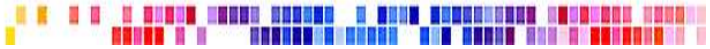
PMMIA 2009: Probabilistic Models for Medical Image Analysis

A MICCAI 2009 Workshop

MICCAI



September 20th, 2009

Imperial College, London, UK

Preface

The workshop Probabilistic Models for Medical Image Analysis (PMMIA 2009) was held in conjunction with the the 12th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2009) on September 20th, 2009 in London, UK. The one day workshop focused on the development, learning and use of probabilistic models for medical image understanding. Probabilistic frameworks are often used for the automatic quantification and generalization of information latent in medical images, which has enabled important work in scientific and disease-oriented research and in surgical guidance. Furthermore, these models frequently are a component of the methodology for such analysis.

The goal of PMMIA was to foster discussions among researchers that are interested in innovative and principled probabilistic models. We were especially interested in developments stimulated by other research communities, such as computer vision, machine learning and biological imaging. Our call for papers resulted in 44 submissions of up to 12 pages. Each paper received at least two reviews. One of our reviews was so detailed and constructive that we decided to include it in the proceedings. Based on these peer reviews, we selected 8 submissions for oral and 19 for poster presentation. To stimulate discussion, each speaker was given 25 minutes for presentation followed by 15 minutes of Q&A. The poster session was held for 2 hours during lunch time.

PMMIA 2009 was only possible due to the contributions of several individuals. First, we would like to thank the relatively high number of researchers that submitted their work to this workshop. Thanks also to the reviewers for providing so many high quality reviews in such a short time span. Finally, the MICCAI organizers provided us with the necessary infrastructure that allowed us to focus our effort on creating an interesting program for the workshop.

William Wells, Sarang Joshi, Kilian Pohl

September 20th, 2009

Conference Committee

Chairs

- Kilian Pohl (IBM)
- Sarang Joshi (University of Utah)
- William Wells (Harvard)

Program Committee

- John Ashburner (University College London)
- Sylvain Bouix (Harvard Medical School)
- Tim Cootes (University of Manchester)
- Jason J. Corso (University of Buffalo)
- Brad Davis (Kitware)
- Stanley Durelman (INRIA)
- Tom Fletcher (University of Utah)
- James Gee (University of Pennsylvania)
- Polina Golland (MIT)
- Casey Goodlett (Kitware)
- Mark Jenkinson (Oxford)
- Alex Leow (UCLA)
- Jose Luis Marroquin (CIMAT)
- Paulo Mendonca (GE)
- Jim Miller (GE)
- Marc Niethammer (UNC)
- Nicos Paragios (Ecole Central de Paris)
- Xavier Pennec (INRIA)
- Marcel Prastawa (University of Utah)
- Michiel Schaap (Erasmus MC)
- Hal Stern (University of California - Irvine)
- Jon Taylor (Stanford)
- Paul Thompson (University of California - Los Angeles)
- Matthew Toews (Harvard Medical School)
- Alain Trouve (Ecole Normale Superieure)
- Zhuowen Tu (University of California - Los Angeles)
- Koen Van Leemput (MIT-Harvard)
- Baba Vemuri (University of Florida)

Contents

Oral Session III

Oral Session IV

Prior knowledge regularization in statistical medical image tasks

Alessandro Crimi[1], Jon Sporring[1], Marleen de Bruijne[2], Martin Lillholm[3] , Mads Nielsen[1,3]

[1] DIKU, University of Copenhagen, Denmark
[2] Erasmus University Medical Center Rotterdam, The Netherlands,
[3] Nordic Bioscience, Denmark.
ac@nordicbioscience.com,

Abstract. The estimation of the covariance matrix is a pivotal step in several statistical tasks. In particular, the estimation becomes challenging for high dimensional representations of data when few samples are available. Using the standard Maximum Likelihood estimation (MLE) when the number of samples are lower than the dimension of the data can lead to incorrect estimation e.g. of the covariance matrix and subsequent unreliable results of statistical tasks. This limitation is normally solved by the well-known Tikhonov regularization adding partially an identity matrix; here we discuss a Bayesian approach for regularizing the covariance matrix using prior knowledge. Our method is evaluated for reconstructing and modeling vertebra and cartilage shapes from a lower dimensional representation and a conditional model. For these central problems, the proposed methodology outperforms the traditional MLE method and the Tikhonov regularization.

1 Introduction

The covariance matrix estimation is required before performing Principal Component Analysis (PCA), Factor Analysis (FA), regressions and several statistical tasks. If the number of samples is small compared to the dimensionality of the data, then the covariance estimation is poor. For solving this problem, several methods were proposed, the most well-known is the Tikhonov regularization [1] where the covariance matrix is boosted by an identity matrix using a regularization parameter (also called the mixing parameter). The regularization parameter is selected to maximize the expected accuracy of the shrunken estimator, using, e.g. cross validation. The resulting covariance matrix can be shown to outperform the standard MLE estimation. The overall aim of this paper is to improve the accuracy of medical diagnosis of diseases, such as Osteoporosis (OP), Osteoarthritis (OA) and Atherosclerosis. This paper is organized as follows: the next section introduces the point distribution model and the main regularization methods, then we analyze two different kinds of priors, the final part contains experiments about vertebrae and cartilage shape reconstruction and a conditional model for aorta location.

2 Background: Statistical shape model

Statistical shape analysis is a geometrical analysis where a set of shapes, often represented with vectors, are measured statistically to describe geometrical properties from similar shapes [2]. In particular, with a Point Distribution Model [2], a shape $\boldsymbol{x} \in \mathbb{R}^d$ relies on a set of p labeled landmark points $\boldsymbol{x} = [\boldsymbol{p}_1^T, \boldsymbol{p}_2^T, \ldots, \boldsymbol{p}_p^T]^T$ with $\boldsymbol{p}_i \in \mathbb{R}^2$ or $\boldsymbol{p}_i \in \mathbb{R}^3$.

Principal component analysis (PCA) is a relevant tool for studying correlations of movement between groups of landmarks among the training set population. It is an orthogonal projection of the data onto a lower dimensional linear subspace, such that the variance of the projected data is maximized. For the data matrix $\mathbf{X} = [(\mathbf{x}_1 - \mu)|(\mathbf{x}_2 - \mu)|\ldots|(\mathbf{x}_n - \mu)]$, where $\mu = E[\mathbf{x}]$ is the mean of the aligned training examples, the PCA coordinates $\mathbf{Y} = [\mathbf{y}_1|\mathbf{y}_2|\ldots|\mathbf{y}_n]$ are given by:

$$\boldsymbol{Y} = \mathbf{V}^T\mathbf{X}. \tag{1}$$

In the previous expression $\boldsymbol{V}$ is the matrix of the eigenvectors on column form obtained with Singular Value Decomposition (SVD) [3] of the estimated covariance matrix $\hat{\boldsymbol{\Sigma}} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$. Summarizing, statistical shape analysis is performed as follows: given n training shapes a linear model is obtained by first aligning them to a common coordinate system using Procrustes analysis [4], which produces the data matrix $\boldsymbol{X}$. This aligned training set forms a cloud in the d dimensional space, which can be considered a sample from a probability density. Estimating the covariance matrix $\hat{\boldsymbol{\Sigma}}$ and calculating its eigenvectors $\boldsymbol{V}$ and eigenvalues $\boldsymbol{\Lambda}$ gives the linear shape model

$$\tilde{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{V}_t\boldsymbol{b}, \tag{2}$$

where $\mathbf{V}_t \in \Re^{d\times t}$, whose columns are the eigenvectors corresponding to the t largest eigenvalues, and $\boldsymbol{b} \in \Re^t$ the vector of shape parameters. The performance of such models can depend on the error of estimation of the covariance matrix $\boldsymbol{\Sigma}$ for the PCA.

An estimate of the covariance matrix, $\hat{\boldsymbol{\Sigma}}$, is traditionally obtained as follows. The shapes $\boldsymbol{x}$ are assumed to be distributed according to the normal density,

$$\begin{aligned} &p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n|\boldsymbol{\Sigma}, \boldsymbol{\mu}) \\ &\quad = \frac{1}{(2\pi)^{\frac{(nd)}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} exp\left[-\frac{1}{2}\Sigma_{i=1}^n(\boldsymbol{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}_i - \boldsymbol{\mu})\right], \end{aligned} \tag{3}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. The density of $\boldsymbol{x}$ is called the Likelihood, and its maximum for varying $\boldsymbol{\Sigma}$ is called the Maximum Likelihood Estimate (MLE) [5]. The point of maximum is found to be,

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T, \tag{4}$$

which is slightly biased, but for large n the bias is negligible.

3 Tikhonov regularization

Using the MLE estimation (4), when the dimensionality of the shape space is high, and the available number of samples is small, the resulting matrix may be rank deficient, implying that some eigenvalues have magnitude close to zero, and that the corresponding eigenvectors are arbitrary. This limitation encourages the introduction of more robust covariance matrix estimators [6–10]. Most common is to use a simple form of Tikhonov regularization [1], where nonzero values are added to the diagonal elements of the covariance matrix, e.g.

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{reg}} = \hat{\boldsymbol{\Sigma}} + \lambda \boldsymbol{I}, \tag{5}$$

where λ is a positive regularization parameter. This regularization parameter is not likely to be known in advance, and finding its optimal value can be a cumbersome and computationally heavy task [10]. The analytical methods for covariance estimation [6–9] propose several methods to estimate the regularization parameter using just the sample covariance matrix. However, these methods rely on the estimation of a specific cost function.

4 Gaussian prior regularization

As already showed in [11], starting from the definition of normal distribution in covariance matrix space, and using the Bayes theorem, it is possible to define a covariance matrix as a random variable:

$$p(\boldsymbol{\Sigma}) = (2\pi s^2)^{-\frac{n^2}{2}} \exp\left(-\frac{\|\boldsymbol{\Sigma} - \boldsymbol{B}\|_2^2}{2s^2}\right), \tag{6}$$

where $\boldsymbol{B}$ and s represent the mean and variance of the covariance matrix. The Gaussian prior should ideally be imposed on the space of symmetric, positive definite matrices, as e.g. approximated by the exponential of the norm of the differences of logarithms of matrices [12], but this makes the equation prohibitively complicated, hence we consider (6) as an approximation. The MAP estimate of (6) is found to be a system of third degree polynomials [11], and we seek the solution by the following iterative scheme, estimating the covariance matrix as shown in [11] as

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \frac{1}{n}\left(\boldsymbol{X}\boldsymbol{X}^T - \frac{2\hat{\boldsymbol{\Sigma}}_t(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}} - \boldsymbol{B})^T\hat{\boldsymbol{\Sigma}}_t}{s^2}\right), \tag{7}$$

which we have found to converge for $t \to \infty$, when starting at

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{0}} = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T. \tag{8}$$

Following [13], in case of landmark shapes as the data set, we find it useful to define the mean $\boldsymbol{B}$ of the prior distribution of $\hat{\boldsymbol{\Sigma}}$ in the following way:

$$\boldsymbol{B}_{ij} = \exp\left(-\boldsymbol{A}_{\boldsymbol{ij}}\right), \tag{9}$$

where $\boldsymbol{A_{ij}} = \left\| E[\boldsymbol{p}_i] - E[\boldsymbol{p}_j] \right\|_2$, $E[\boldsymbol{p}_i]$ being the i'th component the mean of the aligned training examples.

In case of 3D shapes, such as an m-rep model of cartilage, equation (9) can be extended as

$$\boldsymbol{B_{ij}} = \exp\left(-\boldsymbol{A_{ij}}\right) \exp\left(-\boldsymbol{T_{ij}}\right), \tag{10}$$

where $\boldsymbol{T}$ is the matrix with all the angular distance of the normals of the atom points. Since the matrices defined in equations (9) and (10) consider the relationship between variables, are symmetric, by construction their eigenvalues are positive or equal to zero; we can assume that they are covariance matrices.

5 Wishart prior regularization

A shape representation based on a Wishart distribution priors assume independence between points, hence it is a prior of spatial noise and not shape variation where probably a Gaussian prior is more suitable.

Consider shapes consisting of points,

$$\boldsymbol{x} = [\boldsymbol{p}_1^T, \boldsymbol{p}_2^T, \ldots, \boldsymbol{p}_p^T]^T \tag{11}$$

with $\boldsymbol{x} \in \mathbb{R}^d$, a collection of shapes,

$$\boldsymbol{X} = [(\boldsymbol{x}_1 - \ \boldsymbol{\mu})|(\boldsymbol{x}_2 - \ \boldsymbol{\mu})|\ldots|(\boldsymbol{x}_n - \boldsymbol{\mu})], \tag{12}$$

as a $d \times n$ matrix of reals, ignore the stochastic dependency on the mean, and write

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^T. \tag{13}$$

When the landmarks $\boldsymbol{p}_i - E[\boldsymbol{p}_i]$ are independently and normal distributed as $N(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, then $\boldsymbol{S}$ is distributed according to the Wishart distribution [5, Chapter 7],

$$p(\boldsymbol{S}|\boldsymbol{\Sigma}, n) = \frac{|\boldsymbol{S}|^{(n-d-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\right)}{2^{nd/2}\,|\boldsymbol{\Sigma}|^{n/2}\,\Gamma_d\left(\frac{n}{2}\right)}, \tag{14}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 & \ldots & 0 \\ 0 & \boldsymbol{\Sigma}_2 & & \\ \vdots & & \ddots & \\ 0 & & & \boldsymbol{\Sigma}_p \end{bmatrix}, \tag{15}$$

and Γ_d is the multivariate Gamma function,

$$\Gamma_d(n) = \pi^{d(d-1)/4} \prod_{i=1}^{d} \Gamma\left(n - \frac{1}{2}(i-1)\right). \tag{16}$$

Using the Bayes theorem, we can write

$$p(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, m) = \frac{|\boldsymbol{\Psi}|^{m/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\right)\right)}{2^{md/2}\,|\boldsymbol{\Sigma}|^{(m+d+1)/2}\,\Gamma_d\left(\frac{m}{2}\right)}, \tag{17}$$

where $\boldsymbol{\Psi}$ and m are fixed as parameters of the density. The inverted Wishart density originates as the density of $\boldsymbol{S}^{-1}$. Since the evidence is independent on $\boldsymbol{\Sigma}$, we find the MAP estimate as:

$$\hat{\boldsymbol{\Sigma}} = \arg\max_{\boldsymbol{\Sigma}} p(\boldsymbol{S}|\boldsymbol{\Sigma}, n)p(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, m) \tag{18a}$$

$$= \arg\max_{\boldsymbol{\Sigma}} \frac{c\,|\boldsymbol{\Psi}|^{m/2}\,\exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{S}+\boldsymbol{\Psi}\right)\boldsymbol{\Sigma}^{-1}\right)\right)}{|\boldsymbol{\Sigma}|^{(n+m+d+1)/2}} \tag{18b}$$

At this point, we can consider $\boldsymbol{\Psi}$ as independent of $\boldsymbol{\Sigma}$, the solution to (18) is found by differentiation as:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n+m+d+1}\left(\boldsymbol{S}+\boldsymbol{\Psi}\right). \tag{19}$$

Or we can include $\boldsymbol{\Psi} = s^2\boldsymbol{\Sigma}^{-1}$ in the estimation, in which case the maximization results in system of quadratic equations

$$\hat{\boldsymbol{\Sigma}}^2 = \frac{1}{n+2m+d+1}\left(\boldsymbol{S}\hat{\boldsymbol{\Sigma}} + 2s^2\boldsymbol{I}\right). \tag{20}$$

We solve this iteratively as

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \hat{\boldsymbol{\Sigma}}_t - \delta\left(\hat{\boldsymbol{\Sigma}}_t^2 - \frac{1}{n+2m+d+1}\left(\boldsymbol{S}\hat{\boldsymbol{\Sigma}}_t + 2s^2\boldsymbol{I}\right)\right), \tag{21}$$

where δ is a sufficiently small constant to avoid divergence, n and d are respectively still the number of samples and dimension, m is the only user-specified parameter that defines the density of the Wishart distribution and in our experiments was set to 40 for vertebrae and 50 for cartilages.

We call this approach MAP-PCA, and the prior described in (7) as *Normal prior*, while the equation (19) define the *Inverted Wishart* (IWIS) prior, and the equation (20) the *Uncommitted Inverted Wishart* (UIWIS) prior.

6 Experiments

Using high-resolution (full boundary) vertebral shapes from radiographs may lead to reliable results when detecting osteoporotic fragility fractures [14]. However, manual annotation of full boundaries is time consuming. For studies on osteoarthritis, detailed m-rep shape models of cartilage also lead to accurate results. In this case the problem of building a high-resolution model is, however, the computational cost during the automatic segmentation.

In the following sections we give examples of high-resolution shape models reconstructed from the coarse annotations, using the MAP-PCA estimation with the three priors described above, and compare to the result obtained using MLE and Tikhonov estimation.

Given $\boldsymbol{y}$ be an incomplete or lower dimensional shape vector of dimensionality $l < d$, we can obtain the corresponding higher resolution shape $\boldsymbol{x} \in \mathbb{R}^d$ using a linear mapping $\boldsymbol{L} : \mathbb{R}^d \rightarrow \mathbb{R}^l$:

$$\boldsymbol{y} = \boldsymbol{L}\boldsymbol{x}, \tag{22}$$

For our experiments the matrix $\boldsymbol{L}$ is a sampling matrix connected identical points between $\boldsymbol{x}$ and $\boldsymbol{y}$. Since the system is overdetermined, the solution is easily not uniquely defined. In [15], $\boldsymbol{x}$ is obtained minimizing the functional

$$E(\boldsymbol{x}) = \|\boldsymbol{L}\boldsymbol{x} - \boldsymbol{y}\|_2^2 . \tag{23}$$

Since $\boldsymbol{x}$ belongs to the shape model (2), the functional transforms to:

$$E(\boldsymbol{b}) = \|\boldsymbol{Q}\boldsymbol{b} - \boldsymbol{y}\|_2^2 , \tag{24}$$

where $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{V}_t\boldsymbol{\Lambda}_t$, and $\boldsymbol{V}_t$ are the eigenvectors corresponding to the t principal eigenmodes of the covariance matrix estimated with one of the method described in the previous sections. It can be shown that $E(\boldsymbol{b})$ is minimized by $\boldsymbol{b}^* = \boldsymbol{Q}^+\boldsymbol{y}$ [15], where $\boldsymbol{Q}^+$ is the Moore-Penrose pseudo-inverse [3]. Hence $\boldsymbol{x}$ is estimated by

$$\tilde{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{V}_t\boldsymbol{\Lambda}_t\boldsymbol{b}^* . \tag{25}$$

Experiments are based on the reconstruction of incomplete data and are divided into two groups: vertebrae shapes and cartilage shapes. The eigenvalues and eigenvectors of the covariance matrix computed from the training set through the MLE and the Bayesian methods.

The reconstruction error of vertebra and cartilage shapes between a high resolution shape and a reconstructed version of the same from a lower dimensional version is computed for all the p points of the boundary using the expression:

$$E_{full} = \frac{1}{n}\sum_{i=1}^{p} \left\|\boldsymbol{p}_{i,reconst.} - \boldsymbol{p}_{i,orig.}\right\|_2 \tag{26}$$

The performances of MAP-PCA and MLE methods are compared using mean reconstruction error over all the test shapes for different number of principal eigenmodes.

6.1 Reconstruction of vertebra shapes

For clinical studies on osteoporosis for fracture quantification, we use a 6 points representation of a vertebra, due to the fact that three points in the lower border of the vertebra and 3 points in the upper part of the vertebra can describe the heights measure defined by Genant [16] that is the gold standard for fracture quantification. In order to perform more sophisticated shape analysis, a full contour is needed. Therefore, an extrapolation of the full contour from the six points is a useful initialization for a segmentation algorithm like [17]. During the experiments, the shapes in the training set were the full boundaries made of 52 points, while for the test shapes only 6 out of these 52 points were used. Using the shape model, (25), the corresponding high-resolution vertebral boundary was reconstructed from a low-dimensional test shape.

6.2 Reconstruction of cartilage shapes

We tested the improvement of our covariance estimation also for a medial atoms representation of tibial knee-cartilage, in order to produce a higher resolution m-rep model. We have performed this evaluation by removing atoms from the models and by measuring, how well the interpolation allows for reconstruction of the original model. The cartilage data set is composed of 620 knee MRI scans from 159 subjects including both left and right knees and baseline and follow-up scans from a longitudinal 21-month study. The dimensions of the scans are 256 x 256 pixels with around 110 slices. The population includes healthy and diseased knees with varying degree of OA from both men and women at ages from 21 to 78. The test sets were made to be from 10 and 20 samples, and the relative training sets to be the remaining 610 and 600 samples in a leave-one patient-out fashion.

For each knee, we have a three-dimensional m-rep model of the medial tibial cartilage compartment estimated from a fully automatic segmentation [18, 19]. Figure 1 illustrates a cartilage shape model is illustrated for a cropped knee MRI. We produced a low-resolution lattice by removing nodes from the m-rep

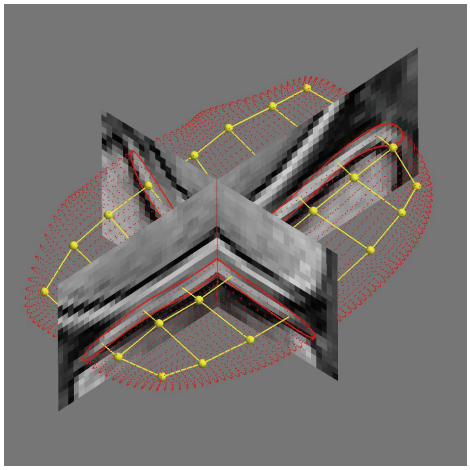


Fig. 1. Sagittal slice from Turb3D T1 MRI. The contours are the manual outlines of femoral and tibial cartilage performed by a radiologist.

representation of the cartilage, and calculated the mean reconstruction error between the original and the reconstructed 3D shapes using (26). To represent a medial lattice for a cartilage sheet, $p = 32$ points were used, hence the dimension is $d = 96$. During the experiments 24 points were removed randomly, and only 8 were left, so the reduced dimensions is $l = 24$.

7 Results

A typical vertebral shape of 52 points (104 dimensions) is depicted in Figure 2(a), while the reduced and reconstructed shape is depicted in Figure 2(b). Figure 2(c)

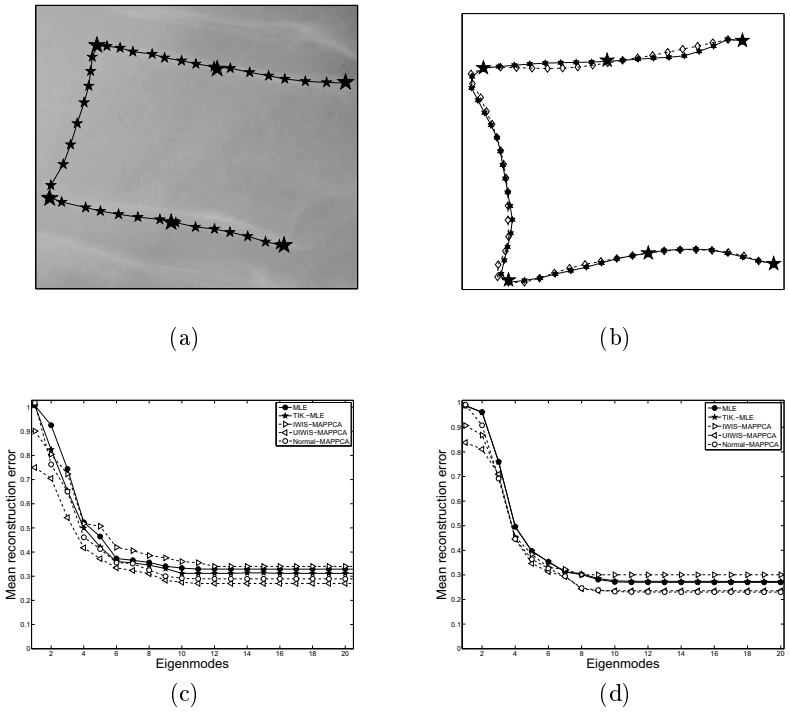


Fig. 2. (a) An image of a vertebra with the shape annotation. (b) An original (continuous line) and reconstructed (dashed line) shape annotation. The shape's 52 points are reconstructed using only the 6 points depicted as the big stars. The vertebra reconstruction error for different number of training shapes: (c) 20 and (d) 40. Here the results with the IWIS and UIWIS are for noisy test shapes with $s = 1$ and $m = 40$ as parameters of the equation (19) and (20). Instead the Normal prior uses $s = 2$ and the Tikhonov regularization parameter is 0.4665. Due to no further changes of the curves, only the first 20 eigenmodes are depicted.

and 2(d) shows the different mean reconstruction errors obtained using (26) for two different sample sizes and when varying t, the number of eigenvalues included in the reconstruction. The vertebra experiments demonstrate that the MAP-PCA method generally improves the reconstruction especially for few samples and a small number of eigenvalues. It seems the Wishart prior performs relatively better than the Normal prior.

For m-rep modeling of tibial knee-cartilage, the result using the first 20 eigenmodes, is shown in Figure 3(a)(b)(c). The mean reconstruction error using the discussed covariance estimators is shown in Figure 3(d)(e). The cartilage shapes show a considerable improvement, when using the Normal prior, and an almost as good improvement, when using the Uncommitted Inverted Wishart prior when using only 10 training shapes.

8 Future experiments

Doing further investigation about regression, we also obtained early good results with a conditional shape model for finding the aorta given the spine. Where the spine model is a collection of the 6 points representation of four lumbar vertebrae, while the aorta model is a sequence of aorta-wall points. For each wall there are 9 points equidistantly sampled along aorta boundaries, hence no sliding along boundaries is necessary to model. An aorta shape can be predicted using an equation similar to (25) where the unknown points of the previous experiments are the aorta points $\mathbf{x_1}$: Figure 4 illustrates the typical result of our method with the lumbar vertebrae (thin continuous line), the annotated aorta (thick continuous line) and the predicted aorta (dashed line).

$$\mathbf{x_1} = \mu_1 + \mathbf{\Sigma_{12}}\mathbf{\Sigma_{22}^{-1}}(\mathbf{S_2} - \mu_2) \tag{27}$$

and where μ_1 and μ_2 are the mean shape of all the aortas and the mean shape of all the spines respectively. $\mathbf{\Sigma_{11}}$ represents the covariance of the aortas, $\mathbf{\Sigma_{22}}$ the covariance of spines, and $\mathbf{\Sigma_{12}}$ and $\mathbf{\Sigma_{21}}$ are the cross-covariance. Here the need is to regularize the covariance matrix $\mathbf{\Sigma_{22}}$ of vertebra.

9 Conclusions

Efficient estimation of covariance matrices is an important task for statistical shape analysis. In this paper we discuss a novel method called MAP-PCA for estimate the covariance matrix in case of small sample size. The matrix obtained with MAP-PCA generally outperforms the traditional MLE method with and without Thikhonov regularization. In addition, the experiments show that the choice of a suitable prior leads to better results. In particular, the Wishart priors assume statistically independent points, and therefore it performs best for variations due to noise and not shape variation. The Normal prior is not limited to zero off-block-diagonal elements and may be used to steer the estimate towards preferred shape variations. Our conclusion is that the choice of prior is related to various factors, by the number of samples available to the shape variance, and that this method is a valid substitute to the Thikhonov regularization where the result is based on the search of the optimal mixing parameter.

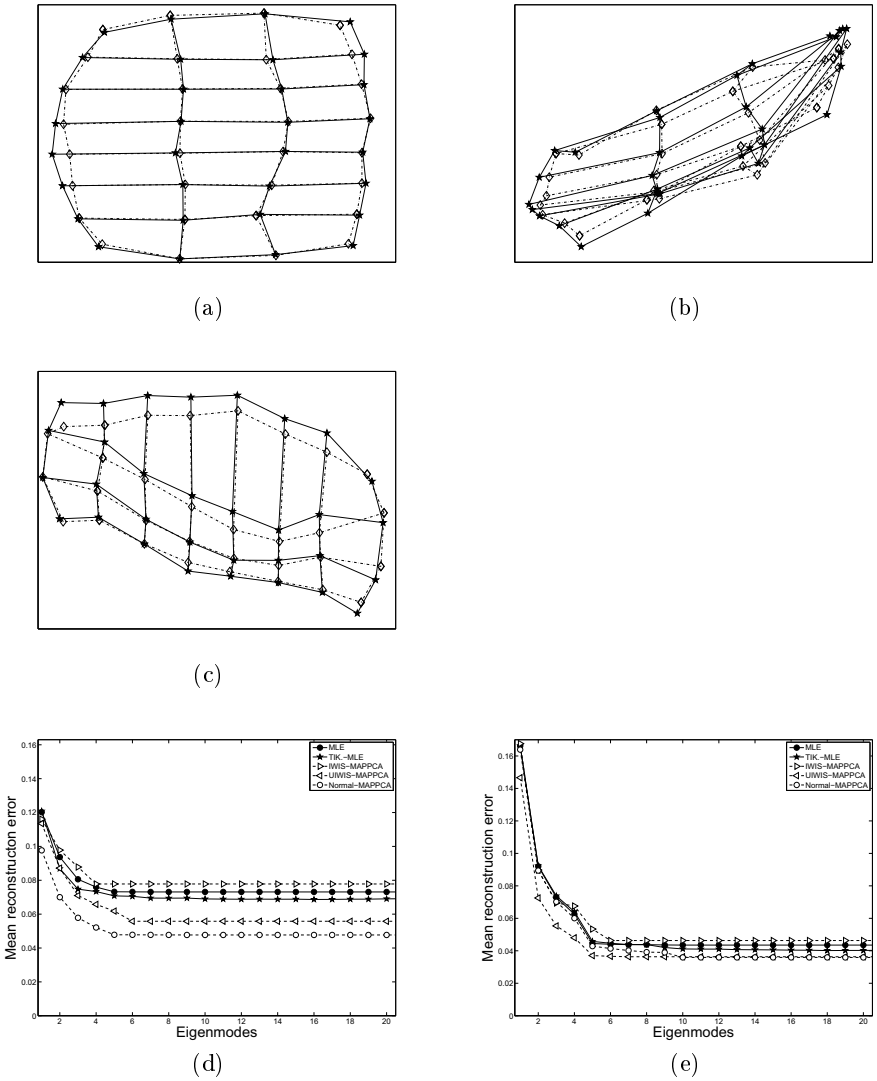


Fig. 3. Overlay of the original and the reconstructed cartilage shape: (a) projection on the plane XY, (b) projection on the plane XZ, and (c) projection on the plane YZ. The continuous line is the original shape and the dashed one is the reconstruction. Here the removed points are chosen randomly. The reconstruction error for different number of training shapes: (d) 10 and (e) 20. Here the results are for noisy test shapes cf. Figure 3, and using $s = 1$ and $m = 50$ in the equation (20). Instead the Normal prior uses $s = 4$ and the Tikhonov regularization parameter is 0.6443. Due to no further changes of the curves, only the first 20 eigenmodes are depicted.

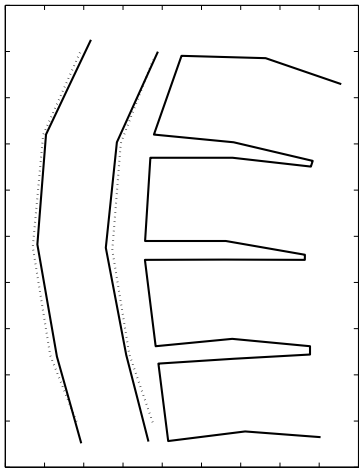


Fig. 4. The typical result of our method: The lumbar vertebrae represented by 6 points(continuous line), the annotated aorta(continuous line) and the predicted aorta (dashed line).

Acknowledgement

We would like to acknowledge the Center for Clinical and Basic Research, Ballerup, Denmark, for providing the annotated vertebra radiographies and the MRI knee scans.

References

1. Tikhonov, A.N., Arsenin, A.N.: Solution of Ill-posed Problems. first edn. Winston & Sons, Washington (1977)
2. Cooper, D., Cootes, T., Taylor, C., Graham, J.: Active shape models, their training and application. Computer Vision and Image Understanding **61** (1995) 38–59
3. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes 3rd Edition: The Art of Scientific Computing. 3rd edn. Cambridge University Press (2007)
4. Bookstein, F.L.: Shape and the information in medical images: A decade of morphometric synthesis. cviu **66**(2) (1997) 97–118
5. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis. 3rd edn. Wiley (2003)
6. Schaefer, J., Strimmer, K.: A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Journal of Statitstical Applications in Genetics and Molecular Biology **4** (2005)

7. Haff, L.R.: Empirical bayes estimation of the multivariate normal covariance matrix. Annals of statistics **8** (1980) 586–597
8. James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the fourth Berkely symposium on mathematical statistics and probability. Volume 1. (1961) 361–379
9. Efron, B., Morris, C.: Multivariate empirical bayes and estimation of covariance matrices. Annals of statistics **4** (1976)
10. Friedman, J.H.: Regularized discriminant analysis. Journal of American Statistics Association **84** (1989) 165–175
11. Crimi, A., Ghosh, A., Sporring, J., Nielsen, M.: Bayes estimation of shape-models with application to vertebrae boundaries. In: SPIE Medical Imaging. (2009)
12. Fillard, P., Pennec, X., Arsigny, V., Ayache, N.: Clinical dt-mri estimation, smoothing, and fiber tracking with log-euclidean metrics. IEEE Trans. Med. Imaging **26**(11) (2008) 1472–1482
13. Sporring, J., Jensen, K.H.: Bayes reconstruction of missing teeth. Journal of Mathematical Imaging and Vision **31** (2008) 245–254
14. de Bruijne M.: Shape particle filtering for image segmentation. In: MICCAI, Lecture Notes in Computer Science. (2004) 186–175
15. Blanz, V., Vetter, T.: Reconstructing the complete 3d shape of faces from partial information. Technical Report report of Computer Graphics no. 1, University of Freiburg (2001)
16. Genant, H.K., Jergas, M., Palermo, L., Nevitt, M., Valentin, R.S., Black, D., Cummings, S.R.: Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures. J Bone Miner Res. **11** (1996) 984–996
17. Iglesias, J., de Bruijne, M.: Semi-automatic segmentation of vertebrae in lateral x-rays using a conditional shape model. Academic Radiology **14** (2007) 1156–1165
18. Dam, E.B., Fletcher, P.T., Pizer, S.M.: Automatic shape model building based on principal geodesic analysis bootstrapping. Medical Image Analysis **12**(2) (2008) 136–151
19. Folkesson, J., Dam, E.B., Olsen, O.F., Pettersen, P., Christiansen, C.: Segmenting articular cartilage automatically using a voxel classification approach. IEEE Transactions on Medical Imaging **26**(1) (2007) 106–115

Parametric estimation for Gaussian operator scaling random fields and anisotropy analysis of bone radiograph textures

Hermine Biermé[1], Claude-Laurent Benhamou[2], and Frédéric Richard[1]

1. University Paris Descartes, MAP5 CNRS UMR 8145
[hermine.bierme,frederic.richard]@parisdescartes.fr,
2. INSERM U658, Hospital of Orleans, France. ⋆

Abstract. In this paper, we consider a stochastic anisotropic model for trabecular bone x-ray images. In [1], a fractal analysis based on isotropic Fractional Brownian Fields was proposed to characterize bone microarchitecture. However anisotropy measurement is of special interest for the diagnosis of osteoporosis [7]. We propose to model trabecular bone radiographs by operator scaling Gaussian random fields which are anisotropic generalizations of the Fractional Brownian Field. We construct consistent estimators for these models and apply them on trabecular bone x-ray images. Our first results suggest that these models are relevant for this modeling.

1 Introduction

Texture analysis is a challenging issue of Image Processing, which is often raised in medical applications. There are several types of texture approaches. Among stochastic approaches, fractal analysis has been largely used in medical applications [1, 7, 8, 10]. The stochastic model beyond fractal analysis is the fractional Brownian field (FBF) which is a multi-dimensional extension of the famous fractional Brownian motion implicitly introduced in [17] and defined in [19]. This field is mathematically defined as the unique centered Gaussian field, null at 0 almost surely, with stationary increments, isotropic, and self-similar of order $H \in (0, 1)$. Its variogram is of the form $v(x) = C_H|x|^{2H}, \forall x \in \mathbb{R}^2$, with $|\cdot|$ the Euclidean norm. Parameter H, called the Hurst index, is a fundamental parameter which is an indicator of texture roughness and is directly related to the fractal dimension of the graph sample paths.

FBF was used for the characterization and classification of mammogram density [8], the study of lesion detectability in mammogram textures [12], and the assessment of breast cancer risk [8, 13]. Fractal analysis has also been used for the radiographic characterization of bone architecture and the evaluation of osteoporotic fracture risk [1]. However, it is well-established that the anisotropy

⋆ We are grateful to the ANR french agency for the financial support to the project ANR MATAIM NT09_441552.

of the bone is an important predictor of fracture risk [7]. Hence fractal analysis with fractional Brownian fields (FBF), which are isotropic by definition, is not completely satisfactory for this medical application. In this paper, our aim is to propose a suitable model which accounts properly for the anisotropy of bone radiograph textures.

The study of random field anisotropy is a wide field of research in the Probability Theory. It covers numerous open issues related to the definition and the analysis of anisotropy, the estimation of anisotropic model parameters, and the simulation of anisotropic fields [4, 3, 6, 11, 20]. In [6], A. Bonami and A. Estrade set a generic framework in which it is possible to define numerous types of anisotropic fields. This framework gathers centered Gaussian fields with stationary increments $\{X(x)\,;\, x \in \mathbb{R}^d\}$, null at 0 almost surely, whose variogram v is characterized by a positive even measurable function f satisfying the relation

$$\forall\, x \in \mathbb{R}^d, v(x) = \mathbb{E}\left(X(x)^2\right) = \int_{\mathbb{R}^d} \left|e^{ix\cdot\zeta} - 1\right|^2 f(\zeta) d\zeta \tag{1}$$

and the condition $\int_{\mathbb{R}^d} \left(1 \wedge |\zeta|^2\right) f(\zeta) d\zeta < \infty$. Within this framework, a field is isotropic whenever the so-called spectral density f of the field is radial, and anisotropic when f depends on the direction $\arg(\zeta)$ of ζ.

In [4, 5], we studied 2-dimensional Gaussian fields with spectral density of the form

$$\forall\, \zeta \in \mathbb{R}^2, f(\zeta) = |\zeta|^{-2h(\arg(\zeta))-2}, \tag{2}$$

where h is a measurable π-periodic function with range $[H, M] \subset (0, 1)$ where $H = \operatorname{essinf}_{[-\pi,\pi)} h$ and $M = \operatorname{esssup}_{[-\pi,\pi)} h$. These fields extend FBF, which are obtained when the function h is almost everywhere constant and equal to the Hurst index H. When h is not constant, the function h depends on the orientation and, consequently, the corresponding field is anisotropic. Another extension of FBF, called operator scaling Gaussian random fields [3], can be obtained by taking spectral densities of the form

$$\forall c > 0,\ \ \forall \zeta \in \mathbb{R}^2\,,\ \ f(c^E \zeta) = c^{-2-\operatorname{tr}(E)} f(\zeta), \tag{3}$$

for some real 2×2 matrix E. The spectral density of an FBF with Hurst index $H \in (0, 1)$, given by $f(\zeta) = |\zeta|^{-2H-2}$, for $\zeta \in \mathbb{R}^2$, satisfies (3) for $E = I_2/H$ with I_2 the identity matrix. In such a model, the anisotropy is characterized by the 2×2 parameters of the matrix E.

In this paper, we focus on a particular class of 2-dimensional operator scaling field for which E is a diagonal matrix. More precisely, we consider 2 dimensional Gaussian fields with spectral density of the form

$$\forall\, \zeta = (\zeta_1, \zeta_2) \in \mathbb{R}^2, f(\zeta) = \left(\zeta_1^2 + \zeta_2^{2a}\right)^{-\beta}, \tag{4}$$

where $\beta = H_1 + (1 + 1/a)/2$ and $a = H_2/H_1$ for some $0 < H_1 \leq H_2 < 1$. Then f satisfies (3) for $E = \operatorname{diag}(1/H_1, 1/H_2)$ and 2 dimensional FBFs are obtained when $H_2 = H_1$, which gives the Hurst index. When $H_1 \neq H_2$ the corresponding field is anisotropic.

There are several ways to analyze the anisotropy of a field. One simple way consists of characterizing the Hölder regularity of rectrictions of the field along oriented lines. However, it was shown that the directional regularity (the regularity of these line restrictions) of any 2-parameter Gaussian random field obtained from (1) is constant, except in at most one direction where it can be larger [11]. In particular, the directional regularity of field model (2) is the same whatever the direction. Hence, the anisotropy of such a model cannot be characterized using line restrictions. For such a model, one can rather study the regularity of windowed Radon transforms [6].

In this paper, we show that fields defined by (4) can have a privileged direction where line restrictions are more regular than in other directions. We also propose some techniques for the estimation of parameters H_1 and H_2. Estimators are constructed using line restrictions and following principles of generalized quadratic variations. Finally, adapting results shown in [4], we prove the convergence of these estimators.

In collaboration with L. Benhamou and M. Rachidi (INSERM U658, Orleans, France) [5], we studied trabecular bone x-ray images. After some preliminary experiments, we came to the conclusion that model (2) was not suitable for the modeling of these images. Due to trabecular structures of the bone, these images have a privileged direction which is detectable from the analysis of line regularity. Such a situation is analogous to the one of sedimentary aquifers whose scaling properties vary according to directions and which were successfully modeled by operator scaling fields [2]. In this paper, we present some preliminary experiments suggesting the adequacy of model (4) to bone radiograph textures.

In Section 2, we recall main properties of operator scaling fields and construct consistent estimators for H_1 and H_2. In Section 3, we present results of estimation on trabecular bone x-ray images, which suggest adequacy to this modeling.

2 Main properties

Let X be a Gaussian field with spectral density (GFSD) on $\mathbb{R}^2$ given by (4) for $0 < H_1 \leq H_2 < 1$. Let us denote E the diagonal 2×2 matrix $E = \mathrm{diag}(1/H_1, 1/H_2)$. We note $q = \mathrm{tr}(E)$ and remark that

$$q = 1/H_1 + 1/H_2 = (1 + 1/a)/H_1,$$

with $a = H_2/H_1 \geq 1$ such that $\beta = H_1\,(1 + q/2)$.

2.1 Operator scaling property

Let us define $\psi(\zeta) = \left(\zeta_1^2 + \zeta_2^{2a}\right)^{H_1/2}$, for $\zeta \in \mathbb{R}^2$, such that ψ is continuous on $\mathbb{R}^2$ with positive values on $\mathbb{R}^d \setminus \{0\}$ and satisfies $\psi(0) = 0$ and $\psi(c^E\zeta) = c\psi(\zeta)$ for all $c > 0$, where c^E is the exponential matrix $c^E = \sum_{n\in\mathbb{N}} \frac{\ln(c)^n}{n!} E^n$. According

to Theorem 4. 1 of [3], the random field X is a Harmonizable operator scaling Gaussian field with respect to E (since $E = E^t$):

$$\forall c > 0, \ \left\{X(c^E x); x \in \mathbb{R}^2\right\} \stackrel{fdd}{=} \left\{cX(x); x \in \mathbb{R}^2\right\},$$

where $\stackrel{fdd}{=}$ means equality for finite dimensional distributions. The operator scaling property is an anisotropic generalization of the well-known self-similarity property.
In particular, for $H_1 = H_2$, the random field X is self-similar of order $H = \min(H_1, H_2)$, which means that

$$\forall c > 0, \ \left\{X(cx); x \in \mathbb{R}^2\right\} \stackrel{fdd}{=} \left\{c^H X(x); x \in \mathbb{R}^2\right\}, \forall c > 0.$$

Moreover, in this case, the spectral density is a radial function, which implies that X is isotropic. Being Gaussian, with stationary increments, null at point zero almost surely, self-similar of order H and isotropic is enough to conclude that X is the famous FBF of Hurst index H. Then, any restriction along straight lines

$$X_{\theta,x_0} = \{X(x_0 + t\theta) - X(x_0) \ ; \ t \in \mathbb{R}\}, \tag{5}$$

for a point $x_0 \in \mathbb{R}^2$ and a unit vector $\theta = (\theta_1, \theta_2)$, will also be a fractional Brownian motion (1 dimensional process) of index H.
When $H_1 \neq H_2$, the stationarity of increments and the operator scaling property with respect to the diagonal matrix E lead to the fact that for any $x_0 \in \mathbb{R}^2$, processes X_{θ,x_0} are fractional Brownian motion of index H_1 when $\theta_2 = 0$ and H_2 when $\theta_1 = 0$. Note that in any other direction θ with $\theta_1 \neq 0$ and $\theta_2 \neq 0$, processes X_{θ,x_0} are not self-similar. Therefore self-similarity parameters are too restrictive to characterize those processes. However, these parameters considered at small scales, are closely linked to Hölder regularity index as we will see in the next section.

2.2 Regularity

Using Kolmogorov-Centsov criterion [15], one can prove that $H = \min(H_1, H_2)$ is the critical Hölder exponent of X. This means that for any $\alpha \in (0, H)$, sample paths of X satisfy a uniform Hölder condition of order α on $[-T, T]^d$, for any $T > 0$: there exists a positive random variable A with $\mathbb{P}(A < +\infty) = 1$ such that

$$\forall x, y \in [-T, T]^d, \ \ |X(x) - X(y)| \leq A|x - y|^\alpha; \tag{6}$$

while for any $\alpha \in (H, 1)$, almost surely the sample paths of X fail to satisfy any uniform Hölder condition of order α. We refer to Theorem 5.4 of [3] for the proof of this result. Actually, global Hölder regularity H does not capture the anisotropy of the field. Therefore one can study regularity properties of the field along straight lines, considering critical Hölder exponent of processes defined by (5). This will provide some additional directional regularity information. Note that when X has stationary increments, the Hölder regularity of the process X_{θ,x_0} will not depend on point $x_0 \in \mathbb{R}^2$ so one only has to study the regularity of $\{X(t\theta) \, ; \, t \in \mathbb{R}\}$ for all directions θ. Let us recall Definition 6 of [6].

Definition 1. *Let $\{X(x); x \in \mathbb{R}^d\}$ with stationary increments and let θ be any direction of the unit sphere. If the process $\{X(t\theta); t \in \mathbb{R}\}$ has Hölder critical exponent $\gamma(\theta)$ we say that X admits $\gamma(\theta)$ as directional regularity in direction θ.*

Note that if X is a GFSD given by f, for any direction θ, the process $\{X(t\theta); t \in \mathbb{R}\}$ is still a Gaussian process with spectral density given by the Radon transform of f, namely,

$$\forall p \in \mathbb{R},\ T_\theta f(p) = \int_{\mathbb{R}} f(p\theta + s\theta^\perp) ds, \tag{7}$$

where $(\theta, \theta^\perp)$ is an orthonormal basis of $\mathbb{R}^2$. It is well known that the asymptotic behavior of the spectral density determines the Hölder regularity of the process, as we recall in the sequel. Let us first introduce some notations. For any $H \in (0,1)$, we note $f(\xi) \asymp_{+\infty} |\xi|^{-2H-1}$, if f satisfies: for any $\varepsilon > 0$ there exists $A, B_1, B_2 > 0$ such that for almost all $\xi \in \mathbb{R}$,

$$|\xi| \geq A \ \Rightarrow\ B_2 |\xi|^{-2H-1-\varepsilon} \leq f(\xi) \leq B_1 |\xi|^{-2H-1+\varepsilon}. \tag{8}$$

Remark that $|\xi|^{-2H-1}$ is, up to a constant, the spectral density of a fractional Brownian motion of Hurst index H. In the same vein, for any $H \in (0,1)$, we note $v(y) \asymp_0 |y|^{2H}$, if v satisfies: for any $\varepsilon > 0$ there exists $\delta, C_1, C_2 > 0$ such that for all $y \in \mathbb{R}$,

$$|y| \leq \delta \ \Rightarrow\ C_2 |y|^{2H+\varepsilon} \leq v(y) \leq C_1 |\xi|^{2H-\varepsilon}. \tag{9}$$

We recall here results proved in [6].

Theorem 1. *Let X be a Gaussian process with spectral density f and variogram v. Let $H \in (0,1)$.*
(a) If $f(\xi) \asymp_{+\infty} |\xi|^{-2H-1}$ then $v(y) \asymp_0 |y|^{2H}$.
(b) If $v(y) \asymp_0 |y|^{2H}$ then X admits H as critical Hölder exponent.

In [4] we prove that better estimates on the spectral density enable to give consistent estimators for H. Therefore we give stronger results for spectral densities $T_\theta f$ of line processes $\{X(t\theta)\,;\, t \in \mathbb{R}\}$.

Theorem 2. *Let f be a spectral density given by (4). Let $\theta = (\theta_1, \theta_2)$ be a unit vector of $\mathbb{R}^2$.*
(a) If $\theta_1 \neq 0$ and $\theta_2 \neq 0$, then

$$T_\theta f(p) = |p|^{-2H_1-1} \left(\int_{\mathbb{R}} (s^{2a} + \theta_1^{-2})^{-\beta} ds \right) / |\theta_1| + O_{|p|\to+\infty}\left(|p|^{-2H_1-1-(1-1/a)} \right).$$

(b) If $\theta_2 = 0$, then $T_\theta f(p) = |p|^{-2H_1-1} \left(\int_{\mathbb{R}} (s^{2a}+1)^{-\beta} ds \right)$.
(c) If $\theta_1 = 0$, then $T_\theta f(p) = |p|^{-2H_2-1} \left(\int_{\mathbb{R}} (s^{2}+1)^{-\beta} ds \right)$.

Proof. Let $\theta = (\theta_1, \theta_2)$, then one can choose $\theta^\perp = (\theta_2, -\theta_1)$ such that

$$T_\theta f(p) = \int_{\mathbb{R}} \left((p\theta_1 + s\theta_2)^2 + (p\theta_2 - s\theta_1)^{2a} \right)^{-\beta} ds.$$

Let us assume that $\theta_1 \neq 0$ and let the change of variables $u = s\theta_1 - p\theta_2$, then

$$T_\theta f(p) = \frac{1}{|\theta_1|} \int_{\mathbb{R}} \left((p/\theta_1 + u\theta_2/\theta_1)^2 + u^{2a} \right)^{-\beta} du. \tag{10}$$

Since $T_\theta f$ is an even function one can assume that $p > 0$ and let the change of variables $u = p^{1/a} s$ such that

$$T_\theta f(p) = \frac{1}{|\theta_1|} p^{1/a - 2\beta} \int_{\mathbb{R}} \left(\left(1/\theta_1 + p^{-(1-1/a)} s\theta_2/\theta_1 \right)^2 + s^{2a} \right)^{-\beta} ds.$$

This concludes for the proof when $\theta_2 = 0$ (and $|\theta_1| = 1$), since $1/a - 2\beta = -1 - 2H_1$. Otherwise, let us consider

$$|\theta_1| p^{-1/a + 2\beta} T_\theta f(p) - \int_{\mathbb{R}} \left(1/\theta_1^2 + s^{2a} \right)^{-\beta} ds = \int_{\mathbb{R}} E_\theta f(p, s) ds,$$

where

$$E_\theta f(p, s) = \left(\left(1/\theta_1 + p^{-(1-1/a)} s\theta_2/\theta_1 \right)^2 + s^{2a} \right)^{-\beta} - \left(1/\theta_1^2 + s^{2a} \right)^{-\beta}.$$

Note that

$$\begin{aligned} |E_\theta f(p, s)| &\leq 2\beta p^{-(1-1/a)} |s\theta_2/\theta_1| \int_0^1 \left(\left(1/\theta_1 + t p^{-(1-1/a)} s\theta_2/\theta_1 \right)^2 + s^{2a} \right)^{-\beta - 1/2} dt \\ &\leq 2\beta p^{-(1-1/a)} |s\theta_2/\theta_1| \left(1/4\theta_1^2 + s^{2a} \right)^{-\beta - 1/2} \text{ if } |s| \leq p^{1-1/a}/2|\theta_2| \\ &\leq 2\beta p^{-(1-1/a)} |\theta_2/\theta_1| |s|^{-2a(\beta+1/2)+1} \text{ if } |s| > p^{1-1/a}/2|\theta_2|. \end{aligned}$$

Therefore, choosing $p > |2\theta_2|^{1/(1-1/a)}$, one has

$$\begin{aligned} &\int_{\mathbb{R}} |E_\theta f(p, s)| \, ds \\ &\quad 2 \leq \beta p^{-(1-1/a)} |\theta_2/\theta_1| \left(\int_{\mathbb{R}} |s| \left(1/4\theta_1^2 + s^{2a} \right)^{-\beta - 1/2} ds + \int_{|s|>1} |s|^{-2a(\beta+1/2)+1} ds \right) \\ &= O_{p \to +\infty} \left(p^{-(1-1/a)} \right), \end{aligned}$$

since $2a(\beta + 1/2) > 2a\beta > 2$. Finally, when $\theta_1 = 0$ (and $|\theta_2| = 1$), we have

$$T_\theta f(p) = \int_{\mathbb{R}} \left(s^2 + p^{2a} \right)^{-\beta} ds.$$

Therefore, the change of variables $s = p^a u$ leads to

$$T_\theta f(p) = p^{a - 2a\beta} \int_{\mathbb{R}} \left(u^2 + 1 \right)^{-\beta} du,$$

which concludes the proof since $2a\beta - a = 2aH_1 + 1 = 2H_2 + 1$.

Following Definition 1 and combining Theorems 1 and 2 we obtain the following results as stated in Theorem 5.4 of [3].

Proposition 1 *For any direction $\theta = (\theta_1, \theta_2)$, the random field X admits H_1 for directional regularity in direction θ such that $\theta_1 \neq 0$. When $\theta_1 = 0$, the random field X admits H_2 for directional regularity in direction θ.*

The next section is devoted to the construction of estimators for H_1 and H_2.

2.3 Estimation

Generalized quadratic variations, studied in [14, 16], have been extensively used to estimate the Hurst parameter of a fractional Brownian motion. More generally they allow the estimation of critical Hölder exponents for Gaussian processes or fields. In [4] we give theoretical results of consitency and asymptotic normality for estimators based on generalized quadratic variations under asymptotic development of spectral densities assumptions. These results will be used in the context of this paper. Let us recall principles of these estimations. Let Y be a Gaussian process with stationary increments and a spectral density f. Let

$$\{Y\left(k/N\right); 0 \leq k \leq N\}$$

be an observed sequence. We consider the stationary sequence formed by second-order increments of Y with step $u \in \mathbb{N} \setminus \{0\}$

$$\forall\, p \in \mathbb{Z},\ Z_{N,u}(Y)(p) = Y\left((p+2u)/N\right) - 2Y\left((p+u)/N\right) + Y\left(p/N\right). \tag{11}$$

The generalized quadratic variations of Y of order 2 are then given by

$$V_{N,u}(Y) = \frac{1}{N-2u+1} \sum_{p=0}^{N-2u} \left(Z_{N,u}(Y)(p)\right)^2. \tag{12}$$

Let us quote that

$$\mathbb{E}(V_{N,u}(Y)) = \mathbb{E}((Z_{N,u}(Y)(0))^2) = \mathbb{E}\left(Y\left(\frac{2u}{N}\right) - 2Y\left(\frac{u}{N}\right) + Y(0)\right)^2,$$

According to Proposition 1.1 of [4], when $N \to +\infty$,

$$\mathbb{E}(V_{N,u}(Y)) \sim c_H N^{-2H} u^{2H},$$

for some $c_H > 0$, whenever the spectral density f satisfies $f(\xi) \sim c|\xi|^{-2H-1}$, when $|\xi| \to +\infty$, with $H \in \left(0, \frac{7}{4}\right)$ and $c > 0$. Intuitively, we can thus define an estimator of H as

$$\widehat{H}_{N,u,v} = \frac{1}{2\log(u/v)} \log\left(\frac{V_{N,u}(Y)}{V_{N,v}(Y)}\right). \tag{13}$$

In [14] the convergence of this estimator to H with asymptotic normality was shown under some appropriate assumptions on the variogram of Y. In Proposition 1.3 of [4], under assumptions on the spectral density, we prove that almost surely $\widehat{H}_{N,u,v} \longrightarrow H$, as $N \to +\infty$, with

$$\sqrt{N}\left(\widehat{H}_{N,u,v} - H\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma_H^{u,v}\right), \text{ with } N\mathbb{E}\left(\left(\widehat{H}_{N,u,v} - H\right)^2\right) \to \gamma_H^{u,v}, \quad (14)$$

for some positive constant $\gamma_H^{u,v}$.

Now let us consider the 2-dimensional random field X and denote by $V_{N,u}(\theta)$ the variations of the line process $Y = X_{\theta,x_0}$ defined by Equations (5) and (12). Let

$$\hat{h}_{N,u,v}(\theta) = \frac{1}{2\log(u/v)} \log\left(\frac{V_{N,u}(\theta)}{V_{N,v}(\theta)}\right). \quad (15)$$

Theorem 3. *Let $\theta = (\theta_1, \theta_2)$ be a unit vector.*
(a) If $\theta_1 \neq 0$ and $\theta_2 \neq 0$, then $\hat{h}_{N,u,v}(\theta) \to H_1$, almost surely as $N \to +\infty$. Moreover, when $a > 2$, (14) holds for $H = H_1$.
When $a \leq 2$,

$$\mathbb{E}\left(\left(\hat{h}_{N,u,v}(\theta) - H_1\right)^2\right) = O_{N\to+\infty}\left(N^{-2(1-1/a)}\right).$$

(b) If $\theta_2 = 0$, then $\hat{h}_{N,u,v}(\theta) \to H_1$, almost surely as $N \to +\infty$. Moreover, (14) holds for $H = H_1$.
(c) If $\theta_1 = 0$, then $\hat{h}_{N,u,v}(\theta) \to H_2$, almost surely as $N \to +\infty$. Moreover, (14) holds for $H = H_2$.

Proof. Let $\theta = (\theta_1, \theta_2)$ be a unit vector. According to Proposition 1.3 of [4] results follow if $T_\theta f$, the spectral density of the process X_{θ,x_0} fulfills assumptions of Propositions 1.1 and 1.2 of [4]. We already know an asymptotic development for $T_\theta f$ from Theorem 2 such that Propositions 1.1 applies. The main additionnal assumption of Propositions 1.2 is concerned with (9) requiring an asymptotic development for the derivative of $T_\theta f$. However, it can be weakened by the following one: $T_\theta f$ is differentiable on $\mathbb{R} \setminus (-r, r)$, for r large enough and

$$(T_\theta f)'(p) = O_{|p|\to+\infty}\left(|p|^{-2H-2}\right), \quad (16)$$

with $H = H_1$ if $\theta_1 \neq 0$ and $H = H_2$ otherwise. It remains to check (16). Let us assume that $\theta_1 \neq 0$ and recall that from (10), for all $p \neq 0$,

$$T_\theta f(p) = \frac{1}{|\theta_1|}\int_{\mathbb{R}} \left((p/\theta_1 + u\theta_2/\theta_1)^2 + u^{2a}\right)^{-\beta} du.$$

Therefore $T_\theta f(p)$ is differentiable on $\mathbb{R} \setminus \{0\}$ with

$$(T_\theta f)'(p) = -\frac{2\beta}{|\theta_1|\theta_1}\int_{\mathbb{R}} (p/\theta_1 + u\theta_2/\theta_1)\left((p/\theta_1 + u\theta_2/\theta_1)^2 + u^{2a}\right)^{-\beta-1} du.$$

Let $p > 0$ and let the change of variables $u = p^{1/a}s$ such that

$$(T_\theta f)'(p) = -\frac{2\beta p^{1/a-2\beta-1}}{|\theta_1|\theta_1} \int_{\mathbb{R}} \left(\frac{1}{\theta_1} + \frac{s\theta_2}{\theta_1 p^{1-1/a}}\right) \left(\left(\frac{1}{\theta_1} + \frac{s\theta_2}{\theta_1 p^{1-1/a}}\right)^2 + s^{2a}\right)^{-\beta-1} ds.$$

Then, as in the proof of Theorem 2, one can show that

$$|(T_\theta f)'(p)| \leq \frac{2\beta p^{1/a-2\beta-1}}{\theta_1^2} \left(\int_{\mathbb{R}} \frac{3}{2|\theta_1|} \left(1/4\theta_1^2 + s^{2a}\right)^{-\beta-1} ds + \int_{|s|>1} s^{-a(2\beta+1)} ds\right).$$

This gives (16) with $H = H_1$, since $(T_\theta f)'$ is odd, $1/a - 2\beta - 1 = -2H_1 - 2$ and $a(2\beta+1) > 1$.
The remaining cases $\theta_1 = 0$ or $\theta_2 = 0$ are straightforward using (b) and (c) of Theorem 2.

3 Application to trabecular bone x-ray images

Results of [1] suggest that fractal analysis of trabecular bone radiographic images is a good indicator of the alteration of the bone microarchitecture. In association with bone mineral density, fractal analysis improves the fracture risk evaluation. However, since this analysis is based on an isotropic model, it does not reveal bone texture anisotropy which is of special interest for the diagnosis of osteoporosis [7, 9].

In this section, we apply our estimation methods to trabecular bone x-ray images. The database contains radiographs of 211 post menopausal women, 165 being control cases and 46 osteoporotic fracture cases. Radiographs were acquired at INSERM U658 (Orleans, France) using a standardized procedure [18]. They were obtained on the calcaneus (a heel bone) with a direct digital X-ray prototype (BMA$^{\text{TM}}$, D3A Medical Systems, Orleans, France) with focal distance 1.15 m and X-ray parameters 55 kV and 20 mAs. The high-resolution digital detector integrated into the device prototype had a 50 μm pixel size, providing a spatial resolution of 8 line pairs per millimeter at 10% modulation transfer function. For each subject, the software device selected a region of interest (ROI) of constant size $1.6 \times 1.6\text{cm}^2$ at a same position using three predefined anatomical landmarks localized by the operator ; see figure 1.

In each image of the ROI, we computed the quadratic variations on lines oriented in four different directions ((1) horizontal direction ($\theta = (1,0)$), (2) vertical direction ($\theta = (0,1)$), (3) first diagonal direction ($\theta = (1,1)/\sqrt{2}$), (4) second diagonal direction ($\theta = (-1,1)/\sqrt{2}$)) and at scales u ranging from 1 to 20 pixels (see Equations (5) and (12)). Log-log-plots of mean variations vs. scale are shown on Figure 1. Scale properties observed in direction 2 differ significantly from those in directions 1, 3 and 4, which are very close. The graph is almost a line in direction 2 (vertical) whereas it is curvilinear in the other directions.

In direction 2, images could be considered as self-similar from the smallest scale to the largest one. In other directions, the self-similarity property is not

valid considering all scales. This property is partially true on two consecutive scale ranges: a small scale range from 1 ($50\mu m$) to 5 pixels ($250\mu m$) and a large scale range above 5 pixels ($250\mu m$). The first range covers scales corresponding to the thickness of trabeculae in the calcaneus. The second range includes scales which are beyond the size of bone structures. In other words, in directions 1, 3, and 4, we can clearly distinguish the scaling properties inside structures from those between the structures. Besides, differences observed between scaling properties in direction 2 and in directions 1, 3 and 4 reflect the presence of longitudinal trabeculae, which are predominant structures in the calcaneus oriented in direction 2.

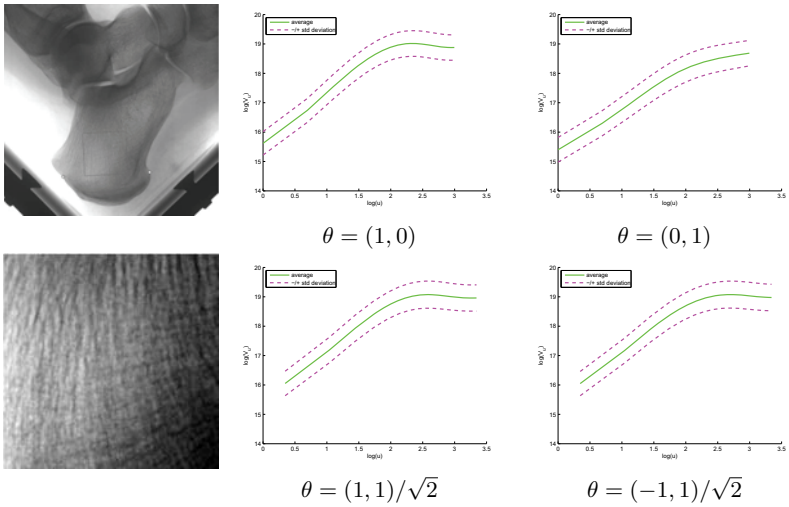


Fig. 1. On the left, a calcaneum ROI delimited in the heel with four anatomical landmarks (top image) and a radiograph of the ROI (bottom image). On the right, plots of the logarithm of the quadratic variations V_u versus the logarithm of the scale u (in pixels) in four different directions θ of the plane.

As described in Equation (15), we estimated the anisotropic index in the four directions on each image by comparing quadratic variations at scales $u = 6$ and $v = 5$ (pixels). On average, we obtained values 0.51 ± 0.08, 0.56 ± 0.06, 0.51 ± 0.08, and 0.51 ± 0.09 for directions 1, 2, 3, and 4, respectively. Comparisons of estimates in pairs of directions on each image are shown on Figure 2. They reveal that the anisotropic index is approximately the same in directions 1, 3, and 4 and higher in direction 2. This observation suggests that bone radiographs would have the same regularity in all directions except one (direction 2). From a theoretical point of view, such a property is consistent with the property of Gaussian operator scaling random fields proven in Theorem 2 of this paper.

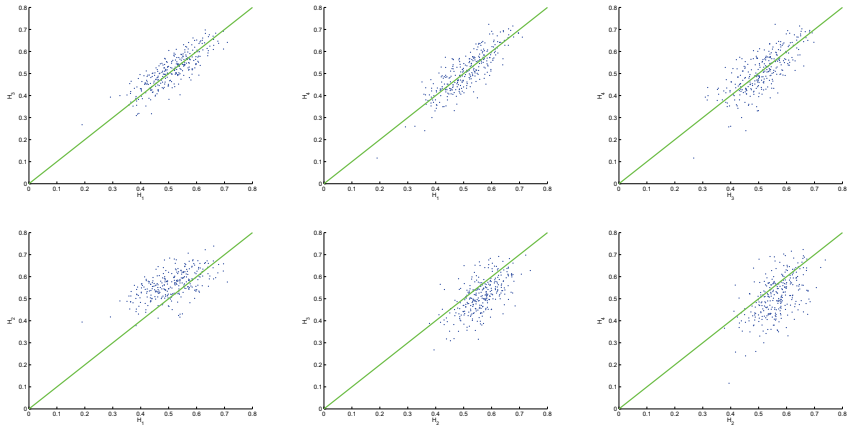


Fig. 2. Comparison of the anisotropic index estimation in pairs of directions. H_1, H_2, H_3 and H_4 are the estimation of the anisotropic index in directions $(1,0)$, $(0,1)$, $(1,1)$, and $(-1,1)$, respectively. In each figure, dots represent the estimation of the indices for a subject.

4 Conclusion

In this paper, we studied some particular operator scaling fields which are anisotropic generalizations of the Fractional Brownian Field. We showed that they have a privileged direction where line restrictions are more regular than in other directions. We also constructed some techniques for the estimation of parameters of these fields, using line restrictions and following principles of generalized quadratic variations. We proved the convergence of these estimators. We then modeled trabecular bone radiographs by operator scaling Gaussian random fields and showed experimentally that images had some properties of the model.

References

1. C.L. Benhamou, S. Poupon, E. Lespessailles, et al. Fractal analysis of radiographic trabecular bone texture and bone mineral density. *J. Bone Miner. Res.*, 16(4):697–703, 2001.
2. D. Benson, M. M. Meerschaert, B. Bäumer, and H. P. Scheffler. Aquifer operator-scaling and the effect on solute mixing and dispersion. *Water Resour. Res.*, 42:1–18, 2006.
3. H. Biermé, M. M. Meerschaert, and H. P. Scheffler. Operator scaling stable random fields. *Stoch. Proc. Appl.*, 117(3):312–332, 2007.
4. H. Biermé and F. Richard. Estimation of anisotropic gaussian fields through radon transform. *ESAIM: Probab. Stat.*, 12(1): 30–50, 2008.

5. H. Biermé, F. Richard, M. Rachidi and C. L. Benhamou. Anisotropic texture modeling and applications to medical image analysis. *ESAIM: Proc.*, Mathematical methods for imaging and inverse problems, 2009.
6. A. Bonami and A. Estrade. Anisotropic analysis of some Gaussian models. *J. Fourier Anal. Appl.*, 9:215–236, 2003.
7. B. Brunet-Imbault, G. Lemineur, C. Chappard, et al. A new anisotropy index on trabecular bone radiographic images using the fast Fourier transform. *BMC Med. Imaging*, 5(4), 2005.
8. C. Caldwell, S. Stapleton, D. Holdsworth, et al. Characterisation of mammographic parenchymal patterns by fractal dimension. *Phys. Med. Biol.*, 35(2):235–247, 1990.
9. C. Chappard, B. Brunet-Imbault, G. Lemineur, et al. Anisotropy changes in post-menopausal osteoporosis: characterization by a new index applied to trabecular bone radiographic images. *Osteoporos Int*, 16: 1193–1202, 2005.
10. C.-C. Chen, J. Daponte, and M. Fox. Fractal feature analysis and classification in medical imaging. *IEEE Trans. Pattern. Anal. Mach. Intell.*, 8(2):133–142, 1989.
11. S. Davies and P. Hall. Fractal analysis of surface roughness by using spatial data. *J. R. Stat. Soc. Ser. B*, 61:3–37, 1999.
12. B. Grosjean and L. Moisan A-contrario detectability of spots in textured backgrounds *J. Math. Imaging Vis.*, 33(3):313–337, 2009.
13. J. Heine and P. Malhorta. Mammographic tissue, breast cancer risk, serial image analysis, and digital mammography: tissue and related risk factors. *Acad. Radiol*, 9:298–316, 2002.
14. J. Istas and G. Lang. Quadratic variations and estimation of the local Holder index of a Gaussian process. *Ann. Inst. Henri Poincaré, Probab. Statist.*, 33(4):407–436, 1997.
15. I. Karatzas and E. Shreve. Brownian Motion and Stochastic Calculus. Springer-Verlag, 1998.
16. J. T. Kent and A. T. A. Wood. Estimating the fractal dimension of a locally self-similar Gaussian process by using increments. *J. R. Stat. Soc. Ser. B*, 59(3):679–699, 1997.
17. A. N. Kolmogorov. Wienersche Spiralen und einige andere interessante Kurven in Hilbertsche Raum.. *C. R. (Dokl.) Acad. Sci. URSS,*26:115–118,1940.
18. E. Lespessailles, C. Gadois, G. Lemineur, J.P. Do-Huu, L. Benhamou. Bone texture analysis on direct digital radiographic images: precision study and relationship with bone mineral density at the os calcis. *Calcif. Tissue Int.*, 80:97–102, 2007.
19. B. B. Mandelbrot and J. Van Ness. Fractional Brownian motion, fractionnal noises and applications. *SIAM Rev.*, 10:422–437, 1968.
20. Y. Xiao. Sample path properties of anisotropic Gaussian random fields. In *A Minicourse on Stochastic Partial Differential Equations*, (D. Khoshnevisan and F. Rassoul-Agha, editors), Lecture Notes in Math. 1962, pp. 145–212, Springer, New York, 2009.

Soft Plaque Detection and Automatic Vessel Segmentation

Shawn Lankton[1], Arthur Stillman[2], Paolo Raggi[2], and Allen Tannenbaum[1,2]

[1] Georgia Institute of Technology, Atlanta, GA, USA
[2] Emory University School of Medicine, Atlanta, GA, USA

Abstract. The ability to detect and measure non-calcified plaques (also known as *soft plaques*) may improve physicians' ability to predict cardiac events. This is a particularly challenging problem in computed tomography angiography (CTA) imagery because plaques may have similar appearance to nearby blood and muscle tissue. This paper presents an effective technique for automatically detecting soft plaques in CTA imagery using active contours driven by spatially localized probabilistic models. The proposed method identifies plaques that exist within the vessel wall by simultaneously segmenting the vessel from the inside-out and the outside-in using carefully chosen localized energies that allow the complex appearances of plaques and vessels to be modeled with simple statistics. This method is shown to be an effective way to detect the minute variations that distinguish plaques from healthy tissue. Experiments demonstrating the effectiveness of the algorithm are performed on eight datasets, and results are compared with detections provided by an expert cardiologist.

1 Introduction

Heart disease remains the leading cause of death in western countries. A number of recent studies have demonstrated that the presence of vascular plaques can be a significant indicator of risk for cardiac events [1]. The advent of multi-detector computed tomography (MDCT) and computed tomography angiography (CTA) allows one to non-invasively image these plaques, which can be categorized as either *calcified* or *non-calcified* based on their composition. Calcified plaques are easily discernible due to their high density and corresponding bright appearance in computed tomography (CT)

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

Fig. 1. Using CTA Imagery shown in (a) and a single-point initialization shown as the dark spot in (b), the vessel is segmented (c) and used to detect non-calcified plaques (d) in the vessel tree.

imagery. Consequently, several techniques have been presented to automatically detect calcified plaques with reasonable accuracy [2, 3]. Alternatively, non-calcified plaques, also known as *soft plaques*, have a CT attenuation similar to blood and myocardial tissue making them difficult to detect, even for trained experts [4, 5]. Detection and segmentation of soft plaques is essential given that non-calcified plaques are much more likely than calcified plaques to rupture and cause a variety of acute coronary syndromes [6].

This paper presents an effective and straightforward technique for automatically detecting soft plaques in CTA imagery based on multiple segmentations of the vessel wall. The segmentations are performed using multiple active contours driven by spatially localized statistical models that allow the complex appearances of plaques and vessels to be described with simple statistics. First, the vessel tree is extracted using a single, user-provided initialization point. Next two surfaces are constructed that lie just inside and just outside the vessel wall. An active contour model driven by a localized energy designed for plaque detection is then employed to simultaneously segment the interior and exterior of the vessel wall. Finally, areas where these two segmentations do not match are identified as potential regions of non-calcified plaque. Figure 1 shows an example of the imagery used, the initialization, the segmented vessel, and the detected plaque.

To the best of our knowledge, there is little published literature on automatic soft plaque detection in CTA imagery. The one recent paper to explicitly addresses this problem by Renard and Yang [7] also utilizes localized information to segment the interior and exterior of the vessel wall. The method in [7] requires pre-and post processing of the volume, describes no way to detect plaques in branching vessels, and uses an adaptive thresholding scheme for segmentation. By contrast, the present work casts the problem in a variational active contour framework that operates directly on the raw imagery thus reducing algorithmic complexity and the number of parameters. Furthermore, the proposed method naturally handles branching vessels and benefits from the geometric properties of active contours.

The segmentation and detection algorithms are described in detail in Section 2. In Section 3, experiments are shown on several vessel trees including the right coronary artery (RCA), left circumflex (LCX), and left anterior descending (LAD). The results are then compared with detections made by an expert cardiologist to validate the method. Finally, conclusions and directions for future work are given in Section 4.

2 Detection Algorithm

In this section, the framework used to guide an active contour segmentation with localized statistics is summarized, and two separate energies are shown. The behaviors and underlying assumptions are described in detail for each of these energies, which are used for vessel segmentation and plaque detection, respectively. Additionally, the process for vessel segmentation, creation of the interior and exterior initialization surfaces, and soft plaque detection are discussed.

2.1 Localized Contours

The detection method presented in this paper makes extensive use of localized active contours as described recently by several authors [8–10]. This technique allows region-based segmentation energies to be spatially localized such that statistical models of the foreground and background adapt to image information as it changes over the domain of the image. This allows for improved modeling accuracy with simplified statistical models. Furthermore, it is particularly powerful for segmenting vessels, which often exhibit changing image intensities over their length, and for the identification of non-calcified plaques, which typically have only slight intensity differences from surrounding structures.

To define these localized active contours, assume that the domain $\Omega \subset \mathbb{R}^n$ of a given intensity image I is partitioned into regions by an evolving surface in 3D or contour C in 2D, where C is embedded in a signed distance function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $C = \{x|\phi(x) = 0\}$ [11, 12]. The interior region is defined by the Heaviside function, $\mathcal{H}\phi$, which is 1 when $\phi < 0$, 0 when $\phi > 0$, and has a smooth transition through 0. Similarly, the interface at the zero level set can be denoted by $\delta\phi$, the derivative of $\mathcal{H}\phi$, which is 1 when $\phi = 0$ and 0 far from the interface. Localized active contours also utilize a characteristic function, $\mathcal{B}(x, y)$ representing a ball of radius r centered at x such that $\mathcal{B}(x, y) = 1$ when $\|x - y\| \leq r$. Given these definitions localized active contours minimize energy functionals of the form

$$E(\phi) = \int_{\Omega_x} \delta\phi(x) \int_{\Omega_y} \mathcal{B}(x, y) \cdot F(I, \phi, x, y)dy + \lambda\delta\phi(x)|\nabla\phi(x)|dx, \qquad (1)$$

where λ is a scalar weight on the arc-length and $F(I, \phi, x, y)$ represents an internal energy functional that is selected based on the application. The sparse level set representation presented by Whitaker in [13] is used for implementation due to its accuracy and speed.

To segment the vessels, an internal energy based on the one presented by Chan and Vese in [14] is used while a separate energy, similar to the one proposed by Yezzi *et al.* in [15], is used for detection of plaques. Both internal energies make use of localized interior and exterior means

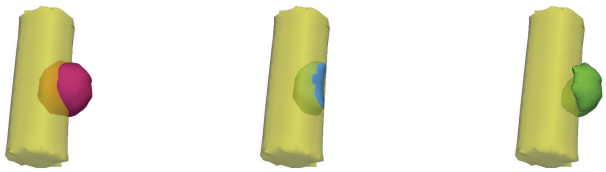
$$\mu_{\text{in}}(x) = \frac{\int_{\Omega_y} \mathcal{B}(x, y) \cdot \mathcal{H}\phi(y) \cdot I(y)dy}{\int_{\Omega_y} \mathcal{B}(x, y) \cdot \mathcal{H}\phi(y)dy} \qquad (2)$$

$$\mu_{\text{out}}(x) = \frac{\int_{\Omega_y} \mathcal{B}(x, y) \cdot (1 - \mathcal{H}\phi(y)) \cdot I(y)dy}{\int_{\Omega_y} \mathcal{B}(x, y) \cdot (1 - \mathcal{H}\phi(y))dy}, \qquad (3)$$

which measure local regions centered around a point x as shown in Figure 2.

More specifically, for vessel segmentation, the *uniform modeling* energy, based on [14] is employed. With this choice of internal energy, Equation (1) is minimized when local interior and exterior regions are well characterized by $\mu_{\text{in}}(x)$ and $\mu_{\text{out}}(x)$, respectively. Hence, the interior energy for uniform modeling is given by

$$F_{\text{UM}} = \mathcal{H}\phi(y)\big(I(y) - \mu_{\text{in}}(x)\big)^2 + (1 - \mathcal{H}\phi(y))\big(I(y) - \mu_{\text{out}}(x)\big)^2. \qquad (4)$$



(a) $\mathcal{B}(x, y)$ (b) $\mathcal{B}(x, y) \cdot \mathcal{H}\phi(y)$ (c) $\mathcal{B}(x, y) \cdot (1 - \mathcal{H}\phi(y))$

Fig. 2. Visualization of the local regions over which statistics are computed. The cylinder represents the surface C. The dark ball in (a) shows the full ball selected by $\mathcal{B}(x, y)$ while dark regions in (b) and (c) show the local interior and exterior regions respectively.

This energy is ideal for vessel segmentation, because it allows the surface to expand into areas of similar local intensity as long as a larger difference exists between local interiors and exteriors. This allows rapid segmentation of vessels despite changing intensities along the length of the vessel.

For soft plaque detection, the *means separation* interior energy is utilized. This energy is based on an energy presented in [15] and is minimized when the difference between $\mu_{\text{in}}(x)$ and $\mu_{\text{out}}(x)$ is maximized. Thus, the internal energy can be expressed as

$$F_{\text{MS}} = -\big(\mu_{\text{out}}(x) - \mu_{\text{in}}(x)\big)^2. \tag{5}$$

This is ideally suited for detection, because expansion into nearby regions that have slightly different intensities is discouraged, even if the local means are similar. This more stringent constraint is quite valuable when attempting to differentiate between vascular plaques and surrounding tissue.

2.2 Vessel Segmentation

The first step in detecting soft plaques is to segment a given vessel tree by minimizing Equation (1) with the uniform modeling internal energy from Equation (4). This energy is particularly powerful for segmenting vessels. By looking locally, it is able to accommodate variations in intensities that occur over the length of vessels, while also remaining sensitive to slight intensity differences that separate vessels from adjacent structures. Hence, the entire vessel can be captured from a single-point initialization while at the same time preventing leaks into nearby contrast-filled heart chambers. In the event that the vessel is completely occluded, additional initialization points may be required to achieve a full segmentation. This localized active contour approach also responds naturally to branches in the vessel, capturing them automatically without any special schemes to detect their presence.

During vessel extraction and plaque detection, a restricted domain $\tilde{\Omega}$ is used, which only includes voxels representing the heart and ignores very dark voxels representing air present in the lungs. This restricted domain is created by excluding from Ω any points that fall below a threshold of -600 Houndsfield units (HU). This threshold is significantly below the ranges for blood, plaques, and myocardial tissues in all CTA

data. Using a single-point initialization within the vessel, the entire vessel structure can be extracted by updating ϕ according to.

$$\frac{d\phi(x)}{dt} = \delta\phi(x) \int_{\tilde{\Omega}_y} \mathcal{B}(x,y) \cdot \delta\phi(y) \cdot \left(\left(I(y) - \mu_{\text{in}}(x)\right)^2 - \left(I(y) - \mu_{\text{out}}(x)\right)^2 \right) dy + \lambda \text{div} \left(\frac{\nabla\phi(x)}{|\nabla\phi(x)|} \right). \quad (6)$$

The parameter values $\lambda = 0.1\text{max}(|\frac{d\phi}{dt}|)$ and $r = 5$ mm are used throughout. This choice for r is reasonable, representing the maximum possible diameter for vessels in the vessel tree. Figure 3 shows the initializations, and resulting vessel segmentations on the LAD, LCX and RCA vessel trees.

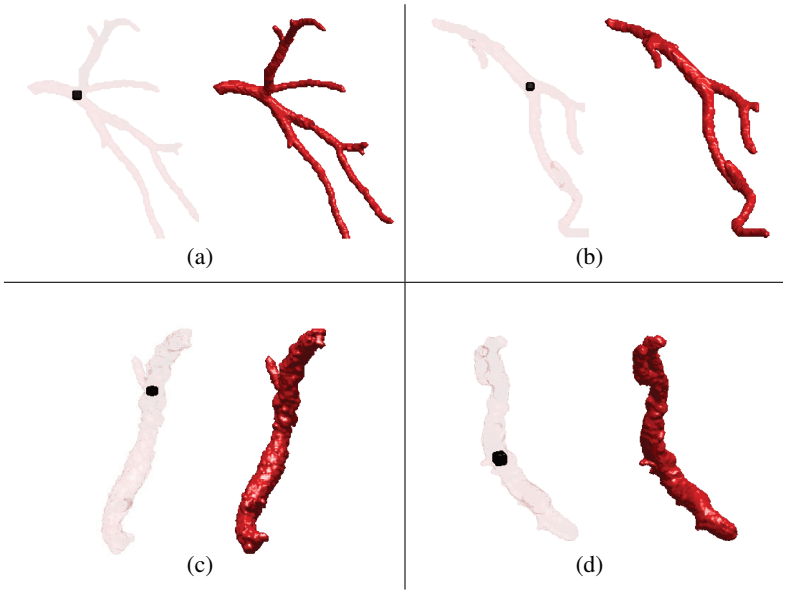


Fig. 3. 3D Renderings of initializations and segmentation results on the (a) LAD, (b) LCX, and (c, d) RCA. For each, the initialization is shown on left as a dark point and the resulting segmentation is shown on the right.

2.3 Constructing the Initialization

Using the vessel segmentation as starting point, two surfaces are created that act as initializations for the detection step. These initializations are formed inside and outside of the initial vessel segmentation so that plaques, which form within the vessel wall, will be located between the two surfaces.

To create the interior surface a third localized energy is employed that relies only upon the geometry of the surface and not image information. Here, the internal energy

defined as $F_{\text{SHRINK}} = \mathcal{H}\phi(y)$ penalizes interior area locally. Thus as the energy is minimized, the segmented vessel surface will thin with each iteration, but the thickest parts of the vessel surface will thin the fastest. Thus by running a few iterations, an interior surface is created that preserves the topology of the original surface and is always inside of it. This is similar to a medial axis (skeleton) except that relative thicknesses of vessel segments are preserved, meaning that although all segments become thinner, thick segments remain thicker than thin segments. Also, the surface resulting from this process will have a definite interior so that image statistics can be computed within it.

The exterior initialization is created by evolving the segmented vessel surface outward along its unit normal for several iterations. In our case, this is functionally equivalent to morphological dilation with a 5 mm$\times 5$ mm ball-shaped structuring element. The end result is a surface which is always outside of the initial vessel segmentation.

2.4 Plaque Detection

The next step involves the application of the means separation internal energy from Equation (5) to pull the two initial surfaces towards each other. Where no soft plaque exists, the two surfaces will meet on the vessel wall. However, if a soft plaque deposit exists between the two evolving surfaces they will each stop on the plaque boundary and remain separated from one another. By identifying areas where the two contours do not meet, soft plaques are detected.

The interior and exterior surfaces created in Section 2.3, are each deformed to minimize the means separation energy using the update function

$$\frac{d\phi(x)}{dt} = \int_{\tilde{\Omega}_y} \mathcal{B}(x,y) \cdot \delta\phi(y) \left(\frac{\big(I(y) - \mu_{\text{out}}(x)\big)^2}{A_{\text{out}}(x)} - \frac{\big(I(y) - \mu_{\text{in}}(x)\big)^2}{A_{\text{in}}(x)} \right) dy + \lambda \text{div}\left(\frac{\nabla\phi(x)}{|\nabla\phi(x)|} \right) \tag{7}$$

where $A_{\text{in}}(x)$ and $A_{\text{out}}(x)$ are defined as the areas of the local interior and exterior regions, respectively. Minimizing this energy deforms the surface such that local means are pulled apart as much as possible.

Initially, the local interior regions of the inner surface will only include bright blood voxels. As the surface deforms, it will expand to capture more voxels containing blood but will not expand into slightly darker soft plaque voxels. Similarly, the exterior surface will not contract if doing so would cause soft plaque voxels, which are slightly brighter than myocardium voxels, to move into its local exterior regions. This behavior allows the two evolving surfaces to capture soft plaques between each other because neither will move into the plaque voxels when driven by the localized means separation energy.

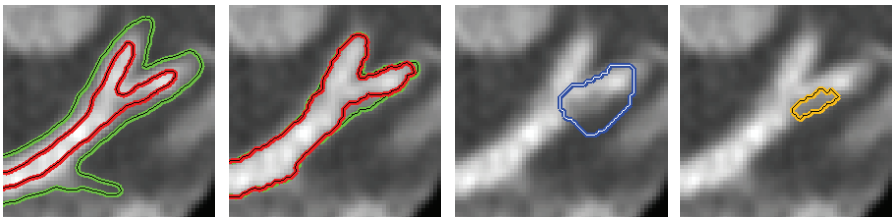
The two surfaces are evolved until convergence, and soft plaques are defined as areas where the distance between the two surfaces is larger than one voxel. A label map P, which is 1 where plaques exist and 0 elsewhere, is defined as

$$P = \big(1 - \mathcal{H}(\phi_{\text{interior}})\big) \cdot \mathcal{H}(\phi_{\text{exterior}} + 1). \tag{8}$$

3 Experiments

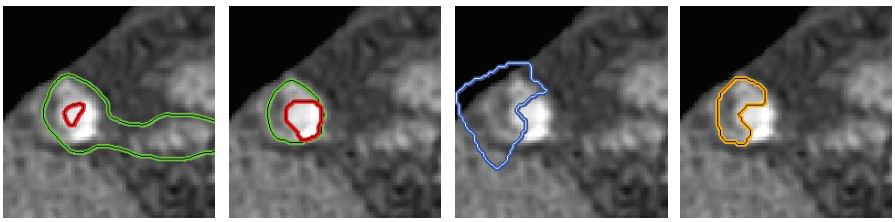
The results of experiments on MDCT scans of 8 patients undergoing CTA, for which we were able to obtain expert markings of soft plaques, are presented in this section. All scans were taken with a Siemens Sensation 64-slice multi detector CT machine and examined by an expert investigator to mark soft plaque deposits. For each dataset, plaques were detected using the algorithm presented in Section 2 based on a single-point initialization within the vessel tree of interest. The output of this detection algorithm was then compared to markings made by an expert investigator. Note that the expert markings denote large areas that are likely to contain plaques while the proposed detection algorithm finds the outline of the detected plaques.

First, we show examples of plaque detection on 2D slices where the behavior of the algorithm is easy to visualize. Figures 4 and 5 show examples of non-calcified plaques that cause negative and positive remodeling of the vessel wall, respectively. Each figure shows the initial interior and exterior surfaces, both surfaces after evolution with the local means separation energy, the expert markings, and the detected plaque.



(a) Initial Surfaces (b) Result of Evolution (c) Expert Marking (d) Detected Plaque

Fig. 4. A 2D view of detection results on the LAD (also shown in Figure 6a). The vessel demonstrates *negative* remodeling in the presence of plaque. The images depict (a) the initial interior and exterior surfaces, (b) the result of evolution with the local means separation energy, (c) the expert markings, and (d) the detected plaque.



(a) Initial Surfaces (b) Result of Evolution (c) Expert Marking (d) Detected Plaque

Fig. 5. A 2D view of detection results on the RCA (also shown in Figure 6d). The vessel demonstrates *positive* remodeling in the presence of plaque. The images depict (a) the initial interior and exterior surfaces, (b) the result of evolution with the local means separation energy, (c) the expert markings, and (d) the detected plaque.

Figure 6 shows rendered 3D views of the results from all 4 datasets that contained soft plaques. For each vessel tree, the segmented vessel is shown with truth markings and detected plaques. The datasets include plaques located in the LAD, LCX, and RCA vessel trees. In these experiments, plaques were identified well based on a single-click input. In total, 8 plaques were marked by the expert investigator within the 4 datasets. The proposed algorithm correctly identified 7 of 8 (87.5%) plaques with 1 false negative and 0 false positives. Additionally, the 4 datasets in which the LAD, LCX, and RCA were determined to be free of plaques were tested and, the algorithm correctly detected 0 plaques.

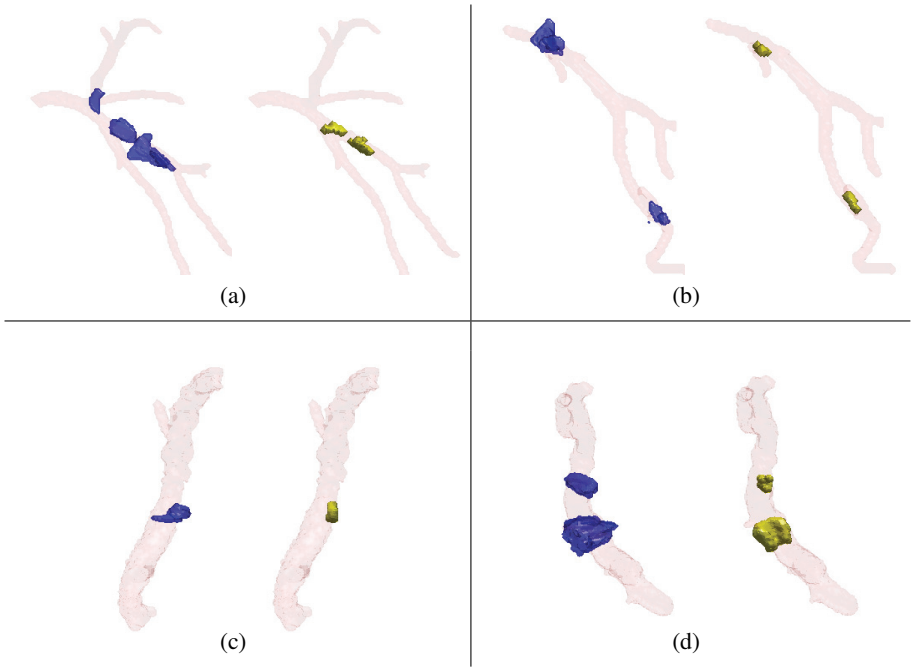


Fig. 6. 3D Renderings of detection results on the (a) LAD, (b) LCX, and (c, d) RCA. For each, the expert markings are shown in blue on the left, and the detection results are shown in yellow on the right.

4 Conclusions

In this note, we presented a technique that is capable of segmenting vessel trees and detecting non-calcified plaques automatically based on a very simple user input. The algorithm is based on a localized active contour framework that employs a scale parameter to restrict the statistical characteristics of the vessel into local regions. This allows simple probabilistic models based on local means to accurately extract the vessel and find areas where soft plaques exist. Experiments have shown good results that match experts' markings regarding the location of plaques.

Future work on this method will include coupling the evolution of the interior and exterior surfaces so that information about local intensities and geometries can be shared in order to detect plaques more robustly. Furthermore, a larger study is planned in which a larger number of datasets will be analyzed, a quantitative analysis will be performed, and the method will be compared with intravascular ultrasound imagery to confirm the presence and composition of detected plaques. We believe this work has the potential of being an important step forward in automatically detecting non-calcified plaques, which have been clearly linked with the occurrence of heart attacks and stroke.

References

1. Schroeder, S., Kopp, A., Burgstahler, C.: Noninvasive plaque imaging using multislice detector spiral computed tomography. Seminars in Thrombosis and Hemostasis **33**(2) (2007) 203–209
2. Saur, S., Alkadhi, H., Desbiolles, L., Szekely, G., Cattin, P.: Automatic detection of calcified coronary plaques in computed tomography data sets. In: Proceedings of Medical Imaging and Computing and Computer Assisted Intervention. (2008) 170–177
3. Brunner, G., Kurkure, U., Chittajallu, D., Yalamanchili, R., Kakadiaris, I.: Toward unsupervised classification of calcified arterial lesions. In: Proceedings of Medical Imaging and Computing and Computer Assisted Intervention. (2008) 144–152
4. Achenbach, S., Moselewski, F., Ropers, D., Ferencik, M., Hoffmann, U., MacNeill, B., Pohle, K., Baum, U., Anders, K., Jang, I., Daniel, W.G., Brady, T.J.: Detection of Calcified and Noncalcified Coronary Atherosclerotic Plaque by Contrast-Enhanced, Submillimeter Multidetector Spiral Computed Tomography. Circulation **109**(1) (2004) 14–17
5. Pohle, K., Achenbach, S., MacNeill, B., Ropers, D., Ferencik, M., Moselewski, F., Hoffmann, U., Brady, T., Jang, I.K., Daniel, W.G.: Characterization of non-calcied coronary atherosclerotic plaque by multi-detector row CT: Comparison to IVUS. Atherosclerosis **190** (2007) 174–180
6. Virmani, R., Burke, A., Farb, A., Kolodgie, F.: Pathology of the vulnerable plaque. Journal of the American College of Cardiology **47**(8-C) (2006) 13–18
7. Renard, F., Yang, Y.: Image segmentation for detection of soft plaques in multidetector CT images. In: Southwest Symposium on Image Analysis and Interpretation. (2008) 121–124
8. Lankton, S., Tannenbaum, A.: Localizing region-based active contours. IEEE Trans. on Image Processing **17**(11) (Nov. 2008) 2029–2039
9. Sum, K., Cheung, P.: Vessel extraction under non-uniform illumination: A level set approach. IEEE Trans. on Biomedical Engineering **55**(1) (Jan. 2008) 358–360
10. Li, C., Kao, C.Y., Gore, J.C., Ding, Z.: Minimization of region-scalable fitting energy for image segmentation. IEEE Trans. on Image Processing **17**(10) (Oct. 2008) 1940–1949
11. Sethian, J.: Level Set Methods and Fast Marching Methods, Second Edition. Springer, New York, NY (1999)
12. Osher, S., Fedkiw, R.: Level Set Methods and Dynamic Implicit Surfaces. Cambridge University Press, New York, NY (2003)
13. Whitaker, R.T.: A level-set approach to 3D reconstruction from range data. International Journal of Computer Vision **29**(3) (1998) 203–231
14. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. on Image Processing **10**(2) (February 2001) 266–277
15. A. Yezzi, J., Tsai, A., Willsky, A.: A fully global approach to image segmentation via coupled curve evolution equations. Journal of Visual Communication and Image Representation **13**(1) (March 2002) 195–216

Level Set Segmentation Using a Point-Based Statistical Shape Model Relying on Correspondence Probabilities

Heike Hufnagel[1], Jan Ehrhardt[1], Xavier Pennec[2], Alexander Schmidt-Richberg[1], and Heinz Handels[1]

[1] Department of Medical Informatics, University Medical Center Hamburg-Eppendorf, Germany, heike.hufnagel@uke.uni-hamburg.de,
[2] Asclepios Project, INRIA, Sophia Antipolis, France

Abstract. In order to successfully perform automatic segmentation in medical images containing noise and intensity inhomogeneities, modern algorithms often rely on a priori knowledge about the respective anatomy. This is often introduced by statistical shape models (SSMs) which are typically based on one-to-one point correspondences. In this work, we propose a unified statistical framework for image segmentation with shape prior information. The shape prior is an explicitly represented probabilistic SSM based on point correspondence probabilities, and the segmentation contour is implicitly represented by the zero level set of a higher dimensional surface. These two aspects are unified in a Maximum a Posteriori (MAP) estimation where the level set is evolved to converge towards the boundary of the organ to be segmented based on the image information while taking into account the prior given by the SSM information. The optimization of the MAP formulation leads to an alternate update of the level set and an update of the fitting of the SSM. We demonstrate the efficiency of our new algorithm with soft tissue segmentation where adaptive weights ensure that the SSM constraint is optimally exploited. Our experimental results show the well-posedness of the approach on noisy kidney CT data impaired by breathing artefacts.

1 Introduction

Segmentation algorithms play a major role in medical image analysis, however, due to typical medical image characteristics as poor contrasts, gray value inhomogeneities, contour gaps and noise, the automatic segmentation of many anatomical structures remains a challenge. To overcome these problems, models incorporating a priori knowledge about mean and variance of shape and gray levels as first proposed by [1] are often employed. However, a SSM is easily too constrained for some segmentation tasks when the number of training observations is too small to represent all the probable shape variabilities. To lighten this constraint, deformable models which balance between SSM and image information are frequently proposed (e.g. [2–4]). These SSMs are typically based on

one-to-one point correspondences and the segmentation is explicitely parameterized which makes them inflexible to topological changes. In this work, we propose an automatic segmentation method that couples an implicit parameterization of the segmentation with a probabilistic SSM based on point correspondence probabilities [5]. We integrate the explicit SSM information into an implicit level set framework where the contour of the segmentation is represented by the zero level set of a higher dimensional function. This front propagation approach was first proposed by [6] and later used for image segmentation by [7]. By choosing an implicit over an explicit representation, our algorithm is kept flexible to different segmentation problems, no remeshing mechanisms have to be implemented, the algorithm can be adapted easily to non-spheric topologies and the integration of regional statistics is straightforward. As a result, the segmentation method does not suffer from the limitations of SSMs while enjoying their benefits in yielding robust and smooth segmentations. An elaborate overview of level set segmentation methods and their advantages can be found in [8]. Methods have already been proposed to make and incorporate statistics on surfaces directly on the level-set surfaces by doing for instance PCA particularly on distance functions as first proposed by Leventon et al. [9] and later adapted by Tsai et al. [10] as well as Rousson et al. [11]. In [12] Cremers extended the approach by dynamical priors for tracking problems. This approach is intuitive and the integration of the priors on shape variation into the level set segmentation is straightforward. However, understanding the variability information on distance functions is not obvious so that it seems difficult to exploit it for a physical modeling of the shape variability. In contrast, by modeling the a priori shape knowledge via an explicitly represented, point-based SSM, we are able to incorporate variation modes with a physical significance which can be controlled directly.

In order to put the implicit representation within a unified statistical framework, we developed a MAP estimation of the level set which is optimized based on the image information as well as the SSM information about probable shapes. The MAP estimation is optimized by alternately updating the level set and updating the SSM parameters to best fit the current zero level set. As our segmentation method is focused on soft tissue in low quality images, we chose the level set formulation presented by Chan and Vese [13]. We further refine this approach by using a prior knowledge about grey value distributions inside and outside the organ as presented in [14] in order to robustify against intensity inhomogeneities across patients as well as inside the respective structures.

The remainder of this paper is organized as follows: The probabilistic SSM and the development of the MAP estimation are presented in section 2. Results of experiments on noisy kidney CT data are shown in section 3. Section 4 discusses the algorithm and results and concludes the paper.

2 Method

2.1 Statistical Shape Model Based on Correspondence Probabilities

In order to develop a comprehensive statistical formulation, we chose to use the probabilistic shape model proposed in [5]. In this model, the model and observation parameters are computed in a unified MAP framework which leads to an optimal adaption of the model to the set of observations. The registration of the model on the observations is solved using an affine version of the Expectation Maximization - Iterative Closest Point (EM-ICP) algorithm which is based on probabilistic correspondences and proved to be robust and fast [15]. The alternated optimization of the MAP explanation with respect to the observation and the generative model parameters leads to very efficient and closed-form solutions for (almost) all parameters. The SSM is explicitly defined by 4 *model parameters* $\Theta = \{\bar{M}, v_p, \lambda_p, n\}$:

- mean shape $\bar{M} \in \mathbb{R}^{3N_m}$ parameterized by N_m points $m_j \in \mathbb{R}^3$,
- variation modes v_p consisting of N_m 3D vectors v_{pj},
- associated standard deviations λ_p which describe - similar to the classical eigenvalues in the PCA - the impact of the variation modes,
- number n of variation modes.

From the parameters Θ of a given structure, the shape variations of that structure can be generated by $M = \bar{M} + \sum_{p=1}^{n} \omega_p v_p$ with $\omega_p \in \mathbb{R}$ being the deformation coefficients. The shape variations along the modes follow a Gaussian probability with variance λ_p:

$$p(\Omega) = \prod_{p=1}^{n} p(\omega_p) = \frac{1}{(2\pi)^{n/2} \prod_{p=1}^{n} \lambda_p} \exp\left(-\sum_{p=1}^{n} \frac{\omega_p^2}{2\lambda_p^2}\right), \ \Omega = \{\omega_1, ..., \omega_n\}. \quad (1)$$

In order to account for the unknown position and orientation of the model in space, we introduce the rigid or affine transformation T consisting of a matrix $A \in \mathbb{R}^{3\times3}$ and a translation $t \in \mathbb{R}^3$. A mean model point $\bar{m}_j$ can then be deformed and placed by $T \star m_j = A(\bar{m}_j + \sum_{p=1}^{n} \omega_p v_p j) + t$.

2.2 Level Set Segmentation Using a MAP Approach

The MAP Formulation Given a shape represented as a set of points with model parameters Θ in our SSM, we first model the probability of the surface best separating the interior and exterior of the object. This amounts to specifying the probability of a function ϕ whose zero level set is the object boundary knowing the SSM deformation parameters $Q = \{T, \Omega\}$. For the second step, we assume the following image formation model: The intensity is assumed to follow a law p_{in} for the voxels inside the object and a law p_{out} outside. Given this generative model, the segmentation is the inverse problem: The MAP method consists of estimating the most probable parameters ϕ and Q given the observation of an

image $I : X \to \mathbb{R}$. Hence, we evolve the level set function ϕ such that $p(\phi, Q|I)$ is maximized.

$$MAP = \operatorname{argmax} p(\phi, Q|I) = \operatorname{argmax} \frac{p(I|\phi, Q)p(\phi|Q)p(Q)}{p(I)}. \tag{2}$$

The shape prior does not add any information when the zero level set of ϕ is known, so I and Q are conditionally independent events $p(I|Q, \phi) = p(I|\phi)$, and we can write

$$p(\phi, Q|I) = p(\phi, T, \Omega|I) = \frac{p(I|\phi)p(\phi|T, \Omega)p(T, \Omega)}{p(I)}. \tag{3}$$

$p(I)$ is constant for a given image. Besides, we assume $p(T)$ to be independent and uniform, so we derive the following energy functional:

$$E(\phi, Q) = -\alpha \log(p(I|\phi)) - \tau \log(p(\phi|Q)) - \kappa \log(p(\Omega)) \tag{4}$$

with introduced weights $\alpha, \kappa, \tau \in \mathbb{R}$ to normalize the scale of the distributions. The first term of equation (4) describes the region-based energy with object specific priors which are given by the normalized grey value distributions p_{in} inside the organ and p_{out} outside the organ as found in the training data set which leads to

$$\log(p(I|\phi)) = -\int_X H_\epsilon(\phi(x)) \log p_{in}(I(x))dx - \int_X (1 - H_\epsilon(\phi(x))) \log p_{out}(I(x))dx.$$

$H_\epsilon(\phi(x))$ is a continuous approximation of the Heaviside function which is close to zero outside the object and close to one inside the object.
The front propagation of ϕ is guided by the probabilistic SSM which models all points x as a mixture of Gaussian measurements of the (transformed) model points m_j. The probability of a point x modeled by the SSM given Q is the normalized sum of correspondence probabilities of x and all m_j and equals

$$p(x|Q) = p_\Theta = \frac{1}{N_m} \sum_{j=1}^{N_m} \exp(-\frac{|x - T \star m_j|^2}{\sigma_\theta^2}). \tag{5}$$

For a contour Γ describing the zero level set of ϕ, the log of the probability is computed by $\log(p(\phi|Q)) = \log(\prod_{x \in \Gamma} p(x|Q)) = \int_{x \in \Gamma} \log p(x|Q)dx$. Integrating over the whole length of the contour is then expressed by

$$log(p(\phi|Q)) = \int_X \delta_\epsilon(\phi(x))|\nabla\phi(x)| \log p(x|Q)dx, \tag{6}$$

with $\delta_\epsilon(\phi(x))$ having a small support > 0. We then add a normalization over the length which leads to $\log(p'(\phi|Q)) = \log(p(\phi|Q)p(\phi|l_0)) = \int_X \delta_\epsilon(\phi(x))|\nabla\phi(x)|$ $(\log p(x|Q) - \beta)dx$ with $\beta = \frac{1}{l_0} \in \mathbb{R}$ where l_0 controls the normalization of the length. (For $p(x|Q) = const$ this equation is generalized to the classical smoothing term $\int_X \delta_\epsilon(\phi(x))|\nabla\phi(x)|dx$ as used by [13].)
The definition of $p(\Omega)$ is given by the Maximum Likelihood in equation (1).

