 1$ and $\lambda = 0.5$. The parameters for epicardium are $\alpha = 1, \beta = 1\ \lambda_S = 0$, $\lambda_A = 1$, $\lambda_L = 0.5$, $\lambda_P = 1$. By taking the manual collections as ground truth, the results by the proposed methods are evaluated using MAD. Among the 147 MADs, for endocardium, the mean MAD is 0.78 pixels for circle-shape based constraint, while for epicardium, mean MAD is 1.12. Figure 15 illustrates the segmentation results of a set of cardiac cine MR images. It is noted that the segmentation strategy is similar to that described in Sect. 5.1 except that the circle-based shape constraint Eq. (24) is replaced with the ellipse-shape constraint Eq. (30).

6.3.2 GVC Snake Based Results

(a) The effectiveness of the similarity based constraints: In this section, we demonstrate the effectiveness of the shape similarity energy for epicardium extraction. Although the GVC derived from usual edge map cannot push the snake contour from the endocardium to the epicardium (Fig. 16a), the proposed strategy in Sect. 5.2.1 works well for this purpose(Fig. 16b), but this new force is not perfect yet, the snake contour with only local constraints would leak out from weak boundaries (Fig. 16c), after the proposed energy based on shape similarities is incorporated, the snake contour behaves satisfactorily, the result is in Fig. 16d. Figure 17 shows another example of this case. In the upper row, the snake contour leaks out since there is no shape constraint and the GVC field is not perfect. In contrast, when the shape similarity

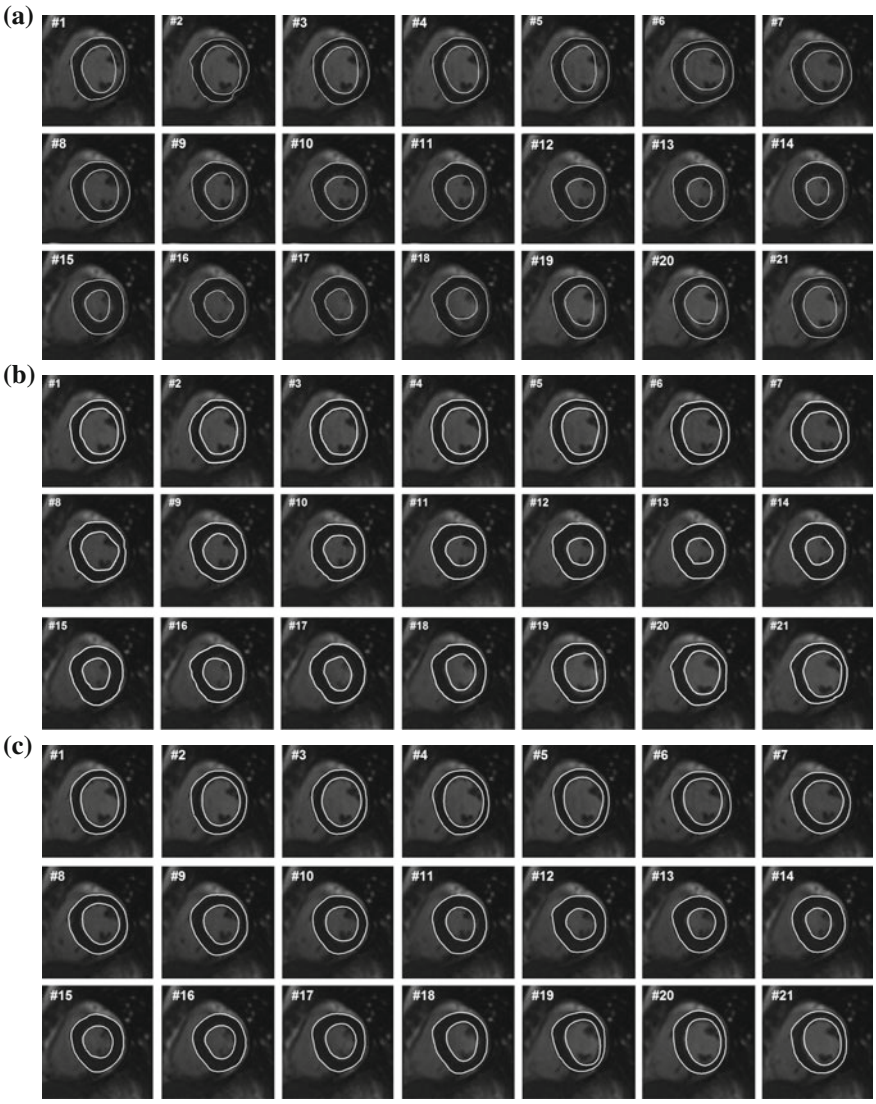


Fig. 20 Qualitative results of one subject on our own dataset. **a** The segmentation results with LSM method [12]. **b** The segmentation results with MFM method [3]. **c** The segmentation results with our method

energy is incorporated, the snake contour works well to delineate the epicardium (see the lower row of Fig. 17).

Figure 18 depicts the segmentation results of LV with and without shape prior information on *MICCAI 2009* dataset. Cardiac cine MRI data of *MICCAI 2009* is provided by the *MICCAI 2009* Cardiac MR LV Segmentation Challenge organizers.

Table 1 Quantitative performance evaluations on our own dataset (147 images) for the LSM [12], MFM [3] and our methods

	LSM	MFM	Ours
Cavity DM	0.80 ± 0.021	0.86 ± 0.011	0.84 ± 0.019
Myocardium DM	0.84 ± 0.032	0.83 ± 0.027	0.85 ± 0.053
Endocardium MAD	3.7667	1.1238	1.1190
Epicardium MAD	2.8048	1.2238	1.1333

The first two rows denote the statistics of the DM given by mean $\pm$ standard deviation. The second two rows are average MAD (in pixels)

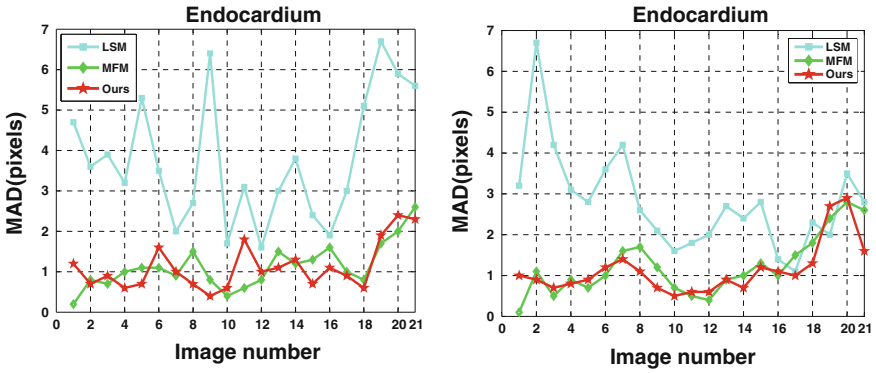


Fig. 21 Quantitative results corresponding to Fig. 20 of one subject on our own dataset

The scanning protocol and evaluation criterion of this datasets are fully described in [75]. Without shape constraint, the external force field pulls the snake to a false contour shown in Fig. 18a. These results are hardly to be accepted. Figure 18b shows that the shape energy functionals presented in Eqs. (24) and (34) are efficient to push the snake contours to the desire solution. Figure 19 illustrates the MAD errors corresponding to Fig. 18 of one subject on *MICCAI 2009* dataset with and without constraints. Overall, the shape constraints integrated into the GVC snake model effectively alleviate the side-effect of papillary muscle and noise, and prevent the snake contour from leaking out from weak boundaries.

(b) Comparison with the state-of-the-arts: To evaluate the performance of the proposed algorithm, we compare our method described in Sect. 5.2 with two published shape-based approaches including the level-set method (referred to as LSM) [12] and max-flow method (referred to as MFM) [3]. The parameters of these three models are unchanged for all the datasets. The parameters of LSM are $\alpha = 1000$, $\beta = 10$, $\lambda = 0.1$ and $c = 10$. The parameters of MFM are $\gamma = 1$, $\lambda = 0.0011$. The parameters of our GVC based model are $\alpha = 0.5$, $\beta = 0.5$, $\lambda = 0.4$, $\rho = 1.2$, $h = 8$ and $n = 2.6$.

From the qualitative comparisons depicted in Fig. 20, we can see that the segmentation results of LSM methods are noticeably worse than those of both MFM and our methods, at some frames (such as around the frames 9, 19, 20 and 21). It can

Fig. 22 Qualitative results of one subject on *MICCAI 2009* dataset. **a** The segmentation results of LSM method [12]. **b** The segmentation results of MFM method [3]. **c** The segmentation results of our approach

be explained that the LSM method is based on the assumption that the *overlap* is approximately constant, thus a high variation of the *overlap* in a given sequence will affect segmentation accuracy. In contrast, The MFM and our methods obtain good results.

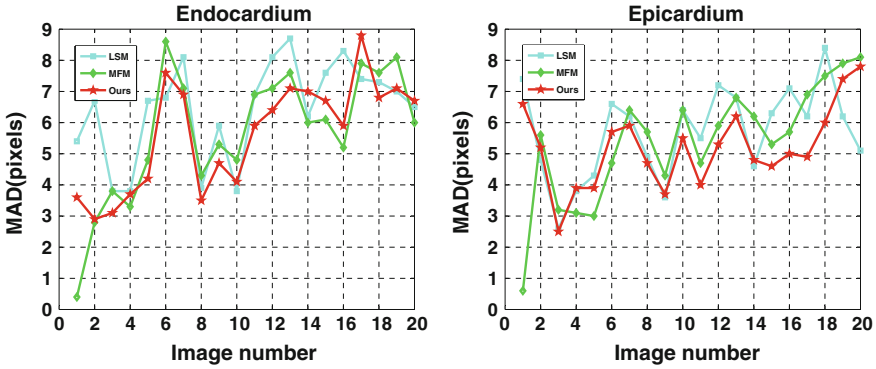


Fig. 23 Quantitative results corresponding to Fig. 22 of one subject on *MICCAI 2009* dataset

Table 1 and Fig. 21 show quantitative evaluations on our own dataset (147 images) for MFM, LSM and our methods. The first two rows in Table 1 report the DM statistics, where DM is given by mean ± standard deviation. For the cavity detection, MFM and our methods led to a region accuracy slightly better than LSM method. The second two rows in Table 1 show average MAD (in pixels) of these models. We see that our method outperforms other approaches.

We also run our algorithm on the *MICCAI 2009* dataset. Visually, these methods obtain the similar results, as shown in Fig. 22. However, our method is able to achieve much more accurate and consistent segmentation results, as shown in Fig. 23. The average MAD of LSM, MFM and our methods for the endocardium segmentation are 7.21, 4.79 and 5.06 pixels, respectively, and those of the epicardium are 5.70, 5.23 and 5.18 pixels, respectively. Overall, the proposed method could conquer image noise, artifacts, weak boundaries and papillary muscles on both endocardial and epicardial boundaries extraction. Nevertheless, our method remains great MAD errors during segmenting the endocardium, especially at the frames 7, 13 and 17. Around at these frames, the blood pool has almost the same intensity profile (e.g. the papillary muscles within the cavity and the myocardium). With the circle constraint and shape similarity energy, we can extract the endocardium and epicardium of the LV from MR images sucessfully. More representative images on *MICCAI 2009* dataset are shown in Fig. 24.

6.4 Discussion

Although the results obtained by our method are desirable, one significant assumption is that the shape of the LV is pre-defined circularly. This assumption limits a more extensive exploration of our method for the LV segmentation. Moreover, A major difficulty in segmentation of MR images is the intensity inhomogeneity due