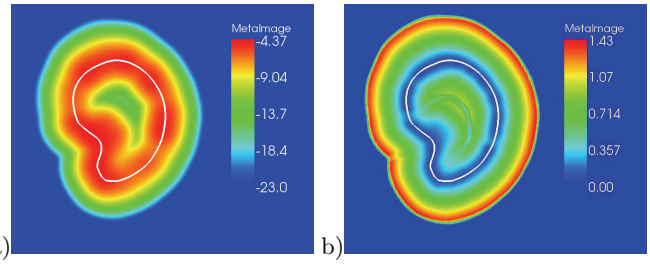


Fig. 1. Illustration of the probabilistic SSM represented by a white contour slice. a) Correspondence probability for image points x. b) Gradient magnitude of probability.

Segmentation Minimization of (4) is done by alternating a gradient decent for the embedding function ϕ with an update of the parameters T and Ω which serves to match the SSM to the current zero level set. The gradient descent for computing $\frac{\partial E(\phi,Q)}{\partial \phi}$ with fixed Q is given by

$$\frac{\partial \phi}{\partial t} = \delta_\epsilon(\phi)\left(\alpha_1 \log(p_{in}) - \alpha_2 \log(p_{out}) - \tau < \nabla(\log p_\Theta), \frac{\nabla\phi}{|\nabla\phi|} > \right.$$
$$\left. + div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)(\beta - \tau \log p_\Theta\ )\right). \tag{7}$$

To fit the SSM to the current zero level set, T is computed by

$$\frac{\partial E(\phi, T, \Omega)}{\partial T} = \frac{\partial}{\partial T}\int_X \delta_\epsilon(\phi(x))|\nabla\phi(x)| \log\left(\frac{1}{N_m}\sum_{j=1}^{N_m} \exp(-\frac{|x - T \star m_j|^2}{\sigma_\theta^2})\right) dx = 0$$

with fixed ϕ and Ω. We employ the affine EM-ICP registration where first the correspondence probabilities between the zero level set and the points of the SSM are established in the expectation step and then T is computed in the maximization step.
Subsequently, we fix ϕ and T and compute the Ω which solve $\frac{\partial E(\phi,\Omega,T)}{\partial \Omega} = 0$. This leads to a matrix formulation in a closed form solution. For a detailed derivation please refer to [5].
The constraints of the SSM on the level set propagation are twofold. The curvature term $\log p_\Theta\ div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)$ ensures that smoothness of the contour is more important at locations of low SSM probability, see figure 1a). Hence, we use a prior whose contour is length minimizing. In addition, the scalar product $< \nabla(\log p_\Theta), \frac{\nabla\phi}{|\nabla\phi|} >$ ensures that the zero level set is actively drawn towards the SSM shape, see figure 1b). The variance σ_θ^2 is a sensitive parameter and has to be carefully adapted to the problem at hand.

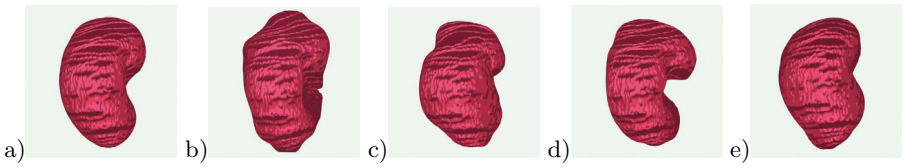


Fig. 2. Statistical Shape Model computed for a training data set of 10 segmented kidneys rendered with vtk. (a) shows the meanshape, (b-e) show the meanshape deformed with respect to first and second mode of variation: $\bar{M} - \lambda_1 v_1$, $\bar{M} + \lambda_1 v_1$, $\bar{M} - \lambda_2 v_2$, $\bar{M} + \lambda_2 v_2$.

2.3 Practical Aspects

Intensity Distribution In order to determine p_{in} and p_{out}, we sample the intensities around the surface and estimate the density functions using a Parzen window approach. We do this on the same training data set we built the probabilistic SSM on. We use a Gaussian kernel and chose a width of $h = 5$ based on empirical tests.

Automatic Initial Positioning In order to determine the initial SSM deformation parameters Q, we apply an evolutionary algorithm. A random population of shapes is built by generating a random set of normally distributed transformations T_k and deformations Ω_k and using them to deform the mean shape $\bar{M}$. In each iteration, the fittest individuals are selected and T_k as well as ω_{kp} are modified randomly to again generate a random set until a good initial position and shape are found. The fitness depends on the sum of distances between SSM points and the nearest voxel with high image gradient magnitude. An example is depicted in figure 3a).

3 Experiments and Results

We apply our method to the segmentation of the left kidney in CT images. The images $(512 \times 512 \times (32-52)$ voxels with resolution $0.98 \times 0.98 \times (2.9-5.0) mm^3)$ as well as the segmentations were kindly provided by the department of Computer Science, UNC, Chapel Hill. The CT images are quite noisy, and the quality lacks because of breathing artefacts.

Experiment Setup: The data set consists of 16 kidney CTs. The probabilistic SSM for the kidney is built using a training data set of 10 segmented observations, see figure 2. The segmentation method is then tested on the remaining 6 kidneys. For the segmentation, we set the weights $\alpha_1 = 1$, $\alpha_2 = 1$, $\kappa = 1$, $\beta = 0$ and $\tau = \{0.1, 0.2\}$. In most cases, the algorithm converged after 150 iterations. Each data set is segmented once with the level set segmentation without shape priors as proposed by [14] and once with the probabilistic SSM prior information integrated in the level set segmentation. For comparison purposes, we also

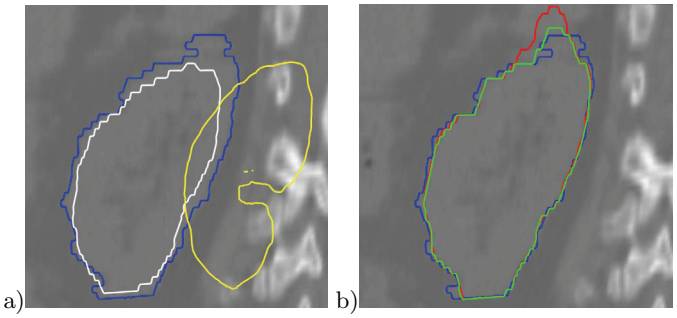


Fig. 3. Segmentation Results on a kidney in CT data, sagittal slice. The blue contour is the gold standard segmentation. Image (a) shows the initial contour in yellow and the contour after applying the automatic evolutionary algorithm as described in section 2.3 in white. Image (b) shows the result of the unconstrained (red) and the result of the SSM constrained (green) level set segmentation.

add the segmentation results on the same data set using the SSM directly in an Active Shape Model (ASM) approach as proposed in [16]. Starting from a well-defined explicit surface mesh (in contrast to an implicit segmentation contour), suitable boundary candidates are searched by evaluating certain grey value and gradient characteristics in normal direction, and the SSM is transformed and deformed in model space in order to optimally approximate the surface spanned by the best fitting candidates.

3.1 Experimental Results

We compare the results with the gold standard segmentations by evaluating the Jaccard coefficient, the Dice coefficient and the Hausdorff distance, see table 1. Using the SSM as an ASM does not lead to satisfying results. This is due to the difficulty of determining reliable contour candidates in the noisy CT images impaired by breathing artefacts. In contrast, the level set framework using a-priori information on the grey level intensities yields good segmentation results overall. The SSM constraint on the level set evolution yields even better results in all cases. The advantage of adding the prior shape information can be seen distinctly for patient 2 where the Hausdorff distance diminished from $9.95mm$ to $5.0mm$ and for patient 6 where the Hausdorff distance diminished from $12.57mm$ to $7.68mm$. This is due to the fact that the evolving zero level is attracted by a neighbouring organ with similar grey value intensities as the kidney. This leakage can be successfully prevented by integrating the SSM prior on shape probabilities. As an example, the effect on patient 2 is shown in figure 3b).

3.2 The Role of the Parameters

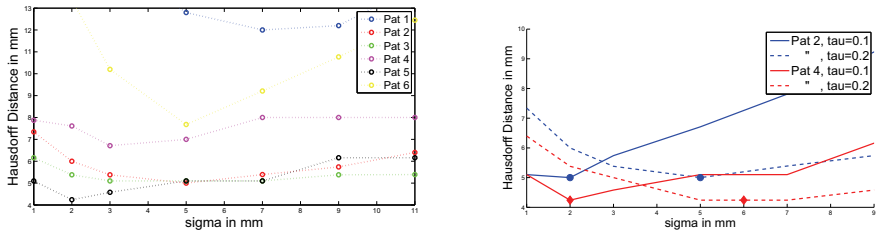
As our functional in eq. (7) is derived by a MAP explanation, in theory all coefficients should be equal to 1. Expanding on this probabilistic analogy, the

		ASM	only LS	LS + SSM
Pat1	D(A,B)	-	0.93	0.93
	J(A,B)	-	0.88	0.87
	H(A,B)	16.66	8.66	6.40
Pat 2	D(A,B)	-	0.91	0.93
	J(A,B)	-	0.83	0.88
	H(A,B)	7.34	9.94	5.0
Pat 3	D(A,B)	-	0.89	0.91
	J(A,B)	-	0.81	0.84
	H(A,B)	7.58	5.83	5.10
Pat 4	D(A,B)	-	0.88	0.89
	J(A,B)	-	0.78	0.80
	H(A,B)	10.11	8.01	6.40
Pat 5	D(A,B)	-	0.92	0.92
	J(A,B)	-	0.86	0.86
	H(A,B)	14.52	4.58	4.24
Pat 6	D(A,B)	-	0.84	0.86
	J(A,B)	-	0.73	0.75
	H(A,B)	-	12.57	7.68

Table 1. Segmentation Results for six different data sets. Left: Segmentation using the SSM as an Active Shape Model. Middle: Level set segmentation without SSM. Right: Level set segmentation using the probabilistic SSM constraint. D(A,B): Dice coefficient. J(A,B): Jaccard coefficient. H(A,B): Hausdorff distance in mm.

traditional coefficients of the variational methods (as e.g. in [13] or [11]) can be seen as powering factors which flatten or peak the density distributions. Concerning the SSM term (eq. (6)), the standard deviation σ_θ of the probabilistic SSM controls the matching of the SSM to the zero level set. This means that in practice, σ_θ should have values around $5mm$ to guarantee a successful matching for the problem at hand. However, the value of σ_θ also controls the strictness of the spatial constraint, so the introduction of the coefficients τ, β and α is necessary in order to position the influence of the SSM with respect to the other terms. What is more, β can be equal to 0 as the smoothness term $div\left(\frac{\nabla\phi}{|\nabla\phi|}\right)$ is also governed by τ as can be seen in eq. (7). Moreover, employing $-\tau \log p_\theta$ as weight has the advantage of using a distance-dependent smoothing term. Figure 4a) shows the influence of the choice of σ_θ for the Hausdorff distances obtained in the segmentation experiments with $\alpha = 1$, $\beta = 0$ and τ fixed to 0.1. These parameters lead to satisfying results for all kidneys except kidney 1. As can be seen, the optimal values for σ_θ are similar for all kidneys and should not exceed $5mm$ in this case.

The relation between the parameters τ and σ_θ are illustrated in figure 4b) where the Hausdorff distances for two kidney segmentations are plotted with respect to σ_θ for different values of τ. As can be seen, for a smaller τ the optimal σ_θ becomes smaller as well which results in a left shift of the curve. This is due to the fact that a smaller σ_θ as well as a greater τ result in a stricter constraint



a) b)

Fig. 4. Hausdorff distances. a) shows the Hausdorff distances of the segmentation results under parameters $\alpha = 1$, $\beta = 0$ and $\tau = 0.1$ for all kidneys with respect to σ_θ. b) illustrates the relation between the parameters τ and σ_θ and their influence on the resulting Hausdorff distances.

of the level set front propagation. However, the best result for the Hausdorff distance remains the same for both choices of τ.

4 Discussion

We proposed a novel algorithm for automatic segmentation of soft tissue. The algorithm employs a probabilistic SSM which is explicitly represented as a point cloud in combination with an implicitly defined evolving contour which makes regridding mechanisms obsolete. The coupling between point-based statistical shape models and level sets as proposed here is new to our knowledge of the literature on this subject and opens new insights on how to take the best of both worlds. We developed a MAP estimation of the level set which is optimized based on the image information as well as the SSM information about probable shapes. The MAP explanation leads to a two-phase formulation where an energy functional is alternately optimized with respect to the embedding level set and the deformation of the underlying SSM. The approach can be used for non-spherical surfaces and can be adapted to applications on data sets with different topologies as the connectivity between points does not play a role. First experiments showed that the new method works well and improves for some cases the approach of using an unconstrained level set segmentation. Especially when the intensity patterns of the organs close by are similar to the organ of interest, the level set segmentation can leak and produce erroneous results. The Hausdorff distance in this case yields a large value. By integrating the SSM probabilities, we reduce this leakage. The leakage problem of level set algorithms can be seen in different segmentation tasks such as the prostate. The proposed algorithm offers a solution to this problem by including the SSMs in a probabilistic framework such that they bring robustness to the segmentation process. Even from a low number of samples a prior on the probabilities can be extracted so that no huge training data set is necessary. From a theoretical point of view, a very powerful

feature of our method is that we are optimizing a unique criterion. However, the practical convergence rate has to be investigated more carefully as it depends on the choice of weights in the functional as well as the variance σ_θ^2 which controls the probability of occurrence with respect to the SSM. In the case of an organ shape which differs greatly from the shapes in the training data set for the SSM, a great sigma is needed in order to not constrain the contour evolution too much (as e.g. for Pat. 1, figure 4a)), so σ_θ is momentarily used somewhat as interactive parameter which is not the optimal solution. In current work, we want to extend the MAP formulation by integrating a priori knowledge about the expected volume V_0 which is given by the probability $p(\phi|V_0)$ and V_0 can be determined by evaluating the training data set. Further evaluation on other data includes the application on a coupled segmentation of acetabulum and femoral head.

References

1. Cootes, T., Taylor, C.: Statistical models of appearance for computer vision. Technical report, University of Manchester (2004)
2. Weese, J., Kaus, M., C, L., Lobregt, S., Truyen, R., Pekar, V.: Shape constrained deformable models for 3D medical image segmentation. In: IPMI 2001. (2001) 380–387
3. Heimann, T., Münzing, S., Meinzer, H.P., Wolf, I.: A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation. In: IPMI 2007. Volume LNCS 4584. (2007) 1–12
4. Kaus, M., von Berg, J., Niessen, W., Pekar, V.: Automated segmentation of the left ventricle in cardiac MRI. In: MICCAI 2003. Volume LNCS 2878. (2003) 432–439
5. Hufnagel, H., Pennec, X., Ehrhardt, J., Handels, H., Ayache, N.: Shape analysis using a point-based statistical shape model built on correspondence probabilities. In: Proceedings of the MICCAI'07. Volume 1. (2007) 959–967
6. Osher, S., Sethian, J.: Fronts propagation with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. Journal of Computational Physics **79** (1988) 12–49
7. Malladi, R., Sethian, J., Vemuri, B.: Shape modeling wit front propagation: A level set approach. IEEE Transactions on Pattern Analysis and Machine Intelligence **17**(2) (1995) 159–175
8. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. International Journal of Computer Vision **72**(2) (2007) 195–215
9. Leventon, M., Grimson, W., Faugeras, O.: Statistical shape influence in geodesic active contours. In: Computer Vision and Pattern Recognition. Volume 1. (2000) 316–323
10. Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A.: A shape-based approach to the segmentation of medical imagery using level sets. IEEE Transactions on Medical Imaging **22**(2) (2003) 137–154
11. Rousson, M., Paragios, N., Deriche, R.: Implicit active shape models for 3d segmentation in mr imaging. In: Medical Image Computing and Computer-Assisted Intervention MICCAI 2004. Volume 3216. (2004) 209–216

12. Cremers, D.: Dynamical statistical shape priors for level set-based tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(8) (2006) 1262–1273
13. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing **10**(2) (2001) 266–277
14. Ehrhardt, J., Schmidt-Richberg, A., Handels, H.: Simultaneous segmentation and motion estimation in 4D CT data using a variational approach. In: SPIE Medical Imaging 2008. Volume 6914. (2008)
15. Granger, S., Pennec, X.: Multi-scale EM-ICP: A fast and robust approach for surface registration. In: Proceedings of the ECCV'02. Volume 2525 of LNCS. (2002) 418–432
16. Hufnagel, H., Ehrhardt, J., Pennec, X., Handels, H.: Application of a probabilistic statistical shape model to automatic segmentation. In: World Congress on Medical Physics and Biomedical Engineering, WC 2009, München. (2009) To appear.

Bayesian Probability Maps For Evaluation Of Cardiac Ultrasound Data

Mattias Hansson[1], Sami Brandt[1,2], and Petri Gudmundsson[3]

[1] Center for Technological Studies, Malmö University, Sweden,
mattias.hansson@mah.se.
[2] Information Processing Laboratory, Oulu University, Finland.
[3] Faculty of Health and Society, Malmö University, Sweden.

Abstract. In this paper we propose a Bayesian approach for describing the position distribution of the endocardium in cardiac ultrasound image sequences. The problem is formulated using a latent variable model, which represents the inside and outside of the endocardium, for which the posterior density is estimated. As the Rayleigh distribution has been previously shown to be a suitable model for blood and tissue in cardiac ultrasound image, we start our construction by assuming a Rayleigh mixture model and estimate its parameters by expectation maximization. The model is refined by incorporating priors for spatial and temporal smoothness, in the form of total variation, preferred shapes and position, by using the principal components and location distribution of manually segmented training shapes. The posterior density is sampled by a Gibbs method to estimate the expected latent variable image which we call the Bayesian Probability Map, since it describes the probability of pixels being classified as either heart tissue or within the endocardium. Our experiments showed promising results indicating the usefulness of the Bayesian Probability Maps for the clinician since, instead of producing a single segmenting curve, it highlights the uncertain areas and suggests possible segmentations.

1 Introduction

Echocardiography is more accessible, mobile and inexpensive compared to other imaging techniques and has become a widely used diagnostic method in cardiology in recent years. Unfortunately ultrasound images struggle with inherent problems which in large part stem from noise, and is often referred to as *speckle contamination*. Speckle is the result of interference between echoes, which are produced when the ultrasound beam is reflected from tissue, and has the properties of a random field, see [1, 2]. The use of the Rayleigh distribution in modeling model speckle in ultrasonic B-scan images is well-established through early works, such as [3, 1], and more recently [4].

There is much previous work done in the field of segmentation of cardiac ultrasound images, of which [5] provides an excellent overview. Here we will only mention those works which, like our algorithm, treat segmentation of blood and tissue as a pixel-classification or region-based problem. Our model makes a dependency assumption of neighboring pixels via total variation. A similar approach is employed in [6–10], where Markov random field (MRF) regularization is used. Like our model [7, 9–11]

uses a Bayesian framework, although the construction of the posterior density function is different. Our approach uses priors on location and shape; of the forementioned, only [9] uses a shape prior. Also in [9] probabilistic pixel class prediction is used, which is reminiscent of the proposed Bayesian Probability Maps.

In this paper, we present a new method of determining the position of the endocardium in ultrasound sequences. This may be used for determining ejection fraction and assessment of regional wall abnormalities of the heart; measures used in diagnosis of ischaemic heart disease. The problem is formulated using a latent variable model, which represents the inside and outside of the endocardium. The method uses priors for spatial and temporal smoothness, in the form of total variation, preferred shapes and location, by using the principal components and location distribution of manually segmented training shapes. The main steps of the method are: 1) We assume a Rayleigh mixture model for the pixel intensities and estimate the parameters by expectation maximization. 2) The posterior distribution of the latent variables is sampled, using the estimated mixture parameters. 3) The mean of the posterior gives us the Bayesian probability map, which describes the position distribution of the endocardium. Instead of giving a single segmenting curve, the certainty of which may vary along the curve, our method provides a more versatile measure.

Our method shares some analogy with other region-based methods, but our approach of describing the position of the endocardium as the expected latent variable image and incorporating priors on location, shape and smoothness in space and time, is in its construction novel to our knowledge.

2 Model

Our goal is to determine the position of the endocardium in an ultrasound sequence. To accomplish this we represent the endocardium by the latent variable model with values one and zero for the inside and outside, respectively and estimate the posterior distribution of the latent variable model

$$P(\mathbf{u}|\mathbf{z},\theta) \propto p(\mathbf{z}|\mathbf{u},\theta)P(\mathbf{u}|\theta), \tag{1}$$

where $\mathbf{u}$ is the vector of latent variables, $\mathbf{z}$ represent image intensities stacked into a single vector and θ are parameters. The Rayleigh distribution has been reported to be an appropriate for modeling blood and tissue in cardiac ultrasound images , see [3, 1, 4]. Therefore to construct the likelihood $p(\mathbf{z}|\mathbf{u},\theta)$, we assume a Rayleigh mixture model for pixels intensities in the ultrasound images, as described in Section 2.1. In Section 2.2, we construct the prior distribution $P(\mathbf{u}|\theta)$ by using prior knowledge such as temporal and spatial smoothness, shape and location.

2.1 Likelihood

We model the ultrasound data as a two component mixture model, one for the object intensities and zero for the background. Denoting the intensity value of pixel k in an ultrasound image by z_k, we assume that

$$p(z_k|\theta) = \alpha p_{\mathrm{rayl}}(z_k|\sigma_1) + (1-\alpha)p_{\mathrm{rayl}}(z_k|\sigma_2), \tag{2}$$

where $\theta = \{\alpha, \sigma_1, \sigma_2\}$ are the mixture model parameters and $p_{\text{rayl}}(z|\sigma) = \frac{z}{\sigma}\exp(-\frac{z^2}{2\sigma})$, $\sigma > 0$ is the Rayleigh probability density function. Pixels are assumed to be independent in the mixture model. The likelihood is then defined as

$$p(\mathbf{z}|\mathbf{u},\theta) = \prod_j P(U_j \in \text{obj}|z_j,\sigma_1)^{u_j} P(U_j \in \text{backgr}|z_j,\sigma_2)^{1-u_j}, \quad (3)$$

where U_j and u_j are the random latent variable j and its realization, respectively, corresponding to z_j and $P(U_j \in \text{obj}|z_j,\theta) = \alpha p_{\text{rayl}}(z_j|\sigma_1)/(\alpha p_{\text{rayl}}(z_j|\sigma_1 + (1-\alpha)p_{\text{rayl}}(z_j|\sigma_2))$ and $P(U_j \in \text{backgr}|\theta) = 1 - P(U_j \in \text{object}|z_j,\theta)$.

2.2 Prior

Our prior model

$$P(\mathbf{u}|\theta) = P_{\text{B}}(\mathbf{u}|\theta)P_{\text{TV|B}}(\mathbf{u}|\theta)P_{\text{shape|B,TV}}(\mathbf{u}|\theta)P_{\text{location|B,TV,shape}}(\mathbf{u}|\theta) \quad (4)$$

consists of four components, where each characterizes different kinds of properties preferred. The Bernoulli component P_{B} is the discrete latent variable distribution following from the Rayleigh mixture model. The total variation $P_{\text{TV|B}}$ enforces spatial and temporal smoothness for latent variable images. Possible shape variations around the mean shape are characterized by trained eigenshapes of manually segmented images through $P_{\text{shape|B,TV}}$. The sequence of ultrasound images is divided into subsequences, to take the temporal variations of the endocardium into account, and so for each part of the ultrasound sequence a corresponding set of eigenshapes and mean is used. The location prior $P_{\text{location|B,TV,shape}}$ is constructed from the mean of the unregistered binary training shapes. The location prior describes the experimental probability value for each pixel location being either inside or outside of the endocardium, thus allowing only similar latent variable values as observed in the training data.

The Bernoulli prior is defined as

$$P_{\text{B}}(\mathbf{u}|\theta) = \prod_j \alpha^{u_j}(1-\alpha)^{1-u_j} \quad (5)$$

and is thus a prior on the proportion of zeros and ones in $\mathbf{u}$ and $j \in \{1,...,N\}$, where N is the total number of latent variables in $\mathbf{u}$.

Let $I_{\mathbf{u}}(\mathbf{x};n)$ be a latent variable image, where $\mathbf{x}$ and its n are spatial and temporal coordinates, respectively. The total variation prior is then given by

$$P_{\text{TV|B}}(\mathbf{u}|\theta) \propto \exp\{-\lambda_{\text{TV}}||I_u(\mathbf{x};n) * h||_{L_1}\},$$

where h is a three dimensional Laplacian kernel and $*$ denotes convolution.

Let $I_{\mathbf{u},\text{r}}(\mathbf{x};n)$ be the spatially registered latent variable image, corresponding to $I_{\mathbf{u}}(\mathbf{x};n)$, where the center of mass has been shifted to the origin; $\mathbf{u}_r^n$ and $\bar{\mathbf{u}}_r^n$ are the corresponding latent variable vectors. The shape prior is defined as

$$P_{\text{shape|B,TV}}(\mathbf{u}|\theta) \propto \prod_n \exp\{-\lambda_{\text{shape}}(\mathbf{u}_r^n - \bar{\mathbf{u}}_r^n)^T(\mathbf{C}_n + \lambda_0\mathbf{I})^{-1}(\mathbf{u}_r^n - \bar{\mathbf{u}}_r^n)\}, \quad (6)$$

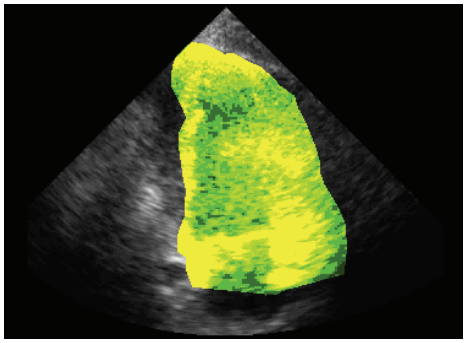


Fig. 1: Location prior (green) for a frame in an ultrasound sequence. The probability of a latent variable being sampled as endocardium is zero outside the colored area.

where $\mathbf{C}_n$ represents the truncated covariance of the training shapes, whose center of mass has been shifted to the origin, and $\lambda_0\mathbf{I}$ is the Tikhonov regularizer [12]. The shape prior cannot strictly impose a shape which does not exist in the data, i.e. a shape which has a very low likelihood. However the shape prior can enhance structures which have low likelihood, which may be due to the effects of noise. The advantage of this is that structures which are e.g. tissue, will never be classified as endocardium.

The location prior is defined as

$$P_{\text{location}|\text{shape,TV,B}}(\mathbf{u}|\theta) \propto \begin{cases} 1 \text{ if } \frac{1}{\sum_j u_j} \sum_n \sum_x h\left(g * \bar{I}_{\mathbf{u}_{\text{train}}}(\mathbf{x}; n)\right) I_{\mathbf{u}}(\mathbf{x}; n) = 1 \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

where $\bar{I}_{\text{u}_{\text{train}}} = \frac{1}{K} \sum_k I^k_{u_{\text{train}}}$ is the mean training image and K is the number of training images. g is a Gaussian kernel and h is the step function s.t. $h(t) = 1$ for $t > 0$, otherwise $h(t) = 0$. This component has the effect that when sampling individual latent variables outside of the (smoothed) mean shape, the result of sampling will be that the latent variable is set to zero. Inside the (unregistered) mean shape the sampling is unaffected.

The three regularization parameters λ_{TV}, λ_{shape} and λ_0 control the influence of the priors. Increasing λ_{TV} makes the sample temporally and spatially smoother, while increasing λ_{shape} makes the impact of the shape prior larger. Finally λ_0 controls the influence of the mean shape in the formation of the shape prior.

3 Algorithm

Our algorithm for generating Bayesian Probability Maps can be divided into three parts. First the mixture model parameters are estimated by the EM algorithm from our ultrasound data, as these parameters are needed to construct the posterior distribution of position of the endocardium. The posterior is then sampled by Gibbs sampling and the samples are used to compute the Bayesian probability map. The algorithm is summarized in Fig. 2.

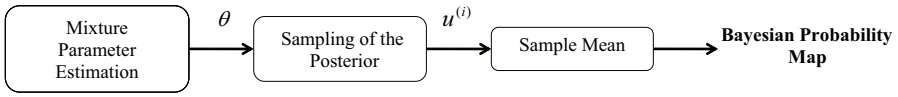


Fig. 2: Summary of the proposed algorithm to construct the Bayesian probability map.

3.1 Estimation of mixture model parameters

The complete data likelihood is represented according to the latent variable model as

$$p(\mathbf{z}, \mathbf{u}|\theta) = \prod_j p_{\mathrm{rayl}}(z_j|\sigma_1)^{u_j} p_{\mathrm{rayl}}(z_j|\sigma_2)^{1-u_j}, \tag{8}$$

where $\mathbf{z}$ are the pixel intensity values and $\mathbf{u} = (u_1, ..., u_N)$ are interpreted as missing data, s.t $u_j = 1$ if x_j is inside the heart chamber, and otherwise $u_j = 0$. The mixture parameters $\theta = \{\alpha, \sigma_1, \sigma_2\}$ are estimated by Expectation Maximization (EM) [13]. That is, on the E-step, we build the expected complete data loglikelihood, conditioned on the measured data and the previous parameter estimates, or

$$\begin{aligned}\chi(\theta, \hat{\theta}^{(n-1)}) &= \mathrm{E}_{\mathbf{u}|\mathbf{z},\hat{\theta}^{(n-1)}}\{\log p(\mathbf{z}, \mathbf{u}|\theta))\} \\ &= \sum_{j=1}^{N} \Big[P(U_j \in \mathrm{obj}|z_j, \hat{\theta}^{(n-1)}) \log p_{\mathrm{rayl}}(z_j|\theta) \\ &\quad + P(U_j \in \mathrm{backgr}|z_j, \hat{\theta}^{(n-1)}) \log p_{\mathrm{rayl}}(z_j|\theta) \Big].\end{aligned} \tag{9}$$

On the M-step, the expected complete data loglikelihood is maximized to obtain an update for the parameters,

$$\hat{\theta}^{(n)} = \underset{\theta}{\mathrm{argmax}}\ \chi(\theta, \hat{\theta}^{(n-1)}) \tag{10}$$

and the steps are iterated until convergence.

3.2 Sampling of the Posterior

The sampling of the posterior (1) was performed by conventional Gibbs sampling [14, 15] i.e. drawing samples from

$$\begin{aligned}&P(u_j|u_1^{(i)}, \ldots, u_{j-1}^{(i)}, u_{j+1}^{(i-1)}, \ldots, u_N^{(i-1)}) \\ &\quad = \left\{ P(u_j = k|u_1^{(i)}, \ldots, u_{j-1}^{(i)}, u_{j+1}^{(i-1)}, \ldots, u_N^{(i-1)}) \right\}_{k=0}^{1},\ j = 1, 2, \ldots, N.\end{aligned} \tag{11}$$

After iteration the center of the heart is calculated, which determines the area of influence of the shape prior. $\bar{I}_{u_{\mathrm{train}}} > 0.1$ defines a region, which contains a large part of the endocardium, but without most of the blood present outside the endocardium. The center of mass of $I_u(\mathbf{y}; n)$ is calculated within this region and is used as an approximation of the center of the heart.

3.3 Sample Mean

To characterize the posterior distribution, we compute estimate conditional mean of the latent variable vector over the posterior

$$\mathrm{E}\{\mathbf{u}|\mathbf{z},\theta\} \approx \frac{1}{M}\sum_i \mathbf{u}^{(i)} = \left(\hat{P}(U_k \in \mathrm{obj}\right)_{k=1}^{N} \equiv \hat{\mathbf{u}}_{\mathrm{CM}} \tag{12}$$

by the latent variable sample vectors $\mathbf{u}^{(i)}$. By the strong law of large numbers $\hat{\mathbf{u}}_{\mathrm{CM}} \rightarrow \mathrm{E}\{\mathbf{u}|\mathbf{z},\theta\}$ when $n \rightarrow \infty$. The corresponding image $I_{\hat{\mathbf{u}}_{\mathrm{CM}}}$ represents the Bayesian probability map.

4 Experiments

4.1 Material

The ultrasound data used in this paper consists of cardiac cycles of two-chamber (2C) apical long-axis views of the heart. The sequences were obtained using the echocardiogram machines Philips Sonos 7500, Philips iE33 or GE Vivid 7, from consecutive adult patients referred to the echocardiography laboratory at the Department of Cardiology at Malmö University Hospital, Sweden, which has a primary catchment area of 250,000 inhabitants. Expert outlines of the endocardium in the sequences have been provided by the same hospital.

4.2 Initialization

We estimate mixture model parameters for pixels in our data lying within the non-zero region of the mean of all training images. This is a natural constraint since we do not sample latent variables outside this region. As an initial estimate of mixture model parameters we set $\alpha^{(0)}$ to the proportion of object pixels in the training images, and σ_1 and σ_2 are set to maximum likelihood estimate $\hat{\sigma} = (\frac{1}{2Q}\sum_{i=1}^{Q} x_i^2)^{\frac{1}{2}}$ of object and background pixels in the training data, where Q is the number of pixels in the training set. Prior parameters $\lambda_{\mathrm{TV}}, \lambda_{\mathrm{shape}}, \lambda_0$ are set manually.

The Gibbs sampling algorithm is seeded by a sample obtained by Bayesian classification of the mean of the annotated images for each category of the heart cycle. The placement of these is determined by correlation of the sample, latent variable images, with masked log probability densities. Specifically, the position of $I_{\mathrm{init}}(\mathbf{x};k)$, the initial latent variable image at time k, is determined by matching it and a masked log probability matrix W by correlation. W is the matrix resulting from termwise multiplication of the mask matrix $\lceil \bar{I}_{\mathbf{u}_{\mathrm{train}}}(\mathbf{x};n) \rceil$ and the probability matrix $p(Z_k|\sigma_1)$, which gives the object probability of each pixel z_j in ultrasound image Z_k in the sequence.

4.3 Evaluation

We divide our data into two sets: training set and validation set. The training set consists of 20 cardiac cycles. The training set is further divided into sets, corresponding to parts of the cardiac cycle. The validation set consists of 2 cardiac cycles.

As evaluation measure the expected misclassification E_{mc} of a pixel, w.r.t the expert outline, is used. Let $I_{\mathrm{true}}(\mathrm{x}; n)$ be ground truth images corresponding to the data $\mathbf{z}$. Then the expected misclassification of a pixel in the examined sequence is given by

$$E_{\mathrm{mc}} = \frac{1}{N} \sum_{n} \sum_{x} \big(1 - I_{\mathrm{true}}(\mathbf{x}; n)\big) P\big(I_{\mathbf{u}}(\mathbf{x}; n) = 0\big) + I_{\mathrm{true}}(\mathrm{x}; n) P\big(I_{\mathbf{u}}(\mathbf{x}; n) = 1\big). \tag{13}$$

A low E_{mc} guarantees that the Bayesian Probability Map is a true reflection of the entire heart cycle, not just a few selected images.

4.4 Results

In Figure (3) and (4) results from two validation sequences are displayed. Eight frames have been selected from each sequence, four from the systole and diastole phase of the cardiac cycle, respectively. Validation sequence A consists of 41 frames, and sequence B of 26.

The Bayesian Probability Map displayed, for both validation sequences, is formed from 50 samples. The probability map spans colors from red to blue with degree of probability,of area being within the endocardium. Hence, red indicates the highest probability.

For sequence A we obtain $E_{\mathrm{mc}} = 0.07$, while $E_{\mathrm{mc}} = 0.11$ for sequence B. The higher expected misclassification for sequence B is clearly due to the fact that a large amount of blood is present outside the endocardium. However the probability map clearly captures the shape of the endocardium.

We compared our results with a Graph Cut method as described in [16–18]. We made this comparison since this method uses MRF, like [6, 7, 9–11]. In Figures (5) and (6) we observe that the Graph Cut method fails to identify the location as clearly as the proposed method.

5 Conclusion and future work

We have presented a novel approach to cardiac ultrasound segmentation, which consists of modeling the endocardium by latent variables. The latent variable distribution is then sampled which yields Bayesian Probability Map, which describes the location of the endocardium. In the future, we plan to introduce a connectivity prior for the latent variables, and to increase the sensitivity of categorization by refining the mixture model. Overall, the proposed Bayesian approach provides a framework into which such improvements can be easily incorporated and further evaluated. Furthermore we will introduce methods for optimizing the sampling process.

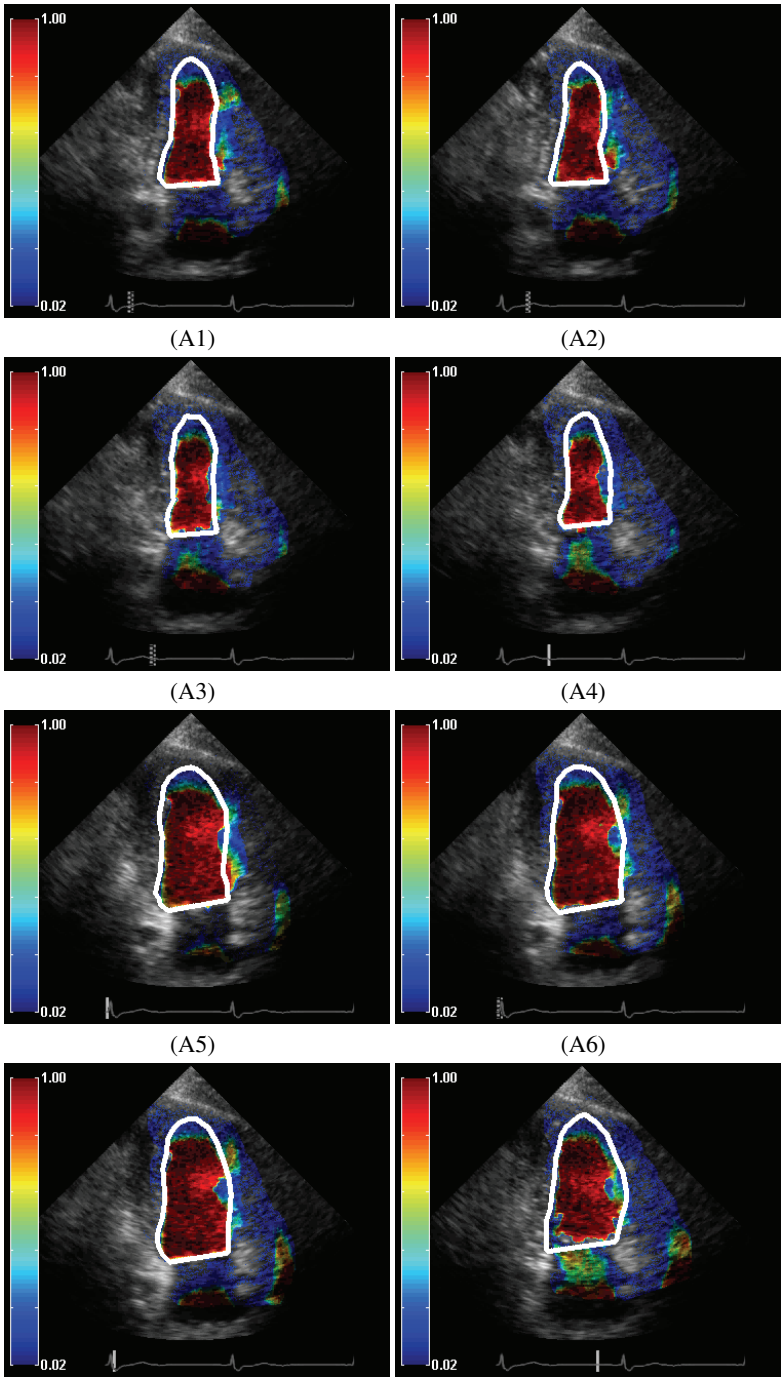


Fig. 3: Validation sequence A (41 frames). BPM with overlaid expert outline (white). Systole (A1-A4) and Diastole (A5-A8). E_{mc}=0.07, λ_{TV}=0.75, λ_{shape}=38.5, λ_0=100.

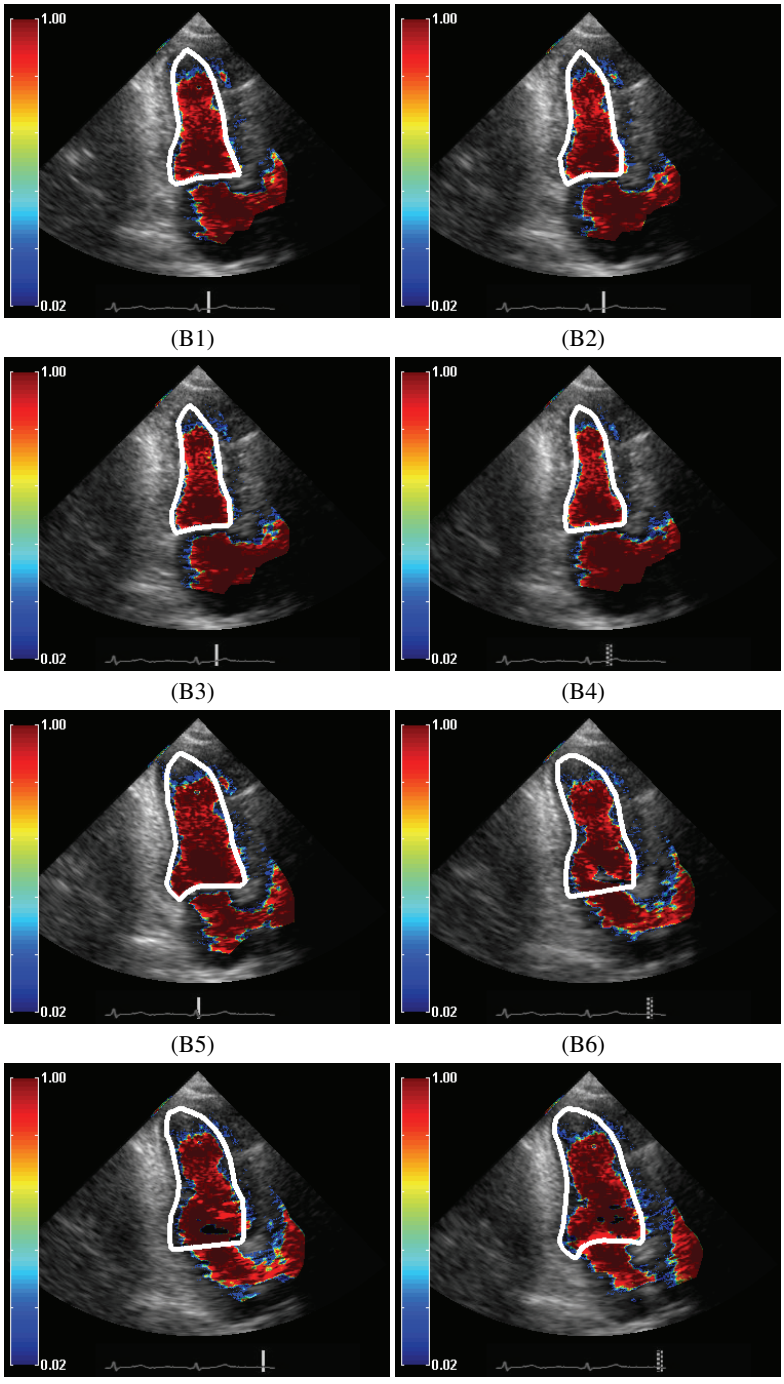


Fig. 4: Validation sequences B (26 frames). BPM with overlaid expert outline (white). Systole (B1-B4) and Diastole (B5-B8). E_{mc}=0.11, λ_{TV}=0.73, λ_{shape}=40, λ_0=100.

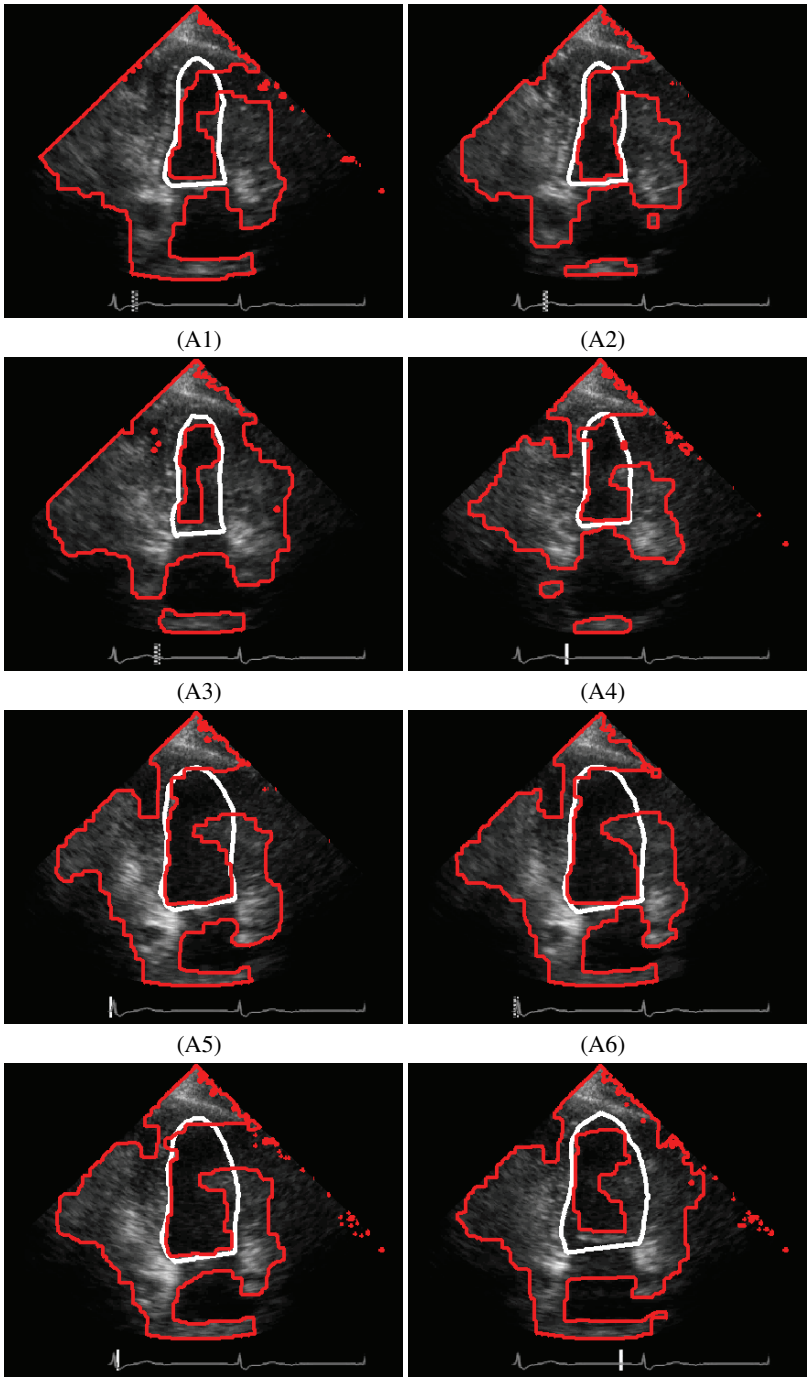
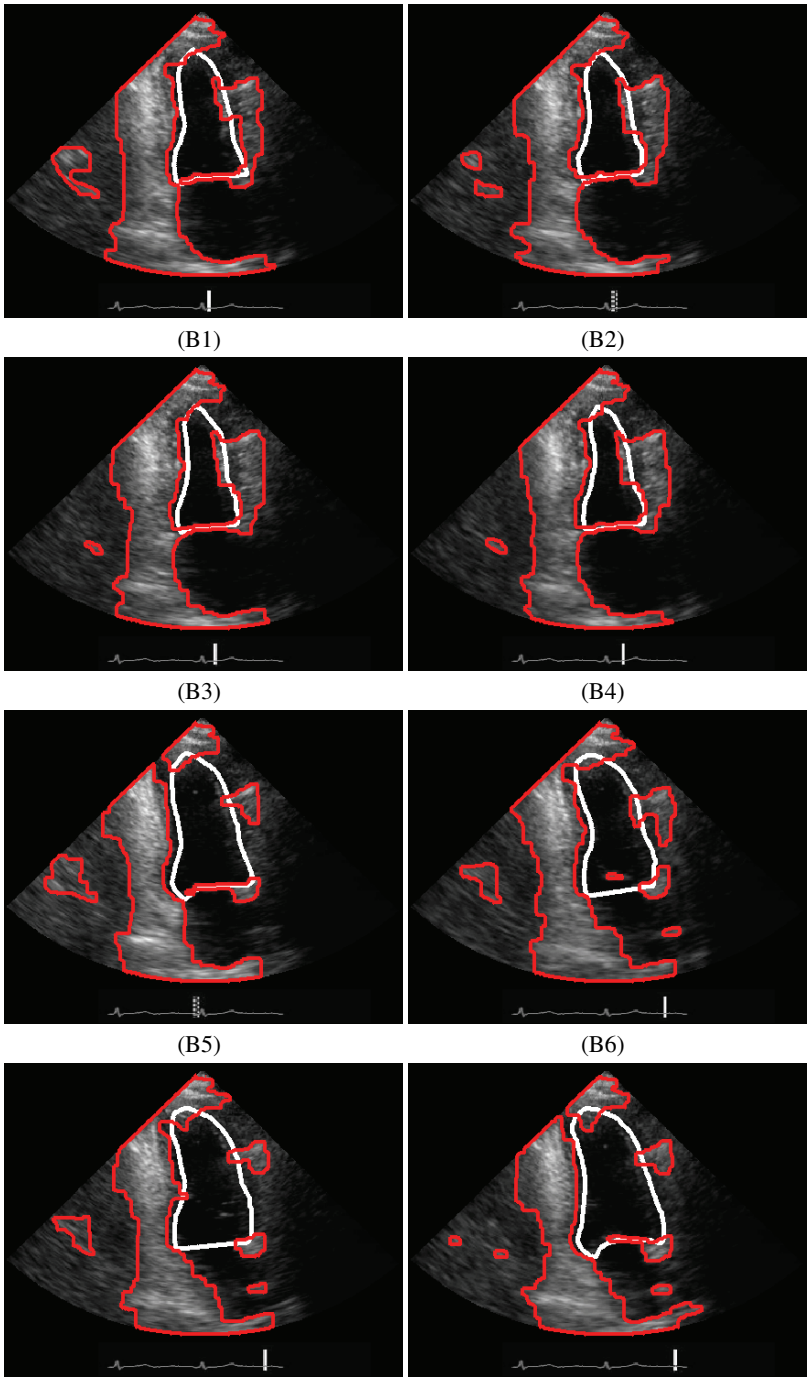


Fig. 5: Graph Cut (red) applied to Validation sequence A with expert outline (white). Systole (A1-A4) and Diastole (A5-A8).



(B1) (B2) (B3) (B4) (B5) (B6) (B7) (B8)

Fig. 6: Graph Cut (red) applied to Validation sequence A with expert outline (white). Systole (B1-B4) and Diastole (B5-B8).

References

1. Burckhardt, C.B.: Speckle in Ultrasound B-Mode Scans. IEEE Trans. Sonics and Ultrasonics **25** (1978) 1–6
2. Wagner, R.F., Smith, S.W., Sandrik, J.M., Lopez, H.: Statistics of Speckle in Ultrasound B-Scans. IEEE Trans. Sonics and Ultrasonics **30**(3) (1983) 156–163
3. Goodman, J.: Laser Speckle and Related Phenomenon. New York: Springer-Verlag (1975)
4. Tao, Z., Tagare, H., Beaty, J.: Evaluation of four probability distribution models for speckle in clinical cardiac ultrasound images. MedImg **25**(11) (2006) 1483–1491
5. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: A survey. Medical Imaging, IEEE Transactions on **25**(8) (2006) 987–1010
6. Friedland, N., Adam, D.: Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated anneal. IEEE Trans. Med. Imag. **8**(4) (1989) 344–353
7. Boukerroui, D., Baskurt, A., Noble, J.A., Basset, O.: Segmentation of ultrasound images: multiresolution 2d and 3d algorithm based on global and local statistics. Pattern Recogn. Lett. **24**(4-5) (2003) 779–790
8. Dias, J.M.B., Leitao, J.M.N.: Wall position and thickness estimation from sequences of echocardiographic images. IEEE Trans. Med. Imag **15** (1996) 25–38
9. Song, M., Haralick, R., Sheehan, F., Johnson, R.: Integrated surface model optimization for freehand three-dimensional echocardiography. Medical Imaging, IEEE Transactions on **21**(9) (2002) 1077–1090
10. Xiao, G., Brady, J.M., Noble, A.J., Zhang, Y.: Contrast enhancement and segmentation of ultrasound images: a statistical method. Medical Imaging 2000: Image Processing **3979** (2000) 1116–1125
11. Figueiredo, M., Leitao, J.: Bayesian estimation of ventricular contours in angiographic images. Medical Imaging, IEEE Transactions on **11**(3) (1992) 416–129
12. Tikhonov, A.: Solutions of Ill Posed Problems. Vh Winston, Scripta series in mathematics (1977)
13. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood form incomplete data via the EM algorithm (with discussion). J. R. Statist. Soc. Ser. B-Stat. Methodol. **39** (1977) 1–38
14. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans. Pattern Analysis and Machine Intelligence **6**(6) (1984) 721–741
15. MacKay, D.: Information Theory, Inference and Learning Algorithms. Cambridge University Press (2003) http://www.inference.phy.cam.ac.uk/mackay/itila/book.html.
16. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Analysis and Machine Intelligence **20** (2001) 1222–1239
17. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Analysis and Machine Intelligence **26** (2004) 147–159
18. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms. IEEE Trans. Pattern Analysis and Machine Intelligence **26** (2004) 1124–1137

Markov Surfaces: A Probabilistic Framework for User-Assisted Three-Dimensional Image Segmentation

Yongsheng Pan[1], Won-Ki Jeong[2], and Ross Whitaker[3]

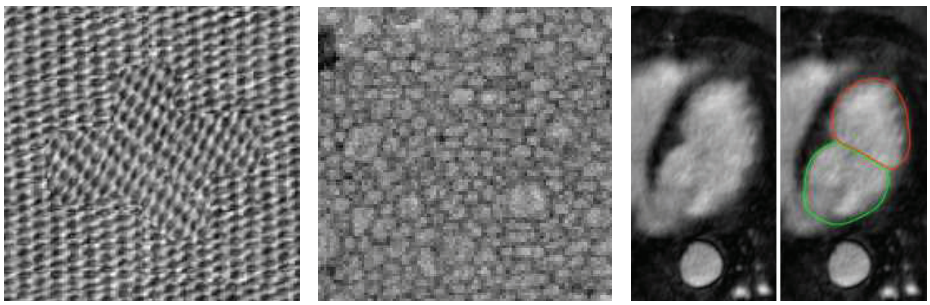
[1] Scientific Computing and Imaging Institute, University of Utah,
ypan@sci.utah.edu
[2] Initiative in Innovative Computing, Harvard University,
wkjeong@seas.harvard.edu
[3] Scientific Computing and Imaging Institute, University of Utah,
whitaker@cs.utah.edu

Abstract. This paper presents *Markov surfaces*, a probabilistic algorithm for user-assisted segmentation of elongated structures in 3D images. The 3D segmentation problem is formulated as a path-finding problem, where path probabilities are described by Markov chains. Users define points, curves, or regions on 2D image slices, and the algorithm connects these user-defined features in a way that respects the underlying elongated structure in data. Transition probabilities in the Markov model are derived from intensity matches and interslice correspondences, which are generated from a slice-to-slice registration algorithm. Bezier interpolations between paths are applied to generate smooth surfaces. Subgrid accuracy is achieved by linear interpolations of image intensities and the interslice correspondences. Experimental results on synthetic and real data demonstrate that Markov surfaces can segment regions that are defined by texture, nearby context, and motion. A parallel implementation on a streaming parallel computer architecture, a graphics processor, makes the method interactive for 3D data.

1 Introduction

Despite many significant advances in machine vision, many 3D image segmentation in radiation oncology, cardiology, and psychiatry still can not be fully automatic. Several examples, especially when the boundary of the object is not clearly separable using intensity differences, are shown in Figure 1. In Figure 1 (a), the background and the cross-shaped object at the center have the same texture patterns with a slightly different orientation. The transmission electron microscopy (TEM) data of retinal ganglia shown in Figure 1 (b) is hard to be segmented, even if the regions there can be distinguished by some combination of texture and dark boundaries. Figure 1 (c) shows an example of a magnetic resonance imaging (MRI) of a heart. The wall between a left atrium and a left ventricle in the left image is usually very thin and fuzzy.

It is usually hard to get good segmentation results without user interaction, especially when the boundary of the object is not clearly separable using intensity differences. Semiautomatic segmentation using user interaction, therefore, seems necessary in these cases.



(a) Texture image (b) TEM data (c) Heart MRI

Fig. 1. Examples that conventional intensity-based image segmentation methods fail to work.

Several methods [1] [2] [3] [4] [5] have been proposed for semiautomatic segmentation. A live wire algorithm is proposed in [1] to formulate boundary extraction as a graph searching problem. It utilizes the start points and the end points specified by users, and generates paths between these points using local gradient features. Falcao *et al* improved the efficiency of this method using live lane, and Schenk *et al* extended the live wire method to 3D based on shape-based interpolation and optimization. Both methods need users to interactively specify points for segmentation.

Ardon *et al* [3] proposed a surface extraction method based on the start and end curves specified by users. A network of minimal paths between these curves are generated using Fast Marching method, and a 3D surface is acquired by the interpolation between minimal paths. A 3D level set algorithm is performed for segmentation using the acquired 3D surface as initialization. Although this approach may provide good results, the topology of the network is often problematic [4]. An implicit method is proposed in [4] for this issue. It segments the object by finding a 3D real function using transport equations such that the network of minimal paths is contained in its zero level set.

All these methods, however, rely only on the gradient information. They may have problems when weak edges are present, e.g., in Figure 1 (c). The method proposed in this paper, Markov surfaces, generalizes these methods by using region information in a probabilistic framework. It allows users to define surfaces or regions in 3D data and to follow object boundaries in a way that does not require any specific formulation of an *edge*. This method is especially

useful for segmentation problems where the objects of interest are elongated in a predefined direction, which occurs regularly in 3D medical images.

The paper is organized as follows. Details of the probabilistic framework are introduced in Section 2. Section 3 discusses implementation issues. Experimental results are shown in Section 4, followed by Section 5, where a summary is presented.

2 Proposed Method

The proposed segmentation system consists of two parts: a preprocessing part, which establishes correspondences between slices by 2D image registration, and an interactive part, which finds the Markov surfaces that connect user defined regions. Once the mapping between slices are acquired, the user can select a start curve to compute cost or probability, for the entire input volume, of attaching every point to the initial conditions via a Markov chain. The user then selects an end curve, and the algorithm backtracks through the cost volume to create a set of curves, each of which has the highest probability of connecting the two sets. This process can be repeated until desired segmentation results are obtained.

2.1 A Probabilistic Formulation for Elongated Structures

The goal of this section is to create a method that allows users to quickly (interactively) define features (curves or regions) on disparate slices of a 3D dataset and connect these regions to form 3D structures in a way that conforms the the data. The strategy is to make it lightweight and general and thereby quickly applicable to a wide range of different applications and data types.

The proposed framework constructs the *most probable* paths between regions using a Markov chain model that incorporates the probability of correspondence between points on two different slices. Here we define the ith *slice* f_i of a volume $f(x, y, z)$ to be the 2D function defined by fixing one of the coordinates, so that we have $f_i(x, y) = f(x, y, i)$.

Denote a particular path $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)$ as having probability $P(\mathbf{W}) = P(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n)$, where $\mathbf{w}_i$ is a position of the path on the slice i of the input data. We model the path as a Markov chain [6], so that probability of each subsequent position along the path depends only on the previous position and probability. This gives

$$P(\mathbf{W}) = P(\mathbf{w}_1) \prod_{i=1}^{n-1} P(\mathbf{w}_{i+1}|\mathbf{w}_i) \tag{1}$$

The proposed strategy is to define the conditional probabilities in terms of a transition function from each pixel on one slice to every pixel on the next. Thus

$$P(\mathbf{w}_{i+1}|\mathbf{w}_i) = F(\mathbf{w}_{i+1}, \mathbf{w}_i), \tag{2}$$

where $F(\mathbf{w}_{i+1}, \mathbf{w}_i)$ may be considered as weights on a directed graph that connects every pixel in one slice to every pixel in the next.

We define the initial probability $P(\mathbf{w}_1)$ in terms of a user-defined region, A (e.g., a starting curve). These probabilities could include some uncertainty around this curve, or alternatively, as in this paper, a binary mask:

$$P(\mathbf{w}_1) = \begin{cases} a > 0 & \mathbf{w}_1 \in A \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where, for curves or points in a continuous domain, this would be, formally, a measure.

The path $\widetilde{\mathbf{W}}$ that maximizes $P(\mathbf{W})$ is defined:

$$\widetilde{\mathbf{W}} = \underset{\mathbf{W}}{\operatorname{argmax}} \left[P(\mathbf{W})\right] = \underset{\mathbf{W}}{\operatorname{argmin}} \left[-\sum_{i=2}^{n} \log F(\mathbf{w}_i, \mathbf{w}_{i-1}) - \log P(\mathbf{w}_1) \right] \tag{4}$$

where $-\log P(\mathbf{W})$ is referred as the *path cost* for the path $\mathbf{W}$.

A variation on Dijkstra's algorithm for dynamic programming, described in Section 2.4, is proposed to find the optimal path to every point in the volume. In practice, users define the start and end curves for an object on different slices in a volume, and the method quickly connects these regions using the most probable paths, as shown in Figure 2. The probability of the paths are derived from a set of automatically determined correspondences, and thus the resulting surfaces follow the structure of the data.

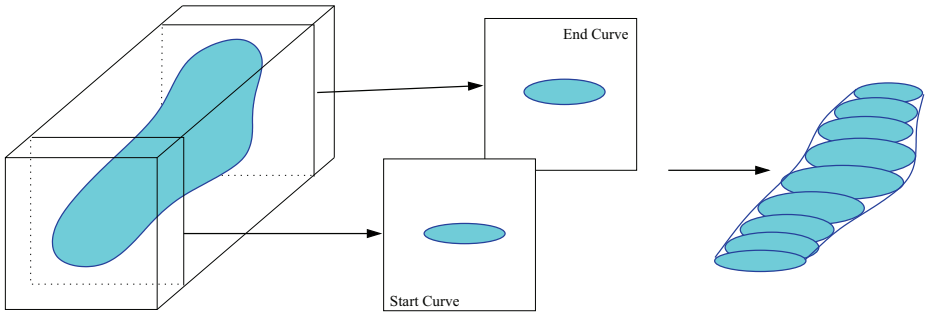


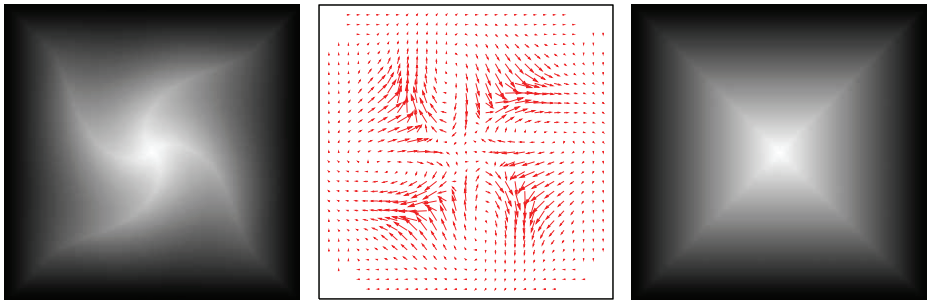
Fig. 2. Overview of the proposed 3D segmentation process using 2D tracking.

2.2 Slice-to-Slice Correspondence Estimation

The first step of the Markov surfaces is to find a dense set of correspondences between 2D slices in a 3D volume. There are a variety of ways that one could find such correspondences, such as patch correlations [7] or feature matching [8].

A deformable image registration method is applied here, which represents the correspondences as a smooth displacement field.

Let I denote a 2D image where $I(\mathbf{x}) : \Omega \mapsto \mathbb{R}$ and $\mathbf{x} \in \Omega \subset \mathbb{R}^2$, and I_i and I_{i+1} be two consecutive slices in the 3D volume. Define a correspondence from image I_i to I_{i+1} as a 2D transformation vector field $\mathbf{v}_{i,i+1}(\mathbf{x}) : \Omega \mapsto \mathbb{R}^2$ that maps each pixel in I_i to I_{i+1}.



(a) Source image (b) Warping vectors (c) Target image

Fig. 3. Vector field for warping between two images. Warping the source image (a) with the vector field (b) gives the target image (c).

Slice-to-Slice correspondence estimation is achieved by minimizing the following the energy E:

$$E = \frac{1}{2} \int_{\mathbf{x} \in \Omega} (\widetilde{I}_i - I_{i+1})^2 + \alpha \|\nabla \mathbf{v}\|^2 \tag{5}$$

where $\widetilde{I}_i = I_i(\mathbf{x} + \mathbf{v}_{i,i+1}(\mathbf{x}))$, $I_{i+1} = I_{i+1}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}_{i,i+1}(\mathbf{x})$, $\mathbf{x} \in \Omega$, and α is a constant parameter. The regularization term $\|\nabla \mathbf{v}\|^2$ helps to produce a smooth vector field and makes the problem well posed.

A gradient flow is used to compute $\mathbf{v}_{i,i+1}$, similar to [9, 10]. The update equation for $\mathbf{v}$ is written as

$$\mathbf{v}^{k+1} = \frac{1}{(I + \Delta t \alpha L)} \left[\mathbf{v}^k + \Delta t (I_{i+1} - \widetilde{I}_i^k) \nabla \widetilde{I}_i^k \right] = G \star \left[\mathbf{v}^k + \Delta t (I_{i+1} - \widetilde{I}_i^k) \nabla \widetilde{I}_i^k \right], \tag{6}$$

where G is a Gaussian kernel of width $\sigma = \sqrt{2\alpha \Delta t}$.

Because the energy function E we want to minimize in this problem is not, generally, convex, the solution of the minimization converges to local minima. Therefore, images with large deformations require better optimization strategies. To overcome such problems, we use cascading multigrid scheme (coarse to fine) with a 4-to-1 averaging D combined with a Gaussian smoothing kernel G (to eliminate aliasing effects) for down sampling the input images, and a 1-to-4 projection kernel U for up sampling the vector images. Figure 3 shows an example of a vector field for warping between two images.

A cascading multigrid scheme (coarse to fine) is applied in the implementation to handle large deformations. A 4-to-1 averaging combined with a Gaussian smoothing kernel G is utilized to downsample and process the input images, followed by a 1-to-4 projection kernel to upsample the vector images.

2.3 Slice-to-Slice Mapping Probability

In this section we define the slice-to-slice mapping probability $F(\mathbf{p}, \mathbf{q})$, defined in Eq. 2, which indicates the transition probability for a point $\mathbf{p}$ on slice i to a point $\mathbf{q}$ on slice $i+1$. A bilateral fall-off function F is applied based on two measures: the distance $d(\mathbf{p}, \mathbf{q}) = |\mathbf{p} + \mathbf{v}_i(\mathbf{p}) - \mathbf{q}|$ from the correspondence given by the registration and the intensity difference $g(\mathbf{p}, \mathbf{q}) = |f_i(\mathbf{p}) - f_{i+1}(\mathbf{q})|$ between the image values on the adjacent slices in the path. Thus, the transition function is

$$F(\mathbf{p}, \mathbf{q}) = \frac{1}{K_p} e^{-\frac{d^2(\mathbf{p},\mathbf{q})^2}{2k_d^2}} e^{-\frac{g(\mathbf{p},\mathbf{q})^2}{2k_g^2}} \tag{7}$$

where k_d and k_g are user-given parameters, and K_p is the normalization constant. The cost computation based on the probability function (7) is therefore the quadratic expression:

$$C^{\text{new}}(\mathbf{x}) = C_{n-1}(\widetilde{\mathbf{x}}) + \log(K_p) + \frac{d^2(\widetilde{\mathbf{x}}, \mathbf{x})}{2k_d^2} + \frac{g^2(\widetilde{\mathbf{x}}, \mathbf{x})}{2k_g^2}, \tag{8}$$

where $C^{\text{new}}(x)$ represents the new cost on grid x computed from the grid $\widetilde{\mathbf{x}}$ in the previous slice.

2.4 Shortest Path Cost Computation

A variant of Dijkstra's algorithm [11] is applied here to compute the optimal path on a directed graph formulated in Equation 4. The strategy is to compute a sequence of optimal paths to every pixel on each successive slice. Let $C_n(\mathbf{x}, \mathbf{y}, \mathbf{n})$ be the cost of the optimal path from the first slice to pixel $(\mathbf{x}, \mathbf{y}, \mathbf{n})$ on slice n. For every neighbor $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}, \mathbf{n}-1)$ of the pixel $(\mathbf{x}, \mathbf{y}, \mathbf{n}-1)$ in the slice $n-1$, the cost C_{new} is computed using $C^{new} = C_{n-1}(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}, \mathbf{n}-1) - \log(F(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}, \mathbf{n}-1; \mathbf{x}, \mathbf{y}, \mathbf{n}))$, and the minimum cost among all the neighbors is taken as the minimum cost of the path $C_n(\mathbf{x}, \mathbf{y}, \mathbf{n})$.

The proposed algorithm computes cost values on each slice in sequence, from the starting to the ending slice. Because of the strict causal relationship and the parallel nature of the method, its implementation on parallel architectures, such as graphics processing units, is straightforward and gives a significant speed-up, allowing the Markov surfaces to be generated and visualized at interactive rates, immediately after the user has defined the starting region (and while they are selecting the ending region).

Curves connecting the regions are generated by finding a path for every point on the end curve that connects to the start curve. This is done by *backtracking* from the ending region through each previous pixel on the optimal path to the starting slice. Figure 4 shows an example of backtracking from $\mathbf{w}_n$ on the end slice n through $\mathbf{W} = \{\mathbf{w}_i \in \Omega | i = 1, ..., n\}$ to slice 1, the start slice.

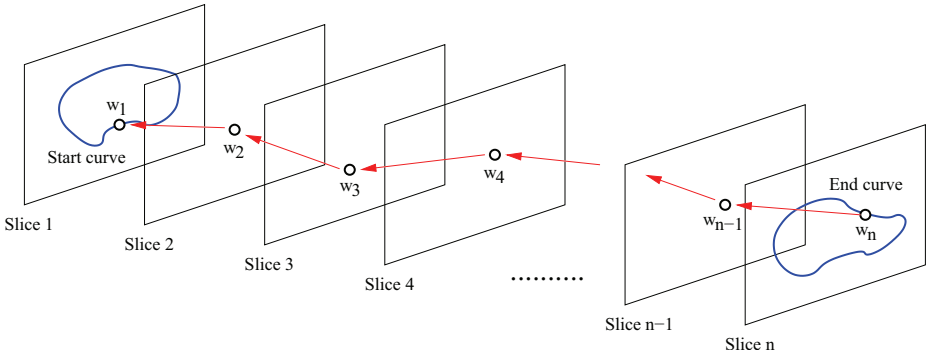


Fig. 4. Backtracking from a point $\mathbf{w}_n$ on the end curve to the start curve.

2.5 Path Interpolation and Forward Update

Paths may merge during backtracking [4]. Path merging generates separate points, which makes surface extraction challenging. A path interpolation method using Bézier curve [12] is applied to generate a continuous curve during backtracking. Bézier curve interpolation regularizes the extracted surface by dynamically adjusting paths to formulate a smooth curve in each backtracking slice. During regularization new paths are added; existing paths are adjusted; abnormal paths are removed. Curve smoothing using kernel averaging may be necessary before interpolation, especially for closed start curves.

The backtracking paths may only reach a small portion of the start curve [4], especially when the specified start curve is open. In this case we record the unmatched points in the start curve, and add their corresponding points to the backtracking curve in the next slice. Each unmatched point in the start curve may have multiple corresponding points. Only the closest one to the backtracking curve in the next slice is selected. This process is repeated from the start slice to the end slice, and Bézier curve interpolation is applied again for continuous curves.

3 Implementation Issues

3.1 Subgrid Cost Computation

The cost computation above assumes that the search on each previous slice is limited to pixel values, and thus paths are limited to the grid, which can result in aliasing and inaccuracies. Here we propose a more accurate subgrid method that solves for continuous locations at each slice using a linear interpolation of the intensity, cost, and correspondence positions from the previous slice. For this interpolation we divide each quadrangle of pixels into four triangles (by adding a new vertex with a value that is the average of its neighbors, as in Figure 5), and interpolate the necessary quantities on these triangles using barycentric

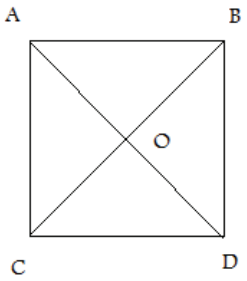


Fig. 5. Illustration for subgrid computation: the algorithm relies on linear interpolation on a set of 4 triangles for each set of 4 adjacent pixels on a slice.

coordinates. The optimal cost is computed for this analytic quantity over all four triangles on every quadrilateral contained in the corresponding neighborhood on the previous slice. The optimal cost and position for each pixel are those associated with the minimum over all these triangles in the previous slice. If the *log* transition probabilities are quadratic in position, cost, and intensity (as they are in this paper), the optimal cost on each triangle has a closed form solution, otherwise optimal costs must be obtained through some root trapping method.

3.2 Acceleration using Graphics Processors

Nonrigid image registration and optimal cost computation in the proposed method are highly parallelizable. They are therefore implemented on a graphics processing unit (GPU). Efficient computation on the GPU entails reusing memory in the access of overlapping neighborhood regions and the reduction of memory latency for random access. For this task, we use texture hardware on the current GPUs because texture memory is cached and interpolation is done for free by hardware. In this way we can get a high cache hit rate and significantly reduce the running time by using texture memory of the GPU.

4 Experimental Results

Experimental results are provided in this section. The proposed segmentation system is implemented on a Windows XP PC equipped with an Intel Core 2 Duo 2.4 GHz CPU, 4 GB main memory, and an NVIDIA Geforce 8800 GT graphics card.

Image registration and cost computation are time consuming processes, and they cannot be done in real-time on a conventional computer. For example, 2D registration of a 300×300 image (e.g., Figure 3) takes about 28 seconds on the CPU after 600 iterations. The same registration can be done only in 0.7 second

on the GPU. Slice-to-slice registration for a large 3D volume could easily take a few minutes even on the GPU, but it needs to be computed only once for each volume along each direction of interest. Thus, we consider a preprocessing step, done just before the interactive segmentation process.

Table 1 compares the running times (in seconds) of computation on the CPU and the GPU on one synthetic and two real 3D datasets. The MRI volume is about four times larger than the other two volumes. The most important factor in computation time is the size in the cost computation, because the algorithm complexity of cost computation is $O(kN)$ where k is the size of neighbor search and N is the size of input data (i.e., the number of voxels). Also, because of the benefit of using local memory (or texture cache [13]), the neighbor search size k affects the running time less significantly in the GPU version. Thus, the speed gain associated with the GPU implementation increases proportionally with neighbor search size.

	Synthetic $(150 \times 150 \times 50)$	Seismic $(301 \times 111 \times 32)$	MRI $(640 \times 460 \times 16)$
CPU time	4.86	21.5	596
GPU time	0.25	0.46	3.8
Search width	1.8	4.3	16
Speedup	19	46	156

Table 1. Comparison of running times for cost computation.

Fig. 6 illustrates the effects of Bézier interpolation and forward update. Backtracking results in separate points in Fig. 6(b) from the end curve in Fig. 6(a). Bézier interpolation of these points provides continuous contour in Fig. 6(c). Forward update is performed when the start curve in Fig. 6(d) is partly reached in Fig. 6(e). Fig. 6(f) shows the results from forward update in the next slice by adding those points corresponding to the unmatched points in the start curve. Fig. 6(g) shows the final results by applying Bézier interpolation.

Results on real and synthetic images demonstrate the effectiveness of the method. These results are better understood by referring to the accompanying video supplements. Each dataset has 30-50 slices, and intermediate results shows the results in the middle slices. Fig. 7 shows the segmentation results of a synthetic 3D texture and the video tracking results of a cup in real application. Fig. 8 displays the effects of the proposed method on real medical images. Volume renderings of the Markov surfaces are provided to show the segmented path.

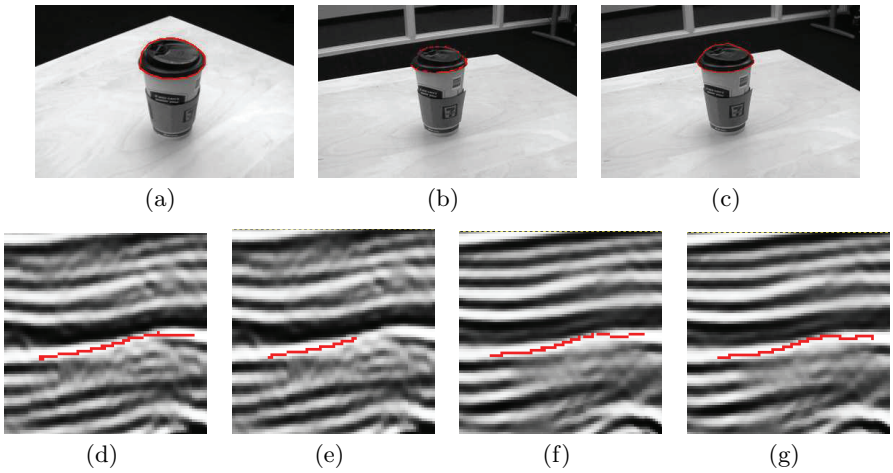


Fig. 6. Illustration of Bezier interpolation and forward update.

5 Summary

This paper addresses a user-assisted segmentation method, Markov surfaces, for elongated structures in 3D images. Markov surfaces are based on a probabilistic framework that finds the optimal paths that connect user-defined regions. Computationally demanding components, such as nonrigid image registration and cost/path computation, are implemented on the GPU, resulting in an interactive technique.

References

1. Barrett, W.A., Mortensen, E.N.: Interactive live-wire boundary extraction. Medical Image Analysis **1** (1997) 331–341
2. Falcao, A.X., Udupa, J.K., Samarasekera, S., Sharma, S.: User-steered image segmentation paradigms: Live wire and live lane. Graphical Models and Image Processing **60**(4) (July 1998) 233–260
3. Ardon, R., Cohen, L.D.: Fast constrained surface extraction by minimal paths. International Journal of Computer Vision **69**(1) (2006) 127–136
4. Ardon, R., Cohen, L.D., Yezzi, A.: Fast surface segmentation guided by user input using implicit extension of minimal paths. Journal of Mathematical Imaging and Vision **25**(3) (2006) 289–305
5.
6. Meyn, S.P., L.Tweedie, R.: Markov Chains and Stochastic Stability. Springer-Verlag (1993)
7. Engin, Z., Lim, M., Bharath, A.: Gradient field correlation for keypoint correspondence. (2007) II: 481–484
8. Wen, G., Lv, J., Yu, W.: A high-performance feature-matching method for image registration by combining spatial and similarity information. IEEE Trans. on Geoscience and remote sensing **46** (2008) 1266–1277

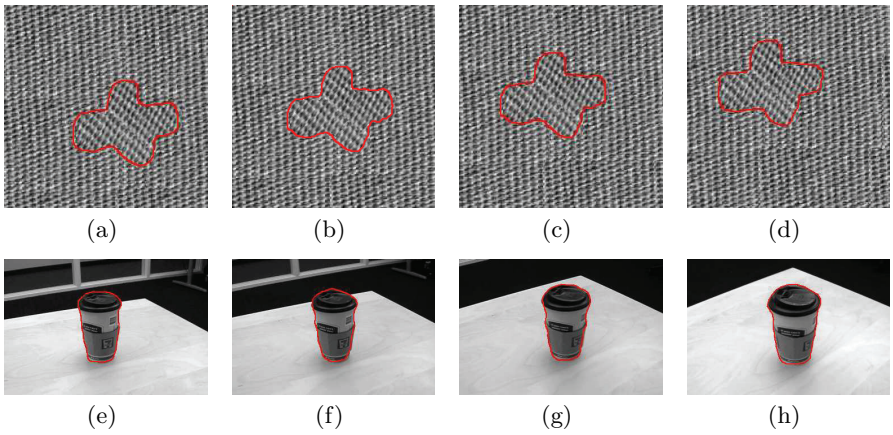


Fig. 7. Segmenting synthetic and real data. (a)(e) Start curves. (b)(c)(f)(g) Intermediate results. (d)(h) End curves.

9. Anandan, P.: A computational framework and an algorithm for the measurement of visual motion. Journal on Computer Vision **2** (1989) 283–310
10. Clarenz, U., Droske, M., Rumpf, M.: Towards fast non–rigid registration. In: Inverse Problems, Image Analysis and Medical Imaging, AMS Special Session Interaction of Inverse Problems and Image Analysis. Volume 313., AMS (2002) 67–84
11. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numerische Mathematik **1** (1959) 269–271
12. Foley, J.D., et al: Computer Graphics: Principles and Practice in C. 2 edn. Addison Wesley (1992)
13. Nguyen, H.: Gpu gems 3. Addison-Wesley Professional (2007)

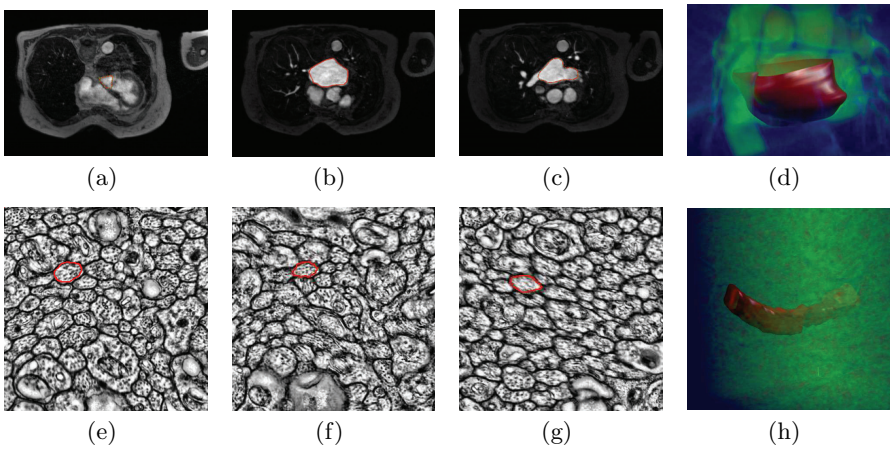


Fig. 8. Segmenting real medical data. (a)(e) Start curves. (b)(c) Intermediate results. (c)(g) End curves. (d)(h) Volume rendering of the Markov surfaces.

Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes

A. Criminisi, J. Shotton, and S. Bucciarelli

Microsoft Research Ltd, Cambridge, UK

Abstract. This paper introduces a new, efficient, probabilistic algorithm for the automatic analysis of 3D medical images. Given an input CT volume our algorithm automatically detects and localizes the anatomical structures within, accurately and efficiently.
Our technique builds upon randomized decision forests, which are enjoying much success in the machine learning and computer vision communities. Decision forests are enriched here with learned visual features which capture long-range spatial context. In this paper we focus on the detection of human organs, but our general-purpose classifier might be trained instead to detect anomalies. Applications include (and are not limited to) efficient visualization and navigation through 3D medical scans.
The output of our algorithm is probabilistic thus enabling the modeling of uncertainty as well as fusion of multiple sources of information (e.g. multiple modalities). The high level of generalization offered by decision forests yields accurate posterior probabilities for the localization of the structures of interest. High computational efficiency is achieved thanks both to the massive level of parallelism of the classifier as well as the use of integral volumes for feature extraction.
The validity of our method is assessed quantitatively on a ground-truth database which has been sanitized by medical experts.

1 Introduction

This paper presents a new, efficient algorithm for the accurate detection and localization of anatomical structures within CT scans. This work represents a significant step towards automatic parsing and understanding of medical images.

Our effort is motivated by recent studies which indicate how the great majority of a radiologist's time is spent searching through scanned volumes (often slice by slice) and navigating through visual data. Even with modern 3D visualization tools locating the organ(s) of interest and selecting optimal views is time consuming. Automatic tools for localizing major anatomical structures within 3D scans promises to speed up navigation and improve the user's work-flow [1]. For instance, a cardiologist may just click on a button to take him/her to the most appropriate view of the heart and its valves. Robust and efficient, probabilistic organ detection is also useful as input to other, more specialized tasks, e.g. detecting the heart to initialize a coronary tree tracer tool.

The two main contributions are: 1) We introduce an efficient algorithm for organ detection and localization which negates the need for atlas registration;

thus overcoming issues related to, e.g. : i) possible lack of atlases, and ii) selecting the optimal model for geometric registration. 2) We introduce new, context-rich visual features which capture long-range spatial correlations efficiently. The simplicity of our features combined with the intrinsic parallelism of our classifier yield high computational efficiency. Finally, our algorithm produces probabilistic output, useful for instance to keep track of uncertainty in the results, to take into account prior information (e.g. about global location of organs) or to fuse multiple sources of information (e.g. different acquisition modalities).

The proposed algorithm is applied here to the task of localizing nine anatomical structures (`head`, `heart`, `left eye`, `right eye`, `l. kidney`, `r. kidney`, `l. lung`, `r. lung`, and `liver`) in CT volumes with varying resolution, varying cropping, different patients, different scanner types and settings, contrast enhanced and not etc. Quantitative assessment is executed on a number of manually labelled ground-truth CT volumes.

Previous work. In the last few years research in object detection and recognition has made huge progress. The published work which is relevant to medical applications may be broadly categorized into the following three groups:

Geometric methods include template matching, and convolution techniques [2]. Geometrically meaningful features are used in [3, 4] for the segmentation of the aorta and the airway tree, respectively. Such geometric approaches often have problems capturing invariance with respect to deformations (e.g. due to pathologies), changes in viewing geometry (e.g. cropping) and changes in intensity. Techniques built upon "softer" geometric models with *learned* spatial correlations have been demonstrated to work well both for rigid and deformable objects [5].

Atlas-based techniques have enjoyed much popularity. Recent techniques for single and multiple organ detection and segmentation based on the use of probabilistic atlases include [6–10]. The apparent conceptual simplicity of such algorithms is in contrast to the need for accurate, deformable registration algorithms. The major problem with n-dimensional registration is in selecting the appropriate number of degrees of freedom of the underlying geometric transformation; especially as it depends on the level of rigidity of each organ/tissue.

Supervised, discriminative classification. Discriminative classification algorithms such as Support Vector Machines (SVM), AdaBoost and Probabilistic Boosting Trees have been applied successfully to tasks such as: automatic detection of tumors [11–14], pulmonary emphysema [15], organs in whole-body scans [19] and brain segmentation [16–18]. Our approach is also a discriminative classification technique. It achieves multi-class recognition efficiently and probabilistically. The classifier employed here is a random decision forest which, in non-medical domains has been shown to be better suited to multi-class problems than SVMs, as well as being more effective than boosting [20, 21]. A model of spatial context is learned by automatically selecting visual features which capture the relative position of visual patterns. Next we describe the details of our technique.

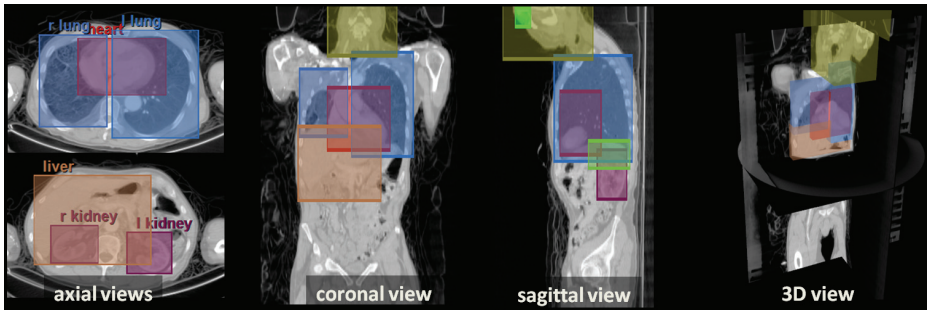


Fig. 1. Constructing labelled ground-truth databases. Organs within 3D CT scans are labelled via 3D, axis-aligned bounding boxes; different colours indicating different organs. Note that the fact that the boxes overlap is not a problem as they are used to indicate the position of the organ centre and the organ's approximate extent.

2 Automatic Parsing of CT Volumes

This section presents our ground-truth database, describes the decision forest classifier in the context of CT images and illustrates the visual features employed.

2.1 Labelled ground-truth database and exemplars

We have 39 CT volumes which have been annotated with 3D bounding boxes centred on each organ using our own annotator tool (shown in fig. 1). The user loads a CT scan, locates the organ of interest and draws a 3D box tightly around the organ. The database is split randomly into training and test sets as outlined in section 3. We focus on the following nine organs: `head`, `heart`, `left eye`, `right eye`, `l. kidney`, `r. kidney`, `l. lung`, `r. lung`, and `liver`. The use of axis-aligned boxes enables speedy manual annotation and is sufficient for tasks such as detection[1]. Our dataset comprises both contrasted and non-contrasted CT data, from different patients, cropped in different ways, with different resolutions and acquired from different scanners.

The goal is to determine the centre of each organ in previously unseen CT scans. A supervised technique such as ours needs to be trained from positive and negative examples of organ centres. Exemplars are provided from the annotation boxes as follows (cf. fig. 2). For each organ (e.g. the right kidney in fig. 2) we denote its annotation box with B^a. The set of positive training example points for the organ centre are defined as the set of points *within* a small box B^+; with B^+ of fixed size and located in the centre of B^a. Similarly, negative examples are all points *outside* the box B^- with same centre and aspect ratio as B^a but 50% in size. The region between B^- and B^+ is ignored.

[1] 2D annotation boxes (with no pixel-wise annotation) are used extensively in the PASCAL VOC challenges: `pascallin.ecs.soton.ac.uk/challenges/VOC/voc2009/`

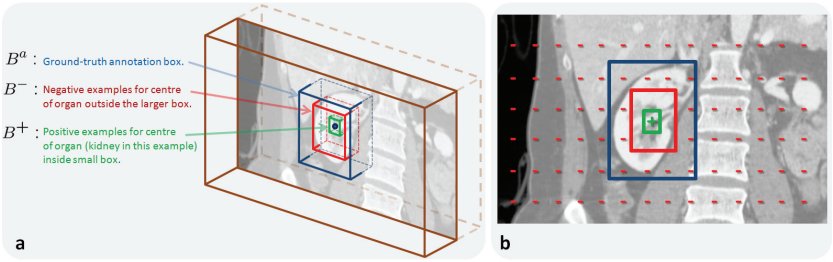


Fig. 2. Positive and negative training examples. (a) A 3D view illustrating the ground truth annotation box B^a, the box of positive examples B^+ and the box of negative examples B^-. **(a)** A 2D view further clarifying the regions where exemplars for the organ centre are taken. Positive examples are sampled within the box B^+ (in green). Negative example points are sampled outside the box B^- (in red).

2.2 Decision forests for recognition in 3D medical images

This section describes our adaptation of decision forests to the task of organ detection and localization in 3D CT scans.

A random decision forest [23, 24] is a collection of deterministic decision trees. Decision *trees* are popular classification algorithms which are known to suffer from over-fitting (poor generalization). Recently, it has been shown that the ensemble of many randomly trained decision trees (a random forest) yields much better generalization while maintaining the advantages of conventional decision trees [23]. Intuitively, where one tree fails the others do well.

We use the following notation. A voxel in a volume V is defined by its coordinates $\mathbf{x} = (x, y, z)$. The forest is composed of T trees denoted $\Psi_1, \cdots, \Psi_t, \cdots, \Psi_T$; with t indexing each tree (fig. 3). In each tree, each internal node (split node) performs a binary test on the input data and based on the result directs the data to the left or right child. The leaf nodes do not perform any action, they just store probability distributions over the organs of interest. Next we describe how the split functions are chosen and how the leaf probabilities are computed.

Forest training. Each point $\mathbf{x}$ of each *training* volume is associated with a known (manually obtained) class label $Y(\mathbf{x})$. The label indicates whether the point $\mathbf{x}$ belongs to the positive set of organ *centres* (see fig. 2) or not. Thus, $Y(\mathbf{x}) \in \{$ `head, heart, left eye, right eye, l. kidney, r. kidney, l. lung, r. lung, liver, background` $\}$, where the `background` label indicates that the point $\mathbf{x}$ is not an organ centre.

During training T is fixed (we use $T = 10$). Then, each point $\mathbf{x}$ is pushed through each of the trees starting at the root. Each split node applies the following binary test: $\xi > f(\mathbf{x}; \boldsymbol{\theta}) > \tau$ and sends the data to the respective child node. $f(\cdot)$ is a function applied to $\mathbf{x}$ with parameters $\boldsymbol{\theta}$. The parameters $\boldsymbol{\theta}$ identify the visual features which needs be computed. Features are described in the next section; for now it suffices to say that f computes some scalar filter response at $\mathbf{x}$. ξ and τ are parameters of the split node. The purpose of training is to optimize

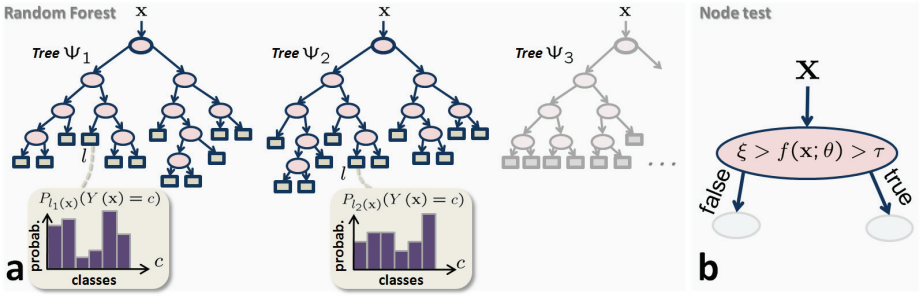


Fig. 3. Random Decision Forests. (a) An example random decision forest made of 3 trees. In each tree the internal nodes (shown with ellipses) perform simple tests on the input data while the leaf nodes (shown as squares) store the posterior probabilities over the classes being trained. During testing, a data point is pushed simultaneously through all T trees until it reaches T leaf nodes. The probability assigned to that point is the average of the probabilities of all the reached leaves (see text). **(b)** Each internal node performs a simple binary tests on the input data $\mathbf{x}$, based on the feature response $f(\mathbf{x};\boldsymbol{\theta})$. The quantities ξ, τ and $\boldsymbol{\theta}$ are parameters of the splitting test in that node.

the values of $\boldsymbol{\theta}, \xi, \tau$ of each split node by maximizing the data information gain, just like in the standard C4.5 tree training algorithm [25].

Injecting randomness for improved generalization. However, unlike standard tree training methods, here the parameters of each split node are optimized only over a randomly sampled subset Θ of all possible features (here $|\Theta| = 500$, details in section 2.3). This is an effective and simple way of injecting randomness into the trees, and it has been shown to improve generalization.

During node optimization all available features $\boldsymbol{\theta}_i \in \Theta$ are tried one after the other, in combination with many discrete values for the parameters ξ and τ. The combination $\xi^*, \tau^*, \boldsymbol{\theta}^*$ corresponding to the maximum information gain is then stored in the node for future use. The expansion of a node is stopped when the maximum information gain is below a fixed threshold. This gives rise to asymmetrical trees which naturally stop growing when no further nodes are needed. In this work the maximum tree depth D is fixed at $D = 15$ levels.

Finally, by simply counting the labels of all the training points which reach each leaf node we can associate each tree leaf with the empirical distributions over classes $P_{l_t(\mathbf{x})}(Y(\mathbf{x}) = c)$, where l_t indexes the leaf node in the t^{th} tree (fig. 4f). This training procedure is repeated for all T component trees.

Testing. During testing each point $\mathbf{x}$ of a previously unseen CT volume is simultaneously pushed through each of the T trees until it reaches a leaf node. Thus, the same input point $\mathbf{x}$ will end up in T different leaf nodes, each associated with a different posterior probability. The output of the forest, for the point $\mathbf{x}$, is the mean of all such posteriors, i.e. :

$$P(Y(\mathbf{x}) = c) \quad = \quad \frac{1}{T}\sum_{t=1}^{T} P_{l_t(\mathbf{x})}(Y(\mathbf{x}) = c)\,. \tag{1}$$

Other ways of combining the tree posteriors have been explored and simple averaging appears to be the most effective (as demonstrated also in the vast literature). Also, analyzing the variability of individual tree posteriors carries useful information about the uncertainty of the final forest posterior.
Organ detection. At this point detecting the presence/absence of an organ c is done simply by looking at the max probability $P_c = \max_{\mathbf{x}} P\left(Y(\mathbf{x}) = c\right)$. The organ c is considered present in the volume if $P_c > \beta$, with $\beta = 0.5$.
Organ localization. The centre of the organ c is estimated by marginalization over the volume V:

$$\mathbf{x}_c = \int_V \mathbf{x} \;\; p(\mathbf{x}|c) \;\; d\mathbf{x}, \tag{2}$$

where the likelihood $p(\mathbf{x}|c) = P\left(Y(\mathbf{x}) = c\right)$ by using Bayes rule and assuming uniform[2] distribution for organs. Furthermore, maximum a-posteriori classification for each voxel $\mathbf{x}$ may also be obtained as: $c^* = \arg\max_c P\left(Y(\mathbf{x}) = c\right)$. After having described our classification algorithm, next we provide details of the visual features employed.

2.3 Visual features and learned spatial context

The problem with identifying anatomical structures in CT images is that different organs may share similar intensity values. Thus, local intensity information is not sufficiently discriminative and further information such as texture, spatial context and topological cues must be used to have any chance of success. The problem then is how to capture and model such information efficiently.

Here we consider visual features which capture both the appearance of anatomical structures as well as their relative position (context) within the decision forest framework. For each location $\mathbf{x}$ context is modeled by integrating information coming from multiple regions which are offset by a quantity Δ in a given direction. Figures 4 explains the main concepts with a 2D illustration. A feature $\boldsymbol{\theta}$ is defined as a reference point $\mathbf{o}$ paired with two boxes F_1, F_2 and two signal channels C_1, C_2. The shapes F_i are just 3D boxes displaced with respect to $\mathbf{o}$. The channels C_i could be for example the CT intensity ($C(\mathbf{x}) = I(\mathbf{x})$), or the magnitude of the 3D gradient ($C(\mathbf{x}) = |\nabla I(\mathbf{x})|$). Given a point $\mathbf{x}$ in a volume, computing the feature response $f(\mathbf{x}; \boldsymbol{\theta})$ corresponds to aligning the reference point $\mathbf{o}$ of the feature $\boldsymbol{\theta}$ with the point $\mathbf{x}$ and computing $f(\mathbf{x}; \boldsymbol{\theta}) = \sum_{\mathbf{q} \in F_1} C_1(\mathbf{q}) - b \sum_{\mathbf{q} \in F_2} C_2(\mathbf{q})$. The parameter $b \in \{0, 1\}$ indicates whether both feature boxes are used or only one (in fig. 4 $b = 0$ for simplicity).

As shown in fig. 4 these features tends to capture the relative layout of visual patterns (e.g. kidney patterns tend to occur a certain distance away, in a certain direction, from liver patterns, fig. 4d). The use of rectangular regions enables efficient *integral volume* processing [29, 30, 16]. Our features may be thought of as a generalization of the Haar-like features used in [26, 30, 16, 17]. In fact, we do not use manually predefined Haar subdivisions of a canonical cuboid. Our classifier is free to select features with very large offsets Δ, which enables

[2] Alternatively one can weight each class based on its own volume in the training set

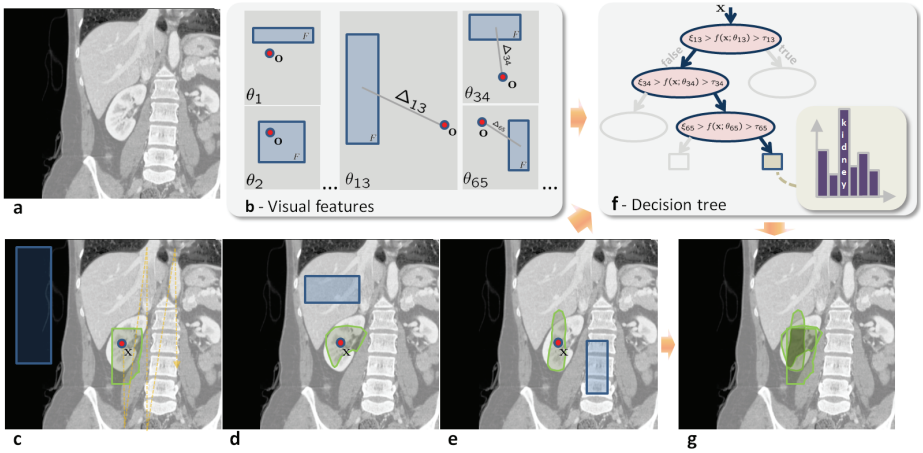


Fig. 4. Context-rich visual features, a 2D illustration. (a) Sagittal view of a patient's abdomen. **(b)** Features (denoted $\boldsymbol{\theta}_i$) are defined as the rigid pairing between a box F and a reference point $\mathbf{o}$. Here we show only some of the infinite possible features. In practice we use 3D axis-aligned boxes. **(c)** Computing the feature response $f(\mathbf{x};\boldsymbol{\theta})$ at position $\mathbf{x}$ within a volume corresponds to aligning $\mathbf{o}$ with $\mathbf{x}$ and computing the sum $f(\mathbf{x};\boldsymbol{\theta}) = \sum_{\mathbf{q}\in F} I(\mathbf{q})$ (*cf.* text. For simplicity here we use intensity as the channel and only one rectangle). For feature $\boldsymbol{\theta}_{13}$ when $\mathbf{o}$ is on the kidney the rectangle F is in a region of low density (air). Thus the value of $f(\mathbf{x};\boldsymbol{\theta})$ is small for those points. During training the algorithm will learn that feature $\boldsymbol{\theta}_{13}$ is discriminative for the position of the right kidney when associated with a small, positive value of the threshold ξ_{13} (with $\tau_{13} = -\infty$). The region for which the condition $\xi_{13} > f(\mathbf{x};\boldsymbol{\theta}) > \tau_{13}$ is true is shown in green. **(d,e)** Similar to (c) but with different features. **(f)** Training associates each node with optimal values of $\xi, \tau, \boldsymbol{\theta}$. In this example, a data point which follows the highlighted path (in blue) gets assigned a high probability of being the centre of a kidney. **(g)** The points which satisfy all three conditions in (f) lie in the intersection of the three regions (c, d, e), highlighted in dark green, inside the organ of interest.

capturing very *long-range* spatial interactions. Inspection of the trained trees reveals that often the Δ of selected features can be as large as the image width. For simplicity, in this paper we only consider intensity and gradient as channels. However, our features are more flexible and general than that as they allow to incorporate complex filters such as SIFT, HOG etc. Multiple modalities may also be exploited; e.g. in the case of MR one may use T1, T2, FLAIR etc. More complex visual cues such as the ones described in [27, 28] or differently shaped aggregation regions may also be employed.

During training, for each split node the set Θ is obtained by randomly generating for each feature the two boxes F_1, F_2 (e.g. their centre and dimensions are randomly selected) and the corresponding channels C_1, C_2. Then all nodes are optimized and once training completes the trees, their nodes and the selected features are frozen and the testing phase proceeds deterministically.

2.4 Discussion and comparisons

The classifier used here is related to the Probabilistic Boosting Tree in [16]. In our case, the tree nodes contain test functions that are simpler than the boosters used in [16], with advantages in terms of speed both during training and testing. Furthermore, as shown in [20], a collection of simple, randomized trees tends to yield better generalization than a single tree of boosters.

In [17] the authors capture context by means of an algorithm which at each iteration uses the posteriors of the previous iteration as features. This produces good results at a cost of multiple iterations. Our algorithm is not iterative and captures spatial correlations of visual patterns, namely "appearance context". Furthermore, our kernels have much longer range. Finally, we do not require preregistration of the CT volumes.

Localizing anatomical structures by atlas registration is a popular option. However, such techniques have to deal with issues such as: i) the optimal choice of degrees of freedom of the registration model (e.g. both fully rigid and fully deformable transformations are bad); ii) the optimal choice of the reference template (e.g. an adult male body? a child? or a woman? contrast enhanced or not?); and iii) robustness to anatomical anomalies (training a classifier on data which presents anomalies allows the system to learn invariance to those).

The work in [19] makes use of information gain to optimize the scheduling of single-organ boosted detectors. In our work we use information gain at the level of feature selection, and detection happens via an ensemble of decision trees simultaneously for all organs. The selected features are organized hierarchically, with the most discriminative ones in the top layers of each tree. This has the advantage of "sharing" the most discriminative features amongst classes (organs) and sets of classes, with positive effects on generalization (e.g. see [31] for details on feature sharing and [20] for a detailed comparison between AdaBoost, decision trees and decision forests). Next we quantify the performance of the proposed algorithm and compare it to some known alternatives.

3 Experimental Results and Validation

This section presents qualitative and quantitative assessment of the accuracy of our algorithm applied to the tasks of organ detection and localization.

3.1 Automatic organ detection and localization

Qualitative results are shown in fig. 5. Our classifier applied to previously unseen CT scans produces accurate posteriors for the location of organ centres. In these visualizations the computed posteriors are used to modulate the transfer function employed during 3D rendering. For instance, notice how the mass of the heart probability (in red) is correctly concentrated around the centre of the heart region. Similarly for the light brown region indicating the liver, etc.
Quantitative evaluation of accuracy. Localization accuracy is assessed here by running training and testing multiple times. In each round the database is split

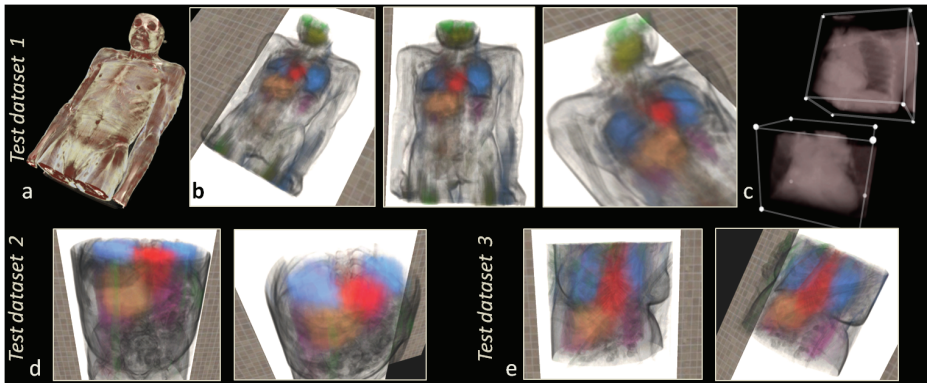


Fig. 5. Results of automatic organ detection and localization. (**a**) The original 3D CT data rendered using a manually-designed colour transfer function. (**b**) Three views of the 3D organ posterior probabilities computed by our algorithm for the localization problem. Different colours indicate different organs. Larger opacities indicate larger probability of a voxel being the organ centre. Notice how well eyes (green), head (yellow), heart, lungs, liver and even kidneys (purple) have been localized. A faint body outline is shown to aid visualization. (**c**) 3D views of the automatically detected bounding boxes including the heart and left lung. (**d,e**) Results on two more test datasets. The different datasets (related to different patients) are cropped differently and have different resolutions.

randomly into a training and a test set (with approximate ratio of 2 : 1). For all algorithms evaluated in this section the same 10 random splits are used. The forest is optimized on the training set only, and then applied on the test set. Then, the location of each organ centre is computed and compared with ground truth. Resulting localization errors collected from 10 runs (with $D = 15, |\Theta| = 500$) are shown below (in mm).

organ	`head`	`heart`	`l. eye`	`r. eye`	`l. kidney`	`r. kidney`	`l. lung`	`r. lung`	`liver`	*mean across organs*
median	25.58	18.31	24.04	25.71	13.52	29.49	22.93	21.94	19.01	**22.28 mm**
mean	29.92	21.32	28.78	27.14	25.42	44.52	27.05	26.75	22.68	28.18 mm
std	12.80	5.67	18.88	18.66	9.82	15.00	7.25	9.44	5.30	11.42 mm

Standard deviations (computed across the means of all runs) are reported here only to provide an indication of stability with respect to different train/test splits. Our algorithm achieves an overall localization error of ~ 2 cm for median. Eyes show the largest localization uncertainty across different runs (largest std), probably due to their smaller size. Furthermore, the use of larger training sets together with global position and shape priors [16] promises to improve generalization across different individuals and anatomies (e.g. missing organs etc.), and increase both the accuracy and its confidence further.

3.2 Comparisons with other algorithms

Gaussian Mixture Models. For comparison we implemented a GMM-based technique where each organ is modelled by fitting a Gaussian Mixture Model to its distribution of CT intensities. During testing, for each voxel $\mathbf{x}$ we evaluate the probability of that point being the centre of a certain organ c. The centre position is then estimated as in (2). Localization errors are reported below:

organ	`head`	`heart`	`l. eye`	`r. eye`	`l. kidney`	`r. kidney`	`l. lung`	`r. lung`	`liver`	*mean acr. organs*
median	53.48	88.54	81.56	85.59	133.04	123.38	89.32	89.59	99.63	93.79 mm
mean	144.27	98.32	174.56	168.42	125.55	128.04	104.88	100.29	98.06	126.93 mm
std	95.63	9.55	121.46	114.19	18.70	15.10	8.48	6.19	14.76	44.90 mm

The table above shows much larger errors than with our technique. An analysis of the posteriors shows that some organ labels are almost uniformly scattered spatially. This induces a bias of the detected centres towards the centre of the volume (thus incorrect), with at times low variance. The reason for such unsatisfactory results is that the GMM approach is based solely on the organs global appearance and fails to capture spatial context; and ways of integrating spatial context *efficiently* within a GMM-based approach are not straightforward. In this case the use of further features such as gradients did not seem to help much.

Template matching. We also compared our technique with a template based method. Here, each organ is represented by a set of 3D templates, extracted from the training volumes and each containing the whole organ. During testing, for each organ c we convolve the input volume with *all* exemplars for that organ and select as centre the point associated with the maximum correlation score over all exemplar templates. Localization errors are reported below.

organ	`head`	`heart`	`l. eye`	`r. eye`	`l. kidney`	`r. kidney`	`l. lung`	`r. lung`	`liver`	*mean acr. organs*
median	167.53	226.54	96.00	98.53	215.31	343.64	230.12	30.89	96.18	167.19 mm
mean	240.31	191.94	238.12	300.05	229.23	303.29	177.18	134.42	150.46	218.33 mm
std	209.08	24.16	51.33	57.20	33.09	67.33	24.41	40.03	55.15	62.42 mm

In this case the results are still worse than with our technique. We believe this is because rigid templates fail to model variations in object's shape, scale and cropping. In this case too the use of gradient features did not help. Finally, as the number of organs of interest increases having to store exemplar templates becomes prohibitive, and the processing burden shifts from training to test.

3.3 Computational efficiency

Training our decision forest model on ~ 26 datasets currently takes around 10 hours on an 8-core Intel desktop. We are planning to port the algorithm onto a High Performance Computing cluster which should reduce training to only about 1 hour. Testing is much faster. In fact, a GPU implementation (following [22]) runs in ~ 2 sec for an approximately 512^3 volume.

4 Conclusion

This paper has introduced a new algorithm for the efficient detection and localization of anatomical structures within Computed Tomography volumes.

We have presented efficient 3D visual features which capture long-range spatial context and help discrimination accuracy. Those features have been incorporated within a random decision forest classifier. The algorithm's parallel nature and the efficiency of its visual features account for the high computational efficiency. The learned model of context accounts for the good localization accuracy.

Next, we plan to extend our technique to other imaging modalities such as MRI, PET-CT and ultrasound. Also, adapting our algorithm to perform hierarchical detection (e.g. thorax $\rightarrow$ heart $\rightarrow$ mitral valve) will help dealing with detailed anatomical structures and will yield richer semantic parsing of medical images. Finally, we would like to extend our work to producing pixel-wise segmentation of complex anatomical structures such as elongated blood vessels. This will necessitate building pixel-wise annotated ground-truth databases and promises to deliver useful results.

References

1. Rubin, G.D.: Data explosion: the challenge of multidetector-row CT. European Journal of Radiology **36**(2) (2000) 74 – 80
2. Linguraru, M.G., Summers, R.M.: Multi-organ automatic segmentation in 4D contrast-enhanced abdominal CT. In: IEEE Intl. Symp. Biom. Im. (ISBI). (2008)
3. Kurkure, U., Avila-Montes, O.C., Kakadiaris, I.A.: Automated segmentation of thoracic aorta in non-contrast CT images. In: IEEE Intl. Symp. Biomedical Imaging (ISBI). (2008)
4. van Ginneken, B., Baggerman, W., van Rikxoort, E.M.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: MICCAI. (2008)
5. Shotton, J., Winn, J., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. In: IJCV. (2009)
6. Shimizu, A., Ohno, R., Ikegami, T., Kobatake, H.: Multi-organ segmentation in three-dimensional abdominal CT images. Int. J CARS **1** (2006)
7. Yao, C., Wada, T., Shimizu, A., Kobatake, H., Nawano, S.: Simultaneous location detection of multi-organ by atlas-guided eigen-orgnmethod in volumetric medical images. Int. J CARS **1** (2006)
8. Han, X., Hoogeman, M.S., Levendag, P.C., Hibbard, L.S., Teguh, D.N., Voet, P., Cowen, A.C., Wolf, T.K.: Atlas-based auto-segmentation of head and neck CT images. In: MICCAI. (2008)
9. Zhuang, X., Rhode, K., Arridge, S., Razavi, R., Hill, D., Hawkes, D., Ourselin, S.: An atlas-based segmentation propagation framework using locally affine registration – application to automatic whole heart segmentation. In: MICCAI. (2008)
10. Fenchel, M., Thesen, S., Schilling, A.: Automatic labeling of anatomical structures in MR fastview images using a statistical atlas. In: MICCAI. (2008)
11. Dolejst, M., Kybic, J., Tuma, S., Polovincak, M.: Reducing false positive responses in lung nodule detector system by asymmetric adaboost. In: ISBI. (2008)

12. Pescia, D., Paragios, N., Chemouny, S.: Automatic detection of liver tumors. In: ISBI. (2008)
13. Wels, M., Carneiro, G., Aplas, A., Huber, M., Hornegger, J., Comaniciu, D.: A discriminative model-constrained graph-cuts approach to fully automated pediatric brain tumor segmentation in 3D MRI. In: MICCAI. (2008)
14. Freiman, M., Edrei, Y., Shmidmayer, Y., Gross, E., Joskowicz, L., Abramovitch, R.: Classification of liver metastases using fMRI images: A machine learning approach. In: MICCAI. (2008)
15. Prasad, M., Sowmya, A.: Multi-level classification of emphysema in HRCT lung images using delegated classifiers. In: MICCAI. (2008)
16. Tu, Z., Narr, K.L., Dollar, P., Dinov, I., Thompson, P.M., Toga, A.W.: Brain anatomical structure segmentation by hybrid discriminative/generaltive models. IEEE Trans. on Medical Imaging **27**(4) (2008)
17. Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Toga, A.W., Thompson, P.M.: Automatic subcortical segmentation using a contextual model. In: MICCAI. (2008)
18. Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., Wells, W.M.: A hierarchical algorithm for MR brain image parcelation. IEEE Trans. on Medical Imaging **26**(9) (2007)
19. Zhan, Y., Zhou, X.S., Peng, Z., Krishnan, A.: Active scheduling of organ detection and segmentation in whole-body medical images. In: MICCAI. (2008)
20. Yin, P., Criminisi, A., Essa, I., Winn, J.: Tree-based classifiers for bilayer video segmentation. In: CVPR. (2007)
21. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: IEEE ICCV. (2007)
22. Sharp, T.: Implementing decision trees and forests on a GPU. In: ECCV. (2008)
23. Breiman, L.: Random forests. Technical Report TR567, UC Berkeley (1999)
24. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. Neural Computation **9** (1997) 1545–1588
25. Quinlan, J.R.: C4.5: Programs for Machine Learning. (1993)
26. Viola, P., Jones, M.J.: Robust real-time face detection. IJCV (2004)
27. Dalal, N., T.B.: Histograms of oriented gradients for human detection. In: IEEE CVPR. (2005)
28. Zambal, S., Buehler, K., Hladuvka, J.: Entropy-optimized texture models. In: MICCAI. (2008)
29. Crow, F.C.: Summed-area tables for texture mapping. In: SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques, New York, NY, USA, ACM (1984)
30. Viola, P., Jones, M.J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. In: ICCV. (2003)
31. Torralba, A., Murphy, K.P., Freeman, W.T.: Sharing visual features for multiclass and multiview object detection. IEEE Trans. PAMI (2007)

Multi-Object Detection and Segmentation of Brain Structures Based on Dynamic Programming

Jue Wu and Albert C.S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Hong Kong

Abstract. This work aims to design a detection and segmentation method using a graphical model in the context of multi-object brain image segmentation. We resort to dynamic programming as the optimization strategy to find the global minimum energy for the relation graph. Compared to other graphical models like tree structures, the proposed approach offers flexibility in accommodating more interactions among objects and thus can inhibit error propagation. Also, the new method is able to detect and segment a larger number of objects by searching for the global optimum energy in an efficient way. Experimental results show that the proposed approach achieves a comparable accuracy to other state-of-the-art methods.

1 Introduction

The goal of anatomical segmentation of human brains is to divide a brain image into anatomical subregions, such as, tissues and structures. Structure segmentation is more difficult than tissue extraction because a structure is usually a further segmentation of a tissue. However, structure segmentation plays a significant role in the study of human brain functioning and brain disease diagnosis [1–3]. Various methodologies have been proposed to tackle the difficult problem of brain structure segmentation [4–8]. Among others, tree-based methods [9–11] are emerging as an interesting way to manage the inter-relations between structures/objects. The weakness of a tree structure lies in the relatively simple interactions among nodes or objects. As a connected graph, a tree has the fewest edges, which equals the number of objects minus one. If more complicated interaction is desirable, then a tree structure is not satisfactory because no loop is allowed in a tree. On the other hand, the spatial relation among the brain structures is complex. In some cases, three or more structures are gathered together so it is hard to establish a hierarchical relationship like a tree among these structures.

In this paper, we present a novel dynamic programming-based graphical model to deal with brain structure segmentation. Dynamic programming (DP) is able to downsize the relation graph to a simple enough graph. As such, the

relation graph can be complicated with many edges as long as DP is applicable. This overcomes the shortcoming of a tree graph with very limited number of edges. Furthermore, the DP-based method excels if the number of objects grows significantly. In this scenario, the tree-based methods tend to propagate errors along a single tree path due to the limited interactions. Note that there is only one path between any two nodes in a tree. Another important advantage of the dynamic programming-based method is the ability to find the global optimum in an efficient way. During the downsizing of the graph, DP keeps the records of the best candidates of the eliminated nodes and incorporates the associated energy in the subgraph. When the subgraph is optimized, the original graph is also optimized. The computational complexity of DP is exponentially lowered compared to an exhaustive search.

2 Methodology

2.1 Multi-Object Template Construction

Without prior information of shape and position, the task of extracting a pool of brain structures is very difficult since the structures often do not have clear boundaries in the brain. The template or atlas of multiple structures in the brain is served as the prior information for the detection and segmentation. It contains two-fold information. First, what are the shapes of the target structures? Second, what are the relations of position among these structures?

A legitimate template can be constructed from the ground truths of a set of segmentations. Multiple descriptions of the shapes and relative positions of the structures can be combined by a certain strategy. In this paper, we adopt a quick and simple strategy to construct the template so we use the ground truth of one segmentation. The ground truth delimitates the shape of each structure and determines position relations among the structures. We assume that different subjects have similar profiles in the two aspects (shape and position).

We create a graph to further define the relations among the structures. The vertices in the graph represent objects/structures. If two objects interact, there is an edge connecting the corresponding vertices. The weight of a vertex is pertaining to the similarity between the object and the superimposed image. The weight of an edge reflects the relative position between two interactive objects. The aim is to minimize the total weights associated with the whole graph. We create a graph that is suitable for dynamic programming to apply. The requirement is that the graph can be reduced into a smaller graph with fewer vertices. When the original graph is optimized, the smaller graph also reaches an optimum. We call such a graph a DP-graph. The vertices in a DP-graph can be eliminated one by one on a certain order until a manageable size of vertices is remained. After one vertex is eliminated, all the weights associated to it are incorporated to the smaller graph through a bookkeeping of the best candidates.

Theoretically, dynamic programming can be applied to many kinds of graphs. In the current context, we apply it to a triangulated graph which strikes a

good balance between complexity and efficiency [12]. We target to detect and segment 12 structures in the brain and the constructed graph has 15 vertices. Three structures (lateral ventricle, the third ventricle and the fourth ventricle) are segmented in advance and serve as the base objects. The other objects are left/right thalamus, caudate, putamen, pallidum, hippocampus and amygdala. The interactions among them are shown in Fig. 1. We build these edges according to the spatial proximity of the objects.

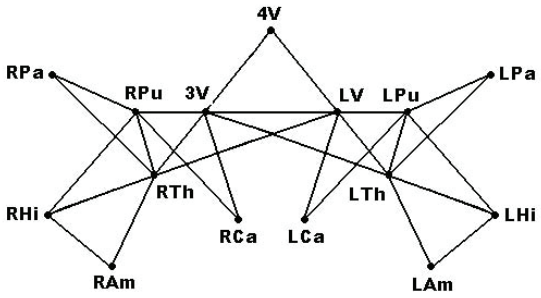


Fig. 1. The relation graph for 15 brain structures. LV: lateral ventricle, 3V: the third ventricle, 4V: the fourth ventricle, L/RTh: left/right thalamus, L/RCa: left/right caudate, L/RPu: left/right putamen, L/RPa: left/right pallidum, L/RHi: left/right hippocampus, L/RAm: left/right amygdala.

2.2 Matching Measures

We formulate the matching process between a template and an image as an optimization process of a specific energy. Each objects in the template can be superimposed on the image in several candidate positions. As such, the multi-object template has many configurations based on the combination of different positions of each object. We design an energy to measure both the similarity between each object and the underlying image region and the proper distances among the objects. The energy is composed of two components. One is the single-object measure and the other is the measure for two interactive objects. The two measures supplement each other. The unary measure is used to draw one object to the position where a brain structure exists probably. The binary measure regulates the inter-position between two interactive objects. The total energy is expressed by,

$$E_T(P) = \alpha \sum_i E_u(p_i|I) + \sum_i \sum_{j \in N_i} E_b(p_i, p_j), \tag{1}$$

where $E_T(P)$ denotes the objective total energy of the model, E_u is a unary measurement for a single object, p_i is the position of part i represented by

spatial coordinates, E_b measures a compatibility between two objects, P is the set of possible positions of all objects, I is the input image, and N_i is the set of objects linking to object i.

The unary measure assumes that (1) the intensity level inside one object is approximately homogeneous and its variation should be small; (2) The voxels around the boundary of the object usually undergo an abrupt change of intensity and the gradient magnitude therein should be large. The two assumptions are valid considering that the brain structures are usually made of the same tissue type, e.g. white matter or grey matter, and they are often surrounded by a different tissue although the boundary may be lowly contrasted. The premises lead to the following formulation for the unary measures:

$$E_u(p_i|I) = \frac{1}{a_i}\iiint_{A_i}(I(x,y,z)-\overline{I}(A_i))^2 dxdydz - \frac{\beta_i}{b_i}\iiint_{B_i}|\nabla I(x,y,x)|dxdydz, \tag{2}$$

where $\overline{I}(A_i)$ is the mean intensity in the image region A_i that is overlapped with object i and a_i is the volume of the region A_i. B_i is the boundary of object i, b_i is the area of the boundary B_i and $I(x,y,z)$ is the intensity of the voxel at coordinate (x,y,z).

The binary measure adjusts the relative position of two objects that are connected by one edge in the graph. It is content free and does not depend on the image data. When two objects are too close or too faraway compared to the beginning status, the measure imposes a large penalty. If two objects have a significant overlapping, the unreal configuration is also punished by the measure. The binary measure is defined as,

$$E_b(p_i,p_j) = ||(C_{p_i}-C_{p_j})-(\hat{C}_{p_i}-\hat{C}_{p_j})|| + L\cdot I\{a_{ij}/a_j > \tau\}, \tag{3}$$

where C_{p_i} is the coordinate vector of the center of mass of object i at position p_i and $\hat{C}_{p_i}$ is the initial center of mass of object i. $||\boldsymbol{v}||$ denotes the norm of vector $\boldsymbol{v}$. $I\{\text{event}\}$ is an indicator function, i.e., if the event is true, $I=1$; otherwise $I=0$. a_j is the volume of object j, a_{ij} is the volume of the overlapping region between objects i and j, and τ is a tolerance factor with respect to the overlapping extent. L is a very large positive penalty.

The parameters α and β determine the weights between terms in the energy. If we allow loose interaction among objects, we can set α to a large value. β helps to reduce the difference in magnitude between the gradient and the variance. It can be set according to the ground truth of the training image as follows,

$$\beta_i = (\frac{1}{a_i^t}\iiint_{A_i^t}(I(x,y,z)-\overline{I}(A_i^t))^2 dxdydz)/(\frac{1}{b_i^t}\iiint_{B_i^t}|\nabla I(x,y,x)|dxdydz), \tag{4}$$

where x^t represents the same quantity $x(x=b_i,A_i,a_i,B_i)$ as aforementioned except with the known position of object i superimposed on the training image.

2.3 Optimization by Dynamic Programming

Once the energy is defined, we can then move the template around the input image to find the optimal position for each object. In the process we regard that positions with smaller energy are better than those with larger energy. The optimization is challenging since the search space is huge. Suppose there are n objects in a triangulated graph and for each object there are k possible positions to choose. If we enforce exhaustive search for a global minimum, we need to calculate the energy k^n times. However, if we apply dynamic programming to the optimization, the energy is calculated only $(n-2)k^3$ times with increased spatial requirement.

Dynamic programming (DP) is a strategy that solves a problem by downsizing. A problem suitable for DP can usually be formulated in a recursive way. DP simplifies the problem by solving a less complicated sub-problem first. When the original problem is optimized, the sub-problem is also optimized on a smaller support. Dynamic programming does the bookkeeping and records the optimal extension from the sub-problem to the original problem. The downsizing continues until a sub-problem with a manageable scale is obtained and can be solved easily. Afterwards, the solution of the smallest sub-problem is used to track back to the solution of the original problem.

Dynamic programming can be applied to problems in various domains. In the present context, a graph is eligible if there is always a cell (a connected subgraph with 3 or more vertices) that has only one shared edge (we call it an anchor edge) with the rest part of the graph before each downsizing. All the edges except the anchor edge are exclusive to the cell. We call such a cell a dangling cell. The simplest dangling cell is a triangle. It is not necessary that the graph has only one kind of cell. The DP strategy is to collapse the dangling cell to the anchor edge. The best configurations in the dangling cell for every possible configuration of the two vertices of the anchor edge are memorized. Collapse the cells one by one and the original graph is downsized to a much smaller scale. In each collapse, the vertices in the dangling cell are eliminated except for the two vertices of the anchor edge.

Take the current setting shown in Fig. 1 as an example. The cells are all triangles. There are more than one dangling cells in the present iteration, e.g, LAm-LHi-LTh, LPa-LPu-LTh, LV-LCa-LPu, and their anchor edges are LTh-LHi, LPu-LTh and LV-LPu, respectively. We can choose any of them to start collapsing. The sequence of vertices to be eliminated is called the elimination order. We adopt for the graph an elimination order of (LAm, LHi, LPa, LCa, LPu, LTh, RAm, RHi, RPa, RCa, RPu, RTh). After the 12 vertices are eliminated, the graph is downsized to a single triangle which can be simply optimized by brute force. The bookkeeping makes it possible to transit from a large graph to a smaller one without loss of information. The bookkeeping is the process of memorizing the best position for an eliminated vertex given the positions of the vertices of the anchor edge in the same dangling cell. Note that all the k^2 position combinations of the vertices of the anchor edge need to be kept for later use. For example, during the elimination of vertex LAm, DP memorizes the

best position for each position combination of LHi and LTh. The corresponding weights for each configuration are also registered, i.e. the sum of three weights on vertex LAm, on edge LAm-LHi and edge LAm-LTh. Then, replace the old weight on the anchor edge LHi-LTh with the above sum after deleting LAm and its edges. Thus a smaller subgraph is generated. The iteration continues until three vertices 4V, 3V and LV are left. We assume the true configuration of the smallest subgraph is easily obtained. The backtracking from the last triangle to the other ones is achievable thanks to the bookkeeping during downsizing. We now know the best choices of the positions of some initial vertices, e.g. 4V, 3V and LV. In the reverse order of elimination, we can find the best position of the other vertices. For instance, the optimal position of RTh can be retrieved since it is recorded in the bookkeeping of 3V and LV. 3V, LV and RTh are once in the same dangling cell during the elimination and recall that the best position of RTh is kept for every possible position combination of 3V and LV during the bookkeeping process. The backtracking is then repeated for RPu, RCa and so on. In the end, the optimal positions for all the eliminated vertices are found.

The pseudocode of the dynamic programming for the graph is shown in Algorithm 1. The notations are as follows. N is the number of objects in the graph. s is the number of possible poses for each object. $\texttt{Edge}(p, q, i)$ is the optimal energy accumulated on edge i when the two objects a and b linked by edge i have poses of p and q. $\texttt{Adj}(i, j)$ is the index (the 3rd dimension in array `Edge`) of the edge between objects i and j, if i and j are adjacent. $\texttt{Adj}(i, j) = 0$ otherwise. $\texttt{ElimOrder}(i)$ is the ith object to eliminate. $\texttt{Tri}(i, 1), \texttt{Tri}(i, 2)$ and $\texttt{Tri}(i, 3)$ are the three objects indices in triangle i. $\texttt{Tri}(i, 1)$ is the first to eliminate. $\texttt{Triangle}(p, q, i)$ is the optimal pose of the object $\texttt{Tri2Obj}(i)$ which is the first-to-eliminate object in triangle i and the two other objects a and b in triangle i have poses of p and q. $\texttt{Obj2Tri}(i)$ is the index (the 3rd dim in array `Triangle`) of the triangle which the first-to-eliminate object is i.

It is worth pointing out that though the current setting of triangular collapsing is similar to some existing methods [13, 14], the dynamic programming-based framework can be applied to more complicated situations since the dangling cell can be far more complex than a triangle.

After the dynamic programming is finished, we fine-tune the coarse segmentation by means of the non-rigid B-spline registration between the input image and the moved trained image [10]. The moved training image has been deformed according to the displaced template obtained from the DP process. The similarity metric for the B-spline registration is mutual information. The final segmentation is thus obtained by propagating the registered template label to the input image.

3 Experiment

In the experiments, we applied the proposed method to the public database, IBSR [15], which contains 18 sets of T1-weighted brain MR images with expert-segmented internal structures. The bias-corrected images are of $256 \times 256 \times 128$

Algorithm 1 : Dynamic programming for a triangulated graph

Require: `ElimOrder`, `Edge`, `Tri`, `Triangle`, `Obj2Tri`
Ensure: `pos`, `optimal`
  for each edge v **do**
    for $s_1 = 1, \cdots, s$ **do**
      for $s_2 = 1, \cdots, s$ **do**
        $\texttt{Edge}(s_1, s_2, v)$ = the binary measure for one object in position s_1 and the other object in position s_2 (the two objects are linked by edge v);
      end for
    end for
  end for
  for $i = 1, \cdots, n-3$ **do**
    $c = \texttt{ElimOrder}(i), a = \texttt{Tri}(\texttt{Obj2Tri}(c), 3), b = \texttt{Tri}(\texttt{Obj2Tri}(c), 2)$;
    for $p = 1, \cdots, s$ **do**
      for $q = 1, \cdots, s$ **do**
        `minEnergy` = a large number;
        for $r = 1, \cdots, s$ **do**
          e = accumulated energy on the edge linking a and c + accumulated energy on the edge linking b and c + unary energy for object c in position r;
          if $e <$ `minEnergy` **then**
            $\texttt{minEnergy} = e$;
            $\texttt{Triangle}(p, q, \texttt{Obj2Tri}(c)) = r$;
          end if
        end for
        Increase $\texttt{Edge}(p, q, \texttt{Adj}(a, b))$ by `minEnergy`;
      end for
    end for
  end for
  $\texttt{optimal} = \min_{p,q}(\texttt{Edge}(\texttt{p}, \texttt{q}, \texttt{1}))$;
  for $i = n-3, \cdots, 1$ **do**
    $c = \texttt{ElimOrder}(\texttt{i})$;
    $t = \texttt{Obj2Tri}(c)$;
    $\texttt{pos}(c) = \texttt{Triangle}(\texttt{pos}(\texttt{Tri}(t, 2), \texttt{pos}(\texttt{Tri}(t, 3), t)$;
  end for
  Return `pos` and `optimal`;

voxels with various voxel sizes. Some of the target structures in the data sets are of low contrast and hard to detect even for human visual inspection.

Before executing the algorithm, we set the parameters of the method once and they were fixed in all the tests. α was set to 4 since the average degree of the vertices in the graph is around 4. We believe the overlapping of more than 10% of the object region is unacceptable so τ was set to 0.1. The grid size of the third-order B-spline was $5 \times 5 \times 5$. The search space for object positions was set to be within the range of $5 \times 5 \times 3$ voxels from the initial positions and hence $k = 75$ poses were explored for each object/structure. The magnitude of the whole search space for 12 objects was 75^{12}. The segmentations of 4V, 3V and LV were acquired from the training. Alternatively, we can adopt an automatic method to segment the ventricles as in [10]. We adopted the commonly-used Dice metric to quantify the difference between the segmentation results and the ground truths. The Dice score is defined as Dice $= (2||A \bigcap B||)/(||A|| + ||B||)$, where A and B are two shapes to be compared. Dice equal to 1 means a perfect match between two shapes and Dice equal to zero means no overlapping. The larger the Dice score is, the more overlapping the two shapes have.

We randomly chose a subject in the data sets as the only template. Then the proposed method was tested on the other 17 data sets. The Dice scores for the segmentation results before and after the application of the proposed approach are shown in Fig. 2. Here the initial segmentations are the regions superimposed by the template before we apply the proposed method. It is observed that the proposed method improves the segmentation of all structures by more than 10 percentage points on average. When the initial segmentations have a large variance, the DP-based method can decrease the variance and obtains more consistent results. The experiments were run in MatLab codes on a 2.13 GH CPU with 1 GB memory. The average running time is around 74 minutes per data set. We also show one example of the original image, ground truth and the segmentation result in Fig. 3.

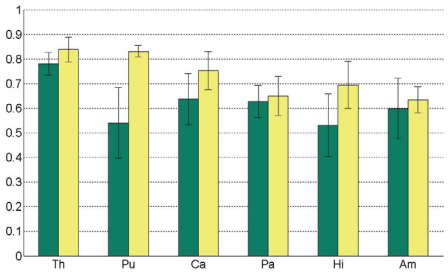


Fig. 2. The average Dice scores and their standard deviations for the initial segmentation (left green bars) and the final segmentation (right yellow bars). Th: thalamus, Pu: putamen, Ca: caudate, Pa: pallidum, Hi: hippocampus, Am: amygdala. Results of the same type of left and right structures are combined.

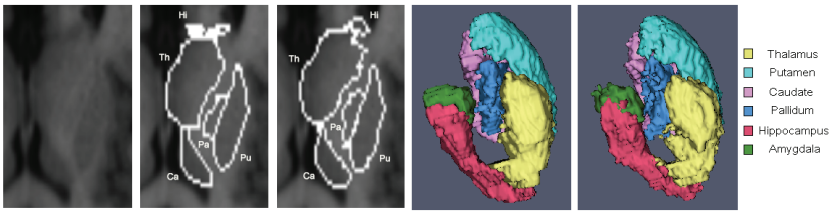


Fig. 3. One slice of a data set and corresponding segmentations. From left to right are original image, ground truth, segmentation result, 3D visualization of the ground truth and 3D visualization of the segmentation result.

4 Discussion and Summary

The same as other graphical models [10], the proposed method has some advantages over other methodologies, such as level sets, fuzzy logic and registration-based segmentation. For example, the graphical methods are usually more efficient than registration-based segmentation because the former processes the structures directly while the latter makes use of the whole brain information. Level set methods require the initialization to be close to the target. Otherwise, the level sets will converge to irrelevant structures in the brain. The proposed method circumvents the initialization problem by finding the global optimum through dynamic programming.

Quantitative comparison of the proposed method with the related methods is not conclusive because the methods were tested on different databases, targeted different sets of structures, or evaluated by different similarity metrics. The proposed method performed slightly better than a related method that was also tested on IBSR [16]. Their Dice scores for caudate, hippocampus, amygdala, putamen and pallidum were 0.76, 0.67, 0.63, 0.78 and 0.71. Ours are 0.76, 0.70, 0.64, 0.83 and 0.65. Generally speaking, in the related methods [5–7], the Dice scores for caudate, putamen and thalamus ranged from 0.75 to 0.90. The scores for pallidum, hippocampus and amygdala ranged from 0.60 to 0.75. The proposed method achieved comparable accuracies.

One limitation about the current method is that although the dynamic programming is capable of finding the global minimal energy, the energy formulation itself is not perfect. It is not ensured that the global minimal energy coincides with the true structure positions. It is an even more complicated case considering that the template possesses moderately different shapes from the true shapes. Theoretically, during the position search, we can allow sophisticated shape search simultaneously but this will make the computational burden formidable. It remains a profound future topic how to design a nearly-perfect energy to accommodate various shape changes while enabling efficient optimizations. All in all, the current work upgrades previous methods [10, 11] based on

simple graphs (e.g. trees) to a more advanced method. This DP-based method allows more interaction in the relation graph and thus remains stable and robust to error propagation in spite of more objects.

References

1. Grenander, U., Miller, M.: Computational anatomy: An emerging discipline. Quarterly of Applied Mathematics **56** (1998) 617 – 694
2. Kandel, E., Schwartz, J., Jessell, T.: Principles of Neural Science. 4 edn. New York: McGraw-Hill (2000)
3. Packard, M., Knowlton, B.: Learning and memory functions of the basal ganglia. Annual Review Neuroscience **25** (2002) 563 – 593
4. Ciofolo, C., Barillot, C.: Brain segmentation with competitive level sets and fuzzy control. In: Information Processing in Medical Imaging. Volume LNCS 3565. (2005) 333 – 344
5. Zhou, J., Rajapakse, J.: Segmentation of subcortical brain structures using fuzzy templates. NeuroImage **28** (2005) 915 – 924
6. Khan, A., Wang, L., Beg, M.: Freesurfer-initiated fully-automated subcortical brain segmentation in mri using large deformation diffeomorphic metric mapping. NeuroImage **41** (2008) 735 – 746
7. Powell, S., Magnotta, V., Johnson, H., Jammalamadaka, V., Pierson, R., Andreasen, N.: Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. NeuroImage **39** (2008) 238 – 247
8. Tu, Z., Narr, K., Dollar, P., Dinov, I., Thompson, P., Toga, A.: Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE Trans. on Medical Imaging **27**(4) (2008) 495 – 508
9. Pohl, K., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R., Kikinis, R., Grimson, W., Shenton, M., Wells, W.: A hierarchical algorithm for mr brain image parcellation. IEEE Trans. Medical Imaging **26** (2007) 1201 – 1212
10. Wu, J., Chung, A.: Markov dependence tree-based segmentation of deep brain structures. In: MICCAI. Volume LNCS 5242. (2008) 1092 1100
11. Wu, J., Chung, A.: A novel framework for segmentation of deep brain structures based markov dependence tree. NeuroImage **46** (2009) 1027 – 1036
12. Amit, Y.: 2D Object Detection and Recognition. The MIT Press (2002)
13. Felzenszwalb, P., Huttenlocher, D.: Efficient matching of pictorial structures. In: Proc. IEEE Computer Vision and Pattern Recognition Conf. (2000) 66–73
14. Felzenszwalb, P., Huttenlocher, D.: Representation and detection of deformable shapes. IEEE PAMI **27**(2) (2005) 208–220
15. Internet Brain Segmentation Repository: www.cma.mgh.harvard.edu/ibsr/.
16. Gouttard, S., Styner, M., Joshi, S., Smith, R., Cody, H., Gerig, G.: Subcortical structure segmentation using probabilistic atlas priors. In: SPIE Medical Imaging. Volume 6512. (2007) 65122j–11

Probabilistic Sequential Segmentation and Simultaneous On-Line Shape Learning of Multi-Dimensional Medical Imaging Data

J. Chiverton[1], X. Xie[2], and M. Mirmehdi[3]

[1] School of Information Technology, Mae Fah Luang University, 57100, Thailand
jpchiverton@gmail.com
[2] University of Wales Swansea, Department of Computer Science, Swansea, SA2 8PP, UK
x.xie@swansea.ac.uk
[3] Department of Computer Science, University of Bristol, Bristol, BS8 1UB, UK
majid@cs.bris.ac.uk

Abstract. Accurate automatic segmentation of anatomical structures is usually considered a difficult problem to solve because of anatomical variability and varying imaging conditions. A prior description of the shape of the anatomical structure to be segmented can reduce the ambiguity associated with the segmentation task. However this prior information has to be prepared specifically for the structure of interest, usually supervised and under favorable imaging conditions. An alternative is to consider the shape of the object sequentially, along a particular dimension of the data. This is the approach taken here, i.e. on-line modeling of sequential shape information which is combined with sequential segmentation of the intensity distributions for the segmented structure and the surrounding region.

1 Introduction

A clinical Magnetic Resonance (MR) scan of a tumor in a human brain contains pixels that represent pathological and non-pathological tissues. The structure of these tissues can be considered on a slice by slice basis, i.e. in a sequential segmentation process. Potential benefits of sequential segmentation include adaptation to changes in the intensity distributions due to e.g. inhomogeneities in the magnetic field and adaptation to other unknown variations e.g. structural. Medical imaging data is often multi-dimensional (including 3 spatial dimensions and a time dimension). The medical analysis of such data is complex, requiring advanced medical and information technology knowledge and skills, see e.g. [1, 2]. Furthermore sequential segmentation approaches require considerably less memory in comparison to segmentation techniques that process large multi-dimensional data sets non-sequentially.

Sequential segmentation and summarization processes may ultimately be medically useful. A sequential segmentation process can be used to sequentially estimate pathological structures such as tumors or even non-pathological structures such as cardiac tissues that may or may not be defined over time. Shape is inherent to the sequential segmentation process, whether implicitly considered (i.e. for initialization purposes on subsequent image slices) or explicitly. We consider an explicit model of shape where

the shape information is not only estimated from the segmentation process but also used in subsequent estimation steps to assist with subsequent image slice segmentations. We consider shape to be an important factor, providing additional information for an intelligent segmentation framework.

An ideal medium for shape modeling are active contour models which have been extensively investigated in conjunction with prior shape knowledge since at least [3, 4]. These spline based approaches are limited by topological constraints unlike level set based active contour approaches that have been extensively investigated with the use of prior shape knowledge since [5]. PCA is often used in these techniques to compress and summarize the important components of a set of characteristic level sets [5, 6] or control points modeled using Active Shape Modeling (ASM) [7]. Many prior shape based segmentation methods have been demonstrated to be quite robust, providing accurate outlines of the shape of the object being segmented. However, preparation of extensive prior shape knowledge is not always convenient and even cumbersome. Furthermore, many methods can encounter difficulties if the structure of interest assumes a new shape that can not be easily approximated by the shape model. This is a realistic problem for medical images that are associated with an inherent variability associated with population differences and potential pathological structures. Sequential estimation techniques, requiring minimal user assistance are a valid alternative where shape and segmentation information is propagated slice to slice.

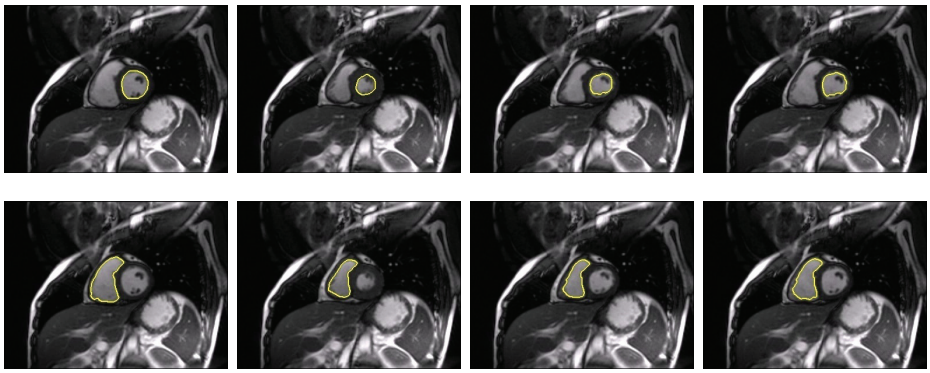


Fig. 1. Example of temporal sequential segmentation in the sagittal plane. Top row shows left atrium outline and bottom row, outline of the right atrium. Images correspond to a cardiac MRI scan acquired at Bristol Royal Infirmary. The algorithm was initialized in both cases by a manually defined region on an initial slice in the sequence for each result.

Sequential segmentation, results illustrated in figure 1 using the methodology described here, has previously been considered by a limited number of authors. An interesting paper by Cho et al. [8] described ways to augment the sequential segmentation process via physical measurements inherent in the MRI scanning process. However the additional physical information is often not available. Senegas et al. [9] utilized se-

quential segmentation techniques for cardiac sequences where shape information was propagated across the temporal dimension from a manually positioned mean location, however the authors utilized a statistical prior model for cardiac shape. Vaswani et al. [10] considered sequential segmentation techniques for both medical data and conventional video. Initialization was provided by manually placed geometric objects with manipulation of parameter settings. The included results appear to be rough estimates of the structures of interest.

We also consider shape in a sequential segmentation framework, where a structure of interest is manually defined for a single 2D image slice. Empirical parameter estimation is also undertaken for the model components. Section 2 describes our methodology, uniquely outlining a model for the sequential shape information in an on-line non-linear setting which is embedded in an image model that considers image information and the overall level set gradient descent based optimization process. Section 3 then presents exemplar results from spatial and temporal sequential estimation processes for pathological and non-pathological structures, respectively. Section 4 closes the work with a short discussion.

2 Methodology

The model of the shape information is described next, in Section 2.1. Section 2.2 then describes the important elements of an image model for n dimensional intensity information which also provides an ideal medium in which to embed the shape model defined earlier. Section 2.3 then summarizes the optimization process used to sequentially estimate the bounds of the structure of interest.

2.1 Shape model

The primary focus of the work here is the shape of a structure of interest S^j that evolves from one image slice j to the next $j+1$. This shape S^j is synonymous to a partition of the image space into two mutually exclusive regions, consisting of foreground $\mathfrak{F}^j = \{\mathbf{x}|\mathfrak{f}^j_{\mathbf{x}}\}$ and background $\mathfrak{B} = \{\mathbf{x}|\mathfrak{b}^j_{\mathbf{x}}\}$ pixels $\mathbf{x}$, where $\Omega^j = \mathfrak{F}^j \cup \mathfrak{B}^j$ and $\mathfrak{f}^j_{\mathbf{x}} \in \{0\ 1\}$ and $\mathfrak{b}^j_{\mathbf{x}} = 1 - \mathfrak{f}^j_{\mathbf{x}}$ are binary foreground and background labels, respectively. The partition of the image space is referred to here with $q^j = \{\mathfrak{F}^j\ \mathfrak{B}^j\}$ for image slice j. The foreground pixels $\mathfrak{F}^j$ define the structure of interest for image slice j.

We consider the level set $\phi^j \triangleq \mathsf{S}^j$ as the primary representation of shape information in our model. ϕ^j enables important geometric information to be conveniently incorporated into the modeling process and it can be considered synonymous to the image partition q^j where pixel level labeling information is fully encapsulated by the level set representation. This can be seen from the properties of the level set which include: $\phi^j_{\mathbf{x}_c} = 0$ on the coterminous foreground and background regions for contour points $\mathbf{x}_c$ and $\phi^j_{\mathbf{x}} = \pm \min |\mathbf{x} - \mathbf{x}_c|\ \forall \mathbf{x}_c | \phi^j_{\mathbf{x}_c} = 0$, i.e. the contour point with minimum Euclidean distance, see e.g. [11]. Also (here in this work) $\phi^j_{\mathbf{x}} \leq 0$ for $\mathbf{x} \in \mathfrak{F}^j$ and $\phi^j_{\mathbf{x}} > 0$ for $\mathbf{x} \in \mathfrak{B}^j$. The shape ϕ^i of the structure of interest for the current image slice can be controlled via comparisons with a set of shapes Φ^{i-1} from a dynamically built space of good shape hypotheses from previous image slices. The comparison of ϕ^i with Φ^{i-1}

should be invariant to translation $\mathbf{T}_{\mathrm{s}}^{i}$, scale $\mathbf{s}_{\mathrm{s}}^{i}$ and rotation $\mathcal{R}_{\mathrm{s}}^{i}$ to enable meaningful comparison resulting in a normalized shape space Ω_{s} representation, where the subscript s refers to the normalized shape space. Thus, Φ^{i-1} and ϕ^{i} have equivalent shape space forms given by $\Theta_{\mathbf{x}}^{i-1} = \Phi_{\mathcal{A}_{\mathrm{s}}(\mathbf{x})}^{i-1}$ and $\theta_{\mathbf{x}}^{i} = \phi_{\mathcal{A}_{\mathrm{s}}(\mathbf{x})}^{i}$ $\forall\ \mathbf{x}$ where

$$\mathcal{A}_{\mathrm{s}}^{i}(\mathbf{x}) = \mathbf{s}_{\mathrm{s}}^{i}\mathcal{R}_{\mathrm{s}}^{i}\mathbf{x} + \mathbf{T}_{\mathrm{s}}^{i} \tag{1}$$

is the similarity transformation from image space $\mathbf{x} \in \Omega$ to shape space $\mathcal{A}_{\mathrm{s}}^{i}(\mathbf{x}) \in \Omega_{\mathrm{s}}$ for the object shape in image slice i. Shape comparisons also have to be in the current image space requiring the inverse transformation of (1), i.e. $\mathcal{A}^{i}(\mathbf{x}) = \mathbf{s}^{i}\mathcal{R}^{i}\mathbf{x} + \mathbf{T}^{i}$ where $\mathcal{A}^{i}(\mathbf{x}) \in \Omega$. As the shape information is learned on-line, without supervision, the resulting estimated shapes will not be perfect representations and hence can be considered to be inherently noisy. Thus we may define a probabilistic shape space with probability distribution $p_m(\theta|\mathcal{M}^{i-1})$ that represents the distribution of the learned noisy shapes over a normalized shape space $\mathcal{M}^{i-1} = \{\theta^{j}|0 \leq j \leq i-1\}$ consisting of shapes up to the current image slice. We can then define a locally weighted shape space expectation $\bar{\Theta}^{i-1}$ to provide a best estimate over the shape distribution $p_m(\theta|\mathcal{M}^{i-1})$ (which acts as a prior) and a local weighting distribution $p_w(\theta|\theta^{i-1})$. This best estimate can then be used to compare the currently evolving shape rather than a global mean or one based on assumptions on the linearity of the shape space or even one based on local integrity. The local weighting is given here by a Gaussian distribution $p_w(\theta|\theta^{i-1})$ centered on the previous image slice object shape θ^{i-1}, so that[4]

$$\bar{\Theta}^{i-1} = \int_{\theta} \theta\ p_m(\theta|\mathcal{M}^{i-1})\ p_w(\theta|\theta^{i-1})\ \mathrm{d}\theta. \tag{2}$$

θ^{j} for $j = 0...i-1$ are considered to be distributed according to $p_m(\theta|\mathcal{M}^{i-1})$, so that the expectation is approximated via

$$\bar{\Theta}^{i-1} = \frac{\sum_{j=0}^{i-1}\left[\theta^{j}\mathcal{W}^{i-1,j}\right]}{\sum_{j=0}^{i-1}\mathcal{W}^{i-1,j}} \tag{3}$$

where $\mathcal{W}^{i-1,j} = \exp(\frac{1}{|\Omega_{\mathrm{s}}|}\sum_{\mathbf{x}\in\Omega_{\mathrm{s}}}(\theta^{i-1} - \theta^{j})^2)$ is the local weighting for θ^{i-1} and θ^{j} image slice shapes and previously identified object shapes are θ^{j} for image slice j. Each weight can then be considered to form an element in a weight matrix that encompasses the similarity of shapes at different image slices. A simple comparison between ϕ^{i} and Φ^{i-1} can then be the sum of squared differences in the current image space:

$$\mathcal{C}_{\mathfrak{s}}(\phi^{i}, \Phi^{i-1}) = \sum_{\mathbf{x}\in\Omega}\left(\phi_{\mathbf{x}}^{i} - \bar{\Phi}_{\mathcal{A}^{i}(\mathbf{x})}^{i-1}\right)^{2}, \tag{4}$$

where $\bar{\Phi}_{\mathcal{A}^{i}(\mathbf{x})}^{i-1} = \bar{\Theta}_{\mathbf{x}}^{i-1}$, and $\mathcal{A}^{i}(\mathbf{x})$, defined earlier, transforms the shape space estimate $\bar{\Theta}^{i-1}$ to a current image space estimate $\bar{\Phi}_{\mathcal{A}^{i}(\mathbf{x})}^{i-1}$. A sum of squared differences calculation implicitly assumes a Gaussian distribution. Therefore taking the exponential of

[4] The expected shape in a probabilistic region of the shape space (c.f. (2)) is $\bar{\Theta}^{i-1} = \int_{\theta\sim p_w} \theta\ p_m(\theta|\mathcal{M}^{i-1})\ \mathrm{d}\theta = \int_{\theta} \theta\ p_m(\theta|\mathcal{M}^{i-1})\ p_w(\theta|\theta^{i-1})\ \mathrm{d}\theta = \int_{\theta\sim p_m} \theta\ p_w(\theta|\theta^{i-1})\ \mathrm{d}\theta.$

(4) results in a Gaussian distribution and considering the partition representation then $p_{\mathfrak{s}}(q^i|\boldsymbol{q}^{i-1})$ can be used to symbolize the distributional form of (4), hence

$$p_{\mathfrak{s}}(q^i|\boldsymbol{q}^{i-1}) \propto \exp\left(-\mathcal{C}_{\mathfrak{s}}(\phi^i, \Phi^{i-1})\right). \tag{5}$$

This defines the shape distribution of the foreground region given past observations via a non-linear shape estimation process. A probabilistic image model is now defined, combined with this shape model, providing a medium for sequential segmentation of image structures in volumetric medical images that simultaneously takes account of both the shape and image information.

2.2 Image Model

Let $I^j_{\mathbf{x}} : \mathbb{R}^2 \times \mathbb{R}^+ \rightarrow \mathbb{R}^n$ be an n dimensional image intensity at pixel $\mathbf{x} \in \mathbb{R}^2$ and image slice $j \in \mathbb{R}^+$ where e.g. $n = 1$ for gray scale images. Considering all image slices j up to a current image slice i, i.e. $\forall j,\, j = 0...i$ then we can use Bayes theorem to calculate a conditional probability density $p(\boldsymbol{q}|\boldsymbol{I})$ for a set of image partitions $\boldsymbol{q}$ given a set of image intensities $\boldsymbol{I}$ up to slice i

$$p\left(\boldsymbol{q} = \{q^j|\forall\, j,\ j = 0...i\}|\boldsymbol{I} = \{I^j|\forall\, j,\ j = 0...i\}\right) = \frac{p(\boldsymbol{I}, \boldsymbol{q})}{p(\boldsymbol{I})}, \tag{6}$$

where $p(\boldsymbol{I})$ is the marginal data probability density which is not dependent on the image partition information and can therefore be ignored for the purposes of optimization. $p(\boldsymbol{I}, \boldsymbol{q})$ is a joint probability density which is expanded assuming Markov first order dependence of the image data, $p(\boldsymbol{q}|\boldsymbol{I}) \propto p(\boldsymbol{q}^{i-1}|q^i)p(q^i)p(I^0|q^0)\prod_{j=1}^{i} p(I^j|q^j, I^{j-1}, q^{j-1})$. $p(I^0|q^0)$ is the initial image's data likelihood term, $p(\boldsymbol{q}^{i-1}|q^i)$ is the probability density of all image partitions, except the current slice i.e. $\boldsymbol{q}^{i-1} = \{q^j|j = 0...i-1\}$ given the current slice partition q^i. $p(q^i)$ is the prior probability density of the current slice partition. $p(\boldsymbol{q}^{i-1}|q^i)$ and $p(q^i)$ are related to shape and contour labeling smoothness respectively.

The data likelihood term $p(I^j|q^j, I^{j-1}, q^{j-1})$ can then be divided into foreground and background terms (assuming conditional independent pixel intensities via the partition terms),

$$p(\boldsymbol{q}|\boldsymbol{I}) \propto p(\boldsymbol{q}^{i-1}|q^i)p(q^i)p(I^0|q^0)\prod_{j=1}^{i}\prod_{\forall \mathbf{x}\in\Omega}\left[p_{\mathfrak{F}}(I^j_{\mathbf{x}}|q^j, \mathfrak{m}^j_{\mathfrak{F}})^{\mathfrak{f}^j_{\mathbf{x}}} p_{\mathfrak{B}}(I^j_{\mathbf{x}}|q^j, \mathfrak{m}^j_{\mathfrak{B}})^{\mathfrak{b}^j_{\mathbf{x}}}\right] \tag{7}$$

where $\mathfrak{m}^j_{\mathfrak{F}} = \{I^{j-1}_{\mathfrak{F}}, q^{j-1}\}$ and $\mathfrak{m}^j_{\mathfrak{B}} = \{I^{j-1}_{\mathfrak{B}}, q^{j-1}\}$ and the powers $\mathfrak{f}^j_{\mathbf{x}}$ and $\mathfrak{b}^j_{\mathbf{x}}$ act as mutually exclusive switches between the foreground and background. $p_{\mathfrak{F}}(I^j_{\mathbf{x}}|q^j, \mathfrak{m}^j_{\mathfrak{F}})$ and $p_{\mathfrak{B}}(I^j_{\mathbf{x}}|q^j, \mathfrak{m}^j_{\mathfrak{B}})$ thus correspond to two different PDFs for the foreground $I^j_{\mathfrak{F}} = \{I_{\mathbf{x}}|\mathbf{x} \in \mathfrak{F}^j\}$ and background $I^j_{\mathfrak{B}} = \{I_{\mathbf{x}}|\mathbf{x} \in \mathfrak{B}^j\}$ image intensities respectively. The initial image's data likelihood can also be similarly expanded. However, to save space the non-expanded form will be retained.

A smooth labeling and a smooth boundary (both of which are defined as synonymous to each other here) separating the foreground and background regions are desirable properties of an image space partition for segmentation applications. These properties can be achieved by minimizing the length of the boundary of the partition q^i. Therefore $p(q^i) \triangleq p(\mathfrak{L}) \propto \exp(-\lambda_\kappa \mathfrak{L})$ where λ_κ is an exponential rate parameter and $\mathfrak{L}$ is the length of the contour defining the partition (c.f. [12]). Substituting this term into (7) gives

$$p(\boldsymbol{q}|\boldsymbol{I}) \propto p(\boldsymbol{q}^{i-1}|q^i)p(\mathfrak{L})p(I^0|q^0)\prod_{j=1}^{i}\prod_{\forall \mathbf{x}\in\Omega} \overbrace{\left[p_{\mathfrak{F}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{F}})^{\mathfrak{f}^j_{\mathbf{x}}} p_{\mathfrak{B}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{B}})^{\mathfrak{b}^j_{\mathbf{x}}}\right]}^{\textit{model based intensity competition}}. \tag{8}$$

The intensity part of this equation has been labeled as "*model based intensity competition*" indicating that the foreground and background terms work in competition with each other. Foreground/background competition is the basis of many active contour techniques e.g. as first proposed in [13, 14]. However, the intensity components used here are first order Markovian, i.e. they remember intensity information from the preceding slice via $\mathfrak{m}^j_{\mathfrak{F}}$ and $\mathfrak{m}^j_{\mathfrak{B}}$.

A Gaussian distribution possesses symmetry about the mean, so that we can define here $p_{\mathfrak{s}}(q^i|\boldsymbol{q}^{i-1}) \triangleq p(\boldsymbol{q}^{i-1}|q^i)$ (see (5)) which is of the form found in (8), so that

$$p(\boldsymbol{q}|\boldsymbol{I}) \propto p(\mathfrak{L})\prod_{\forall \mathbf{x}\in\Omega}\left[p_{\mathfrak{s}}(q^i_{\mathbf{x}}|\boldsymbol{q}^{i-1}_{\mathbf{x}})p(I^0_{\mathbf{x}}|q^0)\prod_{j=1}^{i}\left[p_{\mathfrak{F}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{F}})^{\mathfrak{f}^j_{\mathbf{x}}} p_{\mathfrak{B}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{B}})^{\mathfrak{b}^j_{\mathbf{x}}}\right]\right]. \tag{9}$$

This probabilistic model now incorporates memory based intensity competition terms, a spatial smoothness term and a shape based term. The current image slice partition q^i is estimated here using a gradient descent level set based optimization process which is now described.

2.3 Optimization process

A gradient descent level set based approach is used to optimize (9) and (1) in the experiments that follow, see e.g. [15, 16]. The optimization of (9) is made possible via the variational derivative (see Appendix) given by

$$\frac{\partial \phi^i_{\mathbf{x}}}{\partial t} = -2\lambda_{\mathfrak{s}}\left(\phi^i_{\mathbf{x}} - \bar{\Phi}^{i-1}_{\mathcal{A}(\mathbf{x})}\right) + \delta_0(\phi^i_{\mathbf{x}})\left(\lambda_\kappa \mathcal{K}_{\mathbf{x}} - \ln\left(\frac{p_{\mathfrak{F}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{F}})}{p_{\mathfrak{B}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{B}})}\right) + \lambda_{\mathfrak{a}}^{\ln\frac{\mathfrak{a}^i_{\mathfrak{b}}}{\mathfrak{a}^i_{\mathfrak{f}}}}\right) \tag{10}$$

where $\lambda_{\mathfrak{s}}$ is a shape term parameter which corresponds to the inverse variance of the shape density $p_{\mathfrak{s}}(q^i|\boldsymbol{q}^{i-1})$ and $\mathcal{K}_{\mathbf{x}} = -\nabla\cdot(\nabla\phi_{\mathbf{x}}/|\nabla\phi_{\mathbf{x}}|)$ is the curvature of the level set at point $\mathbf{x}$. The curvature result follows $p(\mathfrak{L}) \propto \exp(-\lambda_\kappa\mathfrak{L})$ (in (9)) and using the definition of length defined in [15], i.e. $\mathfrak{L} \triangleq \int_\Omega |\nabla H(\phi^i_{\mathbf{x}})|\mathrm{d}\mathbf{x}$ where $H(\cdot)$ is the

Heaviside function. An advection term has also been introduced to regularize the foreground area $\lambda_{\mathfrak{a}}^{\ln \mathfrak{a}_{\mathfrak{b}}^i}/\lambda_{\mathfrak{a}}^{\ln \mathfrak{a}_{\mathfrak{f}}^i}$ where $\lambda_{\mathfrak{a}}$ is an area term weight and $\mathfrak{a}_{\mathfrak{b}}^i = \sum_{\mathbf{x}\in\Omega} \Delta_{\mathbf{x}} \mathfrak{b}_{\mathbf{x}}^i$ and $\mathfrak{a}_{\mathfrak{f}}^i = \sum_{\mathbf{x}\in\Omega} \Delta_{\mathbf{x}} \mathfrak{f}_{\mathbf{x}}^i$ are the background and foreground areas respectively. This advection force penalizes large changes in the area of the segmented structure, to counteract the reduction in the object area that might occur due to the curvature based force and any mis-alignment of the template shape.

The evolving data PDFs $p_{\mathfrak{F}}(I_{\mathfrak{F}}^i|q^i, \mathfrak{m}^i)$ and $p_{\mathfrak{B}}(I_{\mathfrak{B}}^i|q^i, \mathfrak{m}^i)$ are approximated here with finite Gaussian mixtures with parameters estimated from the histograms of the image intensities from the currently estimated image regions $I_{\mathfrak{F}}^i$, $I_{\mathfrak{B}}^i$ and the image intensities from the previous slice, $I_{\mathfrak{F}}^{i-1}$ and $I_{\mathfrak{B}}^{i-1}$. The parameters of the finite Gaussian mixture models are estimated here using Expectation Maximization and the number of finite mixtures in the Gaussian mixture model was set to 6. This was empirically found to provide the best results for the test sequences.

In common with many active contour techniques manual parameter adjustment is also required to control the relative contribution of the individual components which have to be tuned for sequences with different properties. The similarity alignment transform (see [16]) also has weight parameters which can be kept constant once suitable values have been determined.

3 Experiments and Results

The sequential segmentation process described here can be applied to spatial and or temporal sequences. A single 2D manually defined segmentation is used here for initialization. Subsequent slices are segmented by automatically propagating the shape and intensity information from one slice to the next. Parameter estimation for the different components in (10) is performed using empirical methods.

3.1 Spatial sequential segmentation

A spatial sequential segmentation process result (3D) that identifies the bounds of a tumor in the transverse plane is illustrated in figure 2. This result uses a central transverse 2D slice for manual initialization. Intensity and shape information is then automatically propagated, via the model described here, to the remaining image slices in the 3D data volume. The bounds of the tumor are successfully located for each image slice in the 3D data volume.

The shape of the spatial pathological structure in figure 2 possesses very little variation other than changes in scale. However the shape of the temporal structure of interest that follows (cardiac MRI) possesses greater variation in shape.

3.2 Temporal sequential segmentation

A temporal sequential segmentation result (3D) was illustrated earlier in figure 1 for a Cardiac MRI sequence where the left and right atriums are successfully segmented. The algorithm was initialized using two 2D image slices, one for the right atrium and another for the left atrium.

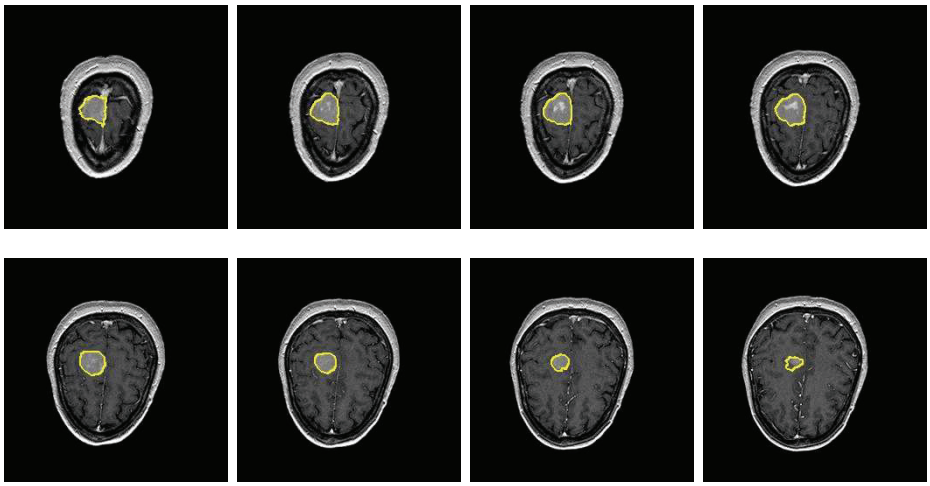


Fig. 2. Result from spatial sequential segmentation process in the transverse plane that identifies the bounds of a tumor for subsequent slices after manual initialization (bottom left). Images correspond to T1 + Gadolinium MRI scan, 59 Year old female at the NMR Center of the Massachusetts General Hospital with a 1.5 Tesla General Electric Sigma and provided by the Center for Morphometric Analysis (http://neuro-www.mgh.harvard.edu/cma/ibsr).

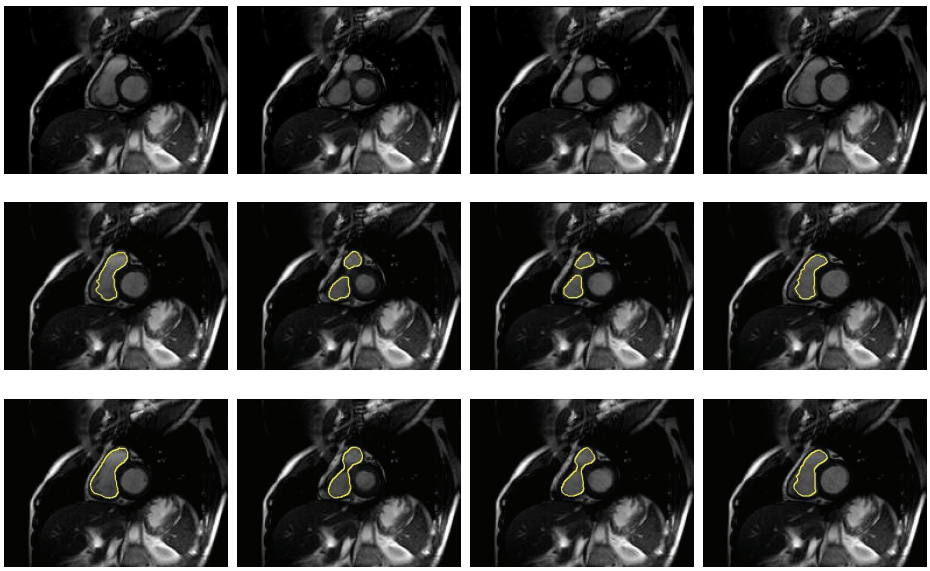


Fig. 3. Comparative temporal sequential segmentation of right atrium in the sagittal plane after single 2D image slice manual initialization. 1st row: original data; 2nd row: sequential segmentation using intensity and curvature information only; 3rd row: sequential segmentation using the proposed method. Images correspond to a cardiac MRI scan acquired at Bristol Royal Infirmary.

A further temporal sequential segmentation result (3D) of the right atrium after manual initialization is shown in figure 3 where the result is compared with a result using intensity information combined curvature but no shape model.

Figure 3 demonstrates an interesting result where the intensity information combined with a curvature based force is not able to identify realistic bounds of the right atrium. However the result corresponding to the work described here using learned on-line shape information enables prior information to be carried across from one image slice to the next, assisting with identifying the boundary of the right atrium.

3.3 4D: Combining temporal and spatial sequential segmentation

In the results above the temporal and spatial sequential segmentations are obtained in three dimensions. The same technique can be used to segment in higher dimensions. An example of 4D sequential segmentation in both the temporal and spatial domains is presented in figure 4.

The 4D sequential segmentation result in figure 4 was obtained by initializing the algorithm with a single 2D image slice, corresponding to a single sagittal section at a particular time instance. After all image slices in the temporal domain are segmented for a particular sagittal location, the algorithm propagates to another sagittally adjacent image slice for a particular time instance and then subsequent time instances are segmented for that sagittal location. The algorithm was able to successfully propagate the shape information to other time instances and other spatial locations in the sagittal plane. A comparison is also made with an algorithm using no shape model. The shape of the tracked atrium using only an intensity based model with curvature degenerates, particularly as new spatially located image slices are encountered. This is further illustrated by sensitivity calculations in comparison to partial volume calculated ground truths, in figure 5. The model described here has a much better sensitivity at the systolic stage of the cardiac cycle despite the significant changes in the shape of the heart.

4 Discussion and Conclusions

A new active contour based sequential segmentation framework has been presented. This utilizes high-level shape information that is learned on-line, adapting to new shape configurations whilst constraining the evolution of the active contour. Results have shown that the combined framework is able to segment structures of interest undergoing complex deformations of shape. The main shortcomings of the method are that it is computationally complex, requiring significant time to segment a single image slice and, similar to many active contour tracking frameworks, successful segmentation is highly dependent on an empirical selection of parameter values that control the relative contribution of the different model components.

Acknowledgments

This research was funded by the UK Leverhulme Trust and carried out at the Department of Computer Science, University of Bristol with some final experiments performed after the first author joined MFU, Thailand.

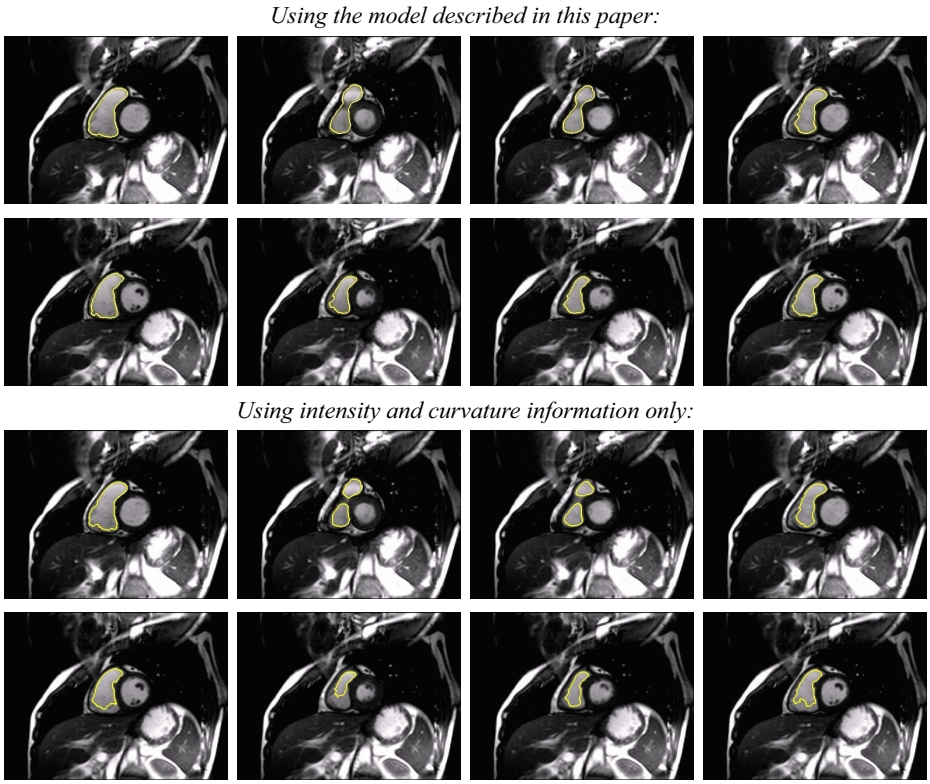


Fig. 4. Results of sequential segmentation in four dimensions in the temporal and spatial domains. Each row is a different sagittal slice through the subject and each column is a different time instance. The right atrium outline is delineated for all image slices for the model described here (top two rows), where the algorithm was initialized using a single 2D image slice in a single sagittal time instance. If only intensity information combined with curvature is used in the modeling process (bottom two rows) then the shape of the tracked region degenerates as more image slices are segmented. Images correspond to a cardiac MRI scan acquired at Bristol Royal Infirmary.

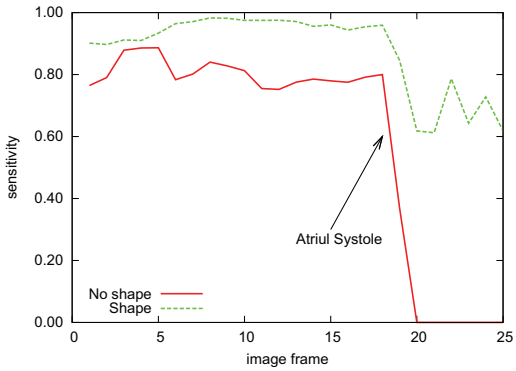


Fig. 5. Quantitative assessment for the 4D cardiac sequence results shown in figure 4.

Appendix

Maximization of (9) can be used to obtain a partition of the image data that most closely matches the given shape model and intensity information. This is also equivalent to minimization of the negative logarithm which is simpler, hence we minimize

$$\mathbb{E}(\boldsymbol{q},\boldsymbol{I}) = -\int_{\Omega}\Bigg[\ln p_s(q^i_{\mathbf{x}}|\boldsymbol{q}^{i-1}_{\mathbf{x}}) + \ln p(q^i_{\mathbf{x}}|q^i_{\mathcal{N}_{\mathbf{x}}}) + \sum_{j=1}^{i}\left[\mathfrak{f}^j_{\mathbf{x}}\ln p_{\mathfrak{F}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{F}}) + \mathfrak{b}^j_{\mathbf{x}}\ln p_{\mathfrak{B}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{B}})\right]\Bigg]\mathrm{d}\mathbf{x}. \tag{A-1}$$

The foreground and background labellings $\mathfrak{f}^j$, $\mathfrak{b}^j$ are equivalent to functions of the Heaviside function on the level sets, i.e. $H(\phi^j_{\mathbf{x}}) = \mathfrak{b}^j_{\mathbf{x}} = (1-\mathfrak{f}^j_{\mathbf{x}})$ hence

$$\mathbb{E}(\boldsymbol{q},\boldsymbol{I}) = -\int_{\Omega}\Big[\ln p_s(q^i_{\mathbf{x}}|\boldsymbol{q}^{i-1}_{\mathbf{x}}) + \ln p(q^i_{\mathbf{x}}|q^i_{\mathcal{N}_{\mathbf{x}}}) + \sum_{j=1}^{i}\Big[(1-H(\phi^j_{\mathbf{x}}))\ln p_{\mathfrak{F}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{F}}) + H(\phi^j_{\mathbf{x}})\ln p_{\mathfrak{B}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{B}})\Big]\Big]\mathrm{d}\mathbf{x}, \tag{A-2}$$

where now $-\ln p(q^i_{\mathbf{x}}|q^i_{\mathcal{N}_{\mathbf{x}}}) = \lambda_\kappa|\nabla H(\phi^j_{\mathbf{x}})|$. Minimization of (A-2) can be performed via gradient descent on the variational derivative w.r.t. a gradient descent time parameter t, $\frac{\partial\mathbb{E}(\boldsymbol{q},\boldsymbol{I})}{\partial\phi^i} = -\frac{\partial\phi^i}{\partial t}$ where the level sets have been parametrized as a function of pixel location $\mathbf{x}$ and the gradient time parameter t, i.e. $\phi(\mathbf{x},t)$. The variational derivative can therefore be shown to be given by (10). An analytical solution to $\frac{\partial\phi^i_{\mathbf{x}}}{\partial t} = 0$ is not available, therefore finite difference approximations are used to iteratively find the solution, i.e.

$$\phi^i_{\mathbf{x}}(t+1) = \phi^i_{\mathbf{x}}(t) + \Delta_t\Bigg(-2\lambda_s\left(\phi^i_{\mathbf{x}} - \bar{\Phi}^{i-1}_{\mathcal{A}(\mathbf{x})}\right) + \delta_0(\phi^i_{\mathbf{x}})\left(\lambda_\kappa\mathcal{K}_{\mathbf{x}} - \ln\left(\frac{p_{\mathfrak{F}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{F}})}{p_{\mathfrak{B}}(I^j_{\mathbf{x}}|q^j,\mathfrak{m}^j_{\mathfrak{B}})}\right)\right) - \lambda_{\mathfrak{a}}\ln\frac{\mathfrak{a}^i_{\mathfrak{b}}}{\mathfrak{a}^i_{\mathfrak{f}}}\Bigg), \tag{A-3}$$

where Δ_t corresponds to the size of the time step. The Dirac delta function restricts the computations to the contour rather than the entire level set. Therefore, following [17], the computations are extended to the entire set of pixels in the image space Ω by replacing $\delta_0(\phi^i_{\mathbf{x}})$ with $|\nabla\phi^i_{\mathbf{x}}|$ and a narrow band ($|\phi^i_{\mathbf{x}}| < \mathfrak{T}$ where $\mathfrak{T} \in \mathbb{R}^+$ is a real positive value) is also used to reduce the number of computations necessary to update the position of the contour. $|\nabla\phi^i_{\mathbf{x}}|$ is approximated here with a non-oscillatory upwind finite difference scheme, see e.g. [18].

The level sets are reinitialized here every t_{reinit} iterations to retain the smoothness and distance properties of the level set using a signed distance transform, [19]. The optimization of (A-3) continues until convergence which is assessed by comparing the sum of squared differences of the level set every t_{reinit} iterations or until a maximum number of iterations have been reached (t_{max}).

References

1. Paladini, D., Vassallo, M., Sglavo, G., Lapadula, C., Martinelli, P.: The role of spatio-temporal image correlation (STIC) with tomographic ultrasound imaging (TUI) in the sequential analysis of fetal congenital heart disease. Ultrasound in obstretics and gynecology **27**(5) (Apr. 2006) 555–561
2. Anderson, R., YenHo, S.: Sequential segmental analysis. In: Continuing Medical Education. Cambridge University Press (1997) 98–116
3. Blake, A., Isard, M., Reynard, D.: Learning to track curves in motion. In: 33rd IEEE Conf. Decision and Control. Volume 4. (1994) 4:3788–3793
4. Baumberg, A., Hogg, D.: An efficient method for contour tracking using active shape models. In: IEEE Workshop on motion of non-rigid and articulated objects. (1994) 194–199
5. Leventon, M., Grimson, W., Faugeras, O.: Statistical shape influence in geodesic active contours. In: Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR'00). (2000) 316–323
6. Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A.: A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Medical Imaging **22**(2) (Feb. 2003) 137–154
7. Cootes, T., Cooper, D., Taylor, C., Graham, J.: A trainable method of parametric shape description. Image Vision Comput. **10** (1992) 289–294
8. Cho, J., Benkeser, P.: Cardiac segmentation by a velocity-aided active contour model. Comp. Med. Imag. Graph. **30** (2006) 2006
9. Senegas, J., Netsch, T., Cocosco, C., Lund, G., Stork, A.: Segmentation of medical images with a shape and motion model: a Bayesian perspective. In: CVAMIA-MMBIA. Volume 3117 of LNCS., Springer (2004) 157–168
10. Vaswani, N., Yezzi, A., Rathi, Y., Tannenbaum, A.: Time-varying finite dimensional basis for tracking contour deformations. In: Proc. 45th IEEE Conf. Dec. Control, San Diego (Dec. 2006) 1665–1672
11. Osher, S., Sethian, J.: Fronts propagating with curvature-dependent speed. J. Comp. Phys. **79** (1988) 12–49
12. Cremers, D.: Nonlinear dynamical shape priors for level set segmentation. In: Proc. IEEE CS Conf. Computer Vision and Pattern Recognition (CVPR'07). (2007)
13. Paragios, N., Deriche, R.: Geodesic active regions: a new framework to deal with frame partition problems in computer vision. J. Vis. Comm. Image Rep. **13**(1-2) (2002) 249–268
14. Zhu, S., Yuille, A.: Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans. Pattern Analysis and Machine Intelligence **18**(9) (1996) 884–900
15. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. Image Processing **10**(2) (2001) 266–277
16. Paragios, N., Taron, M., Huang, X., Rousson, M., Metaxas, D.: On the representation of shapes using implicit functions. In: Statistics and Analysis of Shapes. Springer (2006)
17. Zhao, H., Chan, T., Merriman, B., Osher, S.: A variational level set approach to multiphase motion. J. Comp. Phys. **127** (1996)
18. Aubert, G., Kornprobst, P.: Mathematical problems in image processing. Springer (2006)
19. Borgefors, G.: Distance transformations in digital images. Comp. Vis. Graph. Image Proc. **34**(3) (1986) 344–371

A spatially variant mixture model for diffusion weighted MRI: application to image denoising

Juan Eugenio Iglesias[1], Paul Thompson[2], and Zhuowen Tu[2]

[1] Medical Imaging Informatics, University of California, Los Angeles, jeiglesias@ucla.edu,
[2] Laboratory of Neuroimaging, University of California, Los Angeles, thompson@loni.ucla.edu, zhuowen.tu@loni.ucla.edu

Abstract. High angular resolution diffusion imaging is an increasingly important image modality. The nature of the diffusion data makes mixtures of probability distributions particularly suitable for modeling its signals. In this paper, we introduce Bayesian finite mixture models for studying the diffusion field. We apply a spatially variant mixture model to study prior distributions on the model parameters of the field. The mean vectors and covariance matrices are independent of lattice locations, but the mixture weights are allowed to differ from one location to another. Spatial smoothness is achieved by placing a Markov random field prior on top of the mixture weights. The output is a general model that can be used in different HARDI applications, such as fiber tracking and image denoising. The latter is illustrated in this study, with promising results on a real dataset.

1 Introduction

Diffusion weighted magnetic resonance imaging (DWI-MRI) is a technique that produces in vivo images of biological tissue weighted by the local characteristics of water diffusion. Using the decrease in MRI signal with respect to a baseline T2 scan when a gradient in the magnetic field is applied in a given direction, the water diffusivity in that direction can be estimated. By sampling different directions, a full diffusion profile can be reconstructed for every voxel in the scanned volume. Since water diffuses along axonal fibers in the brain, DWI can be used to follow neural tracts in the white matter and generate a map of the connections between different brain regions. DWI is currently the only way to study these fiber bundles in a non-invasive manner.

The most popular way of reconstructing fiber orientations is diffusion tensor imaging (DTI)[1, 2], in which a zero-mean Gaussian probability distribution function (PDF) is fitted to a relatively small set of data measurements. However, this model fails to explain fiber crosses and bifurcations[3]. One way of overcoming the limitations of DTI is to sample the diffusivity on a high number of directions on a sphere around each voxel. This approach is known as high angular resolution diffusion imaging (HARDI)[3, 4]. From the HARDI data, the

PDF of the water diffusivity on a spherical shell (orientation distribution function, ODF[5]) can be computed.

This paper presents spatially variant Gaussian mixture model (GMM) for DW-MRI. Mixture models have been successfully applied to DW-MRI[3, 6–9]; they suit this task very well because fiber crossings and bifurcations can be modeled as the sum of two or more mixture components. In this study, the mean vectors and covariance matrices are fixed but the mixture weights are allowed to vary across the image. Smoothness is ensured by placing a Markov random field (MRF) prior on top of these weights. This idea was applied to natural image segmentation by Nikou et al. in [10]. King et al. [11] also described a similar system in a general way, though without any mathematical formulation. The output of the algorithm is a general model that can be used in different DW-MRI applications; a denoising technique is illustrated here. Denoising is very attractive in HARDI because acquiring a high number of images with different gradient directions in limited time leads to a low signal to noise ratio (SNR).

2 Methods

2.1 Image model

Let $\boldsymbol{x}(\boldsymbol{r}) = [x_1(\boldsymbol{r})...x_D(\boldsymbol{r})]^t$ be the vector of D features representing the HARDI data at voxel $\boldsymbol{r} = (i, j, k)$. We assume that $\boldsymbol{x}$ is a realization of a Gaussian mixture model (GMM):

$$p(\boldsymbol{x}(\boldsymbol{r})) = \sum_{c=1}^{C} \pi_c(\boldsymbol{r}) G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c) \tag{1}$$

where $G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}^c, \boldsymbol{\Sigma}^c)$ is a Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. The weights of the mixture $\boldsymbol{\pi}(\boldsymbol{r}) = [\pi_1(\boldsymbol{r})...\pi_C(\boldsymbol{r})]^t$ are spatially variant, and they lie in the probability simplex $\sum_{c=1}^{C} \pi_c(\boldsymbol{r}) = 1$, with $\pi_c(\boldsymbol{r}) \geq 0$.

Now, we add a MRF prior on top of the mixture weights to model the fact that neighboring pixels should be characterized by similar Gaussian mixtures:

$$p(\boldsymbol{\Pi}) \propto \prod_{d=1}^{3} \prod_{c=1}^{C} \beta_{d,c}^{-N} \exp\left[-\frac{1}{2} \frac{\sum_{\boldsymbol{r}} \left(2\pi_c(\boldsymbol{r}) - \pi_c(\boldsymbol{r} + \boldsymbol{u}_d) - \pi_c(\boldsymbol{r} - \boldsymbol{u}_d)\right)^2}{\beta_{d,c}^2} \right] \tag{2}$$

where $\boldsymbol{\Pi}$ is the set of weight vectors at all spatial locations, N is the total number of voxels, $\beta_{d,c}^2$ is the variance of the mixture component c in direction $d = \{1, 2, 3\}$, and $\boldsymbol{u}_d$ represents the a unit vector along d. The interpretation of this prior is simple: when one mixture component at one voxel is predicted as the average of the values of its neighbors in direction d, the error follows a zero-mean Gaussian distribution with variance $\beta_{d,c}^2/2$.

2.2 Data space

The HARDI data consists of a set of MRI images corresponding to different gradient directions and intensities. At each location $\boldsymbol{r}$ and direction k, the signal intensity decreases by water diffusion according to the Stejskal-Tanner equation:

$$S_k(\boldsymbol{r}) = S_0(\boldsymbol{r}) \exp[-b_k D_k(\boldsymbol{r})] \Longleftrightarrow D_k(\boldsymbol{r}) = \frac{1}{b_k} \log\left(\frac{S_k(\boldsymbol{r})}{S_k(\boldsymbol{r})}\right)$$

where S_0 is the non-diffusion weighted baseline T2 intensity, $D_k(r)$ is the apparent diffusion coefficient (ADC), and b_k is the Le Bihan's factor, which groups information from the pulse sequence, gradient strength, and other physical constants. The SNR of the baseline $S_0(\boldsymbol{r})$ is typically much higher than the SNR of the gradient images because it is not attenuated by diffusion and because several acquisitions of $S_0(\boldsymbol{r})$ are usually available. In that case, it is fair to assume that $S_0(\boldsymbol{r})$ is a perfect estimate and only the gradient images $S_k(\boldsymbol{r})$ are affected by noise. Then, the data for each voxel is fully characterized by its vector of ADCs. This is the feature set used in this study: $\boldsymbol{x}(\boldsymbol{r}) = \{D_k(\boldsymbol{r})\}$.

2.3 Maximum likelihood estimation

Initialization: choice of C with minimum message length criterion A good starting point for the algorithm that estimates the parameters of the complete image model (described in section 2.3 below) is to carry out the estimation without considering the MRF from equation 2. This amounts to the classical problem of estimating the parameters of a GMM, which can be solved with the expectation maximization algorithm (EM)[12]. The EM can in turn be initialized with the K-means algorithm, which is further initialized with random samples from the training data. Since the EM algorithm is prone to getting stuck in local optima, it is a good idea to run it several times with different initializations and keep the parameters that produce the highest likelihood.

A key design parameter in GMMs is the number of components C. One way of determining C is to assign it a very high initial value, run the EM algorithm, and then merge down components as long as the minimum message length[13] (MML) decreases. MML is a performance function that combines the likelihood of the data and a penalty term for the complexity of the model, which increases with the number of components. At each step, all possible pair-wise merges are considered. If the MML of the model resulting from the best merge is lower than the MML of the current model, the merge is carried out, and the process is repeated. Otherwise, the algorithm terminates.

Optimization with EM algorithm The spirit of the optimization algorithm is similar to that of [10]. The parameters to estimate are the mixture component parameters $\{\boldsymbol{\mu}^c, \boldsymbol{\Sigma}^c\}$ and the class variances $\beta^2_{d,c}$. The functional to maximize combines a term for the GMM (see details in [14]) and a term for the MRF:

$$Q = \sum_{\boldsymbol{r}} \sum_{c=1}^{C} \left(z_c(\boldsymbol{r}) \left[log(\pi_c(\boldsymbol{r})) + log\left(G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)\right)\right]\right) - \ldots$$

$$\ldots - \frac{1}{2}\sum_{\boldsymbol{r}}\sum_{c=1}^{C}\sum_{d=1}^{3}\left[log(\beta_{d,c}^2) + \frac{[2\pi_c(\boldsymbol{r}) - \pi_c(\boldsymbol{r}+\boldsymbol{u}_d) - \pi_c(\boldsymbol{r}-\boldsymbol{u}_d)]^2}{\beta_{d,c}^2}\right] \quad (3)$$

where $z_c(\boldsymbol{r})$ is the posterior probability that the voxel belongs to component c:

$$z_c(\boldsymbol{r}) = \frac{\pi_c(\boldsymbol{r})G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)}{\sum_{p=1}^{C}\pi_p(\boldsymbol{r})G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)} \quad (4)$$

The functional Q from equation 3 can be maximized using EM. In the E step, the posterior probabilities $z_c(\boldsymbol{r})$ are calculated using equation 4. In the M step, the estimates of the parameters $\{\boldsymbol{\mu}^c, \boldsymbol{\Sigma}^c\}, \beta_{d,c}^2$ and the mixture weights at each pixel $\pi_c(\boldsymbol{r})$ must be updated to maximize Q. The update equations for the mixture model parameters are well-known:

$$\boldsymbol{\mu}_c \leftarrow \frac{\sum_{\boldsymbol{r}} z_c(\boldsymbol{r})\boldsymbol{x}(\boldsymbol{r})}{\sum_{\boldsymbol{r}} z_c(\boldsymbol{r})}, \quad \boldsymbol{\Sigma}_c \leftarrow \frac{\sum_{\boldsymbol{r}} z_c(\boldsymbol{r})[(\boldsymbol{x}(\boldsymbol{r}) - \boldsymbol{\mu}_c(\boldsymbol{r})][\boldsymbol{x}(\boldsymbol{r}) - \boldsymbol{\mu}_c(\boldsymbol{r})]^t}{\sum_{\boldsymbol{r}} z_c(\boldsymbol{r})} \quad (5)$$

Updating the mixture weights is more complicated. Simultaneous optimization of all the weights is very difficult and impractical. Instead, coordinate ascent can be used: voxels are visited in random order and their weights are optimized under the assumption that weights are all other locations remain constant. The algorithm usually converges after five or six passes. Every time a voxel is visited, the equation $\partial Q/\partial\pi_c(\boldsymbol{r}) = 0$ must be solved. The derivative introduces terms depending on neighbors of neighbors:

$$\frac{\partial Q}{\partial \pi_c(\boldsymbol{r})} = \frac{z_c(\boldsymbol{r})}{\pi_c(\boldsymbol{r})} - \sum_{d=1}^{3}\frac{6\pi_c(\boldsymbol{r}) - 4\pi_c(\boldsymbol{r}+\boldsymbol{u}_d) - 4\pi_c(\boldsymbol{r}-\boldsymbol{u}_d) + \pi_c(\boldsymbol{r}+2\boldsymbol{u}_d) + \pi_c(\boldsymbol{r}-2\boldsymbol{u}_d)}{\beta_{d,c}^2}$$

Setting this derivative equal to zero gives a second degree equation of the form $a\pi_c(\boldsymbol{r})^2 + b\pi_c(\boldsymbol{r}) + c = 0$:

$$-6\sum_{d=1}^{3}\left(\prod_{\substack{p=1\\p\neq d}}\beta_{d,c}^2\right)\pi_c^2(\boldsymbol{r}) + \sum_{d=1}^{3}\left(\prod_{\substack{p=1\\p\neq d}}\beta_{d,c}^2\right)[4\pi_c(\boldsymbol{r}+\boldsymbol{u}_d) + 4\pi_c(\boldsymbol{r}-\boldsymbol{u}_d) - \ldots$$

$$\ldots - \pi_c(\boldsymbol{r}+2\boldsymbol{u}_d) - \pi_c(\boldsymbol{r}-2\boldsymbol{u}_d)]\pi_c(\boldsymbol{r}) + z_c(\boldsymbol{r})\prod_{d=1}^{3}\beta_{d,c}^2 = 0 \quad (6)$$

Due to the positive nature of variances and probabilities, it is always the case that $a \leq 0$ and $c \geq 0$, which guarantees that one solution will be positive and the other negative; the positive one is kept. Because each weight is optimized independently, the sum of all weights for a given voxel will not in general be equal to one. The weights must hence be projected onto the constraints ($\sum_c \pi_c = 1$, $\pi_c \geq 0$). This is a simple quadratic program that can be efficiently solved[15].

Finally, the updated mixture weights give the new class variances:

$$\beta_{d,c}^2 \leftarrow \frac{1}{N}\sum_{\boldsymbol{r}}[2\pi_c(\boldsymbol{r}) - \pi_c(\boldsymbol{r}+\boldsymbol{u}_d) - \pi_c(\boldsymbol{r}-\boldsymbol{u}_d)]^2 \quad (7)$$

The algorithm is summarized in table 1.

1. Initialize the mixture model components $\{\boldsymbol{\mu}^c, \boldsymbol{\Sigma}^c\}$ with the standard GMM.
2. Initialize the mixture weights for every voxel with those from the standard GMM.
3. While the functional Q from equation 3 keeps increasing significantly, repeat:
a) Calculate the posterior probabilities $z_c(\boldsymbol{r})$ with equation 4.
b) Update the mixture model components with equation 5.
c) Calculate the next mixture weights $\pi_c(\boldsymbol{r})$ at each voxel by solving equation 6.
d) Project the mixtures onto the constraints by solving a quadratic program.
e) Update the class variances $\beta^2_{d,c}$ with equation 7.

Table 1. EM algorithm to estimate the model parameters.

2.4 Application to image denoising

A possible application of the model is image denoising, which is very attractive in HARDI because the large number of DW images to acquire limits the acquisition time and hence the SNR. In a Bayesian framework, the goal to maximize the posterior probability of the underlying noise-free image given the noisy version. Optimizing this probability directly is computationally very demanding. As in the previous section, coordinate ascent can be used. In this context, this strategy is known as iterated conditional modes (ICM)[16]: voxels are visited in random order, and their values are optimized by maximizing the conditional posterior $p(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{x}^*(\boldsymbol{r}), \hat{\boldsymbol{x}}(S \setminus \boldsymbol{r}))$, where $\boldsymbol{x}^*(\boldsymbol{r})$ is the observed value at the given voxel and $\hat{\boldsymbol{x}}(S \setminus \boldsymbol{r})$ is the current reconstruction at all other locations. The algorithm typically converges after five or six iterations.

According to the image model in this study, the joint conditional distribution of the mixture weights $\boldsymbol{\pi}(\boldsymbol{r})$ and the "real" image value $\boldsymbol{x}(\boldsymbol{r})$ at each voxel is:

$$p(\boldsymbol{x}(\boldsymbol{r}), \boldsymbol{\pi}(\boldsymbol{r})|\boldsymbol{x}^*(\boldsymbol{r}), \boldsymbol{\pi}(S \setminus \boldsymbol{r})), \hat{\boldsymbol{x}}(S \setminus \boldsymbol{r})) \propto ...$$

$$p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r}), \boldsymbol{\pi}(\boldsymbol{r})) \cdot p(\boldsymbol{x}(\boldsymbol{r}), \boldsymbol{\pi}(\boldsymbol{r})|\boldsymbol{\pi}(S \setminus \boldsymbol{r})), \hat{\boldsymbol{x}}(S \setminus \boldsymbol{r})) = F$$

Assuming: 1. that the values for the rest of voxels are not noisy, 2. that the image value at one pixel only depends on the "real" value at that pixel, 3. that such noise-free value only depends on the mixture weights at the given location, and 4. that those mixture weights form a Markov random field ($\aleph_{\boldsymbol{r}}$ represents the neighborhood of $\boldsymbol{r}$):

$$\begin{aligned} F &= p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r})) \cdot p(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\pi}(\aleph_{\boldsymbol{r}}), \boldsymbol{x}(\aleph_{\boldsymbol{r}}), \boldsymbol{\pi}(\boldsymbol{r})) \cdot p(\boldsymbol{\pi}(\boldsymbol{r})|\boldsymbol{\pi}(\aleph_{\boldsymbol{r}}), \boldsymbol{x}(\aleph_{\boldsymbol{r}})) \\ &= \underbrace{p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r}))}_{NOISE\ MODEL} \cdot \underbrace{p(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\pi}(\boldsymbol{r}))}_{GMM} \cdot \underbrace{p(\boldsymbol{\pi}(\boldsymbol{r})|\boldsymbol{\pi}(\aleph_{\boldsymbol{r}}))}_{MRF} \end{aligned} \tag{8}$$

The first term in equation 8 corresponds to the noise model (rician in MRI):

$$p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r})) = \prod_{k=1}^{K} \frac{S_0(\boldsymbol{r}) e^{-b_k x_k^*(\boldsymbol{r})}}{\sigma^2} \exp\left(-\frac{S_0^2(\boldsymbol{r}) e^{-2b_k (x_k(\boldsymbol{r}) + x_k(\boldsymbol{r}))}}{2\sigma^2}\right) I_0\left(\frac{S_0^2(\boldsymbol{r}) e^{-b_k (x_k(\boldsymbol{r}) + x_k^*(\boldsymbol{r}))}}{\sigma^2}\right)$$

where I_0 is the modified Bessel function of the first kind with order zero. The second term is just the GMM from equation 1. The third term is a product of

univariate Gaussians corresponding to the MRF. Each Gaussian is the result of multiplying the conditional probabilities given by the neighbors in each of the three spatial directions:

$$p(\boldsymbol{\pi}(\boldsymbol{r})|\boldsymbol{\pi}(\aleph_{\boldsymbol{r}})) = \prod_{c=1}^{C}\prod_{d=1}^{3} G\left(\pi_c(\boldsymbol{r}+\boldsymbol{u}_d)+\pi_c(\boldsymbol{r}-\boldsymbol{u}_d), \beta_{d,c}^2\right) =$$

$$= \prod_{c=1}^{C} G\left[2\pi_c(\boldsymbol{r})|\frac{\sum_{d=1}^{3}(\pi_c(\boldsymbol{r}+\boldsymbol{u}_d)+\pi_c(\boldsymbol{r}-\boldsymbol{u}_d))\prod_{\substack{p=1\\p\neq d}}^{3}\beta_{p,c}^2}{\sum_{d=1}^{3}\prod_{\substack{p=1\\p\neq d}}^{3}\beta_{p,c}^2}, \frac{\prod_{d=1}^{3}\beta_{d,c}^2}{\sum_{d=1}^{3}\prod_{\substack{p=1\\p\neq d}}^{3}\beta_{p,c}^2}\right]$$

Thanks to the fact that the partial derivatives of F (or rather its logarithm, to simplify the expressions) with respect to $\pi_c(\boldsymbol{r})$ and $x_k(\boldsymbol{r})$ can be quickly computed analytically, the image value and mixture weights at each voxel can be easily optimized using gradient ascent. For the same reasons as in the EM algorithm in section 2.3, it is necessary to project the mixture weights onto the constraints after each step (again, the method from [15] can be used).

3 Experiments and results

3.1 Data

100 HARDI images were acquired at the Center for Magnetic Resonance at the University of Queensland using a 4 Tesla Bruker Medspec scanner with a transverse electromagnetic headcoil. Diffusion-weighted scans utilized a single-shot echo planar technique with a twice-refocused spin echo sequence to minimize eddy-current induced distortions. The timing of the diffusion sequence was optimized for SNR. 94 diffusion-sensitized gradient directions and 11 baseline images with no diffusion-sensitization were obtained for every subject. Imaging parameters were: b-value = 1159 s/mm^2, TE/TR = 92.3/8,259 ms, voxel size = 1.8 mm 1.8 mm 2.0 mm. The acquisition time was approximately 15 minutes.

The 11 baseline images were merged down to a single estimate of the reference $S_0(\boldsymbol{r})$ using the method in [17]. The baseline image was used to calculate a mask corresponding to the brain using the BET algorithm[18]. The calculated mask was then applied to all the diffusion images.

The diffusion vector image was downsampled from 94 to 30 directions using a Laplace-Beltrami regularized spherical harmonic expansion[19] of order six, with regularization coefficient $\lambda = 0.006$. The 30 directions were determined using an electrostatic approach[20]. The reason for downsampling is double. First, it lightens the computational load of the algorithms. Second, it provides a version of the images with low noise level. These 30-directional images will be used as ground truth in this study.

Finally, the 100 images were randomly divided into two groups: one for training purposes and one for testing the model.

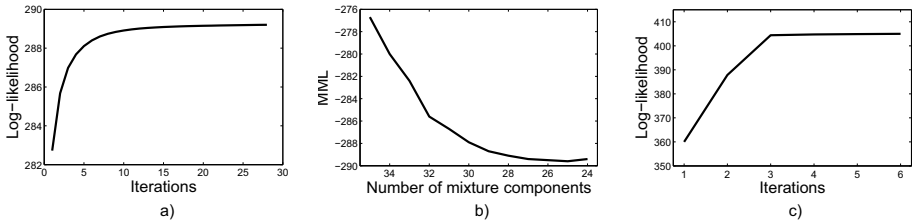


Fig. 1. Metrics in the training stage (a) GMM (best of five) (b) MML (c) MRF-GMM.

3.2 Training the model

Pilot experiments with small chunks of data showed that $25 \sim 30$ components was a reasonable value for C. The GMM (without MRF) was therefore trained using 35 components at first, so that the MML criterion would bring them down to the expected amount. The training process was repeated five times with different random seeds for the K-means algorithm, and the model with the best likelihood was kept. Not all the data were used in the training, but 25x25x20-voxel patches randomly extracted from the 50 test images. The only condition was that at least two thirds of the voxels of each patch should be within the corresponding brain mask. This provides approximately half a million training voxels, approximately 30 per model parameter to estimate. The evolution of the log-likelihood with the number of iterations is displayed in figure 1-a.

Once the GMM was trained for $C = 35$, components were merged until the MML did not decrease anymore, which happened at $C = 25$ (figure 1-b). Finally, the resulting GMM was used as initialization for training the GMM-MRF model. The EM algorithm converges very rapidly, as seen in figure 1-c. Apart from the model, the average over all training pixels of the mixture weights (henceforth $\bar{\pi}_c$) was saved. These values will be used in the initialization for the denoising algorithm. The model components and weights are displayed in figure 2.

3.3 Denoising

The 50 test images were artificially corrupted with Rician noise (σ=40, which gives SNR$\approx$ 9) and then restored using the proposed denoising algorithm. The noisy versions were used to initialize the estimate of the image values. The mixture weights for each voxel were initialized with the following expression:

$$\pi_c(\boldsymbol{r}) = \frac{\bar{\pi}_c G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)}{\sum_{p=1}^{C} \bar{\pi}_p G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)}$$

which is very similar to equation 4, with the difference that the local mixture weights are replaced by the average weights $\bar{\pi}_c$ over all the training data after the last step of the EM algorithm. The value of the parameter σ of the Rician

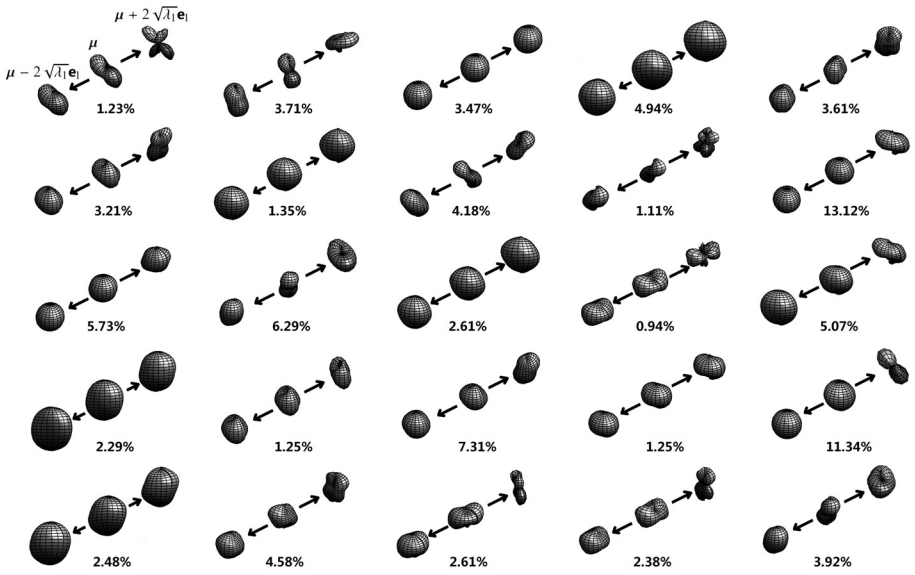


Fig. 2. Trained GMM components. For each component, the average weight over the training data $\bar{\pi}_c$ and the first mode of variation ($\boldsymbol{\mu} \pm 2\sqrt{\lambda_1}\boldsymbol{e}_1$, where $\{\boldsymbol{\mu}, \lambda_1, \boldsymbol{e}_1\}$ are the mean, first eigenvalue and first eigenvector of the covariance matrix) are shown. The components were min-max normalized and smoothed with a Gaussian radial basis function for display.

distribution is assumed to be known exactly. When unknown, this parameter can be efficiently estimated from the background pixels by studying their statistical moments [21] or histogram [22].

Figure 3 shows the 16^{th} DW component, as well as the ADC field, for the original, corrupted and denoised versions of a sample slice. The proposed algorithm is compared with three simple strategies: Gaussian blurring with standard deviation σ_g=3 mm, anisotropic diffusion[23] and Rician adapted non-local means[24]. Rician bias correction was added to the first two in order to take advantage of the fact that the noise power is known exactly; the non-local means filter already makes use of thiss information. The Gaussian filter required a high value for σ_g to eliminate the noise, resulting in too much smoothing. Anisotropic diffusion works slightly better, but it still oversmoothes the image. Non-local means, on the other hand, does not blur the edges as much, but the noise is still very noticeable. Our method clearly outperforms the others, even though a little bit of high-frequency details are still lost.

Assuming that the original images are almost noise free, which is fair because they are the product of evaluating an expansion in spherical harmonics that was fitted to 94 directions, it is possible to quantitatively evaluate the method by calculating the symmetrized Kullback-Leibler divergence between the ODFs (cal-

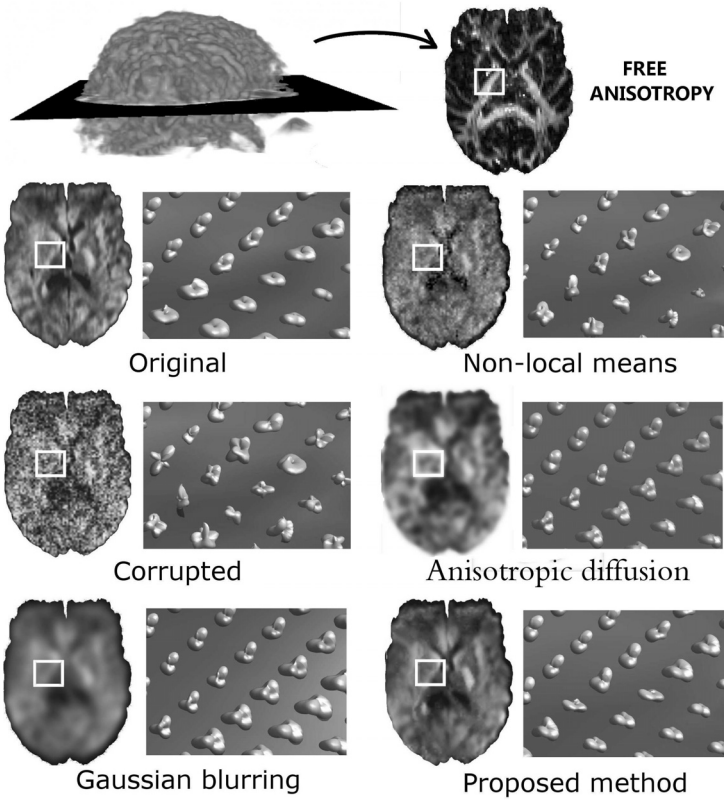


Fig. 3. For the original, corrupted and denoised versions of a sample axial slice: 16^{th} DW component (left) and ADC field around the internal capsule in a region of the slice (right). The white square marks the region for which the ADC field is displayed. The gradient direction for this DW component is zenith 0.9556 rad, azimuth 2.8768 rad.

culated using a spherical harmonic expansion of order four[25]) of the original, corrupted and denoised images. The results in table 2 show that the divergence decreases with the complexity of the method, as one would have expected. Finally, the fact that the mixture weights were optimized jointly with the image values allows us to plot the probability maps for the components of the mixture. Sometimes the maps correlate almost directly with anatomy (see figure 4).

4 Discussion

In this paper we use a statistical approach, Bayesian finite mixture models, to study the DW-MRI data field. A denoising application has been illustrated here, but the proposed framework is general and can be used in other DW-

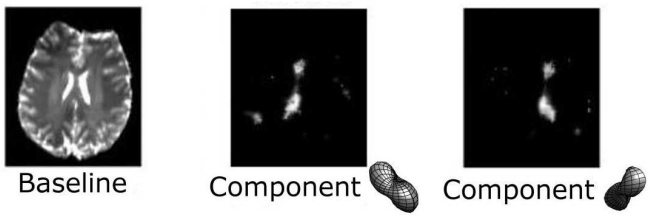


Fig. 4. Probability maps for the components of the GMM that predominate around the corona radiata.

Method	Noisy image	Gauss. blur	NL means	Anis. diff.	This study
Average	$9.23{\cdot}10^{-3}$	$7.31{\cdot}10^{-3}$	$5.86{\cdot}10^{-3}$	$5.28{\cdot}10^{-3}$	$4.40{\cdot}10^{-3}$
p value w.r.t. prev. method	N/A	$3.76{\cdot}10^{-4}$	$3.39{\cdot}10^{-3}$	$1.96{\cdot}10^{-4}$	$7.31{\cdot}10^{-8}$

Table 2. First row: average KL distance to ODF of original image, in order of increasing complexity. Second row: p values for paired t-tests between each method and the previous one (t-test at image level, as opposed to pixel level).