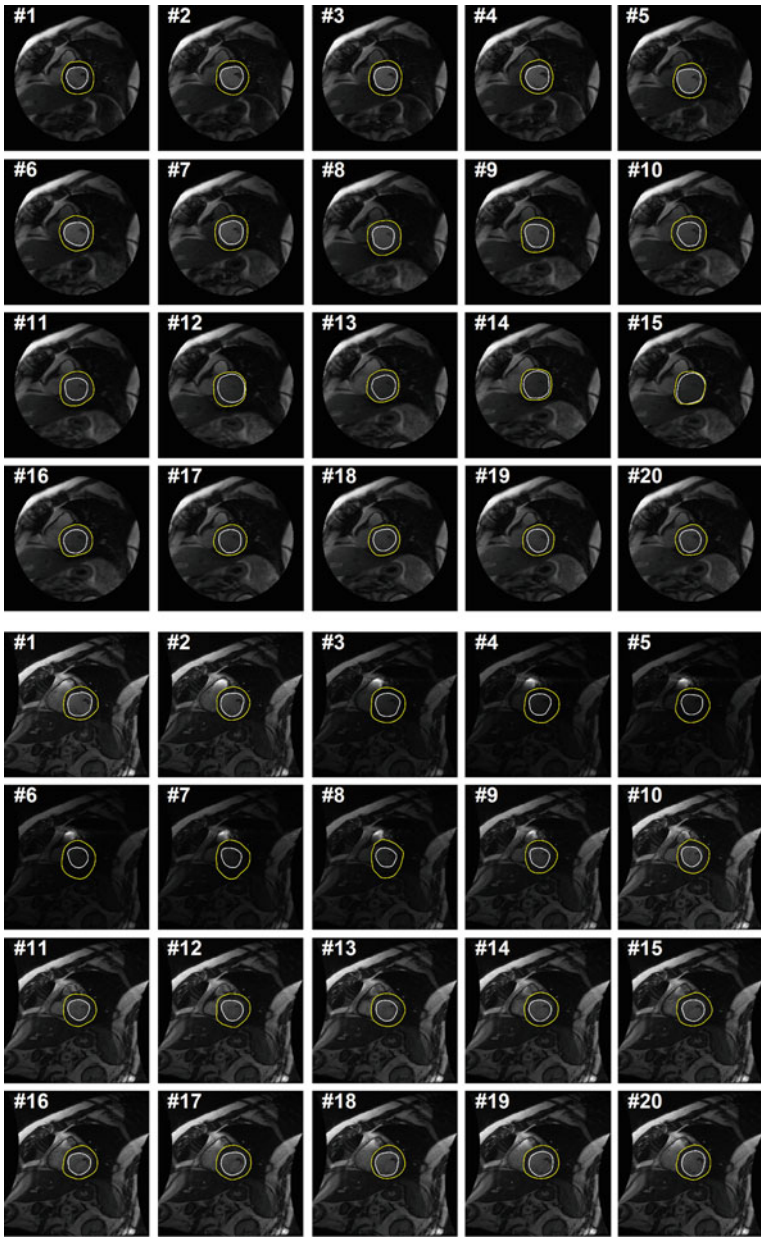


Fig. 24 More experiment results of our method on *MICCAI 2009* dataset

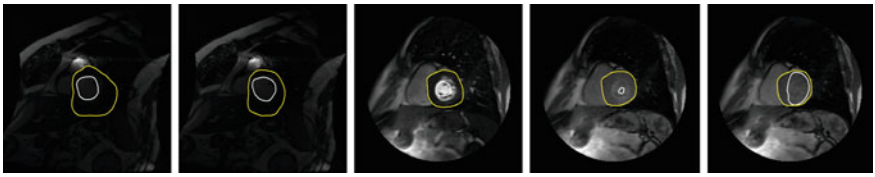


Fig. 25 Failed segmentation cases of our method

to the radio-frequency coils or acquisition sequences. There exists the inefficiency in handling images with *severe* intensity inhomogeneity. Figure 25 shows failed segmentation cases of our method, in which the myocardium and its neighbor organs such as the liver are connected, resulting in the same intensity profile. Segmenting such region of interest is extremely difficult because the contour is almost indistinguishable in this situation, even for human eyes. In addition, it should be emphasized that, different from the MFM method [3], the proposed approach requires a large number of iterative updates of the segmentation, therefore, is computationally onerous. Running on a Intel Core2 2.66 GHz processor with 2GB RAM, on average, our implementation needs 4.76 s to process a frame of 256×256 pixels. These issues motivate us to develop a more efficient LV segmentation algorithm in future.

7 Conclusion

We investigated the shape constraints for extracting the endocardium and epicardium from cine MRI based on snake models. A circle-shape based energy is first integrated into the GVF snake model for extracting the endocardium to conquer papillary muscle and artifacts. Considering the LV is not an exact circle, we employ different weights for different partitions of the LV (i.e., septum, anterior, lateral and posterior) to extract the epicardium when using the circle constraint.. In the framework of GVF snake model, we also generalize the circle-shape constraint to the ellipse-shape constraint which is a more versatile shape model. Although the shape-constraint-based GVF snake model yields promising results, the shortcomings of this strategy, such as computation cost of the GVF, parameters estimation of the ellipse and different weights for the circle constraint for epicardium segmentation, limit its application. Therefore, we apply the proposed GVC snake model to segment the endocardium and epicardium of the left ventricle based on the circle constraint and shape similarity constraint by assuming that the epicardium resembles the endocardium in shape. Comparative results on both our dataset and the *MICCAI 2009* dataset demonstrated a good performance of the proposed segmentation approach over the state-of-the-art methods.

Acknowledgments This work was supported in part by the Natural Science Foundation of China(NSFC) under Grants No.90920009 & No. 60602050, and the Key Program from the Tianjin Commission of Technology of China under Grant No. 11JCZDJC15600.

References

1. Roger VL, Benjamin EJ, Go AS, Mozaffarian D (2013) Heart disease and stroke statistics-2013 update a report from the american heart association. Circulation 127(1): e6–e245
2. Frangi A, Niessen W, Viergever M (2001) Three-dimensional modeling for functional analysis of cardiac images: a review. IEEE Trans Med Imaging 20(1):2–5
3. Ayed I, Chen H, Punithakumar K, Ross I, Li S (2012) Max-flow segmentation of the left ventricle by recovering subject-specific distributions via a bound of the bhattacharyya measure. Med Image Anal 16:87–100
4. Punithakumar K, Ben Ayed I, Islam A, Ross I, Li S (2010) Tracking endocardial motion via multiple model filtering. IEEE Trans Biomed Eng 57(8): 2001–2010
5. Ben Ayed I, Punithakumar K, Li S, Islam A, Chong J (2009) Left ventricle segmentation via graph cut distribution matching. In: Medical image computing and computer-assisted intervention-MICCAI 2009, pp 901–909
6. Cousty J, Najman L, Couprie M, Clément-Guinaudeau S, Goissen T, Garot J (2010) Segmentation of 4 D cardiac MRI: automated method based on spatio-temporal watershed cuts. Image Vis Comput 28(8):1229–1243
7. Cocosco C, Niessen W, Netsch T, Vonken E, Lund G, Stork A, Viergever M (2008) Automatic image-driven segmentation of the ventricles in cardiac cine MRI. J Magn Reson Imaging 28(2):366–374
8. Pednekar A, Kurkure U, Muthupillai R, Flamm S, Kakadiaris I (2006) Automated left ventricular segmentation in cardiac MRI. IEEE Trans Biomed Eng 53(7):1425–1428
9. Kurkure U, Pednekar A, Muthupillai R, Flamm S, Kakadiaris I (2009) Localization and segmentation of left ventricle in cardiac cine- MR images. IEEE Trans Biomed Eng 56(5):1360–1370
10. Kass M, Witkin A, Terzopoulos D (1988) Snakes: active contour models. Int J Comput Vision 1(4):321–331
11. Paragios N (2003) A level set approach for shape-driven segmentation and tracking of the left ventricle. IEEE Trans Med Imaging 22(6):773–776
12. Ben Ayed I, Li S, Ross I (2009) Embedding overlap priors in variational left ventricle tracking. IEEE Trans Med Imaging 28(12): 1902–1913
13. Wang Y, Jia Y (2006) Segmentation of the left ventricle from cardiac MR images based on degenerated minimal surface diffusion and shape priors. In: 18th International conference on pattern recognition, 2006. ICPR 2006, vol. 4. IEEE, pp 671–674
14. Liang J, Ding G, Wu Y (2008) Segmentation of the left ventricle from cardiac MR images based on radial GVF snake. In: International Conference on BioMedical Engineering and Informatics, 2008. BMEI 2008, vol 2. IEEE, pp 238–242
15. Cootes T, Taylor C, Cooper D, Graham J et al (1995) Active shape models-their training and application. Comput Vis Image Underst 61(1):38–59
16. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23(6):681–685
17. Zhang H, Wahle A, Johnson R, Scholz T, Sonka M (2010) 4-D cardiac MR image analysis: Left and right ventricular morphology and function. IEEE Trans Med Imaging 29(2):350–364
18. Carneiro G, Nascimento J, Freitas A (2012) The segmentation of the left ventricle of the heart from ultrasound data using deep learning architectures and derivative-based search methods. IEEE Trans Image Process 21(3):968–982
19. Petitjean C, Dacher J (2011) A review of segmentation methods in short axis cardiac MR images. Med Image Anal 15(2):169–184

20. Wang Y, Jia Y (2006) Segmentation of the left venctricle from MR images via snake models incorporating shape similarities. In: 2006 IEEE International conference on image processing. IEEE, pp 213–216
21. Wang Y, Jia Y (2008) External force for active contours: gradient vector convolution. In: PRICAI, (2008) trends in artificial intelligence, pp 466–472
22. Wang Y, Liang J, Jia Y (2007) On the critical point of gradient vector flow snake. In: Proceedings of the 8th Asian conference on computer vision-volume Part II. Springer-Verlag,pp 754–763
23. Wu Y, Wang Y, Jia Y (2013) Segmentation of the left ventricle in cardiac cine mri using a shape-constrained snake model. Comput Vis Image Underst 117(9):990–1003. http://dx.doi.org/10.1016/j.cviu.2012.12.008
24. Ranganath S (1995) Contour extraction from cardiac MRI studies using snakes. IEEE Trans Med Imaging 14(2):328–338
25. Makowski P, Sørensen T, Therkildsen S, Materka A, Stødkilde-Jørgensen H, Pedersen E (2002) Two-phase active contour method for semiautomatic segmentation of the heart and blood vessels from MRI images for 3 D visualization. Comput Med Imaging Graph 26(1):9–17
26. Cohen LD (1991) On active contour models and balloons. CVGIP: Image underst 53(2):211–218
27. Hautvast G, Lobregt S, Breeuwer M, Gerritsen F (2006) Automatic contour propagation in cine cardiac magnetic resonance images. IEEE Trans Med Imaging 25(11):1472–1482
28. Xu C, Prince J (1998) Snakes, shapes, and gradient vector flow. IEEE Trans Image Process 7(3):359–369
29. Santarelli M, Positano V, Michelassi C, Lombardi M, Landini L (2003) Automated cardiac MR image segmentation: theory and measurement evaluation. Med Eng Phys 25(2):149–159
30. Lee H, Codella N, Cham M, Weinsaft J, Wang Y (2010) Automatic left ventricle segmentation using iterative thresholding and an active contour model with adaptation on short-axis cardiac MRI. IEEE Trans. Biomed. Eng 57(4):905–913
31. Nguyen D, Masterson K, Vallée J (2007) Comparative evaluation of active contour model extensions for automated cardiac MR image segmentation by regional error assessment. Magn Reson Mater Phys, Biol Med 20(2):69–82
32. Paragios N (2002) A variational approach for the segmentation of the left ventricle in cardiac image analysis. Int J Comput Vision 50(3):345–362
33. Folkesson J, Samset E, Kwong R, Westin C (2008) Unifying statistical classification and geodesic active regions for segmentation of cardiac MRI. IEEE Trans Inf Technol Biomed 12(3):328–334
34. Alessandrini M, Dietenbeck T, Basset O, Friboulet D, Bernard O (2011) Using a geometric formulation of annular-like shape priors for constraining variational level-sets. Pattern Recogn Lett 32(9):1240–1249
35. Pluempitiwiriyawej C, Moura J, Wu Y, Ho C (2005) Stacs: new active contour scheme for cardiac mr image segmentation. IEEE Trans Med Imaging 24(5):593–603
36. Chen T, Babb J, Kellman P, Axel L, Kim D (2008) Semiautomated segmentation of myocardial contours for fast strain analysis in cine displacement-encoded MRI. IEEE Trans Med Imaging 27(8):1084–1094
37. Lynch M, Ghita O, Whelan P (2006) Left-ventricle myocardium segmentation using a coupled level-set with a priori knowledge. Comput Med Imaging Graph 30(4):255–262
38. Punithakumar K, Ben Ayed I, Ross I, Islam A, Chong J, Li S (2010) Detection of left ventricular motion abnormality via information measures and bayesian filtering. IEEE Trans Inf Technol Biomed 14(4):1106–1113
39. Lynch M, Ghita O, Whelan P (2008) Segmentation of the left ventricle of the heart in 3- D+ t MRI data using an optimized nonrigid temporal model. IEEE Trans Med Imaging 27(2):195–203
40. Osher S, Sethian J (1988) Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. J Comput Phys 79(1):12–49
41. Zhu Y, Papademetris X, Sinusas A, Duncan J (2010) Segmentation of the left ventricle from cardiac MR images using a subject-specific dynamical model. IEEE Trans Med Imaging 29(3):669–687