MRI applications. While most of the literature in this domain has focused on improving the fit of the individual DW-MRI data points, we provide a different perspective by providing a statistical framework for capturing the relationship of the neighboring diffusion profiles. The intrinsic multi-modal properties of the data make mixture models very suitable. Our model learns the prior on top of the distributions for the different spatial locations, which, to the best of our knowledge, has never been studied in this field.

Future research directions include:

- Studying more efficient algorithms for the learning and computing stages. Training the model takes approximately eight hours with our current Java implementation, whereas denoising a volume of size 128x128x55 voxels takes roughly two hours. Given that Java is an interpreted language and that the code was not optimized for speed, we believe that the execution time of the denoising algorithm could be reduced to 10-15 minutes.
- Finding an efficient way to maximize the marginal probability $p(\boldsymbol{x}(\boldsymbol{r}))$ rather than the joint probability $p(\boldsymbol{x}(\boldsymbol{r}), \boldsymbol{\pi}(\boldsymbol{r}))$ in the denoising algorithm. The problem is that calculating the marginal probability would require statistical sampling to compute the integral $p(\boldsymbol{x}(\boldsymbol{r})) = \int_{\mathbb{R}^C} p(\boldsymbol{x}(\boldsymbol{r}), \boldsymbol{\pi}(\boldsymbol{r}))d\boldsymbol{\pi}(\boldsymbol{r})$, which has no analytical solution. Because the denoising algorithm is already rather slow due to the large size of the data, and because optimizing the joint probability provides good results, $p(\boldsymbol{x}(\boldsymbol{r}), \boldsymbol{\pi}(\boldsymbol{r}))$ was maximized here.
- Studying whether the trained model could be used with HARDI images from another source. In principle, it should be possible to resample the HARDI data to the 30 directions used here and adapt the value of the class variances

to accommodate other image resolutions. However, the impact of changing the Le Bihan's factor is not clear. Validation on a different dataset would be required to address these issues.
- Exploring more general baseline models: we tested an approach based on assuming that the parameters $\{\boldsymbol{\mu}^c, \boldsymbol{\Sigma}^c\}$ are the realization of a Dirichlet process with level of resolution α and base measure G_0[26]. The main advantage is that the number of mixture components does not have to be predefined. Instead, new components are created dynamically to best accommodate the data. The likelihood with which new components are created is controlled by α(see [27] for details). Unfortunately, pilot experiments showed a very high sensitivity of the number of components to the level of resolution. Exploring possible modifications to tackle this problem remains as future work.
- Applying the model to more applications, such as fiber tracking.

Acknoledgements

This work was funded by the National Science Foundation (grant no. 0844566) and the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 RR021813 entitled Center for Computational Biology (CCB). The authors would like to thank Greig de Zubicaray, Kathie McMahon, Margaret Wright from the Center for Magnetic Resonance at the University of Queensland for acquiring the data. The first author would also like to thank the U.S. Department of State's Fulbright program for the funding.

References

1. Basser, P., Mattiello, J., LeBihan, D., et al.: Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. Journal of Magnetic Resonance Series B **103** (1994) 247–247
2. Pierpaoli, C., Jezzard, P., Basser, P., Barnett, A., Di Chiro, G.: Diffusion tensor MR imaging of the human brain. Radiology **201**(3) (1996) 637–648
3. Tuch, D., Reese, T., Wiegell, M., Makris, N., Belliveau, J., Wedeen, V.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine **48**(4) (2002) 577–582
4. Frank, L.: Anisotropy in high angular resolution diffusion-weighted MRI. Magnetic Resonance in Medicine **45**(6) (2001) 935–939
5. Tuch, D.: Q-Ball Imaging. Magnetic Resonance in Medicine **52**(6) (2004) 1358–1372
6. Assaf, Y., Freidlin, R., Rohde, G., Basser, P.: New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. Magnetic Resonance in Medicine **52**(5) (2004) 965–978
7. Tournier, J., Calamante, F., Gadian, D., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. Neuroimage **23**(3) (2004) 1176–1185
8. Behrens, T., Berg, H., Jbabdi, S., Rushworth, M., Woolrich, M.: Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? Neuroimage **34**(1) (2007) 144–155

9. Jian, B., Vemuri, B., Ozarslan, E., Carney, P., Mareci, T.: A novel tensor distribution model for the diffusion-weighted mr signal. NeuroImage **37**(1) (2007) 164–176
10. Nikou, C., Galatsanos, N., Likas, A.: A Class-Adaptive Spatially Variant Mixture Model for Image Segmentation. IEEE Transactions on Image Processing **16**(4) (2007) 1121–1130
11. King, M., Gadian, D., Clark, C.: A random effects modelling approach to the crossing-fibre problem in tractography. Neuroimage **44**(3) (2009) 753–768
12. Dinov, I.: Expectation maximization and mixture modeling tutorial. Statistics Online Computational Resource (December 2008)
13. Figueiredo, M., Jain, A.: Unsupervised Learning of Finite Mixture Models. IEEE Transactions on pattern analysis and machines intelligence (2002) 381–396
14. Sanjay-Gopal, S., Hebert, T.: Bayesian pixel classification using spatially variant finitemixtures and the generalized EM algorithm. IEEE Transactions on Image Processing **7**(7) (1998) 1014–1028
15. Zymnis, A., Kim, S., Skaf, J., Parente, M., Boyd, S.: Hyperspectral Image Unmixing via Alternating Projected Subgradients. In: Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on. (2007) 1164–1168
16. Besag, J.: On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society **48**(3) (1986) 259–302
17. Aja-Fernandez, S., Alberola-Lopez, C., Westin, C.: Signal LMMSE estimation from multiple samples in MRI and DT-MRI. Lecture Notes in Computer Science **4792** (2007) 368
18. Smith, S.: Fast robust automated brain extraction. Human Brain Mapping **17**(3) (2002) 143–155
19. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications. Magnetic Resonance in Medicine **56**(2) (2006) 395–410
20. Jones, D., Horsfield, M., Simmons, A.: Optimal Strategies for Measuring Diffusion in Anisotropic Systems by Magnetic Resonance Imaging. optimization **515–525** (1999)
21. Nowak, R.: Wavelet-based Rician noise removal for magnetic resonance imaging. Image Processing, IEEE Transactions on **8**(10) (1999) 1408–1419
22. Sijbers, J., Poot, D., Dekker, A., Pintjens, W.: Automatic estimation of the noise variance from the histogram of a magnetic resonance image. Physics in medicine and biology **52**(5) (2007) 1335–1348
23. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis and Machine Intelligence **12**(7) (1990) 629–639
24. Descoteaux, M., Wiest-Daessle, N., Prima, S., Barillot, C., Deriche, R.: Impact of Rician Adapted Non-Local Means Filtering on HARDI. In: Medical image computing and computer-assisted intervention. Volume 11. (2008) 122–129
25. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical Q-ball imaging. Magnetic Resonance in Medicine **58**(3) (2007) 497–510
26. Ferguson, T.: A Bayesian analysis of some nonparametric problems. Ann. Statist **1**(2) (1973) 209–230
27. Orbanz, P., Buhmann, J.: Nonparametric Bayesian image segmentation. International Journal of Computer Vision **77**(1) (2008) 25–45

Data-Driven Density Estimation applied to SPECT Subtraction Imaging for Epilepsy Diagnosis

Dieter A. Hahn[1], Volker Daum[1,2], Joachim Hornegger[1] and Torsten Kuwert[3]

[1] Chair of Pattern Recognition, Friedrich-Alexander University Erlangen-Nürnberg (FAU)
[2] Institute of Optics, Information and Photonics (Max-Planck Research Group), FAU
[3] Department of Nuclear Medicine, University Hospital, FAU

Abstract. In this article we present a subtraction imaging approach for the assessment of difference in the cerebral blood flow between intra- and inter-ictal SPECT images of epilepsy patients. The workflow consists of a rigid, automatic image registration of the SPECT images, an intensity normalization, and an alignment of the differences with an MRI. For the registration, the statistical measure of normalized mutual information is applied. The probability density estimation is sensitive to the input data, the sampling approach, and the kernels applied for Parzen-windowing. An adaptive scheme to estimate the required parameters is mandatory, because the system has to work reliably for a large number of images. We propose data-driven estimation techniques for a B-spline Parzen-window density estimation that is adapted to the variation within the random measures by an anisotropic binning approach. The optimal kernel widths are determined by a log-likelihood estimation.
The approach has been integrated into a commercially available software and applied to a collective of 26 epilepsy patients. Results are presented for a blind evaluation study with physicians from the Department of Nuclear Medicine of the University Hospital in Erlangen. The results show a good correlation of 81% between the certain outcomes of the proposed workflow and the standard procedure.

1 Introduction

In epilepsy surgery planning, inter- and intra-ictal SPECT images are acquired between epileptic seizures and closely afterwards. The goal is to find the location of the seizure onset by comparing the cerebral blood flow (CBF) inferred from the images. Standard side-by-side visual assessment techniques are not ideal, because the patient position and orientation may have changed between the acquisitions, and the images lack a normalization to standardized intensity values (e. g. as Hounsfield units in CT).

In [1], we introduced a workflow for SPECT epilepsy imaging. The basic idea is to use image subtraction similar to digital subtraction angiography (DSA) in order to visualize the differences between both images. The subtraction image then depicts changes in the CBF between both acquisitions. In order to be able to perform a difference operation on two SPECT images, we have to compensate

for patient motion and variations of the tracer uptakes between the acquisitions. A previously acquired MRI is integrated into the workflow and used to spatially localize the differences. In this article, we refine the concepts of the prior work and extend the registration, as well as the intensity normalization, by automatic, data-driven density estimation techniques.

The similarity measure used in the registration of the SPECT images has to be invariant to intensity variations, changes in blood flow activity, and structured noise. We apply normalized mutual information (NMI) [2] and use parameters that are adapted to the input data. This principally concerns the discrete probability density functions (PDF), which are computed by a novel, quasi-adaptive Parzen-window estimator based on cubic B-spline kernels. Structured noise in the background is handled by an automatic determination of background thresholds for both images, which can afterwards be reused within the intensity normalization. After the registration, we relate the intensities between both images by fitting an affine model into the joint PDF. The final result of the workflow is a subtraction image that can be fused with an MR scan.

We have applied the algorithm to a collective of 26 epilepsy patients and the results of the proposed workflow have been evaluated in a blind study.

2 Related Work

The workflow, presented in Hahn et al. [1], is based on the application of subtraction methods in the context of SPECT imaging. Other approaches make use of statistical parametric mapping, for example Chang et al. [3] and McNally et al. [4]. Their methods, however, require collectives of norm patients in order to identify those variations that are due to the epileptic disease. Koo et al. [5] proposed settings for the visualization of subtraction results in order to achieve a good correlation in the detection of the focal spots with other, well established techniques.

Image similarities can be modeled in a statistical framework. Here, the intensity values are regarded as random measures of an unknown distribution. Using nonparametric density estimation, efficient PDF estimators can be realized by a discretization of the Parzen-window technique [6, 7]. The density estimation is based on the work by Viola [8] and Hermosillo et al. [9]. Knops et al. [10] and Katkovnic and Shumulevich [11] investigated the effects of the kernel width parameter on the estimator and showed that state-of-the-art, isotropic binning is outperformed by adaptive techniques. Depending on the sampling pattern, numerical problems arise in the discretization of the estimator, as described by Maes [12] and Pluim et al. [13]. Thévenaz et al. [14] proposed quasi-random sampling based on Halton sequences in order to overcome these problems.

The structured noise within reconstructed medical images often poses problems to the registration, as the algorithm tends to align not only the interesting image content, but also the background noise. Some authors have tried to eliminate this problem by using intensity thresholds within the joint PDF [15, 16], or by masking the background region of the images [17]. Although this removes all influences of the background, mis-registrations with the background regions are partly or entirely disregarded.

3 Methods and Materials

This section is organized as follows. After a brief summary of the similarity measure for the registration in section 3.1, we present in part 3.2 an adaptive binning scheme that is used to re-quantize the image intensities for a given number of histogram bins. Based on a similar scheme, we describe in section 3.3 the automatic detection of the intensity thresholds for the background, which is assumed to contain structured noise. The density estimation is subject of part 3.4, where the efficient Parzen-window discretization using histograms is presented. The density estimation techniques are used for the registration, and afterwards, for the normalization of the SPECT intensities, as described in section 3.5.

3.1 Normalized Mutual Information

In an automatic, intensity-based image registration, a distance measure $\mathcal{D}$ is used as an objective function for the alignment between a reference image $\mathcal{R}$ and a template image $\mathcal{T}$. For a spatial transform $\Phi : \mathbb{R}^3 \mapsto \mathbb{R}^3$, which consists of rotation and translation parameters in our case, the term $\mathcal{T}_\Phi$ is used to refer to the transformed template image:

$$\mathcal{T}_\Phi(\boldsymbol{x}) = \mathcal{T}(\Phi(\boldsymbol{x})) \tag{1}$$

During the registration, we search for an optimal transform $\hat{\Phi}$ that minimizes the distance measure:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}}\, \mathcal{D}[\mathcal{R}, \mathcal{T}, \Phi] \tag{2}$$

Distance measures based on image intensity statistics are widely used for multimodal registration tasks and also for single modalities where the intensities are not normalized, for example the SPECT image pairs in our application. Based on Shannon's theory [18], the information content within the images is measured using the entropies of the marginal PDFs $p_\mathcal{R}$ and $p_{\mathcal{T}_\Phi}$, and the joint PDF $\boldsymbol{p}_{\mathcal{R},\mathcal{T}_\Phi}$:

$$\mathcal{H}(\mathcal{R}) = -\int_{\mathbb{R}} p_\mathcal{R}(r) \log p_\mathcal{R}(r)\ \mathrm{d}r \tag{3}$$

$$\mathcal{H}(\mathcal{T}_\Phi) = -\int_{\mathbb{R}} p_{\mathcal{T}_\Phi}(t) \log p_{\mathcal{T}_\Phi}(t)\ \mathrm{d}t \tag{4}$$

$$\mathcal{H}(\mathcal{R}, \mathcal{T}_\Phi) = -\int_{\mathbb{R}^2} \boldsymbol{p}_{\mathcal{R},\mathcal{T}_\Phi}(\boldsymbol{i}) \log \boldsymbol{p}_{\mathcal{R},\mathcal{T}_\Phi}(\boldsymbol{i})\ \mathrm{d}\boldsymbol{i} \tag{5}$$

where r, t, and $\boldsymbol{i} = (r, t)^T$ are intensity random measures of $\mathcal{R}$ and $\mathcal{T}$. In the following, we make use of the normalized mutual information (NMI) [2], which is less variant to overlap effects than the common mutual information, which was introduced by Wells et al. [19] and Maes et al. [20]:

$$\mathcal{D}^{\mathtt{NMI}}[\mathcal{R}, \mathcal{T}_\Phi] = -\frac{\mathcal{H}(\mathcal{R}) + \mathcal{H}(\mathcal{T}_\Phi)}{\mathcal{H}(\mathcal{R}, \mathcal{T}_\Phi)} \tag{6}$$

Here, $\mathcal{D}^{\mathtt{NMI}}$ is written as a distance measure, i. e. smaller values indicate a better result.

A common technique to estimate the intensity PDFs is Parzen-windowing. In a one-dimensional case with n random samples $x_1, x_2, \ldots, x_n$, the Parzen-window PDF estimator yields [7, 21]:

$$p_{\lambda,n}(x) = \frac{1}{n} \sum_{i=1}^{n} K_\lambda(x - x_i) \ , \tag{7}$$

with K_λ being the kernel PDF of width λ. Unfortunately, this approach has a high computational complexity and the storage requirements needed for large numbers of samples are high. If the random samples are discretized into a histogram, a discretization error is introduced on the one hand, but on the other hand, a lot of computations can be saved. The non-parametric estimator then resembles the behavior of a mixture model with as many components as bins. The n samples are stored in a discrete histogram h_n with b bins ($b > 1$). Here, $h_n(x_i)$ provides the fraction of samples that fall into the bin corresponding to x_i. $\hat{p}_{\lambda,n}$ is the discretized PDF estimator that differs from $p_{\lambda,n}$ in (7) by the application of histogram binning:

$$\hat{p}_{\lambda,n}(c_j) = \sum_{i=1}^{b} h_n(c_i) K_\lambda(c_j - c_i) = (h_n \star K_\lambda)(c_j) \approx p_{\lambda,n}(c_j) \ , \tag{8}$$

where c_j is the intensity value corresponding to the center of the j-th bin, and '$\star$' the convolution operator. We assume in the following, that $\hat{p}_{\lambda,n}$ is an approximation of its continuous counterpart, as indicated in (8).

3.2 Adaptive Binning Scheme

In data-driven approaches that estimate the optimal kernel width, one can observe that the result is directly related to the uncertainty within the data. Due to the discrete nature of histograms, this uncertainty is reflected by a varying smoothness or degenerations. Estimators using constant kernel widths cannot distinguish between regions of high and low certainty within one histogram. Therefore, several authors suggest making this parameter spatially variant. A disadvantage of adaptive, anisotropic kernel widths is the increased computational complexity for both the estimator and the formulation of its derivatives. In medical imaging, this increase in complexity is prohibitive. In addition, the efficient evaluation scheme (8) cannot be applied to estimators with varying kernel sizes. Therefore, we propose a trade-off in favor of a higher computational efficiency.

Instead of determining different kernel widths for an equidistantly spaced histogram, the image intensities are initially sampled into a histogram of varying bin widths. The corresponding bin centroids define a quantization characteristic, which can be used to map the input intensities to re-quantized output values. These, in turn, can be represented with an equidistantly spaced histogram. A density estimation on this re-quantized intensity space then does not have to account for different bin widths of the histogram and the convolution-based estimator (8) can be applied. The nonlinear mapping is computed as a preprocessing

step in the beginning, which means that it has to be computed only once for each image, but requires a distance measure that is invariant to this type of intensity transform, a property that is fulfilled by $\mathcal{D}^{\texttt{NMI}}$ (6).

In order to distribute the bin centers for the initial histogram with a minimal quantization error, we apply an approach introduced by Lloyd [22] and Max [23]. It minimizes the noise power $\mathcal{N}$ for a specific number of bins by an iterative refinement of the bin center locations. The spatial region of the i-th bin within the domain of the random variable is defined by the interval $[l_{i-1}; l_i]$ with the centroid c_i. The noise power of the re-quantization with respect to the signal PDF $p(x)$ is:

$$\mathcal{N} = \sum_{i=1}^{b} \int_{l_{i-1}}^{l_i} (c_i - x)^2 p(x) \, \mathrm{d}x \; . \tag{9}$$

Lloyd [22] proposed a fixed point iteration scheme to numerically minimize (9) with respect to the bin intervals and centroids. Again, $p(x)$ is unknown, but can simply be exchanged by the histogram of the entire image with full intensity resolution, or a suitable Parzen-window estimator.

3.3 Background Threshold Detection

Tomographic, medical images are the result of discrete, modality-specific reconstruction methods that are based on physical measurements. In practice, these measurements are affected by detector noise and many physical effects, which may impair the reconstruction result. Problems for image registration algorithms especially arise from structured noise in the reconstructed images. Thévenaz et al. [17] presented a robust technique to distinguish between the object and background region within an image. They used the aforementioned Max-Lloyd quantization algorithm on a low pass filtered version of the image. Combined with the filtering, the algorithm computes the bin widths for a discrete histogram of two bins. The boundary between these two bins is assumed to separate intensities in the background from object values. The authors used the algorithm to determine the background region within PET images. We apply the background values to down-weight the corresponding region within the joint PDF – instead of thresholding it – in order to reduce the influence of the background and the contained noise. The background thresholds are also used for the intensity normalization 3.5.

3.4 Density Estimation

Very common choices for the kernel PDF K_λ are the Gaussian or cubic B-spline [24, 25]. Using a cubic B-spline B yields the following Parzen-window kernel K_λ^{B}:

$$K_\lambda^{\mathrm{B}}(x) = \frac{1}{\lambda} \mathrm{B}\left(\frac{x}{\lambda}\right) \tag{10}$$

The B-spline function is commonly defined recursively by the Cox-de Boor recursion formula, however, in the case of degree three, the kernel may be written

as a non-recursive function:

$$K_\lambda^{\mathrm{B}}(x) = \begin{cases} \frac{1}{\lambda}\left(\frac{4}{3} - 2\frac{|x|}{\lambda} + \frac{x^2}{\lambda^2} - \frac{|x|^3}{6\lambda^3}\right) & , \text{ if } \frac{|x|}{\lambda} \in [1,2[\\ \frac{1}{\lambda}\left(\frac{2}{3} - \frac{x^2}{\lambda^2} + \frac{|x|^3}{2\lambda^3}\right) & , \text{ if } \frac{|x|}{\lambda} \in [0,1[\\ 0 & , \text{ otherwise} \end{cases} \tag{11}$$

K_λ^{B} can be discretized either by sampling of the kernel values or recursive filtering. This specific window function has some advantages over, for instance, a Gaussian: it has a local support and fulfills the partition of unity constraint [26].

A necessary requirement for the implementation of the estimator is the specification of λ. Unfortunately, this parameter is dependent on the data, i. e. the values and the number of the random samples. In order to get an optimal PDF estimator, it is necessary to apply data-driven estimation techniques. A leave-one-out estimator is usually plugged into a log-likelihood function with respect to the kernel width in order to measure how good it resembles the missing data. Let $p^j_{\lambda,n-1}$ be the estimator after deleting the j-th sample. The resulting log-likelihood objective function then yields [27]:

$$\mathcal{L}(\lambda) = \sum_{j=1}^{n} \log p^j_{\lambda,n-1}(x_j) \ . \tag{12}$$

An optimal value $\hat{\lambda}$ for the kernel width maximizes $\mathcal{L}$

$$\hat{\lambda} = \operatorname*{argmax}_{\lambda} \mathcal{L}(\lambda) \ . \tag{13}$$

To optimize (13), it is necessary to search for a zero crossing of the derivative of (12) with respect to λ. The problem can be solved using an iterative, nonlinear optimization scheme, e. g. Newton's method. The derivatives of $\mathcal{L}$ with respect to λ are:

$$\begin{aligned} \frac{\partial}{\partial\lambda}\mathcal{L}(\lambda) &= \sum_{j=1}^{n} \frac{1}{\hat{p}^j_{\lambda,n-1}(x_j)} \frac{\partial}{\partial\lambda}\hat{p}^j_{\lambda,n-1}(x_j) \\ \frac{\partial^2}{\partial^2\lambda}\mathcal{L}(\lambda) &= \sum_{j=1}^{n} -\frac{1}{\left(\hat{p}^j_{\lambda,n-1}(x_j)\right)^2}\left(\frac{\partial}{\partial\lambda}\hat{p}^j_{\lambda,n-1}(x_j)\right)^2 \\ &\quad + \frac{1}{\hat{p}^j_{\lambda,n-1}(x_j)} \frac{\partial^2}{\partial^2\lambda}\hat{p}^j_{\lambda,n-1}(x_j) \end{aligned} \tag{14}$$

For the B-spline kernel function K_λ^{B}, the derivatives of the estimator yield:

$$\begin{aligned} \frac{\partial}{\partial\lambda}\hat{p}^j_{\lambda,n-1}(x_j) &= \sum_{i=1}^{b} h^j_{n-1}(c_i) \frac{\partial}{\partial\lambda} K_\lambda^{\mathrm{B}}(x_j - c_i) \\ \frac{\partial^2}{\partial^2\lambda}\hat{p}^j_{\lambda,n-1}(x_j) &= \sum_{i=1}^{b} h^j_{n-1}(c_i) \frac{\partial^2}{\partial^2\lambda} K_\lambda^{\mathrm{B}}(x_j - c_i) \end{aligned} \tag{15}$$

Similar to the kernel function (11), we formulate the derivatives of K_λ^{B} for three cases.

Case 1: $\frac{|x|}{\lambda} \in [1, 2[$

$$\begin{aligned} \frac{\partial}{\partial\lambda} K_\lambda^{\mathrm{B}}(x) &= \frac{1}{\lambda^2}\left[\frac{2|x|}{\lambda} - \frac{2x^2}{\lambda^2} + \frac{|x|^3}{2\lambda^3} - \lambda K_\lambda^{\mathrm{B}}(x)\right] \\ \frac{\partial^2}{\partial^2\lambda} K_\lambda^{\mathrm{B}}(x) &= \frac{2}{\lambda^3}\left[-\frac{2|x|}{\lambda} + \frac{3x^2}{\lambda^2} - \frac{|x|^3}{\lambda^3} - \lambda^2 \frac{\partial}{\partial\lambda} K_\lambda^{\mathrm{B}}(x)\right] \end{aligned} \tag{16}$$

Case 2: $\frac{|x|}{\lambda} \in [0, 1[$

$$\begin{aligned} \frac{\partial}{\partial\lambda} K_\lambda^{\mathrm{B}}(x) &= \frac{1}{\lambda^2}\left[\frac{2x^2}{\lambda^2} - \frac{3|x|^3}{2\lambda^3} - \lambda K_\lambda^{\mathrm{B}}(x)\right] \\ \frac{\partial^2}{\partial^2\lambda} K_\lambda^{\mathrm{B}}(x) &= \frac{2}{\lambda^3}\left[\frac{3|x|^3}{\lambda^3} - \frac{3x^2}{\lambda^2} - \lambda^2 \frac{\partial}{\partial\lambda} K_\lambda^{\mathrm{B}}(x)\right] \end{aligned} \tag{17}$$

Case 3: $\frac{|x|}{\lambda} \notin [0, 2[$

$$\frac{\partial}{\partial\lambda} K_\lambda^{\mathrm{B}}(x) = \frac{\partial^2}{\partial^2\lambda} K_\lambda^{\mathrm{B}}(x) = 0 \tag{18}$$

The equations (15) are again convolutions with the partial derivatives for the

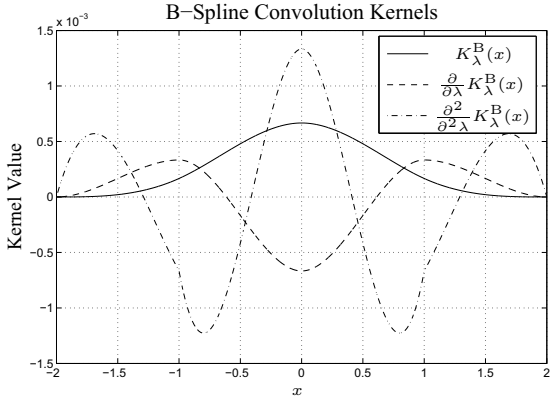
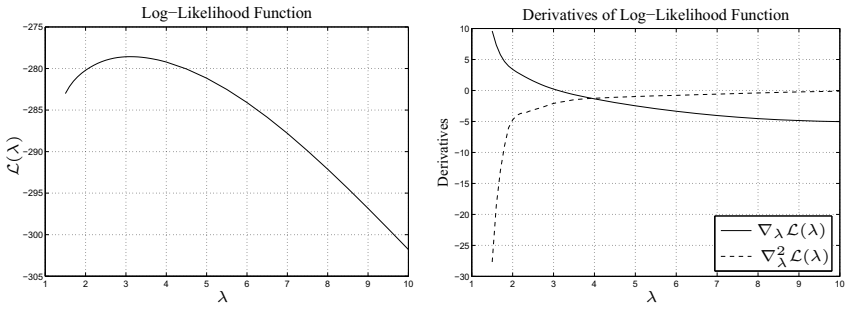


Figure 1. Convolution kernels for the kernel width estimation using a B-spline window function K_λ^{B} printed for a kernel width $\lambda = 1$. The kernel function, the first derivative, and the second derivative with respect to λ are plotted within the locally supported region.

kernel width of the Parzen-window kernel. Due to its local support, the spline kernel is very suitable for an implementation of the log-likelihood optimization using a discrete convolution operator and the kernel functions shown in Figure 1. The multivariate density kernel width estimation is realized analogue to the 1-D



(a) (b)

Figure 2. (a) Parzen-window log-likelihood function for a B-spline kernel K_λ^{B} cross-validation and an estimation of 100 samples drawn from a normal distribution with $\lambda = 4$. (b) Shows the corresponding first and second order derivative with respect to λ.

case and the 1-D kernels can be applied subsequently to each dimension of the histogram, because the multivariate B-spline kernel is separable. Figure 2 shows example curves that can typically be observed for the log-likelihood optimization of the kernel width. In this example, the width of a cubic B-spline kernel has been adapted to 100 sample values drawn from a Gaussian distribution with mean zero and a variance of 4.

3.5 Intensity Normalization

As mentioned above, for a correct interpretation of the differences between the SPECT images, the intensities have to be normalized to a common intensity range. This is necessary due to different acquisition times and changes in the overall tracer uptake within the human body. We model the mapping by an affine intensity transform similar to the proposals of Liao et al. [28]. In order to be invariant to the intensities of the background, we restrict the affine mapping to the region of the joint PDF above the background thresholds, i. e. the probabilities for joint intensities each belonging to intensities corresponding to brain tissue. The thresholds are determined using the approach described in section 3.3. The components of the intensity mapping are computed by linear regression within the joint PDF. This leads to an affine intensity transformation and, in the ideal case, a clustering at the diagonal entries of the joint PDF after the mapping.

4 Results

The proposed method has been applied to a collective of 26 epilepsy patients and assessed by physicians. Each patient underwent the standard diagnostic procedures, and, from the patient charts, the location as well as the number of the focal spots were known. The images have been anonymized within the clinics

and handed blinded to us. The subtraction images were then generated with the proposed workflow and randomly ordered into a set of evaluation protocols. For the registrations, the number of bins has been computed automatically using a threshold for the quantization error of 0.05%. The number of samples has been 10% of the overlap domain, and at least 10.000 for lower resolutions in the multi-level optimization. No misregistrations were encountered during the experiments. Based on a fusion of the subtraction image with the corresponding MRI, the physicians had to specify the locations and the number of the focal spots, their certainty, and a rating of the image quality. No additional data was provided for the evaluation. The results of two physicians from the Department of Nuclear Medicine, University Hospital Erlangen, showed a good correlation of the proposed method with the conventional procedure. On an ordinal scale ranging from *very uncertain*, *uncertain*, *certain* to *very certain*, 45% of the results from the subtraction workflow were classified as *certain*. The observers reached a correlation value of 81% for correctly locating the focal spot in those cases. In the *uncertain* cases, still 53% of the focal spots have been located correctly. The overall image quality was rated *good* on an ordinal scale between *very bad*, *bad*, *good*, and *very good*. The intra-observer variability was 88% for the results computed from two evaluation passes.

An example for the density estimation and the normalization, using an affine model, is shown in Figure 3. Resulting subtraction images fused with the corresponding MR images are presented in Figure 4.

5 Discussion

In this article we have presented data-driven parameter estimation techniques for the Parzen-window estimation of intensity distributions, which are required for statistical image similarity measures. The proposed algorithms have been integrated into a normalized mutual information registration and applied to align intra- with inter-ictal SPECT images of epilepsy patients. After the registration, the estimated densities are used for the intensity normalization between the two SPECTs in order to allow for an image subtraction. The difference image is then fused with an MRI to spatially localize the focal spots.

Regarding the evaluation, in the cases where the physicians indicated a high certainty, the results showed a high correlation with the standard evaluation method. In cases of low certainty, an inspection of the subtraction results revealed no failure of the algorithm, but instead often showed several weak focal spots that gave no clear information about the seizure onset. This might lead to the conclusion that the diagnosis of these patients with only the SPECT images is generally uncertain also in the standard approach. An evaluation based on real ground truth data (i. e. results from surgery) would be necessary to confirm this. Please note that only the *difficult* cases are undergoing a further analysis using multiple SPECT and MR images. A major advantage of the presented algorithm is the low amount of time that is required for the physician to assess the patient data in the beginning of the diagnosis. This suggests to combine the approach with the standard procedure, giving the physician a good indication of interesting regions within the data.

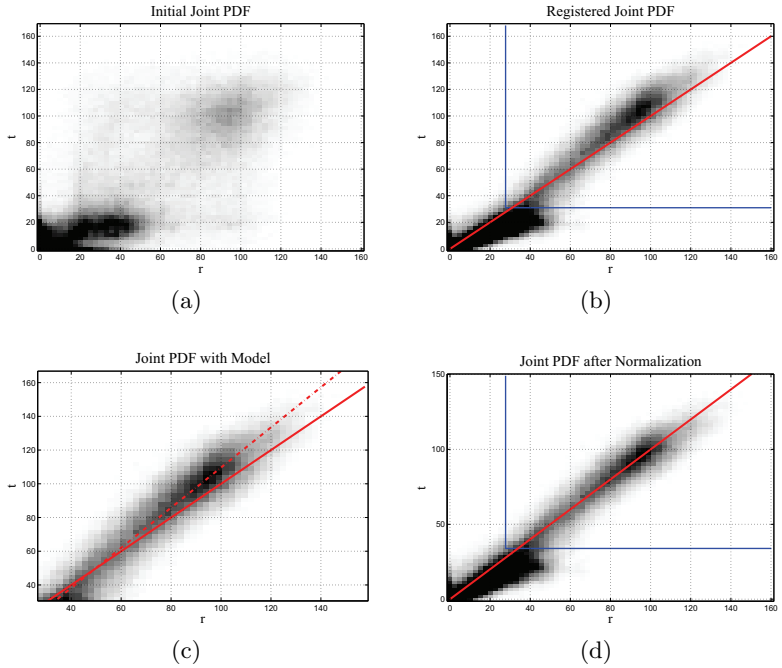


Figure 3. The figures show the joint PDF of two input SPECT images (a) before, (b) after the registration, (c) with the model fitted into the joint PDF without background content, and (d) the joint PDF after applying the intensity normalization. The figures also show the diagonal through the histogram space (solid red line), the fitted affine model (dashed red line), and the threshold region for the background (solid blue line).

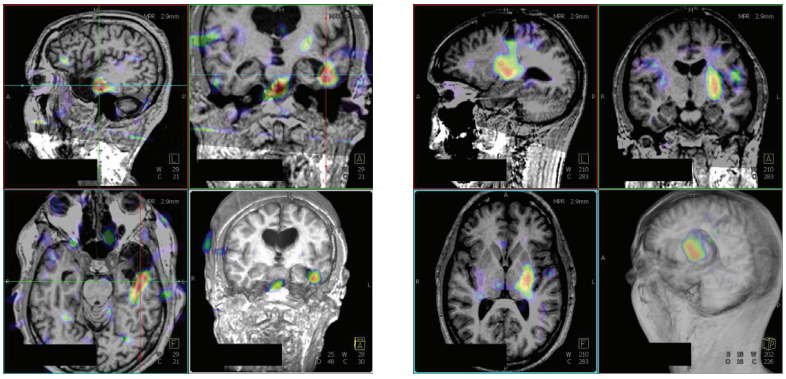


Figure 4. Two example patients taken from the collective. The images show the SPECT subtraction image fused with the corresponding MRI.

6 Ackowledgements

The authors would like to thank Prof. Dr. med. H. Stefan (Department of Neurology, University Hospital, Erlangen) for providing the patient data, PD Dr.

med. R. Linke (Department of Nuclear Medicine, University Hospital, Erlangen) for participating in the evaluation, and Dipl.-Ing. P. Ritt (Department of Nuclear Medicine, University Hospital, Erlangen) for his efforts in the patient data selection. We are also thankful to HipGraphics for providing the volume rendering software InSpace.

References

1. Hahn, D.A., Daum, V., Hornegger, J., Bautz, W., Kuwert, T.: Difference imaging of inter- and intra-ictal SPECT images for the localization of seizure onset in epilepsy. In Frey, E.C., ed.: IEEE Nuclear Science Symposium and Medical Imaging Conference, Honolulu, HI, USA, IEEE Nuclear & Plasma Sciences Society (October 2007) 4331–4335
2. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition **32**(1) (1999) 71–86
3. Chang, D.J., Zubal, I.G., Gottschalk, C., Necochea, A., Stokking, R., Studholme, C., Corsi, M., Slawski, J., Spencer, S.S., Blumenfeld, H.: Comparison of statistical parametric mapping and SPECT difference imaging in patients with temporal lobe epilepsy. Epilepsia **43** (2002) 68–74
4. McNally, K.A., Paige, A.L., Varghese, G., Zhang, H., Novotny, E.J., Spencer, S.S., Zubal, I.G., Blumenfeld, H.: Seizure localization by ictal and postictal SPECT. Epilepsia **46**(9) (2005) 1–15
5. Koo, C.W., Devinsky, O., Hari, K., Balasny, J., Noz, M.E., Kramer, E.L.: Stratifying differences on ictal/interictal subtraction SPECT images. Epilepsia **44**(3) (2003) 379–386
6. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Oxford (1997)
7. Parzen, E.: On the estimation of probability density function and mode. Annals of Mathematical Statistics **33**(3) (1962) 1065–1076
8. Viola, P.: Alignment by Maximization of Mutual Information. PhD thesis, Massachusetts Institute of Technology (MIT) (1995)
9. Hermosillo, G., Chef d'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. International Journal of Computer Vision **50**(3) (2002) 329–343
10. Knops, Z.F., Maintz, J.B.A., Viergever, M.A., Pluim, J.P.W.: Normalized mutual information based registration using k-means clustering and shading correction. Medical Image Analysis **10**(3) (2006) 432–439
11. Katkovnik, V., Shumulevich, I.: Kernel density estimation with adaptive varying window size. Pattern Recognition Letters **23**(14) (2002) 1641–1648
12. Maes, F.: Segmentation and registration of multimodal medical images: From theory, implementation and validation to a useful tool in clinical practice. PhD thesis, Catholic University of Leuven (1998)
13. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Interpolation artefacts in mutual information-based image registration. Comput. Vis. Image Underst. **77**(9) (2000) 211–232
14. Thévenaz, P., Bierlaire, M., Unser, M.: Halton sampling for image registration based on mutual information. Sampling Theory in Signal and Image Processing **7**(2) (2008) 141–171
15. Rohlfing, T., Beier, J.: Improving reliability and performance of voxel-based registration by coincidence thresholding and volume clipping. In Hawkes, D.J., Hill, D.L.G., Gaston, R., eds.: Proceedings of Medical Image Analysis and Understanding 99, King's College (1999) 165–168

16. Rohlfing, T.: Multimodale Datenfusion für die bildgesteuerte Neurochirurgie und Strahlentherapie. PhD thesis, Technical University Berlin (2000)
17. Thévenaz, P., Ruttiman, U.E., Unser, M.: A pyramid approach to subpixel registration based on intensity. IEEE Transactions on Image Processing **7**(1) (1998) 27–41
18. Shannon, C.E.: A mathematical theory of communication (parts 1 and 2). The Bell System Technical Journal **27** (1948) 379–423,623–656
19. Wells III, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis **1**(1) (1996) 35–51
20. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging **16**(2) (April 1997) 187–198
21. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. John Wiley & Sons, New York (2001)
22. Lloyd, S.P.: Least squares quantization in PCM. IEEE Transactions on Information Theory **28**(2) (1982) 129–137
23. Max, J.: Quantizing for minimal distortion. IRE Transactions on Information Theory **6**(1) (1960) 7–12
24. Unser, M., Aldroubi, A., Eden, M.: B-Spline signal processing: Part I - theory. IEEE Transactions on Signal Processing **41**(2) (1993) 821–833
25. Unser, M., Aldroubi, A., Eden, M.: B-Spline signal processing: Part II - efficient design and applications. IEEE Transactions on Signal Processing **41**(2) (1993) 834–848
26. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. IEEE Transactions on Image Processing **9**(12) (2000) 2083–2099
27. Chow, Y.S., Geman, S., Wu, L.D.: Consistent cross-validated density estimation. The Annals of Statistics **11**(1) (1983) 25–38
28. Liao, Y.L., Chiu, N.T., Weng, C.M., Sun, Y.N.: Registration and normalization techniques for assessing brain functional images. Biomed Eng Appl Basis Comm **15** (June 2003) 87–94

PCA Regularized Nonrigid Registration for PET/MRI Attenuation Correction

Volker Daum[1,2], Dieter A. Hahn[1], Joachim Hornegger[1,3], and Torsten Kuwert[4]

[1] Chair of Pattern Recognition, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany
[2] Institute of Optics, Information and Photonics (Max-Planck Research Group), FAU
[3] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Erlangen, Germany
[4] Department of Nuclear Medicine, University Hospital, FAU

Abstract. An open issue in the use of hybrid PET / MRI scanners is the attenuation correction of the PET image. In order to solve this problem, we propose to perform a nonrigid registration of an atlas CT image with the MRI. The registered atlas CT contains the information about the tissue densities necessary for the attenuation correction.
In multi-modal, nonrigid image registration, the correspondence between the intensity values is not known a priori. Statistical, multi-modal distance measures determine this correspondence during the registration solely from the intensity distributions. Without the incorporation of prior knowledge this may lead to wrong results, such as the alignment of the skull with brain tissue, or the skin with fat surrounding the skull. Therefore, we propose a novel, PCA-based regularization of the nonrigid registration. This limits the possible registration results to morphologically plausible deformations. The model is constructed such that it is invariant to global translations in the registration. Thus, the registration is less dependent on the initial, rigid preregistration.
Results are presented on a database of 18 CT datasets for the training of a PCA deformation model. MR images of the same patients have been rigidly registered with the corresponding CT datasets, which are used as ground truth for the tests. The evaluation is performed using a leave-one-out cross-validation by registering an atlas onto the total of 38 MRI datasets and comparing the deformed atlas with the ground truth CT of the patient. Results indicate a better performance of the proposed approach compared to the standard. On average, the mean squared error is decreased by 18% and the sensitivity for correct soft tissue and bone alignment is increased by 4%.

1 Introduction

The advent of hybrid scanners, for example the combination of PET (Positron Emission Tomography) and CT (Computed Tomography) imaging within one machine, has brought many new possibilities to the field of medical imaging, such as the invention of highly specific tumor markers. However, the superior

tissue contrast and the large variety of different sequences offered by MRI make it desirable to replace the CT with an MRI (Magnetic Resonance Imaging) scanner in such hybrid systems. Although there are still technical difficulties to overcome, first combined PET / MRI scanners for acquisition of the human head have already been built, and it is only a matter of time until they enter the market.

One of the difficulties to overcome in PET / MRI hybrid scanners is the attenuation correction of the PET image. In PET, the aim is to measure the concentration of a radioactive marker within the patient body. The quantity that the machine can actually measure is the radiation emitted by the tracer. The rays are attenuated by the anatomical structures while traveling through the human body. Therefore, it is necessary to provide an attenuation map for each acquisition in order to perform an attenuation correction of the dose measured in the PET image. The map can be created, for instance, from a CT, where the relation between the intensities within the image and the tissue densities is known.

The values measured by MRI, however, are not related to the attenuation, therefore, no straightforward solution is currently available for the attenuation correction in case of a hybrid PET / MRI scanner. In the following, we propose a method based on the registration between an atlas CT and the MRI image. The deformed atlas image then replaces the missing CT of the patient and can be used for the attenuation correction.

The multi-modal, nonrigid registration, which is used to perform such an atlas registration, offers many degrees of freedom in the spatial domain. It is more difficult than the mono-modal registration, because the relation between the image intensities is not known a priori. This may lead to mis-registrations where a low value for the distance measure indicates a good alignment of the images, however, the deformation may not be correct in the physical and morphological sense. Hence, we propose a novel regularization of the nonrigid registration process that incorporates prior knowledge in terms of a deformation model. This morphologically-based regularization utilizes a PCA (Principal Component Analysis) of the previously acquired deformation fields computed from CT images of a collective of patients. These registrations performed within the CT modality use a sum-of-squared differences similarity measure. The intra-modal registrations are, in our experience, better conditioned than the multi-modal, nonrigid registrations between CT and MRI. The result of this learning phase is an atlas CT, together with a model for the variations within the deformations. This prior knowledge is used to constrain the registration between the MRI of the hybrid scanner with the atlas CT to penalize morphologically improbable deformations.

This article is organized as follows. First, we introduce related work regarding the attenuation correction, morphological models in image registration, and the nonrigid registration. The methods provide information about the registration, distance measures, and the model generation. In the final section we present a leave-one-out cross-validation of 18 CT images used for the atlas CT creation and applied to MR images, and discuss the results.

2 Related Work

A very recent survey of PET / MRI attenuation techniques can be found in [1]. Usually, the employed methods are categorized into segmentation- or classification-based approaches. In segmentation-based correction, the intensities of the MRI image are directly transformed into an attenuation map. Using an atlas for the correction, an atlas CT is transformed into the space of the MRI to provide the tissue densities. In addition, combined methods can be applied that first perform an atlas registration and use additional knowledge from the atlas to improve the results of a classification approach [2].

Statistical morphological constraints in image registration have already been employed. Wang and Staib [3] describe a method that generates a sparse PCA-based model on a set of boundary points that they use to constrain the dense nonrigid registration. Kim et. al.[4] construct a dense PCA-based deformation model from registrations with a standard registration approach. The model is used to generate a large set of sample images which are then compared to the reference image in order to find a good starting position for a standard registration approach. An alternative to a simple PCA model is proposed by Xue and Shen [5]. They, instead, use a wavelet PCA that has the advantage to also capture very local and fine grained deformations. Nevertheless, the model is only used for an initial registration followed by an unconstrained nonrigid registration.

In this work we focus on a nonrigid, nonparametric atlas registration of a CT dataset with MRI, similar to the approaches presented in [6, 7]. As distance measure we employ the mutual information (MI), based on the works of Viola [8] and Hermosillo [6]. As regularization term we employ a curvature term introduced by Fischer and Modersitzki [9].

3 Methods

In the following, we present the applied nonrigid registration framework 3.1 that is used both to generate the atlas CT, and to register the MRI with the atlas CT. In section 3.2, we describe the applied intensity distance measures, which are used as objective functions for the registration. The regularization of the registrations is based on the curvature of the deformation field 3.3, which is then supplemented by the deformation model created with the PCA on the sample deformation fields. The novel regularization approach is described in 3.4, followed by the modeling of the translation invariance 3.5.

3.1 Nonrigid Registration Framework

In nonrigid, nonparametric registration the dense deformation field $\mathbf{u}$ is calculated between the spatial positions of each voxel. It is determined by minimizing a distance measure $\mathcal{D}$, which evaluates the quality of the match between the moving image M and the fixed image F. Its optimization is subject to a smoothness constraint $\mathcal{S}$ to ensure that the resulting deformation does not contain any

cracks, ridges, or folds. This constraint is usually incorporated as a penalty term weighted by a parameter $\alpha \in \mathbb{R}_{>0}$, i.e. lower values of α will result in a less smooth deformation field, but a better match and vice versa. The optimization of the distance alone is ill-posed and the addition of the smoothness constraint is related to classical Tikhonov regularization [10]:

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmin}}\, \mathcal{E}(F, M, \mathbf{u}) = \underset{\mathbf{u}}{\operatorname{argmin}}\, \mathcal{D}(F, M_{\mathbf{u}}) + \alpha \mathcal{S}(\mathbf{u}) \qquad (1)$$
$$M_{\mathbf{u}}(x) = M(x - \mathbf{u}(x))$$

The optimization problem is to find a minimizer for the functional (1) in the space of all deformation fields $\mathcal{U}$. If we assume that $\mathcal{E}$ is sufficiently smooth and differentiable, and with appropriate boundary conditions, we can apply the calculus of variations to find a minimizer $\hat{\mathbf{u}}$. For the direction $\mathbf{v} \in \mathcal{U}$ of the first variation, the Gâteaux derivative of (1) is defined as:

$$\delta\mathcal{E}(F, M, \mathbf{u} \circ \mathbf{v}) = \lim_{\epsilon \mapsto 0} \frac{\mathcal{E}(F, M, \mathbf{u} + \epsilon\mathbf{v}) - \mathcal{E}(F, M, \mathbf{u})}{\epsilon} = \left.\frac{\mathrm{d}\mathcal{E}(F, M, \mathbf{u} + \epsilon\mathbf{v})}{\mathrm{d}\epsilon}\right|_{\epsilon=0} \qquad (2)$$

For the existence of a minimizer for (1), it is necessary that the Gâteaux derivative vanishes for all variations $\mathbf{v}$: $\delta\mathcal{E}(F, M, \hat{\mathbf{u}} \circ \mathbf{v}) = 0$. If $\mathcal{U}$ is assumed to be a Hilbert space that defines a scalar product, the gradient of the functional with respect to the optimal displacement vanishes, $\nabla_{\mathcal{U}}\mathcal{E}(F, M, \hat{\mathbf{u}}) = 0$, and the minimizer is a solution to the Euler-Lagrange equations associated with this problem:

$$\nabla_{\mathcal{U}}\mathcal{E}(F, M, \mathbf{u}) = \nabla_{\mathcal{U}}\mathcal{D}(F, M_{\mathbf{u}}) + \nabla_{\mathcal{U}}\mathcal{S}(\mathbf{u}) \qquad (3)$$

As solver, we employ a Newton-type method that uses a numeric approximation of the second derivative of $\mathcal{D}$.

3.2 Distance Measures

For the mono-modal registration between the CT images, we apply the widely known sum-of-squared differences measure. It is based on the assumption that the intensities of corresponding tissue within the two images are equal, or differ by noise at the utmost.

$$\mathcal{D}_{\text{SSD}}(F, M_{\mathbf{u}}) = \frac{1}{|\Omega|} \int_{\Omega} \left(M_{\mathbf{u}}(\mathbf{x}) - F(\mathbf{x})\right)^2 \, \mathrm{d}\mathbf{x} \qquad (4)$$

with Ω being the spatial domain of the overlap between F and $M_{\mathbf{u}}$.

Distance measures based on image intensity statistics are widely used for multi-modal registration tasks. Based on Shannon's theory [11], the information content within the images is measured using the entropies of the marginal PDFs

p_F and $p_{M_\mathbf{u}}$, and the joint PDF $\mathbf{p}_{F,M_\mathbf{u}}$:

$$\mathcal{H}(F) = -\int_{\mathbb{R}} p_F(f) \log p_F(f) \, \mathrm{d}f \tag{5}$$

$$\mathcal{H}(M_\mathbf{u}) = -\int_{\mathbb{R}} p_{M_\mathbf{u}}(m) \log p_{M_\mathbf{u}}(m) \, \mathrm{d}m \tag{6}$$

$$\mathcal{H}(F, M_\mathbf{u}) = -\int_{\mathbb{R}^2} \mathbf{p}_{F,M_\mathbf{u}}(\mathbf{i}) \log \mathbf{p}_{F,M_\mathbf{u}}(\mathbf{i}) \, \mathrm{d}\mathbf{i} \tag{7}$$

where f, m, and $\mathbf{i} = (f, m)^T$ are intensity random measures of F and M. In the following, we make use of the MI, which was introduced by Wells et. al.[12] and Maes et. al.[13]:

$$\begin{aligned} \mathcal{D}_{\mathrm{MI}}(F, M_\mathbf{u}) &= -\mathrm{MI}(F, M_\mathbf{u}) \\ &= \mathcal{H}(F, M_\mathbf{u}) - \mathcal{H}(F) - \mathcal{H}(M_\mathbf{u}) \\ &= \int_{\mathbf{R}^2} \mathbf{p}_{F,M_\mathbf{u}}(\mathbf{i}) \log \frac{\mathbf{p}_{F,M_\mathbf{u}}(\mathbf{i})}{p_F(f) p_{M_\mathbf{u}}(m)} \, \mathrm{d}\mathbf{i} \end{aligned} \tag{8}$$

Here, $\mathcal{D}_{\mathrm{MI}}(M_\mathbf{u}, F)$ is written as a distance measure, i.e. smaller values indicate a better result.

3.3 Curvature Regularizer

The choice of a suitable smoother depends on the type of application. Common regularization techniques are based on Dirichlet, elasticity, fluidal, and higher order functionals. Among the latter ones, curvature regularization is an approach that features some advantages for medical image registration [14]:

$$\mathcal{S}_{\mathrm{CURV}}(\mathbf{u}) = \int_\Omega |\Delta \mathbf{u}|^2 \, \mathrm{d}\mathbf{x} \; . \tag{9}$$

This regularization term does not penalize affine transformations and leads to smooth displacement fields.

3.4 PCA Regularization

To generate the proposed morphological model, a series of mono-modal registrations is performed on CT images. The images are rigidly aligned, before a nonrigid registration is employed, which yields the training deformations. The mono-modal is considered to be more robust than the multi-modal registration, especially as one has to deal with less local minima during the optimization. For n input images, the resulting sample deformation fields $\mathbf{w}_i$ with $i = 1, \ldots, n$ are then used to extract the mean deformation and the principal modes of variation

by means of a PCA.

$$\begin{aligned} \bar{\mathbf{w}} &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i \\ \mathbf{W} &= ((\mathbf{w}_1-\bar{\mathbf{w}}),\ldots,(\mathbf{w}_n-\bar{\mathbf{w}})) \\ \mathbf{W}\mathbf{W}^T\mathbf{v}_i &= \lambda_i\mathbf{v}_i \\ \text{s.t. } |\mathbf{v}_i|^2 &= 1 \end{aligned} \tag{10}$$

where $\bar{\mathbf{w}}$ denotes the mean, and $\mathbf{W}\mathbf{W}^T$ the covariance matrix of the deformation. The Eigenvalue / Eigenvector decomposition of $\mathbf{W}\mathbf{W}^T$ (10) is performed as described in Murase and Lindenbaum [15]. The resulting Eigenvectors $\mathbf{v}_i$ form an orthonormal vector space, which is an important property for the following article. Choosing the m components with the largest Eigenvalues λ_i and arranging them in a matrix $\mathbf{V}$ our model consists of the components $\bar{\mathbf{w}}$ and $\mathbf{V}=(\mathbf{v}_1,\ldots,\mathbf{v}_m)$. Using this model, the registration energy (1) is then supplemented by an additional regularization term $\mathcal{P}$, which enforces the result to be close to the model space.

$$\begin{aligned} \min_{\mathbf{u}}\mathcal{E}(F,M,\mathbf{u}) &= \mathcal{D}(F,M_{\mathbf{u}})+\alpha\mathcal{S}(\mathbf{u})+\beta\mathcal{P}(\mathbf{u}) \\ \mathcal{P}(\mathbf{u}) &= \frac{1}{s}\left(\mathbf{u}-\left(\bar{\mathbf{w}}+\mathbf{V}\mathbf{V}^T(\mathbf{u}-\bar{\mathbf{w}})\right)\right)^2 \\ &= \frac{1}{s}\left((\mathbf{I}-\mathbf{V}\mathbf{V}^T)(\mathbf{u}-\bar{\mathbf{w}})\right)^2 \end{aligned} \tag{11}$$

where β is again a weighting factor that governs the strictness with which the morphological model is applied, and s is a normalization factor equal to the number of voxels in the images. In $\mathcal{P}(\mathbf{u})$, we measure the squared difference between $\mathbf{u}$ and its projection onto the PCA model. Thus, $\mathcal{P}(\mathbf{u})$ quadratically penalizes a deviation from the model. For the optimization, the derivative of the new energy term $\mathcal{P}$ is calculated as

$$\begin{aligned} \nabla_{\mathcal{U}}\mathcal{P}(\mathbf{u}) &= \frac{2}{s}(\mathbf{I}-\mathbf{V}\mathbf{V}^T)^T(\mathbf{I}-\mathbf{V}\mathbf{V}^T)(\mathbf{u}-\bar{\mathbf{w}}) \\ &= \frac{2}{s}(\mathbf{I}-2\mathbf{V}\mathbf{V}^T+\mathbf{V}\underbrace{\mathbf{V}^T\mathbf{V}}_{=\mathbf{I}}\mathbf{V}^T)(\mathbf{u}-\bar{\mathbf{w}}) \\ &= \frac{2}{s}(\mathbf{I}-\mathbf{V}\mathbf{V}^T)(\mathbf{u}-\bar{\mathbf{w}}) \end{aligned} \tag{12}$$

The identity $\mathbf{V}^T\mathbf{V}=\mathbf{I}$ is due to the orthonormality of $\mathbf{V}$. The calculation of $\nabla_{\mathbf{u}}\mathcal{P}$ is thus very closely related to calculating $\mathcal{P}$ itself, which saves a lot of computational complexity.

3.5 Translation Invariance

For the generation of the model, as well as its application, the datasets were aligned by a rigid registration. This is a necessary step for the usage of PCA

models, because they are, in general, dependent on a consistent initial positioning. Even though the used rigid registration performed very well, it did not always yield consistent results in the translational alignment. This is mostly due to the variation in the data for inter-patient registration cases. For example, when the facial bone between two datasets matches very well, the rigid registration will tend to align it, and, if the back of the skulls is morphologically similar, the rigid registration is likely to match this part best. The rotation did generally not suffer from these problems. A successful nonrigid registration might incorporate these inconsistencies into the deformation model, which then leads to problems later in the application stage.

To overcome these inconsistencies in the preregistration, we introduce a deformation model in the following that is invariant to global translations, i.e. global translations in a vector field will neither be learned nor penalized. In order to do this, the global translation is removed from the training data. Nonetheless, the global translation can still be incorporated by augmenting the PCA model with additional basis vectors for the global translation.

Let the vector field $\mathbf{u}$ be organized in the components of the coordinate system (x, y, z):

$$\mathbf{u} = (\underbrace{u_1, \ldots, u_s}_{x \text{ components}}, \underbrace{u_{s+1}, \ldots, u_{2s}}_{y \text{ components}}, \underbrace{u_{2s+1}, \ldots, u_{3s}}_{z \text{ components}})^T \tag{13}$$

The calculation of the global translation $\mathbf{t}(\mathbf{u})$ can be written as a matrix vector product:

$$\begin{aligned} \mathbf{e}_0 &= (0, \ldots, 0)^T \in \mathbb{R}^s \\ \mathbf{e}_1 &= \frac{1}{\sqrt{s}}(1, \ldots, 1)^T \in \mathbb{R}^s \end{aligned}$$

$$\mathbf{b}_x = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_0 \\ \mathbf{e}_0 \end{pmatrix} \in \mathbb{R}^{3s} \qquad \mathbf{b}_y = \begin{pmatrix} \mathbf{e}_0 \\ \mathbf{e}_1 \\ \mathbf{e}_0 \end{pmatrix} \in \mathbb{R}^{3s} \qquad \mathbf{b}_z = \begin{pmatrix} \mathbf{e}_0 \\ \mathbf{e}_0 \\ \mathbf{e}_1 \end{pmatrix} \in \mathbb{R}^{3s}$$

$$\begin{aligned} \mathbf{B} &= (\mathbf{b}_x, \mathbf{b}_y, \mathbf{b}_z) \\ \mathbf{t}(\mathbf{u}) &= \mathbf{B}\mathbf{B}^T\mathbf{u} \end{aligned} \tag{14}$$

where the vectors $\mathbf{b}_x, \mathbf{b}_y, \mathbf{b}_z$ describe a global translation along the corresponding coordinate axis. They are mutually orthogonal and normalized to $|\mathbf{b}_{[x,y,z]}| = 1$. Accordingly the model is generated with the modified samples

$$\begin{aligned} \tilde{\mathbf{w}}_i &= \mathbf{w}_i - \mathbf{t}(\mathbf{w}_i) \\ &= \mathbf{w}_i - \mathbf{B}\mathbf{B}^T\mathbf{w}_i \end{aligned} \tag{15}$$

and the mean of the modified samples

$$\bar{\tilde{\mathbf{w}}} = \frac{1}{n}\sum_{i=0}^{n} \tilde{\mathbf{w}}_i \tag{16}$$

The new sample vectors $\tilde{\mathbf{w}}_i$ are therefore orthogonal to the vectors $\mathbf{b}_{[x,y,z]}$ that compose $\mathbf{B}$.

$$\begin{aligned}\mathbf{B}^T\tilde{\mathbf{w}}_i &= \mathbf{B}^T(\mathbf{w}_i - \mathbf{B}\mathbf{B}^T\mathbf{w}_i)\\ &= \mathbf{B}^T\mathbf{w}_i - \underbrace{\mathbf{B}^T\mathbf{B}}_{=\mathbf{I}}\mathbf{B}^T\mathbf{w}_i\\ &= 0 \end{aligned} \tag{17}$$

The regularizing term is adapted in the same way, by subtracting the global translation $\mathbf{t}(\mathbf{u})$ from the current deformation field $\mathbf{u}$ before applying the PCA model.

$$\begin{aligned}\mathcal{P}(\mathbf{u}) &= \frac{1}{s}\left((\mathbf{I}-\mathbf{V}\mathbf{V}^T)(\mathbf{u}-\bar{\tilde{\mathbf{w}}}-\mathbf{t}(\mathbf{u}))\right)^2\\ &= \frac{1}{s}\left((\mathbf{I}-\mathbf{V}\mathbf{V}^T)(\mathbf{u}-\bar{\tilde{\mathbf{w}}}-\mathbf{B}\mathbf{B}^T\mathbf{u})\right)^2 &(18)\\ &= \frac{1}{s}\left((\mathbf{I}-\mathbf{V}\mathbf{V}^T)(\mathbf{I}-\mathbf{B}\mathbf{B}^T)(\mathbf{u}-\bar{\tilde{\mathbf{w}}})\right)^2 &(19)\\ &= \frac{1}{s}\left((\mathbf{I}-\mathbf{V}\mathbf{V}^T-\mathbf{B}\mathbf{B}^T)(\mathbf{u}-\bar{\tilde{\mathbf{w}}})\right)^2 &(20)\\ &= \frac{1}{s}\left((\mathbf{I}-\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T)(\mathbf{u}-\bar{\tilde{\mathbf{w}}})\right)^2 &(21)\\ \tilde{\mathbf{V}} &= (\tilde{\mathbf{v}}_1,\ldots,\tilde{\mathbf{v}}_m,\mathbf{b}_x,\mathbf{b}_y,\mathbf{b}_z)\end{aligned}$$

The step from (18) to (19) is possible because $\bar{\tilde{\mathbf{w}}}$ is a linear combination of $\tilde{\mathbf{w}}_i$, which means that $\bar{\tilde{\mathbf{w}}}$ is orthogonal to $\mathbf{B}$ (i.e. $\mathbf{B}\mathbf{B}^T\bar{\tilde{\mathbf{w}}}=0$) due to (17). The same argument can be applied to the step from (19) to (20): the components $\mathbf{v}_i$ of $\mathbf{V}$ are linear combinations of the training vectors $\tilde{\mathbf{w}}_i$ and, accordingly, orthogonal to the vectors $\mathbf{b}_{[x,yz]}$ of $\mathbf{B}$. Therefore, the product $\mathbf{V}\mathbf{V}^T\mathbf{B}\mathbf{B}^T=0$. Essentially, the whole process of eliminating the global translation in the calculation of $\mathcal{P}$ yields a new basis $\tilde{\mathbf{V}}$ by augmenting the principal components $\mathbf{V}$ artificially with the vectors $\mathbf{b}_{[x,y,z]}$. Note that $\tilde{\mathbf{V}}$ is still orthonormal, which allows the calculation of the derivative of $\mathcal{P}$ exactly as in (12).

$$\nabla_{\mathbf{u}}\mathcal{P}(\mathbf{u}) = \frac{2}{s}(\mathbf{I}-\tilde{\mathbf{V}}\tilde{\mathbf{V}}^T)(\mathbf{u}-\bar{\tilde{\mathbf{w}}}) \tag{22}$$

4 Results

For an evaluation of the proposed approach, 18 CT datasets were used for the training of the model consisting of the mean deformation and the first 10 components of the PCA. The algorithm was then applied to 23 T1- and 15 T2-weighted MRI scans from the same patients in a leave-one-out cross validation. Prior to the experiments, the tables present within the CT images have been segmented and ignored during the segmentation. The CT and MRI of the same patient were rigidly registered in order to provide a ground-truth CT for every patient's

MRI dataset. This rigidly registered, ground-truth CT was used, in combination with the deformed atlas CT, to calculate objective quality measurements for the nonrigid atlas registration results. All patient data was resampled to a common volume size of $129 \times 129 \times 104$ voxels and an isotropic spacing of 1.95 mm in order simplify model generation and evaluation.

Regarding the choice of the α parameter, please note that our algorithm operates in physical space and not on a unit square, or with a unit spacing. Intensity values were also taken into account without any rescaling. The weights for the regularizing terms can, therefore, differ from the values presented in the related work. For the generation of the training deformations, the mono-modal, nonrigid registration that uses the sum-of-squared-differences measure was regularized with a weighting factor of $\alpha = 0.01$. The multi-model registration was driven by the MI distance measure and carried out with and without the PCA regularization. With the PCA regularization enabled, the weighting factors were chosen as $\alpha = 7$ and $\beta = 0.01$. This choice represents a rather low value for α and would result in very large local deformations without the additional morphological regularization. An example of this setting is shown in Fig. 1. Here

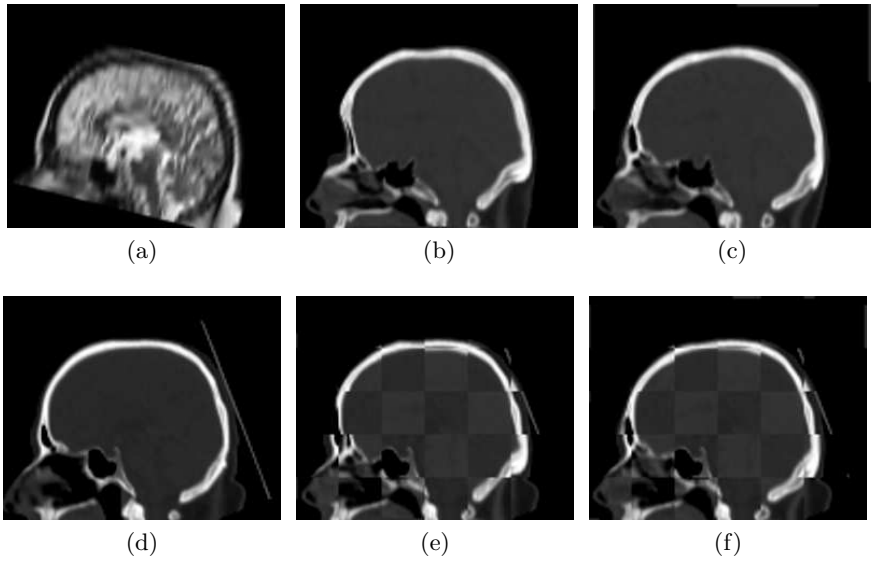


Fig. 1: The images show results for a single example of (a) an MR and its corresponding (d) CT image. The registration result with $\alpha = 7$ and no PCA model regularization is shown as (b) the deformed atlas CT and (e) the checkerboard fusion with the ground truth CT. Corresponding results with the usage of the prior knowledge can be seen in (c) and (f).

$\alpha = 7$ was used with and without PCA regularization. The effect is clearly

visible, especially in the facial region, when comparing the registration results with the ground truth CT. Accordingly, for the nonrigid registration without PCA regularization, a much higher value for α is used in order to regain the necessary stability. Empirically, we determined a value of $\alpha = 60$ as a good choice. Tables 1 and 2 present the results for the comparison between the two

Method	**Measure**				
	MSE	DistBO	SE(ST)	SE(BO)	SE(ST,BO)
No PCA	25440.8 ± 7715.2	0.41 ± 0.25	84.8 ± 2.8	64.0 ± 11.4	81.2 ± 3.2
PCA	21297.9 ± 10466.7	0.25 ± 0.10	84.3 ± 3.6	70.8 ± 6.0	82.0 ± 3.2

Table 1: Results for the leave-one-out cross-validation on T1-weighted MRI data. The values shown consist of the mean and the standard deviation for the corresponding measure calculated over all datasets. For mean squared error in Hounsfield units (MSE) and the distance to the bone mask in mm (DistBO), smaller values indicate better results. The sensitivity measures for segmented soft tissue (SE(ST)), bone (SE(BO)), and the combination of both (SE(ST, BO)), larger values are better.

Method	**Measure**				
	MSE	DistBO	SE(ST)	SE(BO)	SE(ST,BO)
No PCA	29914.7 ± 17514.4	0.55 ± 0.46	83.3 ± 5.3	58.8 ± 13.3	79.0 ± 6.5
PCA	23966.4 ± 10371.6	0.31 ± 0.16	84.3 ± 5.0	66.4 ± 7.9	81.1 ± 5.3

Table 2: Results for the leave-one-out cross-validation on T2-weighted MRI data. For a description of the values, see table 1.

approaches for T1- and T2-weighted MRI scans. The results of the registration are compared with the ground truth CT of the patient, who was left out for the cross-validation, based on a number of measures: The mean square error (MSE) was calculated between the intensities of both images. Since especially bone densities between different patients are usually not directly comparable, this is only a coarse measure. To compensate for these differences, the CT images are segmented into three classes for further comparisons: air, soft tissue (ST), and bone (BO). Based on these segmentations the class-specific sensitivities for soft tissue (SE(ST)), bone (SE(BO)), and the joint sensitivity (SE(ST,BO)) are calculated. A final measure (DistBO) is computed in order to provide a quantitative measure of the spatial distance between the segmentations of the bone. This measure determines the average euclidean distance between each pixel segmented as bone in one image to the nearest pixel segmented as bone in the other image. On average, the MSE measure for T1 and T2 data is decreased by

18% and the DistBO measure by 42% using the proposed algorithm. The overall sensitivity for correct soft tissue and bone alignment is increased by 4%.

5 Discussion

In this article, we have introduced a novel, PCA-based regularizing energy. This morphological term constrains the deformation to be close within the known space of variability that is learned from a training set of deformations. We have shown that this model is invariant to global translations and is able to compensate for morphologically unreasonable deformations when other regularization energies are reduced. The presented results indicate a better performance of the proposed approach with respect to especially the MSE and the distance between the bone segmentations. These two criteria are of special interest for the attenuation correction, because the location of the bones has a high impact on the corrected result. Except for the mean values of the soft tissue sensitivity, the registration with the incorporation of the prior knowledge performed better than the standard algorithm.

6 Acknowledgements

The authors would like to P. Ritt (Department of Nuclear Medicine, University Hospital, Erlangen) for his efforts in the patient data selection. We are also thankful to HipGraphics for providing the volume rendering software InSpace.

References

1. Hofmann, M., Pichler, B., Schölkopf, B., Beyer, T.: Towards quantitative pet/mri: a review of mr-based attenuation correction techniques. European Journal of Nuclear Medicine and Molecular Imaging, European Journal of Nuclear Medicine and Molecular Imaging **36**(Supplement 1) (03 2009) 93–104
2. Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., Aschoff, P., Brady, M., Schölkopf, B., Pichler, B.J.: Mri-based attenuation correction for pet/mri: A novel approach combining pattern recognition and atlas registration. Journal of Nuclear Medicine **49**(11) (10 2008) 1875–1883
3. Wang, Y., Staib, L.H.: Physical model-based non-rigid registration incorporating statistical shape information. Medical Image Analysis **1** (2000) 35–51
4. Kim, M.J., Kim, M.H., Shen, D.: Learning-based deformation estimation for fast non-rigid registration. Computer Vision and Pattern Recognition Workshop **0** (2008) 1–6
5. Xue, Z., Shen, D.: A new statistically-constrained deformable registration framework for mr brain images. International Journal of Medical Engineering and Informatics **1**(3) (2009) 357 – 367
6. Hermosillo, G.: Variational Methods for Multimodal Image Matching. PhD thesis, University of Nice Sophia Antipolis, France (2002)
7. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, Oxford (2004)

8. Viola, P.: Alignment by Maximization of Mutual Information. PhD thesis, Massachusetts Institute of Technology (MIT), Cambridge, USA (1995)
9. Fischer, B., Modersitzki, J.: A unified approach to fast image registration and a new curvature based registration technique. Linear Algebra and its Applications **380** (March 2004) 107–124
10. Tikhonov, A.N., Arsenin, V.Y.: Solutions of ill posed problems. Wiley, New York (1977)
11. Shannon, C.E.: A mathematical theory of communication (parts 1 and 2). The Bell System Technical Journal **27** (1948) 379–423,623–656
12. Wells III, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis **1**(1) (1996) 35–51
13. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging **16**(2) (April 1997) 187–198
14. Fischer, B., Modersitzki, J.: Curvature based image registration. Journal of Mathematical Imaging and Vision **18**(1) (2003) 81–85
15. Murase, W., Lindenbaum, M.: Partial eigenvalue decomposition of large images using spatial temporal adaptive method. IEEE Transactions on Image Processing **4**(5) (1995) 620–629

Consistent Atlas Estimation on BME Template Model: Applications to 3D Biomedical Images.

Stéphanie Allassonnière, Estelle Kuhn, J. Tilak Ratnanather and Alain Trouvé

Abstract. This paper aims at validating a methodology proposed in [1, 2] for estimating a Bayesian Mixed Effect (BME) atlas, i.e. coupled templates and geometrical metrics for estimated clusters, in a statistically consistent way for a sample of images. We recall the generative statistical model applied to the observations which enables the simultaneous estimation of the clusters, the templates and geometrical variabilities (related to the metric) in the population. Following [1–3], we work in a Bayesian framework, use a Maximum A Posteriori estimator and approach its value using a stochastic variant of the Expectation Maximisation (EM) algorithm. The method is validated with a data set consisting of 3D biomedical images of dendrite spines from a mouse model of Parkinson's disease. We show the performance of the method on the estimation of the template, the geometrical variability and the clustering.

1 Introduction

In the field of Computational Anatomy, one aims at segmenting images, detecting pathologies and analysing the normal versus abnormal variability of segmented organs. The most widely used techniques are based on the comparisons between subjects and a prototype image (usually called template in the literature). Such a prototype is an image whose biological properties are known and which - in a sense to be defined - characterises the population being studied. This template contains common features of the population which would not be revealed by multiple inter-subject comparisons.

Regarding the large variability of anatomical structures, a template only may be not able to summarise the diversity of a whole population. For example, two populations can have the same template but can be distributed quite differently around (very like points clouds in a manifold can be concentrated or spread in many different way around their means). Therefore, in addition to the template, a parametrisation of the shape variability around a given template is of importance in producing relevant statistical summary of a population. These two parameters will together be considered as an atlas in the following.

One way to estimate an atlas in a population is to use statistical learning approaches on statistical models. Statistical learning on such models consists of tuning its parameters to maximise the penalised data likelihood of the observed population. Among all the statistical models, generative statistical models make assumptions on how the observed images are derived from the atlas. These models do not only explain data but are also able to randomly generate new data.

When simulating a large number of likely images (according to the model), one can better interpret and even exhibit unexpected behaviours that would not be easily detectable by a visual inspection of a small population (typical case in medical image analysis).

One further step is to consider that the population is composed of several subgroups. The population is therefore summarised by the weight of each cluster and an atlas for each of them. Since the clustering may not be known, the corresponding model enables an estimation of both the distribution of the subgroups in the population and the cluster atlases at the same time.

Our special interest is the **construction** of a **statistically consistent atlas**, called Bayesian Mixed Effect (BME) atlas, as the estimation of the templates and their global geometric variabilities in estimated cluster for a given population in a statistically consistent way.

The usual way to measure the geometrical heterogeneity is to map the template to all the observations (or the other way around) and do some statistics on these deformations (typically PCA). Many registration methods have been developed for this purpose, for example in [4–6]. Based on this, several different approaches have been proposed recently to estimate templates. Some are based on a minimisation of a penalised energy function describing the cost to match the template to the observations [7–9]. Another view, closer to ours, is to propose a statistical model whose parameters are the template and the **mappings** between this template and the observations [10] or [11, 12] and the optimisation is done via maximum likelihood. Even if these methods lead to interesting results and effective computation schemes, they suffer from different limitations. First, in most cases, the deformation is applied to the observations instead of the template. However, these images are only noisy observations known on a discrete fixed grid of voxels. Applying the deformation to these discretly supported images requires interpolating between voxels and therefore creates errors which are difficult to control. The template is computed as an arithmetic mean of the deformed noisy observations which leads to a noisy version of this template. Moreover, the modelling implies inexact matching. One way to model this is to consider that the difference between the deformed image and the template is an independent additive noise. This noise accounts for both the acquisition noise and the fact that the model does not describe the reality (but is only an approximation of the true distribution). Assuming the deformation is invertible, applying the mapping to the observations is equivalent to apply its inverse to both the template and the noise. There is no suitable interpretation of this fact; there is no reason for the noise to be affected by the mapping which is only a mathematical tool we introduce. The last but not least drawback is that the deformations are considered as nuisance parameters which have to be optimised. Knowledge of these elements only gives information subject by subject and nothing about the global nature of the population. Moreover, the convergence of such procedure has not been proved and has even been shown to fail for a phantom example [1].

For these reasons, we consider the model proposed in [1]. Indeed, the authors consider the usual modelling called the Deformable Template model. This assumes that each observation is a random deformation of the template which is then corrupted by an additive Gaussian noise term. This avoids the interpolation problem as well as the lack of meaning of the deformed noise mentioned below. The deformations are *unobserved* random variables whose probabilistic distribution has to be estimated. This generative statistical model defines a global information of the geometrical variability inside the population. This distribution also characterises the metric on the deformation space. Thanks to this model, the estimation of the template is correlated to this estimated metric and vice versa.

To take into account the heterogeneity of the whole population, we use the extended model based on a mixture of the previous models (cf. [1, 3]): each observation belongs to one component of the mixture governed by its parameters (template, noise and metric). The observation memberships are specified through *hidden random* labels whose weights are estimated as well.

We summarise here this efficient methodology, called Bayesian Mixed Effect (BME) template [13], to construct a BME atlas, i.e.clusters distribution, templates and geometrical metrics, via a consistent estimation, given a sample of images. We focus on its validation in the context of 3D biomedical images of dendrite spines which have a large geometrical variability (various shapes) in order to show its performance in terms of estimation and generation of new plausible shapes.

In this paper, the model and the estimator are detailed in Section 2. We then present the algorithm in Section 3. Section 4 is devoted to the experiments. We end this paper with some conclusions and a discussion in Section 5.

2 BME Template Model and MAP Estimation

We consider a population of n gray level images which we aim to automaticaly cluster in a small number of groups called components later. We assume that each observation y belongs to an *unknown* component t. We work within the small deformation framework [10] so that conditional on the image membership to component t, there exists an *unobserved* deformation field $z : \mathbb{R}^3 \to \mathbb{R}^3$ of a continuously defined template $I_t : \mathbb{R}^3 \to \mathbb{R}$ and a Gaussian centred white noise[1] ϵ of variance σ_t^2 such that

$$y(s) = I_t(x_s - z(x_s)) + \epsilon(s) = zI_t(s) + \epsilon(s)\,, \tag{1}$$

where Λ is a discrete grid of pixels and the pixels location is denoted by $(x_s)_{s\in\Lambda}$. Given $(p_k)_{1\leq k\leq k_p}$ a fixed set of uniformly distributed landmarks covering the

[1] This model is relevant for grey level images. One could slightly modify it in order to better interpret binary images. Instead of a Gaussian noise (usually used for image matching with a L^2 penalty term), one can use a Bernoulli distribution whose parameter would be a continuous map $r_t(x)$, analogous to our template $I_t(x)$. However, this model does not belong to the exponential family which make the coding more complicated. The convergence of the algorithm has not been proved in this case either.

image domain, the template functions I_t are parameterised by coefficients $\alpha_t \in \mathbb{R}^{k_p}$ through: $I_t(x) = \mathbf{K_p}\alpha_t(x) \triangleq \sum_{k=1}^{k_p} K_p(x, p_k)\alpha_t(k)$, where K_p is the kernel of the Reproducing Kernel Hilbert Space (RKHS) in which we search the template. The kernel controls the smoothness of the interpolation between landmarks. It is also nicely described as the covariance operator of a Gaussian random field globally defined on the image domain and defining a natural prior for the template. The restriction of these Gaussian fields on the p_k's is an easily tractable finite dimensional zero mean Gaussian vector with explicit covariance matrix. This has the advantage of giving a prior that is essentially independent of the number of landmarks k_p, and that only depends on the global choice made for the RKHS. In this context, the number of landmarks used determines a trade-off between accuracy of the approximations of functions in the respective spaces and the amount of required computation.

The same kind of decomposition with a second set of landmarks $(g_k)_{1\leq k\leq k_g}$ and kernel K_g is used to parametrise the deformation field z by the *unobserved random* vector β such that $z = \mathbf{K_g}\beta$. This random vector is assumed to follow a Gaussian distribution with zero mean and covariance matrix Γ_g^t depending on the component t (which could be the natural prior associated with K_g as a first guess but will be learnt from the data during the estimation process).

The model parameters of each component $t \in \{1, \dots, m\}$ are denoted by $\theta_t = (\alpha_t, \sigma_t^2, \Gamma_g^t)$. We assume that θ belongs to the open parameter space $\Theta \triangleq \{\ \theta = (\alpha_t, \sigma_t^2, \Gamma_g^t)_{1\leq t\leq m} |\ \forall t \in \{1, \dots, m\}\ ,\ \alpha_t \in \mathbb{R}^{k_p},\ \sigma_t^2 > 0,\ \Gamma_g^t \in \Sigma^+_{2k_g,*}(\mathbb{R})\ \}$ and $\rho = (\rho_t)_{1\leq t\leq m}$ to the open simplex ϱ. Here $\Sigma^+_{2k_g,*}(\mathbb{R})$ is the set of strictly positive symmetric matrices. Let $\eta = (\theta, \rho)$, the hierarchical Bayesian structure of our model is :

$$\begin{cases} \rho \sim \nu_\rho, \quad \theta = (\alpha_t, \sigma_t^2, \Gamma_g^t)_{1\leq t\leq m} \sim \otimes_{t=1}^m (\nu_p \otimes \nu_g) \\ \tau_1^n \sim \otimes_{i=1}^n \sum_{t=1}^m \rho_t \delta_t \mid \rho\,, \\ \beta_1^n \sim \otimes_{i=1}^n \mathcal{N}(0, \Gamma_g^{\tau_i}) |\ \tau_1^n,\ \eta \\ \\ y_1^n \sim \otimes_{i=1}^n \mathcal{N}(z_{\beta_i} I_{\alpha_i}, \sigma^2_{\tau_i} Id_\Lambda) \mid \beta_1^n,\ \tau_1^n,\ \eta \end{cases} \tag{2}$$

$$\text{with} \quad \begin{cases} \nu_\rho(\rho) \propto \left(\prod_{t=1}^m \rho_t\right)^{a_\rho}, \quad \nu_g(d\Gamma_g) \propto \left(\exp(-\langle \Gamma_g^{-1}, \Sigma_g\rangle/2) \frac{1}{\sqrt{|\Sigma_g|}}\right)^{a_g} d\Gamma_g, \\ \nu_p(d\sigma^2, d\alpha) \propto \left(\exp\left(-\frac{\sigma_0^2}{2\sigma^2}\right) \frac{1}{\sqrt{\sigma^2}}\right)^{a_p} \cdot \exp\left(-\frac{1}{2}\alpha^t (\Sigma_p)^{-1} \alpha\right) d\sigma^2 d\alpha, \end{cases}$$

where the hyper-parameters are fixed (their effects has been discussed in [1]). All priors are the natural conjugate priors and assumed independent.

The Gaussian distribution set on the observations whose mean is the deformed template is the usual Deformable Template model used in image analysis and in particular image matching. This model is quite natural saying that the observation is, up to an independent noise, close to the deformed template. The Gaussian distribution used to model the deformation vector β is assumed to have zero mean. This assumption corresponds to the intuitive fact that once

we are moving around the template -the "mean shape" of the population-, the mean of all these movements should be close to zero. Therefore, we only estimate its covariance matrix. The last probabilistic distribution for τ is a common distribution on random variables on finite space, namely a finite sum of weighted Dirac measures.

The system of equations (2) can be interpreted top to bottom, which corresponds to the generation of some images. The generation process consists in first drawing the parameters from their prior distributions. Given these parameters, pick a membership according to the weighted distribution. This label points towards a component. For this particular component, draw a deformation with respect to this Gaussian law and apply it to the pointed template. Adding a random Gaussian noise whose variance is given by the membership to each voxel independently gives you a new image. The estimation process takes the images as observed elements and attempts to recover the parameters (giving that they follow some constrains given by the priors). This scheme can be summarised in Figure 1.

For each component t :
- ρ_t : probability of the component
- α_t : template parameter
- Γ_g^t : geometrical covariance matrix
- σ_t^2 : additive noise variance

For each observation y_i :
- τ_i : component label
- β_i : deformation parameters
- ϵ_i : additive noise

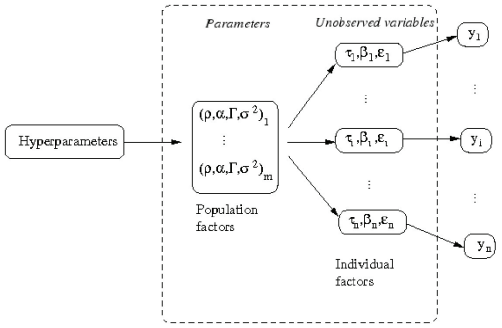


Fig. 1. Latent structure of BME-Template model.

In this context, in order to estimate the model parameters, we use a Maximum A Posteriori estimator, i.e. a value of the parameters which maximises the posterior density on η conditional on y_1^n:

$$\hat{\eta}_n = \underset{\eta}{\operatorname{argmax}}\, q(\eta|y_1^n). \tag{3}$$

It has been proved in [1] that this estimator is consistent.

3 Convergent Algorithm for the Estimation

To solve this maximisation in a non linear context with missing variables in $\mathbb{R}^D$ where D is large (typically $D \leq 3000$), we use a Stochastic Approximation EM algorithm (SAEM) [14] coupled with an MCMC procedure [15]. Our model belongs to the exponential density family which means that the complete likelihood can be put in the following form: $q(y, \beta, \tau, \eta) = \exp\left[-\psi(\eta) + \langle S(\beta, \tau), \phi(\eta)\rangle\right]$, where the sufficient statistic S is a Borel function on $\mathbb{R}^{3k_g} \times \{1, \ldots, m\}$ taking

its values in an open subset $\mathcal{S}$ of $\mathbb{R}^m$ and ψ, ϕ two Borel functions on $\Theta \times \varrho$ (the dependence on y is omitted for sake of simplicity). We introduce the following function: $L : \mathcal{S} \times \Theta \times \varrho \to \mathbb{R}$ as $L(s;\eta) = -\psi(\eta) + \langle s, \phi(\eta) \rangle$. Then, iteration l of this algorithm consists of the following three steps.

Simulation step: Draw the missing data with respect to a transition probability Π_{η_l} of a convergent Markov chain having the posterior distribution as stationary distribution:

$$(\beta_{l+1}, \tau_{l+1}) \sim \Pi_{\eta_l}((\beta_l, \tau_l), \cdot)\,. \tag{4}$$

Stochastic approximation step: Do the stochastic approximation on the sufficient statistics:

$$s_{l+1} = s_l + \Delta_{l+1}(S(\beta_{l+1}, \tau_{l+1}) - s_l)\,, \tag{5}$$

where $(\Delta_l)_l$ is a decreasing sequence of positive step-sizes and using the simulated values $(\beta_{l+1}, \tau_{l+1})$.

Maximization step: Updated: $\eta_{l+1} = \underset{\eta}{\operatorname{argmax}}\, L(s_{l+1}, \eta)$.

We refer to [3] for more details about the algorithm in particular for the choice of Π_η used in the simulation step. The MCMC procedure mainly consists in a hybrid Gibbs sampler for which we use auxiliary Markov chains in the Metropolis-Hastings step. It has been proved in [3], that, under mild assumptions, the sequence $(\eta_l)_l$ generated through this algorithm converges a.s. toward a critical point of the penalised likelihood of the observations.

The theoretical convergence properties of the estimator and the algorithm strengthen the potential of this method. We will now show the numerical results on 3D biomedical images to highlight its practical performance.

4 Experiments

We run the algorithm on a set of murine dendrite spines [16–18]. The data set consists of 50 binary images of microscopic structures, tiny protuberances found on many types of neurons termed dendrite spines. The images are from control mice and knockout mice which have been genetically modified to mimic human neurological pathologies like Parkinson's disease. The acquisition process consisted of electron microscopy after injection of Lucifer yellow and subsequent photo-oxidation. The shapes were then manually segmented on the tomographic reconstruction of the neurons. The images are labelled by experts as belonging to six different categories (called types): double, filopodia, long mushroom, mushroom, stubby and thin. Some of these images are presented in Figure 2. This figure shows a 3D view of some examples among the training set. Each image is a binary (background = 0, object = 2) cubic volume of size 56. We can notice here the large geometrical variability of this population of images.

The study in [16] showed a correlation between the spine type and its shape. This study is based on a template shape and a given metric to compare the spines through the computation of deformations. The estimation here aims at

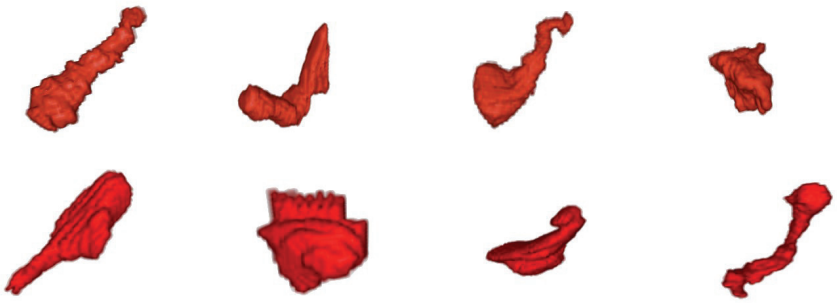


Fig. 2. 3D view of eight samples of the data set of dendrite spines. Each image is a volume leading to a binary image.

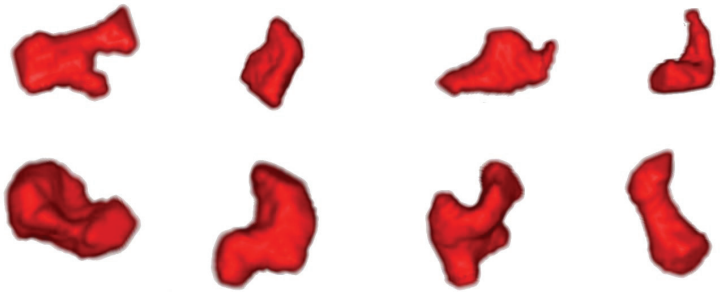


Fig. 3. 3D view of eight synthetic data. The estimated template shown in Figure 4 is randomly deformation with respect to the estimated covariance matrix. The results are then thresholded in order to get a binary volume.

proposing one or more templates with their correlated metric in order to exhibit the common features of the population.

The computation of the Stochastic Approximation EM algorithm coupled with the MCMC procedure is performed in Matlab. Experiments were performed on 64bit system with 16GB of shared memory. Each run takes about a day with the whole data set. The main difficulty concerns the resolution of the linear system in α involved in the maximisation step at each iteration l of the algorithm. The matrix involved in this linear system is very ill-conditioned. The effects are edge effects on the template, i.e. some non-zero values of the voxel grey level on the sides of the template image. Therefore, incomplete LU factorisation as a pre-conditioner is performed to stabilise the numerical inversion. If this is insufficient (in extreme cases), one solution would be to use full or partial pivoting strategies as in Gaussian elimination. This leads to slightly longer algorithm but without numerical issues.

One step further in the optimisation of the processing time is to parallelise the loop on the observations. Indeed, given the current parameters, each observation

is independent from the others. The simulation step can therefore be run on separate processors. This divides the time of processing by the number of images.

4.1 One Component Model

In this section we present the result of the estimation using the single component model. Since the training set shows very different shapes for the six categories, a single template model might not be able to capture this large variability. In order to have a little bit smaller variability, we focused on 30 images of only three spine types to estimate our atlas with a single component model. We choose thin, long mushroom and stubby.

The estimated template is presented in the left column of Figure 4. The estimated image is real valued, in particular here in the segment $[0, 2]$. We do not specify any criteria in order to impose a binary template. This is why the estimated volumes look blurred. For 3D visualisation, one can threshold the estimated image and binarise the values (most of the values are very close to the extrema and it only creates really sharp boundaries). The resulting shape is presented in the right column of Figure 4. As expected, the shape of this estimated spine is a relevant representation of the data set. It is smoother than the observations (as expected for an "average") but it could be one of them.

One crucial improvement coming from our method is that we also get an estimation of the geometrical variability through the covariance matrix Γ_g. In order to visualise the accuracy of this coupled estimation and thanks to the generative model, we simulate new synthetic data using the estimated values of the parameters. Figure 3 shows eight images obtained by applying random deformations (sampled from $\mathcal{N}(0, \Gamma_g)$) to the estimated template. The resulting

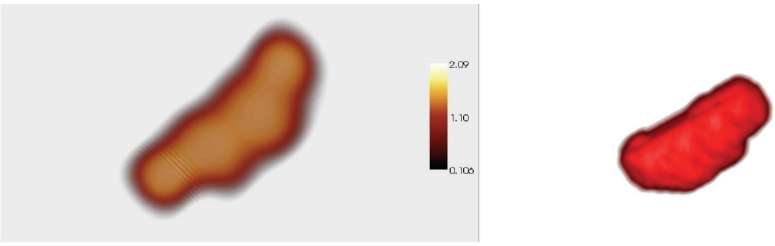


Fig. 4. Estimated template with the one component model: Left: 3D representation of the grey level volume. Right: 3D representation of the thresholded volume.

shapes look like potential dendrite spines. Indeed, we can see some similarities between these synthetic images and some images of the data set as presented in Figure 2. For example, the estimated geometrical variability has taken into account shrinking the template to get a long and thin appearance. It has also learnt to inflate one extremity and contract the other to get what is labelled as long mushroom and to make the shape more or less curved. Considering the

huge dimensionality of the deformation space, this estimation is pretty good. In this model, the deformation is not constrained to be a diffeomorphism. This can affect the estimation in a way that the estimated geometrical variability could create holes or overlaps in the template. In these experiments, this problem did not occur. One way however to correct this would either be to tune the hyper-parameters which controls the deformation regularity or to use diffeomorphisms.

The last parameter which is estimated is the variance of the additive Gaussian noise. This parameter is quite interesting since it helps to see how close the model managed to fit the data. In our experiments, the estimated standard deviation of the noise in the one component case is 0.1387. Since the data set is very heterogeneous, it is very low. Indeed, as a comparison, one can look at the 2D experiments on hand written digits in [2]. The standard deviations of the digits were between 0.1 and 0.3. This suggests that the estimation in this 3D case of dendrite spine is relevant.

4.2 Two Component Model

Fig. 5. Estimated templates of the two components with the 30 image training set: 3D representation after thresholding.

The large geometrical variability of the spine shapes leads to consider several different sub-populations in the data set. However, since the data set is of small size (at most 50 images), the estimated parameters would not be accurate in a mixture model involving more than two sub-groups called components. Indeed, we have to estimate one template and one covariance matrix for each component. This leads to parameters of large dimension. The small number of images in each component would not give enough information to perform the computation of the corresponding atlas accurately. For this reason, we restrict the estimation to two components.

We ran the algorithm on the previous data set of 30 images of the three types used for the single component estimation. We also use the whole data set of 50 images from the six dendrite spines. The estimated templates are shown in Figure 5 for the three categories and in Figure 6 for the whole training set. We only show the thresholded shapes for illustrating the differences between the two component templates.

The two estimated components show very different shapes. Indeed, we can see that the second template has a curved shape with a thin extremity and a larger one on the other side. The other template size is more isotropic. The

curvature of the two shapes is also distinct. These two shapes are quite relevant representatives of the spine population. The first component looks to contain the stubby group which corresponds to plumper shapes. Whereas the second component gathers the thin and long mushroom groups. The estimated weights of the components in the population are respectively 0.32 and 0.68 which actually match the number of such shapes in the data set.

To see the impact of the different spine types on the estimation, we ran the same algorithm with the whole data base of the 50 images with the six different types. This training set has a larger geometrical variability than the previous one since we increase the number of spine types considered. But the estimation may be sharper since more images are available for each component parameters to be estimated. While clustering the data, the algorithm only uses a percentage of the data set for a given component and therefore estimates its high dimensional parameters (template and covariance matrix) using only this sub-sample. This yields a small number of images per cluster and may produce a relatively blurred image since the geometrical variability has not been well-estimated. When more images are available, the number of images per component increases and even if the variability increases, the estimation is supposed to better capture it.

The two sub-groups are expected to be quite different from the previous ones and so their respective templates. These templates are shown in Figure 6. The estimated shapes are again good representations of the whole population. The subdivision is made between more isotropic shapes (similar to the previous stubby type) and longer ones, curved and with irregular boundaries. This summarises the differences which appear in the training set.

Fig. 6. Estimated templates of the two components with the 30 image training set: 3D representation after thresholding.

Concerning the experiment with a data set of 30 images, the estimated standard deviations are 0.1780 and 0.1659 respectively. One would expect a lower value compared to the one component. However, the small number of images leads to less precise parameters and therefore a slightly higher value of the standard deviation. For the last experiment with the whole data set, the values are 0.1521 and 0.1800. These values are quite good again compared to the 2D example of hand written digits. The slightly larger values (compared to the single component) may come from the fact that even if the training set is bigger, the variability increases as well.

It would be interesting to run the algorithm with a larger data base of only these six types and six possible components. It would also be interesting to either repeat the kind of study presented in [16] or to use the model as a classifier. Concerning the second application, we trust this model in particular looking at the classification results obtained in [1] on some hand written digits. The huge geometrical variability is even higher in this "phantom" example since there are some change of topology. We think that this methodology would give interesting results if we had the chance to analyse new data of this type.

5 Discussion and Conclusion

We considered a generative statistical model and a stochastic algorithm to estimate mixtures of deformable templates to construct a BME Atlas. The theoretical statistical properties of the estimator and of the algorithm were established. We validated them by numerical results. Indeed, we ran this estimation on highly variable 3D shapes of murine dendrite spines. The results in the one component model using a sub group of the data involving only three different types of dendrites are relevant on both the estimation of the template image and of the geometrical variability around its template. Using the two component model with the same data set of three different types of spines, we capture more precisely the variability. This leads to two different templates representing characteristic shapes of the data set. We also ran the two component model with the whole training set involving six types of spines. The estimated templates are also quite relevant and face the large heterogeneity of the training sample. This method can be used to estimate different population atlases such as healthy controls and Parkinson's disease populations and then compute likelihood ratios in order to classify new un-labelled images. Another possibility is to compute atlases at different stages of the disease in order to characterise its evolution. These applications may increase the knowledge and understanding of diseases.

Acknowledgements

Murine dendrite spines were initially obtained from Dr. M. Martone of NCMIR at UCSD. They were processed for image analysis at CIS under the support of NSF DMS-0101329.

References

1. Allassonnière, S., Amit, Y., Trouvé, A.: Toward a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society **69** (2007) 3–29
2. Allassonnière, S., Kuhn, E., Trouvé, A.: Bayesian deformable models bulding via stochastic approximation algorithm: A convergence study. in revision
3. Allassonnière, S., Kuhn, E.: Stochastic algorithm for bayesian mixture effect template estimation. to appear in ESAIM Probab.Stat.

4. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. Neuroimage **45** (2009) 61–72
5. Miller, M.I., Trouvé, A., Younes, L.: On the metrics and Euler-Lagrange equations of computational anatomy. Annual Review of biomedical Engineering **4** (2002) 375–405
6. Ashburner, J.: A fast diffeomorphic image registration algorithm. NeuroImage **38** (2007) 95–113
7. Glaunès, J., Joshi, S.: Template estimation form unlabeled point set data and surfaces for computational anatomy. In Pennec, X., Joshi, S., eds.: Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA). (2006) 29–39
8. Twining, C., Cootes, T., Marsland, S., Petrovic, V., Schestowitz, R., Taylor, C.: Information-theoretic unification of groupwise non-rigid registration and model building. In: Proceedings of Medical Image Understanding and Analysis (MIUA). Volume 2. (2006) 226–230
9. Cootes, T., Edwards, G., Taylor, C.: Active appearance model. IEEE Transactions on Pattern Analysis and Machine Intelligence **23(6)** (2001) 681–685
10. Amit, Y., Grenander, U., Piccioni, M.: Structural image restoration through deformable templates. Journal of the American Statistical Association **86** (1989) 376–387
11. Glasbey, C.A., Mardia, K.V.: A penalised likelihood approach to image warping. Journal of the Royal Statistical Society, Series B **63** (2001) 465–492
12. Sabuncu, M., Balci, S.K., Golland, P.: Discovering modes of an image population through mixture modeling. Proceeding of the MICCAI conference **LNCS**(5242) (2008) 381–389
13. Allassonnière, S., Kuhn, E., Trouvé, A.: Map estimation of statistical deformable templates via nonlinear mixed effects models : Deterministic and stochastic approaches. In Pennec, X., Joshi, S., eds.: Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy (MFCA). (2008)
14. Delyon, B., Lavielle, M., Moulines, E.: Convergence of a stochastic approximation version of the EM algorithm. Ann. Statist. **27**(1) (1999) 94–128
15. Kuhn, E., Lavielle, M.: Coupling a stochastic approximation version of EM with an MCMC procedure. ESAIM Probab. Stat. **8** (2004) 115–131 (electronic)
16. Aldridge, G., Ratnanather, J., Martone, M., Terada, M., Beg, M., Fong, L., Ceyhan, E., Kolasny, A., Brown, T., Cochran, E., Tang, S., Pisano, D., Vaillant, M., Hurdal, M., Churchill, J., Greenough, W., Miller, M., Ellisman, M.: Semi-automated shape analysis of dendrite spines from animal models of fragilex and parkinson's disease using large deformation diffeomorphic metric mapping. Society for Neuroscience Annual Meeting, Washington DC (2005)
17. Ceyhan, E., Fong, L., Tasky, T., Hurdal, M., Beg, M.F.and Martone, M., Ratnanather, J.: Type-specific analysis of morphometry of dendrite spines of mice. 5th Int. Symp. Image Signal Proc. Analysis, ISPA (2007) 7–12
18. Ceyhan, E., Ölken, R., Fong, L., Tasky, T., Hurdal, M., Beg, M., Martone, M., Ratnanather, J.: Modeling metric distances of dendrite spines of mice based on morphometric measures. Int. Symp on Health Informatics and Bioinformatics (2007)

Integrating Statistical Models of Bone Density into Shape Based 2D-3D Registration Framework

Gouthami Chintalapani[1], Ofri Sadowsky[1], Lotta M. Ellingsen[2], Jerry L. Prince[2], and Russell H. Taylor[1]

[1] Department of Computer Science
[2] Department of Electrical and Computer Engineering
The Johns Hopkins University, Baltimore, MD, USA
{gouthami,rht}@jhu.edu

Abstract. We present a framework to deformably register simulated X-ray images to a combined statistical model of pelvis anatomical structure, created from a population of CT scans. The primary contributions are: 1) a framework to create and analyze bone density variations, separate from shape variations and 2) an augmented 2D/3D registration framework that couples shape and density priors to create accurate patient specific models. Our statistical model representation consists of a tetrahedral mesh for approximating bone shape and Bernstein polynomials defined within each tetrahedron for bone density. All datasets in the given population are registered deformably to a template CT dataset. The shape and density statistics are extracted using principal component analysis on the corresponding mesh vertices and voxels of the shape-free deformed subjects respectively. In the registration framework, we register the 2D input images to the 3D shape prior and estimate the bone density parameters in a least-squares like setup by projecting the density modes on to the input image space. This approach was tested using leave-n-out experiments, with n = 8, datasets using an atlas of 63 full pelvis CT datasets.

1 Introduction

Statistical modeling and analysis of anatomical shapes is a promising research area in medical imaging with a variety of applications such as segmentation, 2D/3D registration and reconstruction of anatomical structures, pre-operative planning, analyzing population variations [1][2]. Statistical models are primarily built from principal component analysis (PCA) on point distribution models representing anatomical shapes [3] or from deformation fields [4]. More often than not, point distribution models characterize shape variations only. Hence, when used in applications such as 2D/3D registration, the resulting models would match the input images in shape but not in bone density properties. Cootes *et al.* have proposed a method to create active appearance models (AAM) by concatenating shape and intensity vectors into a single vector and computing PCA

on the resulting matrix [3]. A weight factor W_s is introduced to scale the shape vector so that the units of the shape and the intensity vectors are normalized. Following this work, 3D shape-intensity atlases were created by combining shape and intensity vectors, with $W_s = 1$, as seen in [5], [6], [7]. After computing PCA on the combined vector, the shape and intensity modes were separated and re-normalized. The correlation between the bone shape and the density sub-spaces is not clearly explained in these works and needs to be explored if the shape vectors were to be combined with the intensity vectors.

The application of such shape-intensity atlas in a 2D/3D registration framework can be seen in [5], [8], [9]. Hurvitz *et al.* have proposed a combined atlas with three components consisting of a shape model, a CT-like reference image and a reference surface of the bone and an inverse warping method to estimate the intensity properties [9]. Their model consists of only mean intensity and does not incorporate any intensity variations present in the population. Yao *et al.* [5] and Steininger *et al.* [8] use shape and intensity variations in the registration framework. The proposed approach in these works is to project the prior model and estimate model parameters from the shape-intensity priors that maximizes a given similarity metric. This method requires model instantiation and projection in each step of the optimization algorithm which could be computationally expensive.

In this paper, we extend an existing 2D/3D deformable registration method to incorporate bone density variations. The primary contributions are: 1) a framework to create and analyze bone density variations, separate from shape variations and 2) an augmented 2D/3D registration framework that couples shape and density priors to create accurate patient specific models. We create two models, one for shape and one for CT intensities and integrate them into a 2D/3D registration method. We present alternate methods of computing density statistics in both the CT voxel intensity space and the polynomial approximation. In the registration framework, we register 2D patient images to a shape statistical model first, and successively estimate the density parameters in a least-squares setup. This registration method enables us to create CT-like patient specific models from the 2D X-ray images that match the shape and density properties of the patient. The primary difference between the above cited works and our proposed approach is that we recover the density parameters in a single step by linearly projecting the density model on to the 2D input image space. After the rigid and shape registration of the model, projection images are created for each density mode. The density parameters are computed such that the similarity between the linear combination of these projection images and the input images is maximized. The rest of the paper is organized as follows. We present our density atlas construction and registration method in Section 2, experiments and results in Section 3, followed by discussion in Section 4.

2 Method

In this section, we briefly outline our earlier work on shape statistical models, discuss our new framework to create and integrate density models into a 2D/3D registration method.