42. Jolly M (2006) Automatic segmentation of the left ventricle in cardiac mr and ct images. Int J Comput Vision 70(2):151–163
43. Duta N, Jain AK, Jolly MP (1999) Learning 2d shape models. In: Proceedings of the international conference on computer vision and pattern recognition, vol 2, 7–14
44. Lorenzo-Valdés M, Sanchez-Ortiz GI, Elkington AG, Mohiaddin RH, Rueckert D et al (2004) Segmentation of 4d cardiac mr images using a probabilistic atlas and the em algorithm. Med Image Anal 8(3):255–265
45. Van Assen HC, Danilouchkine MG, Frangi AF, Ordas S, Westenberg JJ, Reiber JH, Lelieveldt BP (2006) Spasm: a 3d-asm for segmentation of sparse and arbitrarily oriented cardiac mri data. Med Image Anal 10(2):286–303
46. van Assen HC, Danilouchkine MG, Dirksen MS, Reiber JH, Lelieveldt BP (2008) A 3-d active shape model driven by fuzzy inference: application to cardiac ct and mr. IEEE Trans Inf Technol Biomed 12(5):595–605
47. Matthews I, Baker S (2004) Active appearance models revisited. Int J Comput Vision 60(2):135–164
48. Gopal S, Otaki Y, Arsanjani R, Berman D, Terzopoulos D, Slomka P (2013) Combining active appearance and deformable superquadric models for lv segmentation in cardiac mri. In: SPIE medical imaging, international society for optics and photonics, pp 86690G–86690G
49. Ghose S, Oliver A, Marti R, Llado X, Freixenet J, Mitra J, Vilanova JC, Meriaudeau F (2012) A hybrid framework of multiple active appearance models and global registration for 3d prostate segmentation in mri. In: SPIE Medical imaging, international society for optics and photonics, pp 83140S–83140S
50. Mitchell SC, Lelieveldt BPF, van der Geest RJ, Bosch HG, Reiver J, Sonka M (2001) Multistage hybrid active appearance model matching: segmentation of left and right ventricles in cardiac mr images. IEEE Trans Med Imaging 20(5):415–423
51. Pfeifer B, Hanser F, Seger M, Hintermueller C, Modre-Osprian R, Fischer G, Muehlthaler H, Trieb T, Tilg B (2005) Cardiac modeling using active appearance models and morphological operators. In: Medical imaging, international society for optics and photonics, pp 279–289
52. Zambal S, Hladuvka J, Buhler K (2006) Improving segmentation of the left ventricle using a two-component statistical model. In: Medical image computing and computer-assisted intervention-MICCAI 2006. Springer, pp 151–158
53. Ray N, Acton S (2004) Motion gradient vector flow: an external force for tracking rolling leukocytes with shape and size constrained active contours. IEEE Trans Med Imaging 23(12):1466–1478
54. Caselles V, Kimmel R, Sapiro G (1997) Geodesic active contours. Int J Comput Vision 22(1):61–79
55. Chan T, Vese L (2001) Active contours without edges. IEEE Trans Image Process 10(2):266–277
56. Zhang K, Zhang L, Song H, Zhou W (2010) Active contours with selective local or global segmentation: a new formulation and level set method. Image Vis Comput 28(4):668–676
57. Xie X, Mirmehdi M (2008) MAC: magnetostatic active contour model. IEEE Trans Pattern Anal Mach Intell 30(4):632–646
58. Tang J (2009) A multi-direction GVF snake for the segmentation of skin cancer images. Pattern Recogn 42(6):1172–1179
59. Mishra A, Fieguth P, Clausi D (2011) Decoupled active contour (dac) for boundary detection. IEEE Trans Pattern Anal Mach Intell 33(2):310–324
60. Cheng J, Foo S (2006) Dynamic directional gradient vector flow for snakes. IEEE Trans. Image Process 15(6):1563–1571
61. Wu Y, Jia Y, Wang Y (2010) Adaptive diffusion flow for parametric active contours. In: 20th International conference on pattern recognition, ICPR 2010. IEEE, 2010, pp 2788–2791
62. Xu C, Prince J (1998) Generalized gradient vector flow external forces for active contours. Signal Process 71(2):131–139
63. Lu S, Wang Y (2010) Gradient vector flow over manifold for active contours. In: Proceedings of the 9th Asian conference on computer vision, ACCV 2009. Springer-Verlag, pp 147–156

64. Park H, Chung M (2002) External force of snake: virtual electric field. Electron Lett 38(24):1500–1502
65. Yuan D, Lu S (2002) Simulated static electric field (ssef) snake for deformable models. In: Proceedings of 16th International Conference on Pattern Recognition, vol 1. IEEE,pp 83–86
66. Jalba AC, Wilkinson MH, Roerdink JB (2004) Cpm: a deformable model for shape recovery and segmentation based on charged particles. IEEE Trans Pattern Anal Mach Intell 26(10):1320–1335
67. Li B, Acton S (2007) Active contour external force using vector field convolution for image segmentation. IEEE Trans Image Process 16(8):2096–2106
68. Vidholm E, Sundqvist P, Nystrom I (2006) Accelerating the computation of 3d gradient vector flow fields. In: 18th International conference on pattern recognition, ICPR 2006, vol 3. IEEE, pp 677–680
69. Han X, Xu C, Prince J (2007) Fast numerical scheme for gradient vector flow computation using a multigrid method. IET Image Proc 1(1):48–55
70. Boukerroui D (2012) Efficient numerical schemes for gradient vector flow. Pattern Recogn 45(1):626–636
71. Ren D, Zuo W, Zhao X, Lin Z, Zhang D (2013) Fast gradient vector flow computation based on augmented lagrangian method. Pattern Recogn Lett 2(34):219-225
72. Ray N, Acton S, Ley K (2002) Tracking leukocytes in vivo with shape and size constrained active contours. IEEE Trans Med Imaging 21(10):1222–1235
73. Otsu N (1975) A threshold selection method from gray-level histograms. Automatica 11:285–296
74. Mikic I, Krucinski S, Thomas J (1998) Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates. IEEE Trans Med Imaging 17(2):274–284
75. Radau P, Lu Y, Connelly K, Paul G, Dick A, Wright G (2009) Evaluation framework for algorithms segmenting short axis cardiac mri. the midas journal-cardiac mr left ventricle segmentation challenge, 2009. http://hdl.handle.net/10380/3070

An Optical Flow Approach to Assessment of Ventricular Shape Change Based on Echocardiography

Vahid Tavakoli, Nirmanmoh Bhatia, Rita Longaker, Motaz Alshaher, Marcus Stoddard and Amir A. Amini

Abstract The quantitative analysis of cardiac motion from echocardiographic images helps clinicians in the diagnosis and therapy of patients suffering from heart disease. Quantitative analysis is usually based on TDI (Tissue Doppler Imaging) or speckle tracking. These methods are based on two techniques which to a large degree are independent—the Doppler phenomenon and image sequence processing, respectively. Herein, to increase the accuracy of the speckle tracking technique and to cope with the angle dependency of TDI, a combined approach dubbed TDIOF (Tissue Doppler Imaging Optical Flow) is proposed. TDIOF is formulated based on the combination of B-mode and Doppler energy terms minimized using algebraic equations and is validated on simulated images, a physical heart phantom, and *in-vivo* data. It was observed that the additional Doppler term is able to increase the accuracy of speckle tracking, compared to two popular motion estimation and speckle tracking techniques (Horn-Schunck and block matching methods). This observation was more pronounced when noise was present. . The magnitude and angular error for TDIOF applied to simulated images when comparing estimated motion with ground-truth motion were 15 % and 9.2 degrees/frame, respectively. The magnitude and angular error for images acquired from physical phantoms were 22 % and 15.2 degrees/frame, respectively. As an additional validation, echocardiography-derived strains were compared to tagged MRI-derived myocardial strains in the same subjects. The correlation coefficient (r) between the TDIOF-derived radial strains and tagged MRI-derived radial strains value were 0.83 ($P < 0.001$). The correlation coefficient (r) for the TDIOF-derived circumferential strains compared to the tagged MRI-derived circumferential strains were 0.86 ($P < 0.001$). The comparison

V. Tavakoli · N. Bhatia · M. Alshaher · M. Stoddard · A. A. Amini (✉)
University of Louisville, Louisville, KY, USA
e-mail: amir.amini@louisville.edu

V. Tavakoli · N. Bhatia · R. Longaker · M. Alshaher · M. Stoddard · A. A. Amini
Veterans Affairs Medical Center, Louisville, KY, USA

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1_13,

of TDIOF-derived and block matching speckle tracking and Horn-Schunck optical flow strain values using student t-test demonstrated superiority of TDIOF (95 % confidence interval, $P < 0.001$).

1 Introduction

Cardiovascular Disease (CVD) is the leading cause of death in the modern world. The mortality rate associated with CVD was estimated to be 17 million in 2005 and continues to be ranked as the top killer worldwide. CVD is the result of undersupply of the cardiac tissue and can lead to malfunction of the involved myocardial territories and manifest as hypokinesia or akinesia. Several imaging methods such as X-ray CT, MRI, and Ultrasound have been used for visualization of the heart function. MRI and X-ray CT provide excellent spatial resolution but the cost and lack of wide-spread availability cause challenges in the clinical settings. Echocardiography is a popular technique for cardiac imaging due to its availability, ease of use, and low cost. Echocardiography shows the motion and anatomy of the heart in real time, enabling physicians to detect different pathologies. However, analysis of motion of the myocardium in echocardiographic images is based on visual grading by an observer and suffers from inter and intra-observer variability. To overcome the inter- and intra-observer variability, computerized image analyses can help by quantitatively interpreting the data. To that end, cardiac image processing techniques, mainly categorized as segmentation and registration, have been widely used for assessing the regional function of the heart [1–3]. To perform such analysis, two independent techniques, have been utilized; these are TDI (Tissue Doppler Imaging) and speckle tracking. TDI computes the tissue motion based on the Doppler phenomenon and is dependent on the angle of insonification. Speckle tracking, on the other hand,is an image processing method based on the analysis of the ultrasound B-mode or RF images. B-mode based algorithms are robust to the variation of the transducer angle but rely entirely on the properties of echocardiographic images which may be noisy or inaccurate. The physical principle underlying B-mode and TDI are to a large degree independent and therefore for myocardial motion estimation carry complementary information [4, 5].

Many methods such as optical flow [6], feature tracking [7], level sets [8], block matching [9], and elastic registration [10] have been utilized for quantitative assessment of myocardial motion in B-mode images. Table 1 shows a description of some of the current methods used in motion estimation in echocardiographic images [6–21]. Suhling et al. [6] integrated rigid registration in an optical flow framework in order to detect myocardial motion from 2D echocardiographic images. B-spline moments invariants were applied to echo images to achieve invariance to the translation and rotation. The motion estimation algorithm was then applied to the B-spline moments of the image instead of the image intensity in a coarse to fine strategy and was validated using open chested dogs after ligation of a coronary artery. Additional

Table 1 Description of some of the current methods used in motion detection in echocardiography images

Article	Output	Technique	Validation (# of subjects)
Suhling et al. [6]	Motion	B-spline moments, optical flow	2D dog (6), simulated images, phantom
Yu et al. [7]	Motion	Maximum likelihood, spline based control points	2D Dog (4), Sonomicrometry
Paragios [8]	Endocardium, motion	Level set + learned shape-motion prior	2D Human
Hayat et al. [9]	Motion	Block matching	3D echo, MRI
Elen et al. [10]	Motion	Elastic registration	3D human (normal: 3, patient: 1), simulated images
Esther Leung et al. [11]	Motion	Optical flow and shape model	3d echo
Myronenco et al. [12]	Motion	Motion coherence by temporal regularization	3D human, EB
Duchateau et al. [13]	Motion	Diffeomorphic registration	2D human (normal: 21, patient: 14),
Bachner et al. [14]	Motion	fiber direction	2D human, simulation, phantom
Dydenco et al. [15]	Epicardium, motion	Regional statistics curve evolution	2D Human, TDI
Yan et al. [16]	Epicardium, motion	Finite element model	3D human, implanted marker
De Craene et al. [17]	Epicardium	Diffeomorphic B-spline free form deformation	3D human (normal: 9, patient: 13)
Ashraf et al. [18]	Motion	3D Pig	Sonomicrometry
Papademetris [19, 20]	Motion	Finite element model	3D echo
Kleijn et al. [21]	Motion	Block Matching	3D echo

validations were performed on simulation and phantom images. Ellen et al. [10] used elastic registration on 3D B-mode echocardiography images to extract myocardial motion and strain values. The method was validated using simulated and real ultrasoundimages. Esther-Leung et al. [11] proposed two different methods (1. model-driven, 2. edge-driven) for tracking the left-ventricular wall in echocardiographic images. Their approach was motivated by the fact that in echocardiography images,

visibility of the myocardium depends on the imaging view; so the myocardium may be, partly, invisible to the beam. Their technique relied on a local data-driven tracker using optical flow applied to the visible parts of the myocardium and a global statistical model applied to the invisible parts. It was concluded that the shape model could render good results for both the visible and the invisible tissues in ultrasound images. Myronenco et al. [12] proposed the so-called Coherent Point Drift (CPD) technique for myocardial motion estimation, constraining the motion of the point set in the temporal direction for both rigid and nonrigid point set registration. A set of point distribution was computed based on endocardium and epicardium locations. The point set was modeled with a Gaussian mixture model (GMM). The GMM centroids were updated coherently in a global pattern using maximum likelihood to preserve the topological structure of the point sets. A motion coherence constraint was added based on regularization of the displacement fields. The purpose of regularization was to increase the motion smoothness.

Most of the motion estimation techniques developed thus far, are either based on TDI or B-mode. Recently, Garcia et al. [22] considered the combination of cardiac B-mode images and intra-cardiac blood flow data for computing the blood flow motion in the heart using continuity equation and mass conservation in polar coordinates. Their paper focused on the blood flow computation and did not consider the cardiac tissue displacements. Dalen et al. [23] and Amundsen et al. [24] previously combined TDI with speckle tracking by integrating TDI in the beam direction and speckle tracking in the direction lateral to the beam. However, this method discarded the speckle tracking data in the beam direction. The authors reported that they were unable to improve the motion estimation performance compared to speckle tracking techniques. In this paper, we propose integration of tissue Doppler and speckle tracking within a novel optical flow framework, we call TDIOF (Tissue Doppler Optical Flow). Our experimental results indicate that TDIOF outperforms both TDI and speckle tracking approaches.

The organization of the rest of the paper is as follows: in Sect. 2, we review the mathematical and algorithmic basis for the proposed method. Section 3 is a description of datasets used for validation of the proposed method. These include computational simulations, US data collected in a cardiac phantom, and in vivo data collected in patients recruited from the echocardiography laboratory at the Robley Rex VA Medical Center to our IRB-approved study. In Sect. 4, strain computations are discussed and, in Sect. 5, results from validation of the proposed method are described. Finally, in Sect. 6 we discuss observations related to TDIOF and our findings.

2 Methods and Materials

TDI and B-mode speckle tracking are different in both their physical underpinning and data type. In speckle tracking, tissue motion is determined from motion of speckles in Ultrasound images–typically, using a block matching approach applied to 2-D B-mode images [give references]. Although speckle tracking provides both com-

ponents of the motion in the spatial domain, it is based on noisy B-mode images. TDI on the other hand only computes the velocity of tissue in one direction as moving towards (displayed as red) or away (displayed as blue) from the transducer. This means that the computed motion is the projection of the real motion in the direction of the transducer and therefore TDI is angle dependent. In this section, we first describe a novel energy minimization framework for estimation of myocardial motion from B-mode images which incorporates a velocity constraint from TDI.