2.1 Model Representation

Following Yao's work [5], we use a tetrahedral mesh to represent bone shape and Bernstein polynomials to approximate bone density or CT numbers within each tetrahedron. A tetrahedral mesh is defined by a set of vertices P and a list of tetrahedra T. Each vertex P_i is a point in 3D. The CT densities in each tetrahedron are approximated using Bernstein polynomials. Given a tetrahedron $T_j = (v_0, ..., v_3)$, where v_i is a vertex, they can be arranged in a homogeneous matrix of the form [2]:

$$M_T = \begin{bmatrix} v_0 & v_1 & v_2 & v_3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \tag{1}$$

The density of each cell is approximated using a local Bernstein polynomial function inside the volume of the cell:

$$f_u^j = \sum_{\|k\|=d} \beta_k^j B_k^d(u) = \sum_{\|k\|=d} \beta_k^j \binom{d}{k} u_0^{k_0} u_1^{k_1} u_2^{k_2} u_3^{k_3} \tag{2}$$

Here, $u = (u_0, u_1, u_2, u_3) = M_T^{-1} x$ are the *barycentric coordinates* of a point x in homogeneous coordinates; $k = (k_0, k_1, k_2, k_3)$ is the power index of the basis function B_k^d, with d being the degree of the polynomial; $\binom{d}{k} = \frac{d!}{k_0!k_1!k_2!k_3!}$ is a multinomial factor; and β is a free coefficient. The advantage of this polynomial representation is that it enables us to compute line integrals in a closed form to create digitally reconstructed radiographic (DRR) images in the 2D/3D registration algorithm [2]. Polynomial representation can also be viewed as a dimensionality reduction step.

2.2 Shape Statistical Model

From a population of CT datasets, $V_i, i = 1, 2...n$, we manually select a template dataset, V_T, segment and mesh the anatomy of interest resulting in a template mesh, M_T. We register each CT dataset, V_i, deformably to the template dataset using a 3D grayscale deformable registration method proposed in [10]. This results in a warped volume, V_i^T and a deformation field, D_T^i. We interpolate this deformation field at each vertex location of the template mesh to create mesh instances of the subjects. The vertices of these mesh instances define our point distribution model. We compute PCA on the mesh vertices to create a shape model consisting of mean shape and dominant shape modes. This shape model is further refined by using an iterative bootstrapping technique proposed in [11]. With this method, we update the atlas as datasets are being added and stabilize

the modes by registering the datasets to the atlas iteratively. This method yields consistent shape models. Given the linear parameterization of the shape prior, any new shape instance can be approximated as follows:

$$S = S^0 + \Sigma_{j=1}^{m} \lambda_j S^j \tag{3}$$

where $S^0$is the mean shape, S^j are orthonormal shape mode vectors, and λ_j are mode weights/parameters.

2.3 Density Statistical Model

For each dataset V_i, we have a deformation field, D_T^i, defining the vector flow from the template to the subject space, and a warped volume,V_i^T, obtained by deforming the subject into the template. These warped volumes are shape-independent, i.e., they resemble the template shape, but the intensities are derived from the individual subject datasets. The voxels of these volumes define PDM for analyzing bone density properties. To extract intensity variations, we stack these shape-free volumes into a big matrix and perform PCA on the data matrix resulting in a mean density volume and dominant density modal volumes (see Figure 1(a)).

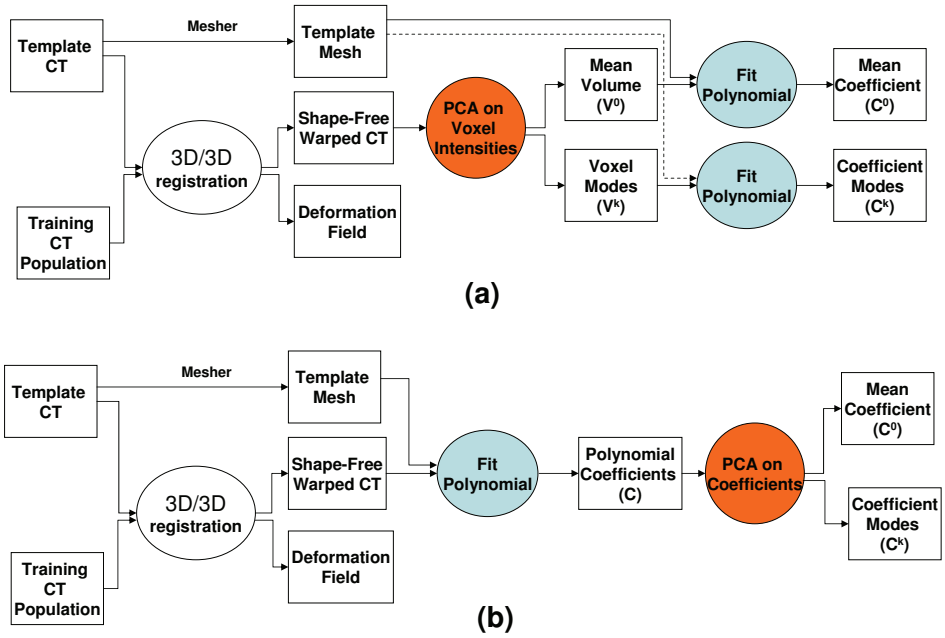


Fig. 1. Density Statistical Model Construction Pipeline: (a) Voxel based density statistical model; (b) Polynomial representation based denstity statistical model;

The parametric linear model of volumetric density model is given as follows.

$$V = V^0 + \Sigma_{k=1}^{n} \mu_k V^k \tag{4}$$

To map voxel-based density modes to polynomial space, we fit 3^{rd} degree Bernstein polynomials to the template mesh and modal density volumes. This is achieved by solving for the unknown coefficients β in equation (2), where f_u^j are derived from the density modes and u from the template mesh.

Alternatively, we fit polynomials to the master mesh M_T and the deformed subjects V_i^T to create polynomial approximation of the intensities of each subject. We then perform PCA on the polynomial coefficients resulting in polynomial density modes (See Figure 1(b)). The linearized parametric form is given as follows:

$$C = C^0 + \Sigma_{k=1}^{n} \mu_k C^k \tag{5}$$

where C^0 is the mean intensity polynomial, C^k are orthonormal intensity mode vectors, and μ_k are mode weights/parameters. To map the polynomial space to voxel space, we transform voxel coordinates to barycentric coordinates, relative to the tetrahedron in which the voxel is contained and then evaluate the polynomial at these points, as shown in equation (2). Here the goal is to compute f_u^j matrix, given the coefficients β and the 3D location, u.

2.4 Registration Framework

We present a new registration framework that combines the shape atlas with a density atlas to create accurate patient specific models (See Figure 2). We use a mutual information based rigid 2D/3D registration method presented in [12], although any such 2D/3D deformable registration method would satisfy our requirements in this work. A detailed analysis of the shape based 2D-3D registration method used in this paper is given in [12]. In this method, projection images or DRRs are created by computing line integrals of the density polynomials along the lines of sight through the space of the mesh cells. This deformable registration method estimates the approximate pose and shape of the patient anatomy by matching the atlas to a set of 2D input images. After estimating the shape, we create projection images of the atlas with various density modes from the intensity atlas. We then solve for the density parameters in a least-square like method that maximizes the similarity between the linear combination of the density projections and the input images. The density parameters are estimated in a single step with this method.

The pseudocode for the algorithm is as follows:

1. *Input:* X-ray images/DRRs I_i, where i is the image number/view angle, camera pose parameters (intrinsic and extrinsic) P_i, mean shape S^0 and shape modes $S^k, k = 1, 2, ..n$, mean density C^0 and density modes $C^l, l = 1, 2, ..m$
2. Register the input images to the shape model, rigidly and deformably to estimate the pose and shape parameters respectively. The output consists of a pose $F(R, T, S)$ and mode weight parameters λ such that

$$(F, \lambda) = \underset{(F,\lambda)}{\operatorname{argmax}} \sum_i MI\left(I_i, DRR\left(F.\left(S^0 + \Sigma_{j=1}^{n} \lambda_j S^j\right)\right)\right) \tag{6}$$

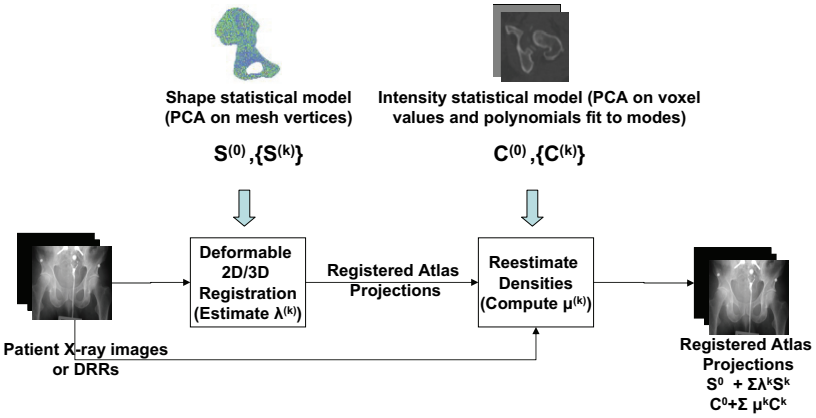


Fig. 2. Registration framework flowchart

3. Transform the shape model and create DRRs of the registered shape model with mean density and density modes.

$$S^{est} = F * \left(S^0 + \Sigma_{j=1}^{m} \lambda_j S^j\right) \tag{7}$$

$$d_i^{mean} = DRR\left(S^{est}, P_i, C_0\right) \tag{8}$$

$$d_i^k = DRR\left(S^{est}, P_i, C_k\right) \tag{9}$$

where d_i^k is the k^{th} density mode projected in i^{th} view direction

4. Formulate a least-squares problem to solve for density mode parameters μ such that

$$\mu = \underset{\mu}{\operatorname{argmin}} \sum_i \left(I_i - \left(d_i^{mean} + \Sigma_{k=1}^{m} \mu_k d_i^k\right)\right)^2 \tag{10}$$

5. Generate the new patient specific model by sampling the estimated shape model on a voxel grid using the estimated coefficients

$$C^{est} = C^0 + \Sigma_{k=1}^{m} \mu_k C^k \tag{11}$$

$$CT_{patient}^{est} = Voxelize\left(S^{est}, C^{est}\right) \tag{12}$$

3 Experiments and Results

3.1 Atlas Experiments

A statistical model of full pelvis anatomy is created from CT scans of 63 healthy individuals. The template mesh model consists of 26875 vertices and 105767 tetrahedra. We have used a $3rd$ order Bernstein polynomials to approximate the CT intensities. The original CT datasets are downsampled from 512x512x82 voxels with voxel size of $0.9375mm^3$ to 256x256x128 voxels and $1.875mm^3$. The

original CT intensities are specified in Hounsfield units (HU) as signed integers. The dynamic range is adjusted such that all the voxels have positive intensity numbers by adding +1000 (HU for air is -1000) to the original units. For volumetric density model, we cropped the volumes to contain 165x95x128 voxels. The projection images of the shape atlas and the density atlas are shown in Figures 3 and 4 respectively.

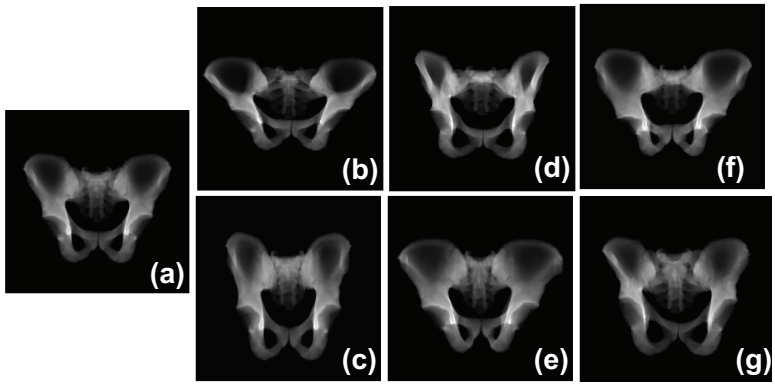


Fig. 3. DRRs of mean shape and first three principal modes. (a) mean shape (S^0); (b)-(c) Mode 1($S^0 \pm 3\lambda_1 S^1$); (d)-(e) Mode 2 ($S^0 \pm 3\lambda_2 S^2$; (f)-(g) Mode 3 ($S^0 \pm 3\lambda_3 S^3$)

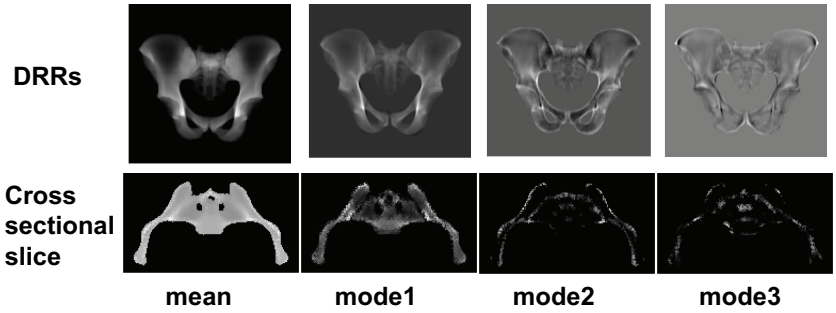


Fig. 4. Top row: DRRs of mean density and first three principal modes (left to right) from polynomial coefficient based density model. Bottom row: An example cross sectional slice of volumes representing mean density and the first three principal modes (left to right) from the volumetric density model

To assess these models, we have performed leave-one-out experiments in which an atlas is created from all but one dataset and the left-out dataset is

reconstructed from the model. For shape atlas, we have measured surface distance between the left out subject and the reconstructed subject from the model. Our analysis shows that we can estimate any given instance with an average accuracy of 1.5 mm using the first 15 dominant shape modes. For density atlas, we have measured the RMS error between the estimated volume and the left out volume in Hounsfield Units.

$$E(V^{true}, V^{est}) = \sqrt{\frac{1}{n_{vox}} \sum_{i=1}^{n_{vox}} \{V^{true} - V^{est}\}^2} \tag{13}$$

Since our registration method requires the density modes in polynomial representation, we have computed the density models in both the volumetric and polynomial space. The leave-out analysis for both volumetric and polynomial based density models is shown in Figure 5. The results show that the density modes from these two representations (voxel and polynomial), are identical and the residual errors from the intensity atlases are comparable. By identical modes, we mean that the intensity mode contributions are similar, i.e., similar anatomical regions are bright in both the images.

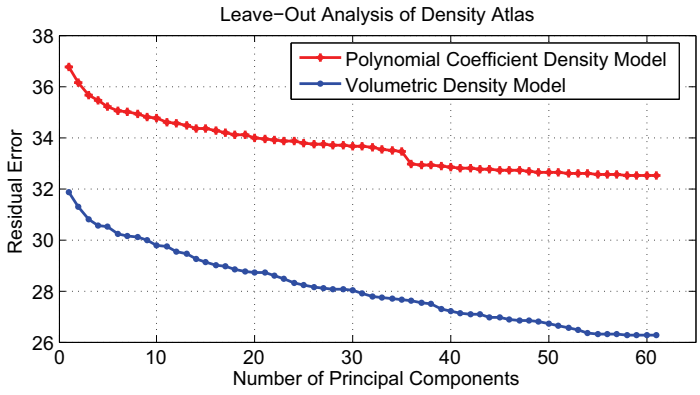


Fig. 5. Root Mean Squared residual errors from leave-out validation using density models from polynomial coefficients and shape free warped volumes

3.2 Registration Experiments

The proposed method is validated using a leave-n-out validation method, with $n = 8$. Eight datasets were randomly selected from a population of 63 datasets. The remaining 55 datasets were used to build the shape and density statistical models. Projection images (DRRs) were created for these 8 datasets with the image size of 512x512 pixels, at views $0°, 90°, 135°$. Based on our analysis presented

above and in [11], we have retained 15 shape modes and 12 density modes along with the mean shape and mean density in the atlas. For each dataset, all three projection images are registered to the shape atlas simultaneously to estimate the shape parameters.

Prior to estimating the density parameters, we render the estimated shape with mean density (1) and density modes (12 modes) for each view angle (3 views) resulting in (12+1) x 3 = 39 DRRs for each left-out subject. After solving for the density parameters in a least-squares setup given in equation (10), we estimate the new polynomial coefficient using equation (5). A CT-like volume is created using the estimated shape and the polynomial. The root mean squared residual errors are computed between the approximated CT and the original CT. Since the goal of the proposed method is to create patient specific models that match both in shape and intensity properties, we chose to use the root mean squared error metric, (See equation 13), between voxel intensities to validate our approach. As we are using DRRs created from the original CT datasets, we could say that the spectrum of CT intensities for both the volumes and the polynomial representation as well as the input DRRs is the same and hence this error metric would be valid. When considering registration with real X-ray images, an intensity calibration step is needed to map the intensities of the input X-ray images to the underlying CT intensities in the density model. The results from these leave-out experiments are shown in Table 1.

(1)	(2)	(3)	(4)	(5)
#	$\mathbf{S^{true}}$– $\mathbf{S^{est}}$ (mm)	**RMS (** $\mathbf{V^{true}}$, $\mathbf{V^{est}_{mean}}$**)** (HU)	**RMS(** $\mathbf{V^{true}}$, $\mathbf{V^{est}_{modes}}$ **)** (HU)	Δ ((3)-(4))/(3) %
1	1.94	109.92	58.88	**46.43**
2	1.62	128.32	96.0	**25.19**
3	1.90	98.4	77.12	**21.63**
4	2.60	51.68	41.6	**19.50**
5	2.48	109.44	84.8	**22.51**
6	1.95	73.44	50.56	**31.15**
7	2.30	72.96	47.52	**34.84**
8	2.93	101.28	85.76	**15.32**
avg	**2.21**	**93.18**	**67.78**	**27.07**

Table 1. Residual errors from leave-out-validation tests of the augmented registration algorithm with the combined atlas of shape and density variations. Columns 3 shows residual errors when using mean density only and column 4 shows residual errors with mean density and density modes. The % reduction in RMS error between columns 3 and 4 is given in Column 5.

The second column shows the accuracy of the shape registration in terms of the surface distance between the registered atlas shape, S^{est} and the original shape S^{true}. The third column presents the root mean squared error between

the original CT volume, V^{true}, and the CT-like approximation of the registered atlas with mean density only, V^{est}_{mean}. The fourth column presents the residual error between the original CT, V^{true}, and the CT-like approximation of the registered atlas with mean density and density modes, V^{est}_{modes}. Column 5 shows the % reduction in the residual errors from using mean density (column 3) to the density modes (column 4). The error values are specified in Hounsfield units.

4 Discussion and Conclusion

We have proposed a framework to analyze bone density variations in a given population. We presented alternate ways of computing density models in either the voxel space or the polynomial space. Our analysis shows that the first 12 eigen modes explain 80% of the variation. The augmented registration framework shows an average reduction of 27% in terms of root mean squared error of the voxel intensities in Hounsfield units, measured after compensating for the CT intensities with the density model. The best case error reduction was 46%. These results show that the bone density properties can be recovered in a single step by linearly projecting the 3D density modes onto the 2D image space. And the resulting 3D registered models exhibit similar bone density properties as the input X-ray images. Although, we have demonstrated the method using a specific 2D/3D registration algorithm, the extension presented here can be applied to any volumetric or mesh-based 2D/3D registration algorithm.

The accuracy of the density atlas depends on the deformable registration method as well as the number of datasets in the population. The sample size is small compared to the dimension of the input vector. We believe that an increase in the population size would result in more compact model generation. We are currently working towards increasing the atlas population to 150 datasets. With a more accurate density model, the registration accuracies can be further improved. Although the accuracy of the density registration depends on the accuracy of shape registration, our results indicate that with a reasonable shape registration, we are still able to achieve a significant reduction in RMS error values with the density model. We are exploring ways to refine the shape registration by estimating the shape parameters and the density parameters in an iterative approach. Other experiments such as using real X-rays instead of DRRs, truncation, partial anatomy etc. to understand the behavior of the density properties in relation to the shape properties need to be performed. However, we believe that the results presented here are promising and validate the hypothesis that the combined shape and intensity atlas would yield better patient specific models.

Acknowledgments

This work was funded in part by NIH grants 1-R01-EB006839-01A1 and 1-R21-EB003616-01, in part by NSF ERC cooperative agreement EEC9731478, and by

Johns Hopkins University internal funds. We would also like to thank Dr. Ted DeWeese and Dr. Lee Myers for providing us with the CT datasets.

References

1. Tang, T., Ellis, R.: 2d/3d deformable registration using a hybrid atlas. In: MICCAI. Volume LNCS 3750, Part II. (2006) 223–230
2. Sadowsky, O.: Image Registration and Hybrid Volume Reconstruction of Bone Anatomy Using a Statistical Shape Atlas. PhD thesis, The Johns Hopkins University (2008)
3. Cootes, T., Beeston, C., Edwards, G., Taylor, C.: A unified framework for atlas matching using active appearance models. In: IPMI. (1999) 322–333
4. Rueckert, D., Frangi, A., Schnabel, J.: Automatic construction of 3d statistical deformation models using non-rigid registration. In: Medical Image Computing and Computer Assisted Intervention. (2001) 77–84
5. Yao, J.: A statistical bone density atlas and deformable medical image registration. Ph. D. Thesis, The Johns Hopkins Univeristy, 2002.
6. Fritscher, K.D., Grunerbl, A., Schubert, R.: 3d image segmentation using combined shape-intensity prior models. International Journal of Computer Assisted Radiology and Surgery **1** (Februaury 2007) 341–350
7. Querol, L., Buchler, P., Rueckert, D., Nolte, L.P., Ballester, M.: Statistical finite element model for bone shape and biomechanical properties. In: R. Larsen, M. Nielsen, and J. Sporning (Eds) MICCAI 2006. Volume 4190. (2006) 405–411
8. Steininger, P., Fritscher, K.D., Kofler, G., Schuler, B., Hanni, M., Schweiger, K., Schubert, R.: Comparison of different metrics for appearance-model-based 2d/3d registration with x-ray images. In: Proc. BMV, Berlin. (2008) 122–127
9. Hurvitz, A., Joskowicz, L.: Registration of a ct-like atlas to fluoroscopic x-ray images using intensity correspondences. International Journal of Computer Assisted Radiology and Surgery **3** (October 2008) 493–504
10. Ellingsen, L., Prince, J.: Deformable registration of ct pelvis images using mjolnir. In: IEEE 7th Nordic Signal Processing Symposium (NORSIG). (2006)
11. Chintalapani, G., Ellingsen, L., Sadowsky, O., Prince, J., Taylor, R.: Statistical atlases of bone anatomy: Construction, iterative improvement and validation. In: MICCAI. Volume 4791/2007. (2007) 499–506
12. Sadowsky, O., Chintalapani, G., Taylor, R.H.: Deformable 2d-3d registration of the pelvis with a limited field of view, using shape statistics. In Ayache, N., Ourselin, S., Maeder, A., eds.: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007. Volume 4792 of LNCS., Springer (2007) 519–526

Barycentric Label Space

Ghassan Hamarneh and Neda Changizi

Medical Image Analysis Lab, School of Computing Science,
Simon Fraser University, Burnaby, Canada
{hamarneh,nca19}@cs.sfu.ca

Abstract. Multiple neighboring organs or structures in medical images are frequently represented by labeling the underlying image (e.g. a brain into WM, GM, CSF). Given the different sources of uncertainties in shape boundaries (e.g. partial volume effect and fuzzy segmentation), it is favorable to adopt a labeling approach that not only encodes uncertainty but also facilitates algebraic label manipulation (e.g. performing PCA). In this work, we extend the label space representation of Malcolm et al. [1] to barycentric label space, in which a proper invertible mapping between probability vectors and label space is proposed. The probability vectors act as barycentric coefficients describing arbitrary labels in label space and a non-singular matrix inversion maps points in label space back to probabilities. The elimination of conversion errors compared to the original label space mapping is demonstrated quantitatively and qualitatively on artificial objects and brain image data, and in the context of smoothing, linear statistics, and uncertainty calculation.

1 Introduction

There are numerous sources of uncertainties in shape boundaries, including graded decomposition [2], image acquisition artifacts, segmentation by multiple-raters, and image segmentation algorithms intentionally designed to output fuzzy results [3, 4]. It is important not to ignore these uncertainties in subsequent analyses and decision-making [2, 5].

There have been numerous works on fuzzy and probabilistic shape representations, speaking to the increasing popularity of manipulating and processing uncertain shapes [6–9]. For example, Sladoje et al. showed that higher precision shape measurements (e.g. boundary length and roundness estimates) are obtained from fuzzy boundaries [10, 11]. Saad et al. demonstrated how measures of label certainty can be leveraged to improve the segmentation [12].

Shape representations come in many different forms; they may be implicit or explicit; boundary or medial-based; binary or multi-shape (or multi-object); and crisp (or hard) or fuzzy (or probabilistic, uncertainty-encoding, etc). Although a full review is beyond the scope of this paper, we highlight some key related works.

Cootes et al.'s seminal work on point distribution models (PDM) was based on representing shapes by (crisp) landmarks followed by performing principle

component analysis (PCA) [13]. Other linear and non linear shape analyses were also performed by many groups on boundary [14–16] as well as medial-based representations [17, 18]. The crisp, impulse-like landmarks were replaced by probabilistic landmarks using Gaussian mixture models (GMM) in [19, 20]. Many approaches represented multi-shapes by labeling the underlying image domain. We highlight main works in this area leading to our proposed approach.

Characteristic function or binary map: $\chi : \mathbb{R}^d \rightarrow \{0, 1\}$ is used to represent a single binary shape, i.e. n=2 (object vs. background), where n is the number of labels or shape classes, and d is the image dimensionality. $\chi(x)$ is 1 if pixel x is interior to the shape and 0 if it's exterior. The set of possible labels is $\mathcal{L}^{n=2} = \{0, 1\}$. Characteristic functions were generalized to multi-shape ($n > 2$) in [21], using vector images $\chi : \mathbb{R}^d \rightarrow \{0, 1\}^n$, where each image layer represents one shape. The labels of this multi-shape crisp representation are the vertices of an $(n-1)$ dimensional simplex with an orthogonal corner, e.g. $\mathcal{L}^3 = \{(0, 0), (1, 0), (0, 1)\}$, $\mathcal{L}^4 = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, etc.

Signed distance function: The characteristic function χ can be represented using a level set of a function ϕ. Typically, ϕ is chosen as the signed distance function (SDF) of the shape boundary, and χ would be the region enclosed within the zero level set of ϕ. By definition, $\phi(x)$ is positive outside the shape (i.e. when $\chi(x) = 0$), negative inside, and zero along the shape boundary $C = \{x \in \mathbb{R}^d | \phi(x) = 0\}$. A multi-shape labelled image ($n > 2$) can be represented by layering n distinct SDFs [21]. In [22], Mansouri et al. used $n-1$ SDFs to represent the n labels, whereas Vese and Chan used a binary encoding of up to n labels using only $ceil(log_2(n))$ level set functions [23]. The reader is referred to [24] and the references within for other alternative multi-object formulations.

Bernoulli and categorical distribution: To allow for fuzzy labels, $\chi(x)$ is replaced by the Bernoulli distribution with probability $p : \mathbb{R}^d \rightarrow [0, 1]$, where $p(x)$ (resp. $1 - p(x)$) measures the probability of pixel x being along or enclosed within the object's boundary (resp. exterior). The set of possible labels becomes $\mathcal{L}^2 = \{(p, 1-p)^t | p \in [0, 1]\}$, where $(.)^t$ denotes transposition. Extending this approach to multi-shape essentially amounts to replacing the Bernoulli with the categorical distribution, where a vector of n probabilities $p : \mathbb{R}^d \rightarrow [0, 1]^n$ is assigned to every pixel, yielding $\mathcal{L}^n = \{(p_1, p_2, \ldots, p_n)^t | p_i \in [0, 1] \wedge \sum_{i=1}^{n} p_i = 1\}$.

The aforementioned representations do not form vector spaces and hence do not facilitate algebraic manipulations. For example, the addition of two SDFs is generally not a valid SDF (e.g. as in [25]), and similarly for the layered characteristic functions. Although feasible solutions are obtained from convex combinations of probability vectors (e.g. [26]), the probability space is not closed under linear combinations. Further, SDFs and characteristic functions are not designed to encode uncertainty. Furthermore, in the layered characteristic functions, one label (typically the 'background') is represented as $\mathbf{0} = (0, 0, 0, \ldots, 0)^t$, i.e. the origin of the orthogonal simplex, which causes a bias (e.g. during smoothing) since $\mathbf{0}$ is a unit distance from any other vertex, whereas any two non-zero labels are separated by a distance of $\sqrt{2}$. The layered approaches also consider

labels independently and their space complexity increases linearly with n (or logarithmically in the case of using a binary encoding, as in [23]).

Unit-hyperspherical labels: To reduce the spatial demand of the layered approaches, Babalola and Cootes distributed the n labels as vectors on the surface of a hypersphere, e.g. $\mathbb{S}^2$ [27]. Besides not forming a vector space, this approach presents a difficulty when deciding which vector to assign to each label, especially since linear combinations of pairs or triplets of labels at boundaries and junctions of different classes should not be mistaken with any other label.

The logarithm of odds ratio (LogOdds): This representation not only encodes label uncertainty but is also closed under addition and scalar multiplication [28]. The first $n-1$ entries of the probability vector are mapped to $logit(p_i) = log(\frac{p_i}{1-\sum_{i=1}^{n-1} p_i})$, i.e. $\mathcal{L}^n = \{(logit(p_1), logit(p_2), \ldots, logit(p_{n-1}))^t | p_i \in (0,1)\}$. Maximally uncertain labels ($p_i = 1/n \ \forall i$) end up as zero level sets of the *logit* function. Since the generalized logistic function is the inverse of the *logit*, a homeomorphism exists between the probability space and LogOdds space, and the vector space structure of the latter is induced on the former. Although powerful, LogOdds suffers from not being suitable for representing certain (or crisp) labels, because whenever $\exists i \ni p_i = 1$, $log(0/1)$ and $log(1/0)$ have singularities. LogOdds also requires normalization and intermediate mapping which may be problematic [28, 1, 29].

Label space: Proposed by Malcolm et al. [1, 29], label space forms a vector space facilitating algebraic manipulations. It does not require normalization or arbitrary mappings and does not produce bias towards any label. It is capable of encoding label uncertainty as well as completely certain labels without singularities. Labels are mapped to a regular simplex in $n-1$ dimensions and the certain labels ($\exists i \ni p_i = 1$), denoted $\{l_i\}_{i=1}^n$, are mapped to the n equally spaced vertices of the simplex. The uncertain labels ($p_i \neq 1 \ \forall i$) are mapped to a linear combination of the crisp labels' positions. In general, any probability vector p is mapped to a unique point l in label space[1] $\mathcal{L}^n \subset \mathbb{R}^{n-1}$ using

$$l = \sum_{i=1}^{n} p_i l_i \in \mathcal{L}^n \tag{1}$$

The label space is thus defined as $\mathcal{L}^n = \{\sum_{i=1}^n p_i l_i \mid p_i \in [0,1] \wedge \sum_{i=1}^n p_i = 1\}$. To calculate the probability for $z \in \mathcal{L}^n$ being label $l \in \mathcal{L}^n$, Malcolm et al. apply:

$$P(z = l) = \exp(-\|z - l\|^2) / \sum_{i=1}^{n} \exp(-\|z - l_i\|^2) \tag{2}$$

To map a point $z \in \mathcal{L}^n$ back to a probability vector $p = (p_1, p_2, \ldots, p_n)$, we must then calculate the probabilities $P(z = l_i) \ \forall i = 1, 2, \ldots, n$, according to (2).

In this paper, we argue that (2) does not provide a proper inverse mapping of (1). If we refer to the mapping in (1) by $f : p \to l$ (probabilities to label space) and to that in (2) by $g_M : l \to p$ (M for Malcolm), then ideally we would like

[1] From this point onward, $\mathcal{L}^n$ refers to label space.

to have $g_M = f^{-1}$, hence $(g_M \circ f)(p) = g_M(f(p)) = p$. However, equation (2) does not in general guarantee this. To demonstrate this, as a simple example, take $n = 3$ and $p = (1, 0, 0)$. Then, according to (1), we obtain $z = l = l_1$. If we now apply (2), we expect to obtain $P(z = l_1) = 1$, since the label of z *is* l_1, but instead we obtain $1/(1 + \exp(-\|l_1 - l_2\|^2) + \exp(-\|l_1 - l_3\|^2)) = 1/(1 + 2\varepsilon) \neq 1$, where $\varepsilon = \exp(-L^2)$, with L being the length of any side of the regular simplex. We also expect $P(z = l_2) = P(z = l_3) = 0$, but instead we obtain $\varepsilon/(1 + 2\varepsilon)$. Similar calculations for crisp labels for any $n \geq 2$, i.e. $p = (1, 0, 0, \ldots, 0)^t$, or a permutation thereof, can be easily shown to give $\hat{p} = (1, \varepsilon, \varepsilon, \ldots, \varepsilon)^t/(1+(n-1)\varepsilon)$, or a respective permutation thereof. We note that for maximally uncertain labels the correct inverse is obtained. As we will see in Section 3, the error in g_M is proportional to the label certainty.

We adopt Malcolm et al.'s forward mapping $f : p \rightarrow l \in \mathcal{L}^n$ in (1). However, in this paper, we propose to replace $g_M : l \rightarrow p$ in (2) with a proper inverse function based on barycentric coordinates: $g_B : l \rightarrow p$ (B for barycentric).

2 Barycentric Label Space

The theory of barycentric coordinates states that the Cartesian coordinates of any n dimensional point located on a $n-1$ dimensional simplex can be calculated as a weighted linear sum of the coordinates of the simplex vertices [30]. In particular, the coefficients of this linear sum are called the barycentric coordinates. From (1), we observe that the probabilities $p_1, p_2, \ldots, p_n$ are also coefficients of a weighted linear sum of simplex vertices: the n certain labels $l_1, l_2, \ldots, l_n \in \mathcal{L}^n$. Therefore, p_i, $i = 1, 2, \ldots, n$, are barycentric coordinates. Therefore, we can regard (1) as the mapping that converts barycentric coordinates p to Cartesian label space coordinates l. In other words, the barycentric representation of $l \in \mathcal{L}^n$ is equivalent to its corresponding probability vector.

We now shift our attention to the procedure for converting the Cartesian coordinates of a point on a simplex to its barycentric coordinates, as we will adopt this same procedure to convert points in label space to their corresponding probability vectors. We denote the Cartesian coordinates of a point in label space as $l = (l_{x_1}, l_{x_2}, \ldots, l_{x_{n-1}})^t \in \mathcal{L}^n \subset \mathbb{R}^{n-1}$ and the coordinates of the i-th certain label as $l_i = (l_{i,x_1}, l_{i,x_2}, \ldots, l_{i,x_{n-1}})^t$.

For $n = 2$, (1) gives $l = p_1 l_1 + p_2 l_2$. Since $p_1 + p_2 = 1$ and $l \in \mathcal{L}^2 \subset \mathbb{R}^1$ is described by a single coordinate, we have $l_{x_1} = p_1 l_{1,x_1} + (1 - p_1) l_{2,x_1}$, which can be re-written as $(l_{1,x_1} - l_{2,x_1}) p_1 = (l_{x_1} - l_{2,x_1})$ and hence $p_1 = (l_{1,x_1} - l_{2,x_1})^{-1}(l_{x_1} - l_{2,x_1})$. For $n = 3$, $l = p_1 l_1 + p_2 l_2 + p_3 l_3 \in \mathcal{L}^3 \subset \mathbb{R}^2$ is rewritten as

$$\begin{pmatrix} l_{x_1} \\ l_{x_2} \end{pmatrix} = p_1 \begin{pmatrix} l_{1,x_1} \\ l_{1,x_2} \end{pmatrix} + p_2 \begin{pmatrix} l_{2,x_1} \\ l_{2,x_2} \end{pmatrix} + (1 - p_1 - p_2) \begin{pmatrix} l_{3,x_1} \\ l_{3,x_2} \end{pmatrix} \tag{3}$$

Re-arranging gives

$$\begin{aligned} p_1 (l_{1,x_1} - l_{3,x_1}) + p_2 (l_{2,x_1} - l_{3,x_1}) + (l_{3,x_1} - l_{x_1}) = 0 \\ p_1 (l_{1,x_2} - l_{3,x_2}) + p_2 (l_{2,x_2} - l_{3,x_2}) + (l_{3,x_2} - l_{x_2}) = 0 \end{aligned} \tag{4}$$

which can be rewritten in matrix form as $\mathbf{T}_3 p = l - l_3$, where the subscript 3 in $\mathbf{T}_3$ indicates the value of n, and where

$$\mathbf{T}_3 = \begin{pmatrix} l_{1,x_1} - l_{3,x_1} & l_{2,x_1} - l_{3,x_1} \\ l_{1,x_2} - l_{3,x_2} & l_{2,x_2} - l_{3,x_2} \end{pmatrix};\ p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}; l - l_3 = \begin{pmatrix} l_{x_1} - l_{3,x_1} \\ l_{x_2} - l_{3,x_2} \end{pmatrix} \quad (5)$$

For any n, we obtain the following linear system of equations

$$\mathbf{T}_n p = l - l_n \quad (6)$$

$$\mathbf{T}_n = \begin{pmatrix} l_{1,x_1} - l_{n,x_1} & l_{2,x_1} - l_{n,x_1} & \dots & l_{n-1,x_1} - l_{n,x_1} \\ l_{1,x_2} - l_{n,x_2} & l_{2,x_2} - l_{n,x_2} & \dots & l_{n-1,x_2} - l_{n,x_2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{1,x_{n-1}} - l_{n,x_{n-1}} & l_{2,x_{n-1}} - l_{n,x_{n-1}} & \dots & l_{n-1,x_n} - l_{n,x_n} \end{pmatrix}; \quad (7)$$

$$p = (p_1, p_2, \dots, p_{n-1})^t;\ l - l_n = (l_{x_1} - l_{n,x_1}, l_{x_2} - l_{n,x_2}, \dots, l_{x_{n-1}} - l_{n,x_{n-1}})^t$$

where $\mathbf{T}_n$ is of size $(n-1)^2$ and its i-th column is $l_i - l_n$. $\mathbf{T}_n$ is non-singular since its $n-1$ columns are essentially the edges of the simplex in $n-1$ dimensions. The probabilities p corresponding to $l \in \mathcal{L}^n$ can now we be obtained using

$$p = \mathbf{T}_n^{-1}(l - l_n); \quad p_n = 1 - p_1 - p_2 - \dots - p_{n-1} \quad (8)$$

3 Results

3.1 Error in the inverse mapping

Adopting (1) as the forward mapping $f : p \to l$, we compare two alternative inverse mappings ($g : l \to p$): Malcolm et al.'s g_M captured by (2) and g_B of our proposed approach summarized in (8). We use a relative inverse composition error $\varepsilon_{inv}(p) = \|(g \circ f)(p) - p\|/\|p\|$, which we calculate for both methods as ε_M^{inv} and ε_B^{inv}. Ideally, $(g \circ f)(p) = p$ and hence $\varepsilon_{inv}(p) = 0$. In Figure 1(a), we examine ε_{inv} as a function of n. Note that ε_B^{inv} is practically zero whereas ε_M^{inv} increases with n and plateaus at roughly 0.7 at n =100. Note that the variance in ε_M^{inv} is decreasing with n. Since ε_{inv} is dependent on p, the mean and standard deviation in Figure 1(a) are the result of a Monte Carlo simulation with 10,000 random probability vectors for each n sampled from a Dirichlet distribution with parameters $\alpha_1 = ... = \alpha_n = 1$, giving a uniform distribution within the $n-1$ dimensional open simplex and zero elsewhere [31].

3.2 Relationship between error and uncertainty

We explore the relationship between the inverse mapping error and the uncertainty. In Figure 1(c), we note that ε_B^{inv} is practically zero (10^{-16}), whereas ε_M^{inv} is only zero at the centre of the simplex ($p_i = 1/3,\ i = 1, 2, 3$) and increases outwards reaching maximal error at the vertices (the certain labels) (Figure 1(d)). In Figure 1(b), we quantify ε_M^{inv} as a function of uncertainty, measured as the normalized (by the maximum entropy for each n) Shannon's entropy $\frac{-\sum_{i=1}^n p_i \log_2 p_i}{-\sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{n}} = -\frac{1}{\log_2 n} \sum_{i=1}^n p_i \log_2 p_i$. Note the monotonically decreasing non-linear behavior of ε_M^{inv} vs. entropy.

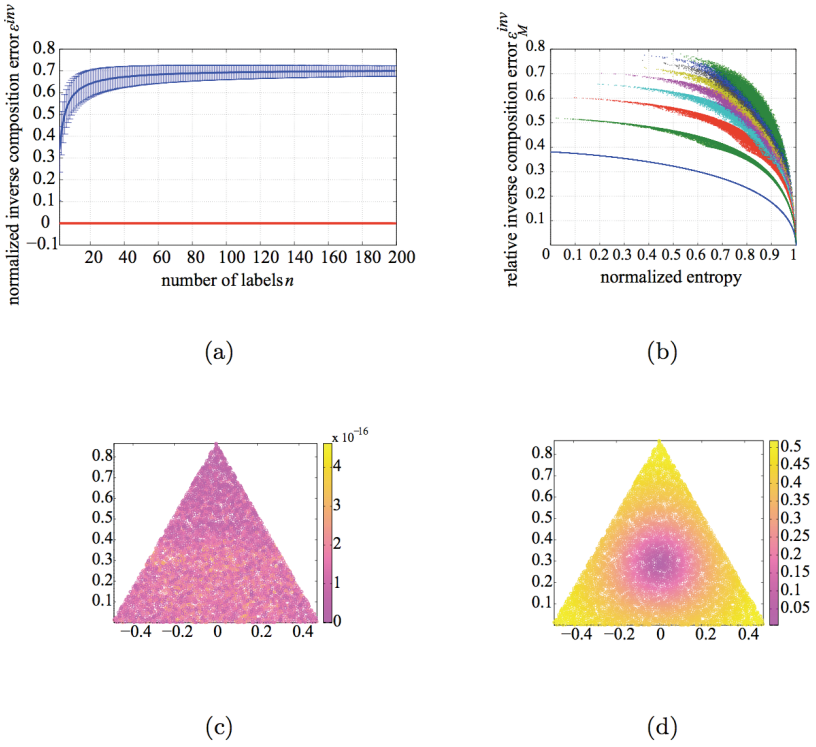


Fig. 1: Relative inverse composition error ε^{inv}. (a) ε_M^{inv} (blue with vertical error bars) and ε_B^{inv} (red) as a function of n. (b) ε_M^{inv} as a function of normalized Shannon's entropy for $n = \{1$=lowest blue curve, $2, 3, \ldots, 10$=topmost green plot$\}$. (c) ε_B^{inv} and (d) ε_M^{inv} at the different points in $\mathcal{L}^3$.

3.3 Effect of error on fuzzy brain images

We applied g_B and g_M to probabilistic maps of 20 subjects from BrainWeb [32]. For each subject and at each pixel, 12 probability values capture the pixel's fuzzy membership to one of 12 classes: cerebrospinal fluid (CSF), gray matter (GM), white matter (WM), etc. (full list in Table 1). For each subject, we calculated $\varepsilon^{msp} = \frac{1}{N}\sum_{i=1}^{N} \varepsilon^{inv}(x_i)$, where N is the number of pixels x_i in the mid-sagittal plane (MSP). The average ε^{msp} over 20 subjects using g_B was $\varepsilon_B^{msp} = 0.780 \times 10^{-16}$ vs. $\varepsilon_M^{msp} = 0.835$ for g_M. In Table 1, we present a similar comparison but with a region-specific error ε^r, $r = \{1, 2, \ldots, 12\}$. ε_B^r (using g_B) is practically zero whereas ε_M^r (using g_M) is ca. 0.83 $\forall r$. Figure 2 presents a qualitative comparison of the probability maps of CSF, GM, and WM for a single subject, averaged over 20 subjects in probability space, and averaged in $\mathcal{L}^{12}$ followed by using either g_B or g_M. Note the similarity in the results obtained by g_B to the average probabilities and the erroneously attenuated probabilities resulting from g_M.

Label	ε_B^r	ε_M^r
1. background	0.212e-16±0.063e-16	0.837 ± 0.000
2. CSF	1.161e-16±0.103e-16	0.834 ± 0.001
3. grey matter	0.989e-16±0.105e-16	0.833 ± 0.001
4. white matter	1.789e-16±0.235e-16	0.833 ± 0.001
5. fat	2.649e-16±0.160e-16	0.836 ± 0.001
6. muscle	1.172e-16±0.121e-16	0.832 ± 0.001
7. muscle/Skin	0.354e-16±0.146e-16	0.836 ± 0.000
8. skull	0.368e-16±0.229e-16	0.836 ± 0.000
9. vessels	1.048e-16±0.044e-16	0.828 ± 0.001
10. around fat	0.852e-16±0.107e-16	0.831 ± 0.002
11. dura matter	0.852e-16±0.071e-16	0.833 ± 0.001
12. bone marrow	0.706e-16±0.061e-16	0.836 ± 0.001

Table 1: ε^r (mean±s.d.) for 12 brain regions averaged over 20 subjects.

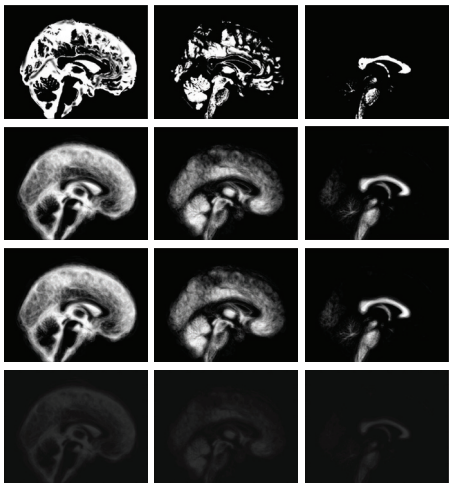


Fig. 2: Comparison between probability maps of CSF (left column), GM (middle column), and WM (right column), obtained in alternative ways (the different rows): the probability map of a single BrainWeb subject (first row), the average probability map of all subjects (2^{nd} row), the average performed in label space and mapped back to probabilities using our barycentric approach, g_B (3^{rd} row), and using Malcolm et al.'s method, g_M (last row). Probability value $p(x) = 1$ is shown as white and $p(x) = 0$ as black.

3.4 PCA in label space

Since our approach shares the same mapping $f : p \to l$ as in [1], performing PCA in label space yields the same variation modes for both approaches. However, we show the effect of errors in g_M when exploring the modes in probability space, compared to accurate results using g_B. In the first experiment, we created a training set of probability maps containing only two types of 50×50-pixel images (Figure 3). One set contains a crisp rectangle, i.e. $p(x) = (1, 0)^t$ inside and $(0, 1)^t$ outside, and the other contains the same rectangle except that its lower side is characterized by uncertainty (the probability changes gradually from inside to outside: $(1, 0), (0.9, 0.1), \ldots, (0, 1)$. The maximum likelihood labeling of both probability maps produce the same labeling. The probability maps are converted to label space $\mathcal{L}^2$ (using (1)), on which PCA is performed. To explore the principle modes of variation in probability space, we used either g_M or g_B

(Figure 3). Note how using g_B results in probability values accurately relating to what is exhibited in the training set, whereas poor probability values result when using g_M. In the second experiment, we followed similar steps but on BrainWeb data, where PCA was performed on $\mathcal{L}^{12} \subset \mathbb{R}^{11}$ of the MSP of all 20 subjects. The results in Figure 4 show improved results when using g_B over g_M.

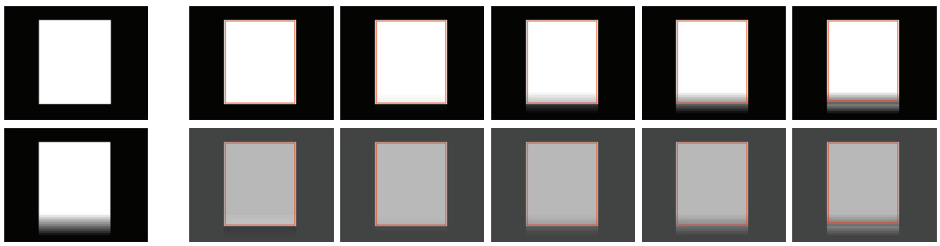


Fig. 3: PCA in $\mathcal{L}^2$ on rectangular shapes. (leftmost column) The two types of images in the training set. (5 rightmost columns) Mean probability map displaced $\{-1, -0.5, 0, 0.5, \text{and } 1\} \times \sqrt{\lambda}$ along the the first mode of variation, where λ is the variance explained by that mode, using g_B (1^{st} row) and g_M (2^{nd} row). $p(x) = 1$ is shown as white and $p(x) = 0$ as black. Note the attenuated probabilities resulting from g_M that do not reflect the values in the training data.

3.5 Smoothing in label space

We re-created the smoothing experiment from [1], in which the label space and binary vector ([21]) representation of an image with 4 labels (including background) are smoothed. As demonstrated in [1], Figure 5(f) shows background labels appearing near the junction between non-background labels when the binary representation is used, whereas smoothing in $\mathcal{L}^4$ smoothes out the sharp corners at the junction without erroneous background labels (Figure 5(e)). The fuzzy labels appearing after smoothing in $\mathcal{L}^4$ are converted to crisp labels using the nearest crisp label. Similar results are obtained if the maximum likely label is chosen. Figure 5(a-d) compares the probability maps obtained using either g_B or g_M following the label space smoothing iterations. Note that g_M inappropriately attenuates the probability values.

3.6 Effect of error on uncertainty

An important advantage of probabilistic representations of multi-shape fields is the ability to calculate uncertainty values in the image domain. Uncertainty calculations have the potential to influence end users' confidence in the results (e.g. probabilistic or multi-rater segmentation results) and may affect decision-making for clinical tasks such as diagnosis or therapy. Further, uncertainty values

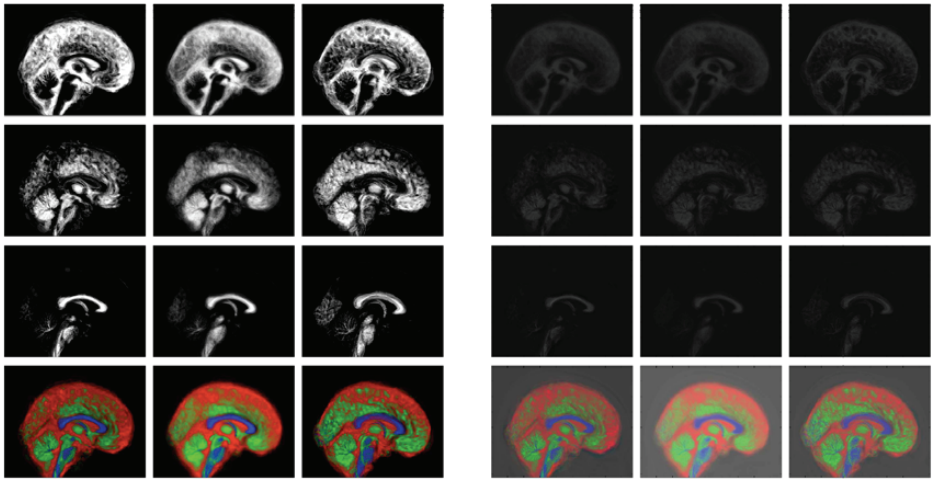


Fig. 4: PCA in $\mathcal{L}^{12}$ on BrainWeb data. Rows 1-3 show the variability in the resulting probability maps for CSF, GM, and WM. $p(x) = 1$ is shown as white and $p(x) = 0$ as black. The 4^{th} row uses the first three rows as RGB channels. Columns 1-3 are obtained using g_B, whereas columns 4-6 are obtained using g_M. The columns 1,2, and 3 correspond the mean probability map displaced by $\{-2, 0, \text{and } 2\} \times \sqrt{\lambda}$, respectively, along the the first mode of variation, and similarly for columns 4,5, and 6, where λ is the variance explained by that mode. Note the low probability values when using g_B.

are useful for self-learning algorithms as demonstrated in [12]. Figure 6 compares the uncertainty values obtained after converting probability fields to label space and back to probability using either our g_B or g_M. The normalized uncertainty at every pixel x is calculated as $U(p(x)) = \frac{-1}{\log_2 n} \sum_{i=1}^{n} p_i(x) \log_2 p_i(x)$, and the absolute error in uncertainty when using g_M is calculated as $\varepsilon_M^{uncert}(x) = |U_M - U|$, where $U_M = U(g_M \circ f)(p(x))$, and similarly for $\varepsilon_B^{uncert}(x)$ when using g_B. The results are obtained on the smoothed data set from Figure 5 (the first smoothing iteration) and on a fuzzy membership map of the MSP of a BrainWeb subject in Figure 2. Note how g_M results in inaccurate uncertainty calculations.

4 Conclusions

Richly labelled fields are common in medical imaging for representing multiple organs or substructures. Given different sources of uncertainty in the labeling (e.g. graded decomposition, multiple raters, fuzzy processing, probabilistic segmentation), it is desirable to have a multi-shape representation that encodes uncertainty. It is also desirable that this representation forms a vector space facilitating algebraic manipulation (e.g. smoothing or statistical analysis), be capable of representing certain as well as uncertain labels, is devoid from arbitrary normalization or other ad-hoc operations, and provides an invertible

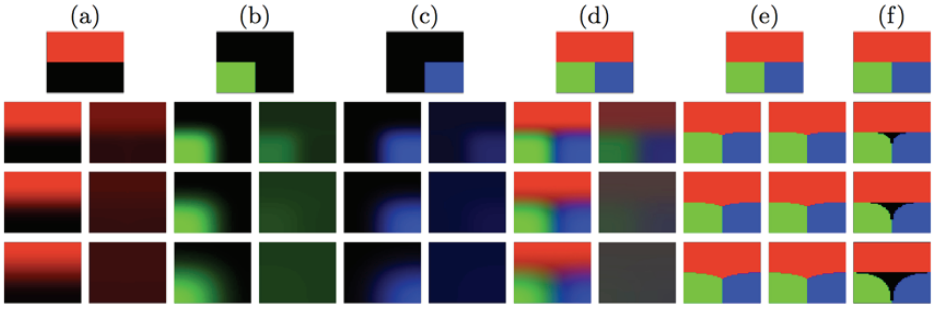


Fig. 5: Smoothing in label space. The top row shows (a) the initial (before smoothing) probability map of region 1 (the top half of the image), (b) region 2 (lower left quadrant), (c) region 3 (lower right quadrant), (d) the three probability maps combined into one RGB rendering, (e) each pixel is labelled according to the nearest crisp label in label space (similar results obtained with the maximum likely label), and (f) using the binary vector representation of [21]. The remaining rows show increasing levels of smoothing performed in label space (except for (f)). Note that there are two, left and right, sub-columns under each main column corresponding to using our method (g_B) and Malcolm et al's (g_M), respectively. In (a-d), pure red, green, and blue correspond to probability 1 and black to zero. In (e), the pure colors correspond to the three labels, with a forth black (or background) label erroneously appearing under (f). Note how our method returns improved probability maps (in (a-d)), even though the crisp labels are the same (in (e)) and are better than (f).

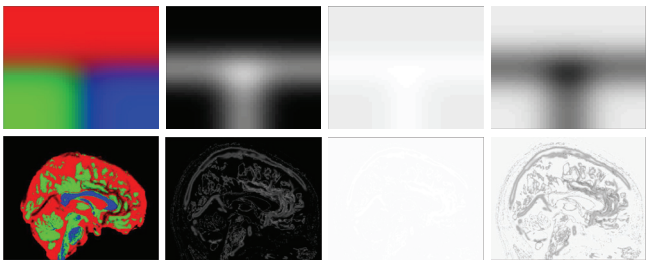


Fig. 6: Effect of the inverse mappings g_B and g_M on uncertainty calculation. (a) Probability map $p(x)$ (three probability entries are visualized as RGB channels). (b) Uncertainty $U(p(x)) = U(g_B \circ f)(p(x))$. (c) $U(g_M \circ f)(p(x))$. (d) Absolute error in uncertainty ε_M^{uncert} ($\varepsilon_B^{uncert} = 0$ not shown). The 1^{st} row uses the first smoothed image of rectangles from Figure 5 whereas the MSP of a BrainWeb subject is used in the 2^{nd} row. The RGB channels in the brain example correspond to CSF, GM, and WM, respectively. In all images pure white, red, green, or blue corresponds to 1 and black to zero.

bijective mapping with probabilistic atlases. Label space is a powerful approach but whose inverse exponential mapping falls short of satisfying that last requirement of invertibility. To address this deficiency, we substituted the exponential inverse mapping (from label space to probability vectors) with a mapping that is based on solving a non-singular system of linear equations, which in turn is based on ideas from the theory of barycentric coordinates. In contrast to the mapping of [1], we showed that in our barycentric approach the composition of the mapping and its inverse guarantees the identity transformation. We quantified these inversion errors for different types of images (e.g. brain, rectangular shapes) and image manipulation scenario (e.g. smoothing, linear statistics, and uncertainty calculation).

References

1. Malcolm, J., Rathi, Y., Shenton, M., Tannenbaum, A.: Label space: A coupled multi-shape representation. MICCAI (2008) 416–424
2. Udupa, J.K., Grevera, G.J.: Go digital, go fuzzy. Pattern Recognition Letters **23**(6) (2002) 743 – 754
3. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging **20**(1) (2001) 45–57
4. Grady, L.: Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(11) (Nov. 2006) 1768–1783
5. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. Medical Imaging, IEEE Transactions on **23**(7) (2004) 903–921
6. Rosenfeld, A., Haber, S.: The perimeter of a fuzzy set. Pattern Recognition **18**(2) (1985) 125–130
7. Saha, P., Udupa, J.: Iterative relative fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (2000) 28–35
8. Bloch, I.: Fuzzy spatial relationships for image processing and interpretation: a review. Image and Vision Computing **23**(2) (2005) 89–110
9. Chanussot, J., Nyström, I., Sladoje, N.: Shape signatures of fuzzy star-shaped sets based on distance from the centroid. Pattern Recogn. Lett. **26**(6) (2005) 735–746
10. Sladoje, N., Lindblad, J.: High-precision boundary length estimation by utilizing gray-level information. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(2) (2009) 357–363
11. Sladoje, N., Nyström, I., Saha, P.K.: Measurements of digitized objects with fuzzy borders in 2D and 3D. Image and Vision Computing **23**(2) (2005) 123–132
12. Saad, A., Hamarneh, G., Moeller, T., Smith, B.: Kinetic modeling based probabilistic segmentation for molecular images. In: MICCAI. (2008) 244–252
13. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Comput. Vis. Image Underst. **61**(1) (1995) 38–59
14. Stegmann, M.B., Sjöstrand, K., Larsen, R.: Sparse modeling of landmark and texture variability using the orthomax criterion. Volume 6144., SPIE (2006) 61441G
15. Dambreville, S., Rathi, Y., Tannen, A.: Shape-based approach to robust image segmentation using kernel pca. IEEE CVPR **1** (2006) 977–984

16. Sozou, P., Cootes, T., Taylor, C., Mauro, E.D., Lanitis, A.: Non-linear point distribution modelling using a multi-layer perceptron. Image and Vision Computing **15** (June 1997) 457–463(7)
17. Ward, A., Hamarneh, G.: GMAT: The groupwise medial axis transform for fuzzy skeletonization and intelligent pruning. IEEE Transactions Pattern Analysis and Machine Intelligence (2009)
18. Fletcher, P., Lu, C., Pizer, S., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging **23**(8) (2004) 995–1005
19. Wang, F., Vemuri, B., Rangarajan, A., Eisenschenk, S.: Simultaneous nonrigid registration of multiple point sets and atlas construction. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(11) (2008) 2011–2022
20. Peter, A.M., Rangarajan, A.: Information geometry for landmark shape analysis: Unifying shape representation and deformation. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(2) (2009) 337–350
21. Tsai, A., Wells, W., Tempany, C., Grimson, E., Willsky, A.: Mutual information in coupled multi-shape model for medical image segmentation. Med Image Anal **8**(4) (2004) 429–445
22. Mansouri, A., Mitiche, A., Vazquez, C.: Multiregion competition: A level set extension of region competition to multiple region image partitioning. Computer Vision and Image Understanding **101**(3) (March 2006) 137–150
23. Vese, L.A., Tony, Chan, F.: A multiphase level set framework for image segmentation using the mumford and shah model. International Journal of Computer Vision **50** (2002) 271–293
24. Fan, X., Bazin, P., Bogovic, J., Bai, Y., Prince, J.: A multiple geometric deformable model framework for homeomorphic 3D medical image segmentation. In: IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. (2008) 1–7
25. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical shape influence in geodesic active contours. In: IEEE CVPR. Volume 1. (2000) 316–323
26. Cremers, D., Schmidt, F., Barthel, F.: Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. IEEE CVPR (June 2008) 1–6
27. Babalola, K., Cootes, T.: Registering richly labelled 3D images. Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on (2006) 868–871
28. Pohl, K.M., Fisher, J., Bouix, S., Shenton, M., McCarley, R.W., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Using the logarithm of odds to define a vector space on probabilistic atlases. Medical Image Analysis **11**(5) (2007) 465–477
29. Malcolm, J., Rathi, Y., Tannenbaum, A.: Label space: A multi-object shape representation. Combinatorial Image Analysis (2008) 185–196
30. Möbius, A.F.: Der barycentrische Calcul. Johann Ambrosius Barth, Leipzig (1827)
31. Connor, R.J., Mosimann, J.E.: Concepts of independence for proportions with a generalization of the dirichlet distribution. Journal of the American Statistical Association **64**(325) (1969) 194–206
32. Cocosco, C.A., Kollokian, V., Kwan, R.K.S., Pike, G.B., Evans, A.C.: Brainweb: Online interface to a 3d mri simulated brain database. NeuroImage **5** (1997) 425

A Hybrid Generative/Discriminative Method for Classification of Regions of Interest in Schizophrenia Brain MRI

D. S. Cheng[1], M. Bicego[1], U. Castellani[1], M. Cristani[1], S. Cerruti[1], M. Bellani[1], G. Rambaldelli[1], M. Aztori[1], P. Brambilla[2], and V. Murino[1]

[1]University of Verona, Verona, Italy
[2]University of Udine, Udine, Italy

Abstract. Schizophrenia research based on magnetic resonance imaging (MRI) traditionally relies on the volumetric analysis of brain matter, either characterizing the whole intracranial volume or studying the attributes of small regions of interest (ROI), corresponding to well-known functional parts in the brain. In this work, we addressed the second scenario, proposing a novel approach able to automatically distinguish schizophrenic patients from normal controls using multiple ROIs. We explore a hybrid generative/discriminative approach, exploiting state of the art generative models via Fisher kernel and support vector machines (SVM). Experimental results, on a dataset of 124 subjects and 7 ROIs, are really encouraging, also in comparison with pure discriminative methods. Moreover, our results find some agreements with previous medical studies in schizophrenia research.

Key words: Hybrid generative/discriminative methods, Fisher kernel, support vector machine, schizophrenia research, brain MRI

1 Introduction

Computational neuroanatomy using magnetic resonance imaging (MRI) is a growing research field that employs image analysis methods to quantify morphological characteristics of different brains [1]. The ultimate goal is to identify structural brain abnormalities by comparing normal subjects with patients affected by a certain disease.

Roughly speaking there are two main categories of methods: (i) methods based on the analysis of regions of interest (ROI), and (ii) methods based on Voxel-based-Morphometry (VBM)[2]. ROI based methods focus on a limited set of brain subparts which are manually traced by experts. Methods based on VBM use the whole brain after a normalization procedure which maps the current brain onto a standard reference, namely the *stereotaxic* space. In this fashion, a voxel-by-voxel correspondence is available among the analyzed subjects.

In this work, we apply pattern recognition techniques to the problem of discriminating subjects affected by schizophrenia. We build our framework on top

of several previous investigations that confirmed the presence of abnormalities in these subjects [3,4,5,6,7,8,9] and extend it to classify healthy (i.e., controls) and unhealthy (i.e., patients) subjects.

Several works have been proposed recently for human brain classification in the context of schizophrenia research [10,11,12]. Beside standard volumetric methods [2,4], the most promising approaches focus on: (i) shape characterization [11], (ii) surface computation [12], and (iii) high dimension pattern classification [10]. In [11] a ROI-based morphometric analysis is introduced by defining a spherical harmonics and a 3D skeleton as shape descriptors. Improvement of such shape-descriptor-based approach with respect to classical volumetric techniques is experimentally shown. In [12] a support vector machine (SVM) has been proposed to classify cortical thickness which has been measured by calculating the Euclidean distance between linked vertices on the inner and outer cortical surfaces. In [10] a new morphological signature has been defined by combining deformation-based morphometry with SVM. In this fashion, multivariate relationships among various anatomical regions have been captured to characterize more effectively the group differences.

In this work, we go beyond volumetric measurements by classifying intensity histograms of the given ROIs. In order to be able to compare intensity values effectively, we perform a preliminary scale normalization based on landmark matching between intensity histograms [13]. The main goal of this study is to verify whether a significant improvement in brains classification can be obtained by exploiting more sophisticated pattern recognition techniques, instead of investigating more complex morphological features from MRI data. Inspired by recent trends in machine learning and pattern recognition research, we explore a hybrid generative/discriminative approach using the Fisher Score Space [14] to represent our data and employing support vector machines (SVM) [15,16] as classifiers. In particular, we based our framework on the so called *constellation* generative model which has been recently successfully applied for object recognition [17].

Generative and discriminative approaches are the two broad categories within which learning and classification methods fall: a generative approach will estimate the joint probability density function (pdf) of the data and class labels and will classify using the posterior probabilities obtained by Bayes' rule, while a discriminative approach will estimate a classification function directly. Generally, methods falling in the latter category obtain lower asymptotic errors. However, generative models remain popular for their ability to capture explicit data attributes and to incorporate missing features. Fisher kernels are designed to get the best of both worlds. In [14], it was shown that it is possible to extract Fisher scores from a generative model and convert them into a Fisher kernel, which may be used for classification by a kernel method, such as the SVM.

The rest of the paper is organized as follows. Section 2 describes our proposed hybrid approach, detailing both the generative and the discriminative parts. Section 3 describes the brains dataset, highlighting the standard medical protocol

which has been involved. Section 4 shows our experimental results, and finally conclusions are drawn in Section 5.

2 The proposed approach

In this paper, we propose to classify between healthy and diseased subjects by using an hybrid generative-discriminative framework. The most known and applied class of hybrid methods relies on the so called generative kernels [14], to be employed with a Support Vector Machine: the basic idea is to employ a generative model to define feature vectors and project objects to the resulting feature space. Therefore, a meaningful similarity/distance measure is defined, leading to a kernel. In the following all the parts of the proposed approach are detailed.

2.1 The generative part

For this part of the approach, the general idea is to choose a generative model capable of considering all the ROI at the same time, together with the relations between them. To this end, we based our framework on the same concepts behind the constellation probabilistic model [17], which foresees the encoding of one object in terms of a fixed number N of object subparts M_j, and relative spatial relationships. In general, object subparts are represented by their appearance A_i, while the spatial relations are encoded by the *shape* X_i (i.e., the relative subparts positions) of the overall configuration.

Here, we apply this intuition to MRI brain scans, by looking at the ROIs as subparts with a definite spatial configuration within the cerebral volume. Ideally, we expect the content and configuration of the ROIs in each subject to be informative enough to recognize patients from controls. It is worth to note that here X_i is not encoding morphological shape properties of subparts which instead are implicitly captured by the appearance A_i. In the following, we will call X_i as *relations* in order to avoid such ambiguity. Note that, while in [17] the correspondences between subparts in different objects is considered missing data, this is not our case, since ROIs identities (e.g., amygdala or thalamus) and matching between subjects are pre-determined. Consequently, instead of having to evaluate multiple combinatorial hypotheses, our model has to evaluate a single combination, incorporating this prior information by design.

Ultimately, we define the following expression for the log-likelihood of a particular class of L subjects $\{O_i\}_{i=1}^{L}$ (assumed to be independent and identically distributed):

$$\log p(\{O_i\}) = \log \prod_{i=1}^{L} p(O_i) = \sum_{i=1}^{L} \log p(O_i) = \sum_{i=1}^{L} \log \left[p(A_i \mid \theta_a) p(X_i \mid \theta_s) \right], \quad (1)$$

where $\theta = \{\theta_a, \theta_s\}$ is the set of appearance and relations model parameters, respectively. Training is performed by estimating the Maximum Likelihood solution θ^{MLE}.

Based on preliminary experiments, we observed that the relations information was not significant since they encode relative positions of ROIs which are the same for all the brains. Therefore we further simplified the model by using only the appearance model part. Note that, similar to [17], the appearances A_i are represented by PCA components obtained from the ROIs intensity histograms, in order to reduce data dimensionality and highlight consistent variations in the distribution of MRI values. As in [17], the appearance model $p(A_i \mid \theta_a)$ is then a Gaussian over the PCA components.

2.2 The discriminative part: the Fisher kernel

Fisher kernels [14] allow an effective general way of mixing generative and discriminative models for classification. In particular, the Fisher kernel approach measures the similarity between the objects by comparing them in the tangent space induced by the trained generative model, which is considered as a point in the Riemannian manifold defined by the chosen family of generative models. In practice, each object is represented by a feature vector, whose components are called Fisher scores, defined by the evaluation of the gradient of the model log-likelihood on the MLE solution. The dimensionality of this space equals the number of parameters. More in detail, given a probabilistic model, the Fisher scores $\phi(O_i)$ are defined through the following derivatives:

$$\phi(O_i) = \frac{\partial}{\partial \theta} \log p(O_i \mid \theta). \tag{2}$$

In particular, starting from Equation 1 and discarding the relations contribution, the Fisher score we obtain is

$$\phi(O_i) = \frac{\partial}{\partial \theta_a} \log p(A_i \mid \theta_a), \tag{3}$$

where θ_a represent the mean and variance parameters of the Gaussian appearance model. Following [14], we employ and train one generative model for both classes.

A kernel can be defined in various ways in the resulting space: the inner product was used in [14], while RBF and polynomial kernel have been proposed in [17].

3 Data and feature extraction

Quantitative data collection and processing in MRI based research implies facing several methodological issues to minimize biases and distortions. The standard approach to dealing with these issues is following well established guidelines, dictated by international organizations, such as the World Health Organization (WHO), or codified by respected institutions, such as leading universities. See [18] for further details.

	Group mean (and SD)*		Statistics		
Characteristic	Control ($n = 60$)	Schizophrenia ($n = 64$)	Test	*df*	*p*
Age, yr	39.95 (11.25) [range 23-60]	38.84 (11.96) [range 18-62]	$t = 0.53$	122	0.60
Male/female	32/28	43/21	$\chi^2 = 2.49$	1	0.11
Age at onset, yr		26.28 (9.17)			
Duration of illness, yr		13.37 (10.30)			

SD = standard deviation; *df* = degrees of freedom; *p* = value of significance.

* Unless otherwise indicated.

Table 1. Some demographic and clinical characteristics of the study groups. The Student's t-test of the age means rejects (at a two-tailed significance level of $p < 0.05$) the hypothesis that the study groups are significantly different in age, and Pearson χ^2 confirms the same for the gender differences.

The dataset used in this work is composed by MRI brain scans of 64 patients affected by schizophrenia and 60 healthy control subjects. Table 1 shows some demographic and clinical characteristics of the study groups. This database has been investigated several times, for example to produce large sample studies aimed at confirming previous reports of physiological abnormalities associated with the given mental illnesses [3,4,8]. Each of these studies focuses on a particular subpart of the brain, a so-called *region of interest* (ROI), whose abnormal activity is noted to affect cognitive processes. Images were acquired and transferred to PC workstations in order to be processed for ROI *tracing*. This latter procedure is the manual annotation of the images, performed by drawing contours enclosing the intended region. It is carried out by a trained expert following a specific protocol for each ROI. The raters generally achieved high interrater reliability, as defined by intra-class correlation coefficient of between 0.94 and 0.97 (see [18] for further details).

The ROIs traced in this dataset are 7 pairs (for the left and the right hemisphere respectively) of disconnected image portions describes as the following:

- Amygdala (*l_amyg* and *r_amyg* in short);
- Dorso-lateral PreFrontal Cortex (*l_dlpfc* and *r_dlpfc*);
- Entorhinal Cortex (*l_ec* and *r_ec*);
- Heschl's Gyrus (*l_hg* and *r_hg*);
- Hippocampus (*l_hippo* and *r_hippo*);
- Superior Temporal Gyrus (*l_stg* and *r_stg*);
- Thalamus (*l_thal* and *r_thal*).

In Fig. 1, we show a sample from the dataset, specifically the ROI volume of *r_stg* for subject 11. This volume is made up of 35 slices of size 41×40 and can be arranged as a montage of images (ordered from left to right, top to bottom). Within this bounding box, the ROI itself is irregularly shaped, as can be clearly seen from the corresponding binary masks on the right, artificially colored to highlight the ROI shape.

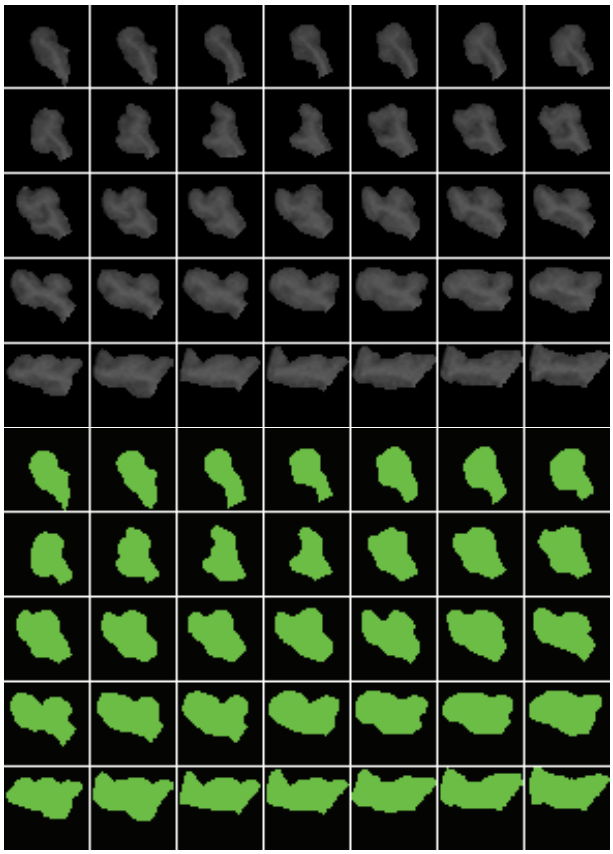


Fig. 1. Montage of the slices in the ROI volume ($41 \times 40 \times 35$) of *r_ stg* for subject 11. On the left, the MRI values; on the right, the corresponding binary masks.

Additionally, another important ROI that is traced is the *intracranial volume* (ICV), that is the volume occupied by the brain in the cranial cavity leaving out the brainstem and the cerebellum. This information is extremely useful for normalizing volume values against differing overall brain sizes.

3.1 MRI Intensity Scale Normalization

A major disadvantage of MRI compared to other imaging techniques is the fact that its intensities are not standardized. Even MR images taken for the same patient on the same scanner with the same protocol at different times may differ in content due to a variety of machine-dependent reasons, therefore, image intensities do not have a fixed meaning [13]. This implies a significant effect on the accuracy and precision of the following image processing, analysis, segmentation and registration methods relying on intensity similarity.

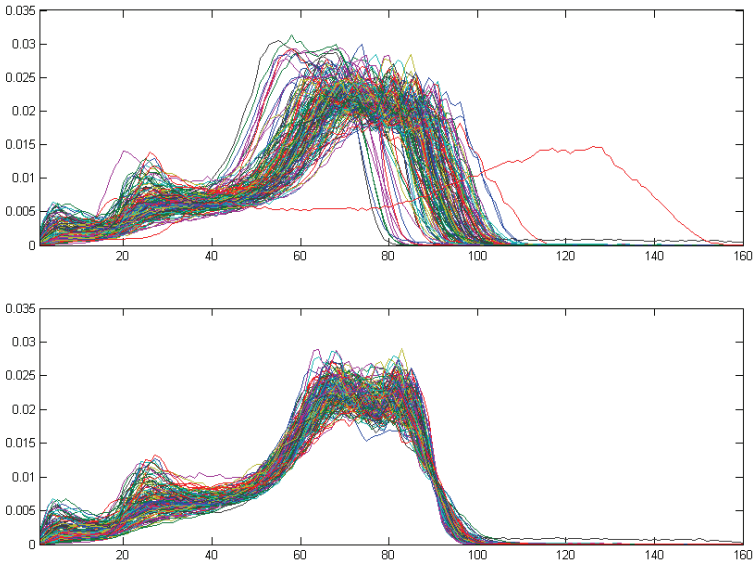


Fig. 2. ICV intensity histograms (treated like probability density functions), before and after the normalization process.

A successful technique used to calibrate MR signal characteristics at the time of acquisition employs *phantoms* [19], by placing physical objects with known attributes within the scanning frame. Unfortunately, this technique is not always exploited, which is our present case. Alternatively, it is possible to obtain good results by retrieving deformation mappings for the image intensities, that is, by developing histogram mappings [20,13].

In this work, we have decided to retrieve the rescaling parameters from the ICV histograms (see Fig. 2). In this way, we focus on the interesting content of the images, which usually contain "noise" in the form of bone and muscle tissue surrounding the brain matter proper. It is also easier to identify landmarks on the histograms that match the canonical subdivision of intracranial tissue into white matter, gray matter and cerebrospinal fluid. We have opted to select a simple rescaling mapping that conserves most of the signal in the gray matter - white matter area, corresponding to the two highest bumps in the range 60-90, since ROIs primarily contain those kinds of tissue.

4 Experimental results

In this part, we will show the effectiveness of the proposed approach using the above described dataset. The goal of the experimental evaluation is twofold: on one hand, we want to provide evidence that using all ROIs at the same time is advantageous with respect to using individual ROIs. On the other hand, we

want to show that an hybrid generative/discriminative model may outperform a simply discriminative approach. To this end, the proposed approach has been compared with two different techniques:

- Single ROI - SVM (RBFkernel): in this case the classification has been carried out using a single ROI. We used the same descriptor employed in the proposed approach, namely histograms whose dimensionality has been reduced with the PCA analysis.
- Multiple ROI - SVM (RBF kernel): in this case, the information coming from all the ROIs is merged. There are many methods for fusing information from different sources (see the huge Multi Classifier System Theory). In this case, we performed a feature level fusion, obtained by simply concatenating the vectors coming from different ROIs. This solution, even if simple, provided optimal results in several contexts (e.g. in Biometrics [21]). Subsequently, the concatenated PCA-reduced vectors have been classified using again a SVM (RBF kernel).

In all cases, the libSVM library [16] has been employed, with optimal parameters chosen via a cross validation analysis. Experiments were carried out in MATLAB and C, whereas accuracies figures for each test run where obtained through leave-one-out (LOO) cross-validation.

Results are proposed in Table 2. From the table, it is evident that most ROIs do not possess significant discriminative powers, but that using all of them at the same time achieves higher accuracy than the individual best ROI. Moreover it is evident that the hybrid generative discriminative approach outperforms the purely discriminative approach, confirming the findings obtained in other fields. Overall, results are suggestive, encouraging in a way, in fact they seem to support the main scientific claim that it is possible to identify schizophrenic patients from healthy people.

From the medical point of view, we can observe that the abnormalities in the amygdala, dorsolateral prefrontal cortex and hippocampus (the three individually most discriminative ROIs in our study) in particular in the left side, are among the most consistent findings in MRI studies on schizophrenia [22,23], suggesting that these structures play a major role for the pathophysiology of the disease [24]. In particular, the dorsolateral prefrontal cortex, along with the thalamus and the hippocampus, is a critic component of the brain circuitry underlying higher cognitive functions, such as attention, executive function and context processing [25]. The amygdala plays a critical role in the neural system that is involved in emotional and in fear-related responses [26]; and the hippocampus is involved in long term memory and in regulating stress response [27,28].

5 Conclusions

In this paper, we proposed a novel approach aimed at discriminating between schizophrenic patients and healthy people based on analyses of brain MR images.

Method	ROI	Accuracy
SVM Single ROIs	*l_amyg*	70.97%
	r_amyg	58.87%
	l_dlpfc	68.55%
	r_dlpfc	47.58%
	l_ec	59.68%
	r_ec	58.06%
	l_hg	58.06%
	r_hg	60.48%
	l_hippo	62.10%
	r_hippo	50.00%
	l_stg	59.68%
	r_stg	56.45%
	l_thal	61.29%
	r_thal	59.68%
SVM Multiple ROIs	*all*	77.42%
Hybrid Approach Multiple ROIs	*all*	80.65%

Table 2. Leave-one-out cross-validation accuracies. Our hybrid approach performs best. Taking all the ROIs performs better than considering them individually.

The proposed approach combines the contribution of different ROIs by exploiting a hybrid generative discriminative method, able to merge the descriptive power of a generative model with the classification accuracies of a discriminative approach. Experimental evaluations on a rather large dataset confirm the appropriateness of the proposed approach, also in comparison with other techniques. Moreover, we have shown that significative improvements can be obtained by focusing on effective classification strategies rather than on the search of complex MRI features. As future work, we envisage a more complex probabilistic modeling, eventually introducing clinical data (e.g., age, gender, illness duration, etc.), to explain variabilities in the data that influence the individual ROIs and hence their overall configuration.

Acknowledgments We acknowledge financial support from the FET programme within the EU-FP7, under the SIMBAD project (contract 213250).

The dataset used in this work is part of a larger database cared by the Research Unit on Brain Imaging and Neuropsychology (RUBIN) at the Department of Medicine and Public Health-Section of Psychiatry and Clinical Psychology of the University of Verona.

References

1. Giuliania, N.R., Calhon, V.D., Pearlson, V.D., Francisd, A., Buchanan, R.W.: Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. Schizophr. Res. **74** (2005) 135–147

2. Ashburner, J., Friston, K.J.: Voxel-based morphometry - the methods. Neuroimage **11** (2000) 805–821
3. Agarwal, N., Rambaldelli, G., Perlini, C., Dusi, N., Kitis, O., Bellani, M., Cerini, R., Isola, M., Versace, A., Balestrieri, M., Gasparini, A., Mucelli, R., Tansella, M., Brambilla, P.: Microstructural thalamic changes in schizophrenia: a combined anatomic and diffusion weighted magnetic resonance imaging study. Journal of Psychiatry & Neuroscience **33**(5) (2008) 440–448
4. Baiano, M., Perlini, C., Rambaldelli, G., Cerini, R., Dusi, N., Bellani, M., Spezzapria, G., Versace, A., Balestrieri, M., Mucelli, R., Tansella, M., Brambilla, P.: Decreased entorhinal cortex volumes in schizophrenia. Schizophr. Res. **102** (2008) 171–180
5. Bernasconi, N., Bernasconi, A., Andermann, F., Dubeau, F., Feindel, W., Reutens, D.C.: Entorhinal cortex in temporal lobe epilepsy: a quantitative MRI study. Neurology **52**(9) (June 1999) 1870–1876
6. Brambilla, P., Harenski, K., Nicoletti, M., Sassi, R.B., Mallinger, A.G., Frank, E., Kupfer, D.J., Keshavan, M.S., Soares, J.C.: MRI investigation of temporal lobe structures in bipolar patients. Journal of Psychiatric Research **37** (2003) 287–295
7. Emmorey, K., Allen, J.S., Bruss, J., Schenker, N., Damasio, H.: A morphometric analysis of auditory brain regions in congenitally deaf adults. Proceedings of the National Academy of Sciences of the United States of America **100** (2003) 10049–10054
8. Potkin, S.G., Turner, J.A., Brown, G.G., McCarthy, G., Greve, D.N., Glover, G.H., Manoach, D.S., Belger, A., Diaz, M., Wible, C.G., Ford, J.M., Mathalon, D.H., Gollub, R., Lauriello, J., O'Leary, D., Erp, T.G.V., Toga, A.W., Preda, A., Lim, K.O.: Working memory and DLPFC inefficiency in schizophrenia: The FBIRN study. Schizophr. Bulletin **35**(1) (2009) 19–31
9. Prasad, K.M., Sahni, S.D., Rohm, B.R., Keshavan, M.S.: Dorsolateral prefrontal cortex morphology and short-term outcome in first-episode schizophrenia. Psychiatry Research **140**(2) (2005) 147–155
10. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: COMPARE: classification of morphological patterns using adaptive regional elements. IEEE Trans. on Medical Imaging **26**(1) (2007) 93–105
11. Gering, G., Styner, M., Lieberman, J.: Shape versus size: Improved understanding of the morphology of brain structures. In: Medical Image Comp. Computer-Assisted Intervention (MICCAI). (2001)
12. Yoon, U., Lee, J., Im, K., Shin, W., Cho, B.H., Kim, I.Y., Kwon, J.S., Kim, S.I.: Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. Neuroimage **34** (2007) 1405–1415
13. Nyúl, L.G., Udupa, J.K., Zhang, X.: New variants of a method of mri scale standardization. IEEE Trans. Med. Imaging **19**(2) (2000) 143–150
14. Jaakkola, T., Haussler, D.: Exploiting generative models in discriminative classifiers. In: Advances in Neural Information Processing Systems 11, MIT Press (1998) 487–493
15. Vapnik, V.: The nature of Statistical Learning Theory. Springer (1995)
16. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.
17. Holub, A.D., Welling, M., Perona, P.: Hybrid generative-discriminative visual categorization. Int. J. Comput. Vision **77**(1-3) (2008) 239–258
18. Cheng, D.S., Bicego, M., Castellani, U., Cerruti, S., Bellani, M., Rambaldelli, G., Aztori, M., Brambilla, P., Murino, V.: Schizophrenia classification using regions of

interest in brain MRI. Technical report, Dipartimento di Informatica, University of Verona, Italy (2009)
19. Edelstein, W.A., Bottomley, P.A., Pfeifer, L.M.: A signal-to-noise calibration procedure for NMR imaging systems. Medical Physics **11** (1984) 180–185
20. Jager, F., Hornegger, J.: Nonrigid registration of joint histograms for intensity standardization in magnetic resonance imaging. IEEE Transactions on Medical Imaging **28**(1) (January 2009) 137–150
21. Ross, A., Nandakumar, K., Jain, A.K., Zhang, D.: Handbook of multibiometrics. Springer (2006)
22. Meisenzahl, E.M., Koutsouleris, N., Bottlender, R., Scheuerecker, J., Jäger, M., Teipel, S.J., Holzinger, S., Frodl, T., Preuss, U., Schmitt, G., Burgermeister, B., Reiser, M., Born, C., Möller, H.J.: Structural brain alterations at different stages of schizophrenia: A voxel-based morphometric study. Schizophrenia Res. **104**(1-3) (Sep 2008) 44–60
23. Shenton, M.E., Dickey, C.C., Frumin, M., McCarley, R.W.: A review of mri findings in schizophrenia. Schizophr. Res. **49**(1-2) (Apr 2001) 1–52
24. Lopez-Garcia, P., Aizenstein, H.J., Snitz, B.E., Walter, R.P., Carter, C.S.: Automated roi-based brain parcellation analysis of frontal and temporal brainvolumes in schizophrenia. Psychiatry Res **147**(2-3) (Oct 2006) 153–161
25. Procyk, E., Goldman-Rakic, P.S.: Modulation of dorsolateral prefrontal delay activity during self-organized behavior. J Neurosci **26**(44) (Nov 2006) 11313–23
26. Swanson, L.W., Petrovich, G.D.: What is the amygdala? Trends in Neurosciences **21**(8) (1998) 323–331
27. Sala, M., Perez, J., Soloff, P., di Nemi, S.U., Caverzasi, E., Soares, J.C., Brambilla, P.: Stress and hippocampal abnormalities in psychiatric disorders. Eur. Neuropsychopharmacology **14**(5) (Oct 2004) 393–405
28. Tulving, E., Markowitsch, H.J.: Episodic and declarative memory: role of the hippocampus. Hippocampus **8**(3) (1998) 198–204

Probabilistic Time Series Models for Statistical Shape Priors in Region Contouring

T. Shepherd and D. C. Alexander

Centre for Medical Image Computing. Department of Computer Science,
University College London, UK.
T.Shepherd@cs.ucl.ac.uk *

Abstract. This paper introduces the benefits of probabilistic time series analysis to the field of medical region contouring. By reviewing where time series methods have been previously used in shape description, we argue that time series analysis has much more to offer than current autoregressive or Markov random field models, through probabilistic methods that lead to full statistical models of shape. We then introduce two examples, namely *Langevin* and *Gaussian Process* models. The combination of generative probabilistic methods allows novel segmentation frameworks. Moreover, we show how the probability distributions can be conditioned by image models and by run-time interactions. The resulting frameworks allow multiple solutions to a single segmentation task, where the end result is ultimately governed by modest user interactions. This, and the underlying nonlinearity of the shape priors, makes the methods particularly suited to tumours and lesions that have irregular form. Experiments validate the segmentation tools and prove the benefits of the global shape information in terms of accuracy and user demand.

1 Introduction

Regions such as tumours and lesions in medical images have poorly defined shape. Due to the complex biological processes that form these regions, there is no obvious shape similarity between examples. In addition, these regions are often associated with high levels of boundary ambiguity due to partial volume effects and low image contrast. In practice, tumour and lesion segmentation calls for semiautomatic methods with often high amounts of user demand. Given the problems of boundary ambiguity and poor shape definition, we believe that segmentation should remain semiautomatic, but that any semiautomatic tool should possess two qualities. First, a tool should give the user complete control over the segmentation, with minimal interaction. Second, the automated part of the semiautomatic framework should be capable of producing alternative solutions, whereby the user ultimately governs the result.

Statistical shape models (SSMs) have proven to benefit medical region segmentation [1, 2] and naturally reduce demand on the user of a semiautomatic

* Work funded by EPSRC. We also thank Dr. D. Tozer at the Institute of Neurology, London, for providing MS lesion contours.

tool. The success of these methods is partly due to their use of generative models to drive an optimisation scheme. Moreover, probabilistic SSMs lend themselves to efficient interactivity, where probability distributions are conditioned on information provided during run-time [3–5]. To apply these ideas to tumour and lesion contouring, we seek probabilistic SSMs that do not assume the same level of shape similarity as the models in [1, 2]. For this purpose, we call on another family of shape models, which represent a region boundary as a *time series*. Previous time series shape models [6–9] have proven to be discriminative, but they do not provide generative, probabilistic, statistical models and their impact on the segmentation community has been relatively low. We revisit the time series literature to vedise new SSMs that are both discriminative and generative, as well as forming the basis of novel, interactive segmentation frameworks by virtue of their probabilistic nature.

The rest of this paper is organised as follows. Section 2 reviews earlier use of time series for shape modelling and motivates a new look at probabilistic methods. Section 3 introduces time series models of *Langevin* and *Gaussian Process* and section 4 defines SSMs based on these, describing their training methods, discriminative and generative techniques and how they are built into segmentation frameworks. Experiments in section 5 test whether using the shape priors reduces demand on the user and increases accuracy for liver tumour and multiple sclerosis (MS) lesion contouring.

2 Previous Work

The field of statistical shape modelling is dominated by methods using the point distribution model (PDM) popularised by Cootes *et al.* [1] in 'Active Shape Model' segmentation. The limitations of these methods stem from their assumption that all examples of a region can be aligned in a meaningful manner, with correspondence points around their boundary. Another family of SSMs [6–9] relax this assumption by representing a 2-dimensional region boundary as a vector of radial distances $\mathbf{r} = \{r_0, \ldots, r_{N-1}\}$, measured from a fixed location $\mathbf{x}_c = \{x_c, y_c\}$ inside the region (figure 1). We refer to $\mathbf{r}$ generally as a *radial time series*, where specific types differ by what 'time' represents. Two examples are boundary arc-length s (eg. [8]) and the angle θ between radial vectors (eg. [6]). Let us define a shape in each case by a parameter set $\mathbf{Q}$ comprising

$$\begin{aligned} \mathbf{Q}_{\text{gen}} &= \{\mathbf{r}, \mathbf{s}, \mathbf{x}_c\} = \{\{r_0, \ldots, r_{N-1}\}, \{s_0, \ldots, s_{N-1}\}, \{x_c, y_c\}\} \quad \text{(a)} \\ \mathbf{Q}_{\text{star}} &= \{\mathbf{r}, \theta, \mathbf{x}_c\} = \{\{r_0, \ldots, r_{N-1}\}, \{\theta_0, \ldots, \theta_{N-1}\}, \{x_c, y_c\}\} \quad \text{(b)}, \end{aligned} \tag{1}$$

where the generalised case $\mathbf{Q}_{\text{gen}}$ in 1a can represent any two-dimensional shape, while the polar representation 1b is limited to the 'star-shaped' set where all boundary points of a shape, denoted $\mathbf{Q}_{\text{star}}$, are visible from $\mathbf{x}_c$. Two main families of radial time series models are the circular autoregressive (CAR) model [6, 7, 10] and Markov models [11, 9, 8]. The CAR model introduced by Kashyap and Chellappa [6] expresses each point r_i as a weighted sum of the radii at

earlier time points on the boundary, plus a noise term, giving the generalised CAR equation

$$r_i = \alpha + \sum_{j=0}^{m-1} p_j r_{i-j} + \sigma\omega_i, \tag{2}$$

where α is proportional to the mean radius, $\mathbf{p} = \{p_0, \ldots, p_j, \ldots, p_{m-1}\}$ is a vector of weights and $m \leq N$ is the number of 'lag terms' in the model, also referred to as the 'order' of the model. The model is used for shape classification, by treating the parameter vectors $\mathbf{p}$ as feature vectors. The method has proven useful for classification and recognition of star-shaped and non star-shaped boundaries that are not too complex, including medical examples (eg. [10]). However, the linearity of the model causes it to struggle in classifying more complex shapes [7]. The CAR model also breaks down when used to characterise boundaries that are noisy [12] or occluded [8]. There are no examples in the literature, of using CAR models in segmentation, either by developing generative models or adapting the classifiers for shape regularisation.

The 1D Cyclic Markov Random Field (1D-CMRF), first seen in [11], treats the radial time series as a vector of N discrete random variables r_i, and each radius r_i as a site in a Markov Random Field (MRF). After choosing a point inside the region, the 1D-CMRF is defined by

$$\Pr(\mathbf{r} = \rho) \geq 0 \tag{3}$$

and

$$\Pr(r_i = \rho_i | r_j = \rho_j, j \neq i) = \Pr(r_i = \rho_i | r_j = \rho_j, j \in \mathcal{W}_i), \tag{4}$$

where $\rho = \{\rho_0, \ldots, \rho_i, \ldots, \rho_N\}$ is a possible configuration for $\mathbf{r}$, ρ_i is a 'hidden variable' or sample point for r_i and $\mathcal{W}_i$ is a 'clique' or neighbourhood of i. MRF methods that if equation 3 holds then the joint probability $\Pr(\mathbf{r})$ is uniquely determined by conditional probabilities in equation 4 and that these follow a Gibbs distribution, i.e. $\Pr(\mathbf{r}) \propto \exp[-\sum_{j \in \mathcal{W}_i} U_j]$ where U_j is an energy function or 'clique potential' that embodies the *a priori* information, combining image and shape priors. In [9] the clique potential is a weighted sum of both 'low-level' (smoothness) and 'high-level' constraints. The high-level information is a crude shape model, which penalises differences between a radial time series and those in a 'library' of training examples. This is not strictly a statistical shape model.

In summary, radial time series models show promise in describing medical regions without assuming high level shape similarity, but there is a need for probabilistic models that are generative, nonlinear, give a true statistical representation of training data and lend themselves to supervised segmentation. The next section describes dynamical models that offer these properties.

3 Background

3.1 Langevin models

Langevin models describe the dynamics of a time dependent state vector $\mathbf{v}(t)$ as a stochastic process. The models are characterised by a deterministic part

$A(\mathbf{v}(t))$ known as the *drift* function and a stochastic part $B(\mathbf{v}(t))$ known as the *diffusion* function. The generalised Langevin equation is given by

$$\mathrm{d}\mathbf{v}/\mathrm{d}t = A(\mathbf{v}(t)) + B(\mathbf{v}(t))\omega(t), \tag{5}$$

where $\omega(t)$ is uncorrelated, time dependent noise with zero mean. Langevin models assume a Markov property, encoded in the *transition density*

$$\Pr(\mathbf{v}(t)) = \Pr(\mathbf{v}(t)|\mathbf{v}(t - \Delta t)), \tag{6}$$

where Δt is a constant delay parameter. The transition density evolves according to a Fokker-Planck equation. The first and second conditional moments in the Fokker-Planck equation, denoted $\mathbf{D}^{(1)}$ and $\mathbf{D}^{(2)}$, relate to the Langevin drift and diffusion functions by

$$A(\mathbf{v}(t), t) \sim \mathbf{D}^{(1)}(\mathbf{v}(t), t)/\Delta t \quad \text{and} \quad B(\mathbf{v}(t), t) \sim \mathbf{D}^{(2)}(\mathbf{v}(t), t)/\sqrt{\Delta t}\,. \tag{7}$$

Training a Langevin model involves finding parametric functions for the drift $A(\mathbf{v}(t), \mathbf{a})$ and diffusion $B(\mathbf{v}(t), \mathbf{b})$, where $\mathbf{a}$ and $\mathbf{b}$ are parameter vectors, by choosing simple functional forms and estimating the parameters from training data.

In section 4.1 we will see that shape scoring can take advantage of the conditional probabilities (transition densities) central to Langevin models. For this we notice that the joint probability of a series of N points under a Langevin model is given by

$$\Pr(\mathbf{v}) = \Pr(\mathbf{v}_0) \prod_{i=0}^{N-1} \Pr(\mathbf{v}(t + \Delta t)|\mathbf{v}_{i-1} = \mathbf{v}(t)) \tag{8}$$

and use equation 8 in an objective function

Finally, generative Langevin models approximate the time evolution of the state vector as the limit of a stochastic difference equation (SDE), following Itô's interpretation [13]. A Langevin SDE has the form

$$\mathrm{d}\mathbf{v}(t) = A(\mathbf{v}(t))\mathrm{d}t + B(\mathbf{v}(t))\omega(t) \tag{9}$$

where $\mathrm{d}t$ is an integration time step, which can be solved numerically by stochastic integration using the explicit Euler-Maruyama scheme

$$v(t + \mathrm{d}t) = \mathbf{v}(t) + \mathrm{d}t\ A(\mathbf{v}(t)) + \sqrt{\mathrm{d}t}\ B(\mathbf{v}(t))\omega(t). \tag{10}$$

3.2 Gaussian Process models

A Gaussian Process (GP) treats a time series as a N-dimensional random vector of outputs $\mathbf{v} = \{v_0, \ldots, v_i, \ldots, v_{N-1}\}$ corresponding to inputs at discrete time points $\mathbf{t} = \{t_0, \ldots, t_i, \ldots, t_{N-1}\}$. The output at time t_i has an associated probability $\Pr(v_i|t_i)$, which follows a normal distribution and the vector of outputs

has a multivariate normal distribution $\mathbf{v} \sim \mathcal{N}_{\mathrm{N}}(\mu, \mathbf{\Sigma}(\mathbf{v}, \mathbf{v}))$, where μ is a discrete mean function and $\mathbf{\Sigma}(\mathbf{v}, \mathbf{v})$ is the $N \times N$ matrix of covariances $\varepsilon(v_i, v_j)$ between pairs of outputs. Each covariance is a function of the corresponding inputs $\varepsilon(v_i, v_j) = f(t_i, t_j)$ where f is a kernel function. The mean function μ and the covariance function $\varepsilon(v_i, v_j)$ completely define a GP [14].

Training a GP involves fitting a discrete function μ and a parametric kernel function $\varepsilon(v_i, v_j, \mathbf{a})$ to training data, where $\mathbf{a}$ is a vector of parameters. Ideally the functions and parameters are estimated together by Bayesian model selection [14]. It is common however, to choose a sensible mean function and form of the covariance kernel, and the task reduces to estimating the parameters $\mathbf{a}$.

As in the Langevin case, we wish to 'score' unseen radial time series according to the probability that they belong to a shape model. In section 4.2 we will see that a probabilistic shape score can take advantage of the fact that GPs are based on multivariate normal distributions.

Finally, generative GP models simulate a time series by drawing a random vector of outputs $\mathbf{v}'$ from the prior distribution. This involves solving

$$\mathbf{v}' = \mu + \mathbf{A}\mathbf{z}, \tag{11}$$

where $\mathbf{A}$ is the Cholesky decomposition of $\mathbf{\Sigma}(\mathbf{v}, \mathbf{v})$ and $\mathbf{z}$ is a N-dimensional vector of independent values drawn from $\mathcal{N}(0, 1)$.

4 Probabilistic Models in Supervised Contouring Frameworks

This section introduces common components of the new time series frameworks followed by the shape modelling techniques specific to Langevin (section 4.1) and GP(section 4.1) SSMs.

Starting with a series derived from a training contour, we take the centre $\bar{r}$ of the occupied state space and subtract this from the whole series, translating it into a *zero-mean field*, and separately store $\bar{r}$ as the scale parameter. Figure 1 illustrates the relationship between zero-mean field, radial time series and shape for the star-shaped parametrisation.

Next, we define an image observation model in the time series framework, given here for the star-shaped case. We introduce a *radial profile model* , using a boundary measure as a function of radius. The boundary measure is based on both the magnitude $|g|$ and direction ψ of the image gradient as shown in figure 2. We define an estimate $\mathbf{x}'_c$ of the region centre by a pixel selected manually by the user of an interactive tool. After obtaining $\mathbf{x}'_c$ we sample $|g|$ and ψ along the each radial vector θ_i. In the case of the gradient magnitude we fit a Gaussian function with mean $\hat{g}_i$ at the first peak of $|g|$ and standard deviation σ_i^g given by the full width at half maximum. Next we take an estimate $\bar{r}'$ of the scale parameter from the mid-point of all the profile means $\hat{g}_i, i = 0, \ldots, N_{\mathrm{obs}} - 1$ where N_{obs} is the number of observation angles. Finally, we form the likelihood ratio $p_i^{\mathrm{on}}/p_i^{\mathrm{off}}$, where p^{on} and p^{off} represent the probabilities that the local

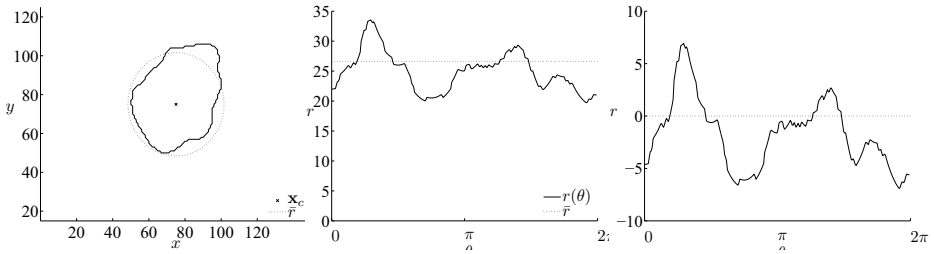


Fig. 1. Example training contour from a liver tumour (a), with estimates of the centre $\mathbf{x}_c$ and scale parameter $\bar{r}$, and the corresponding radial time series (b) in original polar co-ordinates and (c) translated into the zero-mean field.

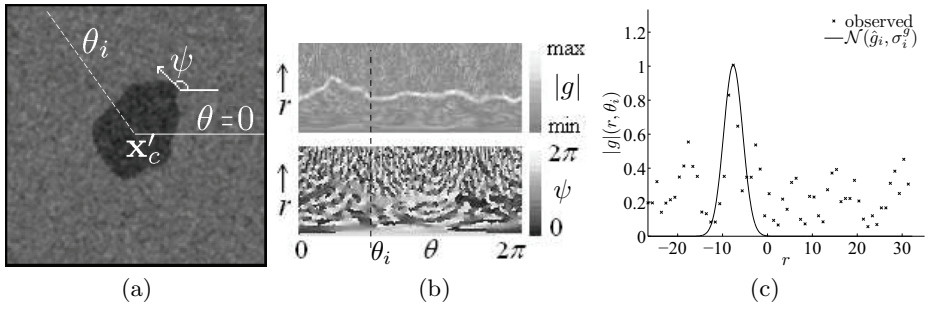


Fig. 2. Observation model from an example synthetic image. (a) Synthetic region with boundary given by a liver tumour contour, showing an estimate of the centre $\mathbf{x}'_c$, local boundary direction φ and radial vector at arbitrary angle θ_i. (b) Greyscale representations of the magnitude $|g|$ (top) and direction φ (bottom) of image gradient sampled along radial vectors, with angle θ_i marked. (c) Radial profile of gradient magnitude corresponding to angle θ_i, with Gaussian fit after translating into the zero-mean field and re-scaling to the range $\{0 \dots 1\}$.

section of a generated shape corresponding to $\{r_i, \theta_i\}$ is on or off the region boundary, given by

$$\begin{aligned} p_i^{\text{on}}(r) &= \exp\left[-(\psi_i(r) - \phi_i(r))^2\right], \text{and} \\ p_i^{\text{off}}(r) &= 1 - \exp\left[-(\frac{r - \hat{g}_i}{\sigma_i^g})^2\right], \end{aligned} \tag{12}$$

and ϕ_i is the angle with respect to the horizontal, made by the contour section from point $\{r_{i-1}, \theta_{i-1}\}$ to $\{r_i, \theta_i\}$.

The observation model will be used in conjunction with generative SSMs to form the basis of a probabilistic optimisation scheme in segmentation. Sections 4.1 and 4.2 present Langevin and GP SSMs using the general time series notation $r(t)$, and demonstrate for polar parametrisation $r(\theta)$.

4.1 Langevin SSMs

We train a Langevin model by adapting the method in [15] to learn from M series $\mathbf{r}^m$, $m = 1 \dots M$, created as above. We divide the range $\{r_{\min}, \dots, r_{\max}\}$ of the whole training set into bins of equal width Δr centred on discrete values r_n. For the nth bin, we identify all points on all series for which $|r(t) - r_n| < \Delta r/2$. We then construct a histogram of radii at the displaced positions $r(t + \Delta t)$ from all these points, allowing wrapping of the closed boundary, to obtain an approximation of the transition density $\Pr(r(t + \Delta t)|r(t) - r_n < \frac{\Delta r}{2})$, which we model with a normal distribution by estimating the mean μ_n and standard deviation σ_n. Repeating for all bins gives a discrete approximation of the drift and diffusion functions

$$\begin{aligned} &A(r_n(t)) = \mu_n - r_n \quad &&\text{and} \\ &B(r_n(t)) = \sigma_n &&\text{where } r_n \in \{r_{\min}, \dots, r_n, \dots, r_{\max}\}. \end{aligned} \tag{13}$$

Finally, we use a Levenberg-Marquardt routine to estimate parameters $\mathbf{a}$ and $\mathbf{b}$ for a chosen family of drift and diffusion functions that best fit the discrete estimations. We find that exponential, quadratic and polynomial functions work well in practice. We choose the function with the lowest χ^2 error to model a given region type with a given contour parametrisation. Figure 3 shows examples of trained Langevin models.