The proposed method is based on optimization of three energy functions: (1) intensity constancy assumption, (2) velocity smoothness, and (3) similarity with Doppler data. The framework is minimized using an incremental algebra in method [25, 26] as described in Appendix A. In order to show the performance of TDIOF, it is compared to two popular motion detection techniques Horn-Schunck optical flow [27] and block matching [28] (Appendix B). Block matching is utilized in several commercial software platforms [29].

3 Validations

3.1 Simulated Computerized Phantom

Echocardiographic images are the result of the mechanical interaction between the ultrasound field and the contractile heart tissue. Previously, we reported on development and use of an ultrasound cardiac motion simulator [29]. In the current study, we utilized the COLE convolution based simulation technique reported in [30]. The significance of an Ultrasound cardiac motion simulator is the availability of both echocardiographic images as well as the actual ground-truth vector field of deformations.

A moving 3D heart was modeled based on a pair of prolate-spheroidal representations and used for the ultrasound simulation. The 3D forward model of cardiac motion was simulated using 13 time-dependent kinematic parameters of Arts et al. [30] (see Table 2). The evolution of the 13 kinematic parameters was previously derived by Arts following a temporal fit to actual location of tantalum markers in a canine heart [31]. In Arts' model, seven time-dependent parameters are applied to define the ventricular shape change, torsion, and shear while six parameters are used to model the rigid-body motions. To simulate the Ultrasound imaging process, scatterers were randomly distributed in the simulated LV wall and the motion prescribed by Arts' model was used to move the ultrasound scatterers. To determine Ultrasonic B-mode intensities, the COLE method was used [30]. COLE is an efficient convolution-based method in the spatial domain, producing US simulations by convolving the segmental PSF (point spread function) with the projected amplitudes of the scatterers [29] with the segmental PSF derived using Field II [30, 32]. In order to model the Doppler Effect, the frequency of the RF signal was shifted in the frequency domain based on the attributed ground truth motion vector and mixed with

Table 2 The 13 k-parameters of the Art's kinematic model for left-ventricular deformation used in our cardiac US motion simulator

Non-rigid body motion	
k_1	Radially dependent compression
k_2	Left ventricular torsion
k_3	Ellipticalization in long-axis (LA) planes
k_4	Ellipticalization in short-axis (SA) planes
k_5	Shear in x direction
k_6	Shear in y direction
k_7	Shear in z direction
Rigid body motion	
k_8	Rotation about x-axis
k_9	Rotation about y-axis
k_{10}	Rotation about z-axis
k_{11}	Translation along x-axis
k_{12}	Translation along y-axis
k_{13}	Translation along z-axis

additive Gaussian noise. If the velocity of the particle is v, ultrasound velocity is c, and transducer frequency is f, then the frequency shift is:

$$\Delta f = \frac{2vf}{c} \tag{1}$$

The resolution of the first simulated sequence was 0.1 mm/pixel for both B-mode and TDI images and included 14 mid-ventricular temporal frames in the axial orientation. In order to analyze the robustness of the method to noise, another set of simulated images were produced by adding Gaussian noise of 1.12 db to the noiseless data

3.2 Physical Cardiac Phantom

As described in [30], a physical cardiac phantom was built in-house, suitable for validation of echocardiographic motion estimation algorithms. Here, a brief description of this phantom is provided. To manufacture the phantom, a cardiac computerized model was used to build an acrylic based cardiac mold. A 10 % solution of Poly Vinyl alcohol (PVA) and 1 % enamel paint were used as the basic material. PVA has the ability to mimic cardiac elasticity, ultrasound and magnetic properties. The solution was heated up to 90 °C. Consequently, it was poured into the cardiac mold and gradually exposed to the temperature of −20 °C until it froze. The mold and the solution were kept in that temperature for 24 h. Finally, the mold and the frozen gel were gradually exposed to the room temperature. At this point, the normal heart phantom has passed one freeze-thaw cycle.

An additional model consisting of the left and right ventricles but with a segmental thin wall in the LV was used to build an additional mold for a pathologically scarred heart. The thinner wall was designed to mimic an aneurysmal, dyskinetic wall. Three PVA-based inclusions were separately made as a circle; slab and cube using nine, six and three freeze-thaw cycles respectively. Each freeze-thaw cycle decreases the elasticity of the heart mimicking scarred myocardium. The attenuation of the PVA and speed of sound increase after each freeze-thaw cycle.The cylindrical, slab like and cube like objects were placed in the mold in different American Heart Association cardiac segments [33]. Subsequently, the PVA solution was added to fill the rest of the space in the mold. After one freeze-thaw cycle, the abnormal heart consisted of a background of normal texture with one freeze-thaw cycle plus three infarct-mimicking inclusions having 10, 7 and 4 freeze-thaw cycles. The speed of sound in PVA is 1527, 1540, 1545, and 1550 m/s and ultrasound attenuation is 0.4, 0.52, 0.57, and 0.59 db/cm for 1, 4, 7 and 10 freeze-thaw cycles. The parameters of the synthetic phantom was adjusted based on the previous phantom studies [34].

A mediastinal phantom that provides the ability to acquire trans-esophageal images was manufactured using another mold. A solution of 50 % water and 50 % glycerol was used to mimic the blood. Finally, a syringe was used to manually force the fluid inside and outside the phantom for contraction and expansion. The enamel paint particles are robust scatterers and can generate reliable markers on the B-mode image. Since each marker is not restricted to just one pixel, the center of the mass of each manually segmented marker is considered as landmark. The displacements of the landmarks are compared to the computed motion field for the validation purposes. Figure 1 shows the cardiac phantom and the acquired phantom images using ultrasound and MRI.

3.3 Patient Studies

Two separate sets of data were utilized for in vivo validations (sets A and B). Set A contained 15 patients and was used for manual tracking validation. Set B was a joint echo and tagged MRI set and was used for both manual tracking and comparison with tagged MRI (as will be discussed in Sect. 3.3.2) (Fig. 2).

3.3.1 Set A: Echocardiography Studies

Data from fifteen subjects who had already undergone echocardiographic imaging as part of their diagnostic evaluation were deidentified and transferred to the laboratory following IRB approval. The data included 13 male, 4 female, average age 52.9 ± 7.3, consisting of hypertension (8 cases), Coronary Artery Disease (4 cases), Left Ventricular Hypertrophy (4 cases), Congestive Heart Failure (1 case), Chronic Obstructive Pulmonary Disease (2 cases), Diabetes Mellitus (2 cases) and smokers (1 case). 2D echocardiography (short-axis, long-axis, four-chamber, two-chamber

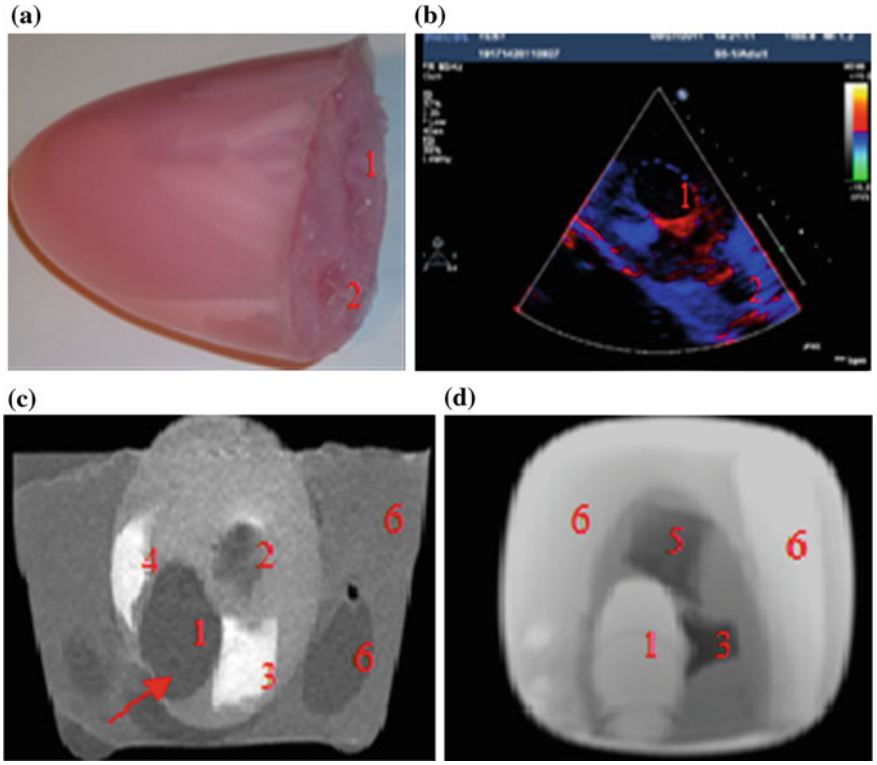


Fig. 1 Seventeen AHA prescribed segments for the heart. **a** Basal SAX view, **b** mid-LV SAX view, **c** apical SAX view (*AS* antero-septal, *A* anterior, *L* lateral, *P* posterior, *I* inferior, *IS* infero-septal) [35, 37, 38]

B-mode with TDI. At the University of Louisville Hospital's echocardiography laboratory, Echocardiographic images are acquired with a commercially available system (iE33, Philips Health Care, Best, The Netherlands) using a S5-1 transducer (3 MHz frequency) and the operator is free to change the gain and filter as needed. The full data set included two-chamber, three-chamber, four-chamber, and long-axis views.

3.3.2 Set B: Echocardiography-MRI Studies

The prospective protocol for patient selection and imaging was approved by the Institutional Review Board of the Robley Rex Veterans' Affairs Medical Center, and a written informed consent was obtained from patients. Eight male subjects were prospectively recruited to the study with average age 54.6 ± 8.5. The subjects included hypertension (5 cases), Coronary Artery Disease (2 cases), Congestive Heart Failure (1 case), Chronic Obstructive Pulmonary Disease (1 case), Diabetes

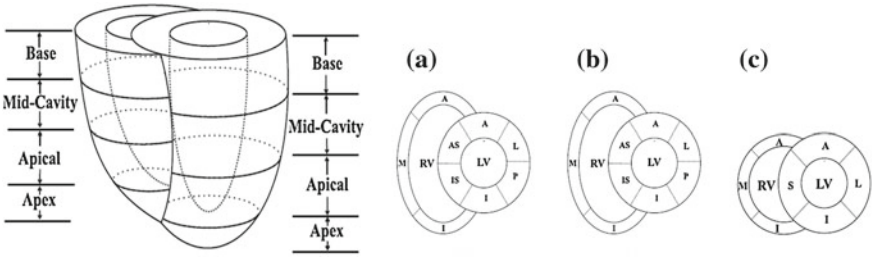


Fig. 2 **a** A picture of the two-chamber model. **b** TDI image of the moving phantom during balloon inflation. **c** A static slice of the phantom using T1 weighted FFE. The *arrow* points to the aneurysm (the thin ventricular wall). **d** A static slice of the phantom using balanced FFE (*1* LV, *2* RV, *3* cylindrical inclusion, *4* slab-like inclusion, *5* cube like inclusion, *6* mediastinum and mediastinal structures)

Mellitus (3 cases) and smokers (4 cases).The imaging protocol included a primary 2D echocardiography including short-axis, long-axis, three-chamber, four-chamber and two-chamber B-mode and TDI imaging as well as simultaneous B-mode/TDI imaging (two-chamber, three-chamber, four-chamber, long-axis). At the Robley Rex Veterans Affair Medical Center's echocardiography laboratory, Echocardiographic images are acquired with an iE33 commercial echocardiography system (Philips Health Care, Best, The Netherlands) using a S5-1 transducer (3 MHz frequency) and the operator is free to change the gain and filter as needed.

Following Ultrasound imaging, cine and tagged MRI data were collected in all subjects. Tagged MRI data acquisition was performed using Philips Achieva, TFE/GR sequence, TE/TR 2/4 ms, Flip Angle 15, spatial resolution 1.25×1.25 mm, slice thickness 8 mm, and spatial size $256 \times 256 \times 8$ pixels. In all subjects, both echocardiography and MR imaging were performed within two hours to decrease any confounding events that could cause discrepancy between wall motion studies in echocardiography and MRI. MRI was performed immediately after the echocardiography. In order to ensure that the B-mode and TDI images were matched, B-mode and TDI images were simultaneously acquired. Additionally, subjects were asked to hold their breath during data collection.

3.4 In-Vivo Comparison of TDIOF-Derived Strains with Strains from Tagged MRI

Tagged MRI [35] is known to provide highly accurate displacement fields in the systolic portion of the cardiac cycle while the tags last. We analyzed the strain field in echocardiography and tagged MR images of slices similar in location in the two modalities in set B. In selecting corresponding slices, qualitative anatomical landmarks such as the papillary muscles and cardiac contours as well as cine MRI

images were utilized. Anatomical landmarks such as endocardial shape and papillary muscle were used to locate the appropriate short axis sections of the heart. The recently proposed SinMod technique [36] was then used to derive displacements from tagged MRI data in the first few systolic phases of the cardiac cycle, while the tags persisted. SinMod is an automated motion estimation technique for tagged MRI that models the pixels as a moving sine wavefront.Since no pixel to pixel mapping between echo and MR images was known, the ventricular geometry from 2-D echo and tagged MRI was divided into 17 segments following the American Heart Association's recommendations. Subsequently the averaged Lagrangian strain for each of the 17 heart segments were compared between the two modalities. Since the frame rate of echo and MRI is not the same and the heart rate may change, it was necessary to align the images in the temporal dimension. This was done by spline interpolation of the measured strain data in the time domain.