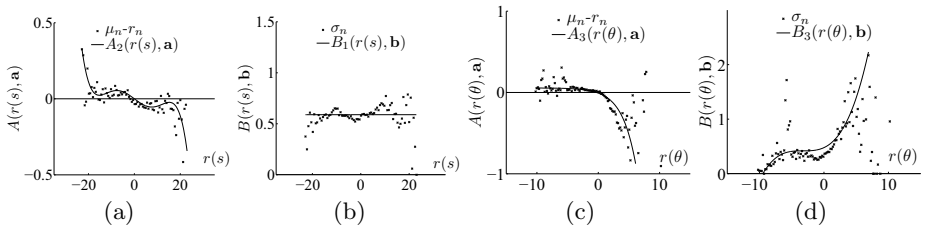


Fig. 3. Drift/diffusion functions (a)/(b) for liver tumours with generalised parametrisation and (c)/(d) for MS lesions with star-shaped parametrisation. A_2 is 5^{th} order polynomial, B_1 is constant, A_3 is exponential and B_3 is cubic.

We derive a probabilistic 'score' for unseen shapes from the log of the joint conditional probabilities. The score is given by

$$S_{\text{LAN}} = \frac{1}{N}\left(\log \Pr(r(t_0)) + \left[\sum_{i=1}^{N-1} \log \Pr(r(t_i + \Delta t)|r(t_i), \mathbf{a})\right] + \log \Pr(\bar{r})\right). \tag{14}$$

Next we present a generative model based on the Euler-Maruyama scheme for shape generation in polar co-ordinates

$$r(\theta_{i+1}) = r(\theta_i) + \mathrm{d}\theta\ A(r(\theta_i)) + \sqrt{\mathrm{d}\theta_i}\ B(r(\theta_i))\omega(\theta_i) + \Delta x \cos(\theta_i) + \Delta y \sin(\theta_i), \tag{15}$$

where $\{\Delta x, \Delta y\}$ allow uncertainty on the estimate of the region centre $\mathbf{x}'_c$, which we build into a further adaptation of the Euler-Maruyama scheme. We combine equation (15) with a boundary tracking algorithm in polar co-ordinates. The algorithm repeats the SDE solution many times at each step θ_{i+1}, which draws several samples from the transition density $\Pr(r_{i+1}|r_i, \mathbf{a}, \mathbf{b})$. This forms the prior distribution in a particle filtering algorithm that is otherwise similar to that in [16]. The algorithm uses factored sampling to compute posterior densities

$$\Pr(r_{i+1}|r_i, \mathbf{a}, \mathbf{b}, \mathcal{D}) \propto \Pr(r_i|\mathbf{a}, \mathbf{b}, \mathcal{D}) \times q(r_{i+1}|r_i, \mathbf{a}, \mathbf{b}) \times l(\mathcal{D}|r_{i+1}), \tag{16}$$

where $q(r_{i+1}|r_i, \mathbf{a}, \mathbf{b}) = \mathcal{N}(r_i - A(r_{i+1}, \mathbf{a}), B(r_{i+1}, \mathbf{b}))$ is the shape prior arising from equation (15) and $l(\mathcal{D}|r_{i+1})$ is a data likelihood term derived from the radial profile model above. For a given $\{\Delta x, \Delta y\}$ we solve equation (16), terminating when a series of N radii forms a closed loop with negligible discontinuity. We repeat for several $\mathbf{x}'_c$ drawn from $\mathcal{N}_2(\mathbf{x}'_c, \sigma_c)$ where the variance σ_c models the uncertainty on $\mathbf{x}_c$. We weight each hypothesis by equation (14) and estimate the maximum *a-posteriori* (MAP) solution by factored sampling.

4.2 Gaussian Process SSMs

We present a general GP SSM using a periodic kernel function $\varepsilon(t_i - t_j, \mathbf{a}) = \exp[-a \sin^2((t_j - t_i)/2)]$, with a single free parameter a related to the length scale of correlation. We use a constant mean function for rotation invariance. To train the GP model we estimate kernel parameters using Markov Chain Monte Carlo (MCMC) methods. Starting with M training series, we seek the parameter a that maximises the joint probability density function $\Pr(\mathbf{r}^{1:M}|a)$. For convenience we take the sum of the log probabilities $\sum_{m=1}^{M} S^m_{\mathrm{GP}}$ where

$$S^m_{\mathrm{GP}} = \frac{N}{2}\log(2\pi) - \frac{1}{2}\log(|\mathbf{\Sigma}(\mathbf{a})|) - \frac{1}{2}\big((\mathbf{r^m} - \mu)^T \mathbf{\Sigma}^{-1}(\mathbf{a})(\mathbf{r^m} - \mu)\big) + \log\Pr(\bar{r}). \tag{17}$$

We also use equation 17 as a probabilistic score for unseen shapes.

Next we constrain the generative model by treating the radial profiles in $\mathcal{D}$ as *noisy observations* and using these to condition the prior following the description in [14]. This yields the posterior covariance matrix $\mathbf{\Sigma}_{\mathrm{post}}$ replacing $\mathbf{\Sigma}$ in equation (11). Finally the MAP solution is given by the posterior mean μ_{post} which we calculate analytically. As in the Langevin case we repeat for several centre point estimates $\mathbf{x}'_c$, scoring each solution by (17) and selecting the final contour by factored sampling. Finally, we incorporate a novel method of user-supervision shown in figure 4, which exploits the probabilistic nature of a GP. The user marks a point on the region boundary missed by the contour. The polar-co-ordinates of such a point define a low-noise observation that further conditions the prior model.

5 Experiments

We perform user trials to test the segmentation frameworks and evaluate the SSMs for the chosen applications of liver tumour and MS lesion contouring. We

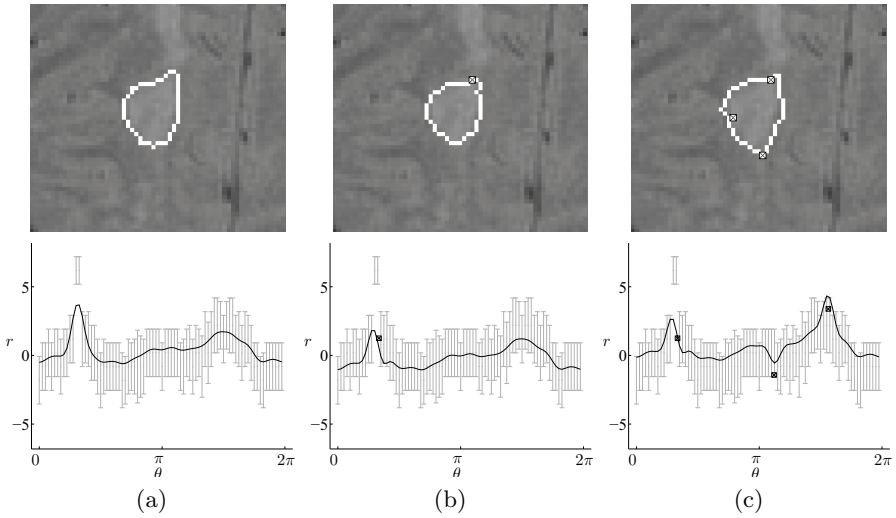


Fig. 4. *Top row:* shapes generated by the GP SSM during supervised contouring, showing the initial contour (a) followed by MAP estimates from the refined model when the user selects (b) one, and (c) three boundary points marked '⊠'. *Bottom row:* corresponding radial time series in the zero-mean field, where black lines show the MAP solution, grey points show noisy observations $\hat{g} \pm \sigma^g$ from the radial profile model and '⊠' are polar representations of user-identified boundary points.

start with 241 star-shaped training contours from manual liver tumour segmentations [17] and 1307 from MS lesions. We remove three test contours from each set and train SSMs on the remaining contours. We use the test contours to create synthetic images with known ground truth by assigning different greylevel histograms inside and outside a contour, then smoothing the result with a 3×3 pixel averaging kernel (figure 5). In the case of MS lesions we also use the PD weighted MR images from where the training data were extracted, so in total there are 9 images to segment.

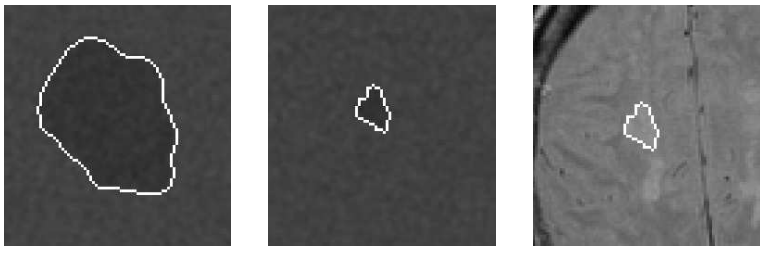


Fig. 5. (a)/(b) Example synthetic liver tumour/MS lesion. (c) An MRI MS lesion.

To isolate the role of the shape priors we compare each framework with a second version, where the learned shape information is replaced with a *normal* prior. In the Langevin case we replace drift and diffusion functions with stationary transition densities $\mathcal{N}(0,1)$. In the GP case we replace the covariance matrix with the $N \times N$ identity matrix.

We asked 10 volunteers to delineate the 9 regions in a randomised sequence, using both tools with learned and normal priors. Segmentation comprises two stages of initialisation by choosing a centre point $\mathbf{x}_c$, which the user can change until satisfied, and post-editing. In the Langevin framework, post-editing involves 'dragging' contour points onto the region boundary. In the GP framework, users can mark boundary points as low-noise observations (section 4.2) and the updated model is displayed in real-time. We evaluate the tools in terms of user demand, accuracy and variability, and perform paired-samples t-tests to identify significant differences between a tool used with learned and normal priors.

We measure **user demand** by the number of interactions N_{int}, where one interaction is a contour point 'dragged' for the Langevin tool or a boundary point selected for the GP tool. we also use the Hausdorff distance d_H [18] to measure the similarity between a final contour accepted by the user and the initial contour before post editing. We take the mean N_{int} or d_H over 10 users and compare a tool using learned and normal priors. Tables 1 and 2 show the results for the Langevin and GP tools respectively. These results reveal increased use-

	Measure	Prior	Langevin SSM		
			synth. LT	synth. MS	MRI MS
use.	N_{int}	normal	16.08±6.45	8.68±2.71	11.58±5.62
		learned	10.52±5.90* (0.002)	8.18±3.90	10.25±3.93
	d_H	normal	6.20±1.27	3.01±0.43	2.81±0.48
		learned	4.41±1.47* (0.019)	2.76±0.73	2.77±0.70
acc.	DSC	normal	0.96±0.01	0.87±0.02	0.80±0.02
		learned	0.96±0.01	0.88±0.02	0.80±0.02
	MMD	normal	0.77±0.13	0.63±0.15	1.23±0.04
		learned	0.79±0.13	0.63±0.11	1.15±0.05 * (0.050)

Table 1. Differences in useability (use.) and accuracy (acc.) between Langevin segmentation with and without learned shape priors, where p-values (in brackets) indicate significant improvement for the learned priors.

ability in all cases where the tools use learned shape knowledge, with significant improvements for all region types with the GP tool and for the (larger) liver tumour shapes with the Langevin tool. We also count the number of times a user repeats an initialisation before accepting or editing a contour. The learned shape priors lead to a significant reduction in this number for both tools and all region types with $p = 0.017$ for Langevin and $p = 0.039$ for GP tools.

	Measure	Prior	Gaussian Process SSM		
			synth. LT	synth. MS	MRI MS
use.	N_{int}	normal	9.88±3.08	3.66±2.34	4.23±1.81
		learned	4.93±1.96* (< 0.001)	2.43±1.55* (0.010)	2.73±1.30* (0.001)
	d_H	normal	4.26±0.91	1.62±0.50	1.98±0.35
		learned	3.69±0.74* (0.021)	1.54±0.58	1.88±0.49
acc.	DSC	normal	0.93±0.01	0.85±0.01	0.76±0.04
		learned	0.95±0.01* (0.003)	0.86±0.02	0.78±0.04* (0.004)
	MMD	normal	1.37±0.13	0.73±0.08	1.47±0.05
		learned	1.09±0.14* (0.003)	0.70±0.11	1.38±0.07

Table 2. Differences in useability (use.) and accuracy (acc.) between Gaussian Process segmentation with and without learned shape priors, where p-values (in brackets) indicate significant improvement for the learned priors.

We measure **accuracy** by the similarity between a segmentation and the ground truth, using one region-based similarity measure, namely the Dice similarity coefficient (DSC) [19], and one boundary-based similarity, the mean minimum distance between contours (MMD) which is equivalent to the modified Hausdorff distance [20]. Tables 1 and 2 show the results for the Langevin and GP tools respectively. All the significant results reveal a positive effect of the shape priors. We also measure the inter- and intra-operator variability by the similarity between the contours created for the same region by one user on two occasions, and by two separate users. The shape priors do not reduce these variabilities, as the level of user control still allows for idiosyncratic segmentation.

6 Conclusions and Future Work

We have designed SSMs for global shape priors in applications lacking correspondence points, and shown them to improve the useability and accuracy of supervised contouring tools. The work shows that probabilistic time series models have a lot more to offer the field of shape modelling than current autoregressive or Markov random field models.

Future work will extend the GP model to use novel kernels specific to different applications, and generalise the models beyond star-shapes using the $\mathbf{Q}_{\text{gen}}$ parametrisation. We propose two extensions to 3-dimensions [21]. The first uses the radial profile model to propogate observations through 2D slices and the second uses the 'spiral' transform as in [22], which turns a 3D surface into a radial signature similar to the time series used above.

References

1. Cootes, T.F., Taylor, C.: Active shape models – smart snakes. In: Proceedings, BMVA Conference on British Machine Vision. (1992) 266–276
2. Cremers, D., Kohlberger, T., Schnörr, C.: Shape statistics in kernel space for variational image segmentation. Pattern Recognition **36** (2003) 1929–1943

3. Hug, J., Brechb uhler, C., Székely, G.: Model-based initialisation for segmentation. In: Proceedings, European Conference on Computer Vision. (2000) 290–306
4. van Ginneken, B., de Bruijne, M., Loog, M., Viergever, M.A.: Interactive shape models. In: Proceedings, SPIE Medical Imaging. (2003) 1206–1216
5. Fan, A.J., Fisher, J., Wells, W., Levitt, J., Willsky, A.: MCMC curve sampling for image segmentation. In: Proceedings, Medical Image Computing and Computer-Assisted Intervention. (2007)
6. Kashyap, R.L., Chellappa, R.: Stochastic models for closed boundary analysis: Representation and reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (1981) 627–637
7. Kartikeyan, B., Sarkar, A.: Shape description by time series. IEEE Transactions on Pattern Analysis and Machine Intelligence **11** (1989) 977–984
8. He, Y., Kundu, A.: 2-D shape classification using hidden Markov model. IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991) 1172–1184
9. Friedland, N.S., Rosenfeld, A.: Compact object recognition using energy-function-based optimisation. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992)
10. Mir, A.H., Hanmandlu, M., Tandon, S.N.: Description of shapes in CT images. IEEE Engineering in Medicine and Biology **18** (1999) 79–84
11. Friedland, N.S., Adam, D.: Ventricular cavity boundary detection from sequential ultrasound images using simulated annealing. IEEE Transactions on Medical Imaging **8** (1989)
12. Kauppinen, H., Seppanen, T., Pietikainen, M.: An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. IEEE Transactions on Pattern Analysis and Machine Intelligence **17** (1995)
13. Itô, K.: On stochastic differenial equations in a differentiable manifold. Nagoya Mathematical Journal **1** (1950)
14. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press (2006)
15. Friedrich, R., Peinke, J.: Description of a turbulent cascade by a Fokker-Planck equation. Physical Review Letters **78** (1997) 863
16. Pérez, P., Blake, A., Gangnet, M.: Jetstream: Probabilistic contour extraction with particles. In: Proceedings, IEEE International Conference on Computer Vision. (2001) 524–531
17. Heimann, T., van Ginneken, B.: 3D liver tumor segmentation challenge 2008, part of *3D segmentation in the clinic: A Grand Challenge II*, at 11^{th} international conference on Medical Image Computing and Computer Assisted Intervention, New York (2008)
18. Huttenlocher, D., Klanderman, G., Rucklidge, W.: Comparing images using the Haussdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence **15** (1993) 850–863
19. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26** (1945) 97//302
20. Dubuisson, M., Jain, A.K.: A modified Hausdorff distance for object matching. In: Proceedings, International Conference on Pattern Recognition. (1994) 566–568
21. T.Shepherd: Dynamical Models and Machine Learning for Supervised Segmentation. PhD thesis, University College London (2009)
22. Wang, J., Engelmann, R.: Segmentation of pulmonary nodules in three-dimensional ct images by use of a spiral-scanning technique. Medical Physics **34** (2007) 4678–4689

3-D Segmentation of Knee Bones on MR images by Constrained Branch-and-Mincut

Soochan Lee[1], Hackjoon Shim[1,2], Sang Hyun Park[1],
Il Dong Yun[3], and Sang Uk Lee[1]

[1] School of Electrical Eng. and Computer Sci., Seoul Nat'l Univ., Korea,
redhouse@cvl.snu.ac.kr,shpark@cvl.snu.ac.kr,sanguk@ipl.snu.ac.kr
[2] BK21 Research Division for Information Technology, Seoul Nat'l Univ., Korea,
hjshim@diehard.snu.ac.kr
[3] School of Electronic and Information Eng., Hankuk University of Foreign Studies,
Korea, yun@hufs.ac.kr ⋆

Abstract. In this paper, we propose a fully automatic method to segment three bone compartments in MR images of a knee joint based on previous segmentation results as a priori information. The proposed method utilizes the branch-and-mincut technique which unifies graph-cut for optimization and branch-and-bound tree search. In spite of its efficiency, branch-and-mincut is not readily applicable to high resolution MR volume images, due to computation complexity and limitations in storage space. To alleviate these problems, we propose an efficient framework where we first simultaneously obtain an initial segmentation of MR images and a coarse rigid registration of the mean shape template by constrained branch-and-mincut, and then refine the result by determining the optimal shape prior among the database. Our main contribution is the constrained branch-and-mincut technique, which by utilizing not only the bound but also the segmentation results at each mincut computation dramatically enhances the efficiency of branch-and-mincut applied to a set of translation based shape priors with little sacrifice in accuracy. In addition, we also exploit intensity priors to boost both efficiency and accuracy. The performance was evaluated cross-sectionally, where eight baseline MR images having pre-segmentation of bones were provided for leave-one-out validation. The experiments demonstrate the effectiveness of the proposed framework.

1 Introduction

Osteoarthritis (OA) is a leading cause for disability and about 40% of adults aged 55 or older have frequent knee pain [1]. Strategies to prevent knee OA are currently very limited, and degradation of articular cartilage is believed as one of its hallmarks [2]. Assessment of cartilage volume/thickness using MR

⋆ This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (20090074888)

images has been shown to provide an accurate measure for OA progression [3]. To derive these measurements, cartilage must be accurately segmented. This task is nontrivial and automation has been shown to be exceedingly difficult [3] due to the complexity of cartilage geometry, small size, and the low contrast with surrounding tissues. On the other hand, knee bones are of considerable size and have consistent intensities. Furthermore, the distinctive contrast at bone-cartilage-interfaces (BCI) can be used as the reference for assessment of changes in thickness of cartilage over time. In this paper, we propose a fully automatic segmentation method of knee bones in MR images provided by the osteoarthritis initiative (OAI, http://www.oai.ucsf.edu). We believe that the proposed method will help segmentation of cartilage and quicken the pace of research on knee OA.

There have been many approaches [4–6] to segment knee bones using various techniques, such as geodesic active contours [4], support vector machine with phase information [5], and interactive graph-cut [6]. Recently, numerous binary segmentation methods have been based on energy optimization where the solution is much less likely to be stuck in local minima. The proposed framework is based on the branch-and-mincut method [7], which integrates graph cuts [8] and branch-and-bound tree search. Graph-cut determines the optimal state of a binary Markov random field (MRF) mainly relying on the gray level uniformity of regions and the continuity of contours. This may suffer from over-segmentation or over-smoothed boundaries. Although high level a priori knowledge would help overcome these limitations, the processes of learning and applying this knowledge inherently suffers from uncertainties. In the literature of similar approaches [9–11], [9] manually provided the position of a known object, [10] approximated both the shape position of the object while learning, and [11] relied on a manually given bounding box and approximated color distributions of the foreground and background. Unlike these methods, branch-and-mincut [7] can efficiently obtain the optimal value of the latent parameters by establishing aggregate potentials to represent this uncertainty.

Although branch-and-mincut efficiently deals with multiple shape priors, required computation greatly increases with the number of priors. Considering the number of shape templates, translations, and rotations, this may grow huge, especially in 3-D. Furthermore, required storage is also much larger, rendering this method inapplicable to 3-D MR images. The proposed method alleviates these problems by partitioning the solution into a two step approach. In the first step, an initial segmentation along with the optimal translation of the mean shape template is obtained by constrained branch-and-mincut. In the second step, the optimal rotation along with the optimal among the set of shape templates is determined and segmentation is refined. Preprocessing is included as an option to handle cases when the intensity levels vary for different MR images. Figure 1 shows a flowchart of the proposed method, while figure 2 show an example of the segmented bone components by our method.

The main contribution of this paper lies in the constrained branch-and-mincut technique, where the efficiency of branch-and-mincut applied to a set of translation based shape priors is greatly improved by utilizing the segmen-

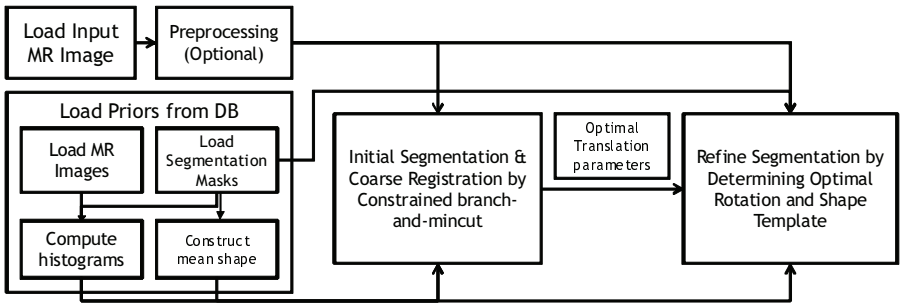
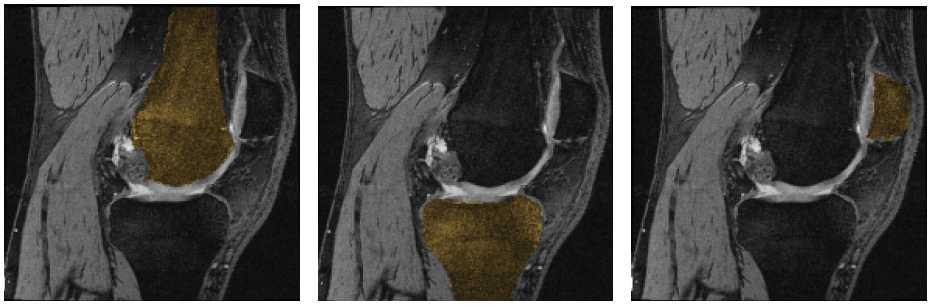


Fig. 1. Flowchart of the proposed method.



(a) (b) (c)

Fig. 2. Examples of segmentation results obtained by the proposed method of (a) femur, (b) tibia, and (c) patella.

tation results at each mincut computation with little sacrifice in accuracy. In addition, by combining shape and intensity priors, we were able to boost both efficiency and accuracy.

2 Method

2.1 Review of Branch-and-Mincut

Many recent methods that deal with binary segmentation [10, 11, 9] have been based on minimizing an energy on an MRF in the form of (1) (following the notation of [7]) via graph cuts [8],

$$E(\mathbf{x}, \omega) = \sum_{p \in \mathcal{V}} F^p(\omega) x_p + \sum_{p \in \mathcal{V}} B^p(\omega)(1 - x_p) + \sum_{p,q \in \mathcal{E}} P^{pq}(\omega)\, |x_p - x_q|, \qquad (1)$$

where $\mathbf{x}$ represents the vector of labels constituted by binary labels x_p for each voxel p in the whole set of voxels $\mathcal{V}$, and $F^p(\omega)$ and $B^p(\omega)$ are foreground and background unary potentials of p, respectively. $P^{pq}(\omega)$ is the pairwise potential

of neighboring voxels pairs $p, q \in \mathcal{E}$, and all potentials are based on the non-local prior ω. Here, ω governs the behavior of the potentials, thereby ultimately controlling the segmentation results.

Though we can determine the global minimum of (1) by the min-cut algorithm [12] when ω is fixed, determining the optimal ω and $\mathbf{x}$ jointly is much more difficult. When ω is altered, potential functions are changed, and the graph is redefined. Thus, finding the optimal ω_{opt} that induces the minimum energy becomes increasingly difficult as the set Ω of non-local priors grows since min-cut for all $\omega \in \Omega$ must be computed.

Branch-and-mincut avoids such brute force computation by defining the lower bound

$$L(\Omega) \leq \min_{x \in 2^{\mathcal{V}}, \omega \in \Omega} E(x, \omega) \tag{2}$$

of the minimum of eq (1) over both the vector of labels $\mathbf{x}$ and non-local prior ω as

$$L\left(\Omega\right) = \min_{x \in 2^{\mathcal{V}}} \left[\sum_{p \in \mathcal{V}} \min_{\omega \in \Omega} F^p(\omega) x_p + \sum_{p \in \mathcal{V}} \min_{\omega \in \Omega} B^p(\omega)(1 - x_p) + \sum_{p,q \in \mathcal{E}} \min_{\omega \in \Omega} P(p,q)(\omega) \left| x_p - x_q \right| \right], \tag{3}$$

by introducing the concept of aggregate potential as $F^p_\Omega = \min_{\omega \in \Omega} F^p(\omega)$, $B^p_\Omega = \min_{\omega \in \Omega} B^p(\omega)$, $P^{pq}_\Omega = \min_{\omega \in \Omega} P(p,q)(\omega)$ which are the minimum values of pixelwise and pairwise potentials over the whole non-local prior set Ω. These terms can be obtained by interchanging the order of computing the minimum. $L(\Omega)$ can likewise be computed by min-cut, and we can efficiently search Ω by excluding subsets with high lower bounds based on branch-and-bound tree search. Here, the child nodes represent a subdivision of the current non-local prior set. The whole process starts with the full set of non-local priors Ω, and at each iteration the the current node with the subset corresponding to the lowest lower bound is branched. The lower bound for the child nodes are successfully computed, and this process is repeated until the current node is comprised of a single shape prior, i.e., the global optimum among Ω. The required number of min-cut computations correspond to the number of traversed nodes, which is typically much smaller than the number of all shape priors.

For segmentation with shape priors, the non-local prior set Ω is comprised of shape templates $\delta \in \Delta$ with specific positions. By defining the unary potentials for each pixel p as the Hamming distance with the template, ($F^p(\omega) = 0$ if p is assigned as foreground and $F^p(\omega) = 1$ if p is background; background potentials defined similarly), the aggregate potentials can be efficiently computed by utilizing translational invariance. Pairwise potentials are defined as

$$P(p,q) = \lambda_2 \left(\frac{e^{-|I_p - I_q|^2/\sigma}}{|p - q|} \right) \tag{4}$$

based on pixel gradients $|I_p - I_q|$. We note that σ is defined as that in [8]. Thus, minimizing (1) on Ω will be equivalent to determining the δ and its position that best coincides with high gradient edges as the optimal.

2.2 Initial Segmentation by Constrained Branch-and-Mincut

Upon analysis of the given problem and the characteristics of the branch-and-mincut method, we have made the following observations. From hereon, we refer to cropped 3-D segmentation masks as shape templates and the positioned shape mask inside the MR image voxel space as shape priors.

For segmenting knee bones from MR images from the OAI: (1) The OAI has a fixed scanning protocol, thereby limiting the rotation and the position of bones in OAI MR images. (2) Bones in the knee joint have limited variation in shape compared to other organs in the body.

For branch-and-mincut: The efficiency of branch-and-mincut is maximized when the overlap between the priors of leaf nodes with the same parent is maximized. Consequently, shape priors with a single shift in the position of the shape template are clustered when constructing the tree for translation. Overlap between different templates is typically much less than the overlap between templates with a single shift, thereby making branch-and-mincut on only the set of translations most efficient. Following [13], each node is hierarchically clustered following the oct-tree structure.

These observations motivated the first step of our framework, where initial segmentation of the input MR image and coarse registration of a single mean shape template is done by branch-and-mincut where the set of shape priors Ω are comprised by translated templates ignoring rotation and scale variations. Here, the mean shape template is constructed from the whole set of shape templates by first aligning the templates and creating a union volume, where each voxel is assigned as filled if that voxel is filled in more than half of the templates.

Direct application of branch-and-mincut is very inefficient in our specific case, and the ratio of nodes traversed was very high. Since the possible translation interval is relatively small compared to the size of bones in the given MR images, the difference between each shape prior becomes small and the problem of determining the optimal translation becomes more ambiguous. Nonetheless, the node with the set containing the optimal translation is repeatedly branched early on in the tree search. Therefore, by excluding nodes with subsets that are unlikely to contain the optimal translation early on, we can prune the set of candidate priors more effectively. We also noticed that only the max-flow values were utilized in the branch-and-mincut process, while the specific segmentation results were completely overlooked. Therefore, we develop a very simple yet highly effective method to constrain branching of nodes depending on the segmentation results obtained at each particular node.

Our method starts from the observation that since the shape template is invariant to translations, so too are the aggregate potentials for different intervals with equal length, thus making the total sum of both foreground and background unary potentials equal. This is shown in figure 3.

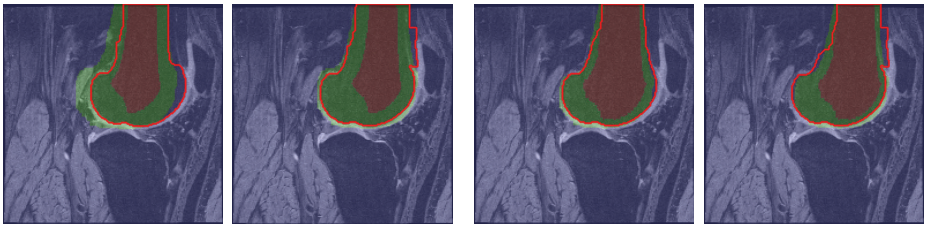
(a) (b)

Fig. 3. (a) Left: two shape priors (represented in yellow and blue, respectively) based on different translations of the same template. Right: the aggregate potentials of shape priors within the interval formed by the two translated priors shown left. (b) Same as (a) but translated rightward. The shape of the aggregate potentials are translation invariant.

Therefore, the difference between the unary energy $E_{unary}(\mathbf{x}, \omega)$ (sum of the first and second terms of the RHS of (3)) after min-cut computation for different ω occurs purely from the difference in the number of voxels that have been assigned contrary labels to the labels in ω. This occurs when the increase in $E_{unary}(\mathbf{x}, \omega)$ due to the contrary assignment is smaller than the decrease in the sum of pairwise potential $E_{pairwise}(\mathbf{x}, \omega)$ (the third term of the RHS of (3)), thereby reducing total energy. Therefore, we assume that if $E_{unary}(\mathbf{x}, \Omega_1)$ is larger than $E_{unary}(\mathbf{x}, \Omega_2)$, $E_{pairwise}(\mathbf{x}, \Omega_1)$ must be smaller than $E_{pairwise}(\mathbf{x}, \Omega_2)$ in order for the two subsets $\Omega_1, \Omega_2 \subset \Omega$ to be comparable. Otherwise, the boundaries found for Ω_1 and its pairwise energy is worse than pairwise energy even with the additional sacrifice made in unary energy, making it highly unlikely for Ω_1 to contain the optimal prior. This is shown in figure 4. To make our argument more concrete, we compare the normalized pairwise potential $\overline{E}_{pairwise} = \frac{E_{pairwise}}{\partial B}$ since $E_{pairwise}$ depends on the area of the boundary surface ∂B. Thus, at each branching step, we compute $E_{unary}(\mathbf{x}, \Omega_i)$ and $\overline{E}_{pairwise}(\mathbf{x}, \Omega_i)$, and compare them with those of the prior sets with equal translation intervals currently added to the active front of nodes. If both values are the largest, that node is not included in the active front, thereby excluding the subtree based on that node. This process is very simple to implement and requires very little additional computation or storage space.

2.3 Refinement by Rigid Registration of Shape Templates

Since only a single mean shape template was applied in the initial segmentation step by constrained branch-and-mincut, a supplementary refinement step is necessary to account for the whole set of shape templates separately. Although applying the original unconstrained branch-and-mincut to a small translation and rotation range may seem like a good way to do this, it is actually quite a bad idea since the efficiency of branch-and-mincut decreases rapidly as the optimization problem becomes more ambiguous. Thus, branch-and-mincut is not suitable for comparing between shape priors with limited variation.



(a) (b)

Fig. 4. Aggregate potentials (red: $F^p_\Omega = 0$, $B^p_\Omega = c$; green: $F^p_\Omega = 0$, $B^p_\Omega = 0$; blue: $F^p_\Omega = c$, $B^p_\Omega = 0$) and segmentation results (red contour) overlapped to the MR image. Left and right aggregate potentials are based on translation intervals of equal length for both (a) and (b). (a) both the unary energy and normalized pairwise energy of the left is larger than the right and therefore the subtree of left translation interval is pruned. (b) Although the unary energy of the left is smaller than that of the right, the normalized pairwise energy of the right is smaller than that of the left due to the large gradient in the right part of the segment boundary. Thus both nodes are branched.

To efficiently determine the optimal template among the whole set while refining segmentation results, we directly register each shape template δ in the template set Δ via iterative closest points (ICP) [14] to the segmentation results previously obtained by constrained branch-and-mincut. We then separately apply min-cut for each template with the registered template as the shape prior to compute the minimum energy. This process is iterated for each δ until the energy does not decrease, and the segmentation result as well as the corresponding optimal template is selected as the one with the minimum energy. This process is represented in figure 5.

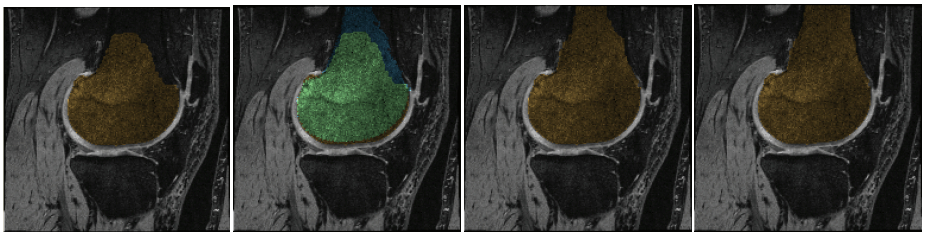
2.4 Combined Shape and Intensity Prior Based Potentials

To take advantage of all available information in our database, we augment the Hamming distance unary potentials based on shape priors to include intensity information. Thus, foreground and background potentials are defined respectively as:

$$F^p(w) = -\lambda_1 \ln\left(\frac{P(I_p|F)}{P(I_p|F) + P(I_p|B)}\right) \times (cy^\omega_p + 1), \tag{5}$$

$$B^p(w) = -\lambda_1 \ln\left(\frac{P(I_p|B)}{P(I_p|F) + P(I_p|B)}\right) \times (c(1 - y^\omega_p) + 1), \tag{6}$$

The *a posteriori* probabilities $P(I_p|F)$ and $P(I_p|B)$ are based on all of the original segmentation results corresponding to the templates of Ω and their intensities in the MR images in our database, and y^ω_p is the Hamming distance based on shape prior ω described in section 2.2. We expressed either potential of $F^p(w)$ and $B^p(w)$ as a product of the intensity and shape prior terms and introduced an additional term c which controls the relative effect of shape to intensity. The $+1$ term is added so that the intensity priors are symmetrically enforced. ωs depend



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

Fig. 5. Illustration of the refinement process. (a) The segmentation result obtained from constrained branch-and-mincut with a single mean shape template overlapped to the corresponding MR image. Yellow voxels represent foreground, i.e., voxels labeled as bone. (b) Registered template overlapped on (a), where cyan voxels represent bone voxels in the registered template. (c) Segmentation results obtained from min-cut with the registered template of (b) as the shape prior. (d) Final optimal segmentation results obtained by iterating (b) and (c).

only on shape templates and rigid motion parameters, and intensity priors are fixed.

3 Experimental Results and Discussion

3.1 Description of Dataset and Quantitative Analysis

Eight 3-D sagittal double-echo and steady-state MR images with 384x384x160 voxels, and 0.36x0.36x0.70 mm3 voxel resolution were obtained from the OAI database. Ground truth segmentations were obtained by a semi-automatic method [15] in full resolution and experiments for the proposed method were conducted on images which were downsampled by a factor of 2 for efficiency. The femur, tibia, and patella were each segmented with a separate pass of the proposed method.

The performance of our method was evaluated by cross-sectional cases. Eight baseline MR images having pre-segmentation of bones were provided for leave-one-out validation.

The segmentation results are measured by the Dice Similarity Coefficient (DSC), defined as the ratio of the overlapping volume to the mean of the two volumes. As the measure of computational efficiency, we used the number of min-cut computations, since this is the the core of the computations in branch-and-mincut. We note that we applied the method of [12] and its implementation offered in V. Kolmogorov's web site [16], and all experiments were run on a 2.40GHz core2 Quad CPU with 2GB RAM, and $c = 8$, $\lambda_1 = 30$, and $\lambda_2 = 1000$.

3.2 Evaluation of Constrained Branch-and-Mincut

We first evaluate the effectiveness of constrained branch-and-mincut (CBNM) in terms of efficiency and accuracy by comparing the results to original un-

Table 1. Comparison between constrained branch-and-mincut (CBNM) and branch-and-mincut (BNM). MC: number of min-cut computations; $time$: computation time (sec); $\mathbf{t}$: optimal translation vector; E: relative total energy, the global minimum computed by original branch-and-mincut is set to 100; SP: segmentation performance (DSC measure)

case	MC		$time$		$\mathbf{t}$		E		SP	
	CBNM	BNM	CBNM	BNM	CBNM	BNM	CBNM	BNM	CBNM	BNM
1	333	649	233.2	361.3	(67,-4,13)	(67,-4,13)	100	100	0.976	0.976
2	675	1979	303.2	950.3	(62,6,8)	(61,6,9)	100.1	100	0.974	0.970
3	1061	6163	1131.4	3101.9	(63,10,10)	(64,8,10)	100.2	100	0.817	0.824
4	485	1227	271.8	537.6	(54,5,9)	(54,5,9)	100	100	0.978	0.978
5	245	887	131.5	317.9	(51,-2,9)	(51,-1,9)	100.0	100	0.953	0.956
6	137	1111	92.8	410.4	(62,-5,8)	(62,-5,8)	100	100	0.912	0.912
7	619	2109	355.8	1035.1	(50,7,11)	(50,7,11)	100	100	0.958	0.958
8	265	829	157.7	330.3	(53,-4,5)	(52,-4,5)	100.0	100	0.885	0.883

constrained branch-and-mincut (BNM). Table 1 presents the number of min-cut computations MC, computation time in seconds, obtained final translation $\mathbf{t}$, computed total energy E, and segmentation performance SP for both constrained and unconstrained branch-and-mincut applied to the dataset with a shape prior set comprised of $40 \times 40 \times 12$ possible translations for a single mean shape template of the femur. We note that for each case, the shape template database applied was constructed in a leave one out fashion and templates were aligned by translation when constructing the mean template. Rigid alignment via PCA was also tested, but many cases resulted in overfitting, and alignment by translation sufficed due to the fixed protocol of the OAI. The unary potentials were computed based on (6) and (5) for both cases.

MC and $time$ represent the efficiency while $\mathbf{t}$, E, and SP represent the accuracy of each method. Since branch-and-mincut determines the global optimal energy and its corresponding shape prior, $\mathbf{t}$, E, and SP for BNM are globally optimal.

In terms of efficiency, our method reduced the number of min-cut computations to an average 30.4% of that of the unconstrained branch-and-mincut while reducing the computation time to 37.2% on average. In terms of accuracy, we can see that in half of the cases the constrained method obtained the global minimum values, while for the rest some degradation in accuracy occurred. Nonetheless, this decrease in accuracy was very small relative to the optimum, supporting the performance of our constrained method. Especially, by observing the obtained translation positions, we can see that the suboptimal results obtained by our method differs to the optimal by only one or two voxels.

We note that the proposed method enhances efficiency not only in terms of computation, but also in terms of memory. In [7] rotated and scaled templates are assumed as separate templates, and only translated templates are considered invariant. Thus memory for rotated and scaled templates, up to a number of $n_r x \times n_r y \times n_r z \times n_s$, is needed, which is excessive even for advanced systems.

Table 2. Segmentation performance for femur, tibia and patella by the proposed CBNM with a single mean shape template and combined intensity and shape potentials.

case	1	2	3	4	5	6	7	8
femur	0.976	0.974	0.817	0.978	0.953	0.912	0.958	0.883
tibia	0.976	0.983	0.970	0.982	0.943	0.908	0.985	0.889
patella	0	0.844	0.674	0.960	0.761	0.969	0.559	0.961

Table 3. Segmentation performance for femur, tibia and patella obtained by adding preprocessing (applying histogram matching to the input MR image) and condition to avoid overlap in bones to the method of table 2.

case	1	2	3	4	5	6	7	8
femur	0.967	0.979	0.939	0.972	0.946	0.926	0.938	0.897
tibia	0.977	0.984	0.972	0.972	0.970	0.903	0.984	0.926
patella	0.966	0.847	0.679	0.940	0.760	0.965	0.938	0.966

By utilizing a mean shape template, required memory is greatly reduced to that storing a single template.

As for the atypically inferior segmentation performance #3 and #8, we found that this was due to differences in intensity levels, where the #3 MR image was unusually brighter than others while #8 was darker. This motivated us to include a preprocessing step, which will be described in the next section.

We present the segmentation performance of constrained branch-and-mincut with a single mean shape template and combined intensity and shape priors (identical to that in table 1 for femur) in table 2. We can see that for tibia our method gives adequate results, but lacks some accuracy for patella. Although this is somewhat inevitable due to its small size and higher variability in shape, the 0 result for case #1 was alarming. We found that the segmentation results obtained were actually part of the femur, and we thus included a condition that segmentation be done in the order of bone size, i.e., femur, then tibia, and then patella, and that regions where bones were previously identified are avoided. The results by adding this condition (along with preprocessing described in the subsequent subsection) is given in table 3.

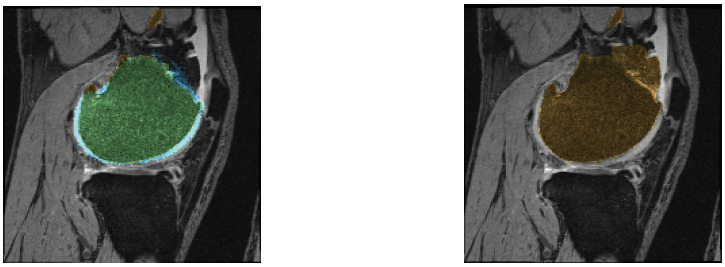
3.3 Preprocessing and Refinement

As mentioned in the previous section, we included a preprocessing step to consider differences in intensity level among different MR images. Specifically, we applied histogram matching to the input MR image, with the mean histogram computed from the MR images in the database as the target. Table 3 represents the performance when preprocessing is included.

We can see that although for some cases the performance decreased slightly, the segmentation performance increased for cases including #3 and #8. The performance especially increased dramatically for the femur of #3, and the patella of #7 thereby supporting the overall utility of the preprocessing step.

Table 4. Segmentation performance for femur, tibia and patella by the proposed framework of preprocessing, constrained branch-and-mincut, and refinement based on multiple shape templates.

case	1	2	3	4	5	6	7	8
femur	0.950	0.971	0.973	0.957	0.954	0.931	0.943	0.897
tibia	0.977	0.981	0.974	0.973	0.984	0.941	0.980	0.968
patella	0.965	0.846	0.688	0.954	0.764	0.969	0.951	0.978



(a) (b)

Fig. 6. Illustration of when refinement fails. (a) The yellow voxels are labeled as foreground from the constrained branch-and-mincut step, and the cyan voxels are of the registered shape template. (b) The result of applying min-cut with the shape template corresponding to cyan voxels in (a) as the shape prior. Due to the high gradient in the upper right side, the energy of result (b) is lower than the CBNM results of (a).

The results after refinement, i.e., the results of the full proposed framework, are presented in table 4. The effect of refinement can be analyzed by comparing the results of table 4 with table 3. The results deteriorated in cases #1, #2, #4 for femur, cases #2 and #7 for tibia, and in cases #1, #2 for patella, while results greatly increased in case #3 for femur, and cases #6 and #8 for tibia. Figure 6 presents how refinement fails by showing the segmented femur after refinement for case #4. Specifically, due to regions with high gradient not actually on the bone boundary, the segmentation with minimum energy does not correctly represent the true segmentation. This can also be seen in cases #2 and #8 of table 1 where despite the higher energy of the constrained branch-and-mincut results, the performance was higher. This is related to how much emphasis is given on matching boundaries to high gradients, and warrants future research. As for the cases where results greatly improve, case #3 for femur shown in figure 5 is a good example.

4 Conclusion

In this paper, a framework based on a new constrained branch-and-mincut method which improves efficiency of both computational complexity and storage space with little sacrifice in accuracy compared to original branch-and-mincut was proposed. The framework enables automatic and efficient segmentation of

bones in high resolution 3-D MR images of knee joints used in diagnoses of osteoarthritis. Evaluation of the proposed method was done on eight baseline MR images having pre-segmentation of bones in a leave-one-out fashion, and experimental results showed that computation was on average only 30.4% of that of original branch-and-mincut considering translation parameters only. Since our method deals with multiple shape templates in a separate refinement step, this ratio will most likely decline even further when multiple templates are applied to original branch-and-mincut. Furthermore, by registering shape templates and further comparing corresponding segmentation energy, the proposed method successfully determined the optimal template and improved robustness of the segmentation performance. We hope this method will help segmentation of cartilage or reviewing of cartilage thickness changes over time, which is the direction for our further research.

References

1. Felson, D., Zhang, Y.: An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. Arthritis Rheum **41**(8) (1998) 1343–55
2. Felson, D., et al.: Osteoarthritis: new insights. part 2: treatment approaches. Ann Intern Med **133**(9) (2000) 726–37
3. Eckstein, F., et al.: Magnetic resonance imaging (mri) of articular cartilage in knee osteoarthritis (oa): morphological assessment. Osteoarthritis Cartilage **14**(9) (2006) Suppl A:A46–75
4. Lorigo, L., et al.: Segmentation of bone in clinical knee mri using texture-based geodesic active contours. In: MICCAI. (1998) 1195–1204
5. Bourgeat, P., et al.: Mr image segmentation of the knee bone using phase information. Medical Image Analysis **11**(4) (2007) 325–335
6. Shim, H., et al.: 3-d segmentation of articular cartilages by graph cuts using knee mr images from osteoarthritis initiative. In: SPIE Medical Imaging. (2008)
7. Lempitsky, V., et al.: Image segmentation by branch-and-mincut. In: ECCV. (2008) IV: 15–29
8. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient n-d image segmentation. IJCV **70(2)** (Feb. 2006) 109–131
9. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: CVPR. (2005) I: 755–762
10. Kumar, M., et al.: Obj cut. In: CVPR. (2005) I: 18–25
11. Rother, C., et al.: Grabcut: interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**(3) (August 2004) 309–314
12. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE trans. PAMI **26**(9) (September 2004) 1124–1137
13. Lempitsky, V., et al.: Image segmentation by branch-and-mincut. In: Microsoft Technical Report. (2008) MSR–TR–2008–100
14. Besl, P., McKay, N.: A method of registration of 3-d shapes. IEEE Trans. PAMI **14(2)** (1992) 239–256
15. Shim, H., et al.: Simultaneous 3-d segmentation of three bone compartments on high resolution knee mr images from osteoarthritis initiative (oai) using graph-cuts. In: Proceedings of SPIE Medical Imaging. (2009)
16. http://www.cs.ucl.ac.uk/staff/V.Kolmogorov/software.html

Weighted Iterated Conditional Modes for Random Fields: Application to Prostate Cancer Detection

James Monaco, Satish Viswanath, and Anant Madabhushi

Department of Biomedical Engineering, Rutgers University, USA.
{jpmonaco,anantm}@rci.rutgers.edu

Abstract. Random fields (RFs) provide a valuable means for modeling large numbers of dependent random variables. However, estimating RFs is nontrivial, requiring sophisticated techniques such as stochastic relaxation. Unfortunately, these techniques provide no means for adjusting classifier performance (e.g. sensitivity/specificity). Instead, they produce a single, hard classification at a static operating point, thus precluding, for example, the construction of receiver operator characteristic (ROC) curves. Addressing this deficiency, we introduce weighted maximum *a posteriori* (WMAP) estimation, a generalization of MAP estimation that allows certain classes to be weighted more heavily than others. We also introduce weighted iterated conditional modes (WICM), a novel adaptation of ICM capable of WMAP estimation on RFs. We demonstrate the use of WICM by integrating it into two separate Markov random field (MRF) based classification systems capable of detecting prostate cancer (CaP) in 1) whole-mount histological sections and 2) multi-protocol MRI. Specifically, we show how WICM can be used to arbitrarily adjust the CaP detection sensitivity of these systems, yielding ROC curves.

1 Introduction

Many estimation tasks require the ability to classify multiple objects simultaneously. For example, these objects could be calcifications in a mammogram or the pixels of a magnetic resonance image. Within a Bayesian framework each object is modeled as a random variable, and the collection of these random variables is called a random field (RF)[1]. If the random variables are assumed independent, we can estimate each in isolation. This estimation typically involves an exhaustive search. For example, obtaining the maximum *a posteriori* (MAP) estimate (for a single random variable) entails calculating the *a posteriori* probability for each possible class, and then choosing the class with the largest probability. However, if the random variables are not independent, the entire RF must be estimated collectively. Since the number of possible states of the random field

[1] Though a Markov random field implies that the the inter-variable dependencies are restricted to local neighborhoods, a Markov random field and a random field are mathematically equivalent.

is prohibitively large, an exhaustive approach is untenable[2]. Consequently, more sophisticated schemes, such as iterated conditional modes [1] (ICM), simulated annealing [2], or maximum posterior marginals [3], become necessary. These schemes employ iterative techniques to converge to a single, hard labeling.

The capability of adjusting classifier performance (e.g. sensitivity/specificity) with respect to specific classes is essential for many applications. For example, the performance of commercial systems for detecting mammographic abnormalities is typically adjusted to the highest detection sensitivity that incurs no more than two false positive per image. In situations amenable to exhaustive searches, means for modifying performance, such as thresholding or weighting the *a posteriori* probabilities, are well established [4]. Unfortunately, analogous methods compatible with the techniques required for classifying RFs have yet to be proposed. Consequently, most RF-based classification systems restrict their performance to a single, static operating point. To our knowledge, the only previous attempts [5,6] to adjust performance have resulted in *ad hoc* schemes that leveraged the peculiarities of the ICM algorithm: Since ICM converges to a local maximum (mode) of the *a posteriori* probability, varying the initial conditions (i.e. the initial state of the RF) can vary the classification results. However, there is no reason to believe that the modes of the *a posteriori* probability are associated with meaningful classifications; and thus, such techniques are heuristic, lacking any mathematical justification.

In this paper we introduce the first mathematically sound means for adjusting the performance of an estimator compatible with RFs. Specifically, we introduce weighted MAP (WMAP) estimation, a generalization of MAP estimation that allows certain classes to be weighted more heavily than others. Furthermore, we introduce weighted ICM (WICM), a novel adaptation of ICM capable of WMAP estimation on RFs. We demonstrate their applicability in two separate classification systems based on Markov random fields (MRFs): 1) for detecting prostate cancer (CaP) in whole-mount histological sections (WMHSs) [7] and 2) for detecting CaP in multi-protocol (T2-weighted and dynamic-contrast enhanced) 3 Tesla *in vivo* magnetic resonance imaging (MRI) [8]. Within this context, we illustrate how WICM can be used to vary classification performance, enabling the construction of receiver operator characteristic (ROC) curves.

The remainder of the paper is organized as follows: Section 2 reviews the necessary nomenclature and introduces the WMAP formulation. In Section 3 we derive WICM. Section 4 provides an evaluation of WICM in the context of our CaP detection systems. Section 5 offers concluding remarks.

2 Weighted Maximum *a Posteriori* Estimation for MRFs

2.1 Markov Random Field Nomenclature and Assumptions

Let the set $S=\{1,2,\ldots,N\}$ reference N sites to be classified. Each site $s\in S$ has two associated random variables: $X_s\in\Lambda\equiv\{\omega_1,\omega_2,\ldots,\omega_L\}$ indicating its state

[2] If a random field contains N random variables, each of which can assume one of L classes, the total number of possible states is L^N.

(class) and $Y_s \in \mathbb{R}^D$ representing its D-dimensional feature vector. Particular instances of X_s and Y_s are denoted by the lowercase variables $x_s \in \Lambda$ and $y_s \in \mathbb{R}^D$. Let $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ refer to all random variables X_s and Y_s in aggregate. The state spaces of $\mathbf{X}$ and $\mathbf{Y}$ are the Cartesian products $\Omega = \Lambda^N$ and $\mathbb{R}^{D \times N}$. Instances of $\mathbf{X}$ and $\mathbf{Y}$ are denoted by the lowercase variables $\mathbf{x} = (x_1, x_2, \ldots, x_N) \in \Omega$ and $\mathbf{y} = (y_1, y_2, \ldots, y_N) \in \mathbb{R}^{D \times N}$. Let $G = \{S, E\}$ establish an undirected graph structure on the sites, where S and E are the vertices (sites) and edges, respectively. A neighborhood η_s is the set containing all sites that share an edge with s, i.e. $\eta_s = \{r : r \in S, r \neq s, \{r, s\} \in E\}$. If P is a probability measure defined over Ω then the triplet (G, Ω, P) is called a random field.

The conclusions in this paper are predicated on two assumptions. First, the states $\mathbf{X}$ are assumed to constitute a Markov random field (MRF). That is, $\mathbf{X}$ is a random field whose local conditional probability functions satisfy the Markov property: $P\left(x_s \middle| \mathbf{x}_{-s}\right) = P\left(x_s \middle| \mathbf{x}_{\eta_s}\right)$, where $\mathbf{x}_{-s} = (x_1, \ldots, x_{s-1}, x_{s+1}, \ldots, x_N)$, $\mathbf{x}_{\eta_s} = \left(x_{\eta_s(1)}, \ldots, x_{\eta_s(|\eta_s|)}\right)$, and $\eta_s(i) \in S$ is the i^{th} element of the set η_s. Second, we assume that each feature vector Y_s is conditionally independent and identically distributed given its associated state X_s: $P\left(\mathbf{y} \middle| \mathbf{x}\right) = \prod_{s \in S} P\left(y_s \middle| x_s\right) = \prod_{s \in S} f\left(y_s \middle| x_s\right)$, where the use of the single probability density function f indicates that $P\left(y_s \middle| x_s\right)$ is identically distributed across S.

2.2 Weighted Maximum *a Posteriori* Estimation

Given an observation of the feature vectors $\mathbf{Y}$, we would like to estimate the states $\mathbf{X}$. Bayes rule advocates selecting the estimate $\widehat{\mathbf{x}} \in \Omega$ that minimizes the conditional risk [4]

$$R\left(\mathbf{X} \middle| \widehat{\mathbf{x}}, \mathbf{y}\right) = E\left[\mathbf{C}\left(\mathbf{X}, \widehat{\mathbf{x}}\right) \middle| \mathbf{y}\right] = \sum_{\mathbf{x} \in \Omega} \mathbf{C}\left(\mathbf{x}, \widehat{\mathbf{x}}\right) P\left(\mathbf{x} \middle| \mathbf{y}\right), \tag{1}$$

where E indicates expected value and $\mathbf{C}\left(\mathbf{x}, \widehat{\mathbf{x}}\right)$ is the cost of selecting labels $\widehat{\mathbf{x}}$ when the true labels are $\mathbf{x}$. For most classification tasks it is sufficient to assume that the cost of mislabeling any individual site is 1) independent of the remaining sites and 2) identical for every site. This implies the following: $\mathbf{C}\left(\mathbf{x}, \widehat{\mathbf{x}}\right) = \prod_{s \in S} C(x_s, \widehat{x}_s)$, where $C(x_s, \widehat{x}_s)$ is the cost of selecting label $\widehat{x}_s$ when the true label is x_s.

The most prevalent cost function for MRFs (though this cost is rarely expressed explicitly) is $\mathbf{C}_{MAP}\left(\mathbf{x}, \widehat{\mathbf{x}}\right) = 1 - \prod_{s \in S} \delta\left(x_s - \widehat{x}_s\right)$, where δ is the Kronecker delta. That is, mislabeling any of the sites results in an identical cost of 1. To allow certain decisions to be weighted more heavily than others $\mathbf{C}_{MAP}$ can be generalized as follows: $\mathbf{C}_{WMAP}\left(\mathbf{x}, \widehat{\mathbf{x}}\right) = \left[\prod_{s \in S} \alpha\left(x_s\right)\right]\left[1 - \prod_{s \in S} \delta\left(x_s - \widehat{x}_s\right)\right]$. In this case mislabeling a system of sites whose true labels are $\mathbf{x}$ has an associated cost of $\prod_{s \in S} \alpha\left(x_s\right)$.

The connection between cost functions and estimation becomes clear when $\mathbf{C}_{WMAP}$ is inserted into (1)

$$R_{WMAP}\left(\mathbf{X} \middle| \widehat{\mathbf{x}}, \mathbf{y}\right) = \sum_{\mathbf{x} \in \Omega} \left\{ P\left(\mathbf{x} \middle| \mathbf{y}\right) \left[\prod_{s \in S} \alpha\left(x_s\right)\right] \left[1 - \prod_{s \in S} \delta\left(x_s - \widehat{x}_s\right)\right] \right\}$$

$$= \sum_{\mathbf{x}\in\Omega} \left[P\big(\mathbf{x}\big|\mathbf{y}\big) \prod_{s\in S} \alpha\left(x_s\right) \right] - P\big(\widehat{\mathbf{x}}\big|\mathbf{y}\big) \prod_{s\in S} \alpha\left(\widehat{x}_s\right). \qquad (2)$$

Since the first term in (2) is not a function of $\widehat{\mathbf{x}}$, minimizing (2) over $\widehat{\mathbf{x}}$ is equivalent to maximizing the second term. For convenience we will refer to this term as the conditional benefit $B\left(\mathbf{x}|\mathbf{y}\right)$, i.e

$$B\left(\mathbf{x}|\mathbf{y}\right) = P\big(\mathbf{x}\big|\mathbf{y}\big) \prod_{s\in S} \alpha\left(x_s\right). \qquad (3)$$

(Since it will no longer be necessary to differentiate the true labels from their estimates, the notation $\widehat{\mathbf{x}}$ is henceforth dropped.) Note that if $\alpha(\cdot) \equiv 1$ then the conditional benefit reduces to the *a posteriori* probability, and the minimization of risk becomes maximum *a posteriori* (MAP) estimation. In this context maximizing the weighted *a posteriori* probability in (3) over $\mathbf{x} \in \Omega$ can be seen as weighted MAP (WMAP) estimation, where $\alpha(x_s)$ are the weights.

3 Weighted Iterated Conditional Modes

In [1] Besag introduced iterated conditional modes (ICM) as a means for maximizing $P\big(\mathbf{x}\big|\mathbf{y}\big)$. By adapting ICM we can introduce a method for maximizing the weighted *a posteriori* probability in (3). We begin the derivation of weighted ICM (WICM) by reformulating $B\left(\mathbf{x}\big|\mathbf{y}\right)$ as follows:

$$\begin{aligned} B\left(\mathbf{x}\big|\mathbf{y}\right) = P\big(\mathbf{x}\big|\mathbf{y}\big) \prod_{s\in S} \alpha\left(x_s\right) &= \alpha\left(x_s\right) P\big(x_s\big|\mathbf{x}_{\text{-}s}, \mathbf{y}\big)\, P\big(\mathbf{x}_{\text{-}s}\big|\mathbf{y}\big) \prod_{r\in S, r\neq s} \alpha\left(x_r\right) \\ &\propto \alpha\left(x_s\right) P\big(x_s\big|\mathbf{x}_{\eta_s}, y_s\big)\, P\big(\mathbf{x}_{\text{-}s}\big|\mathbf{y}\big) \prod_{r\in S, r\neq s} \alpha\left(x_r\right) \quad (4) \end{aligned}$$

The final step follows from the two assumptions of Section 2.1. Increasing the first two terms of (4) necessarily increases $B\big(\mathbf{x}\big|\mathbf{y}\big)$. This suggests a global optimization strategy that sequentially visits each site $s \in S$ and determines the label $x_s \in \Lambda$ that maximizes $\alpha\left(x_s\right) P\big(x_s\big|\mathbf{x}_{\eta_s}, y_s\big)$. The maximization of this quantity is straightforward since $P\big(x_s\big|\mathbf{x}_{\eta_s}, y_s\big) \propto P\big(y_s|x_s\big)\, P\big(x_s|x_{\eta_s}\big)$. Note that WICM converges to a local maximum of $B\big(\mathbf{x}\big|\mathbf{y}\big)$. The WICM algorithm is as follows:

Weighted Iterated Conditional Modes
Input: Initial labeling $\mathbf{x}^0$, weights $\alpha(\omega)$
Output: Final labeling $\mathbf{x}^k$ after iteration k
1. $k = 0$
2. **do**
3. $\quad k = k + 1$
4. $\quad \mathbf{x}^k = \mathbf{x}^{k-1}$
5. $\quad$ **for** $\forall\ s \in S$ **do**
6. $\quad\quad x_s^k = \arg\max_{\omega\in\Lambda} \left[\alpha(\omega) P\big(\omega|\mathbf{x}_{\eta_s}^k, y_s\big)\right]$

7. **end for**
8. **while** $\mathbf{x}^k \neq \mathbf{x}^{k-1}$

It is illuminating to consider the case of binary classes, i.e. $\Lambda = \{\omega_1, \omega_2\}$. In this situation step 6 in the above algorithm simplifies to the following:

$$x_s^k = \begin{cases} \omega_1 \text{ if } P\big(\omega_1 | \mathbf{x}_{\eta_s}^k, y_s\big) > \frac{\alpha(\omega_2)}{\alpha(\omega_1)+\alpha(\omega_2)} = T_{wicm} \\ \omega_2 \text{ otherwise.} \end{cases} \tag{5}$$

Consequently, increasing $T_{wicm} \in [0, 1]$ results in a greater preference for ω_2, while decreasing T_{wicm} increases the tendency toward ω_1. Note that when $T_{wicm} = 1/2$, WICM is equivalent to ICM.

4 Experimental Results and Discussion

In this section we evaluate WICM by incorporating it into two separate MRF-based classification systems for detecting prostate cancer (CaP). The goal of both systems is to classify their respective sites S (specifically glands or pixels) into one of two classes: $X_s \in \Lambda \equiv \{\omega_1, \omega_2\}$, where ω_1 and ω_2 indicate malignancy and benignity. The basic procedure for both systems is similar: 1) Using the distribution $f(y_s|x_s)$, a Bayesian classifier assigns each site s a probability of malignancy $P(x_s|y_s)$ based solely on its feature vector y_s. 2) If this probability exceeds the threshold T_f, the state x_s of site s is labeled malignant; otherwise it is labeled benign. 3) Using these labels as the initial conditions $\mathbf{x}^0$, WICM produces a final labeling $\mathbf{x}^k$. The performance (e.g. sensitivity and specificity) of this procedure is a function of the two thresholds T_f and T_{wicm}. T_f is fixed at an empirically chosen value. (We have observed that system performance is remarkably consistent over a wide range of T_f.) We vary T_{wicm} to adjust classifier performance. In the following subsections we use this approach to construct ROC curves for our two classification systems. These ROC curves and the corresponding areas under them reflect the ability of each system to detect CaP.

It is worth mentioning that classification performance could conceivably be adjusted by fixing T_{wicm} and varying T_f. That is, since modifying T_f alters the initial conditions $\mathbf{x}^0$, it can cause WICM to converge to a different local maximum of (3). However, there is no reason to assume that the individual modes (local maxima) of the weighted *a posteriori* distribution correspond to meaningful classifications; and consequently, this method has no obvious justification.

4.1 Detecting Cancerous Glands on WMHS

Methodology: The goal of this classification system is to detect malignant glands in whole-mount histological sections (WMHSs) of the prostate. Figure 1(a) illustrates a prostate WMHS. The black circle delimits the approximate spatial extent of CaP as delineated by a pathologist. The numerous white regions

are the glands, which our system automatically identifies and segments. Figure 1(b) illustrates the segmented gland boundaries in blue. Figure 1(c) provides a magnified view of the white box in Figure 1(b). Let the set $S = \{1, 2, \ldots, N\}$ reference the N segmented glands in a WMHS. The random variable $Y_s \in \mathbb{R}$ indicates the area of gland s. Two glands are neighbors if the distance between their centroids is less than 0.7 mm. The MRF is implemented using pairwise probabilistic Markov models [7]. The distribution $f(y_s|x_s)$ is described using a parametric model—specifically, a mixture of Gamma functions. All results were produced using leave-one-out cross-validation over 20 WHMSs from 19 patients.

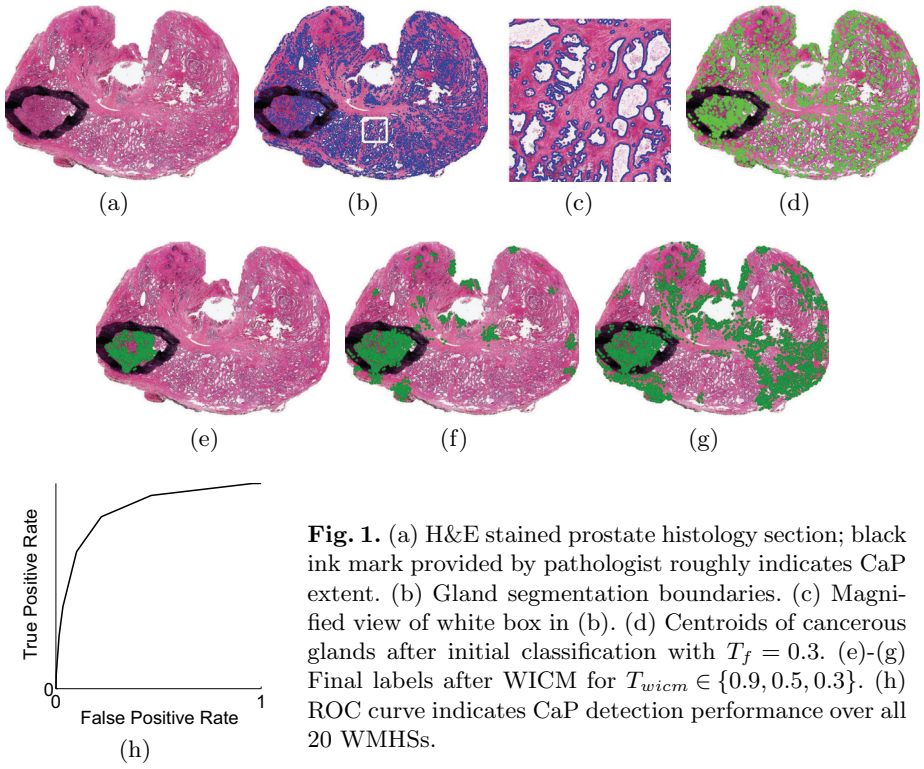


Fig. 1. (a) H&E stained prostate histology section; black ink mark provided by pathologist roughly indicates CaP extent. (b) Gland segmentation boundaries. (c) Magnified view of white box in (b). (d) Centroids of cancerous glands after initial classification with $T_f = 0.3$. (e)-(g) Final labels after WICM for $T_{wicm} \in \{0.9, 0.5, 0.3\}$. (h) ROC curve indicates CaP detection performance over all 20 WMHSs.

Results: The light green dots in Figure 1(d) indicate the centroids of those glands initially classified as malignant (i.e. $\mathbf{x}^0$) with $T_f = 0.3$. Figures 1(e)-1(g) illustrate the final labels (i.e. $\mathbf{x}^k$)—dark green dots indicate the centroids of the malignant glands—for $T_{wicm} \in \{0.9, 0.5, 0.3\}$. Notice that as T_{wicm} decreases, indicating a increase in the weighting of the cancer class as compared to the

benign class, the sensitivity increases. The black line in Figure 1(h) indicates system performance[3] over all 20 WMHSs as T_{wicm} varies from 0 to 1.

4.2 CaP Detection in Multi-protocol MRI

Methodology: In this section we consider a classification system for detecting CaP in multi-protocol *in vivo* MRI. This system combines functional and structural information from dynamic-contrast enhanced (DCE) and T2-weighted (T2-w) 3 Tesla MRI from six patients (18 2D slices). Figures 2(a) and 2(b) illustrate a T2-w MR image and a DCE MR image sampled at the first time point. The green overlays indicate the cancerous extent as specified by a radiologist. Let the set $S = \{1, 2, \ldots, N\}$ reference the N pixels in a T2-w MR image that reside within the prostate. The random vector $Y_s \in \mathbb{R}$ represents the 14 features associated with pixel s. These features are comprised of the T2-w image intensity, six textural features [8] extracted from the T2-w image, and the DCE image intensity sampled at seven time points. The neighborhood η_s of a pixel s is the typical 8-connected region. The MRF is implemented using pairwise probabilistic Markov models. A random forest (i.e. bagging multiple decision trees classifiers) was used to determine the distribution $f(y_s|x_s)$. All results were produced using leave-one-out cross-validation.
Results: The intensities in Figure 2(c) indicate the probability of malignancy at each pixel. Those pixels labeled as malignant during the initial classification with $T_f = 0.1$ are indicated by the red overlay in Figure 2(d). Figures 2(e)-(g) illustrate the malignant labels after WICM for $T_{wicm} \in \{0.8, 0.275, 0.008\}$. The black ROC curve in Figure 2(h) indicates system performance over all 18 studies as T_{wicm} varies from 0 to 1.

5 Concluding Remarks

The inter-variable dependencies within random fields (RFs) necessitate the use of sophisticated estimation strategies. Unfortunately, these strategies provide no clear means for varying classification performance (e.g. sensitivity/specificity). Instead, they produce a single, hard labeling at a static operating point. Addressing this deficiency, we introduced weighted maximum *a posteriori* estimation (WMAP), a generalized form of MAP estimation that allows certain classes to be weighted more heavily than others. We also introduced weighted iterated conditional modes (WICM), an RF-compatible classification strategy capable of WMAP estimation. To illustrate the value of WICM we applied it to separate MRF-based classification systems for detecting prostate cancer (CaP) in whole-mount histological sections and multi-protocol MRI. We demonstrated how WICM could arbitrarily vary the CaP sensitivity of these systems by appropriately altering their associated cost functions.

[3] The true positive rate is the ratio of malignant sites correctly classified to the total number of malignant sites. The false positive rate is the ratio of benign sites *incorrectly* classified to the total number of benign sites.

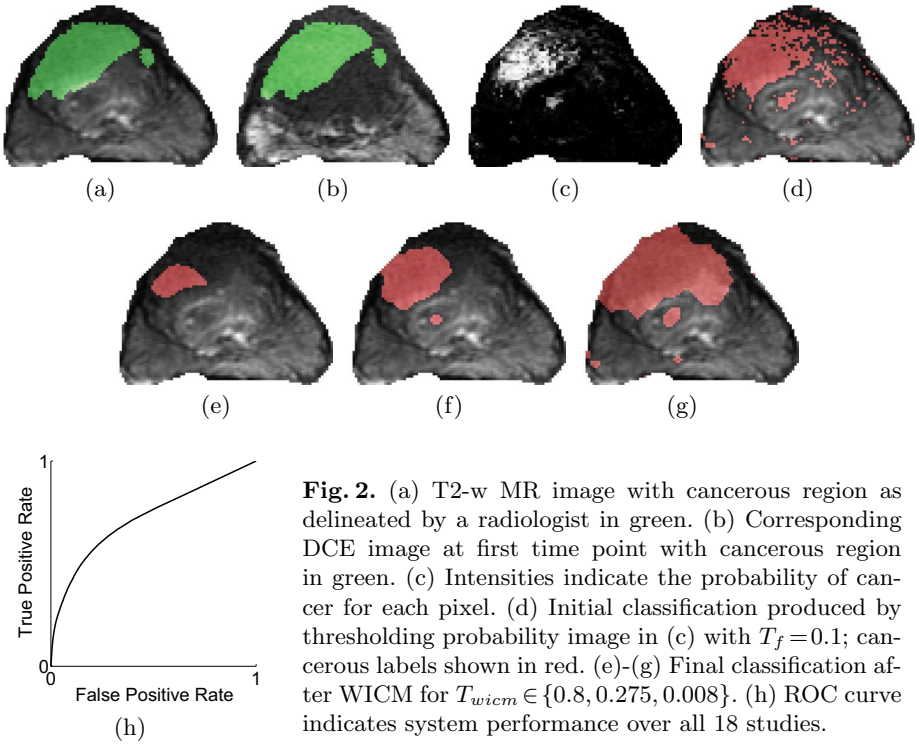


Fig. 2. (a) T2-w MR image with cancerous region as delineated by a radiologist in green. (b) Corresponding DCE image at first time point with cancerous region in green. (c) Intensities indicate the probability of cancer for each pixel. (d) Initial classification produced by thresholding probability image in (c) with $T_f = 0.1$; cancerous labels shown in red. (e)-(g) Final classification after WICM for $T_{wicm} \in \{0.8, 0.275, 0.008\}$. (h) ROC curve indicates system performance over all 18 studies.

Though the concept of weighting probabilities as a means for adjusting classifier performance is well established, its extension to RFs is completely novel. This extension, in retrospect, may seem obvious; yet, previous authors [5, 6], requiring such capabilities, have instead resorted to using *ad hoc* methods.

Acknowledgment

This work was made possible via grants from the Wallace H. Coulter Foundation, New Jersey Commission on Cancer Research, National Cancer Institute (R01CA136535-01, ARRA-NCI-3 R21 CA127186-02S1, R21CA127186-01, R03CA128081-01), the Society for Imaging Informatics in Medicine (SIIM), the Cancer Institute of New Jersey, and the Life Science Commercialization Award from Rutgers University. Additionally, we would like to thank Michael D. Feldman and John E. Tomaszewski from the Dept. of Surgical Pathology at the University of Pennsylvania for the whole mount histology sections, and Dr. Nicholas B. Bloch and Dr. Neil Rofsky from the Beth Israel Deaconess Medical Center for the prostate MRI data.

References

1. Besag, J.: On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society. Series B (Methodological) **48**(3) (1986) 259–302
2. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. IEEE Transactions on Pattern Recognition and Machine Intelligence **6** (1984) 721–741
3. Marroquin, J., Mitter, S., Poggio, T.: Probabilistic solution of ill-posed problems in computational vision. Technical Report AIM-897, MIT Artificial Intelligence Laboratory (March 6 1987)
4. Duda, R., Hart, P., Stork, D.: Pattern Classification. John Wiley & Sons (2001)
5. Salli, E., Aronen, H.J., Savolainen, S., Korvenoja, A., Visa, A.: Contextual clustering for analysis of functional MRI data. IEEE Transactions on Medical Imaging **20**(5) (May 2001) 403–414
6. Maitra, R., Roys, S.R., Gullapalli, R.P.: Test-retest reliability estimation of functional MRI data. Magnetic Resonance in Medicine **48**(1) (2002) 62–70
7. Monaco, J., Tomaszewski, J., Feldman, M., Moradi, M., Mousavi, P., Boag, A., Davidson, C., Abolmaesumi, P., Madabhushi, A.: Probabilistic pairwise Markov models: Application to prostate cancer detection. SPIE Medical Imaging **7260** (2009)
8. Viswanath, S., Bloch, B., Rosen, M., Chappelow, J., Toth, R., N. Rofsky, R.L., Genega, E., Kalyanpur, A., Madabhushi, A.: Integrating structural and functional imaging for computer assisted detection of prostate cancer on multi-protocol in vivo 3 tesla MRI. SPIE Medical Imaging **7260** (2009)

Commentary on *Weighted Iterated Conditional Modes for Random Fields: Application to Prostate Cancer Detection*

Paulo R. S. Mendonça

GE Global Research, One Research Circle, Niskayuna, NY 12309, USA
mendonca@crd.ge.com

Abstract. The paper "Weighted Iterated Conditional Modes for Random Fields: Application to Prostate Cancer Detection," in *Prob. Models for Med. Image Analysis,* 2009, by J. Monaco et al., presents an extension of maximum a posteriori (MAP) estimation in a Markov random field (MRF) by introducing weights in the cost function optimized by MAP, and applies this idea to the generation of receiver operating characteristic curves for classification. Perhaps unbeknown to the authors, this conceptually simple generalization of MAP estimation of MRFs has an interesting connection with other areas of decision theory. A discussion of this connection provides theoretical insight into the authors's contribution and allows for practical elements of the proposed algorithm to be implemented on firmer footing.

Commentary

In their excellent paper, Monaco et al. [1] demonstrate how a simple extension to maximum a posteriori (MAP) estimation in a Markov random field (MRF) allows for a natural mechanism for producing receiver operating characteristic (ROC) curves on the MRF's performance for a given classification task. This contrasts with the static operating point typically found in the literature.

This extension is related to a powerful connection, unexplored in the paper, between Bayesian methods and techniques from decision theory based on utility functions. In particular, the introduction of the weights that generalize equation (1) into equation (2) makes clear that $\mathbf{C}_{MAP}$ (as well as $\mathbf{C}_{WMAP}$) can be interpreted as a log-additive utility function [2]. Moreover, for a log-additive utility function U given by $U = \alpha c_{\mathrm{I}} + \beta c_{\mathrm{II}}$, where α and β are the probabilities of a *type I error* (false rejection of a true null hypothesis) and a *type II error* (failure to reject a false null hypothesis), respectively, and c_{I} and c_{II} are application-dependent costs associated with each type of error, the minimum of U is achieved at a specific point along the ROC curve of the associated Bayesian classifier [3,4]. Therefore, the proposed method works by varying the probabilities of the loss incurred by type I and type II errors. Although this connection is well known, the typical approach for sweeping ROC curves for a Bayesian classifier is to use different thresholds on *Bayes factors* [5], and in doing so the connection between these different approaches is lost.

A practical consequence of this observation is the provision of a theoretical justification for varying thresholds T_f and T_{wicm} simultaneously. One could already ask why

the decision thresholds would be different depending on whether neighborhood information is ($P(x_s|\mathbf{x}_{\eta_s}, y_s)$) or is not ($P(x_s|y_s)$) used. In view of the discussion above this question could be reformulated as: why would $U = U_f$ obtained from $P(x_s|y_s)$ be kept fixed while $U = U_{wicm}$ obtained from $P(x_s|\mathbf{x}_{\eta_s}, y_s)$ is allowed to vary? From a theoretical standpoint is seems natural to assume c_{I} and c_{II} the same in both cases and produce the ROC curve by jointly varying T_f and T_{wicm}. Given the authors's important observation that "system performance is [...] consistent over a wide range of T_f" (Section 4), this change will have limited effect on performance on the data in which the proposed method was tested. Nevertheless, this may not be the case for different data, and it certainly adds to the elegance of the approach.

References

1. Monaco, J., Viswanath, S., Madabhushi, A.: Weighted iterated conditional modes for random fields: Application to prostate cancer detection. In: Prob. Models for Med. Image Analysis. (2009)
2. Pollak, R.A.: Additive von Neumann-Morgenstern utility functions. Econometrica **35**(3) (1967) 485–494
3. Berger, J.O.: Statistical Decision Theory and Bayesian Analysis. 2nd edn. Springer Series in Statistics. Springer-Verlag (1993)
4. DeGroot, M.H., Schervish, M.J.: Probability and Statistics. Addison-Wesley Series in Statistics. Addison-Wesley, Reading, MA, USA (2001)
5. Kass, R.T., Raftery, A.E.: Bayes factors. J. Amer. Statist. Assoc. **90**(430) (1995) 773–795

Incorporating Prior Knowledge on Class Probabilities into Local Similarity Measures for Intermodality Image Registration

Matthias Hofmann[1,2,3], Bernhard Schölkopf[2],
Ilja Bezrukov[2,3], and Nathan D. Cahill[1,4]

[1] Department of Engineering Science, University of Oxford, Oxford, UK
[2] Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[3] Department of Radiology, University Hospital Tübingen, Germany
[4] School of Mathematical Sciences, Rochester Institute of Technology, NY, USA

Abstract. We present a methodology for incorporating prior knowledge on class probabilities into the registration process. By using knowledge from the imaging modality, pre-segmentations, and/or probabilistic atlases, we construct vectors of class probabilities for each image voxel. By defining new image similarity measures for distribution-valued images, we show how the class probability images can be nonrigidly registered in a variational framework. An experiment on nonrigid registration of MR and CT full-body scans illustrates that the proposed technique outperforms standard mutual information (MI) and normalized mutual information (NMI) based registration techniques when measured in terms of target registration error (TRE) of manually labeled fiducials.

1 Introduction

Intermodality image registration describes the task of aligning two images that have been acquired using different modalities and thus have values with different meanings. Typical registration methods are designed in terms of three components: the allowable space of geometric transformations, the similarity measure relating two images, and the optimization technique for finding the optimal value of the similarity measure over the space of allowable transformations.

Mutual information (MI) [1, 2] is one of the most commonly used similarity measures in intermodality image registration. The underlying assumption is that if images values are treated as realizations of some underlying probability distribution function, then the mutual information between both images is maximized when they are aligned. The reason for the success of MI and related measures such as normalized mutual information (NMI) [3] lies in the fact that these measures do not make any underlying assumptions about the modalities. This allows them to be used for a whole range of inter-modality registration problems. In practice, however, most volume registration problems in medical imaging applications use images acquired from a limited number of modalities, including MR, CT, PET, and Ultrasound. When prior knowledge about the images, modalities, or optimal transformations is available, this knowledge can be

exploited to enable better registration results than would be possible with MI or NMI.

Prior knowledge has been previously exploited in a number of ways for image registration [4–9]. Ashburner et al. [4] have shown how incorporating prior knowledge about an object's shape and scale can significantly improve accuracy and reliability for affine registration. Leventon and Grimson [5], Guetter et al. [6], Chung et al. [7], and Sabuncu and Ramadge [8] have all proposed variations on the idea of learning the optimal joint distribution from exemplar aligned images and then designing a similarity measure that reflects the difference between the learned joint distribution and an observed joint distribution. Lee et al. [9] have focused directly on learning an optimal similarity measure for a specific pair of image modalities, assuming exemplar aligned images from those modalities are provided.

While these techniques are useful for certain intermodal registration problems, they seem to rely directly on the image intensities; hence, they may not be directly generalizable to all types of intermodal registration. Whereas CT scanners are calibrated to yield images with an accuracy of within a few Hounsfield units, this is not possible with MR scanners and MR typical field inhomogeneities lead to severe intensity bias. Furthermore, it is not necessarily true that exemplar aligned images will always be available; this is especially true in situations when nonrigid registration is required.

Other previous work has attempted to use prior information in the form of tissue class probabilities to aid in registration. D'Agostino et al. [10] describe how to align images by minimizing the Kullback-Leibler divergence between the actual joint class distribution and an ideal joint class distribution, although their formulation assumes that individual class probabilities in pairs of corresponding voxel positions in two images are independent. Lorenzen et al. [11] avoid making this assumption and propose an approach based on local KL divergence. Although Lorenzen's approach is applicable for the more general problem of *multiple* image registration, it requires the additional step of atlas construction even when only two images must be registered.

In this paper, we show how to design a nonrigid registration algorithm based on vector valued images. Such a nonrigid registration algorithm exploits prior knowledge about the imaging modalities, but unlike other techniques, it does not require exemplar aligned images, independence assumptions, or atlas construction. We compare the performance of the proposed nonrigid registration technique versus standard MI-based registration on full-body MR/CT images, measuring the results in terms of target registration error (TRE) on manually labeled fiducial points.

2 Prior Knowledge on Class Probabilities

In medical images, the intensity at a given voxel depends on the underlying tissue class at the voxel's position. If it is possible to infer the underlying tissue class at each position in an image, this knowledge could potentially be exploited by

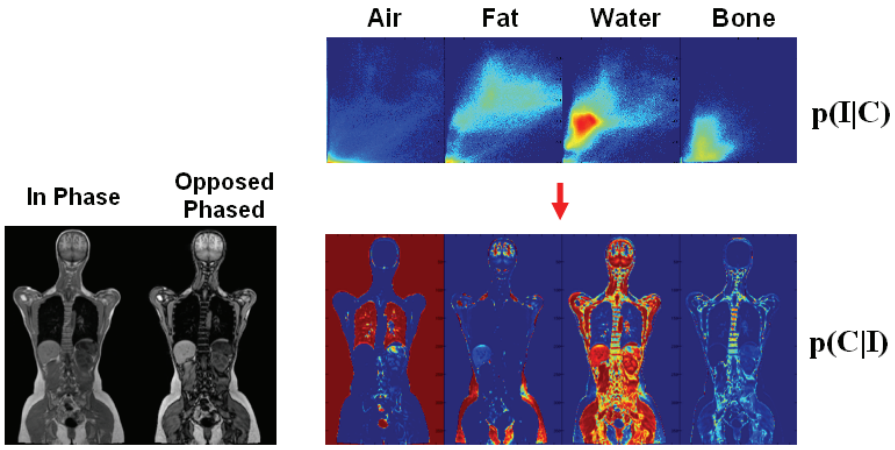


Fig. 1. Estimating class probabilities using Bayes' rule and pre-segmented images. Red/blue colors indicate high/low probabilities, respectfully.

a registration algorithm. In many situations, however, it is difficult to uniquely identify the underlying tissue class at a voxel solely from the intensity value at that voxel. This can be remedied by using additional information that may be available directly from the imaging device, from a pre-segmentation, or from a probabilistic atlas.

One example of additional information that may be provided directly from an imaging device is the phase information in MR imaging that is available after MR reconstruction [12]. This information is not usually used for diagnosis, but may be useful in inferring the underlying tissue class. If no additional direct source of information is available, class probabilities can be estimated based on the intensity values by using pre-segmented images. Assuming that pre-segmented images and the image to be registered have similar intensity characteristics, the probability that a particular voxel has tissue class C, given the voxel intensity I, is given by $p(C|I)$. According to Bayes' rule, $p(C|I) \propto p(I|C) \cdot p(C)$, where $p(C)$ is a prior probability on how often class C is expected to occur. Figure 1 provides an illustration of the process of estimating class probabilities at each voxel in a full-body MR scan.

Alternatively, if there is no direct source of additional information available, and pre-segmentation is not desirable, probabilistic atlases can be used to provide prior information about which tissue class is expected at different positions in the patient [13, 14]. Since the anatomy of different subjects shows structural similarities, the images of different subjects can be brought into alignment through intersubject registration. For example, a template MR image for which tissue labels are available can be registered to a patient MR image. Applying the same transformation on the label image yields a tissue label prediction for the patient MR image. Performing this procedure for several template MR image and label pairs can be used to derive a position-depedendent prior probability $p(C|X)$.

This prior probability can be combined with the intensity-derived class estimate using Bayes' rule:

$$p(C|I,X) \propto p(C|X) \cdot p(I|C) \tag{1}$$

Whether estimated directly from the imaging modality, or indirectly from pre-segmented images or probabilistic atlases, once class probabilities have been derived for each voxel, they can be used to form distribution-valued images. These distribution-valued images can be subsequently used for registration, hopefully yielding better results than registration based solely on the original intensity-valued images.

3 Nonrigid Registration Preliminaries

In this section, we describe a variational framework for registering two images, using MI or NMI as a similarity measure. In the next section, we will extend this framework and introduce new similarity measures to allow for vector-valued and distribution-valued images.

Consider two images, a reference image R and a floating image F, both as functions mapping $\Omega \subseteq \mathbb{R}^d \mapsto \mathbb{R}^N$. Define a *deformation* $\mathbf{\Phi} : \mathbb{R}^d \mapsto \mathbb{R}^d$ by $\mathbf{\Phi}(\mathbf{x}) = \mathbf{x} - \mathbf{u}(\mathbf{x})$, and call $\mathbf{u}$ the *displacement*. The general form of the variational registration problem is given by:

$$\min_{\mathbf{u}} \qquad \mathcal{E}(R,F,\mathbf{u}) := \mathcal{S}(\mathbf{u}) + \alpha \mathcal{J}(R, F^{\mathbf{u}}) \quad , \tag{2}$$

where $\mathcal{J}$ is a similarity measure that quantifies the similarity between the reference image R and the deformed floating image $F^{\mathbf{u}} := F(\mathbf{\Phi})$, $\mathcal{S}$ is a regularizer that ensures that the minimization problem is well-posed and that the solution is smooth in some sense, and α is a weighting parameter.

In this paper, we will assume the use of the curvature regularizer [15]:

$$\mathcal{S}(\mathbf{u}) := \int_{\Omega} \sum_{j=1}^{d} (\Delta \mathbf{u}_j(\mathbf{x}))^2 d\mathbf{x} \ . \tag{3}$$

Normally, variational registration proceeds by identifying the Euler-Lagrange equations associated with (2) and embedding them in an artificial time. When the curvature regularizer is used, this process yields the evolution equation:

$$\partial_t \mathbf{u}(\mathbf{x},t) + \Delta^2 \mathbf{u}(\mathbf{x},t) = -\alpha \mathbf{f}(\mathbf{x}, R, F^{\mathbf{u}}) \,, \quad \mathbf{x} \in \Omega, \quad t > 0, \tag{4}$$

$$\mathbf{u}(\mathbf{x},0) = \mathbf{u}^{(0)}(\mathbf{x}) \quad , \tag{5}$$

where $\mathbf{u}^{(0)}(\mathbf{x})$ is typically chosen to be 0, and where Dirichlet or Neumann boundary conditions are typically imposed on the boundary of Ω. The equilibrium solution of (4)–(5) yields a stationary point of (2).

To handle large deformations, we follow a strategy similar to fluid registration [16], and solve the evolution equation in terms of the velocity field $\mathbf{v}$ instead of

the displacement field $\mathbf{u}$; i.e.,

$$\partial_t \mathbf{v}(\mathbf{x},t) + \Delta^2 \mathbf{v}(\mathbf{x},t) = -\alpha \mathbf{f}(\mathbf{x}, R, F^{\mathbf{u}}), \quad \mathbf{x} \in \Omega, \quad t > 0, \tag{6}$$

$$\mathbf{v}(\mathbf{x},0) = \mathbf{0} \ , \tag{7}$$

where the velocity and displacement are related by the material derivative:

$$\mathbf{v}(\mathbf{x},t) = \partial_t \mathbf{u}(\mathbf{x},t) + (\nabla \mathbf{u}(\mathbf{x},t))^{\mathrm{T}} \mathbf{v}(\mathbf{x},t). \tag{8}$$

The vector field $\mathbf{f}(\mathbf{x}, R, F^{\mathbf{u}})$ that forms the nonhomogeneous term in (4) is called the *force vector*, and it arises from the Gâteaux derivative of $\mathcal{J}$. Force vectors can typically be determined analytically from the similarity measure at hand. For the MI and NMI, we first present the similarity measures and then their corresponding force vectors. The MI and NMI are defined by:

$$MI(R, F^{\mathbf{u}}) := H(R) + H(F^{\mathbf{u}}) - H(R, F^{\mathbf{u}}) \ , \tag{9}$$

$$NMI(R, F^{\mathbf{u}}) := \frac{H(R) + H(F^{\mathbf{u}})}{H(R, F^{\mathbf{u}})} \ , \tag{10}$$

where $H(A)$ and $H(A,B)$ are the marginal and joint entropies, defined by:

$$H(A) = -\int_{-\infty}^{\infty} p_A(a) \log p_A(a)\, da \ , \tag{11}$$

$$H(A,B) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{A,B}(a,b) \log p_{A,B}(a,b)\, dadb \ , \tag{12}$$

and where $p_A(a)$ and $p_{A,B}(a,b)$ are the values of the probability density function of A at a, and the joint probability density function of (A,B) at (a,b), respectively.

The corresponding force vectors for MI and NMI are given by:

$$\mathbf{f}_{MI}(\mathbf{x}, R, F^{\mathbf{u}}) = \mathcal{P}_H(\mathbf{x}; F^{\mathbf{u}}) - \mathcal{P}_H(\mathbf{x}; R, F^{\mathbf{u}}) \ , \tag{13}$$

$$\mathbf{f}_{NMI}(\mathbf{x}, R, F^{\mathbf{u}}) = \frac{\mathcal{P}_H(\mathbf{x}; F^{\mathbf{u}}) - NMI(R, F^{\mathbf{u}})\, \mathcal{P}_H(\mathbf{x}; R, F^{\mathbf{u}})}{H(R, F^{\mathbf{u}})} \ , \tag{14}$$

where

$$\mathcal{P}_H(\mathbf{x}; F^{\mathbf{u}}) = \frac{1}{|\Omega|} \left(\frac{p'_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))}{p_{F^{\mathbf{u}}}(F^{\mathbf{u}}(\mathbf{x}))} \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \ , \tag{15}$$

$$\mathcal{P}_H(\mathbf{x}; R, F^{\mathbf{u}}) = \frac{1}{|\Omega|} \left(\frac{\frac{\partial}{\partial f} \left[p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), f) \right]_{f = F^{\mathbf{u}}(\mathbf{x})}}{p_{R,F^{\mathbf{u}}}(R(\mathbf{x}), F^{\mathbf{u}}(\mathbf{x}))} \right) \nabla F^{\mathbf{u}}(\mathbf{x}) \ . \tag{16}$$

The force vector for MI is derived in [17], and the force vector for NMI follows a similar derivation.

Once the force vector has been chosen, the evolution equations (4) or (6) can be approximately solved by introducing a discretization that is backward in time and centered in space, and then by using Fourier methods [18] to resolve the linear system that arises at each time step. If (6) is being solved, the displacement field $\mathbf{u}$ can be computed from the velocity field $\mathbf{v}$ at each time step by Euler integration of (8).

4 Nonrigid Registration with Vector-Valued Images

If the images being registered are vector-valued or distribution-valued, the same nonrigid registration framework can be utilized as long as appropriate similarity measures and force vectors are defined.

We now denote the reference and floating images as functions mapping $\Omega \subseteq \mathbb{R}^d \mapsto \mathbb{R}^N$; namely, $\mathbf{R} = (R_1, \ldots, R_N)^{\mathrm{T}}$ and $\mathbf{F} = (F_1, \ldots, F_N)^{\mathrm{T}}$. The bold font indicates a vector quantity, whereas the regular font indicates a scalar quantity.

The variational registration problem (2) can still be used, but with R and F replaced by $\mathbf{R}$ and $\mathbf{F}$, respectively. We must now define similarity measures that operate on these vector-valued images. One such measure can be envisioned as the integral of the point-wise sum of squared differences between vector components. We call this similarity measure the local SSD (LSSD), and define it by:

$$LSSD(\mathbf{R}, \mathbf{F}^{\mathbf{u}}) := \int_{\Omega} \sum_{j=1}^{N} \left(R_j(\mathbf{x}) - F_j^{\mathbf{u}}(\mathbf{x})\right)^2 d\mathbf{x} \ . \tag{17}$$

The corresponding force vector is derived in the Appendix and is given by:

$$\mathbf{f}_{LSSD}(\mathbf{x}, \mathbf{R}, \mathbf{F}^{\mathbf{u}}) = 2\left[\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x})\right]\left(\mathbf{R}(\mathbf{x}) - \mathbf{F}^{\mathbf{u}}(\mathbf{x})\right) \ . \tag{18}$$

Note that the matrix $\nabla \mathbf{F}^{\mathbf{u}}$ is the $d \times N$ Jacobian matrix of $\mathbf{F}^{\mathbf{u}}$ defined componentwise by $\left[\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x})\right]_{jk} = \frac{\partial}{\partial x_j} F_k^{\mathbf{u}}(\mathbf{x})$. The force vector (18) can be inserted into the evolution equations (4) or (6), and nonrigid regisration can proceed as described in the previous section.

If our vector-valued images contain vectors of class probabilities at each voxel, we can make the more restrictive assumption that not only are $\mathbf{R}$ and $\mathbf{F}$ vector-valued images, they are distribution-valued; i.e., $\mathbf{R}, \mathbf{F} : \Omega \subseteq \mathbb{R}^d \mapsto \mathbb{P}_N$, where $\mathbb{P}_N$ is the open probability simplex defined in [19]:

$$\mathbb{P}_N := \left\{ \mathbf{p} \,|\, \mathbf{p} = (p_1, \ldots, p_N) \in [0,1] \wedge \sum_{i=1}^{N} p_i = 1 \right\} \ . \tag{19}$$

Since the images are distribution-valued, other choices for the dissimilarity measure can be made based on probability metrics or premetrics [20]. In this article, we choose to design a similarity measure based on the Kullback-Leibler divergence.

The Kullback-Leibler divergence [21] is a premetric that describes the relative entropy between two distributions. For each position in the reference and floating images, it is defined by:

$$KL(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x})) := \sum_{j=1}^{N} F_j^{\mathbf{u}}(\mathbf{x}) \log\left(\frac{F_j^{\mathbf{u}}(\mathbf{x})}{R_j(\mathbf{x})}\right) \ , \tag{20}$$

An image similarity measure can be constructed by integrating (20) over the image domain. We call this similarity measure the local KL divergence (LKL),

and define it according to:

$$LKL(\mathbf{R}, \mathbf{F}^{\mathbf{u}}) := \int_{\Omega} KL(\mathbf{R}(\mathbf{x}), \mathbf{F}^{\mathbf{u}}(\mathbf{x}))\, d\mathbf{x} \ . \tag{21}$$

The corresponding force vector is derived in the Appendix and is defined componentwise by:

$$[\mathbf{f}_{LKL}(\mathbf{x}, \mathbf{R}, \mathbf{F}^{\mathbf{u}})]_j := -\sum_{k=1}^{N} [\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x})]_{j,k} \left(\log\left(\frac{F_k^{\mathbf{u}}(\mathbf{x})}{R_k(\mathbf{x})} \right) + 1 \right) \ . \tag{22}$$

As with the LSSD similarity measure, the LKL force vector can be inserted into the evolution equations (4) or (6), and nonrigid registration can proceed as described in the previous section.

Note that the LSSD and LKL are not the only similarity measures that can be designed for vector-valued or distribution-valued data. Theoretically, any distance measure that can be defined in a vector space or over probability distributions can be developed into a similarity measure for image registration. This shows a potential advantage over the registration approach of Lorenzen et al. [11]. Even though our LKL similarity measure is similar to the one used by Lorenzen, it does not require construction of an atlas; Lorenzen's atlas construction step appears to be possible only when KL divergence is used, and it does not appear to be easily generalizable to other similarity measures.

5 Registration Experiment

To validate the use of distribution-valued images in a practical registration, we developed an experiment on a dataset of 4 human whole body MR and CT images, where patients were scanned with their arms above the head for both the MR and the CT scan. Class probabilities for the classes air, fat, soft tissue and bone were computed for the MR and the CT image. For the MR images they were constructed from a normalized product of $p(I|C)$ and $p(C|I, X)$. Here, $p(I|C)$ was determined by histogram binning the intensities of all voxels that belong to each of the four classes, where the class assignment was determined by using a previously obtained segmentation. The quantity $p(C|I, X)$ was determined by averaging the result of several atlas registrations. For the acquisition of the MR images we used a protocol that acquires opposed phased and inphase images in one scan. Resolution was $2.6 \times 2.6 \times 2.6\text{mm}^3$ and TA = 18s, which allowed for breath hold acquisition. In the CT images, the intensity alone was informative enough and no atlas registration was necessary as an additional source of information. We assumed that for the CT images, the intensity distribution $p(I|C)$ could be approximated by a mixture of Gaussians.

We performed nonrigid registration, both with MI and NMI on the original MR and CT images, and with LSSD and LKL on class probability images. For the MI and NMI computations, histograms (and joint histograms) were estimated using 32 (and 32×32) bins. Linear (or bilinear) interpolation was used to

accumulate partial weights in neighboring bins. Registration in all cases was done over three levels in a multiresolution pyramid, with approximately isotropic sampling at 2.6mm in each dimension at the finest level, and 10.4mm in each dimension at the coarsest level. We varied the α parameter that trades off the contribution of the similarity measure and regularizer, and we experimented with various amounts of smoothing at each resolution level.

In order to assess errors in a physically meaningful sense, we manually placed approximately 30 landmarks of corresponding points in each MR and CT image pair, examples of which are shown in Fig. 2. These landmarks were unknown to the registration algorithms; they were only used retrospectively to compute target registration error (TRE).

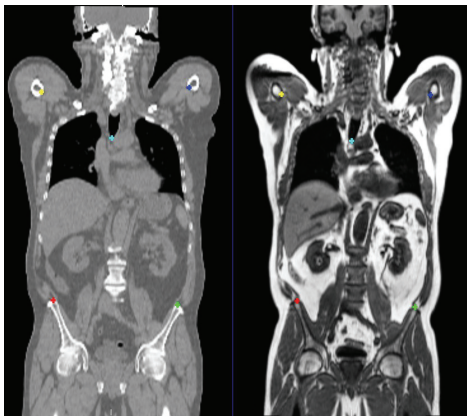


Fig. 2. Manually placed landmarks in CT and MR image.

6 Results

For each similarity measure, we selected the best performing set of values for α and smoothing amount at each resolution level. The resulting target registration errors (TRE) are reported for each patient in Table 1. Results indicate that there is an appreciable improvement in TRE for three out of the four patients when using the vector-valued similarity measures over the MI/NMI.

To illustrate an example registration, Fig. 3 shows fused images before and after a nonrigid registration algorithm using LKL was applied to one of the patient datasets. The left images are slices from an MR dataset that are treated as the reference image and remain unchanged; the right images are slices from a CT dataset that is in its original position (upper) and warped aligned position (lower). In the middle show the fused MR/CT slices, both before (upper) and after (lower) nonrigid registration.

	MI	NMI	LSSD	LKL
Patient 1 Median TRE	17.8	17.6	13.4	13.4
Patient 1 St. Dev. TRE	10.7	10.3	8.0	8.0
Patient 2 Median TRE	13.2	14.4	9.4	8.8
Patient 2 St. Dev. TRE	7.0	6.5	6.0	5.9
Patient 3 Median TRE	16.1	16.9	10.8	10.7
Patient 3 St. Dev. TRE	10.4	9.8	10.6	10.4
Patient 4 Median TRE	13.5	13.9	13.0	13.2
Patient 4 St. Dev. TRE	14.7	14.7	15.3	15.0

Table 1. TRE statistics (in mm) for each patient for each similarity measure.

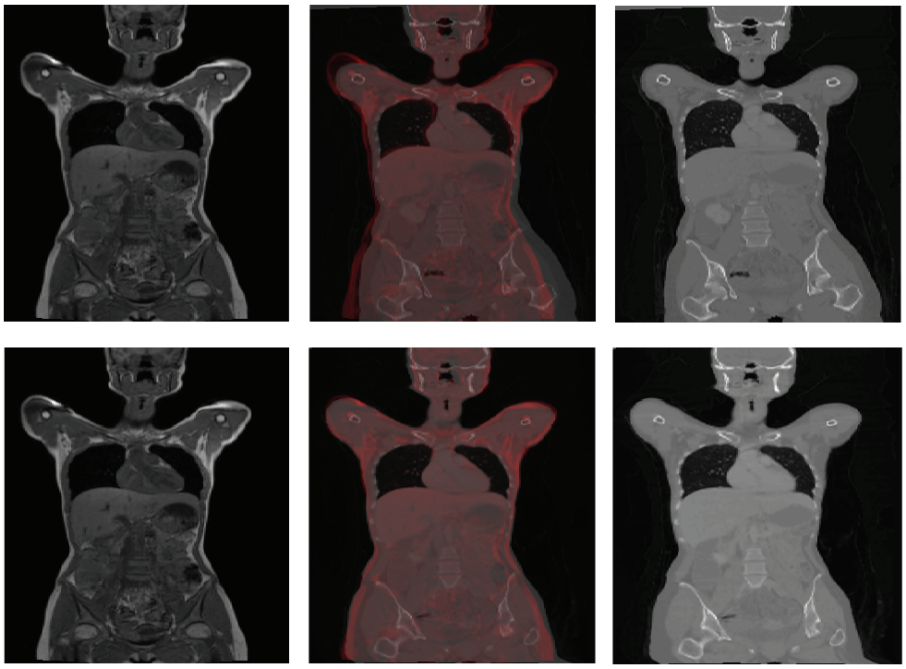


Fig. 3. Example visualization of nonrigid registration. Left images are identical MR slices; right images are CT slices before (upper) and after (lower) nonrigid registration. Middle images are fused versions of the MR/CT slices before (upper) and after (lower) registration.

7 Conclusion

In this article, we have shown that prior knowledge on class probabilities can be incorporated into intermodality image registration. Class probability vectors at each voxel can be constructed from extra information provided by an imaging device, from pre-segmentations, and/or from probabilistic atlases. Incorporation of vector-valued or distribution-valued images into registration requires defining new image similarity measures. In this paper, we presented two such measures, the local SSD, and the local KL divergence, which do not rely on independence assumptions or intermediate atlas formation. When tested on the problem of nonrigid registration of MR and CT full-body images, registration using the proposed measures outperformed standard MI/NMI-based registration a majority of the time when measured in terms of target registration error on manually identified fiducial markers.

Appendix

In this appendix, we derive the force vectors for the LSSD and LKL similarity measures defined in (17) and (21), respectively. We note that for a general similarity measure $\mathcal{J}(\mathbf{R}, \mathbf{F^u})$, the force vector $\mathbf{f}_{\mathcal{J}}(\mathbf{x}, \mathbf{R}, \mathbf{F^u})$ is related to the Gâteaux derivative of $\mathcal{J}$ in the following way:

$$d\mathcal{J}(\mathbf{R}, \mathbf{F^u}; \mathbf{w}) = \int_\Omega \left\langle \mathbf{f}_{\mathcal{J}}(\mathbf{x}, \mathbf{R}, \mathbf{F^u}), \mathbf{w}(\mathbf{x}) \right\rangle d\mathbf{x} \ . \tag{23}$$

To aid in the derivation of the force vectors for the LSSD and LKL similarity measures, we present and prove the following Lemma.