4 Strain Analysis

Strain is a measure of deformation of the cardiac tissue. With I representing the identity matrix, the Lagrangian strain tensor at a given myocardial point and for a specific time point can be expressed as:

$$E = \frac{1}{2}(F^T F - I) \tag{2}$$

where the elements of the deformation gradient tensor, F, are:

$$F = \begin{pmatrix} \frac{\partial x}{\partial X} & \frac{\partial x}{\partial Y} & \frac{\partial x}{\partial Z} \\ \frac{\partial y}{\partial X} & \frac{\partial y}{\partial Y} & \frac{\partial y}{\partial Z} \\ \frac{\partial z}{\partial X} & \frac{\partial z}{\partial Y} & \frac{\partial z}{\partial Z} \end{pmatrix} \tag{3}$$

while $x = X + V(X)$, X represents the spatial coordinates in the undeformed coordinates (typically taken to be the end-diastolic frame), and $V(X)$is the accumulated motion vector at the corresponding spatial location relative to the undeformed state . For the echocardiography data, the reference frame for the strain computation was considered to be the end diastolic frame and was selected based on ECG trigger. The deformation field was then computed between each two frames and was added to the motion field from the previous frame in order to measure the accumulated deformation and strain . Since the deformation field of the consecutive frames do not represent the motion of the same pixels, spline interpolation was used to align the deformation fields. For the tagged MRI data, the end-diastolic frame was always the first acquired image which is collected immediately after the R-wave trigger.

The normal strain in the direction of the unit vector n can be calculated from the Lagrangian strain tensor through the quadratic form $n^T n$, where n is a unit vector and can point to any direction on the unit sphere. Due to the geometry of the left ventricle,

the normal strains are usually calculated in radial, circumferential, and longitudinal directions.

Regional analysis is performed on 17 American Heart Association (AHA) prescribed segments. Figure 7 shows the different segments. The acronyms stand for antero-septal (AS), anterior (A), lateral (L), posterior (P), inferior (I), and infero-septal (IS). For a review of topics related to determination of strain from cardiac images, the reader is referred to [35, 37, 38].

5 Results

As noted in Sect. 3, TDIOF was applied to three different datasets: simulated images, data collected in a physical phantom, and in vivo data (both set A and set B). To further elucidate the effect of the Doppler term, results from TDIOF were compared to Horn-Schunck (HS) optical flow (beta = 0 in Eq. (A.11)) and block-matching (BM) (see Appendix) with the latter being the basis for most commercial speckle tracking methods [13]. Since the performance of each technique depends on the parameters of the method, it was necessary to optimize the parameters. Based on simulated images, an exhaustive search was performed over the parameters of TDIOF, HS optical flow, and BM speckle tracking method (a large range was considered for each parameter) and the best values were selected experimentally. To analyze the performance of the techniques with different parameter settings, simulated images were compared to the next simulated frame after being warped using the estimated motion field. Relative mean absolute error was used for the comparison. Relative mean absolute error was computed as $\frac{1}{N}\sum_{i,j}\left\|\hat{I}-I\right\|/I$; where I and $\hat{I}$ are the first and subsequent warped images, and N is the total number of points. Figure 3 shows the performance of the TDIOF technique using different parameters. The methods were then applied to all the datasets using the resulting parameters: number of scales for multiscale implementation: 5, α (smoothness weight): 2000, β (TDI similarity weight): 0.001, and σ (penalizer parameter): 0.1. The parameters for the HS technique were set as follows: number of scales 5, and smoothness weight 2000.

5.1 Validations on Simulated Images

The simulated 3D cardiac model built based on Arts' et al. [31] is shown in Fig. 4a. The deformation shown in the figure is that of a systolic motion. The 3D B-mode image deformation was computed based on [29] and was shown in Fig. 4b. Figure 4c shows the computed TDI using the simulated sequence—the red colors represent motion towards the transducer and the blue colors represents motion away from the transducer. Figure 5 shows application of TDIOF to simulated data and comparison with ground truth. Angular and magnitude error metrics were used for validation of

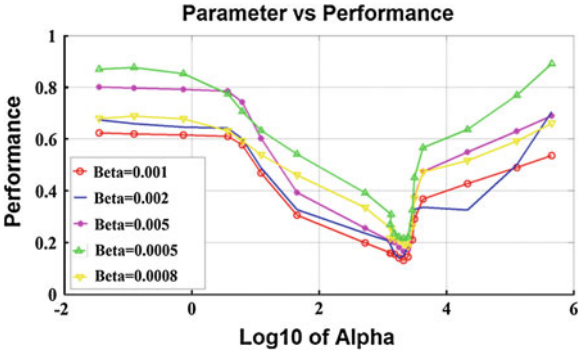


Fig. 3 Performance of TDIOF using different parameters based on relative mean absolute error: X-axis is shown with a logarithmic scale in order to report a wide range of parameter settings. The performance of TDIOF is plotted versus smoothness coefficient for different TDI similarity coefficients (β). Changes of performance is evident when smoothness parameter (α) changes. As seen from the plots, performance was more dependent on the smoothness and insensitive to the scale for the TDIOF term

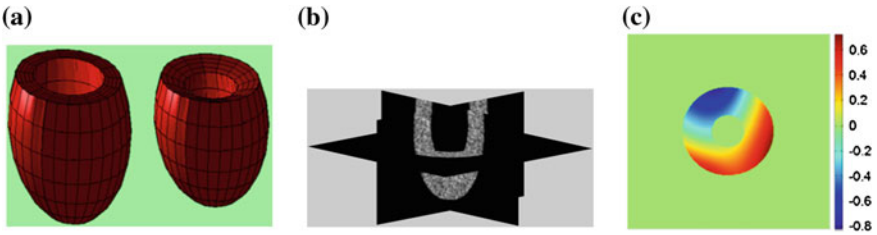


Fig. 4 **a** Simulated cardiac model in diastole and systole. **b** A 3D simulated B-mode image based on COLE. **c** The computed tissue Doppler image using the simulated sequence

the proposed technique as stated in Eqs. (4) and (5):

$$Magnitude\ Error = \frac{1}{N}\sum_{i,j}\left|\frac{\|\hat{V}\| - \|V\|}{\|V\|}\right| \tag{4}$$

$$Angular\ Error = \frac{1}{N}\sum_{i,j}\left|Arc\cos\frac{\left\langle V, \hat{V}\right\rangle}{\|V\| \cdot \|V\|}\right| \tag{5}$$

where V and $\hat{V}$ are the true and estimated displacement vectors and N is the total number of vectors.

To quantitatively analyze the proposed method, averaged performance of TDIOF, Horn-Schunck optical flow, and block matching speckle tracking are reported in Table 3. The methods were applied to 14 simulated cardiac frames (one full cardiac

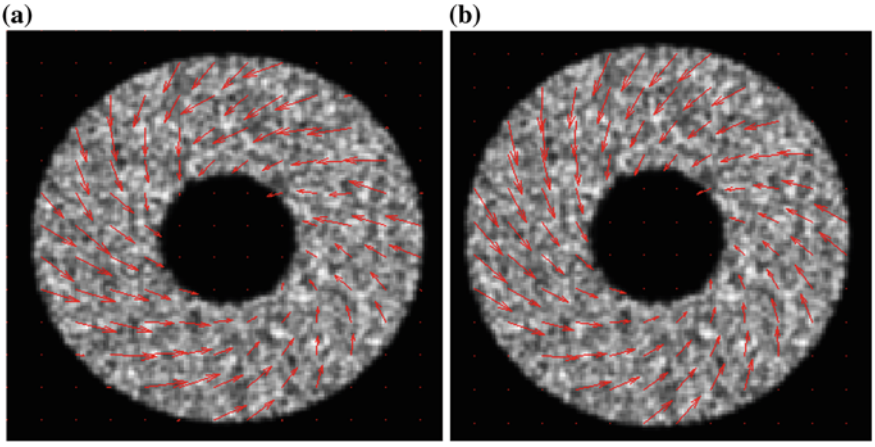


Fig. 5 **a** Results of TDIOF from a mid ventricular section of the 3D simulated echo data. **b** Ground truth motion field

Table 3 TDIOF versus HS optical flow and BM speckle tracking when applied to simulated images

Data	Simulation (no noise)			Simulation (SNR 1.12 db)		
Method	TDIOF	HS	BM	TDIOF	HS	BM
Magnitude error (pixel/frame)	$\mathbf{0.15 \pm 0.09}$	0.20 ± 0.13	0.20 ± 0.14	$\mathbf{0.22 \pm 0.12}$	0.34 ± 0.16	0.31 ± 0.15
Angular error (degrees/frame)	$\mathbf{9.2 \pm 3.8}$	11.2 ± 5.2	11.3 ± 5.6	$\mathbf{10.0 \pm 5.5}$	12.5 ± 6.8	12.7 ± 6.0

BM block-matching, *HS* Horn-Schunck

cycle) of size $300 \times 300 \times 150$ pixels with and without noise. The error represents the angular or magnitude error averaged over all 100 slices and over all 14 temporal frames (averaged in both space and time). Please note that TDIOF was applied to 100 short-axis cross sections of the simulated heart. Table 3 illustrates that TDIOF has markedly improved performance on noisy images. Figure 6 shows the magnitude and angular error over one cardiac cycle for the 3 techniques—note that for each time point, the errors have been averaged over all spatial positions and all slices.It is evident that for all methods the errors are more pronounced in systolic frames compared to diastolic frames. This, we believe, is due to larger out of plane displacements causing errors for the 2-D method. From the figure, it can also be observed that, TDIOF outperforms Horn-Schunck optical flow and BM speckle tracking more significantly on noisy images.

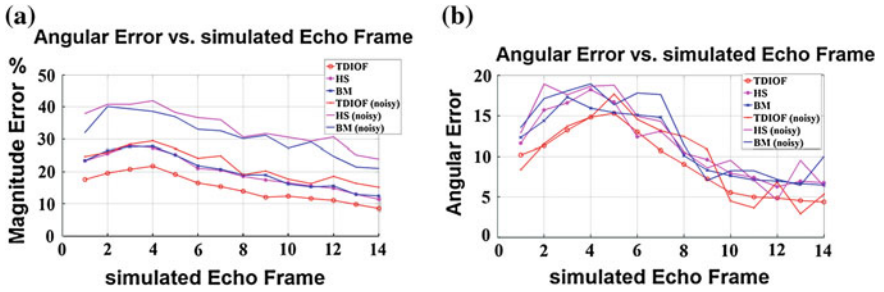


Fig. 6 Comparison of (**a**) magnitude (Eq. (4)) and (**b**) angular error (Eq. (5)) over one cardiac cycle for different techniques

5.2 Validation on Data Collected in a Physical Phantom

In order to validate TDIOF on phantom data, the enamel markers on the B-mode images were manually segmented and the centers of mass of the markers were considered as landmarks. The error was computed on 128 landmarks over one cardiac cycle with 54 2D echocardiographic frames. As with simulated images, angular and magnitude errors were used to analyze the performance. The averaged magnitude and angular error of the landmarks for TDIOF, HS optical flow, and BMspeckle tracking are shown in Table 4. Figure 7 shows application of TDIOF and TDI to the physical phantom.

5.3 Validation on In Vivo Images

The algorithm was also evaluated in a similar way using in vivo images with 519 landmarks selected by an expert over 106 sets acquired from 23 patients. Landmarks were prominent regions in in vivo images such as speckles that could easily be detected. Each landmark was delineated and the center of mass of the landmark was defined to be the actual location. The average error for each of the three methods (TDIOF, Horn-Schunck, block matching) applied to in vivo data was classified per segment and is reported in Table 5. Figure 8a shows the application of the TDIOF technique to one four-chamber in-vivo B-mode study in systole. Figure 8b shows the end-systolic longitudinal strain map for the same patient.

Table 4 Comparison of performance of TDIOF with HS optical flow and BM speckle tracking on physical phantom data

Data	Phantom (slow heart rate)			Phantom (fast heart rate)		
Method	TDIOF	HS	BM	TDIOF	HS	BM
Magnitude error (pixel/frame)	$\mathbf{0.22 \pm 0.13}$	0.37 ± 0.19	0.32 ± 0.17	$\mathbf{0.35 \pm 0.15}$	0.49 ± 0.21	0.47 ± 0.22
Angular error (degrees/frame)	$\mathbf{15.2 \pm 4.7}$	29.0 ± 9.7	27.3 ± 9.4	$\mathbf{26.5 \pm 9.8}$	47.6 ± 12.1	42.0 ± 11.6

BM block-matching, *HS* Horn-Schunck

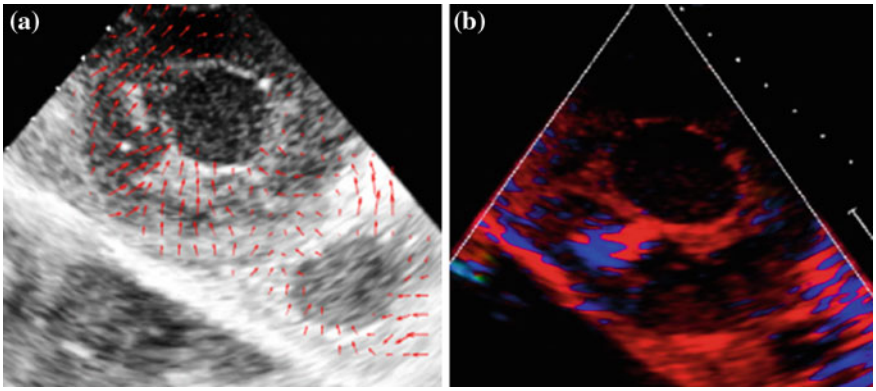


Fig. 7 **a** TDIOF applied to phantom data in "systole", **b** TDI forthe same phase

5.4 Preliminary Comparison of Strains from TDIOF and Tagged MRI

In this part of the study, radial and circumferential strains derived from TDIOF, HS optical flow, and BM speckle tracking were computed from B-mode echoand were compared to tagged MRI strains. Anatomical landmarks such as endocardial shape and papillary muscle locations were used to locate the corresponding short axis sections of the heart in tagged MRI and echocardiography. In addition, since the papillary muscles could not be easily visualized in the tagged studies, non-tagged cine MR images were used to better define the papillary muscles locations. Despite these efforts to ensure correspondence of the data, as alignment of the data based on landmarks could only be approximate (due to differences in slice thickness and identical view orientation in echo and MRI), and the results reported here should only be qualitatively interpreted.