Lemma 1 *If the dissimilarity measure $\mathcal{J}(\mathbf{R}, \mathbf{F^u})$ can be expressed as a functional of the form:*

$$\mathcal{J}(\mathbf{R}, \mathbf{F^u}) = \int_\Omega \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F^u}(\mathbf{x}))\, d\mathbf{x} \ , \tag{24}$$

then the Gâteaux derivative $d\mathcal{J}(\mathbf{R}, \mathbf{F^u}; \mathbf{w})$ can be expressed as:

$$d\mathcal{J}(\mathbf{R}, \mathbf{F^u}; \mathbf{w}) = \int_\Omega \left\langle -\nabla \mathbf{F^u}(\mathbf{x}) \left[\nabla_{\mathbf{F^u}} \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F^u}(\mathbf{x})) \right], \mathbf{w} \right\rangle d\mathbf{x} \ . \tag{25}$$

Proof. By the definition of the Gâteaux derivative, we have:

$$\begin{aligned} d\mathcal{J}(\mathbf{R}, \mathbf{F^u}; \mathbf{w}) &= \lim_{h \to 0} \frac{1}{h} \left[\int_\Omega \mathcal{M}\big(\mathbf{R}(\mathbf{x}), \mathbf{F^{u+hw}}(\mathbf{x})\big)\, d\mathbf{x} - \int_\Omega \mathcal{M}(\mathbf{R}(\mathbf{x}), \mathbf{F^u}(\mathbf{x}))\, d\mathbf{x} \right] \\ &= \lim_{h \to 0} \frac{1}{h} \int_\Omega \int_0^1 \frac{d}{ds} \mathcal{M}\big(\mathbf{R}(\mathbf{x}), \mathbf{F^{u+hsw}}(\mathbf{x})\big)\, ds\, d\mathbf{x} \\ &= \lim_{h \to 0} \frac{1}{h} \int_\Omega \int_0^1 \left\langle \frac{d}{ds} \mathbf{F^{u+hsw}}(\mathbf{x}), \nabla_{\mathbf{F^u}} \mathcal{M}\big(\mathbf{R}(\mathbf{x}), \mathbf{F^{u+hsw}}(\mathbf{x})\big) \right\rangle ds\, d\mathbf{x} \end{aligned}$$

$$= \lim_{h\to 0} \frac{1}{h} \int_\Omega \int_0^1 \left\langle \left\langle \nabla \mathbf{F}^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x}) \, , -h\mathbf{w} \right\rangle , \nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}\big(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}+hs\mathbf{w}}(\mathbf{x})\big) \right\rangle ds \, d\mathbf{x}$$

$$= \int_\Omega \int_0^1 \left\langle \left\langle -\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x}) \, , \mathbf{w} \right\rangle , \nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}}(\mathbf{x})) \right\rangle ds \, d\mathbf{x}$$

$$= \int_\Omega \int_0^1 \left\langle -\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x}) \left[\nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}}(\mathbf{x})) \right] , \mathbf{w} \right\rangle ds \, d\mathbf{x}$$

$$= \int_\Omega \left\langle -\nabla \mathbf{F}^{\mathbf{u}}(\mathbf{x}) \left[\nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}}(\mathbf{x})) \right] , \mathbf{w} \right\rangle \, d\mathbf{x} \qquad \square$$

Now, the LSSD similarity measure (17) can be defined in terms of (24) with:

$$\mathcal{M}(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = \sum_{j=1}^{N} \left(R_j(\mathbf{x}) - F_j^{\mathbf{u}}(\mathbf{x}) \right)^2 \ . \tag{26}$$

For this case, we have:

$$\nabla_{\mathbf{F}^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = -2 \left(\mathbf{R}(\mathbf{x}) - \mathbf{F}^{\mathbf{u}}(\mathbf{x}) \right) \ , \tag{27}$$

and (18) is a direct result of Lemma 1 and (23).

The LKL similarity measure (17) can also be defined in terms of (24) with:

$$\mathcal{M}(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = KL(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = \sum_{j=1}^{N} F_j^{\mathbf{u}}(\mathbf{x}) \log \left(\frac{F_j^{\mathbf{u}}(\mathbf{x})}{R_j(\mathbf{x})} \right) \ . \tag{28}$$

Now for this case, we have:

$$\nabla_{F_k^{\mathbf{u}}} \mathcal{M}(\mathbf{R}(\mathbf{x}) \, , \mathbf{F}^{\mathbf{u}}(\mathbf{x})) = \log \left(\frac{F_k^{\mathbf{u}}(\mathbf{x})}{R_k(\mathbf{x})} \right) + 1 \ , \tag{29}$$

and (22) is a direct result of Lemma 1 and (23).

References

1. Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical image analysis **1**(1) (1996) 35–51
2. Collignon, A.: Multi-modality Medical Image Registration by Maximization of Mutual Information. PhD thesis, Catholic University of Leuven, Leuven, Belgium (1998)
3. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition **32** (1999) 71–86
4. Ashburner, J., Neelin, P., Collins, D., Evans, A., Friston, K.: Incorporating prior knowledge into image registration. Neuroimage **6**(4) (1997) 344–352
5. Leventon, M., Grimson, W.: Multi-modal volume registration using joint intensity distribution. In: Proc. Medical Image Computing and Computer Assisted Intervention. (1998) 1057–1066

6. Guetter, C., Xu, C., Sauer, F., Hornegger, J.: Learning based non-rigid multi-modal image registration using Kullback-Leibler divergence. In: Proc. Medical Image Computing and Computer Assisted Intervention. (2005) 255–262
7. Chung, A.C.S., Gan, R., Wells III, W.M.: Robust multi-modal image registration based on prior joint intensity distributions and minimization of kullback-leibler distance. HKUST CSE Technical Report, HKUST-CS07-01 (2007)
8. Sabuncu, M., Ramadge, P.: Using spanning graphs for efficient image processing. IEEE Transactions on Image Processing (2008) 788–797
9. Lee, D., Hofmann, M., Steinke, F., Altun, Y., Cahill, N.D., Schlkopf, B.: Learning the similarity measure for multi-modal 3d image registration. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2009) to appear.
10. D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: An information theoretic approach for non-rigid image registration using voxel class probabilities. Medical Image Analysis **10** (2006) 413–431
11. Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., Joshi, S.: Multi-modal image set registration and atlas formation. Medical Image Analysis **10** (2006) 440–451
12. Bourgeat, P., Fripp, J., Stanwell, P., Ramadan, S., Ourselin, S.: MR image segmentation of the knee bone using phase information. Medical Image Analysis **11**(4) (2007) 325–335
13. Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J.: A probabilistic atlas of the human brain: theory and rationale for its development the international consortium for brain mapping (ICBM). Neuroimage **2**(2PA) (1995) 89–101
14. Hofmann, M., Steinke, F., Scheel, V., Charpiat, G., Farquhar, J., Aschoff, P., Brady, M., Schlkopf, B., Pichler, B.J.: MR-based Attenuation Correction for PET/MR: A Novel Approach Combining Atlas Registration and Recognition of Local Patterns. Journal of Nuclear Medicine **49**(11) (2008) 1875–1883
15. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press (2004)
16. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable templates using large deformation kinematics. IEEE Transactions on Image Processing **5**(10) (October 1996) 1435–1447
17. Hermosillo Valadez, G.: Variational Methods for Multimodal Image Matching. PhD thesis, Université de Nice - Sophia Antipolis (2002)
18. Cahill, N.D., Noble, J.A., Hawkes, D.J.: Fourier methods for nonparametric image registration. In: Proceedings of the CVPR Workshop on Image Registration and Fusion. (June 2007) 1–8
19. Pohl, K.M., Fisher, J., Bouix, S., Shenton, M., McCarley, R., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Using the logarithm of odds to define a vector space on probabilistic atlases. Medical Image Analysis **11** (2007) 465–477
20. Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. International Statistical Review **70** (2002) 419–435
21. Kullback, S., Leibler, R.A.: On information and sufficiency. Annals of Mathematical Statistics **22**(1) (1951) 79–86

Medial Models of Populations of Nearly Tubular Objects*

Rohit R. Saboo[1], Joshua H. Levy[3], Edward M. Chaney[2], and Stephen M. Pizer[1,2]

[1] Computer Science and [2] Radiation Oncology, The University of North Carolina at Chapel Hill
[3] Formerly [1], presently Morphormics Inc.

Abstract. Many structures in the world and several in the human body are nearly tubular in shape, i.e., have approximately circular cross-sections. Examples are portions of blood vessels, the colon and the whole head and neck. The method of modeling populations of slab-like objects as medial objects and segmentation using statistical shape and appearance models has been shown to be successful for several structures. However, the medial surface of a tubular object degenerates to a curve, and the statistics of even nearly tubular objects represented as slabs will typically be unstable. In this paper, we detail the representation, geometry and means of computing statistics of a population of nearly tubular medial models. We test our method on CTs of real rectums.

1 Introduction

In the human body, the blood vessels, the bronchi and the colon are examples of nearly tubular objects. Segmenting these structures is an important task in medical imaging and learning probability distributions on their populations is useful to segmentations [1]. Most of them can be thought of as a tube at the large scale with smaller scale changes understood as deviations from the tube. Some of these are shown in Fig. 1.

There are several definitions of tubular objects in the literature. Koenderink [2] defines a tube as the envelope of a set of spheres centered on a space curve. This is the medial definition of a tube. Swept surfaces and generalized cylinders model tubes as skeletal structures. A tubular generalized cylinder has a circular cross-section that may vary in size and have a possibly bent axis. In this paper, we discuss geometry and statistics for the former definition of tubes and then extend it to support deviations from the tubular structure. In Section 1.1 we discuss prior work done on modeling tubular objects. Section 1.2 presents the segmentation objective that serves as the driving problem.

1.1 Prior Work

Generalized cylinders, also known as generalized cones, were fist proposed by Binford [3] with special instances studied extensively in computer vision. A straight homogenous generalized cone [4] is the surface obtained by sweeping a fixed cross-section along a straight axis while possibly scaling it, whereas a straight homogenous generalized cylinder may have a cross-section that can change shape. Huang et al. [5,6] discuss

* This work was supported under NIH grant number P01 EB02779. We would also like to thank other members of the Medical Image Display and Analysis Group at UNC Chapel Hill.

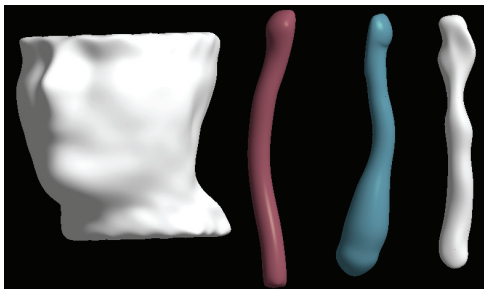


Fig. 1. Renderings of quasi-tube models fitted to different structures. From the left to right the structures are sections of the skin surface extracted from a 3D CT scan, carotid artery, internal jugular vein and the pharynx.

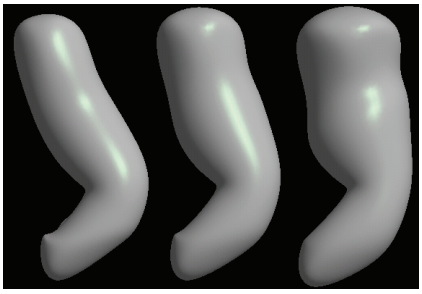


Fig. 2. mean model of a rectum (center) deformed by ± 1.5 standard deviations along the first mode of variation (left and right), which resembles the anatomical shape change due to gas.

generalized tubes that are constructed by sweeping a fixed cross-section along an axis with certain constraints. Terzopoulos et al's [7] physically motivated deformable model uses image-based and regularity forces to deform the model. In our applications, it is difficult to get image-based forces to work due to poor contrast and noise.

Several of these center-line based methods are agnostic to the choice of the center-line. However, when modeling populations, there needs to be a principled way of finding the center-line so that no unwanted variation introduced due to the modeling process is reflected in the statistics. O'Donnell et al. [8] discuss a novel method of generalized cylinders that works around part of this issue by starting with a base cross-section that may be anisotropically scaled. Although they use only two scaling parameters, their method can be extended to produce arbitrary scaling. Further, they allow for local deformations of the cross-section by a spline function on the surface. However, they have not discussed any method to compute statistics of their structures.

The subtle differences between a skeletal and a medial axis are sometimes overlooked; the two are used interchangeably in some of the cited work. Having a true medial axis representation overcomes the issue of finding a unique center-line. The class of generalized cylinders whose medial axis is a curve is restricted to those with a circular cross-section. A non-circular cross-section results in a 2D medial surface.

A generalized cylinder and a structure with a well-defined medial axis are closely related. When we sweep a constant circular cross-section along a curved axis, the curved axis is the medial axis. However, if we sweep a non-constant circular cross-section along a curved axis, then such a structure may not have a curve as its medial locus, though some such objects (generated by sweeping spheres of varying sizes) will have a curve as the medial locus. Even when a generalized cylinder does not have a curve as its medial axis, it is useful to find an approximate medial axis for that object.

A quasi-tubular object can be thought of as a structure that is modeled as deviations from a tubular object. In the general case, it is a structure with a cross-section that may not be close to circular but does not vary much along the axis.

A sweep of a cross-section along an axis may result in two adjacent cross-sections crossing each other near a sharp bend on the axis. Such instances of the generalized cylinder are illegal, restricting the range of permissible cross-sections and axes. Damon [9] has described a method in the swept surface paradigm using a shape operator that can be used to detect these illegal generalized cylinders.

Several alternative methods exist that focus on extraction of the center-line from image data. Examples include the cores methods (height ridges of medial strength) developed by Aylward et al. [10] and by Fridman et al. [11].

No specialized means of performing statistics have been developed for the generalized cylinders and swept surface models discussed above, so they are best suited for modeling *individual* quasi-tubular objects versus *populations* of them. Such statistical descriptions on populations are useful if objects are to be segmented from images in which they have low contrast at their boundaries. There has been some work on modeling tubes with the help of a statistical shape model. The generalized stochastic tubes developed by Huang et al. [12] aid in the segmentation of blood vessels but are specialized for this application. De Bruijne et al. [13] have adapted the method of Active Shape Models with center-line based methods.

With statistical shape models a special concern is their robustness against the number of training samples, since in medicine these training samples can be very expensive. As mentioned by Joshi et al. [14], the orientation of the narrow medial sheet of objects with a nearly circular cross-section is sensitive to small changes in the boundary and will result in a population with broad variation. By avoiding this variability, the method we describe uses statistical shape models for which the probability estimation is particularly robust against the number of training samples.

1.2 The Driving Problem: Segmentation of Quasi-tubes

The method of segmentation via posterior optimization of m-reps developed by Pizer et al. [15] has been successful in dealing with slab-shaped objects with a lot of variability and poor contrast. We develop a new method that draws on the ideas from these methods but represents a tube-like object with a discretely sampled medial space curve and then models quasi-tubes as deviations from these tubes.

The segmentation method can be divided into two parts: training and the actual segmentation itself. During training, a rough m-rep model of the object is allowed to vary inside an optimizer that favors smooth models with a regularly spaced discrete medial mesh and that match well with the image data. The resulting models are known as training or fitted models.

These training models are then statistically analyzed. The variation in the shape space of the models is studied using Principal Geodesic Analysis (PGA), developed by Fletcher et al. [16], which is a variation of Principal Component Analysis (PCA) suited for non-linear spaces. The result is a mean shape $\overline{\mathbf{m}}$ and a prior $p(\mathbf{m})$ for segmentation. At the same time, the region around the object is divided into small parts and the distribution of intensities in each region is studied with the help of local region intensity quantile functions, developed by Broadhurst [17] and Stough [18]. We then apply PCA on these quantile functions to produce a likelihood function $p(\mathbf{I}|\mathbf{m})$ for segmentation.

When an image is to be segmented, the mean model is placed close to the real organ with the help of landmarks or manually. The model is allowed to deform along its principal modes of variation in an optimizer that favors likely shapes and intensity distributions around and within the object. The objective function maximized is the weighted sum of $\log p(\mathbf{m})$ and $\log p(\mathbf{I}|\mathbf{m})$ with the weights chosen to make the two terms have equal variance. This is a variant of the method of posterior optimization.

The remainder of the paper is organized as follows. In Section 2 we describe the representation and geometry for tubular medial models. In Section 3, we describe the way in which we estimate probability distributions on these models. We then describe the modeling of the deviations from a tubular to a quasi-tubular model in Section 4. Section 5 gives more details of our training and segmentation approaches. Finally in Section 6, we test our method on real data obtained from CTs of rectums and provide both quantitative and qualitative results on the same.

2 Medial Models for Tubes

A first order tube m-rep is a continuous space curve with a cone placed at every point along the curve. The axis of the cone is tangential to the space curve at the tip of the cone. Sweeping the edges of the cone bases gives the boundary of the modeled object, which is orthogonal to the rays from the cone tip to the cone base. The cones may have a half cone angle greater than $\pi/2$ but less than π. They are not allowed to intersect each other. Damon [19] has provided us with tools that can be used to measure local self-intersection (folding) of the object implied by the medial surface of a slab m-rep. In Section 2.1 we adapt these tools to do the same for tubular m-reps.

In practice, we represent the medial model of a tube by discretely sampling the space curve of cone tips. Each sample, shown in Fig. 3, is called an atom. Associated with each sample is its position in space, $\mathbf{p} = (x, y, z)$, and a cone with its tip positioned on the sample. The cone in turn is represented by its bisector, $\widehat{\mathbf{U}}_0 = (U_{0,1}, U_{0,2})$, the half cone angle, θ, and the length of its inclined surface rays, r. The bisector of the cone always points along increasing arc length. Thus, the bisector points in the same direction when θ changes across $\pi/2$. To keep the discrete samples regularly spaced, while developing the models, we impose a penalty, called *irregularity penalty*, on the model that penalizes atoms moving away from the average of its neighbors.

The atoms at the two ends of the chain have an additional parameter describing the curvature of the cap at that end. However, when we are modeling open tubes, the end-atoms don't have any special properties and are just like any other atom.

A continuous medial curve $\gamma(u)$ is interpolated from these atom positions and cone bisector vectors with the help of piecewise cubic Hermite splines. The cone bisector vectors, scaled by the mean of the distance between the position of the atom and its two neighbors, are used as the tangents in the Hermite interpolation.

To resolve the rotational symmetry, we have a parameter ϕ that rotationally orients the entire tube along its length. One of the atoms in the tube is designated as a base atom. Usually this atom is close to a feature in the object that can help fix the rotational orientation. This feature can be an anatomic entity such as a part of a bone, a certain neighboring organ or tissue that can be easily identified in the entire population.

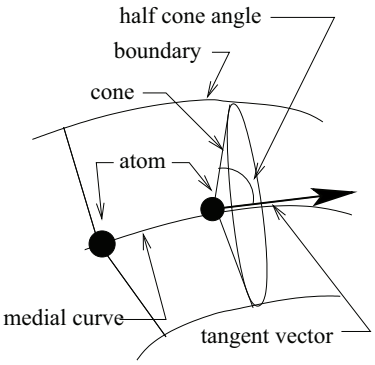


Fig. 3. Representation of a tube atom

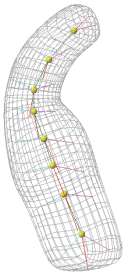


Fig. 4. A mean model of a rectum from one of our studies showing the medially implied surface as a wireframe.

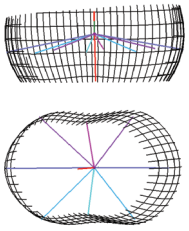


Fig. 5. A quasi-tube atom with spokes of varying length and a cut-away section of the medially implied surface shown in two different orientations.

Whenever the cone for this atom is rotated around its bisector, all the other cones are sympathetically rotated. This is needed for correspondences that depend upon the position along the circumference of the tube such as those required for quasi-tube statistics.

The surface implied by the tubular medial model is called as the *medially implied surface* and generated as follows. The circles at the base of the cones are discretely sampled. The medially implied surface is the fine and smooth mesh obtained by interpolating the set of surface points generated from the cones of all the atoms. The medially implied surface for a tube model of a rectum is shown in Fig. 4.

In Section 2.1, we highlight a geometric operator that is useful for detecting models with self-intersecting surfaces and interpolating between atoms in the medial axis. In Section 2.2, we describe a geometric measure, which we call curviness, that is useful in avoiding wavy medial axes.

2.1 Geometry on tubular models

Consider the medial atom cone to be composed of a set of spokes U extending from the tip to the base. These spokes are parameterized by the arc angle $\phi \in [0, 2\pi)$ on the circumference of the base made with respect to a zero point on the circumference. Define the 1D radial shape operator for a tube as

$$S_{rad}(\phi) = -proj_U\left(\frac{\partial \mathbf{U}(\phi)}{\partial s}\right), \tag{1}$$

where $proj_U$ is projection along the spoke vector U, $\mathbf{U}$ is the corresponding unit spoke vector and s is an arc-length parametrization of the medial curve. Note that S_{rad} and r are also functions of u but we will ignore this for brevity. The derivative may be written in the form

$$\frac{\partial \mathbf{U}(\phi)}{\partial s} = a \cdot \mathbf{U} - \kappa_{rad} \cdot \gamma'(s), \tag{2}$$

where κ_{rad} is the principal radial curvature. Thus $S_{rad}(\phi) = \kappa_{rad}(\phi)$.

An important use of this shape operator is in detecting models that are *illegal*, i.e., some of the spokes are crossing each other and the surface has folded onto itself. Damon [19] shows us that spokes will cross each other if and only if $r \times \kappa_{rad} > 1$. Unlike Damon's shape operator, our shape operator is a function of the angle ϕ. However, it suffices to evaluate this for the angle corresponding to the direction of the curve normal.

The model is illegal if $\exists\ u$ s.t. $r \times S_{rad} >= 1$. This condition can be relaxed as a penalty more suitable for an optimizer, which expects a continuous objective function, that is the p-norm of the individual measurements $\max(0, r \times S_{rad} - \beta)$ along the tube for a certain threshold $\beta \in [0, 1)$. Experimentally, we have seen that $p = 6 - 8$ and a threshold of $0.8 - 0.9$ produces good results. Larger values of p are sometimes useful when we wish to make the aggregate measure more sensitive to local problems.

The S_{rad} operator can also be used to interpolate between two consecutive atoms on the same medial manifold. Han et al. [20] have used this operator to interpolate atoms in 2D medial manifolds. We have adapted the interpolation method to generate interpolated atoms for tubes. An important use is in interpolating atoms to improve correspondence between models.

2.2 Geometric Penalty - Curviness

A wavy medial sheet results in crooked-looking models. It also necessitates the use of significantly more samples in the computation of the illegality penalty. Penalizing fitted m-reps according to an aggregate measure of curviness can alleviate these problems. Apart from being anatomically improbable, there is nothing wrong with crooked-looking models. We define *curviness* by the p-norm of the total curvature over the entire length of the medial curve.

$$C = \left(\frac{1}{n-1} \int_0^{n-1} T^p du \right)^{1/p}, \tag{3}$$

where the total curvature T is related to the geometric curvature κ and the geometric torsion τ of the interpolated medial axis $\gamma(u)$ by $T = \sqrt{\kappa^2 + \tau^2}$. Different values of p between 2 and 10 are appropriate depending upon how much we need to emphasize individual sharp bends. However, larger values of p create steeper gradients that may not be good in an optimizer.

The curviness measure is rotation and translation invariant. Dividing by the arc length or the average inter-atom distance makes it scale-invariant too.

3 Shape Space and Statistics

A tube atom $\mathbf{m}$ can be represented by the tuple $\mathbf{M} = \langle \mathbf{P}, \hat{\mathbf{U}}_0, \theta, r \rangle$. A tube consisting of n atoms can be represented by n such tuples concatenated together. Here the cone vertex $\mathbf{P}$ belongs to the group $\mathbb{R}^3$, the cone axis $\hat{\mathbf{U}}_0$ is a point on the two dimensional sphere S^2, the half cone angle θ ranges from 0 to π and is related to the group $\mathbb{RP}^1$, and

the cone length r belongs to $\mathbb{R}^+$. All of the groups are Lie groups and except for $\mathbb{R}^3$, they are not Euclidean manifolds. A Lie group has a differentiable group operator, an inverse element and the identity element. Using the group operators and by the action of the inverse on an element of the group on the element itself, we obtain the identity element. The other things needed are a distance metric in this space and the ability to project back and forth between this shape space and a tangent plane. Several of the results presented in this section are similar to those worked out by Fletcher et al. [16] for slabular m-reps.

The path with the shortest distance between two points in a manifold is known as the geodesic between them. The length of this path is called the geodesic distance. If we have a suitable mapping between the manifold and a tangent space, then this mapping can be used to measure the geodesic distance on the tangent plane. The map that takes us from the manifold to the tangent space is called as the Logarithmic chart and the reverse map is known as the Exponential chart. The maps of the individual components are given by the following equations. The map for the atom is simply the direct product of these maps. The maps for the position are identity functions and that for the radius are logarithm and exponential functions. In the following equations, $\theta \in (0, \pi)$, $\mathbf{U} = \langle u_1, u_2, u_3 \rangle \in \mathbb{S}^2$ and $\mathbf{U}' = \langle u_1', u_2' \rangle \in \mathbf{T}_{(0,0,1)}\mathbb{S}^2$.

$$\mathrm{Log}(\theta) = \tan\left(\theta - \frac{\pi}{2}\right), \tag{4}$$

$$\mathrm{Exp}(\theta') = \tan^{-1}(\theta) + \frac{\pi}{2}, \tag{5}$$

$$\mathrm{Log}(\mathbf{U}) = \left(u_1 \frac{\alpha}{\sin\alpha}, u_2 \frac{\alpha}{\sin\alpha}\right), \tag{6}$$

$$\mathrm{Exp}(\mathbf{U}') = \left(u_1' \frac{\sin|u'|}{|u'|}, u_2' \frac{\sin|u'|}{|u'|}, \cos|u'|\right), \tag{7}$$

$$\tag{8}$$

where $\alpha = \cos^{-1}(u_3)$ and $|u'| = \sqrt{u_1'^2 + u_2'^2}$. All the above maps are taken centered at the identity element. The identity element for the group $\mathbb{R}^3$ is 0, for $\mathbb{R}^+$ it is 1 and for S^2 it is the point $(0, 0, 1)$. For the group of θ, the identity element is $\pi/2$. To obtain the chart for a tangent plane centered at a point $\mathbf{m}$ different from the identity element, we need to apply the inverse of that point to the element in order to move the tangent plane to the identity element. We use $\mathrm{Log}_{\mathbf{m}}$ and $\mathrm{Exp}_{\mathbf{m}}$ as the notation in this case.

To make the units of all the components commensurate in the Log map, we multiply the unitless quantities with the mean radius taken over all the corresponding atoms in the population. The geodesic distance is then defined as the norm of the difference of these normalized atoms projected into the tangent space.

We define the Fréchet mean μ of a set of atoms as the one that minimizes the sum of squared geodesic distances from all the corresponding atoms from the population.

$$\mu = \underset{\mathbf{m} \in \mathbf{M}}{\mathrm{argmin}} \sum_i |\mathrm{Log}_{\mathbf{m}}(\mathbf{m}_i)|^2. \tag{9}$$

We then compute the Log map of all these atoms and project them on the tangent plane centered on the mean. We do PCA on these projected atoms and keep the first

few modes that represent more than 90% of the total variation. Fig. 2 shows the shape variation along the first mode of variation of the rectum from our study.

4 Quasi-tubes

In this section, we show how we model the deviations from a tube to a quasi-tube.

Several objects in the real world can be modeled as deviations from a tubular structure. Take the head and neck for example. Start with a cylinder, make the cylinder bulge out in certain regions to produce features such as the nose and the lips, and make it cave in to produce the eye sockets and other cavities. Sections of the colon, blood vessels, the bronchial tree and many other organs in the body can be thought of as quasi-tubes.

A tube atom is represented by a cone of spokes. All of the spoke ends lie in a single plane, α, forming a cross-section of the tube. Our approach is to change the cross-section in this plane. This can be accomplished by inclining each spoke in the plane between it and the axis in such a way that the tips of all the spokes continue to lie in the plane α. Changing the cross-sectional shape in this way makes the computation of the shape operator in the circumferential direction straightforward and the component along the axis is still given by the 1D shape operator defined by equation 1 as $\mathbf{U}$ is a function of the spoke angle ϕ.

The above method results in varying length spokes in an atom. Therefore the structure is no longer Blum medial but skeletal. Damon's proofs concerning the use of the shape operator S_{rad} in illegality measurements are valid for skeletal structures too.

The change in the length of the spokes is represented as a multiplicative parameter that belongs to the group $\mathbb{R}^+$. The Log and Exp maps for this parameter have the same form as that for the radius r.

The segmentation step is divided into two scales. At the large scale, a mean of each quasi-tube atom across all population samples is used. This gives a large scale model with each cross-section having a different shape. In the small scale, the individual quasi-tube atoms are allowed to vary to allow for small changes in each cross-sectional shape.

5 Training and Segmentation

We have discussed how we geometrically and statistically model tubes and quasi-tubes. We now discuss their training and the appearance model we use.

To fit quasi-tube models to training cases and derive a probability distribution on that object we follow the general approach described in Merck [21]. Expert outlines designating the target object in several 3D images are converted into binary images. An initial model that somewhat resembles the object is created. Sometimes landmarks are added to initialize our models. The parameters for these models are then varied inside an optimizer. The objective function is set to a sum of a combination of image match and geometric penalty terms. The image match we use is the average of the sum of squared distances between a point on the object's surface and the closest point on the expertly outlined object. In certain places such as the crest, the reverse distance is used as the original distance is artificially low. When landmarks are used, the distance

between them and the corresponding positions on the model are taken into account. The geometric penalty terms consist of a combination of the irregularity, S_{rad} and the curviness penalties mentioned in Sections 2, 2.1 and 2.2 respectively. The weights for the geometric penalties are relaxed proportionally to the quality of the fit of the model.

We then compute the statistical shape model from all the models in the population. The log map of the atoms is computed. The spoke deviations are represented as a multiplicative parameter that belongs to the group $\mathbb{R}^+$. The Log map of the deviations is the Log map for the group $\mathbb{R}^+$ as described in Section 3. We then compute the mean across the entire population to give us the mean shape. The atoms are projected at the tangent plane centered on the mean. The statistical shape model is computed as described in Section 3 except that the spoke deviations are only used in the computation of the mean and not in the computation of the modes of variation of the shape. These comprise the large scale shape variation. The spokes are allowed to vary in the smaller scale. This separation of variation at two different scales gives more stable statistics.

At segmentation time, the appearance model we use has been developed by Broadhurst [17] and Stough [18]. The object's surface is divided into several regions. An intensity histogram for each region is computed and converted into quantile distributions. The distribution of these quantile functions is then analyzed with the help of Principal Component Analysis (PCA). As tubes can be arbitrarily cut-off, we optionally allow for the ends of the tubes to not contribute to these appearance models.

During segmentation, we use the method of posterior optimization where the statistical shape model gives the prior and the appearance model gives the likelihood.

6 Application and Results

Segmenting rectums from 3D CT scans is important for adaptive radiotherapy treatment for prostate cancer. It is important for the patient's health and quality of life that the rectum does not receive too much radiation. The rectum changes shape a lot from day to day due to the presence of gas and faeces pushing the abutting prostate around. The presence of gas also creates a large variation in the intensity distributions and necessitates special handling of the interior intensity distributions in the rectum.

In our experiment the data comes from different days of several patients. For each patient, the statistical shape model is built with the help of data from just that patient. The study is done in a leave-one-out manner, i.e., we build the statistical shape model for each day of a patient by using data from the other days. This is clinically not possible but it suffices for the purposes of our study. In the availability of more training data, one can incorporate cross-patient statistics to get over this limitation.

We obtained 3D CT images with a resolution of $0.98 \times 0.98 \times 3\ \mathrm{mm}^3$ of 5 patients with 13–18 images for each patient for a total of 79 images. We then trained quasi-tubular medial models on manual segmentations of rectums from this data. For our training, we used a hierarchical training process where we first fit a coarsely sampled tubular medial model to the data. We then allowed the individual spokes to vary, generating quasi-tubular models. After this we subdivided the model by generating interpolated atoms halfway between previously existing atoms and then refined this finely

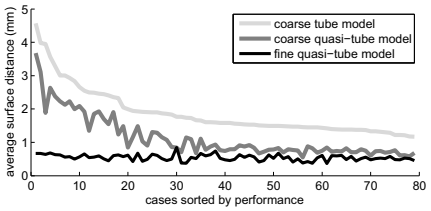


Fig. 6. Average surface distance of the training models at different stages.

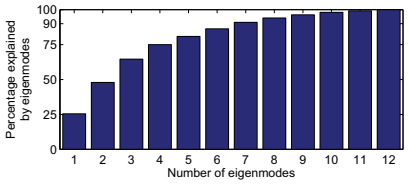


Fig. 7. Graph showing cumulative shape variation captured by the first few eigenmodes: the first 7 modes capture about 90% of the shape variation.

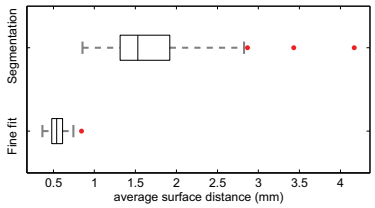


Fig. 8. Boxplot of distribution of segmentation results versus training results. The box goes from 25% to 75% quartile with a line for the median. The two lines at the end go 1.5 times the intra-quartile range (except when there is no more data). The red dots are outliers.

Fig. 9. Each row shows the outline (white or black) of our segmentation on two different axial slices of the same image. Note the poor contrast in the slices in the right column. The slices on the right are inferior with respect to those on the left. The first three rows are typical results and the last row is one of the better segmentations.

sampled model to fit the binary image better. Further subdivision yielded only marginal improvement.

The quantitative results of our training process are shown in Fig. 6. The median average surface distances for the coarse tube, coarse quasi-tube and the fine quasi-tube

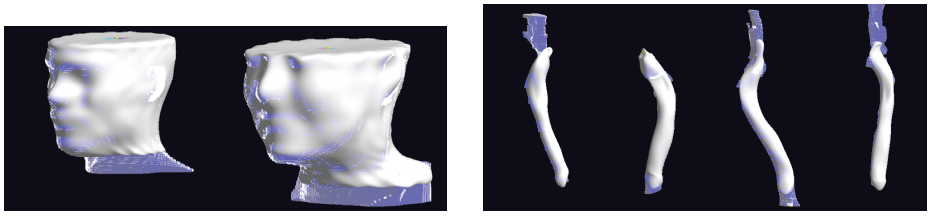


Fig. 10. Quasi-tubular medial models (in white) fit to a section of the head and neck skin surface (left) and the pharynx (right) vs. the manually segmented structures (in translucent blue)

models are 1.58, 0.84 and 0.54 mm with standard deviations of 0.70, 0.62 and 0.09 mm. We can conclude that we can model rectums fairly well and in cases where the tube model is lacking, the quasi-tubular model does a much better job.

In Fig. 7, we show the cumulative shape variation captured by the first few eigenmodes for a quasi-tubular model trained over the data of one of the patients. In about 7 modes, we can effectively capture most of the shape variation. In Fig. 2, we show the variation of the rectum along the first principal direction, which resembles anatomical shape changes due to filling of gas.

During segmentation, we used a similar hierarchical approach. The model was first initialized semi-automatically. We then segmented the model in the shape space of the coarse models. The result was then used to drive the segmentation in the shape space of the finely sampled trained medial models. In Fig. 8, we show the segmentation results versus the quality of fine training fits for all the patients. The median of the segmentation results is 1.53 mm. The segmentation results for a few of these is shown in Fig. 9. Notice the complete lack of contrast in some of the inferior slices.

We have also fit quasi-tubular models to various structures of the head and neck. In Fig. 10, we show quasi-tubular models fit to sections of the skin surface and the pharynx. The expertly contoured outline is shown in translucent blue. The average surface distance for the fitted models of the skin surface, common carotid artery, internal jugular vein and the pharynx are $1.22, 0.8, 1.3$ and 1.13 voxels.

7 Conclusion

We have developed a new method for modeling populations of nearly tubular objects as a tubular medial model with deviations from the perfect tubularity described by local changes and demonstrated the same over real world rectum data. We have seen that this representation is also effective for elongated objects with distinctly noncircular but slowly varying cross-sections.

Further, we show how we can study population variations by doing statistics in the non-linear space in which these quasi-tubular medial models lie. We also discuss the radial shape operator that is needed for studying medial geometry on these models.

The rectum is a challenging organ due to immensely varying shape and poor image contrast. Several structures in the head and neck – skin surface, pharynx, jugular veins

and carotid arteries – provide difficult modeling challenges. We have shown that our models can be trained to within sub-voxel accuracy and give reasonable segmentations.

References

1. Cootes, T.F., Hill, A., Taylor, C.J., Haslam, J.: The use of active shape models for locating structures in medical images. In: IPMI. (1993) 33–47 1
2. Koenderink, J.J.: Solid Shape. MIT Press (1990) 1
3. Binford, T.O.: Visual perception by computer. In: Proceedings of the IEEE Conference on Systems and Control. (1971) 1
4. Ulupinar, F., Nevatia, R.: Shape from contour: straight homogeneous generalized cones. In: Proceedings of Computer Vision. (Dec 1990) 582–586 1
5. Huang, Q., Stockman, G.: Model-based automatic recognition of blood vessels from mr images and its 3D visualization. In: Proceedings of the IEEE International Conference on Image Processing. Volume 3. (Nov 1994) 691–695 1
6. Huang, Q., Stockman, G.C.: Generalized tube model: recognizing 3D elongated objects from 2D intensity images. In: CVPR. (Jun 1993) 104–109 1
7. Terzopoulos, D., Witkin, A., Kass, M.: Constraints on deformable models: recovering 3D shape and nongrid motion. Artif. Intell. **36**(1) (1988) 91–123 2
8. O'Donnell, T., Boult, T., Fang, X.S., Gupta, A.: The extruded generalized cylinder: a deformable model for object recovery. In: CVPR. (21-23 Jun 1994) 174–181 2
9. Damon, J.: Swept regions and surfaces: Modeling and volumetric properties. Theoretical Computer Science **In Press, Accepted Manuscript** 3
10. Aylward, S., Pizer, S., Eberly, D., Bullitt, E.: Intensity ridge and widths for tubular object segmentation and description. MMBIA (1996) 131 3
11. Fridman, Y., Pizer, S.M., Aylward, S., Bullitt, E.: Segmenting 3D branching tubular structures using cores. In: MICCAI. (November 2003) 570–577 3
12. Huang, Q., Stockman, G.: Generalized stochastic tube model: tracking 3D blood vessels in mr images. In: ICPR. Volume 2. (9-13 Oct 1994) 156–160 3
13. de Bruijne, M., van Ginneken, B., Viergever, M., Niessen, W.: Adapting active shape models for 3D segmentation of tubular structures in medical images. IPMI **2732** (2003) 136–147 3
14. Joshi, S., Pizer, S., Fletcher, P., Yushkevich, P., Thall, A., Marron, J.: Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. IEEE Transactions on Medical Imaging **21**(5) (May 2002) 538–550 3
15. Pizer, S., Fletcher, T., Fridman, Y., Fritsch, D., Gash, A., Glotzer, J., Joshi, S., Thall, A., Tracton, G., Yushkevich, P., Chaney, E.: Deformable m-reps for 3D medical image segmentation. IJCV - Special UNC-MIDAG issue **55**(2) (Nov-Dec 2003) 85–106 3
16. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.: Principal geodesic analysis for the study of nonlinear statistics of shape. Medical Imaging, IEEE Transactions on **23** (2004) 995–1005 3, 7
17. Broadhurst, R., Stough, J., Pizer, S., Chaney, E.: A statistical appearance model based on intensity quantiles histograms. ISBI (2006) 3, 9
18. Stough, J., Broadhurst, R.E., Pizer, S.M., Chaney, E.L.: Regional appearance in deformable model segmentation. In: IPMI. (2007) 3, 9
19. Damon, J.: Determining the geometry of boundaries of objects from medial data. International Journal of Computer Vision **63** (June 2005) 45–64 4, 6
20. Han, Q., Pizer, S.M., Damon, J.N.: Interpolation in discrete single figure medial objects. In: Proceedings of Computer Vision and Pattern Recognition. (2006) 85 6
21. Merck, D., Tracton, G., Saboo, R., Levy, J., Chaney, E., Pizer, S., Joshi, S.: Training models of anatomic shape variability. Medical Physics **35**(8) (2008) 3584–3596 8

Joint Segmentation via Patient-Specific Latent Anatomy Model

T. Riklin Raviv[1], B.H. Menze[1,5], K. Van-Leemput[1,2,3], B. Stieltjes[6], M.A. Weber[6,7], N. Ayache[5], W. M. Wells III[1,4] and P. Golland[1]

[1] Computer Science and Artificial Intelligence Laboratory, MIT, USA
[2] Department of Information and Computer Science, Helsinki University of Technology, Finland
[3] Department of Radiology, MGH, Harvard Medical School, USA
[4] Brigham and Womens Hospital, Harvard Medical School, USA
[5] Asclepios Research Project, INRIA Sophia Antipolis, France
[6] German Cancer Research Center (DKFZ), Heidelberg, Germany
[7] Diagnostic Radiology, University Hospital, Heidelberg, Germany

Abstract. We present a generative approach for joint 3D segmentation of patient-specific MR scans across different modalities or time points. The latent anatomy, in the form of spatial parameters, is inferred simultaneously with the evolution of the segmentations. The individual segmentation of each scan supports the segmentation of the group by sharing common information. The joint segmentation problem is solved via a statistically driven level-set framework. We illustrate the method on an example application of multimodal and longitudinal brain tumor segmentation, reporting promising segmentation results.

Key words: patient-specific latent anatomy, spatial parameters, tumor segmentation, level-set framework

1 Introduction

Modeling patient-specific anatomy is essential in longitudinal studies and pathology detection. We present a generative approach for joint segmentation of MR scans of a specific subject, where the latent anatomy, in the form of spatial parameters is inferred concurrently with the segmentation. While the methodology can be applied to a variety of applications, here we focus on segmentation of pathological tissues. Specifically, we demonstrate our algorithm on a problem of multimodal segmentation of brain tumors in longitudinal studies. Patient-specific datasets acquired through different modalities at a particular time point are segmented simultaneously, yet individually, based on the specific parameters of their intensity distributions. The spatial parameters that are shared among the scans facilitate the segmentation of the group.

While probabilistic atlases are commonly used as priors for segmentation of MR scans of normal tissues or structures [1, 8, 20–22, 31] the standard methods fail to handle pathologies or to detect subtle anatomical deformation, for example, due to aging. A few methods generate patient-specific atlases by iteratively refining the normal template model [13, 17, 24]. Other methods detect tumors

from differences in images acquired at different time points [26, 29]. Both approaches rely heavily on priors such as tumor shape, intensities, growth and expected evolution [5, 13, 16, 17, 24, 26, 29, 34]. Discriminative approaches [4, 9–11, 25, 32, 33] segment lesions by constructing feature distributions that characterize healthy subjects, so that the pathology can be specified as outliers. However, the variability of normal brain scans and the effects some tumors have on their surrounding healthy tissues lead to a high false positive detection rate. Moreover, mild anomalies can be wrongly classified as normal.

Here we propose and demonstrate a fully automatic groupwise segmentation method. No prior knowledge or external information is required but a couple of mouse clicks at approximately the center and the boundary of a single tumor slice (out of the few dozen volumes to segment) that are used to initialize the segmentations of the images acquired at the first time point. All model parameters, spatial and intensity, are inferred from the patient scans alone. Tumor segmentations at a given time point are used to initialize the segmentations at the next time point for scans of corresponding modalities. The output of the algorithm consist of individual segmentations for each modality and time point. This is in contrast to many discriminative methods, e.g., [32], that use multimodal datasets for multivariate feature extraction, assuming spatial coherence of the tumor outlines in different image modalities. Here we relax this assumption and search for systematic, structural differences of the visible tumor volume acquired by different imaging protocols.

Our latent anatomy segmentation model is based on probabilistic principles but is solved using partial differential equations (PDEs) and energy minimization criteria. We describe a statistically-driven level-set algorithm that expresses segmentation uncertainty via the logistic function of the associated level-set values, similar to [23]. We relate the image likelihood term to the region based constraint that relaxes the piecewise smoothness assumption of [18], in the spirit of [3, 19, 35]. We also draw the connection between a Markov random field (MRF) prior on the individual segmentations and two continuous-form energy terms: the commonly used smoothness constraint, originally proposed in [12] and the spatial constraint, associated with the latent anatomy parameters. We developed this approach in [27] and validated the algorithm on joint segmentation of cortical and subcortical structures in a population. Here we investigate its application to a patient-specific tumor data set.

The paper is organized as follows. Section 2 defines the problem of latent-anatomy segmentation. In Section 3 we derive our level-set framework for fitting the probabilistic model to the image data. The alternating minimization algorithm is presented in Section 4. Section 5 reports the experimental results followed by a discussion in Section 6.

2 Problem definition and probabilistic model

This section summarizes the formulation of [27] for the joint segmentation of N aligned images. The images can, for example, represent N scans of a specific

patient acquired via different imaging protocols. Our objective is to segment a particular region of interest, a brain lesion for example, that may appear slightly differently across the images. Let $I_n \colon \Omega \to \mathbb{R}^+$, be a gray level image with V voxels, defined on $\Omega \subset \mathbb{R}^3$ and let $\Gamma_n \colon \Omega \to \{0,1\}$ be the unknown segmentation of the image I_n, $n = 1, \ldots, N$. We assume that each segmentation Γ_n is generated iid from a probability distribution $p(\Gamma;\ \theta_\Gamma)$ where θ_Γ is the set of the unknown spatial parameters. We also assume that Γ_n generates the observed image I_n, independently of all other image-segmentation pairs, with probability $p(I_n|\Gamma_n;\theta_{I,n})$ where $\theta_{I,n}$ are the parameters corresponding to image I_n. Since the images are acquired by different imaging protocols we assign a different set of intensity parameters to each of them.

Let $\{I_1 \ldots I_N\}$ be the given set of aligned images that form the observed variable in our problem and let $\Gamma = \{\Gamma_1, \ldots, \Gamma_N\}$ be the corresponding unknown segmentations. The joint distribution $p(I_1 \ldots I_N, \Gamma_1 \ldots \Gamma_N; \Theta)$ is governed by the composite set of parameters $\Theta = \{\theta_\Gamma, \theta_{I,1} \ldots \theta_{I,N}\}$. Our goal is to estimate the segmentations Γ. This, however, cannot be accomplished in a straightforward manner since the model parameters are also unknown. We therefore jointly optimize Γ and Θ:

$$\{\hat{\Theta}, \hat{\Gamma}\} = \arg \max_{\{\Theta,\Gamma\}} \log p(I_1 \ldots I_N, \Gamma_1 \ldots \Gamma_N; \Theta) \tag{1}$$

$$= \arg \max_{\{\Theta,\Gamma\}} \sum_{n=1}^{N} \left[\log p(I_n|\ \Gamma_n; \theta_{I,n}) + \log p(\Gamma_n; \theta_\Gamma)\right]. \tag{2}$$

We alternate between estimating the maximum a posteriori (MAP) segmentations and updating the model parameters. For a given setting of the model parameters $\hat{\Theta}$, Eq. (2) implies that the segmentations can be estimated by solving N separate MAP problems:

$$\hat{\Gamma}_n = \arg \max_{\Gamma_n} \left[\log p(I_n|\ \Gamma_n; \theta_{I,n}) + \log p(\Gamma_n;\ \theta_\Gamma)\right]. \tag{3}$$

We then fix $\hat{\Gamma}$ and estimate the model parameters $\Theta = \{\theta_\Gamma, \theta_{I,1}, \ldots \theta_{I,N}\}$ by solving two ML problems:

$$\hat{\theta}_{I,n} = \arg \max_{\theta_{I,n}} \log p(I_n;\ \Gamma_n, \theta_{I,n}), \tag{4}$$

$$\hat{\theta}_\Gamma = \arg \max_{\theta_\Gamma} \sum_{n=1}^{N} \log p(\Gamma_n;\ \theta_\Gamma). \tag{5}$$

3 Probabilistic view of the level-set framework

Now we draw the connection between the probabilistic model presented above and a level-set framework for segmentation. Let $\phi_n \colon \Omega \to \mathbb{R}$ be the level-set function associated with image I_n. The zero level $C_n = \{\mathbf{x} \in \Omega|\ \phi_n(\mathbf{x}) = 0\}$

defines the interface that partitions the image space of I_n into two disjoint regions ω and $\Omega \setminus \omega$. Similar to [20, 23] we define the level-set function ϕ_n using the log-odds formulation instead of the conventional signed distance function:

$$\phi_n(\mathbf{x}) \triangleq \epsilon \ \text{logit}(p) = \epsilon \log \frac{p(\mathbf{x} \in w)}{1 - p(\mathbf{x} \in \omega)} = \epsilon \ \log \frac{p(\mathbf{x} \in \omega)}{p(\mathbf{x} \in \Omega \setminus \omega)}, \tag{6}$$

where $p(\mathbf{x} \in \omega)$ can be viewed as the probability that the voxel in location $\mathbf{x}$ belongs to the foreground region. The constant ϵ determines the scaling of the level-set function ϕ_n with respect to the ratio of the probabilities. The inverse of the logit function for $\epsilon = 1$ is the logistic function:

$$H_\epsilon(\phi_n) = \frac{1}{2}\left(1 + \tanh\left(\frac{\phi_n}{2\epsilon}\right)\right) = \frac{1}{1 + e^{-\phi_n/\epsilon}}. \tag{7}$$

Note, that $H_\epsilon(\phi_n)$ is similar, though not identical, to the regularized Heaviside function introduced by Chan and Vese [3]. We use this form of Heaviside function and its derivative with respect to ϕ in the proposed level-set formulation. To simplify the notation, we omit the subscript ϵ in the rest of the paper.

3.1 Cost functional for segmentation

The joint estimation problem of the hidden variables Γ, or ϕ_n (using the level-set notation) and the unknown model parameters Θ can be solved as an energy minimization problem, where

$$E(\phi_n) = -\log p(I_n|\ \Gamma_n; \theta_{I,n}) - \log p(\Gamma_n; \theta_\Gamma).$$

As in [27], we establish the correspondence between the log probability and the level-set energy terms. Let $E_I(\phi_n, \Theta)$ denote the term corresponding to the image likelihood in Eq. (3). Then

$$\begin{aligned} E_I(\phi_n, \Theta) = -\int_\Omega & [\log p_{\text{in}}(I_n; \theta_{I,n}) H(\phi_n(\mathbf{x})) \\ & + \log p_{\text{out}}(I_n; \theta_{I,n})\,(1 - H(\phi_n(\mathbf{x})))]\, d\mathbf{x}, \end{aligned} \tag{8}$$

where, p_{in} and p_{out} denote the probability distributions of the foreground and background intensities of a particular image I_n, respectively. If we use, for example, Gaussian densities for p_{in} and p_{out} we get the familiar minimal variance term [3, 19, 35]. Here, we use a Gaussian mixture to model the background, as described later in the paper.

Let us now consider the prior probability $p(\Gamma_n;\ \theta_\Gamma)$ in Eq. (2) and its corresponding energy terms. Specifically, we construct an MRF prior for segmentations:

$$\begin{aligned} \log p(\Gamma_n; \theta_\Gamma) = & \sum_{v=1}^{V} [\Gamma_n^v \log(\theta_\Gamma^v) + (1 - \Gamma_n^v) \log(1 - \theta_\Gamma^v)] \\ & - \sum_{v=1}^{V} f(\Gamma_n^v, \Gamma_n^{\mathcal{N}(v)}) - \log Z(\theta_\Gamma), \end{aligned} \tag{9}$$

where $Z(\theta_\Gamma)$ is the partition function and $\mathcal{N}(v)$ is the set of the closest neighbors of voxel v. The function

$$f(\Gamma_n^v, \Gamma_n^{\mathcal{N}(v)}) = \sum_{v' \in \mathcal{N}(v)} w_{(v,v')}(\Gamma_n^v - \Gamma_n^{v'})^2,$$

accounts for the interactions between neighboring voxels. It can be configured, by setting the values of $w_{(v,v')}$, to act as a finite difference operator approximating the gradient of Γ_n at the voxel v [15]. This approximation allows us to represent the discrete term $\sum_{v=1}^{V} f(\Gamma_n^v, \Gamma_n^{\mathcal{N}(v)})$ as an approximation of the continuous term

$$E_{\text{LEN}}(\phi_n) = \int_\Omega |\nabla H(\phi_n(\mathbf{x}))| d\mathbf{x}, \tag{10}$$

which is the commonly used length term. Note that if we omit the pairwise term in Eq. (9), the prior on segmentations $p(\Gamma_n|\theta_\Gamma)$ reduces to a Bernoulli distribution, where the parameters θ_Γ represent the probability map for the structure of interest. The introduction of the pairwise clique potentials complicates the model but encourages smoother labeling configurations.

We define the spatial energy term E_S based on the singleton term in Eq. (9). Using the level-set formulation we obtain:

$$E_S(\phi_n, \Theta) = -\int_\Omega \left[\log \theta_\Gamma(\mathbf{x}) H(\phi_n(\mathbf{x})) + \log(1 - \theta_\Gamma(\mathbf{x}))\,(1 - H(\phi_n(\mathbf{x})))\right] d\mathbf{x}. \tag{11}$$

Note, that ignoring the partition function in the equations that follow Eq. (9) has no effect on the estimation of Eq. (3), but it changes Eq. (5) to be maximum pseudo likelihood [2], rather than maximum likelihood.

We construct the cost functional for $\phi_1 \dots \phi_N$ and the parameters Θ by combing Eq. (8), (10) and (11):

$$E(\phi_1 \dots \phi_N, \Theta) = \gamma E_{\text{LEN}} + \beta E_I + \alpha E_S \tag{12}$$

where $\alpha = 1 - \beta - \gamma$. As in [28], we adaptively tune the weights such that the contributions of the energy terms E_{LEN}, E_I and E_S to the overall cost are balanced.

4 Gradient descent and parameter estimation

We optimize Eq. (12) by a set of alternating steps. For fixed model parameters Θ, the evolution of each level-set function ϕ_n is determined by the following gradient descent equation:

$$\phi_n(\mathbf{x}, t + \Delta t) = \phi_n(\mathbf{x}, t) + \frac{\partial \phi_n}{\partial t} \Delta t, \tag{13}$$

where $\frac{\partial \phi_n}{\partial t}$ is obtained from the first variation of $E(\phi_n, \Theta)$. Using the Euler-Lagrange equations we get:

$$\begin{aligned} \frac{\partial \phi_n}{\partial t} = \delta(\phi_n) \Big\{ & \gamma \text{ div } \left(\frac{\nabla \phi_n}{|\nabla \phi_n|}\right) + \beta \left[\log p_{\text{in}}(I_n(\mathbf{x}); \theta_{I,n}) - \log p_{\text{out}}(I_n(\mathbf{x}); \theta_{I,n})\right] \\ & + \alpha \left[\log \theta_\Gamma - \log(1 - \theta_\Gamma)\right] \Big\}, \end{aligned} \tag{14}$$

where $\delta(\phi_n)$ is the derivative of $H(\phi_n)$ with respect to ϕ_n:

$$\delta_\epsilon(\phi_n) = \frac{1}{2\epsilon}\text{sech}(\frac{\phi_n}{2\epsilon}) = \frac{1}{\epsilon\cosh(\frac{\phi_n}{\epsilon})}.$$

For fixed segmentations ϕ_n, the model parameters are recovered by differentiating the cost functional in Eq. (12) with respect to each parameter.

4.1 Intensity parameters

We assume that the intensities of the structure of interest are drawn from a normal distribution, i.e., $p_{\text{in}}(I_n;\theta_{I,n}) = \mathcal{N}(I_n;\mu_n,\sigma_n^2)$. The intensities of the background tissues are modeled as a K-Gaussian mixture:

$$p_{\text{out}}(I_n;\theta_{I,n}) = \text{GMM}(\mu_n^1\cdots\mu_n^K,\sigma_n^1\cdots\sigma_n^K,\lambda_n^1\cdots\lambda_n^K),$$

where λ_n^k is the mixing proportion of component k in the mixture. We estimate the Gaussian mixture model parameters using the expectation maximization (EM) method [6].

4.2 Spatial parameters

We estimate the spatial function $\theta_\Gamma(\mathbf{x})$, which represents a dynamically evolving latent atlas, by optimizing the sum of the energy terms that depend on θ_Γ :

$$\hat{\theta}_\Gamma = \arg\max_{\theta_\Gamma}\sum_{n=1}^{N}\int_\Omega[\tilde{H}(\phi_n(\mathbf{x}))\log(\theta_\Gamma(\mathbf{x})) + (1-\tilde{H}(\phi_n(\mathbf{x})))\log(1-\theta_\Gamma(\mathbf{x}))]d\mathbf{x},$$

yielding

$$\hat{\theta}_\Gamma(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}\tilde{H}(\phi_n(\mathbf{x})). \tag{15}$$

4.3 Algorithm

We summarize the proposed latent-anatomy segmentation algorithm assuming the following setup. The input consist of N aligned volumes $\{I_{\tau,m}\}$, where $I_{\tau,m}$ is a volume acquired at time τ and modality m.

Initialization The user selects one of the volumes acquired at the first time point and identifies a single sagittal, axial or coronal slice where the tumor or the structure of interest is clearly seen. The user marks with a couple of mouse clicks the approximate location of the tumor center and one of its boundary points. This input determines a sphere that is used to initialize the segmentations of all the volumes acquired at the first time point. We denote their number by M_1. In our implementation M_1 identical level-set functions that are defined by the signed distance function of this sphere are used for initialization.

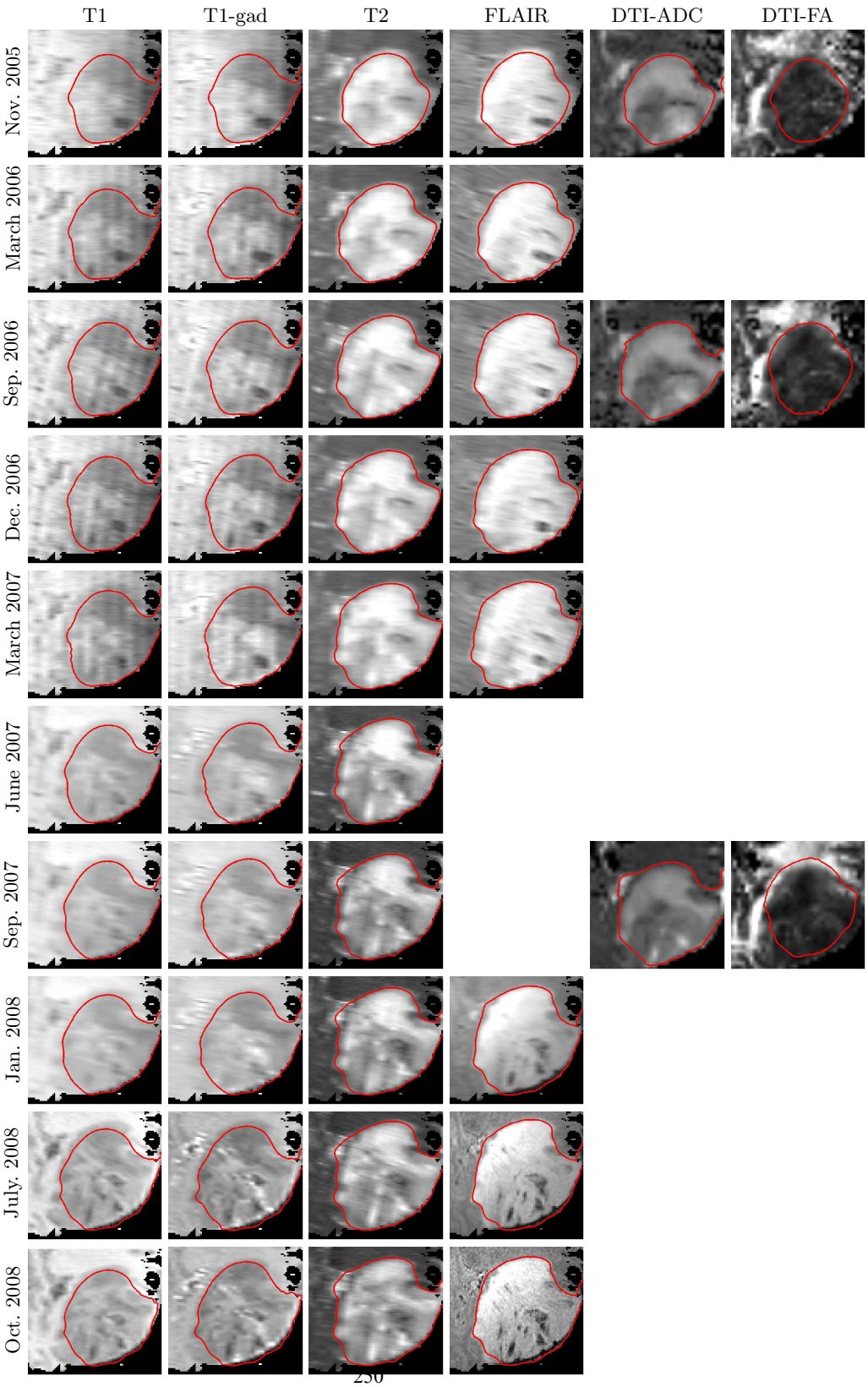


Fig. 1. Axial slice of the tumor volumes and the automatic 3D segmentations (red outlines) across 6 modalities and 10 time points. Not all the modalities were acquired at each time point.

1. Set $\tau = 1$. Note that τ is the index of the actual time in which the scans were acquired. It should not be confused with t in Eq. (13) that denotes the number of gradient descent iterations.
2. Calculate the background and foreground intensity parameters, for each of the volumes acquired at time τ based on the current estimates of their corresponding level-set functions ϕ_n, according to subsection 4.1.
3. Calculate the latent anatomy parameters θ_Γ based on the current estimate of the level-set functions ϕ_n, corresponding to the image volumes acquired at time τ (Eq. (15)).
4. Use Eq. (14) (gradient descent) to evolve the level-set functions associated with the volumes acquired at time τ based on the current estimates of the respective intensity parameters θ_I and the spatial parameters θ_Γ.
5. Repeat steps 2-4 until convergence.
6. Use the final state of the level-set functions ϕ_n associated with the volumes acquired by modality m at time τ to initialize the corresponding level-set associated with the volumes acquired at time $\tau + 1$. Set $\tau := \tau + 1$
7. Repeat steps 2-6 sequentially, for all the time points.

5 Experimental Results

We applied the proposed method to a set of 44 image volumes of a patient with histologically confirmed low-grade glioma, acquired at 10 different time points at the German Cancer Research Center (Heidelberg, Germany) using 1.5T Siemens Magnetom and 3T Siemens TRIO MR scanners. The volumes were acquired via six imaging protocols: T1, T2, FLAIR, DTI, and contrast-enhanced T1 sequences (T1gd). We note that not all acquisition modalities were used at each time point, as illustrated in Fig. 1. We aligned the images using the MedINRIA registration software [30]. We calculated fractional anisotropy (FA) and apparent diffusion coefficient (ADC) maps from the diffusion tensor images (DTI) using the same software. To enable quantitative evaluation, three manual segmentations of three orthogonal slices that pass through the center of the tumor were provided for each volume.

Fig. 1 presents axial slices of the available image volumes together with the boundaries of the automatic 3D segmentation. Fig. 2 shows the manual segmentations for three lateral slices through the tumor, together with the automatic segmentation.

Table 1 provides quantitative evaluation of the overlap between the automatic and the manual segmentations as measured by the Dice coefficients [7]. We compared the automatic segmentations with the corresponding triplets of manually segmented slices. The first number in each cell in Table 1 reports the mean and the standard deviation of these nine Dice scores. The second number in each cell in Table 1 reports the mean and the standard deviation of the Dice scores obtained by comparing one of the manual segmentations with the other two in the three slices. We marked with asterisks cells that show similar Dice scores. The overall average Dice score for the automatic segmentation is

Time of acquisition	**T1**	**T1gd**	**T2**	**Flair**	**DA**	**ADC**
Nov. 2005	.71±.11*	.49±.14*	.87±.01	.91±.03	.84±.02	.73±.08
	.80±.07	.52±.15	.95±.01	.94±.02	.94±.01	.91±.07
March 2006	.78±.04	.77±.09*	.93±.02*	.84±.02		
	.92±.02	.85±.06	.95±.02	.96±.01		
Sep. 2006	.87±.03*	.85±.08*	.84±.02	.94±.01*	.88±.03	.81±.04
	.90±.04	.82±.12	.95±.02	.94±.02	.94±.02	.93±.03
Dec. 2006	.87±.02	.89±.03*	.93±.01	.86±.02		
	.91±.03	.90±.04	.95±.01	.94±.01		
March 2007	.84±.02	.82±.11*	.93±.02*	.84±.02		
	.91±.04	.87±.11	.94±.03	.94±.01		
June 2007	.81±.09*	.85±.09*	.93±.02*			
	.86±.11	.81±.11	.93±.03			
Sep. 2007	.87±.04	.84±.08*	.87±.02		.86±.03	.87±.02
	.92±.03	.86±.07	.94±.02		.93±.01	.94±.02
Jan. 2008	.88±.02*	.89±.02*	.91±.03*	.83±.04		
	.90±.01	.90±.02	.91±.01	.94±.02		
July 2008	.87±.03	.85±.02	.91±.03	.86±.03		
	.93±.02	.93±.02	.94±.01	.94±.02		
Oct. 2008	.88±.03	.85±.03	.91±.03*	.84±.04		
	.93±.03	.93±.01	.94±.01	.93±.02		

Table 1. Dice coefficients for 44 volumes in the study. The first number in each cell reports the mean and the standard deviation of the Dice scores of the automatic segmentation with respect to three manual segmentations. The second number in each cell reports the mean and the standard of the Dice scores calculated between one of the manual segmentations and the average of the other two (see text for detail). Automatic segmentations that did not differ significantly from the manual ones are marked by the asterisk.

above 0.85 while the average Dice scores obtained for the manual segmentations is 0.91. The top plot in Fig. 3 presents the average Dice score over all modalities at a given time point obtained by our method (red) and the Dice scores of the multivariate tissue classification [32] (green). The plot shows that the Dices scores obtained via the latent anatomy method are consistently higher. The bottom plot in Fig. 3 compares the overlap among the manual segmentations for each individual modality ('intra-modal') with the overlap among the manual segmentations for all the modalities together ('inter-modal'). We define overlap as the mean Dice score among the three manual segmentations, as described above. This plot suggests that even the manual segmentations vary significantly across modalities for the same time point, justifying our approach of generating separate tumor segmentation for each volume.

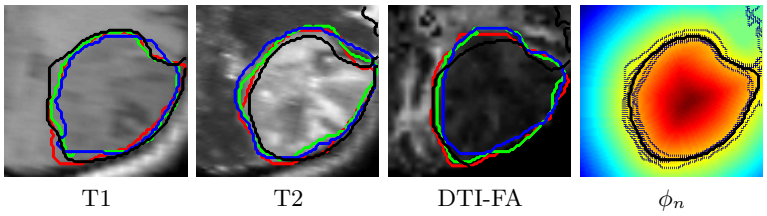


Fig. 2. Manual segmentations (red, green, blue) and automatic segmentation (black) for lateral T1, T2 and DTI-FA images acquired at the same time point. The forth image show the corresponding section of the average of the associated 3D level-set functions. The black line indicates the zero level. Gray dashed lines indicate the tumor boundaries of all the modalities available for that time point.

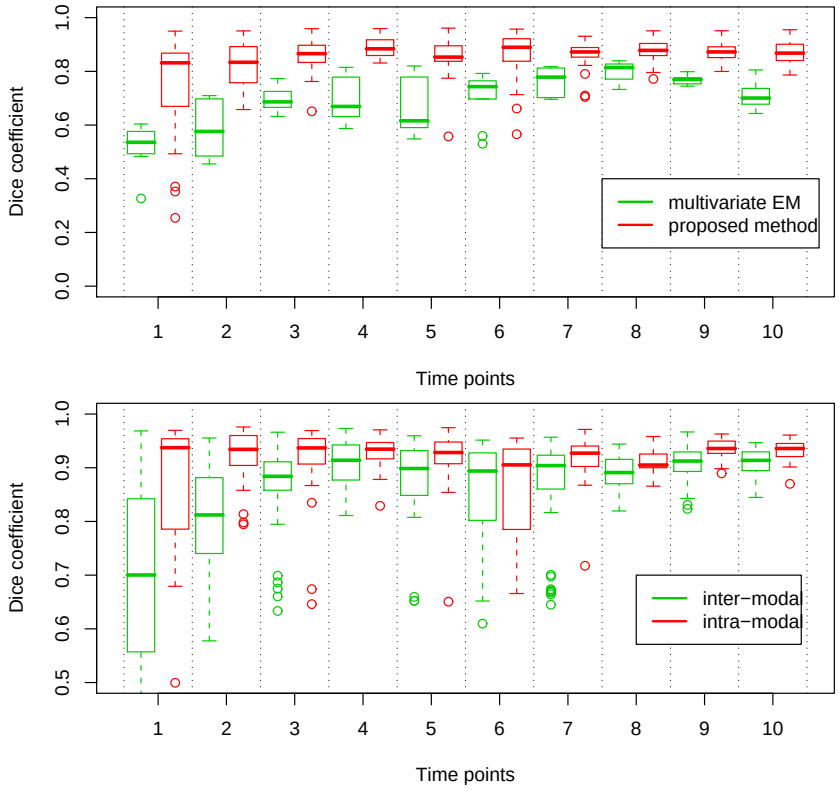


Fig. 3. Top: A comparison of the average Dice scores of the proposed latent anatomy method (red) and the Dice scores of the multivariate EM for lesion segmentation of [32] (green). Note that the segmentation results obtained by the proposed latent anatomy method are consistently better. Bottom: comparison of the correspondence between the manual segmentations for each individual modality (red) with the correspondence between the manual segmentations for all the modalities together (green). Correspondence, here, is defined as the mean Dice score between the three manual segmentations (see text for detail).

6 Discussion and future directions

We presented a statistically driven level-set approach for joint segmentation of subject-specific MR scans. The latent patient anatomy, which is represented by a set of spatial parameters is inferred from the data simultaneously with the segmentation through an alternating minimization procedure. Segmentation of each of the channels, or modalities, is therefore supported by the common information shared by the group. The method is demonstrated by addressing the problem of multi-modal brain tumor segmentation across $5 - 10$ time points. Promising segmentation results were obtained. An on-going research is now conducted to construct a functional model of the tumor growth based on the image sequences [14] that will used to predict the evolution of the tumor outlines.

Acknowledgments. This work was supported in part by the Leopoldina Fellowship Programme (LPDS 2009-10), NIH NIBIB NAMIC U54-EB005149, NIH NCRR NAC P41-RR13218, NIH NINDS R01-NS051826 grants and NSF CAREER Award 0642971.

References

1. J. Ashburner and K. Friston. Unified segmentation. *NeuroImage*, 26:839–851, 2005.
2. J. Besag. Statistical analysis of non-lattice data. *The Statistician*, 24(3):179–195, 1975.
3. T.F. Chan and L.A. Vese. Active contours without edges. *IEEE TIP*, 10(2):266–277, 2001.
4. Cobzas *et al.* 3D variational brain tumor segmentation using a high dimensional feature set. In *ICCV*, 1–8, 2007.
5. Cuadra*et al.* Atlas-based segmentation of pathological brain MR images. *IEEE TMI*, 23:1301–1314, 2004.
6. A. Dempster, N. Laird, and D. Rubin. Maximal likelihood form incomplete data via the EM algorithm. *Proceedings of the Royal Statistical Society*, 39:1 – 38, 1977.
7. L. Dice. Measure of the amount of ecological association between species. *Ecology*, 26(3):297–302, 1945.
8. Fischl *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.
9. D. Gering, W.E.L. Grimson, and R. Kikinis. Recognizing deviations from normalcy for brain tumor segmentation. In *MICCAI*, vol. 2488, 388 – 395, 2002.
10. Grlitz *et al.* Semi-supervised tumor detection in magnetic resonance spectroscopic images using discriminative random fields, 2007.
11. S. Ho, E. Bullitt, and G. Gerig. Level-set evolution with region competition: automatic 3-D segmentation of brain tumors. In *ICPR*, vol. 1, p532–535, 2002.
12. M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, Jan. 1988.
13. Kaus *et al.* Automated segmentation of MR images of brain tumors. *Radiology*, 218:586 – 591, 2001.
14. Konukoglu *et al.* A recursive anisotropic fast marching approach to reaction diffusion equation: Application to tumor growth modeling. In *IPMI*, p687–699, 2007.
15. S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.
16. Mohamed *et al.* Deformable registration of brain tumor images via a statistical model of tumor-induced deformation. *Medical Image Analysis*, 10:752–763, 2006.

17. Moon *et al.* Model-based brain and tumor segmentation. In *ICPR*, vol. 1, 528–531, 2002.
18. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.
19. N. Paragios and R. Deriche. Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. *JVCIR*, 13:249–268, 2002.
20. K.M. Pohl and R. Kikinis and W.M. Wells. Active Mean Fields: Solving the Mean Field Approximation in the Level Set Framework. *IPMI*, 4584:26–37, 2007.
21. Pohl *et al.* A bayesian model for joint segmentation and registration. *NeuroImage*, 31(1):228 – 239, 2006.
22. Pohl *et al.* A hierarchical algorithm for MR brain image parcellation. *TMI*, 26(9):1201–1212, 2007.
23. Pohl *et al.* Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis*, 11(6):465–477, 2007.
24. Prastawa *et al.* Automatic brain tumor segmentation by subject specific modification of atlas priors. *Academic Radiology*, 10:1341–1348, 2003.
25. Prastawa *et al.* A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis*, 8:275–283, 2004.
26. Rey *et al.* Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. *Medical Image Analysis*, 6(2):163–179, 2002.
27. Riklin Raviv *et al.* Joint segmentation of image ensembles via latent atlases. In *MICCAI*, 2009. accepted.
28. T. Riklin-Raviv, N. Sochen, and N. Kiryati. Shape-based mutual segmentation. *International Journal of Computer Vision*, 79:231–245, 2008.
29. J.P. Thirion and G. Calmon. Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences. *IEEE TMI*, 18(5):429–441, 1999.
30. N. Toussaint, J.C. Souplet, and P. Fillard. Medinria: Medical image navigation and research tool by inria. In *Proc. of MICCAI Workshop on Interaction in medical image analysis and visualization*, 2007.
31. Van Leemput *et al.* Automated model-based tissue classification of MR images of the brain. *IEEE TMI*, 18(10):897–908, 1999.
32. Van Leemput *et al.* Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE TMI*, 20:677–688, 2001.
33. Wels *et al.* A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3-D MRI. In *MICCAI*, vol. 5241, 67 – 75, 2008.
34. Zacharaki *et al.* Orbit: A multiresolution framework for deformable registration of brain tumor images. *IEEE TMI*, 27(8):1003–1017, 2008.
35. S.C. Zhu and A.L. Yuille, Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *PAMI*,18(9):884–900, 1996.

Investigating the uncertainty in multi-fiber estimation in High Angular Resolution Diffusion Imaging

Liang Zhan[1] , *Alex D. Leow[1] , Marina Barysheva[1], Albert Feng[1], Arthur W. Toga[1], Guillermo Sapiro[2], Noam Harel[2], Kelvin O. Lim[2], Christophe Lenglet[2], Katie L. McMahon[3], Greig I. de Zubicaray[3], Margaret J. Wright[4], Paul M. Thompson[1]

[1] Laboratory of Neuroimaging, Department of Neurology, University of California, Los Angeles, USA
[2] Center for Magnetic Resonance Research, University of Minnesota, USA
[3] Functional MRI Laboratory, Centre for Magnetic Resonance, University of Queensland, Brisbane, Australia
[4] Queensland Institute of Medical Research, Brisbane, Australia

Abstract. In this paper, we investigated the reconstruction accuracy and information uncertainty in multi-fiber estimation to better understand the trade-off between scanning time and angular precision in High Angular Resolution Diffusion Imaging (HARDI). Reconstruction accuracy was measured using the Kullback-Leibler divergence (sKL) on the orientation density functions (ODFs) first in simulations with varying b-values and variable additive Rician noise. ODFs were computed analytically from tensor distribution functions (TDFs) which model the HARDI signal at each point as a unit-mass probability density on the 6D manifold of symmetric positive definite tensors. Reconstruction accuracy rapidly increased with additional gradients at lower SNR. The information uncertainty was quantified by the Exponential Isotropy (EI), a TDF-derived measure of fiber integrity that exploits the full multidirectional HARDI signal. Simulations and empirical results both found that information uncertainty decreased as angular resolution increased, and plateaued at around 70~80 gradients. Furthermore, in high magnetic field (7 Tesla) HARDI, the reconstruction accuracy and information uncertainty index decreased at higher b-values.

Keywords: High Angular Resolution Diffusion Imaging, Tensor Distribution Function, multi-fiber reconstruction, Kullback-Leibler divergence, Exponential Isotropy

1 Introduction

Diffusion weighted MR imaging is a powerful tool to study water diffusion in tissue, providing vital information on white matter microstructure, such as fiber connectivity and composition in the healthy and diseased brain. To date, most diffusion imaging studies (especially in clinical applications) still employ the diffusion tensor imaging (DTI) model [1], which describes the anisotropy of water diffusion in tissues by

* Correspondence should be addressed to feuillet@ucla.edu.

estimating, from a set of *K* diffusion-sensitized images, the 3x3 diffusion tensor (the covariance matrix of a 3-dimensional Gaussian distribution). Each voxel's signal intensity in the *k*-th image is attenuated, by water diffusion, according to the Stejskal-Tanner equation [2]: $S_k = S_0 \exp[-b\mathbf{g}_k^T\mathbf{D}\mathbf{g}_k]$, where S_0 is the non-diffusion weighted signal intensity, $\mathbf{D}$ is the 3x3 diffusion tensor, $\mathbf{g}_k$ is the direction of the diffusion gradient and b is Le Bihan's factor containing information on the pulse sequence, gradient strength, and physical constants. Although 7 gradients are mathematically sufficient to determine the diffusion tensor, MRI protocols with higher angular and radial resolutions, such as the high-angular resolution diffusion imaging (HARDI) or diffusion spectrum imaging (DSI) techniques, can resolve more complex diffusion geometries that a single-tensor model, as employed in standard DTI, fails to capture, e.g., fiber crossings and intermixing of tracts.

Recent technical advances have made HARDI more practical. A 14-minute scan can typically sample over 100 angles (with 2 mm voxels at 4 Tesla). HARDI's improved signal-to-noise ratio may be used to reconstruct fiber pathways in the brain with extraordinary angular detail, identifying anatomical features, connections and disease biomarkers not seen with conventional MRI. If more angular detail is available, fiber orientation distribution functions (ODFs) may be reconstructed from the raw HARDI signal using the Q-ball imaging technique [3]. Deconvolution methods [4,5] have also been applied to HARDI signals, yielding mathematically rich models of fiber geometries as probabilistic mixtures of tensors [6], fields of von Mises-Fisher mixtures [7], or higher-order tensors (i.e., 3x3x…x3 tensors) [8,9]. Stochastic tractography [10, 11] can also exploit HARDI's increased angular detail, and fluid registration methods have also been developed to align HARDI ODFs using specialized Riemannian metrics [12]. In most deconvolution-based methods, however, restrictive prior assumptions are typically imposed on the allowable fibers, e.g., all fiber tracts are considered to have the same anisotropy profile.

A novel approach, the Tensor Distribution Function (TDF), was recently proposed by Leow et al. in [13] to model multidirectional diffusion at each point as a probabilistic mixture of all symmetric positive definite tensors. The TDF models the HARDI signal more flexibly, as a unit-mass probability density on the 6D manifold of symmetric positive definite tensors, yielding a TDF, or continuous mixture of tensors, at each point in the brain. From the TDF, one can derive analytic formulae for the orientation distribution function (ODF), tensor orientation density (TOD), and their corresponding anisotropy measures. Because this model can accurately resolve sharp signal peaks in angular space where fibers cross, we studied how many gradients are required in practice to compute accurate orientation density functions - as more gradients require longer scanning times. In this paper, we assessed how many diffusion-sensitized gradients were sufficient to (1) accurately resolve the diffusion profile, measured by the Kullback-Leibler divergence (sKL) and (2) achieve a satisfactory information uncertainty index, quantified by the exponential isotropy (EI), a TDF-derived measure of fiber integrity that exploits the full multidirectional HARDI signal. We used simulation data generated from two-fiber systems crossing at 90 degrees with varying Rician noise, as well as 4T human HARDI94 data.

2 Methods

2.1 Image acquisition

Three datasets were used in this study. The first one was simulated: we created various models of two-fiber systems, crossing at 90 degrees with equal volume fractions (w_1=w_2=0.5). Here we chose λ_1=10x10^{-4} mm^{2}s^{-1} and λ_2=2x10^{-4} mm^{2}s^{-1} as the eigenvalues for each individual component tensor (with FA=0.77, typical for white matter in the brain) and we added Rician noise of different amplitudes (with signal-to-noise ratio, SNR=5, 15, 25, and with a standard deviation of S(0)/SNR) to generate simulations using discrete mixtures of Gaussian distributions. The simulated data were sampled at 94 points evenly distributed on the hemisphere with an angular distribution computed from a partial differential equation (PDE) based on electrostatic repulsion [14]; we chose 94 as it was the same as the number of gradients used in the human 4T HARDI experiment, which was the source of the second dataset analyzed in this study.

One young healthy human subject was scanned using a diffusion-sensitized MRI protocol on a Bruker Medspec 4 Tesla MRI scanner, with a transverse electromagnetic (TEM) headcoil. The timing and angular sampling of the diffusion sequence was optimized for SNR [14, 15]. The protocol used 94 diffusion-sensitized gradient directions, and 11 baseline scans with no diffusion sensitization (b-value: 1159 s/mm^2; TE/TR: 92.3/8250 ms; FOV=230x230; in-plane resolution: 1.8mmx1.8mm; 55 x 2mm contiguous slices; acquisition time: 14.5 minutes).

Finally, a third HARDI dataset came from a monkey scanned using diffusion imaging on a 7 Tesla MRI scanner at the Center for Magnetic Resonance Research, at the University of Minnesota, using 100 gradients and 3 different b-value settings (1000, 2000, 3000 s/mm^2), TR/TE of 4600/65 ms, and an imaging matrix of 128x128x50 with isotropic voxels of 1 mm^3 (acquisition time: 23.5 minutes).

2.2 Data Processing

Several angular sampling schemes, with 20 to 94 directions, were sub-sampled from the original 94 angular locations to maximize a measure of the total angular energy. The angular distribution energy between point *i* and point *j* is denoted by E_{ij}, and defined as the inverse sum of the squares of the least spherical distance between point *i* and point *j* and the squares of the least spherical distance between point *i* and point *j*'s antipodally symmetric point *J* (Eq. 1):

$$E_{ij}^{-1} = \mathrm{dist}^2(i,j) + \mathrm{dist}^2(i,J) \quad (1)$$

Here, i,j are two different points in the spherical surface, J is the antipodally symmetric point to j, and dist(i,j) is the least spherical distance between point i and point j (see **Figure 1**). The total angular distribution energy for one gradient subset with *N* diffusion-sensitized gradients was defined as the summation of angular

distribution energy between all points in all pairs, using geodesic distances on the sphere (Eq. 2):

$$E(N) = \sum_{i=1}^{N} \sum_{j=1}^{N} E_{ij} \quad (i \neq j) \tag{2}$$

We first chose one seed point from the original 94 points, which (without loss of generality) was chosen to be (1, 0, 0) in our study. We then found another 5 points from the remaining 93 points to maximize E (6), since 6 diffusion-sensitized gradients are the minimum required for tensor estimation (so long as a non-diffusion-sensitized reference signal is also collected). In this way, the first subset with 6 diffusion-sensitized gradients was produced. After this initial subset, we artificially increased the angular sampling one gradient at a time, by maximizing E (N) (where N is the total number of diffusion sensitized gradients).

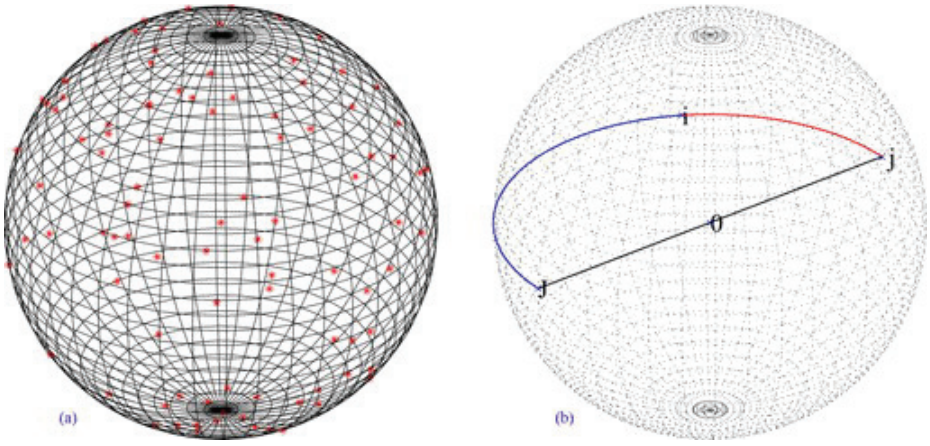


Figure 1. (a) Spherical distribution of diffusion gradient encoding angles. Red points on the sphere indicate the spherical distribution of angles at which diffusion-sensitized gradient images were collected for the 105-gradient HARDI sequence, which consists of 94 diffusion-sensitized gradients and 11 non-sensitized gradients. Each red dot in this figure represents one gradient direction, so there are 94 points in total on the unit sphere. In areas that appear to be relatively sparsely sampled, there is typically a sampled point on the opposite side of the sphere. Also, equidistribution problems sometimes lead to apparent clusters of points in some regions (see e.g., Friedman E. "Circles in Circles." http://www.stetson.edu/~efriedma/cirincir/), as the minimum point separation is only the same for all points for certain specific sample sizes. (b) Angular distribution energy calculation. In this figure, O is the original point, i, j represent two different points on the spherical surface, J is the antipodally symmetrical point to j. Based on Equation 1, the angular distribution energy between i and j is contributed based on the least spherical distance between point i and point j - denoted by dist(i,j) - and the least spherical distance between point i and point J - denoted by dist(i, J). dist(i,j) is illustrated by the red curve while dist(i,J) is represented by blue curve on the sphere on the right.

Using these optimized subsets of angular points, we sub-sampled the original HARDI94 data, and applied the framework in [13] to all these sub-samples. We denote the space of symmetric positive definite 3x3 matrices by $\mathbb{D}$. The probabilistic

ensemble of tensors, as represented by a tensor distribution function (TDF) P, is defined on the tensor space $\mathbb{D}$ that best explains the observed diffusion-weighted images (Eq. 3):

$$S_{calculated}(q) = \int_{D\in\bar{\mathbb{D}}} P(D)\exp(-tq^t Dq)dD \quad (3)$$

To solve for an optimal TDF P^*, we use the multiple diffusion-sensitized gradient directions q_i and arrive at P^* using the least-squares principle (Eq. 4):

$$P^* = argmin_p \sum_i (S_{obs}(q_i) - S_{calculated}(q_i))^2 \quad (4)$$

From the TDF, the orientation density function (ODF) may be analytically computed from Eq. 5. These ODFs were rendered using 642 points, determined using a seventh-order icosahedral approximation of the unit sphere.

$$ODF(\tilde{x}) = C\int_{r=0}^{\infty} p(r\tilde{x})dr = C\int_{D\in\bar{D}} P(D)(\det(D)\,\tilde{x}^t D^{-1}\tilde{x})^{-\frac{1}{2}}dD \quad (5)$$

To assess how accurately the diffusion profiles could be reconstructed from subsampled data based on different angular sampling schemes, the Kullback-Leibler (sKL) divergence, a commonly used measure from information theory, was used to measure the discrepancy between the reconstructed and ground truth ODFs. Reconstruction error was calculated from Eq. 6, in which *p(x)* is the ODF derived from the subsampled schemes with additive Rician noise of various amplitudes, while *q(x)* is the noise-free ODF derived from the ground truth data.

$$sKL(p,q) = \frac{1}{2}\int_{\Omega} \left\{p(x)\log[\frac{p(x)}{q(x)}] + q(x)\log[\frac{q(x)}{p(x)}]\right\}dx \quad (6)$$

We also computed another measure of fiber integrity proposed in the original TDF framework, the exponential isotropy (EI; Eq.7). Given any TDF P, EI quantifies the overall isotropy of diffusion at any given voxel, and highlights the gray matter instead of white matter as in FA (since gray matter voxels tend to have low anisotropy, or high isotropy, and thus high EI values). EI is defined as the exponential function of the Shannon Entropy, so EI can also be used to quantify the information uncertainty:

$$EI\big(P(D)\big) = e^{Shannon\ Entropy} = e^{-\int_{D\in\bar{D}} P(D)logP(D)dD} \quad (7)$$

3 Results and Discussion

3.1 How reconstruction accuracy was affected by angular resolution in the presence of variable additive Rician Noise.

Figure 2 shows several characteristic ways in which the additive Rician noise affected the reconstructed ODFs. The effects of image noise on the reconstructed

HARDI ODF included combinations of (1) local diffusion coefficient swelling, (2) incorrect rotations of the dominant fiber directions, and (3) mixing or omission of maximum diffusivity peaks in the radial fiber profile.

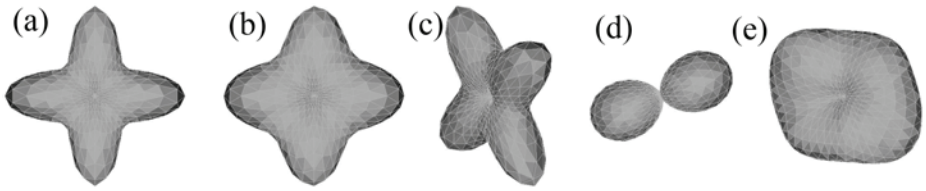


Figure 2. Noise effects on the HARDI ODF.
These glyphs show characteristic types of reconstruction errors that resulted from adding Rician noise to a simulated 2-fiber system, followed by deriving an ODF from the fitted tensor distribution function. (a) Ground truth ODF; (b) swelling of the local diffusion coefficient; (c) incorrect rotations of the dominant fiber directions (this is a rotation out of the plane of the page); (d) total omission of a dominant fiber direction; (e) mixing of the dominant directions. All these ODF are calculated based on Eq. 5 in the TDF framework without any regularization. Overall, the effect of noise on the HARDI ODF will most likely induce combinations of each of these types of distortion.

Next we assessed how the angular resolution affects the HARDI ODF reconstruction. **Figure 3** shows that, as expected, the higher the angular resolution, the more accurately the ODF can be recovered; even so, reconstruction errors vary from angular smearing and coalescing of the ODF peaks between 30 and 60 gradients to incorrect recovery of the dominant fiber direction at 20 gradients, which could be problematic for ODF-based tractography.

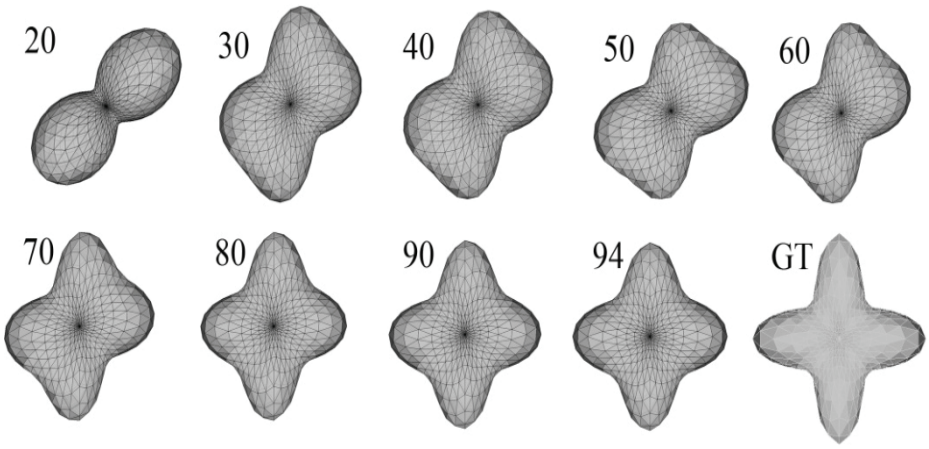


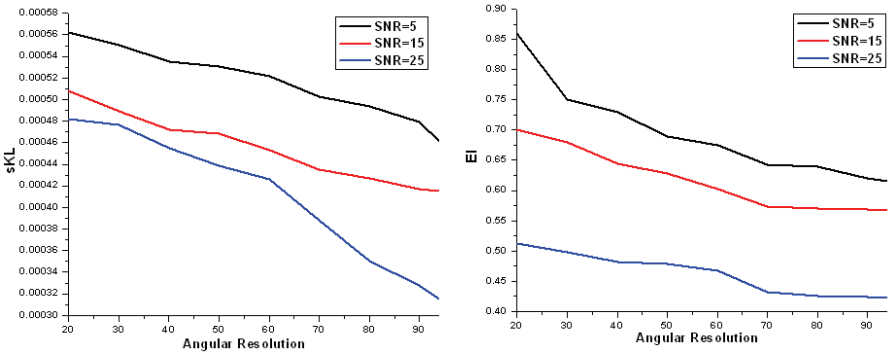
Figure 3. Angular Resolution effects on the HARDI ODF
This figure illustrates how angular resolution affects the HARDI ODF, which was calculated in the TDF framework based on Eq. 6, without any regularization. ODFs are reconstructed from sets of progressively more gradients, in directions that optimize the angular distribution energy (Eq. 3): the number on the upper left of each panel is the number of diffusion-sensitized gradients used to reconstruct the ODF. GT denotes ground truth.

In **Figure 3,** it is not immediately clear why the smaller number of gradient directions always coalesces the two peaks in the same direction (bottom-left to top-right); this most likely occurs because we use induction to define the gradient sets, so there cannot be perfect symmetry in the gradient set for all *n*, and some subsets will have a net excess of gradients in one quadrant (i.e., the point set will have a well-defined principal axis), which may lead the 2 dominant ODF peaks to coalescence into one in a specific quadrant, as the angular detail is reduced.

To quantify the accuracy of ODF recovery at different SNR levels and at different angular resolutions, we calculated the reconstruction error, represented by the sKL divergence between the recovered and the ground truth signal. As expected, the sKL error decreased with increasing SNR, and when more scanning directions were used (**Figure 4(a)**). The reconstruction accuracy of a 90-direction low-SNR sequence was about the same as a 30-direction sequence with five times the SNR. Our simulation studies showed that when SNR is low, adding directions has greater benefit. Moreover, higher angular resolution is needed for low SNR sequences to achieve reconstruction accuracy comparable to those obtained with higher SNR.

3.2 How the information uncertainty index was affected by the angular resolution

Information uncertainty was quantified here by EI which is a measure of fiber integrity related to FA (but avoiding the limitations of the single-tensor model).



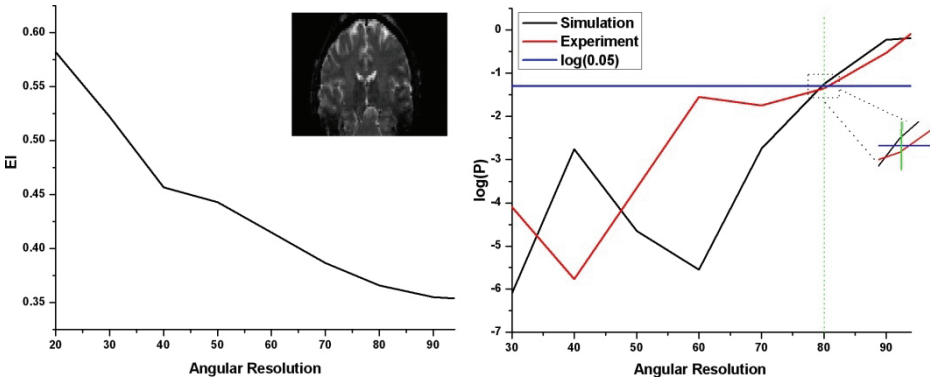
(a) sKL vs. Angular resolution (b) EI vs. Angular resolution

Figure 4 Simulation Studies with different SNR level

For each angular resolution scheme, 1000 simulations of two-tensor systems (equal volume fractions; 90º crossing) were computed with different SNR. (a) sKL divergence (reconstruction error) decreases with increasing SNR level, and with higher angular resolution sampling schemes. This means that the accuracy of the computed ODF improves as SNR increases and angular resolution increases. (b) EI decreases as the SNR level increases, and with more detailed angular sampling. In this figure, EI values have been normalized by the corresponding isotropic term, so that all EI values lie between 0 and 1 (which is the range for the more common anisotropy measure, FA). Also, EI tends to stabilize when the angular resolution reaches ~70 gradients. This is in line with the observation that standard FA measures are biased (too low) in regions where fibers mix or cross.

Other common measures of HARDI diffusion would also be used, such as generalized FA or total diffusion, but here we used EI as it has a direct link to the information content of the signal as defined by information theory.

As expected, EI decreased with increasing angular resolution. Simulation results show the EI stabilized by ~70 directions (**Figure 4(b)**). This is in line with the finding that fractional anisotropy, derived from DTI, is generally underestimated when fibers cross. Also, this result is consistent with **Figure 3,** which shows that HARDI70 has satisfying results when reconstructing a two-fiber system crossing at 90 degrees; these diagrams make it clearer why the isotropy falls (i.e., anisotropy rises) when the two fiber peaks no longer coalesce.



(a) EI *vs.* Angular Resolution (b) Paired *t* test result

Figure 5. 4 Tesla Human HARDI results

(a) EI *vs*. Angular Resolution in 4 T human HARDI94 data. We computed the average EI at different angular resolutions for one brain slice (the inset image is the corresponding T2-weighted slice). All EI values were normalized with respect to an isotropic diffusion profile to ensure that the EI values are between 0 and 1. We chose the average EI value in the cerebrospinal fluid (CSF) as the normalization constant since CSF has the highest diffusion isotropy in the brain. (b) Paired *t* test results. In this simulation data, the probability exceeds the threshold (p=0.05) when N is increases from 70 to 80, while for the empirical data, the probability exceeds the threshold when N is increases from 80 to 90. This answers the question, "does adding 10 more gradients improve the information in the signal?" Although these thresholds are to some extent arbitrary, they show that the information content converges within the standard range of gradients used in a HARDI study (~100).

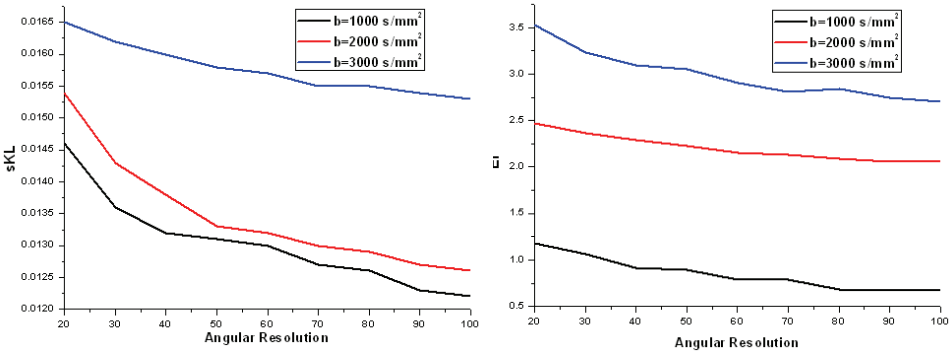
Figure 5(a) shows how EI is affected by the angular resolution in the 4 Tesla human data. EI indicates the information uncertainty, so the smaller the EI value is, the less uncertainty there is in the multi-fiber estimation. The EI decrease with increasing angular resolution does slow down, but we did not find a plateau at around 70 gradients in **Fig. 5(a)**, as was seen in **Figure 4(b).** To better understand whether EI has converged, we performed a paired Student's *t* test on the EI values, assessing the effect of adding additional gradients, in increments of 10, (e.g. 40 *vs* 30, 30 *vs* 20) at all voxels in the brain. If the *t* test result is significant ($p<0.05$) then this test confirms that adding 10 more gradients does indeed lead to lower EI (i.e., lower information uncertainty). When this *t* test is not significant, there is no evidence that adding 10

more gradients to the acquisition protocol is helpful, so the signal may be said to be saturated. The result of this test clearly depends on the number of voxels (here 1000 for simulation and 3255 for the experiment), Even so, this is a reasonable and intuitive operational definition of saturation for practical purposes. We note that this test could be slightly improved by incorporating a multiple comparisons correction into the *p*-value, to control the false discovery rate, but we did not do so as the tests were intended as a heuristic to compare successive increments in gradient numbers.

Figure 5(b) shows paired *t* test results assessing whether the EI significantly decreases when adding more gradients (i.e., $EI_{N+10}<EI_N$) for both simulations and the 4-Tesla human data. In the simulation data, EI decreases as angular resolution increases; this progressive decrease is also statistically significant when initially adding batches of 10 additional gradients, then after 70 gradients are reached, the probability exceeds the threshold (p=0.05) and the information uncertainty no longer shows a stastistically significant improvement, consistent with Figure 4(b). Here, we may refer to this ceiling effect on EI, at 70, as the Statistical Saturation (SS) number (i.e., SS=70). We defined the meaning of this number to be that successively increasing the angular resolution always leads to statistically significant improvements in EI until the statistical saturation number of gradients is reached. This definition of incremental information gain clearly depends on the batch size (adding 10 gradients each time). For our empirical HARDI data collected at 4-Tesla, this statistical saturation number was 80. As a qualification, we note that our simulation is based on only two fibers crossing at 90 degrees with equal weighting. In the more complex case of human brain data, the voxels in each slice have varying numbers of crossing fibers, varying numbers of detectable dominant fibers, and inevitably, a different weighting for each single component fiber within each voxel. Thus, for the experimental data, more gradient directions may be needed to cause statistical saturation in the EI (our information uncertainty index).

3.3 How reconstruction accuracy was affected by multiple b values with Rician Noise.

Similarly, reconstruction accuracy was assessed with simulations with b values varying from 1000 s/mm^2 to 3000 s/mm^2 - which would be within a typical range used in diffusion spectrum imaging (DSI) studies. Rician noise was added at a SNR of 10, a level similar to real MRI images. From **Figure 6a**, we note that the sKL-divergence (reconstruction error) increases with increasing b values, but it decreases with increasing angular resolution. The explanation for this is that $S(q)/S(0)=\exp(-b\mathbf{q}^T\mathbf{D}\mathbf{q})$, so the higher b-value is, the smaller the value of S(q)/S(0) will be. This value will therefore be more greatly affected by noise, if the noise characteristics are set independently of the b-values. So increasing the b-value leads to an increasing effect of noise in the final composite data, and thus higher reconstruction error. Even so, the additional b-value shells may be used to provide additional information on the diffusion propagator that would not be obtainable using only a single b-value.



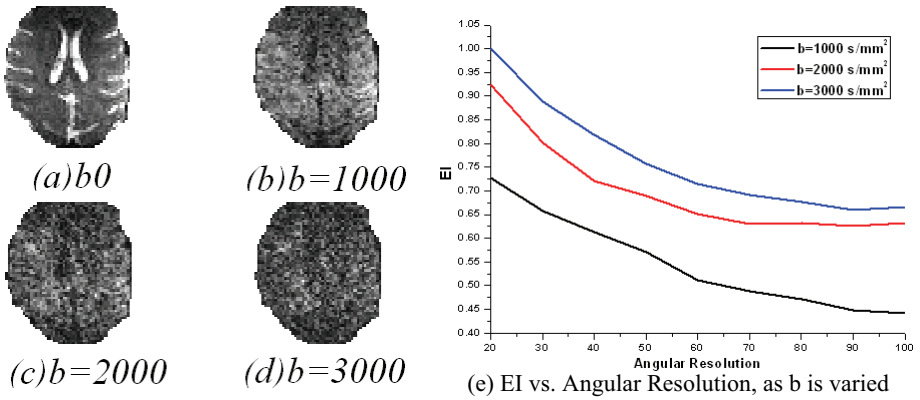
(a) sKL vs. Angular resolution , as b is varied (b) EI vs. Angular resolution, as b is varied

Figure 6. Simulation Study using different b- value settings

For each angular resolution scheme, 1000 simulations of two-tensor systems (equal volume fractions; 90º crossing) were computed with different b values (b=1000, 2000 and 3000 s/mm^2). Rician noise was added at a SNR level of 10. (a) sKL divergence (reconstruction error) increases with increasing b values, while decreasing using higher angular sampling schemes. This means that the accuracy of the computed ODF improves as b value decreases and angular resolution increases. (b) The EI behaves in the same fashion as sKL in (a).

3.4 How the information uncertainty index was affected by the b-values

To investigate the effect of b-value settings at high magnetic field (7 Tesla), we analyzed a 100-direction 7 T monkey HARDI dataset, exactly as in Section 3.2.



(e) EI vs. Angular Resolution, as b is varied

Figure 7. 7 Tesla HARDI scanning results

In this figure, 100-direction 7 Tesla HARDI data from a monkey was analyzed using EI to measure the information uncertainty. Illustrative slices are shown from the (a) T2 reference image (b) DWI at b=1000 s/mm^2 (c) DWI at b=2000 s/mm^2 (d) DWI at b=3000 s/mm^2, and (e) EI plot vs. Angular Resolution at the different b-values. As expected, higher b-value shells give noisier data.

Figure 7 shows a T2 image and DWI images taken at three separate b-values. The plots show how EI was affected by the angular resolution at the three different b-

value settings. EI was also affected by changing the angular resolution, at different b-values. Visualizations in **Figures 7(b)-(d**) show that the diffusion weighted images at higher b-values are quite noisy, which may be due to the combination of the higher magnetic field and the higher b value (suppressing the diffusion-weighted signal relative to the noise). Here, the EI versus angular resolution plot, **Figure 7(e)**, exhibits the same pattern as that seen in simulations (**Figure 6(b)**), suggesting that b-values higher than 2000 may be suboptimal when acquiring ultra-high magnetic field strength DWI images.

4 Conclusion

HARDI scanning allows better diffusion reconstruction than DTI, and provides new insight into fiber architecture and connectivity that cannot be achieved, even in principle, using a smaller number of diffusion-sensitized gradients. These advantages come at the expense of longer scanning times, but the trade-off may be worth it for studies assessing fiber connectivity and for fine-scale mapping of anatomy, and to avoid errors in routine clinical studies. We identified several types of ODF reconstruction errors that are typical when smaller numbers of gradients are used, and studied their asymptotics in optimized angular sets. To improve diffusion reconstruction accuracy and remove bias from the derived anisotropy measures, it is more effective to acquire additional angular samples than to repeatedly sample the same directions for purposes of signal averaging [16-18]. We found that, with a reasonable intuitive definition of saturation, the information uncertainty cannot be statistically improved when the number of diffusion-sensitized gradients exceeds 80. Also, from our preliminary study at 7 Tesla, the b-values should not be set too high, in order to obtain satisfactory EI values. Thus, our study may be of interest in designing future DTI and HARDI acquisition protocols for assessing fiber integrity in the living brain.