The image-derived strain values related to the same cardiac phase and the same sections of the same patient were compared by averaging the corresponding radial and circumferential strain values for each of the 17 AHA segments. For alignment, the short-axis tagged MR images were visually matched to the corresponding short axis echocardiographic images acquired from basal, mid-ventricular, and apical slices. Since the tag lines fade after systole, only the first 3–4 systolic tagged frames and the corresponding temporal extent in echo was considered in this analysis. Furthermore, since the number of the frames in echocardiography is several times that of tagged MRI data (roughly 20 tagged MR frames versus 4 echocardiographic frames during the cardiac cycle), the strain fields in echo images were interpolated using spline interpolation to match the systolic tagged MRI frames. Finally, 2-D strain maps from corresponding echocardiography and tagged MRI were computed and averaged over 17 segments.

Table 5 Comparison of displacement errors for TDIOF, HS optical flow, and BM speckle tracking on 17 AHA segments averaged over 23 patient data sets

Method		Apical-Ant	Apical-LAT	Apical-INF	Apical-septal	Mid-Ant	Mid-AL	Mid-IL	Mid-INF	
Magnitude error (pixel/frame)	TDIOF	0.25 ± 0.09	0.23 ± 0.09	0.25 ± 0.10	0.26 ± 0.10	0.18 ± 0.07	0.19 ± 0.06	0.19 ± 0.06	0.17 ± 0.07	
	HS	0.40 ± 0.14	0.39 ± 0.12	0.44 ± 0.16	0.41 ± 0.14	0.36 ± 0.13	0.36 ± 0.14	0.34 ± 0.13	0.37 ± 0.12	
	BM	0.37 ± 0.14	0.35 ± 0.13	0.43 ± 0.14	0.40 ± 0.09	0.32 ± 0.14	0.34 ± 0.11	0.33 ± 0.09	0.32 ± 0.14	
Angular error (degrees/frame)	TDIOF	22.7 ± 0.09	27.5 ± 0.10	25.4 ± 0.09	23.5 ± 0.08	17.7 ± 0.08	16.3 ± 0.07	19.0 ± 0.08	20.1 ± 0.08	
	HS	34.2 ± 0.11	35.5 ± 0.13	39.1 ± 0.12	41.7 ± 0.14	32.4 ± 0.09	31.6 ± 0.08	32.0 ± 0.14	35.4 ± 0.09	
	BM	26.5 ± 0.09	25.7 ± 0.09	24.5 ± 0.10	27.5 ± 0.09	23.8 ± 0.10	29.2 ± 0.14	25.9 ± 0.10	24.0 ± 0.09	
Method		Mid-IS	Mid-AS	Basal-Ant	Basal-AL	Basal-IL	Basal-INF	Basal-IS	Basal-AS	Apex
Magnitude error (pixel/frame)	TDIOF	0.20 ± 0.08	0.17 ± 0.07	0.25 ± 0.09	0.24 ± 0.10	0.23 ± 0.10	0.28 ± 0.09	0.24 ± 0.10	0.26 ± 0.09	0.61 ± 0.19
	HS	0.35 ± 0.14	0.33 ± 0.11	0.41 ± 0.14	0.39 ± 0.12	0.42 ± 0.14	0.43 ± 0.10	0.44 ± 0.11	0.46 ± 0.15	0.78 ± 0.22
	BM	0.33 ± 0.12	0.34 ± 0.13	0.37 ± 0.12	0.38 ± 0.09	0.40 ± 0.14	0.41 ± 0.10	0.44 ± 0.14	0.40 ± 0.13	0.72 ± 0.25
Angular error (degree/frame)	TDIOF	19.3 ± 0.09	17.8 ± 0.08	23.5 ± 0.10	21.8 ± 0.08	26.7 ± 0.10	28.7 ± 0.11	24.5 ± 0.08	23.2 ± 0.09	52.9 ± 0.23
	HS	37.3 ± 0.14	31.0 ± 0.13	37.8 ± 0.12	39.5 ± 0.14	32.6 ± 0.13	35.4 ± 0.10	33.7 ± 0.09	37.1 ± 0.14	58.6 ± 0.21
	BM	24.8 ± 0.10	26.5 ± 0.08	27.2 ± 0.09	29.3 ± 0.11	29.7 ± 0.14	31.5 ± 0.12	29.1 ± 0.09	34.0 ± 0.14	56.6 ± 0.22

AS antero-septal, *ANT* anterior, *LAT* lateral, *IL* infero-lateral, *IL* infero-lateral, *INF* inferior, *IS* infero-septal

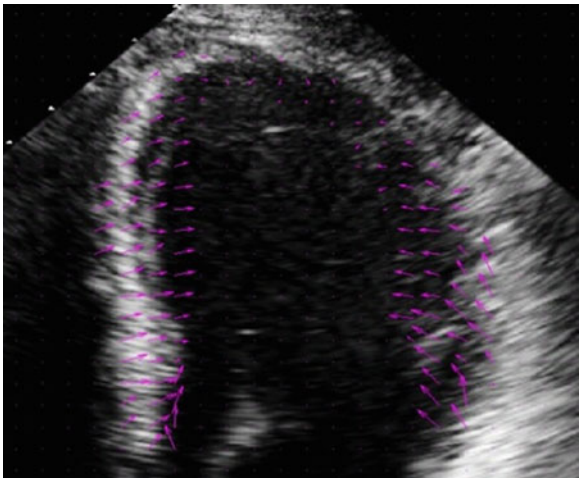


Fig. 8 Application of TDIOF to four-chamber B-mode data during diastole. As expected, TDIOF-derived displacements are larger for the basal segments when compared to the apical segments

Figure 9a shows B-mode and TDI images in early systole at the high papillary muscle level of a subject. The computed motion of the heart between these two frames is shown in Fig. 9b and c based on HS optical flow and TDIOF, respectively. The cardiac strain maps for the same cardiac phase and same slice are shown in Figs. 10 and 11. Figure 11 compares the radial strain map with the tagged MRI radial strain map. As expected and observed from the tagged MRI results, the radial strains from TDIOF are positive and gradually increase during systole. The increased radial strain is more pronounced in AL and IL segments in both SinMod derived and TDIOF strain maps. The increased radial strain is also prominent in the AS and IS segments. Figure 11 shows the circumferential strain map compared to the tagged MRI circumferential strain map. As expected and observed from the tagged MRI results, the circumferential strains from TDIOF are negative and gradually increase in magnitude during systole. This increase is more pronounced in AL and ANT segments in both SinMod tagged MRI-derived and TDIOF strain maps.

To compare the performance of TDIOF and HS, statistical analysis of the strain map results are helpful. Figure 12 shows correlation studies of the radial and circumferential strain values compared to tagged MRI. The correlation coefficient (r) for the TDIOF radial strain values compared to the tagged MRI radial strain values was 0.83 ($P < 0.001$); while the correlation coefficient (r) for the HS and BM radial strain values compared to the tagged MRI radial strain values were 0.71($P < 0.001$) and 0.75($P < 0.001$), respectively. The correlation coefficient (r) for the TDIOF circumferential strain values compared to the tagged MRI circumferential strain values was 0.86 ($P < 0.001$); while the correlation coefficient (r) for the HS and BM circumferential strain values compared to the tagged MRI circumferential strain values

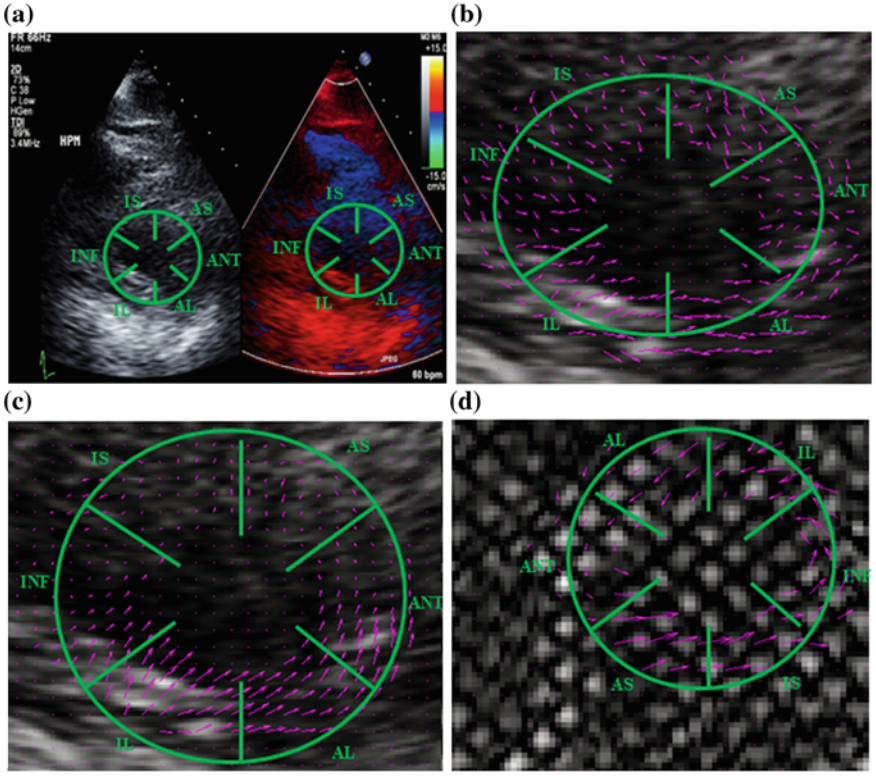


Fig. 9 **a** A short axis B-mode image at the high papillary muscle level of a subject during early systole compared to TDI image for the same phase. **b** Horn-Schunck motion field for the same phase and in the same subject as **a**. **c** Corresponding TDIOF motion field. **d** Tagged MRI motion field for the same approximate slice location in systole (*AS* antero-septal, *ANT* anterior, *IL* infero-lateral, *AL* antero-lateral, *P* posterior, *INF* inferior, *IS* infero-septal)

were 0.77 ($P < 0.001$) and 0.79 ($P < 0.001$). Therefore, it may be concluded that for both radial and circumferential strains, TDIOF analysis achieves a more significant correlation with the tagged MRI in comparison to HS and BM analysis. This effect is believed to be due to the additional Doppler term that is added to the TDIOF framework. The comparison of TDIOF and HS radial strain using student t-test showed superiority of TDIOF (95 % confidence interval, $P < 0.001$). Similarly, the comparison of TDIOF and HS circumferential strain using student t-test showed superiority of TDIOF (95 % confidence interval, $P < 0.001$).The comparison of TDIOF and BM radial strain using student t-test was statistically meaningful (95 % confidence interval, $P < 0.001$).The comparison of TDIOF and BM circumferential strain using student t-test was prominent as well (95 % confidence interval, $P < 0.001$).

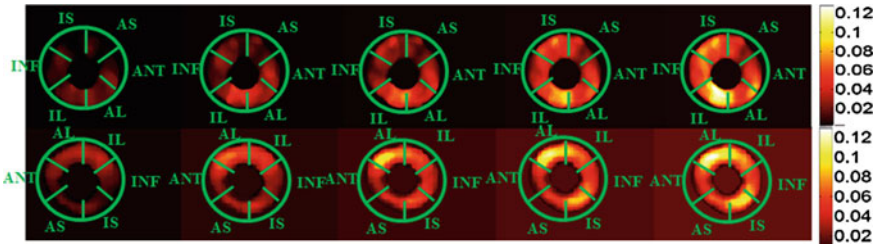


Fig. 10 *Top row* Lagrangian radial strain maps computed from TDIOF. *Lower row* Lagrangian radial strain maps computed with SinMod from tagged MRI during the same cardiac phase at the high papillary muscle level in one subject (*AS* antero-septal, *ANT* anterior, *IL* infero-lateral, *AL* antero-lateral, *P* posterior, *INF* inferior, *IS* infero-septal). The tagged MR images are resized to match the echo images with respect to the size

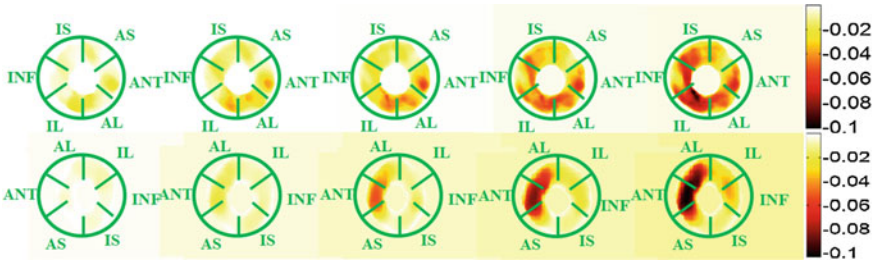


Fig. 11 *Top row* Lagrangian circumferential strain maps computed from TDIOF. *Lower row* Lagrangian circumferential strain maps computed with SinMod from tagged MRI during the same cardiac phase at the high papillary muscle level in one subject (*AS* antero-septal, *ANT* anterior, *IL* infero-lateral, *AL* antero-lateral, *P* posterior, *INF* inferior, *IS* infero-septal). The tagged MR images have been resized to match the echo images with respect to size

6 Discussion

In order to increase the accuracy of motion estimation and speckle tracking techniques and to overcome the angle dependency of TDI, fusion of the techniques has been proposed. TDIOF makes use of the combination of B-mode and Doppler energy terms, minimized using linear algebraic methods. It was demonstrated that TDIOF outperforms the Horn-Schunck optical flow technique and block matching speckle tracking when applied to simulated, physical phantom, and real data. In this paper, we demonstrated that the additional Doppler term is able to increase the accuracy of the intensity (B-mode) based methods in tracking left-ventricular wall motion. The additional Doppler term may very well be added to other cardiac Ultrasound image registration techniques and we expect a corresponding improvement in performance. As demonstrated in the simulation study, the improvement in performance is more pronounced on noisy images.

TDIOF had better performance when compared to HS and Block matching in simulated, phantom, and in vivo data. Due to increased thickness of the wall, the

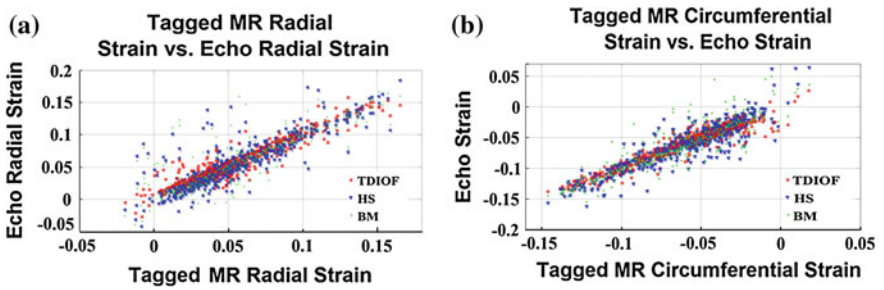


Fig. 12 **a** TDIOF ***radial strain*** value versus tagged MRI radial strain is shown as *red*; while HS radial strain value versus tagged MRI radial strain value is shown as *blue stars*; and BM radial strain vs tagged MRI radial strain is shown as *green dots*. **b** TDIOF ***c***ircumferential strain value versus tagged MRI circumferential strain is shown as *red*; while HS circumferential strain value versus tagged MRI circumferential strain value is shown as *blue stars*; and BM circumferential strain versus tagged MRI circumferential strain is shown as *green dots*. For both cases, it is evident that the *blue dots* (HS strain values) are more scattered compared to the tagged strain values. The plots include corresponding average strain quantities for 17 segments in 8 patients over 4 tagged MRI frames

results were better in mid-ventricular slices for all three methods. Nevertheless, results at the basal and apical slices were still acceptable. Due to poor acquisition at the apex, results for apical segments,were not as good for all three techniques compared. Similarly, in comparison to HS and BM, results from TDIOF correlated more significantly with tagged MRI. It is evident from Figs. 11 and 12, that both radial and circumferential strains increase over the cardiac systole, while the heart is contracting and peaks attend systole and then as the heart recoils back to the original length the cardiac strain decreases to about zero at end diastole. It should be noted that the strain values for TDIOF and tagged MRI are not exactly the same because it is not possible to perfectly align the images in space and time due to differences in image slice thickness, resolution, and precise image orientation.

6.1 Comparison with Previous Work

A comparison of correlative strain results for TDIOF reported in this paper can further illustrate the performance of the proposed technique. In [39], a comparison of MRI-derived strains and speckle tracking-derived strains were reported. The authors collected data in patients using a commercially available system (Vivid 7, GE Vingmed Ultrasound AS, Horten, Norway) and performed off-line analysis (EchoPac BT04, GE Vingmed Ultrasound AS). Subsequently, the same group of patients underwent tagged MRI scan and HARP off-line analysis to determine the regional strains. The correlation between radial strain based on B-mode speckle tracking and tagged MRI was reported to be ($r = 0.60, p < 0.001$) while the correlation between circumferential and longitudinal strain values based on B-mode speckle tracking and tagged MRI was reported to be ($r = 0.51, p < 0.001$) and ($r = 0.64, p < 0.001$). The

authors concluded that there is a modest correlation between echocardiographic and tagged-MRI-derived strains.

6.2 Limitations

The present study has several limitations that should be stated. At this time it is not possible to extend TDIOF to three dimensions because TDI is only possible in two dimensions.

Another limitation is lack of availability of ground truth applicable to in vivo images which makes the validation more difficult. Tagged MRI is a good surrogate but it is not perfect. Tagged MRI slices do not exactly overlap on echocardiographic slices and there is no accurate pixel to pixel mapping from the cardiac tissue in tagged MRI to the cardiac tissue in echocardiography. Additionally, the orientation of the Ultrasound transducer is not exactly the same as image orientation in tagged MRI. Furthermore, Echocardiography and tagged MRI have different resolutions in space and time.

An additional potential issue is that MRI and echocardiography cannot be performed simultaneously. In our study, since MRI was performed immediately after echocardiography, the cardiac physicologic changes are felt to be less significant. However, heart rate variability may cause alignment problems between the images. We attempted to overcome these issues by careful image acquisition and matching of the slices in space and time.

7 Conclusion

In order to increase the accuracy of the speckle tracking technique and to cope with the angle dependency of TDI, a combined approach dubbed TDIOF (Tissue Doppler Imaging Optical Flow) has been proposed. TDIOF is formulated based on the combination of B-mode and Doppler energy terms minimized using linear algebraic methods. TDIOF was validated extensively based on simulated images, physical cardiac phantom, and *in-vivo* data. The performance of TDIOF was demonstrated to be better than popular motion estimation and speckle tracking techniques in echocardiography.

Appendix A: Mathematical Framework for TDIOF

To determine myocardial motion, we propose a novel optical flow energy function which combines three energy terms: B-mode intensity constancy, Doppler/B-mode velocity similarity, and motion smoothness.

(1) B-mode intensity constancy: If we assume that $p = (x, y, t)$and the flow field is $w(p) = (u(p), v(p), 1)$ where u and v are the motion vectors and x, y and t are the spatial and temporal dimensions, B-mode intensity constancy term assumes that the pixel intensity is the same along the motion vector. When $I(p)$ is the pixel intensity at location p and $I(p + w)$ is the pixel intensity in the subsequent frame at location p + w,

$$E_{data} = |I(p + w) - I(p)|^2 \tag{A.1}$$

Although optical flow is usually solved using calculus of variation, we use the recent incremental flow framework [25] which provides significant computational savings. The incremental flow assumes that an estimate of flow is already known (iteration 0) and then, the best increment will be found at each iteration,.With inclusion of an incremental motion vector, the intensity constancy is then revised to be:

$$E_{data} = |I(p + w + dw) - I(p)|^2 \tag{A.2}$$

The above equation can be linearized using Taylor series expansion:

$$I_t(p + w + dw) - I(p) = I_t(p) + I_x(p)du(p) + I_y(p)dv(p) \tag{A.3}$$

with

$$I_x(p) = \frac{\partial I(p + w)}{\partial x} \tag{A.4}$$

$$I_y(p) = \frac{\partial I(p + w)}{\partial y} \tag{A.5}$$

$$I_t(p) = I(p + w) - I(p) \tag{A.6}$$

(2) The smoothness energy term forces the flow field to be continuous:

$$E_s = |\nabla(u + du)|^2 + |\nabla(v + dv)|^2 \tag{A.7}$$

with

$$|\nabla(u + du)|^2 = \left(\frac{\partial(u + du)}{\partial x}\right)^2 + \left(\frac{\partial(u + du)}{\partial y}\right)^2 \tag{A.8}$$

(3) TDI velocity term. The 2D motion when projected in the direction of the transducer should be similar to the computed velocity. If $\vec{\mathbf{v}} = (u, v)$ is the B-mode velocity, $\vec{\mathbf{v}}_t = (u_t, v_t)$ is the transducer orientation and w_{tdi} is the TDI velocity acquired from the echo machine, then the constraint is formulated as:

$$E_{tdi} = \left(\vec{\mathbf{v}}^{\mathbf{T}}\vec{\mathbf{v}}_{\mathbf{t}} - w_{tdi}\right)^2 = (u_t u + v_t v - w_{tdi})^2 \tag{A.9}$$

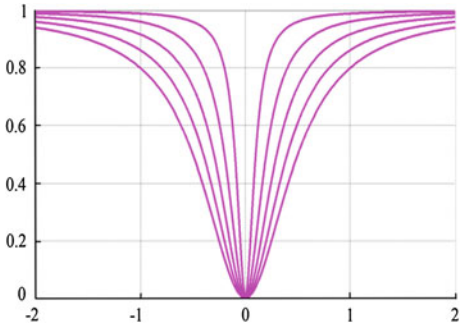


Fig. A.1 Geman-Mcclure penalizer for different σ parameters (0.1, 0.3, 0.5, 0.7, 0.9)

In order to keep the range of E_{tdi} between 0 and 1, and to reject outliers, we utilized a Geman-Mcclure penalizer (ψ) [26]:

$$\psi(\mathrm{s}) = \frac{\mathrm{s}^2}{\mathrm{s}^2 + \sigma^2} \tag{A.10}$$

In (A.10), s is the input data and σ is the scaling parameter. The behavior of Geman-Mcclure equation is shown in Fig. A.1.

The total energy function to be minimized is:

$$\begin{aligned} E_(u, v) &= E_{data} + \alpha E_s + \beta\psi(E_{tdi}) \qquad \text{(A.11)} \\ &= \int_{\Omega} (|I(P + w + dw) - I(p)|^2 + \alpha(|\nabla(u + du)|^2 + |\nabla(v + dv)|^2) \\ &\quad + \beta.\psi((u_t(u + du) + v_t(v + dv) - w_{tdi})^2)) \end{aligned}$$

where α is the smoothness weight and β is the TDI/velocity correspondence parameter—we note that setting beta to zero essentially results in the Horn and Schunck optical flow frame case in the incremental flow framework.

Next, we vectorize u, v, du, dv as U, V, dU, dV.

$$\mathrm{I_x} = diag(I_x)\mathrm{I_y} = diag(I_y)$$

$\mathrm{D_x}$ and $\mathrm{D_y}$ are denoted as matrices related to the x and y derivative filters such that: $\mathrm{D_x}U = u \otimes [0 - 11]$. The derivative operator is used to compute the gradient of the image in each direction. Additionally the column vector δ_p is defined as a Dirac function with the only nonzero element at location p such that$\delta_p I_x = I_x(p)$. Now the discretized version of the energy function (Eq. (A.11)) becomes:

$$E = \sum_{p} (\delta_p^T (I_t + I_x dU + I_y dV))^2 + \delta_p^T D_x (U + dU)^2 + (\delta_p^T D_y (U + dU)^2) + (\delta_p^T D_x (V + dV)^2 + (\delta_p^T D_y (V + dV)^2)] + \beta((\delta^T (u_t (U + dU) + v_t (V + dV) - w_{tdi}))^2) \quad (6)$$

To minimize (A.12), Iterative Reweighted Least Squares (IRLS) was used with the stopping criterion that $[\frac{\partial E}{\partial dU}; \frac{\partial E}{\partial dV}] = 0$. Here, it is noteworthy to state that since for matrix A and vectors x, b:

$$\frac{d}{dx} x^T A x = 2Ax \frac{d}{dx} x^T b = b$$

Therefore:

$$\frac{\partial E}{\partial dU} = 2 \sum_{p} \left(I_x \delta_p \delta_p^T (I_y dV + I_t) + I_x \delta_p \delta_p^T I_x dU \right) + \alpha[(D_x^T \delta_p \delta_p^T D_x + D_y^T \delta_p \delta_p^T D_y)(U + dU)] + \beta \psi'(E_{tdi})[u_t \delta_p (\delta_p^T ((U + dU) u_t + (V + dV) v_t - w_{tdi}))] = 2((I_x^2 + \alpha L + \beta \psi' u_t^2) dU + (I_x I_y + \beta \psi' u_t v_t) dV + (\alpha L + \beta \psi' u_t^2) U + \beta \psi' u_t v_t V + (I_x I_t - \beta \psi' u_t w_{tdi}) \quad (A.13)$$

:

$$\frac{\partial E}{\partial dU} = 2((I_x^2 + \alpha L + \beta \psi' u_t^2) dU + (I_x I_y + \beta \psi' u_t v_t) dV + (\alpha L + \beta \psi' u_t^2) U + \beta \psi' u_t v_t V + (I_x I_t - \beta \psi' u_t w_{tdi}) \quad (A.14)$$

where:

$$L = D_x^T \psi' D_x + D_y^T \psi' D_y \quad (A.15)$$

$$\psi' = \mathrm{diag}(\psi'(E_{tdi})) \quad (A.16)$$

and $\sum_p \delta_p^T \delta_p$ is the identity matrix.
Similarly

$$\frac{\partial E}{\partial dV} = 2((I_x I_y + \beta \psi' u_t v_t) dU + (I_y^2 + \alpha L + \beta \psi' v_t^2) dV + \beta \psi' u_t v_t U + (\alpha L + \beta \psi' v_t^2) V (I_y I_t - \beta \psi' v_t w_{tdi}) \quad (A.17)$$

Finally, the following linear equation is derived as:

$$\begin{pmatrix} I_x^2 + \alpha L + \beta \psi' u_t^2 & I_x I_y + \beta \psi' u_t v_t \\ I_x I_y + \beta \psi' u_t v_t & I_y^2 + \alpha L + \beta \psi' v_t^2 \end{pmatrix} \begin{pmatrix} dU \\ dV \end{pmatrix} \tag{A.18}$$
$$= \begin{pmatrix} I_x I_t + \alpha LU - \beta \psi' u_t w_{tdi} + \beta \psi' u_t^2 U + \beta \psi' u_t v_t V \\ I_y I_t + \alpha LV - \beta \psi' v_t w_{tdi} + + \beta \psi' u_t v_t U + \beta \psi' v_t^2 V \end{pmatrix}$$

In practice, u, v, dU and dV are initialized as zero with dU and dV iteratively updated using linear least squares. In order to cover a wide range of displacements and to reduce the computational time, the algorithm is applied in a multi-scale strategy. The coarse scale is tackled in the first step, while the fine scale is computed in the last stage.

Appendix B: Block Matching

The main idea in this type of motion estimation is that each block of a frame moves toward a block with similar intensity in the next frame, when the time interval between the two frames is small. The general strategy is to slide each block of the first frame over the next frame in order to locate the most similar match. To find the best matching block, it is necessary to have a similarity metric that measures the similarity between two blocks. There are several well-known block matching algorithms based on different cost functions such as Mean Absolute Difference (MAD), Mean Square Error (MSE), or correlation. MAD is utilized in this paper because of its accuracy and computational efficiency [28]. MAD is formulated as:

$$MAD = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left| C_{ij} - R_{ij} \right| \tag{A.19}$$

In (A.9), N is the size of macro-block while C_{ij} and R_{ij} define the pixel locations within the blocks. The index i is the shift in x and j is the shift towards y when the main block is sliding over the image [28].

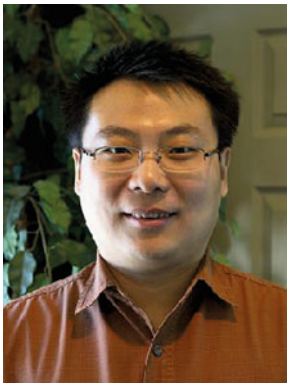
References

1. Kasper DL, Braunwald E, Fauci A (2008) Harrison's principles of internal medicine, 17th edn. McGraw-Hill, New York
2. Fuster V, O'Rourke R, Walsh R, Poole-Wilson P (2007) Hurst's the heart, 12th edn. McGraw Hill, New York
3. American Heart Association, Heart disease and stroke statistics—(2009), update (at-a-glance version). http://www.americanheart.org/presenter.jhtml?identifier=3037327

4. Webb A (2003) Introduction to biomedical imaging. Wiley, Hoboken
5. Hedrick WR, Hykes DL, Starchman DE (2004) Ultrasound physics and instrumentation, 4th edn. Mosby, Chicago
6. Suhling M, Arigovindan A et al (2005) Myocardial motion analysis from B-mode echocardiograms. IEEE Trans Image Process 14(2):525–553
7. Yu W, Yan P, Sinusas AJ, Thiele K, Duncan JS (2006) Towards point-wise motion tracking in echocardiographic image sequences—comparing the reliability of different features for speckle tracking. Med Image Anal 10(4):495–508
8. Paragios N (2003) A level set approach for shape-driven segmentation and tracking of the left ventricle. IEEE Trans Med Imaging 22(6):773–776
9. Hayat D, Kloeckner M, Nahum J, Ecochard-Dugelay E, Dubois-Rande JL, Jean-Francois D et al (2012) Comparison of real-time three-dimensional speckle tracking to magnetic resonance imaging in patients with coronary heart disease. Am J Cardiol 109:180–186
10. Elen A, Choi HF, Loeckx D, Gao H, Claus P, Suetens P, Maes F, D'hooge J (2008) Three-Dimensional cardiac strain estimation using spatio-temporal elastic registration of ultrasound images: a feasibility study. IEEE Trans Med Imaging 27(11):1580–1591
11. Esther Leung KY, Danilouchkine MG, Stralen M van, Jong N. de, Steen AFW van der, Bosch JG (2010) Probabilistic framework for tracking in artifact-prone 3D echocardiograms. Med Image Anal 14(6):750–758
12. Myronenko A, Song X (2010) Point set registration: coherent point drift. IEEE Trans Pattern Anal Mach Intell 32(12):2262–2275
13. Duchateau N, De Craene M, Piella G, Silva E, Doltra A, Sitges M, Bijnens BH, Frangi AF (2011) A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. Med Image Anal 15(3):316–328
14. Bachner-Hinenzon N, Ertracht O, Lysiansky M, Binah O, Adam D (2011) Layer-specific assessment of left ventricular function by utilizing wavelet de-noising: a validation study. Med Biol Eng Comput 49(1):3–13
15. Dydenko I, Jamal F, Bernard O, D'hooge J, Magnin IE, Friboulet D (2006) A level set framework with a shape and motion prior for segmentation and region tracking in echocardiography. Med Image Anal 10(2):162–177
16. Yan P, Sinusas A, Duncan JS (2007) Boundary element method-based regularization for recovering of LV deformation. Med Image Anal 11 (6):540–554
17. De Craene M, Piella G, Camara O, Duchateau N, Silva E, Doltra A, D'hooge J, Brugada J, Sitges M, Frangi A (2012) Temporal diffeomorphic free form deformation application to motion and strain estimation from 3D echocardiography. Med Image Anal 16(2):427–450
18. Ashraf M, Myronenko A, Nguyen T, Inage A, Smith W, Lowe RI et al (2010) Defining left ventricular apex-to-base twist mechanics computed from high-resolution 3D echocardiography: validation against sonomicrometry. JACC Cardiovasc Imaging 3:227–234
19. Papademetris X, Sinusas AJ, Dione P, Constable RT, Duncan JS (2002) Estimation of 3-D left ventricular deformation from medical images using biomechnical models. IEEE Trans Med Imaging 21(7):786–800
20. Papademetris X, Sinusas AJ, Dione DP, Duncan JS (2001) Estimation of 3D left ventricular deformation from echocardiography. Med Image Anal 5(1):17–28
21. Kleijn SA, Brouwer WP, Aly MF, Russel IK, de Roest GJ, Beek AM et al (2012) Comparison between three-dimensional speckle-tracking echocardiography and cardiac magnetic resonance imaging for quantification of left ventricular volumes and function. Eur Heart J Cardiovasc Imaging 13:834–839
22. Garcia D, delA'lamo JC, Tanne' D et al (2010) Two-dimensional intraventricular flow mapping by digital processing conventional color-doppler echocardiography images. IEEE Trans Med Imaging 29(10):1701–1713
23. Dalen H, Thorstensen A, Aase SA, Ingul CB et al (2010) Segmental and global longitudinal strain and strain rate based on echocardiography of 1266 individuals: the HUNT study in Norway. Eur J Echocardiogr 11(2):76–83

24. Amundsen BH, Crosby J, Steen PA, Torp H, Slordahl SA, Stoylen A (2009) Regional myocardial long-axis strain and strain rate measured by different tissue Doppler and speckle tracking echocardiography methods: a comparison with tagged magnetic resonance imaging. Eur J Echocardiogr 10:229–37
25. Liu C (2009) Beyond pixels: exploring new representations and applications for motion analysis. Doctoral thesis. Appendix A, Massachusetts Institute of Technology, Cambridge
26. Geman S, McClure DE (1987) Statistical methods for tomographic image reconstruction. Bull Int Statist Int 52:5–21
27. Horn BKP, Schunck BG (1981) Determining optical flow. Artif Intell 17:185–203
28. Abolhassani MD, Tavakoli V (2009) Optimized thermal change monitoring in renal tissue during revascularization therapy. J Ultrasound Med 28(11):1535–1547
29. Gao H, Choi HF, Claus P, Boonen S, et al (2009) A fast convolution-based methodology to simulate 2-D/3-D cardiac ultrasound images. IEEE Trans Ultrason Ferroelectr Freq Control 56(2):404–409
30. Arts T, Hunter WC, Douglas A, Muijtjens AMM, Reneman RS (1992) Description of the deformation of the left ventricle by a kinematic model. J. Biomech 25(10):1119-l 127
31. Tavakoli V, Kemp J, Dawn B, Stoddard M, Amini AA (2010) Comparison of myocardial motion estimation methods based on simulated echocardiographic B-mode and RF data. SPIE Med Imaging 76260N
32. Tavakoli V, Kendrick M, Alshaher M, Amini A (2012) A two-chamber multi-modal (MR/Ultrasound) cardiac phantom for normal and pathologic hearts. In: International society of magnetic resonance in medicine (ISMRM)
33. Jensen JA (19991) A model for the propagation and scattering of ultrasound in tissue. J Acoust Soc Am 89:182–191
34. ASE Guidelines and Standards—American Society of Echocardiography. http://www.asecho.org/guidelines/
35. Tavakoli V, Negahdar MJ, Kendrick M, Alshaher M, Stoddard M, Amini AA (2012) A biventricular multimodal (MRI/ultrasound) cardiac phantom. In: 2012 Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp. 3187–3190, 28 Aug 2012–1 Sept 2012
36. Lesniak-Plewinska B, Cygan S, Kaluzynski K, D'hooge J, Zmigrodzki J, Kowali E, Kordybac M, Kowalski M (2010) A dual-chamber, thick-walled cardiac phantom for use in cardiac motion and deformation imaging by ultrasound. Ultrasound Med Biol 36(7):1145–1156
37. Amini A, Prince J (eds) (2001) Measurement of cardiac deformations from MRI: physical and mathematical models. Kluwer Academic Publishers, Dordrecht
38. Arts T, Prinzen FW, Delhaas T, Milles J, Rossi A, Clarysse P (2010) Mapping displacement and deformation of the heart with local sine wave modeling. IEEE Trans Med Imag 29(5):1114–23
39. Tustison NJ, Roman VGD, Amini AA (2003) Myocardial kinematics from tagged MRI based on a 4-D B-spline Model. IEEE Trans Biomed Eng 50(8):1038–1040
40. Jasaityte R, Heyde B, D'hooge J (2013) Current state of three-dimensional myocardial strain estimation using echocardiography. J Am Soc Echocardiogr Official Publ Am Soc Echocardiogr 26(1):15–28
41. Jensen JA, Svendsen NB (1992) Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. IEEE Trans Ultrason Ferroelec Freq Contr 39:262–267
42. Tustison NJ, Amini AA (2006) Biventricular myocardial strains via nonrigid registration of anatomical NURBS models. IEEE Trans Med Imaging 25(1):94–112
43. Cho G-Y, Chan J, Leano R, Strudwick M, Marwick TH (2006) Comparison of two-dimensional speckle and tissue velocity based strain and validation with harmonic phase magnetic resonance imaging. Am J Cardiol 97(11):1661–1666 1 June 2006

Editors Biography



Dr. Shuo Li is a research scientist and project manager in general electric (GE) healthcare, Canada. He is also an adjunct research professor in the University of Western Ontario and adjunct scientist in Lawson health research institute. He is currently leading the digital imaging group of London (http://digitalimaginggroup.ca/) as the funding scientific director. He received his Ph.D. degree in computer science from Concordia University 2006, where his PhD thesis won the doctoral prize, which gives to the most deserving graduating student in the faculty of engineering and computer science. He obtained his master and bachelor degrees both from China. He is the recipient of several GE internal awards. He serves as guest editor and associate editor in several prestigious journals in the field. His current interest is in intelligent medical imaging systems with a main focus on automated medial image analysis and visualization.

S. Li and J. M. R. S. Tavares (eds.), *Shape Analysis in Medical Image Analysis*,
Lecture Notes in Computational Vision and Biomechanics 14,
DOI: 10.1007/978-3-319-03813-1,

João Manuel R. S. Tavares
Faculdade de Engenharia da Universidade do Porto (FEUP)
Rua Dr. Roberto Frias, s/n
4200-465 Porto
PORTUGAL
Email: tavares@fe.up.pt
url: www.fe.up.pt/~tavares

João Manuel R. S. Tavares is graduated in Mechanical Engineering from the University of Porto, Portugal (1992). In 1995, he obtained a M.Sc. in Electrical and Computers Engineering, in the field of Industrial Informatics, also at the University of Porto. In 2001, he obtained a Ph.D. degree in Electrical and Computers Engineering from the same University. Since 2001, he has been senior researcher and project coordinator at the Laboratory of Optical and Experimental Mechanics (LOME) of the Institute of Mechanical Engineering and Industrial Management (INEGI). He was Assistant Professor in the Department of Mechanical Engineering (DEMec) of the Faculty of Engineering of the University of Porto (FEUP) between 2001 and 2011, and since then he has been Associate Professor in the same department. He has been co-chair of several scientific national and international conferences and workshops, including CompIMAGE 2006/2010/2012/2014, VipIMAGE 2007/2009/2011/2013, ICEEBS 2013, CIBEM 2011, BioDENTAL 2009/2012/2014, TMSi 2010, IMAGAPP 2009 and EUROMEDIA 2008, founder and editor-in-chief of the Taylor & Francis international journal "Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization" (CMBBE: Imaging & Visualization), founder and co-editor of the Springer book series "Lecture Notes in Computational Vision and Biomechanics" (LNCV&B), editor and reviewer of several scientific journals, co-author than more of 400 scientific publications, including the co-editing of scientific books and of special issues of scientific journals, and co-author of 3 international patents. In the last years, he has been participating in several scientific research projects, some as coordinator, and the supervisor and co-supervisor of several M.Sc. and Ph.D. thesis. His main research areas include Computational Vision, Medical Imaging, Scientific Visualization, Biomechanics, Human-Computer Interaction and New Product Development.