References

1. Basser PJ, Pierpaoli C. Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI. **J. Magn. Reson.,** vol. B 111, no. 3, pp.209-219 (1996).
2. Stejskal, EO, Tanner JE. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. **J. Chem. Phys.** 42:288-292 (1965).
3. Tuch DS. Q-Ball Imaging. **Mag. Res. Med.** 52:1358-1372 (2004).
4. Tournier JD, Calamante F, Gadian DG, Connelly A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. **NeuroImage**: 23: 1176-1185 (2004).
5. Kaden E, Knosche TR, Anwander A. Parametric spherical deconvolution: Inferring anatomical connectivity using diffusion MR imaging. **Neuroimage**: 37: 474-486 (2007).
6. Jian B, Vemuri BC, Özarslan V, Carney PR, Mareci TH. A novel tensor distribution model for diffusion-weighted MR signal. **NeuroImage** 37: 164-176. (2007)

7. McGraw T, Vemuri BC, Yezierski B, Mareci T. Von Mises-Fisher mixture model of the diffusion ODF. Biomedical Imaging: Nano to Macro, **3rd IEEE International Symposium on Biomedical Imaging**: 65-68 (2006).
8. Ozarslan E, Vemuri BC, Mareci TH. Higher rank tensors in diffusion MRI, **Visualization and Image Processing of Tensor Fields**, (2005).
9. Barmpoutis A, Vemuri BC. Exponential Tensors: A framework for efficient higher-order DT-MRI computations. In Proceedings of ISBI07: **IEEE International Symposium on Biomedical Imaging**: Page(s): 792-795 (2007).
10. Perrin M, Poupon C, Rieul B. Validation of q-ball imaging with a diffusion fibre-crossing phantom on a clinical scanner. **Phil Trans R Soc Lond B Biol Sci** 360:881–891 (2005).
11. Jbabdi S, Woolrich MW, Andersson JLR et al. A Bayesian framework for global tractography. **NeuroImage** 37:116-129. (2007).
12. Kaden E, Knösche TR, Anwander A. Parametric spherical deconvolution: Inferring anatomical connectivity using diffusion MR imaging. **Neuroimage**. 37: 474-488 (2007).
13. Leow AD, Zhu S, Zhan L, McMahon KL, de Zubicaray GI, Meredith M, Wright MJ, Toga AW, Thompson PM (2009). The Tensor Distribution Function, **Magnetic Resonance in Medicine**, 2009 Jan. 18; 61(1):205-214 (2009).
14. Jones DK, Horsfield MA, Simmons A. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. **Magnetic Resonance Imaging** 42(3): 515-525 (1999).
15. Pend H, Arfanakis K. Diffusion tensor encoding schemes optimized for white matter fibers with selected orientations. **Magnetic Resonance Imaging** 25: 147-153 (2007).
16. Goodlett C, Fletcher PT, Lin W, Gerig G. Quantification of measurement error in DTI: theoretical predictions and validation. **MICCAI** 2007; 10(Pt 1): 10-17 (2007).
17. Jones DK, Basser PJ. Squashing peanuts and smashing pumpkins: How noise distorts diffusion-weighted MR data. **Magnetic Resonance in Medicine** 52(5), 979–993 (2004).
18. Jones DK. The Effect of Gradient Sampling Schemes on Measures Derived From Diffusion Tensor MRI: A Monte Carlo Study. **Magnetic Resonance in Medicine** 51(4), 807–815 (2004).

Stochastic Tractography in 3-D Images via Nonlinear Filtering and Spherical Clustering

H. Ertan Çetingül[1], Gernot Plank[2], Natalia Trayanova[3], and René Vidal[1]

[1] Center for Imaging Science, Johns Hopkins University, USA
[2] Institute of Physiology, Medical University of Graz, Austria
[3] Institute for Computational Medicine, Johns Hopkins University, USA
{ertan,ntrayanova,rvidal}@jhu.edu, gernot.plank@meduni-graz.at

Abstract. This paper presents a stochastic tractography algorithm to identify branching fibrous structures in 3-D images. Specifically, we employ a Bayesian formulation that involves sequential importance resampling of multiple fiber trajectories (particles). At each step, we sample local orientation candidates from a discrete importance density that is estimated by utilizing a nonlinear oriented filter. The weights of the particles are subsequently obtained by using an observation density, which models the intensity coherence along the particle segments of inferred widths, and a prior density that enforces local smoothness of the trajectories. At the same time, we successfully build multiple fiber trajectories along all branches of a fiber by detecting dominant local fiber orientations via the mean shift algorithm. The performance of the proposed method is evaluated by tracking synthetic fibers at different noise levels as well as by extracting selected free-running cardiac Purkinje fibers in structural magnetic resonance images. Our experiments show that the method achieves a mean tracking error of about 4 voxels in the case of Purkinje fibers and remains robust to moderate amount of noise.

1 Introduction

The development of robust processing methods to quantitatively characterize fibrous structures constitutes an important yet challenging problem in medical image analysis. Specifically, the extraction of complex fiber networks in intensity data finds a wide range of applications in biological studies (see [1, 2] and references therein). For instance, by extracting the anatomical topology of different structures such as the human vasculature or pulmonary airways, one could perform a quantitative assessment of various pathologies. Furthermore, advanced simulations of arrhythmias could benefit from an electrophysiological model of the heart with a realistic conduction system [3]. This would involve extracting the Purkinje system, which is responsible for the propagation of the electrical impulse initiating the contraction of the ventricular myocardium. Modern ex vivo magnetic resonance imaging (MRI) techniques provide sufficient resolution to identify the *free-running* Purkinje fibers, which activate endocardial structures such as the papillary muscle. However, tracking these fibers in an automated manner is difficult due to the presence of numerous bifurcations and image noise.

From an image processing standpoint, a fibrous structure comprises a spatially coherent appearance pattern that can be quantified via different feature-based or model-based approaches. Typical examples reviewed in [4] include skeletons, matched filters, region growing, active contours, and image Hessian-guided streamline tracking. However, most of these methods are *deterministic* in the sense that the uncertainties associated with the resulting fibers are not inferred. In addition, they do not usually incorporate a priori geometric knowledge. Similarly, the majority of the existing tracking techniques on diffusion weighted imaging (DWI) estimate fiber tracts to be proportional to principal diffusion directions [5]. However, despite their practicality for visualization, they also do not estimate the aforementioned uncertainties, which are critical for eliminating anatomically incorrect trajectories caused by local tracking errors and their accumulation.

These issues have inspired recent works [6–10], which aim at *stochastically* tracking fibrous structures at the expense of increased computational complexity. These methods choose the most probable fiber trajectory among multiple possible trajectories that are generated by statistical sampling. For instance, [7] proposes a bootstrap filtering scheme to track the cross-sections of non-branching cerebral arteries. In order to extract non-branching vessels, [8] employs a particle filtering formulation with a sophisticated observation density that requires several parameters to be tuned, whereas [9] proposes a modified Bayesian formulation that requires fewer particles. For the analysis of branching fibers, [6] and [10] employ the k-means algorithm to cluster the trajectories of a particle filter to segment coronary arteries and brain vasculature, respectively. However, the extracted vessel topology is limited since k-means requires the number of clusters (branches) to be known beforehand. Moreover, there exist several works providing different stochastic tracking approaches on DW images [11–14]. In particular, [12] proposes a Bayesian formulation that draws samples from the posterior density, whereas [14] employs sequential importance resampling to track white matter fibers. [15] estimates complex fiber orientation distributions from high angular resolution diffusion images and performs random walks to infer brain connectivity.

We believe that the efficiency of stochastic tractography methods can be substantially improved by developing 1) practical probabilistic models for sampling, and 2) accurate clustering techniques for detecting bifurcations. We thereby present such a stochastic approach to track branching fibrous structures in 3-D images. Our contribution is to fuse the notions of filter-based local orientation extraction and multimodal fiber orientation distributions into a particle filtering formulation to infer the uncertainties of the fiber trajectories. Specifically, we utilize a nonlinear *pivoting filter* that estimates a reliable discrete importance density for sampling trajectories (particles). Furthermore, we use this density for detecting bifurcations by clustering the local fiber orientations via *spherical mean shift*. We recursively compute the weights of the particles using an observation density that quantifies intensity coherence along the particle segments of inferred widths and a prior density that enforces locally smooth trajectories. We evaluate our tracking method on synthetic fibers at different signal-to-noise ratios (SNRs) as well as on selected free-running cardiac Purkinje fibers in structural MRI.

2 Preliminaries on Stochastic Tracking

Our tracking scheme follows a well-known particle filtering technique called *sequential importance resampling* (SIR), which implements a recursive Bayesian filter by Monte Carlo simulations [16, 17]. In the SIR formulation, we represent a fiber trajectory as a sequence of 3-D points (states) $\boldsymbol{x}_{0:T} \doteq \{\boldsymbol{x}_t\}_{t=0}^{T}$ in the image domain $\Upsilon \subset \mathbb{R}^3$. Without loss of generality, we assume that the root of the fiber of interest is determined, either by the user or through some prior automatic detection. We define the observations as a sequence of image intensities $y_{1:T} \doteq \{y_t\}_{t=1}^{T} = \{I(\boldsymbol{x}_t)\}_{t=1}^{T}$, where $I(\boldsymbol{p})$[4] denotes the intensity value at voxel $\boldsymbol{p} \in \Upsilon$. We further assume that 1) the states are modeled as a first-order Markov process, 2) the observations are mutually independent, and 3) the observation at time t only depends on the current state, i.e., $p(y_t|\boldsymbol{x}_{0:t}) = p(y_t|\boldsymbol{x}_t)$.

The key idea in SIR is to represent the required posterior density $p(\boldsymbol{x}_{0:t}|y_{1:t})$ by a set of N weighted random samples $\{\boldsymbol{x}_{0:t}^{(n)}\}_{n=1}^{N}$ and the associated weights $\{w_t^{(n)}\}_{n=1}^{N}$, which are chosen using the principle of *importance sampling*. This specifically involves defining an *importance density* $q(\cdot)$ from which the random samples can be easily drawn. Accordingly, the weights measure the reliability of the samples as $w_t^{(n)} \propto p(\boldsymbol{x}_{0:t}^{(n)}|y_{1:t})/q(\boldsymbol{x}_{0:t}^{(n)}|y_{1:t})$.

In a sequential setting, we consider N samples that represent $p(\boldsymbol{x}_{0:t-1}|y_{1:t-1})$ and then approximate $p(\boldsymbol{x}_{0:t}|y_{1:t})$ with a new set of samples. Using Bayes' rule, a recursive formula for the posterior density can be written in terms of the *observation density* $p(y_t|\boldsymbol{x}_t)$ and the *prior density* $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ as

$$p(\boldsymbol{x}_{0:t}|y_{1:t}) \propto p(\boldsymbol{x}_{0:t-1}|y_{1:t-1})p(y_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}). \tag{1}$$

If the importance density satisfies $q(\boldsymbol{x}_{0:t}|y_{1:t}) = q(\boldsymbol{x}_{0:t-1}|y_{1:t-1})q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, y_{1:t})$, the formulation only requires a filtered estimate $p(\boldsymbol{x}_t|y_{1:t})$ at each time step. Consequently, we can recursively compute the weights as

$$w_t^{(n)} \propto w_{t-1}^{(n)} \frac{p(y_t|\boldsymbol{x}_t^{(n)})p(\boldsymbol{x}_t^{(n)}|\boldsymbol{x}_{t-1}^{(n)})}{q(\boldsymbol{x}_t^{(n)}|\boldsymbol{x}_{t-1}^{(n)}, y_{1:t})}. \tag{2}$$

Then the posterior filtered density $p(\boldsymbol{x}_t|y_{1:t})$ can be approximated as

$$p(\boldsymbol{x}_t|y_{1:t}) \approx \hat{p}(\boldsymbol{x}_t|y_{1:t}) = \sum_{n=1}^{N} w_t^{(n)} \delta(\boldsymbol{x}_t - \boldsymbol{x}_t^{(n)}), \tag{3}$$

where $\delta(\cdot)$ denotes the Dirac delta function and the weights are normalized such that $\sum_{n=1}^{N} w_t^{(n)} = 1$. As the number of samples $N \to \infty$, the approximate posterior $\hat{p}(\boldsymbol{x}_t|y_{1:t})$ approaches the true posterior $p(\boldsymbol{x}_t|y_{1:t})$.

It is also worth noting that during SIR, the distribution of the weights becomes more skewed as t increases [16]. To avoid this degeneracy, a resampling procedure

[4] When the point $\boldsymbol{p}$ lies outside the discrete grid, we compute the corresponding intensity value by trilinear interpolation.

is usually included at each step. More specifically, if the *effective sample size* $N_{\text{eff}} \doteq \lceil 1/\sum_{n=1}^{N}(w_t^{(n)})^2 \rceil$ is less than a fixed threshold τ, we resample N particles from the discrete density $\hat{p}(\boldsymbol{x}_t|y_{1:t})$ and set $w_t^{(n)} = 1/N, \forall n$. The reader is referred to [16, 17] for further details on sequential Monte Carlo methods.

3 Estimation of the Density Functions in SIR

This section explains the estimation of the importance, observation and prior densities in the SIR formulation. These densities are primarily used to recursively compute the importance weights of the particles. For the sake of clarity, we drop the particle superscripts in the following discussions.

3.1 Importance Density

In the SIR formulation, modeling the importance density with the prior and solely using the observation density to compute the weights from (2) constitute a popular yet trivial choice. However, we believe that sampling from a density that models local fiber structure is more appropriate for tracking purposes. We thereby estimate the importance density $q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, y_{1:t})$ using a nonlinear *pivoting filter*, which is a 3-D variant of the matched filter introduced in [18]. Its efficacy in extracting local fiber orientations has been recently demonstrated in [19].

As depicted in Fig. 1(a), the pivoting filter is centered at a point of interest $\boldsymbol{x}$ with a fixed backward point $\boldsymbol{b}$ and a pivoting forward point $\boldsymbol{f}$ located at a distance l from $\boldsymbol{x}$. The segments $\overline{\boldsymbol{b}\boldsymbol{x}}$ and $\overline{\boldsymbol{x}\boldsymbol{f}}$ define the main segments of the filter and are designed to align with the fiber of interest. The purpose of the remaining points $\{\boldsymbol{b}_k, \boldsymbol{f}_k\}_{k=1}^{2K}$ is to fully encapsulate the fiber. Specifically, the points $\{\boldsymbol{b}_k\}$ (and $\{\boldsymbol{f}_k\}$) are placed by rotating the $2w$-apart antipodal pair $(\boldsymbol{b}_k, \boldsymbol{b}_{k+K})$ (and $(\boldsymbol{f}_k, \boldsymbol{f}_{k+K})$) by an angular step α until one circular round is completed.

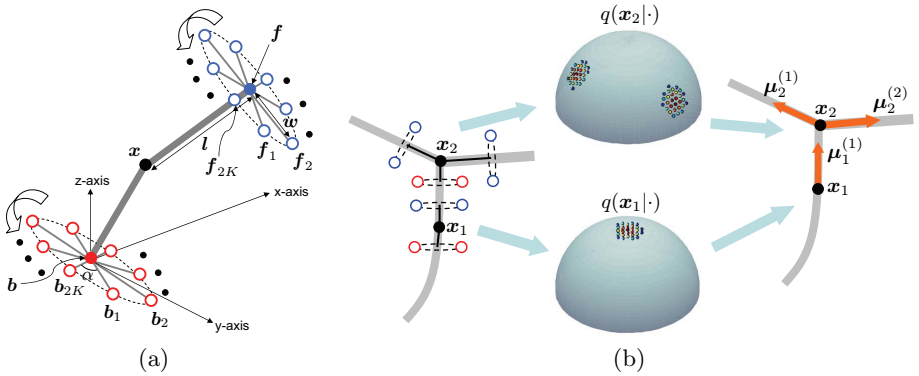


Fig. 1. (a) 3-D pivoting filter with key parameters, (b) Discrete importance densities estimated at two analysis points and the modes detected via spherical mean shift.

The operation of the filter can be summarized as follows: After fixing the backward segment $\overline{\boldsymbol{b}\boldsymbol{x}}$ along the known portion of the fiber, the forward segment

$\overline{\boldsymbol{xf}}$ is allowed to rotate for scanning multiple local orientations. Specifically, consider a particular orientation $\boldsymbol{s}$ in $\mathcal{S} \doteq \{\boldsymbol{s} : \langle \boldsymbol{s}, \boldsymbol{x} - \boldsymbol{b} \rangle > 0, \boldsymbol{s} \in \mathbb{S}^2\}$ such that $\boldsymbol{f} = \boldsymbol{x} + l\boldsymbol{s}$. For each pair of antipodes $(\boldsymbol{b}_k, \boldsymbol{b}_{k+K})$ and $(\boldsymbol{f}_k, \boldsymbol{f}_{k+K})$, we compute

$$h_k(\boldsymbol{s}; \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{b}, I) = \begin{cases} 1 & \text{if } |I(\boldsymbol{b}) - I(\boldsymbol{f})| \leq \min\limits_{j=k,k+K} \{|I(\boldsymbol{b}) - I(\boldsymbol{b}_j)|, |I(\boldsymbol{f}) - I(\boldsymbol{f}_j)|\} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

In other words, the filter detects an oriented structure if the absolute value of the intensity variation along the structure is less than the minimum absolute intensity variation orthogonal to the structure. The expression in (4) can be considered as a *partial* filter response for a fixed $k \in \{1, 2, \ldots, K\}$. The overall filter response is subsequently computed by summing (4) over all pairs of antipodes as $h(\boldsymbol{s}; \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{b}, I) = \sum_{k=1}^{K} h_k(\boldsymbol{s}; \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{b}, I)$.

In the sequential setting, having fixed the backward segment along the preceding orientation, i.e., $\boldsymbol{s}_{t-1} \doteq (\boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-2})/l$, the filter response at $\boldsymbol{x}_{t-1}$ gives a coarse estimate of the probability of having a structure oriented along any $\boldsymbol{s} \in \mathcal{S}$. This is further refined by using $g(\boldsymbol{s}; \boldsymbol{x}, \boldsymbol{x} - \boldsymbol{b}, I) = \int_0^1 |I(\boldsymbol{x} + \lambda l \boldsymbol{s}) - I(\boldsymbol{x})|^2 \mathrm{d}\lambda$, which quantifies the intensity coherence along the forward segment [4]. The importance density is then estimated as

$$q(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, y_{1:t}) \propto \Big(h(\boldsymbol{s}_t; \boldsymbol{x}_{t-1}, \boldsymbol{s}_{t-1}, I) \times \exp(-g(\boldsymbol{s}_t; \boldsymbol{x}_{t-1}, \boldsymbol{s}_{t-1}, I)) \Big). \tag{5}$$

Given $q(\cdot)$, we can sequentially sample N local orientations $\{\boldsymbol{s}_t^{(n)}\}_{n=1}^N$ that also define the points $\{\boldsymbol{x}_t^{(n)}\}$ given $\{\boldsymbol{x}_{t-1}^{(n)}\}$. The resulting discrete density can be further used to detect branches via spherical clustering, as described in §4.

3.2 Observation Density

Given a current sample $\boldsymbol{s}_t$ (and $\boldsymbol{x}_t$), we obtain an estimate of the observation density $p(y_t | \boldsymbol{x}_t)$. For this purpose, we assume that 1) the fiber portion along $\boldsymbol{s}_t$ has the same width (diameter) as the preceding portion, and 2) the intensities of the voxels in a neighborhood of $\boldsymbol{x}_t$, denoted by $\Omega_{\boldsymbol{x}_t}$, are also observed.

We first find the cylinder that optimally encapsulates the preceding fiber portion and then use it to find the voxels around the segment $l\boldsymbol{s}_t = \boldsymbol{x}_t - \boldsymbol{x}_{t-1}$. Specifically, we encapsulate the fiber portion along $\boldsymbol{s}_{t-1}$ with (oriented) cylinders of height l and of different radii, and estimate the optimal radius r^e as

$$r^e = \underset{r \in \mathcal{R}}{\operatorname{argmax}} \ \frac{1}{2\pi l r} \int_{\mathcal{F}(r)} \Big(\langle \nabla I(\boldsymbol{p}), \overrightarrow{N}(\boldsymbol{p}) \rangle \Big)^2 \mathrm{d}\boldsymbol{p}. \tag{6}$$

Notice that the expression in (6) measures the alignment of the image gradients with the normals at the cylindrical surface, where $\mathcal{R}$ denotes the set of radii, $\boldsymbol{p}$ is a point on the lateral surface $\mathcal{F}(r)$ of the cylinder of radius r, $\nabla I(\boldsymbol{p})$ is the image gradient and $\overrightarrow{N}(\boldsymbol{p})$ is the outward normal to the surface $\mathcal{F}(r)$ at point $\boldsymbol{p}$.

We subsequently use the inferred radius to form a cylinder (of radius r^e and height l) oriented along $\boldsymbol{s}_t$, define $\Omega_{\boldsymbol{x}_t}$ as the voxels inside this cylinder, and compute the observation density as

$$p(y_t|\boldsymbol{x}_t) \propto \exp\left(-\frac{(I(\boldsymbol{x}_t) - \bar{I}_{t;r^e})^2}{2\sigma^2_{t;r^e}}\right). \tag{7}$$

Here $\bar{I}_{t;r^e}$ and $\sigma_{t;r^e}$ denote the mean and standard deviation of the intensities of the voxels inside the cylinder, respectively.

3.3 Prior Density

In order to enforce a certain level of smoothness in the fiber trajectories, we choose to model the prior density based on the von Mises-Fisher (vMF) distribution over the 2-sphere [20]. It constitutes a unimodal parametric distribution for directional data with the probability density function (pdf) of the form

$$p_{\text{vMF}}(\boldsymbol{s}; \boldsymbol{\mu}, \kappa) = \frac{\kappa}{4\pi \sinh \kappa} \exp(\kappa \langle \boldsymbol{\mu}, \boldsymbol{s} \rangle), \tag{8}$$

where $\boldsymbol{\mu} \in \mathbb{S}^2$ is the mean direction and $\kappa > 0$ is a parameter regulating the concentration around the mean direction. In our experiments, we manually set the value of κ and estimate the prior density as

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) \propto p_{\text{vMF}}(\boldsymbol{s}_t; \boldsymbol{s}_{t-1}, \kappa). \tag{9}$$

4 Algorithm Overview and Implementation Details

Our method is initiated at two user-specified seed points $\{\boldsymbol{x}_0, \boldsymbol{x}_1\}$, which place the backward segment $\overline{\boldsymbol{bx}}$ of the filter along $\boldsymbol{s}_1 = (\boldsymbol{x}_1 - \boldsymbol{x}_0)/l$. We subsequently estimate the width of this fiber segment from (6). In order to have a finite number of search orientations, we discretize the unit sphere at 642 predefined vectors obtained by a threefold tessellation of an icosahedron. Accordingly, at each step, we form the set of candidate orientations $\mathcal{S}_t \doteq \{\boldsymbol{s} : \langle \boldsymbol{s}, \boldsymbol{s}_{t-1} \rangle > 0, \boldsymbol{s} \in \mathbb{S}^2\}$, estimate the discrete importance density as described in §3.1, and use the resulting pdf to analyze local fiber structure, i.e., to detect branches at the point of interest.

The rationale behind analyzing local fiber structure is that in the case of branching fibers, one of the branches may cause the loss of the remaining ones by attracting most of the particles. To avoid this problem, before sampling the particles, we employ the mean shift (MS) algorithm [21], which automatically detects the number and directions of the branches. Specifically, we utilize a weighted spherical MS formulation with the Fisher kernel for clustering local orientations using their importance density values as weights [19]. This scheme converges to the modes $\{\boldsymbol{\mu}_t^{(c)}\}_{c=1}^{C}$ of the importance density $q(\boldsymbol{x}_t|\cdot)$ and the resulting modes are identified as the directions of the branches rooted at $\boldsymbol{x}_t$. Fig. 1(b) illustrates the operation of the filter and our branch detection strategy.

Algorithm 1 SIR-based tractography in 3-D images

1. At the t-th step, given $N = 1000$ weighted particles $\{(\boldsymbol{x}_{0:t-1}^{(n)}, w_{t-1}^{(n)})\}_{n=1}^{N}$ and the mean trajectory $\{\bar{\boldsymbol{x}}_{0:t-1}\}$, consider $\boldsymbol{s}_{t-1} = \bar{\boldsymbol{x}}_{t-1} - \bar{\boldsymbol{x}}_{t-2}$.
2. Place the pivoting filter such that $\boldsymbol{b} = \bar{\boldsymbol{x}}_{t-2}$ and $\boldsymbol{x} = \bar{\boldsymbol{x}}_{t-1}$.
3. Obtain the set $\mathcal{S}_t = \{\boldsymbol{s} : \langle \boldsymbol{s}, \boldsymbol{s}_{t-1} \rangle > 0, \boldsymbol{s} \in \mathbb{S}^2\}$.
4. Estimate the discrete importance density $\hat{q}(\boldsymbol{x}_t | \bar{\boldsymbol{x}}_{t-1}, y_{1:t})$ from (10) using (5).
5. Perform branch analysis by detecting the modes $\{\boldsymbol{\mu}_t^{(c)}\}_{c=1}^{C}$ of the importance density $\hat{q}(\boldsymbol{x}_t|\cdot)$ via spherical MS and check the number of branches C:
   - If $C > 1$, stop tracking the "parent" fiber, go to step 9 and generate separate SIR schemes for the branches.
   - If $C = 1$, sample N points $\{\tilde{\boldsymbol{x}}_t^{(n)}\}$ from the importance density.
6. For each $\tilde{\boldsymbol{x}}_t^{(n)}$, compute the weight $w_t^{(n)}$ from (2) using (5), (7), and (9) with $\kappa = 3$.
7. Calculate N_{eff} as described in §2, set $\tau = 2N/3$, and check N_{eff}:
   - If $N_{\text{eff}} \geq \tau$, then for $n = 1, 2, \ldots, N$, set $\boldsymbol{x}_t^{(n)} = \tilde{\boldsymbol{x}}_t^{(n)}$ and update the n-th trajectory as $\boldsymbol{x}_{0:t}^{(n)}$.
   - If $N_{\text{eff}} < \tau$, then for $n = 1, 2, \ldots, N$, sample an index $z(n)$ from the discrete distribution of $\{w_t^{(n)}\}_{n=1}^{N}$, set $\boldsymbol{x}_t^{(n)} = \tilde{\boldsymbol{x}}_t^{z(n)}$, $w_t^{(n)} = 1/N$ and update the n-th trajectory as $\boldsymbol{x}_{0:t}^{(n)}$.
8. Iterate between 1-8 by setting $t = t + 1$ until $\{w_t^{(n)}\}$ are small, a user-defined stopping criterion is met and/or a branching point is detected.
9. Identify the MAP fiber (or branch) as $\boldsymbol{x}_{0:t}^{(\zeta)}$, where $\zeta = \text{argmax}_{n=1,2,\ldots,N}\, w_t^{(n)}$, i.e., the trajectory with the maximum importance weight.

It depicts the importance densities, where the probabilities of the vectors are color-coded (blue∼low, red∼high), at two points as well as the resulting modes.

At this point, it is worth noting that the pivoting filter needs to be used N times per iteration to estimate the importance density given N preceding particles. This computational load can be reduced with the approximation

$$q(\boldsymbol{x}_t^{(n)} | \boldsymbol{x}_{t-1}^{(n)}, y_{1:t}) \approx \hat{q}(\boldsymbol{x}_t^{(n)} | \bar{\boldsymbol{x}}_{t-1}, y_{1:t}), \tag{10}$$

where $\bar{\boldsymbol{x}}_{t-1}$ is the point reached along the mode of interest, i.e., $\boldsymbol{\mu}_{t-1}^{(c)}$ for some c. After estimating the approximate importance density $\hat{q}(\boldsymbol{x}_t|\cdot)$ at the t-th step and clustering the local orientations, $N = 1000$ samples are drawn from this discrete density if the number of the modes, denoted by C, is equal to 1. However, if $C > 1$, the newly identified branches are treated as separate fibers to be tracked via new SIR schemes. Finally, we compute the importance weight $w_t^{(n)}$ from (2) using (5), (7), and (9) with $\kappa = 3$, and then eliminate the degeneracy described in §2 using *systematic resampling* with $\tau = 2N/3$. We repeat the same procedure by shifting the filter according to the modes of the importance density, i.e., keeping track of the "mean" trajectory $\{\bar{\boldsymbol{x}}_{0:t}\}$, and generate N different trajectories $\{\boldsymbol{x}_{0:T}^{(n)}\}_{n=1}^{N}$ for each branch. The maximum a posteriori (MAP) estimate of the true fiber/branch is identified as the trajectory with the maximal importance weight. In summary, our tracking method proceeds as outlined in Algorithm 1.

5 Method Validation

The performance of our method is evaluated via experiments on synthetic fibers at different noise levels as well as on selected free-running cardiac Purkinje fibers on structural MR images. In order to quantify the spatial tracking error, we compute the symmetrized Chamfer distance (in voxels) between the true fiber trajectory $\mathcal{X}^t \doteq \boldsymbol{x}^t_{0:T} = \{\boldsymbol{x}^t_i\}$ and its MAP estimate $\mathcal{X}^e \doteq \boldsymbol{x}^e_{0:T'} = \{\boldsymbol{x}^e_j\}$ as $\epsilon(\mathcal{X}^t, \mathcal{X}^e) = [d(\mathcal{X}^t, \mathcal{X}^e) + d(\mathcal{X}^e, \mathcal{X}^t)]/2$, where

$$d(\mathcal{X}^t, \mathcal{X}^e) = |\mathcal{X}^t|^{-1} \sum_{\boldsymbol{x}^t_i \in \mathcal{X}^t} \min\{\|\boldsymbol{x}^t_i - \boldsymbol{x}^e_j\| : \boldsymbol{x}^e_j \in \mathcal{X}^e\}. \tag{11}$$

5.1 Experiments on Synthetic Fibers

In order to evaluate the performances of the fiber width estimator (6), orientation detection and tracking methods as a function of image noise, we generate 120 synthetic fibers of radii $r^t \in \{1, 2, 3\}$ by fitting cubic splines through four randomly selected points in a $80\times80\times80$ lattice. In the case of branching fibers, bifurcations are randomly selected on previously generated fibers to add further branches. The centerlines of the resulting fibers constitute the true trajectories $\mathcal{X}^t$. The binary images are then corrupted by Rician noise to obtain image data at five different SNRs. We set the filter parameters $\{l, w, \alpha\} = \{3, 2, 20^\circ\}$ (or $\{4, 4, 20^\circ\}$ depending on the value of r^t), the search radii $\mathcal{R} = \{1, 1.5, 2, \ldots, 5\}$, and subsequently obtain the width estimates $\{r^e_j\}$ and the MAP trajectory $\mathcal{X}^e$ of each fiber.

For performance evaluation, we compute the width estimation error rate $\xi(\mathcal{X}^t, \mathcal{X}^e) \doteq |\mathcal{X}^e|^{-1} \sum_j |r^t - r^e_j|/r^t$ along with tracking error ϵ. In fact, the value of the optimal radius r^e is also critical in order to obtain an accurate observation density. Although a slight underestimation of the fiber width is tolerable unless the resulting filter is totally "buried" into the fiber, it should be noted that the more the fiber width is overestimated, the worse the observation density is modeled. Therefore, we also compute the rate ξ_+ of overestimating the fiber width. Finally, in order to quantify the reliability of our orientation detection strategy at different noise levels, we calculate the orientation detection error rates as follows: In the analysis of non-branching fibers, this rate, denoted by γ_1, is computed as the ratio of the number of points (in $\mathcal{X}^e$) at which the MS locates more than one mode over the total number of points, i.e., the cardinality of $\mathcal{X}^e$. However, in the analysis of branching fibers, we solely focus on the identification of bifurcations. Specifically, we place the filter at the bifurcations, which constitute the set of points at which the MS locates two modes in the noise-free case, and compute the rate γ_2 of misidentified bifurcations at different SNRs.

Analysis of non-branching fibers: Table 1 shows the mean and the standard deviation of the tracking error ϵ and the width estimation error rate ξ along with the average values of the width overestimation rate ξ_+ and the orientation detection error rate γ_1 over 60 non-branching fibers at different SNRs. The symbol "$\boldsymbol{\infty}$" for SNR represents the noise-free case. First, we observe that the

mean width estimation error rates are about 0.09 with $\xi_{\max} = 0.36$ for SNR > 5 dB. These results validate the accuracy of our flux-based estimator at moderate amount of noise. Furthermore, we overestimate the true fiber widths at rates (ξ_+) less than 0.25 for SNR > 10 dB. In terms of tracking accuracy, we observe that for SNR > 10 dB, the mean tracking errors are about 2.80 voxels with $\epsilon_{\min} = 1.00$ and $\epsilon_{\max} = 13.78$. However, the performance starts degrading at higher levels of noise, especially in the case of tracking fiber segments with high curvature. Finally, for SNR > 5 dB, the average values of the rates of erroneously detected orientations (γ_1) are less than 0.07, which demonstrate the reliability of our filter-based orientation detection strategy under noisy conditions.

Table 1. Experimental results on non-branching fibers

Error parameter	**SNR (dB)**					
	∞	**30**	**20**	**15**	**10**	**5**
ξ	0.08±0.07	0.08±0.07	0.08±0.07	0.09±0.07	0.11±0.09	0.20±0.26
$(\xi_{\min}, \xi_{\max})$	(0,0.36)	(0,0.36)	(0,0.31)	(0,0.31)	(0,0.32)	(0,1.42)
ξ_+	0.23	0.24	0.23	0.25	0.31	0.39
ϵ	2.39±1.21	2.82±2.11	3.20±2.39	2.68±1.32	3.96±2.70	5.18±3.83
$(\epsilon_{\min}, \epsilon_{\max})$	(0.87,6.35)	(0.78,13.68)	(0.65,13.78)	(0.90,6.81)	(1.00,11.94)	(1.67,18.92)
γ_1	0.04	0.04	0.04	0.05	0.07	0.17

Analysis of branching fibers: We further perform tracking experiments on 60 remaining branching fibers as well as focus on testing our orientation detection strategy at the bifurcations. At this point, note that extracting such fibers in noisy data may result in misleading tracking errors if a bifurcation point is undetected. Therefore, we estimate the widths of the fibers and obtain their MAP trajectories only in the noise-free case. Table 2(a) shows the mean and the standard deviation of the width estimation error rate ξ and the tracking error ϵ over 60 branching fibers. We observe that the mean tracking error is 4.81 voxels with $\epsilon_{\min} = 2.46$ and $\epsilon_{\max} = 8.21$. Notice that the tracking performance is worse than the one obtained for non-branching fibers by approximately 2.50 voxels. This is mainly due to 1) a number of branches that are generated very close to each other, and 2) differences between the coordinates of the actual bifurcations and the detected ones. Furthermore, we overestimate the true fiber widths in only 18% of the fiber segments and achieve a width estimation error rate of 0.07±0.05 with $\xi_{\min} = 0$ and $\xi_{\max} = 0.22$. Finally, Table 2(b) shows the average rates of misidentified bifurcations (γ_2). Notice that for SNR > 10 dB, less than 20% of the bifurcations are misidentified, but the error rate γ_2 increases at lower SNRs.

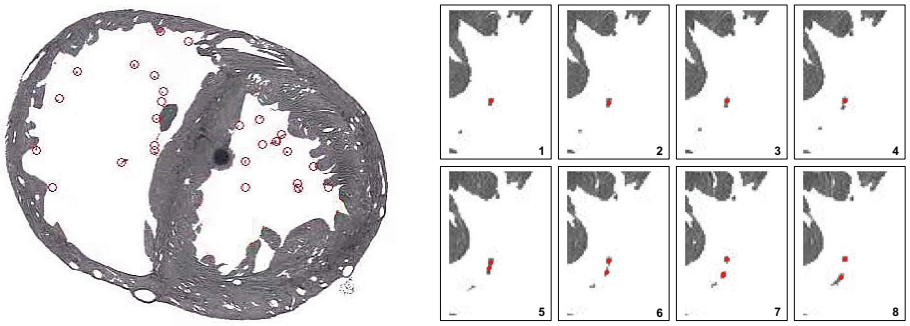
Table 2. Experimental results on branching fibers

(a) Tracking and width estimation errors

ϵ	$(\epsilon_{\min}, \epsilon_{\max})$	ξ	$(\xi_{\min}, \xi_{\max})$
4.81±1.42	(2.46,8.21)	0.07±0.05	(0,0.22)

(b) Bifurcation misidentification

SNR (dB)	**30**	**20**	**15**	**10**	**5**
γ_2	0.10	0.15	0.18	0.25	0.33



(a) (b)

Fig. 2. (a) An MR image slice with manually extracted Purkinje fibers (red circles), (b) Selected region of interests illustrating the trajectory of a branching Purkinje fiber.

5.2 Experiments on Cardiac Purkinje Fibers

We conduct additional tracking experiments on a structural MR image of a healthy rabbit heart, which was acquired on an 11.7 T MR system at an in-plane resolution of 26.5 μm $\times$ 26.5 μm and an out-of-plane resolution of 24.5 μm [3]. Denoising and rescaling steps are performed to obtain a 3-D image of size 512×512×850 (Fig. 2(a)). Our method is tested on selected subvolumes that contain 53 free-running Purkinje fibers. They are either non-branching fibers running from one Purkinje-myocardial junction (PMJ) to another or branching fibers with at least one Purkinje-Purkinje junction (PPJ), as depicted in Fig. 2(b). In addition, we manually extract the centerlines $\mathcal{X}^t$ of the true fibers and obtain the MAP trajectories $\mathcal{X}^e$ using the filter parameters $\{l, w, \alpha\} = \{4, 4, 20^\circ\}$.

In our experiments, we obtain a tracking error of 3.93±3.15 voxels with $\epsilon_{\min} = 0.78$ and $\epsilon_{\max} = 19.74$. Specifically, our method achieves such promising results that 16 fibers are tracked with errors of less than 2 voxels, and 32 fibers with errors of less than 4 voxels. In addition, all the bifurcations are correctly detected along with the corresponding branch directions to be followed. However, 2 fibers are tracked with errors greater than 10 voxels and we observe that, in those cases, tracking is affected by the presence of nearby fibers/cardiac wall. Specifically, the pivoting filter fails at encapsulating the fibers of interest and this causes an inaccurate estimation of the importance densities. Fig. 3 illustrates our tracking results by showing the surface renderings (green) of selected subvolumes, the seed points (black), and the resulting MAP trajectories (red) of the fibers.

6 Conclusions

The aim of this work is to stochastically track branching fibrous structures in 3-D images. Specifically, we present an SIR-based tracking framework that employs an oriented filter for the estimation of a practical discrete importance density. We also describe a flux-based fiber width estimator to obtain an observation density that models the intensity coherence in the fiber segment of interest.

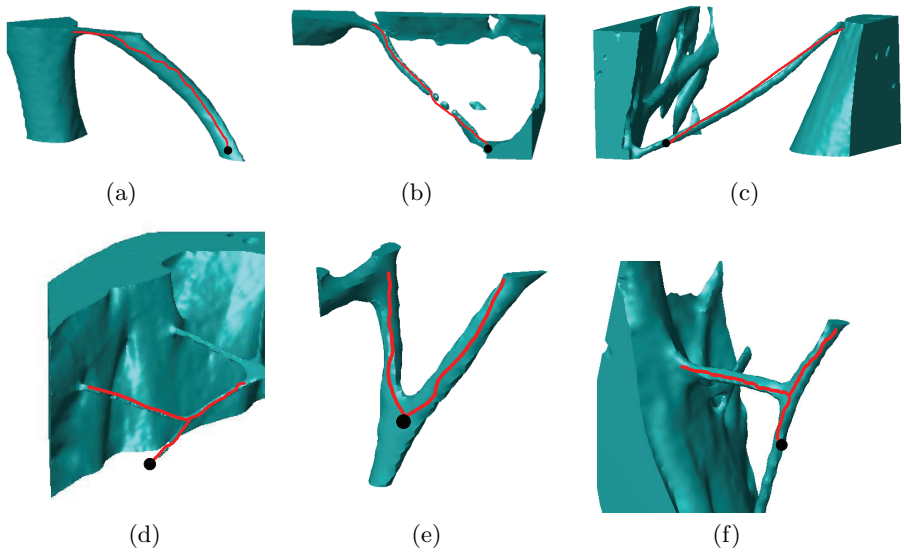


Fig. 3. Visualization of 6 free-running Purkinje fibers and their MAP trajectories (red).

Along with a prior density, which enforces locally smooth fiber trajectories, the aforementioned densities are used to recursively compute the certainties of the trajectories. Furthermore, in order to correctly analyze branching structures, we sample trajectories along all branches by detecting dominant branch directions via spherical mean shift. Comprehensive experiments in synthetic and real data demonstrate that the proposed method achieves promising results in terms of fiber width estimation and tracking accuracy at moderate amount of noise. Nevertheless, tracking fibers in dense fibrous regions as well as fiber segments with high curvature under noisy conditions may be problematic due to the (linear) geometry of the filter. Therefore, as a future work, we will perform a sensitivity analysis on the filter parameters and subsequently use the extracted free-running Purkinje system in cardiac conduction simulations.

Acknowledgments. The authors would like to thank Yu Ouyang for his help with the manual extraction of the Purkinje fibers and Dr. Volkan Cevher for useful discussions. This work has been funded by the grant NIH HL082729 and by startup funds from the JHU Whiting School of Engineering.

References

1. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms. ACM Computing Surveys **36**(2) (2004) 81–121
2. Palágyi, K., Tschirren, J., Hoffman, E., Sonka, M.: Quantitative analysis of pulmonary airway tree structures. Computers in Biology and Medicine **36** (2006) 974–996

3. Vadakkumpadan, F., Rantner, L., Tice, B., Boyle, P., Prassl, A., Vigmond, E., Plank, G., Trayanova, N.: Image-based models of cardiac structure with applications in arrhythmia and defibrillation studies. Journal of Electrocardiology **42**(2) (2009) 157.e1–157.e10
4. Qian, X., Brennan, M., Dione, D., Dobrucki, W., Jackowski, M., Breuer, C., Sinusas, A., Papademetris, X.: A non-parametric vessel detection method for complex vascular structures. Medical Image Analysis **13** (2009) 49–61
5. Mori, S., van Zijl, P.: Fiber tracking: Principles and strategies - A technical review. NMR in Biomedicine **15** (2002) 468–480
6. Florin, C., Paragios, N., Williams, J.: Globally optimal active contours, sequential Monte Carlo and on-line learning for vessel segmentation. In: European Conference on Computer Vision. LNCS 3953 (2006) 476–489
7. Shim, H., Kwon, D., Yun, I., Lee, S.: Robust segmentation of cerebral arterial segments by a sequential Monte Carlo method: Particle filtering. Computer Methods and Programs in Biomedicine **84** (2006) 135–145
8. Schaap, M., Smal, I., Metz, C., van Walsum, T., Niessen, W.: Bayesian tracking of elongated structures in 3D images. In: Information Processing in Medical Imaging. LNCS 4584 (2007) 74–85
9. Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G.: Medial-based Bayesian tracking for vascular segmentation: Application to coronary arteries in 3D CT angiography. In: IEEE International Symposium on Biomedical Imaging. (2008) 268–271
10. Allen, K., Yau, C., Noble, J.: A recursive, stochastic vessel segmentation framework that robustly handles bifurcations. In: Conference on Medical Image Understanding and Analysis. (2008)
11. Lazar, M., Alexander, A.: Bootstrap white matter tractography (BOOT-TRAC). NeuroImage **24**(2) (2005) 524–532
12. Friman, O., Farnebäck, G., Westin, C.F.: A Bayesian approach for stochastic white matter tractography. IEEE Trans. on Medical Imaging **25**(8) (2006) 965–978
13. Jbabdi, S., Woolrich, M., Andersson, J., Behrens, T.: A Bayesian framework for global tractography. NeuroImage **37**(1) (2007) 116–129
14. Zhang, F., Hancock, E., Goodlett, C., Gerig, G.: Probabilistic white matter fiber tracking using particle filtering and von Mises-Fisher sampling. Medical Image Analysis **13** (2009) 5–18
15. Descoteaux, M., Deriche, R., Knoesche, T., Anwander, A.: Deterministic and probabilistic tractography based on complex fiber orientation distributions. IEEE Trans. on Medical Imaging **28**(2) (2009) 269–286
16. Doucet, A., de Freitas, N., Gordon, N., eds.: Sequential Monte Carlo Methods in Practice. Springer-Verlag (2001)
17. Arulampalam, S., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. on Signal Processing **50**(2) (2002) 174–188
18. Geman, D., Jedynak, B.: An active testing model for tracking roads in satellite images. IEEE Trans. on Pattern Analysis and Machine Intelligence **18** (1996) 1–14
19. Çetingül, H., Plank, G., Trayanova, N., Vidal, R.: Estimation of multimodal orientation distribution functions from cardiac MRI for tracking Purkinje fibers through branchings. In: IEEE Int. Symposium on Biomedical Imaging. (2009)
20. Mardia, K., Jupp, P.: Directional Statistics. Wiley Series in Probability and Statistics (2000)
21. Subbarao, R., Meer, P.: Nonlinear mean shift over Riemannian manifolds. International Journal of Computer Vision **84** (2009) 1–20

Towards Application-specific Multi-modal Similarity Measures: a Regression Approach.*

Olivier Pauly[1], Nicolas Padoy[1,2], Holger Poppert[3], Lorena Esposito[3], Hans-Henning Eckstein[4] and Nassir Navab[1]

[1] Computed Assisted Medical Procedures, Technische Universität München, Germany
[2] LORIA-INRIA Lorraine, France
[3] Neuro-Kopf Zentrum, Klinikum Rechts der Isar, Germany
[4] Klinik für Gefäßchirurgie, Klinikum Rechts der Isar, Germany
{*pauly*, *padoy*, *navab*}@cs.tum.edu

Abstract. In multi-modal registration, similarity measures based on intensity statistics are the current standard for aligning medical images acquired with different imaging systems. In fact, the statistical relationship relating the intensities of two multi-modal images is constrained by the application, defined in terms of anatomy and imaging modalities. In this paper, we present the benefits of exploiting application-specific prior information contained in one ***single*** pair of registered images. By varying the relative transformation parameters of registered images around the ground truth position, we explore the manifold described by their joint intensity distributions. An adapted measure is fitted using support vector regression on the training set formed by points on the manifold and their respective geometric errors. Experiments are conducted on two different pairs of modalities, MR-T1/MR-TOF and MR-T1/SPECT. We compare the results with those obtained using mutual information and Kullback-Leibler distance. Experimental results show that the proposed method presents a promising alternative for multi-modal registration.

1 Introduction

Image registration is a crucial processing step in all image analysis tasks in which information from various imaging sources needs to be combined. Establishing correspondences between images acquired with different medical imaging modalities is a challenging task known as multi-modal registration. Objective functions that evaluate the quality of alignment, known as similarity measures, are optimized to identify the geometric transformation that maps the coordinate system of one modality to the other [8]. The choice of the appropriate measure is not straightforward, because it implicitly models the relationship between the different images to register [4]. Classical measures such as sum of square differences (SSD) or correlation coefficient (CC) make the assumption of a linear functional mapping between the intensities of the images to align. But this hypothesis is far from being realistic according to the physics of different

* This research is partially funded by the Munich Center for Advanced Photonics and by the Gefäßchirurgie Department, Klinikum Rechts der Isar, Germany

imaging systems. Modeling the real relationship between different imaging modalities is very difficult and this explains why statistical measures have become more and more popular. Since its introduction by Viola and Wells [6] and Collignon et al [2], mutual information remains the state of the art of multi-modal registration of medical images.

Even though the statistics relating intensities of two multi-modal images is modality-specific, there were only few attempts to incorporate prior knowledge in such similarity measures. Chung et al. [1] proposed to use as prior information a reference joint probability distribution of registered images from different modalities. Images are then aligned by minimizing the Kullback-Leibler distance between an observed and the expected joint histogram. Leventon et al. [3] compared two methods to model this reference histogram from a training set of registered images, namely a mixture of Gaussians and Parzen windowing. The distance to this expected histogram is then estimated by using log likelihood. In these works however, the use of prior information remains limited to one reference joint distribution.

Zhou et al.[7] propose an approach based on Adaboost to learn local similarity measures for anatomic landmarks detection in echocardiac images. It uses an atlas of the left ventricle containing pairs of local patches with their relative displacements. In a monomodal scenario, the method shows that incorporating prior information can improve the detection results. This approach requires however extensive initial supervision.

Joint histograms of multi-modal images warped with different relative transformations describe a manifold embedded in the joint distribution space. Our contribution is to define a similarity measure relating the topology of such manifolds to the registration error. This yields an application-specific similarity measure, which requires one single pair of registered images as prior information. Using a set of relative transformations between the two images, we generate a training set of data points from the corresponding joint histograms and their associated geometric error values defined in section 2. The similarity measure is then learned by performing a support vector regression on this data.

The remainder of the paper is organized as follows: Section 2 presents our regression approach to define an application-specific similarity measure. Section 3 reports experiments performed in two different and challenging applications in comparison to classical methods such as mutual information and Kullback-Leibler distance. Results show that our approach presents a promising alternative for multi-modal registration. Section 4 concludes the paper and gives an outlook on future work.

2 Methods

2.1 Problem statement

The goal of multi-modal image registration is to identify the geometric transformation that maps the coordinate system of one modality to the other. Let us consider two 2D images defined on the domains Ω_1 and Ω_2 with intensity functions $I_1 : \Omega_1 \subset R^2 \rightarrow R$ and $I_2 : \Omega_2 \subset R^2 \rightarrow R$. The two dimensional case is discussed for better readability, the extension to three dimensions being straightforward. The registration task can be defined as a maximization problem, in which we want to estimate the best transformation T according to a chosen similarity measure S computed on the discrete overlap

domain $\Omega = \Omega_1 \cap T(\Omega_2)$:

$$T = \underset{T}{\operatorname{argmax}}\ S_\Omega(I_1, T(I_2)). \tag{1}$$

The joint intensity distribution $p(I_1, I_2)$ of both images can be evaluated by histogramming or parzen windowing. In most of statistical measures, the similarity S_Ω is a mapping from the joint distribution space $\mathcal{J}$ into $\mathbb{R}$. While Mutual Information (MI) gives a measure of the distance between the joint histogram of both images and what it would be if their intensity distributions were independant, the Kullback-Leibler distance (KL) [1] evaluates the distance between an observed p_o and an expected p_e joint histogram:

$$MI(I_1, I_2) = D(p(I_1, I_2)||p(I_1)p(I_2)) \tag{2}$$
$$KL(I_1, I_2) = D(p_o(I_1, I_2)||p_e(I_1, I_2)), \tag{3}$$

where D in its general form is defined on two histograms p and q as:

$$D(p||q) = \sum_x p(x) log\left(\frac{p(x)}{q(x)}\right). \tag{4}$$

The statistical relationship relating the intensities of two different multi-modal images is constrained by the application. With "application", we mean the combination of the modalities to relate and the different tissues appearing in the imaged anatomy, e.g. blood, bones or muscles. Joint histograms of images warped with different relative transformations describe a manifold $\mathcal{M}$ embedded in $\mathcal{J}$ which is application-specific. In [1], Chung et al. makes use of one expected joint histogram, corresponding to one single reference point on such a manifold. The used Kullback-Leibler divergence is however not adapted to its topology.

Instead, we propose to model an application-specific similarity Ψ taking into account how the topology of $\mathcal{M}$ relates to the registration error. By using a set of relative geometric transformations $\{T_i\}_{i\in\mathbb{N}}$ between a source and a target image, we sample $\mathcal{M}$ by the joint histograms $J_{I_1,T_i(I_2)}$. Each of these "points" is then associated to a geometric error derived from the corresponding transformation parameters, generating thereby a set of data points. Finally, the similarity Ψ is defined by performing a regression on these points. The following section presents how to generate data points to relate this manifold $\mathcal{M}$ to the geometric error.

2.2 Data points generation

Our objective is to model a similarity Ψ learned on the full manifold $\mathcal{M}$:

$$\Psi : \mathcal{M} \rightarrow \mathbb{R}, \tag{5}$$

which has favorable characteristics for registration purposes, namely convexity, smoothness and the ability to estimate the geometric error. To model an accurate mapping Ψ,

the manifold $\mathcal{M}$ must be sampled thoroughly as a function of the transformation T, whose space is parameterized as follows:

$$T(t_x, t_y, \theta) \text{ where } \begin{cases} t_x \in [-M, +M] \\ t_y \in [-N, +N] \\ \theta \in [-\phi, +\phi] \end{cases} \tag{6}$$

By sampling the space of transformations, a set $\{T_i\}_{1 \leq i \leq Q}$ of Q transformations is generated. Then, by using a pair of registered images from different modalities, joint histograms are computed according to these $\{T_i\}_{1 \leq i \leq Q}$. As illustrated by Fig. 1, each joint histogram is then associated to a geometric error value. In medical image registration, the *target registration error* (TRE) permits the evaluation of error in translation and orientation between corresponding structures or organs appearing in both modalities. The TRE is computed by comparing the positions of a set of points $\{p_i,\ 1 \leq i \leq P\}$ after being mapped by the estimated transformation T and by the ground truth transform G:

$$\mathcal{E}(T) = \frac{1}{P} \sum_{i=1}^{i=P} \|T(p_i) - G(p_i)\| \,. \tag{7}$$

This procedure permits us to generate following couples:

$$\left\{ (J_{I_1, T_i(I_2)}, \mathcal{E}(T_i)) \right\}_{1 \leq i \leq Q} , \tag{8}$$

which we denote $\{(J_i, \mathcal{E}_i)\}_{1 \leq i \leq Q}$ for better readability.

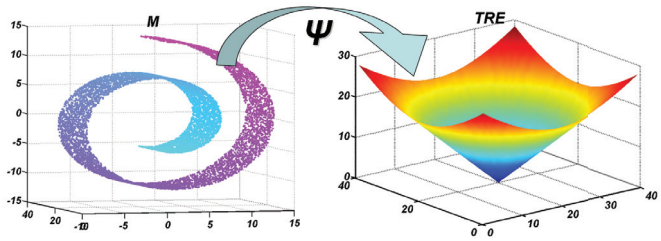


Fig. 1. Our regression approach: learn a similarity Ψ mapping each point of the manifold $\mathcal{M}$ (abtract representation on the left) to a value of the geometric error (on the right).

2.3 Fitting the similarity model through support vector regression

We propose to learn the similarity by approximating the function Ψ with the previously generated data points. Since this function is a high dimensional non-linear mapping, we use support vector regression for its ability of modeling complex non-linear functions. We consider the problem of fitting a similarity function on the set of Q data points $\{(J_i, \mathcal{E}_i)\}_{1 \leq i \leq Q}$. The $\{J_i\}$, as discrete approximations of the joint intensity distributions, consist of $B \times B$ bins. They are linearized into a vector of dimensionality B^2. Let φ be a non-linear mapping from $\mathcal{M}$ into a hidden feature space H with dimensionality $\dim(H) > B^2$ used to model non-linear relationships between joint histograms

and their corresponding geometric error values. The mapping Ψ is modeled by the following function:

$$\Psi(J) = w \cdot \varphi(J) + b, \tag{9}$$

where w is a linear separator of dimensionality $\dim(H)$ and b a bias. The optimal regression function is then given by the minimum of the following functional [5]:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^{Q} \left(\xi_i^+ + \xi_i^-\right), \tag{10}$$

where C controls the flexibility of the model. This functional aims at minimizing the norm of w and the regression errors on the data points, characterized by the slack variables ξ_i^+ and ξ_i^-. The optimal vector w_0 can be written as a linear combination of the training vectors in H with weights $\{\alpha_i\}_{1 \leq i \leq Q}$:

$$w_0 = \sum_{i=1}^{Q} \alpha_i \, \varphi(J_i). \tag{11}$$

The regression function becomes then:

$$\Psi(J) = \sum_{i=1}^{Q} \alpha_i \, \varphi(J_i) \cdot \varphi(J) + b = \sum_{i=1}^{Q} \alpha_i \, K(J_i, J) + b, \tag{12}$$

where K is the kernel associated to φ in H. To handle non-linear relations between the manifold $\mathcal{M}$ and the TRE, K is chosen as a RBF kernel, giving thus the following similarity model:

$$\Psi(J) = \sum_{i=1}^{Q} \alpha_i \, \exp\left(\frac{-|J_i - J|^2}{\sigma^2}\right) + b. \tag{13}$$

3 Experiments and Results

Our regressed similarity measure is evaluated on two challenging applications for multimodal registration: rigid registration of MR-T1 and MR-TOF (Angiography) images of the carotid artery, and of MR-T1 and SPECT images of the brain. In this paper, we focus on 2D rigid-body experiments to prove the concept of our novel approach.

This permits in particular to show that the approach is not limited to pairs of images with a tissue distribution similar to the image pair used for training. Indeed, in the following experiments, a pair of corresponding images from a 3D dataset is used for training. The obtained similarity measure is then evaluated on pairs of images taken from the 3D datasets of the other patients. For statistical relevance, the pairs are chosen randomly and the tests are repeated. It must be noted that the tissue distribution varies depending on the randomly chosen slices, which can originate from the neck or from the head.

3.1 Experimental Setup

Our similarity measure will be compared to normalized mutual information (NMI) and Kullback-Leibler distance (KL) in terms of *success rate*, *accuracy* and *capture range*.

We consider a registration experiment as *successful* when the final target registration error is inferior to a given threshold t_e. In fact, this permits to quantify the ability of an approach to converge in the neighborhood of the right solution. We then define the *accuracy* as the mean target registration error on all registered images after the removal of such outliers. *Capture range* is evaluated by assessing the success rate as function of an increasing initial TRE. Knowing the ground truth position of each dataset, an initial random perturbation is applied to each pair of images according to a given value of TRE. Experiments are then repeated with an increasing initial target registration error.

The objective of our experiments is to highlight the benefits of a similarity measure taking advantage of prior information. Since the convergence to the right solution depends on the topography of the search space offered by a similarity measure, we use a Downhill-Simplex optimizer, that does not require any gradient information. For fair comparison, all measures have the same number of joint histogram bins (32×32) and are tested in the same conditions.

In both experimental setups T1/TOF and T1/SPECT, a cross-validation of N tests is performed on a set of P patients. A test consists of one ***regression step*** performed on a random pair of slices from a given patient and one ***validation step*** consisting of $P - 1$ evaluations performed on the other $P - 1$ patients. During the regression step, our similarity measure and the expected joint histogram needed by KL are computed on the same pair of images. During an evaluation, all measures are tested in the same conditions on a random pair of slices taken from another patient with the same initial perturbation. By using 10 initializations with an increasing TRE per evaluation, we can investigate the ability of each measure to converge towards the right solution and thereby assess their capture range.

The transformation space is sampled as follows: $-40 \leq t_x \leq +40$ (in pixels), $-40 \leq t_y \leq +40$ and $-40 \leq \theta \leq +40$ (in degrees) with a step of 4 for each parameter, generating thereby 9261 data points. For the choice of the hyperparameters σ and C, a grid-search has been performed. All experiments are performed with the Spider environment for MATLAB on an Intel Core 2 Duo CPU 2.40 GHz.

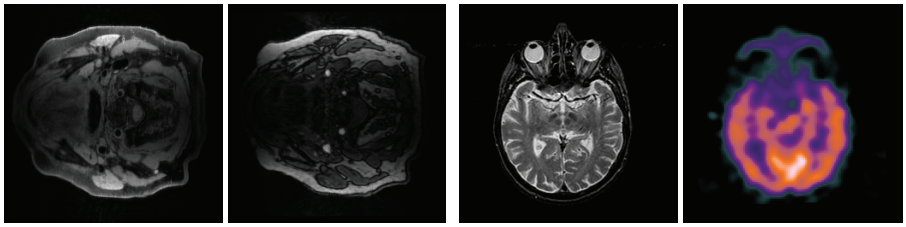


Fig. 2. From left to right: T1 and TOF MR Angiography of the neck of the same patient. Then T1 and SPECT-Tc images of the brain of a healthy patient

MR-T1 and MR-TOF Angiography images: experiments are conducted on images (refer to Fig. 2) taken from $P = 8$ patients (48 pairs of images) with different staging of

atherosclerosis. Both sequences were consecutively acquired, patients were positioned on a vacuum pillow and the acquisition was ECG gated to ensure perfect alignment. Images have a resolution of 128x128 with a pixel size of 2.5mm x 2.5mm. The threshold t_e is set to 1cm which corresponds to 4 pixels. A cross-validation of $N = 32$ tests has been performed, which then corresponds to $N \times (P - 1) \times 10 = 2240$ registration experiments.

MR-T1 and SPECT-Tc images: experiments are conducted on images (refer to Fig.2) taken from $P = 5$ patients (73 pairs of images): a healthy patient, one with a glioma, one with a carcinoma, one with a stroke and finally one with an encephalopathy. These already registered datasets are taken from the publicly available Whole Brain Atlas database. Images have a resolution of 128x128 with a pixel size of 1.67mm x 1.67mm. The threshold t_e is set to 1cm which corresponds to 6 pixels. A cross-validation of $N = 40$ tests has been performed, which then corresponds to $N \times (P-1) \times 10 = 1600$ registration experiments.

3.2 Results

The objective of our experiments is to show the benefits of a similarity measure taking full advantage of prior information. As shown on Fig. 3, the optimal regression model provides a smooth and convex search space, which is very close to the original TRE surface to approximate. Moreover, the global optimum has been preserved at the right position. In fact, smoothness and convexity are crucial characteristics to prevent the optimizer of being stuck in a local optimum and to ensure its convergence to the global one. The great advantage of our approach is its ability to model the convexity, the smoothness and the capture range of the similarity measure. Indeed, its convexity can be changed by choosing another function of the geometric error. The choice of hyperparameters C and σ influences the flexibility of the regression and thus the smoothness of the resulting function. During the regression process, increasing the sampling range of the transformation space permits to increase the capture range of the trained similarity. A high capture range is crucial when no good initialization parameters are available. Results presented in Fig. 3 shows the overall success rate and the final TRE as functions of the initial TRE. While the success rate of other measures sinks with an increasing initial TRE, our regressed similarity measure shows a good behaviour. This highlights its greater capture range and this, for a better accuracy. In the T1-TOF experiments, KL provided once a better accuracy for an initial TRE of 22.5 mm. This comes from the fact that KL was only successful on three registration experiments: Ψ and MI were actually better than KL in these specific experiments, but in the displayed results their accuracy is averaged on many more experiments as they have much higher sucess rates.

Our method was robust face to different tissue distributions, e.g. coming from patients affected by different kinds of disease or from different locations of the head that were not learned during the regression phase. For example, while slices from the top of the skull contain mostly skin, bone, cerebrospinal fluid, grey and white matter, slices in the middle of the head also consists of muscles and eyes. This could suggest that the manifold on which the similarity was learned is not strictly dependent on the anatomy. This needs however to be extensively studied with further experimentations.

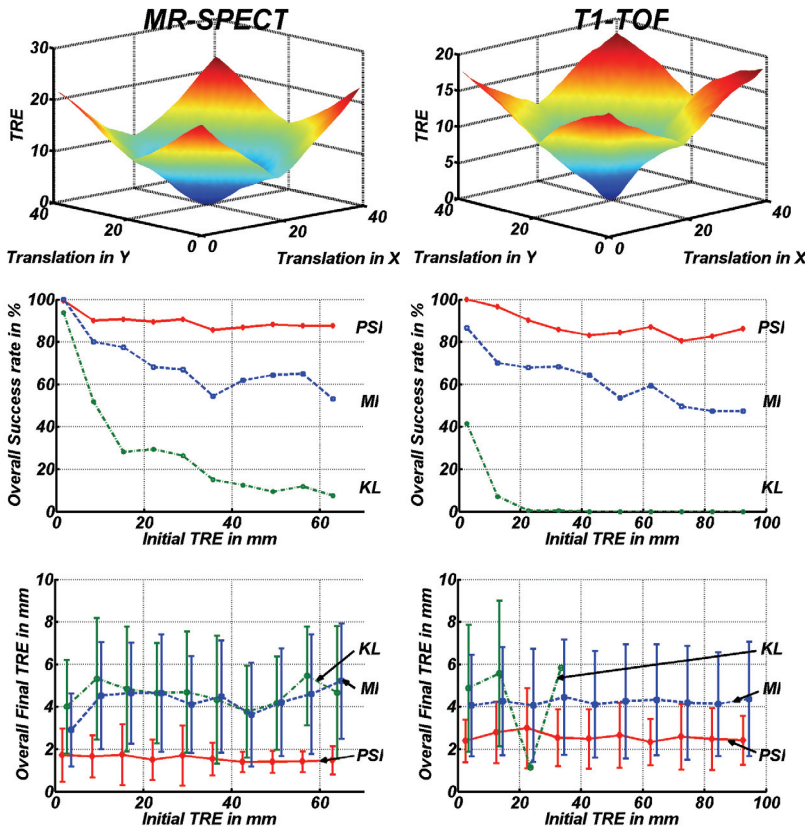


Fig. 3. From top to bottom: Plot of the similarity Ψ (PSI) for variations in translation in x and y between -20 and $+20$ pixels. Plot of the success rate (in percent) and final TRE (mean and standard deviation in mm) according to an increasing initial TRE. Left MR-SPECT, right T1-TOF

4 Discussion and Conclusion

In this work, we propose to take advantage of prior information, namely a registered pair of images, in order to improve results in multi-modal registration. Our contribution is to define, with a regression approach, a new similarity measure relating the manifold described by joint histograms of two different modalities to the registration error. Experiments conducted on MR-T1/MR-TOF and MR-T1/SPECT images show that the presented method is a promising alternative for multi-modal registration. We empirically demonstrated that these manifolds are not dependant on the choice of the particular training pair within the dataset. This means that such an adapted application-specific measure can be defined by using a single pair of manually registered images from the specific application. Moreover, its robustness to different or new tissue distributions suggests that such manifolds could be modality-specific. In future work, we will further study their dependence to the variations of tissue distribution within the images.

References

1. A. Chung, W. Wells, A. Norbash, and E. Grimson. Multi-modal image registration by minimising kullback-leibler distance. *Proc. of MICCAI Conf.*, pp:525–532, 2002. 2, 3
2. A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3d multi-modality medical image registration using feature space clustering. *Proc. of Computer Vision, Virtual Reality and Robotics in Medicine Conf.*, pp:195–204, 1995. 2
3. M. E. Leventon and E. Grimson. Multi-modal volume registration using joint intensity distributions. *Proc. of MICCAI Conf.*, pp:1057–1066, 1998. 2
4. A. Roche, G. Malandain, N. Ayache, and S. Prima. Towards a better comprehension of similarity measures used in medical image registration. *Proc. of MICCAI Conf.*, pp:555–567, 1999. 1
5. A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14:199–222, 2004. 5
6. P. Viola and W. Wells. Alignment by maximization of mutual information. *IJCV*, 24(2):137–154, 1997. 2
7. S. K. Zhou, J. Zhou, and D. Comaniciu. A boosting regression approach to medical anatomy detection. *Proc. of CVPR Conf.*, pp:1–8, 2007. 2
8. L. Zollei, J. Fisher, and W. Wells. A unified statistical and information theoretic framework for multi-modal image registration. *Proc. of IPMI Conf.*, pp:366–377, 2003. 1

Automatic Region Template Generation for Shape Particle Filtering based Image Segmentation*

Lukas Fischer[1,2], Rene Donner[1,3], Franz Kainberger[1], Georg Langs[4,1]

[1] Medical University of Vienna, Department of Radiology,
Computational Image Analysis and Radiology Lab
{lukas.a.fischer, rene.donner, franz.kainberger,
georg.langs}@meduniwien.ac.at
[2] Vienna University of Technology, Institute of Computer Aided Automation,
Pattern Recognition and Image Processing Group
[3] Graz University of Technology, Institute for Computer Graphics and Vision
[4] Massachusetts Institute of Technology,
Computer Science and Artificial Intelligence Lab

Abstract. Segmentation approaches based on sequential Monte Carlo Methods deliver promising results for the localization and delineation of anatomical structures in medical images. Also known as Shape Particle Filters, they are used for the segmentation of human vertebræ, lungs and hearts, and are especially well suited to cope with the high levels of noise encountered in MR data and overlapping structures with ambiguous appearance in radiographs.
They require a region template of the appearance features which allow to estimate the confidence in the hypotheses generated during the search. Currently these templates are created manually, which introduces a bias, and leads to particularly sub-optimal results in complex anatomy.
In this work we propose a Differential Evolution based Shape Particle Filter segmentation scheme where the optimal distribution and number of template regions is derived automatically from a set of training images. The method adapts to complex data and finds consistent features in the training examples. Experiments on two medical data sets (radiographs of metacarpal bones and MRI slices of hearts) show that this yields considerably higher accuracy with fewer outliers.

1 Introduction

Motivation Statistical methods such as sequential Monte Carlo Methods were proposed for detection, segmentation [1, 2, 3, 4, 5] and tracking [6, 7, 8] of objects. A similar approach, called Shape Particle Filters was introduced in [2, 3] for the segmentation of vertebræ, lungs and hearts. In Fig. 1 an illustration of a Shape Particle Filter approach is depicted: Based on a global shape model

* This work has been supported by the Austrian National Bank Anniversary Fond projects COBAQUO and BIOBONE.

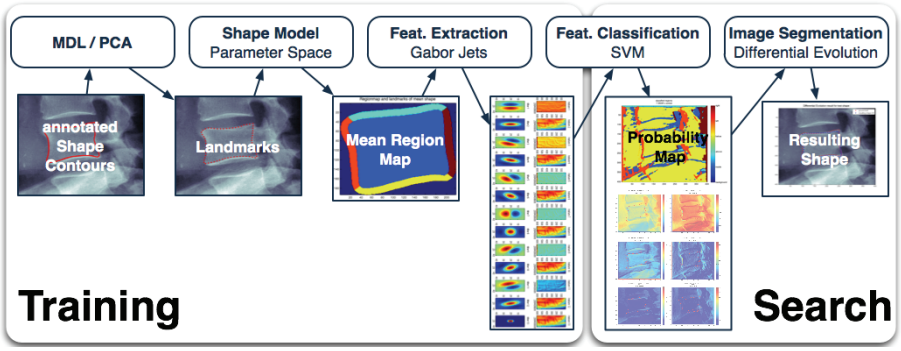


Fig. 1. Flowchart of a Shape Particle Filter. Landmarks are obtained from manually annotated contours, forming the basis for the generated Shape Model. A region map based on the shape models mean shape is generated that in turn is used to determine image region features. Test images are then classified yielding probability maps that are used in the final image segmentation step.

a *class template* or *labeling map* for the following feature extraction is defined according to the respective medical object of interest (e.g. interior of the contour and one or more regions within a certain border on the outside). The number and location of these regions however is currently defined manually. Using these regions and their corresponding distributions in the feature space particle filtering approaches are used to estimate the most probable point in the shape parameter space (corresponding to a segmentation of the image) for a given test image. This is achieved by sampling from the image features according to shape hypothesis and computing corresponding confidence values / posterior probabilities. Estimating this posterior probabilities over the parameter space allows to find values of maximum confidence, i.e. to optimize the fit of the model to the object in the test image.

Particle Filters Particle filtering was introduced with the intention to implement recursive Bayesian filters [9]. It is also known as Sampling Importance Resampling (SIR), Bayesian bootstrap filter or *sequential Monte Carlo Methods*. In contrast to other filters that use Monte Carlo Methods to get estimates of the mean and covariance of the posterior, particle filters approximate the complete posterior. They aim to approximate posterior densities using swarms of points (so called *particles*) in a sample space. A weight is assigned to each particle and using a discrete distribution of the particles the posterior distribution can be approximated. This results in particle probabilities which are proportional to the particle weights. Several algorithms exist differing mainly in the way how the particle swarms evolve and adapt to input data [10].

Contribution In this paper we propose an approach to automatically select the subregions of the object required to compute the expected feature distributions

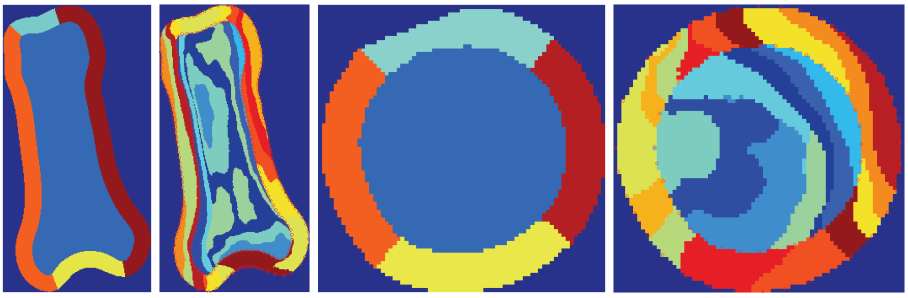


Fig. 2. Manually predefined region maps (a,c) with five regions and examples of automatically generated region maps (b,d) with 17 (8 inner and 9 outer) regions for the metacarpal bones and the hearts.

and the confidences for each shape hypothesis. The automatic region map is based on the shape of the objects to be classified. Two areas are defined, the interior of the shape and to constrain the representation to a neighborhood of the object a corridor on the outside of the shape. The subregions are estimated for each of the two areas by clustering in feature space.

The optimal number and location of inner and outer regions (clusters) is found by leave one out cross validation. A comparison of manually predefined region maps and examples of automatically generated region maps for each of the processed data sets is shown in Fig. 2. This has several advantages over the manual choice of template regions: the main feature is a partitioning adapted to complex but consistent anatomical structures, which hold information relevant to the segmentation process, but cannot be determined optimally by manual segmentation. It removes bias and finally it opens the possibility of autonomous learning approaches, that do not require manual supervision on a training population.

We evaluate our method on radiographs of human metacarpal bones and MRI slices of human hearts. The results show the drastically improved search results compared to a manual region definition, with both higher accuracy and a smaller number of outliers.

Paper Structure Sec. 2 gives a description of the fundamental methods of the proposed approach, namely the concept of Shape Models 2.1, Differential Evolution 2.2 and the resulting Shape Particle Filter 2.3. Sec. 3 details the automatic region estimation and the experiments and results are presented in Sec. 4 followed by the conclusion and an outlook in Sec. 5.

2 Methods

Shape Particle Filters rely on a model of the objects' shapes and a means of estimating the posterior distribution of parameter estimates given an input im-

age. These are estimated by means of an Markov Chain Monte Carlo (MCMC) method, where Differential Evolution MCMC is used in this work.

2.1 Shape Models

Objects in images can be represented using statistical models of the objects' shape. A Point Distribution Model (PDM) [11] constructs a shape model by computing the significant eigenmodes of a shape population, e. g. assuming a multivariate Gaussian distribution of the shape parameters.

These shapes consist of a set of n points or landmarks. The points can have any dimension, but throughout this paper they are considered to be 2-dimensional. Before modeling the non-rigid shape variation we normalize for similarity transformations, i. e. translation, rotation and scaling (the *pose*). Thus the parameters to define a unique shape in an image are shape and pose parameters.

Using this shape information the goal is to build models which provide the abilities to represent shapes and to generalize to new shapes within the distribution of the shapes in a predefined training set.

The first step to create a training set for shape model generation is to define landmarks for several objects. To obtain landmarks for a shape usually a human expert annotates several images containing the corresponding object. Landmarks are derived from this manually annotated image contours by applying Minimum Description Length (MDL) [12]. Then a vector $\mathbf{v}_i$ for all $i \in 1, ..., N_S$ annotated shapes for d-dimensional landmarks is defined as

$$\mathbf{v}_i = (d_{1_1}, ..., d_{1_d}, d_{2_1}, ..., d_{2_d}, ..., d_{n_1}, ...d_{n_d}) \tag{1}$$

The training set is then aligned using Procrustes Analysis, which minimizes $\sum |\mathbf{v}_i - \bar{\mathbf{v}}|^2$, where $\mathbf{v}_i$ is the i^{th} point position vector and $\bar{\mathbf{v}}$ is the mean of all vectors.

To be able to generate new shapes out of the training set a parameterized model $\mathbf{v} = M(\mathbf{b})$ of the distribution of the N_S point position vectors $\mathbf{v}_i$ is defined. $\mathbf{b}$ is a vector containing the model parameters (Eq. 3). With the help of this model it is possible to generate new shapes $\mathbf{v}$ and to estimate the distribution $p(\mathbf{v})$ of these new vectors. The model is finally built by applying Principle Component Analysis (PCA) to the data, yielding eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_e$ with $e = min(nd, N_S)$. Using $e^* < e$ eigenvectors (thereby neglecting the modes with small variance, which are considered to model noise only) any shape $\mathbf{v}$ within the subspace spanned by the training set can be represented by:

$$\mathbf{v} \approx \bar{\mathbf{v}} + \mathbf{F}\mathbf{b} \tag{2}$$

where $\mathbf{F} = (\mathbf{e}_1, ..., \mathbf{e}_{e^*})$ is the basis of the eigenspace and $\mathbf{b}$ is a vector of length e^* which defines the parameters for the deformable model [13]:

$$\mathbf{b} = \mathbf{F}^T(\mathbf{v} - \bar{\mathbf{v}}) \tag{3}$$

The resulting model $\mathbf{F}$ represents the shape variation of the modeled objects utilizing a single parameter vector $\mathbf{b}$. Each element of $\mathbf{b}$ controls one mode of shape variation, with the first modes being responsible for the highest variation, in descending order.

To ensure that the generated shapes are similar to those in the training set the parameter vectors are always limited to the range $\pm 3\sqrt{\lambda_i}$, where λ_i is the eigenvalue and also the variance of the i^{th} parameter b_i in the training set.

In addition to these modes the transformations translation, scaling and rotation need to be taken into account. Therefore a new linear parameter vector $\mathbf{t} = (s_x, s_y, t_x, t_y)^T$ is introduced, controlling rotation θ, scaling s and translation (t_x, t_y), with $s_x = s\cos\theta - 1$ and $s_y = s\sin\theta$.

Combining the parameter vector of the PCA $\mathbf{b}$ and the parameter vector for translation, rotation and scaling $\mathbf{t}$ results in the combined parameter vector

$$\mathbf{c} = (\mathbf{b}^T, \mathbf{t}^T). \tag{4}$$

2.2 Differential Evolution

We approach particle filtering using the DE-MCMC formulation introduced by [14], that uses Differential Evolution (DE) [15] for the sampling step in sequential Monte Carlo Methods. DE is a genetic algorithm and aims to optimize functions based on populations in parameter space, which in our case is the subspace (restricted to plausible models) of the model parameters $\mathbf{c}$.

DE is a parallel direct search method that uses N_x d-dimensional parameter vectors $\mathbf{x}_i$ $(i = 1, ..., N_x)$ (i.e. N_x d-dimensional Markov chains [14]) as members of a population $\mathbf{X}_g$ for each generation g.

Starting with parameter vectors randomly drawn from the training distribution at generation $g = 1$, during each generation $(g + 1)$, N_x new parameter vectors / shape hypotheses $\mathbf{x}^h$ are then generated by adding the weighted difference vector between two population members to a third member

$$\mathbf{x}^h = \mathbf{x}_1 + \gamma(\mathbf{x}_2 - \mathbf{x}_3), \tag{5}$$

where $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are randomly selected without replacement from the population $\mathbf{X}_g$ and γ is a constant factor weighting the differential variation $\mathbf{x}_2 - \mathbf{x}_3$ ($\gamma = 0.85$ in our experiments). If the confidence $\pi(\mathbf{x}^h)$ in hypothesis $\mathbf{x}^h$ is higher than $\pi(\mathbf{x}_1)$, $\mathbf{x}^h$ replaces $\mathbf{x}_1$ in generation $\mathbf{X}_{g+1}$, otherwise $\mathbf{x}_1 \in \mathbf{X}_{g+1}$. After I iterations ensuring convergence (200 in our experiments) the hypothesis / population member $\mathbf{x}^*$ with the highest confidence is considered to represent the best solution.

To guarantee a detailed balance of proposal and acceptance with respect to the fitness function $\pi(.)$ Eq. 5 is modified to

$$\mathbf{x}^h = \mathbf{x}_1 + \gamma(\mathbf{x}_2 - \mathbf{x}_3) + \mathbf{k} \tag{6}$$

where $\mathbf{k}$ is drawn from the normal distribution $\mathbf{k} \sim N(0, \mathbf{a})$ with variance $\mathbf{a}$ small compared to the variance of population $\mathbf{X}_g$ [14].

2.3 Shape Particle Filter

Shape Particle Filters as proposed in [2, 3] contain the following steps. First the mean shape is derived from hand annotated training images. Then a region map is manually defined on this mean shape, representing regions which are presumed to be distinct and of significance to the segmentation process (see below). An example of these manually defined regions can be seen in Fig. 2(a) and 2(c). A small corridor (20 pixel for the metacarpal bones and 10 pixel for the hearts) is created defining the region around the shape. This area outside the shape contains information about surrounding neighboring structures that is taken into account during classification.

Local image descriptors (e.g. gaussian derivative filters, Gabor filters [16]) are computed for all training images. The region map is warped back to each training shape using Thin Plate Spline Warping (TPS) [17], so that for each region a distribution of the corresponding descriptors/features can be estimated by sampling. The shape model, the region map and the learnt feature descriptors for the regions thus constitute the prior knowledge of Shape Particle Filters.

During search on a test image, a k-NN classifier is used to classify the image's pixels $j \in 1 \dots N_J$ resulting in one region probability map $P(j \mid l)$ per region l. The actual segmentation step uses this probabilities to optimize a fitness function encoding the belief in the segmentation corresponding to a given shape model parameter vector.

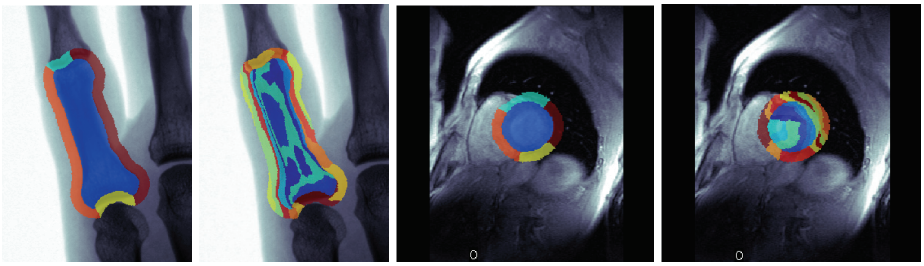
This optimization is performed using particle filtering by importance resampling processes [2, 3], estimating the posterior distribution of the shapes given the image by means of the following fitness function (Eq. 7). Initially, a random set of shapes, the particles, represented by shape parameter vectors $\mathbf{c_i}$ are generated according to the distribution of the prior shape model.

The region map is deformed according to the shape parameters and for each region the cumulative probability is computed by summing up the probabilities from the corresponding probability map.

By this, a weight can be applied to each particle considering their likelihood,

$$\pi(\mathbf{c}_i) = \frac{\sum_{l=1}^{L} \sum_{j=1}^{n} P(j \mid l)}{L}, \tag{7}$$

where L is the number of regions, n is the number of pixels in the region map and $P(j \mid l)$ is the probability of pixel j belonging to region l. New particles are generated from the current set of particles by weighting them with their likelihood $\pi(\mathbf{c}_i)$ and randomly sampling in parameter space around these particles with probabilities proportional to the weights. While this importance resampling process is repeated the initial sparse particles evolve into a distribution with high density around the most likely shapes.



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

Fig. 3. Manually predefined and automatically generated region maps for the metacarpal bones (a),(b) and the hearts (c),(d). The automatic results show how the Gabor features are clustered into regions which are considerably different from the ones chosen by humans.

3 Automatic region map generation

The manual definition of the region map represents the introduction of a strong bias, as there is no guarantee that their definition is beneficial or at least suited to the convergence and accuracy of the optimization scheme. In contrast to previous approaches we thus propose a method which derived an optimal region map directly from the training image data. It takes into account the discriminative power of the computed image features and reflects their distribution in the region map.

Similar to the approach in Sec. 2.3 an average feature vector $\mathbf{f}_i$ (Gabor Jets) is computed for each pixel i within the mean shape (as well as within a border around the mean shape) from the corresponding features $\mathbf{f}_i^n$ from each training image $n \in 1 \dots N_S$. The task is now to estimate a sensible partition, i. e. region map, of the inner and outer/border of the mean shape such that the probabilities $P(j \mid l)$ of the image classification using this partition convey information which makes the particle filter converge at a good segmentation. The number of regions within the mean shape is denoted by N_{inner} and the number of regions within the border by N_{border}.

First the feature matrix $\mathbf{F}_n$ for each training image is extracted. Then the mean feature matrix

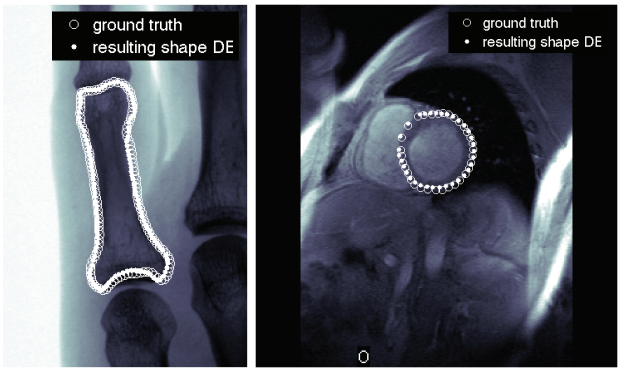
$$\bar{\mathbf{F}} = \frac{\sum_{n=1}^{N} \mathbf{F}_n}{N_S} \tag{8}$$

is calculated from the individual feature maps after they have been warped onto the mean shape. The area inside the mean shape is clustered into N_{inner} regions and the corridor is clustered into N_{border} regions. The resulting region map with $L = N_{inner} + N_{border}$ regions is for training the classifier and finally for the segmentation step of the Shape Particle Filter.

To obtain the optimal number of regions leave one out cross validation (LOOCV) is used. For each test shape all 100 possible region number combi-

	k-NN	kd-tree + k-NN	linear SVM	speed gain vs. k-NN	speed gain vs. kd-tree + k-NN
Metacarpal bones	641.4	363.8	70.5	9.1x	5.2x
Hearts	239.8	134.8	19.9	12.1x	6.8x

Table 1. Mean classification speed over 10 runs for one test image in seconds for the k-NN, the kd-tree + k-NN and the linear SVM. In the rightmost column the speedup achieved by using SVMs compared to the other classification algorithms is shown (SVM times faster than k-NN and kd-tree + k-NN).



(a) (b)

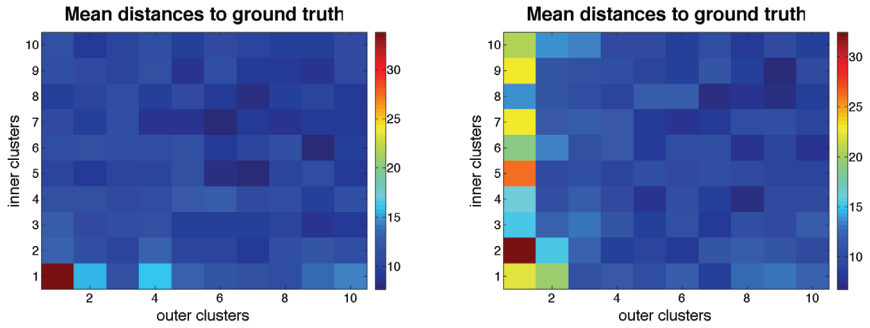
Fig. 4. Example images of the metacarpal bones (a) and hearts (b) data set with both ground truth and segmentation result. The objects were both automatically localized in the image and delineated.

nations ($N_{inner} \times N_{border}$ where $N_{inner}, N_{border} \in 1, \ldots, 10$) are generated and the segmentation results are computed. By comparing the segmentation results for each region pair and selecting those with the minimal landmark error to ground truth, the optimal number of regions for the region map is determined. Examples of these automatic region maps and manual region maps are shown in Fig. 3.

During the segmentation process the pixels of the target test image are classified using a linear Support Vector Machine (SVM [18, 19, 20, 21]), yielding probabilities for each pixel for belonging to the L regions. By using linear SVMs a major speed up compared to a simple k-Nearest Neighbors (k-NN) approach as well as compared to k-NN in combination with a kd-tree [22] based distance computation could be achieved (see Tab. 1).

4 Experiments

In this section the experimental setup and results, as well as a discussion of the proposed results are presented.



(a) metacarpal bones (b) hearts

Fig. 5. Visualization of the mean landmark error for all evaluated combinations of region numbers over three LOOCV runs for the metacarpal bones (a) and the hearts (b). As can be expected, using a too small number of regions results in poor results as the resulting region probability estimates do not convey enough geometric information for the optimization.

Setup Evaluation was performed on radiographs of metacarpal bones and on MRI slices of the heart. The shapes in the data sets were annotated manually forming the ground truth. For each data set and parameter combination $3N_S$ segmentation runs were performed, i.e. three full leave-one-out cross validations. For the first experiment manually predefined region maps and for the second automatically generated region maps were used as presented in Sec. 3. A SVM with linear kernel was trained for the classification task. The decision for a linear kernel was made to achieve a tradeoff between computational performance and classification result quality. Furthermore the SVM was used to compute probability estimates instead of definite class labels as required for the particle filter segmentation. To eliminate all other factors, the identical optimization scheme using Differential Evolution was used for both experiments, such that the results illustrate the influence of manual vs. automatic region definition only. For each resulting segmentation (see Sec. 2) the mean euclidean distance (mean landmark error) to the ground truth was calculated.

The data sets used for evaluation were **1)** 15 radiographs of human metacarpal bones with the resolution of approximately 500×400 pixels each. **2)** 14 short-axis, end-diastolic cardiac MRI slices of the human heart with the resolution of 256×256 pixels with manually placed landmarks on the epicardial and endocardial contours [23]. Example images for both data sets including ground truth and exemplary segmentation results are shown in Fig. 4.

Results In the comparison of the results that were generated using the manual and automatic region maps the main focus was on the landmark error. As mentioned above the automatic region map leading to the minimal landmark error was obtained using LOOCV over all region combinations. The resulting mean landmark errors for all region pairs of the respective data set are visual-

		median	mean	std
Metacarpal bones	manual	10.12	12.51	9.88
	auto	**4.96**	**7.21**	**6.92**
Hearts	manual	5.80	8.96	11.26
	auto	**4.36**	**5.10**	**3.59**

Table 2. Resulting median, mean and standard deviation of the landmark error in pixel for the metacarpal bones with 12 (5 inner, 7 border) regions and for the hearts with 18 (9 inner, 9 border) regions in the auto region map (our approach) compared to the results obtained using the manually defined region map. Note how the accuracy of the resulting segmentation is considerably increased.

ized in Fig. 5. It can be observed that the results for the automatic regions show that both the number and location of the manual regions are not sufficient to capture the spacial feature distribution in the images. In fact for both data sets good segmentation results were achieved with region combinations from $5 - 9$ inner and border regions.

In Fig. 6 boxplots of the resulting landmark errors for all four distinctive runs (manual vs. automatic region map each for the metacarpal bones and the hearts) are shown. For the metacarpal bones a auto region map with 5 inner and 7 border regions and for the hearts a region map with 9 inner and 9 border regions led to the best segmentation results and therefore to the minimal landmark error. Therefore the following key values in pixel for the different data sets could be achieved: The median of the landmark error for the metacarpal bones could be reduced from 10.12 to 4.96 and for the hearts from 5.8 to 4.36. The mean landmark error did also decrease for the metacarpal bones from 12.51 to 7.21 and for the hearts from 8.96 to 5.10. The standard deviation of the landmark error for the metacarpal bones was reduced from 9.88 to 6.92 and for the hearts from 11.26 to 3.59. A summary of these values is shown in Table 2.

Discussion The quality of the generated results depends on four major factors besides the used region map. These are the image quality of the used input data (that in turn influences the extracted image features), overlaps of or no space between nearby similar structures, the accuracy of the manually annotated ground truth as well as the accuracy of the used image classification scheme. Input image quality is used as a collective term for the amount of noise, shadows of overlapping anatomical structures or other distortions due to the image creation process.

The approach using automatically generated region maps outperformed the manually defined version especially on images with low input image quality. Due to the better incorporation of the feature information of the structure and its surroundings, the auto region maps lead to more precise classifications and therefore to more exact segmentation results. Furthermore the manual region maps could not provide the necessary level of detail to cope with nearby or overlapping similar structures because of the lack of distinctive regions in crucial image areas. In particular this was observed on the metacarpal bones data set, where

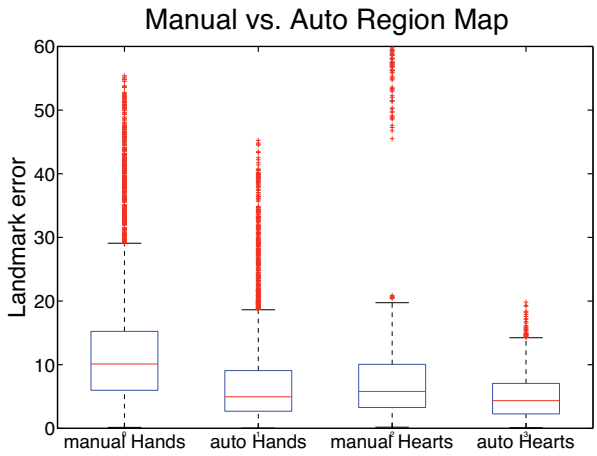


Fig. 6. Boxplots of the results for the image segmentation using manually and automatically generated region maps for the metacarpal bones and the hearts data sets. Using the automatic region estimation results in fewer outliers and higher accuracy, displaying the effectiveness of the proposed approach.

the main problem arose at the distal bone region i.e. the joint area where extreme narrow inter-bone spaces appeared. A selection of different image features incorporating local structures or even automatically learned features for these areas could lead to a better segmentation accuracy.

5 Conclusion

In this work we have shown that using manually defined region maps for shape particle filtering introduces a bias which reduces segmentation accuracy considerably. By using region maps which are automatically derived through clustering in the feature space the optimal number and distribution of inner and outer regions is found, leading to substantially increased accuracy, as shown on the two medical data sets. Furthermore, the laborious estimation of a suitable manual region map trough trial and error is eliminated, paving the way to rapid application of Shape Particle Filters in clinical scenarios. Future work will focus on the investigation of the optimal selection of feature types for specific target objects. The method's extension to 3D will allow to use its capability to localize and segment anatomical structures to further modalities.

References

[1] Seise, M., Mckenna, S.J., Ricketts, I.W., Wigderowitz, C.A.: Parts-based segmentation with overlapping part models using markov chain monte carlo. Image and Vision Computing **27**(5) (Feb 2009) 504–513

[2] de Bruijne, M., Nielsen, M.: Shape particle filtering for image segmentation. Proc. MICCAI 2004 **3216** (Jan 2004) 168–175
[3] de Bruijne, M., Nielsen, M.: Image segmentation by shape particle filtering. Proc. ICPR **3** (2004) 722–725
[4] Bruijne, M.D., Nielsen, M.: Multi-object segmentation using shape particles. Lecture Notes in Computer Science: Information Processing in Medical Imaging **3565** (Jan 2005) 762–773
[5] de Bruijne, M.: Shape particle guided tissue classification. Mathematical Methods in Biomedical Image Analysis (MMBIA) (Apr 2006) 8
[6] Doucet, A., Freitas, N.D., Gordon, N.: Sequential monte carlo methods in practice. Springer (Jan 2001) 581
[7] Liebelt, J., Schertler, K., Germany, E.: Precise registration of 3d models to images by swarming particles. CVPR (2007)
[8] Sørensen, L., Østergaard, J., Johansen, P., Bruijne, M.D.: Multi-object tracking of human spermatozoa. Proceedings of SPIE (Jan 2008)
[9] Gordon, N., Salmond, D., Smith, A.: Novel approach to nonlinear/non-gaussian bayesian state estimation. IEEE Proceedings for Radar and Signal Processing **140**(2) (Jan 1993) 107–113
[10] Fearnhead, P.: Sequential monte carlo methods in filter theory. (Jun 2008)
[11] Cootes, T., Taylor, C., Cooper, D., Graham, J.: Training models of shape from sets of examples. In Proc. British Machine Vision Conference (Jan 1992) 9–18
[12] Thodberg, H.H.: Minimum description length shape and appearance models. Proceedings of Information Processing and Medical Imaging (2003)
[13] Cootes, T., Taylor, C.: Statistical models of appearance for computer vision. World Wide Web Publication (Jan 2001)
[14] Braak, C.: A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. Stat Comput **16**(3) (Jan 2006) 239–249
[15] Storn, R., Price, K.: Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11**(4) (Jan 1997)
[16] Yoshimura, H., Etoh, M., Kondo, K., Yokoya, N.: Gray-scale character recognition by gabor jets projection. Proc. ICPR **2** (Jan 2000) 335–338
[17] Bookstein, F.: Principal warps: thin-plate splines and the decomposition of deformations. Pattern Analysis and Machine Intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(6) (Jun 1989) 567 – 585
[18] Vapnik, V.N.: The nature of statistical learning theory. Springer (2000)
[19] Burges, C.: A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery (Jan 1998)
[20] Scholkopf, B., Smola, A., Williamson, R., Bartlett, P.: New support vector algorithms. Neural Computation (Jan 2000)
[21] Bishop, C.M.: Pattern recognition and machine learning. Springer (Jan 2006)
[22] Bentley, J.: Multidimensional binary search trees used for associative searching. Communications of the ACM **18**(9) (Jan 1975) 509 – 517
[23] Stegmann, M., Fisker, R., Ersboll, B.: Extending and applying active appearance models for automated, high precision segmentation in different image modalities. Proceedings of the Scandinavian Converence on Image Analysis (Jan 2001)

Nonparametric Mixture Models for Supervised Image Parcellation

Mert R. Sabuncu[1], B.T. Thomas Yeo[1], Koen Van Leemput[1,2,3], Bruce Fischl[1,2], and Polina Golland[1]

[1] Computer Science and Artificial Intelligence Lab, MIT, USA
[2] Department of Radiology, Harvard Medical School, USA
[3] Dep. of Information and Comp. Sci., Helsinki University of Technology, Finland

Abstract. We present a nonparametric, probabilistic mixture model for the supervised parcellation of images. The proposed model yields segmentation algorithms conceptually similar to the recently developed label fusion methods, which register a new image with each training image *separately*. Segmentation is achieved via the fusion of transferred manual labels. We show that in our framework various settings of a model parameter yield algorithms that use image intensity information differently in determining the weight of a training subject during fusion. One particular setting computes a single, global weight per training subject, whereas another setting uses locally varying weights when fusing the training data. The proposed nonparametric parcellation approach capitalizes on recently developed fast and robust pairwise image alignment tools. The use of multiple registrations allows the algorithm to be robust to occasional registration failures. We report experiments on 39 volumetric brain MRI scans with expert manual labels for the white matter, cerebral cortex, ventricles and subcortical structures. The results demonstrate that the proposed nonparametric segmentation framework yields significantly better segmentation than state-of-the-art algorithms.

1 Introduction

Supervised image parcellation (segmentation) tools traditionally use atlases, which are parametric models that summarize the training data in a single coordinate system [1–9]. Yet, recent work has shown that more accurate segmentation can be achieved by utilizing the entire training data [10–16], by mapping each training subject into the coordinates of the new image via a pairwise registration algorithm. The transferred manual labels are then fused to generate a segmentation of the new subject. There are at least two advantages of this approach: (1) across-subject anatomical variability is better captured than in a parametric model, and (2) multiple registrations improve robustness against occasional registration failures. The main drawback of the label fusion (multi-atlas) approach is the computational burden introduced by the multiple registrations and the manipulation of the entire training data.

Early label fusion methods proposed to transfer the manual labels to the test image via nearest neighbor interpolation after pairwise registration [12, 14].

Segmentation labels of the test image were then estimated via majority voting. Empirical results suggested that errors in the manual labeling and in registration are averaged out during label fusion, resulting in accurate segmentation. More recent work has shown that a weighted averaging strategy can be used to improve segmentation quality [11]. The basic idea is that training subjects more similar to the test subject should carry more weight during label fusion. The practical advantages of various strategies based on this idea have lately been demonstrated [11, 13, 16]. Some of these strategies use the whole image to determine a single, global weight for each training subject [11, 15, 16], whereas others use local image intensities for locally adapting the weights [11, 13].

This paper presents a novel unified probabilistic model that enables local and global weighting strategies within a label fusion-like segmentation framework. We formulate segmentation using MAP, where a nonparametric approach is used to estimate the joint density on the image intensities and segmentation labels of the new subject. The proposed framework generalizes a model we recently presented at MICCAI 2009 [16], which is based on the assumption that the test subject is generated from a single, unknown training subject. That specific model leads to a segmentation algorithm that assigns greater importance to the training subjects that are globally more similar to the test subject and can be viewed as a particular instantiation of the more general approach presented in this paper. In addition, the proposed approach further enables two possible variants within the same framework. First, we present a local mixture model that assumes each voxel in the test image is generated from some training subject with a uniform prior, independently of other voxels. This local model yields a pixel-wise weighting strategy in segmentation. Second, we develop a semi-local mixture model that relaxes the independence assumption of the local model with a Markov Random Field prior. This model leads to a weighting strategy where intensity information in a local neighborhood is pooled in a principled manner.

In related literature, soft weighting of training subjects was recently used for shape regression [17], where the weights depended on the subjects' age. The proposed nonparametric parcellation framework is also parallel to STAPLE [18], which fuses multiple segmentations of a single subject. In contrast, our framework handles multiple subjects and accounts for inter-subject variability through registration.

The paper is organized as follows. The next section presents the non-parametric generative model for image segmentation. In section 3, we discuss three instantiations of the framework. In section 4, we present inference algorithms for these instantiations. We conclude with experiments in section 5. We report experiments on 39 brain MRI scans that have corresponding manual labels, including the cerebral cortex, white matter, and sub-cortical structures. Experimental results suggest that the proposed nonparametric parcellation framework achieves better segmentation than the existing state-of-the-art algorithms.

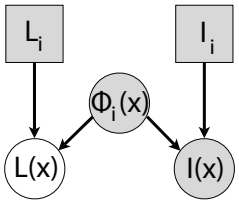


Fig. 1. Generative model for $(L(x), I(x))$ given $M(x) = i$ and (L_i, I_i, Φ_i). Φ_i is the mapping from the image coordinates to the template coordinates. Squares indicate non-random parameters, while circles indicate random variables. Shaded variables are assumed to be observed.

2 Theory

Let $\{I_i\}$ be N training images with corresponding label maps $\{L_i\}$, $i = 1, \ldots, N$. We assume the label maps take discrete values from 1 to $\mathcal{L}$ that indicate the label identity at each spatial location. We treat these training images as spatially continuous functions on $\mathbb{R}^3$ by assuming a suitable interpolator. Let $I : \Omega \mapsto \mathbb{R}$ denote a new, previously unseen test image defined on a discrete grid $\Omega \subset \mathbb{R}^3$. Let $\Phi_i : \Omega \mapsto \mathbb{R}^3$ denote the spatial mapping (warp) from the test image coordinates to the coordinates of the training image i. We assume that $\{\Phi_i\}$ have been precomputed using a pairwise registration procedure, such as the one described in Section 4.1.

Our objective is to estimate the label map $\hat{L}$ associated with the test image I. One common formulation to compute $\hat{L}$ is via MAP:

$$\hat{L} = \underset{L}{\operatorname{argmax}}\, p(L|I, \{L_i, I_i, \Phi_i\}) = \underset{L}{\operatorname{argmax}}\, p(L, I|\{L_i, I_i, \Phi_i\}), \tag{1}$$

where $p(L, I|\{L_i, I_i, \Phi_i\})$ denotes the joint probability of the label map L and image I given the training data.

Rather than using a parametric model for $p(L, I|\{L_i, I_i, \Phi_i\})$, we employ a non-parametric estimator, which is an explicit function of the entire training data, not a summary of it. Let $M : \Omega \mapsto \{1, \ldots N\}$ denote an unknown (hidden) random field that, for each voxel in test image I, specifies the training image I_i that generated that voxel. Given M, the training data, and warps, and assuming the factorization depicted in the graphical model of Fig. 1, we can construct the conditional probability of generating the test image and label map:

$$\begin{aligned}
&p\left(L, I|M, \{L_i, I_i, \Phi_i\}\right) \\
&= \prod_{x\in\Omega} p\left(L\left(x\right), I\left(x\right)|M\left(x\right), \{L_i, I_i, \Phi_i\}\right) &(2)\\
&= \prod_{x\in\Omega} p_{M(x)}\left(L\left(x\right), I\left(x\right)|L_{M(x)}, I_{M(x)}, \Phi_{M(x)}\left(x\right)\right) &(3)\\
&= \prod_{x\in\Omega} p_{M(x)}\Big(L\left(x\right)|L_{M(x)}, \Phi_{M(x)}\left(x\right)\Big) p_{M(x)}\Big(I\left(x\right)|I_{M(x)}, \Phi_{M(x)}\left(x\right)\Big), &(4)
\end{aligned}$$

where $p_{M(x)}(L(x), I(x)|L_{M(x)}, I_{M(x)}, \Phi_{M(x)}(x))$ is the conditional probability of $(L(x), I(x))$ given that voxel $x \in \Omega$ of the test subject was generated from training subject $M(x)$. Note that Eq. (4) assumes that $(L(x), I(x))$ are conditionally

independent given the membership $M(x)$, corresponding warp $\Phi_{M(x)}$ and training data. Given a prior on M, we can view $p(L, I|\{L_i, I_i, \Phi_i\})$ as a mixture:

$$p(L, I|\{L_i, I_i, \Phi_i\}) = \sum_{\mathcal{M}} p(M) p\left(L, I|M, \{L_i, I_i, \Phi_i\}\right), \tag{5}$$

where $\sum_{\mathcal{M}}$ denotes the marginalization over the unknown random field M. Substituting Eq. (4) into Eq. (5) yields:

$$\hat{L} = \underset{L}{\operatorname{argmax}} \sum_{\mathcal{M}} p\left(M\right) \prod_{x \in \Omega} p_{M(x)} \left(L\left(x\right) | L_{M(x)}, \Phi_{M(x)}\left(x\right)\right) \times p_{M(x)} \left(I\left(x\right) | I_{M(x)}, \Phi_{M(x)}\left(x\right)\right). \tag{6}$$

In the next section, we present instantiations of the individual terms in Eq. (6).

3 Model Instantiation

3.1 Image Likelihood

We adopt a Gaussian distribution with a stationary variance σ^2 as the image likelihood:

$$p_i(I(x)|I_i, \Phi_i(x)) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(I\left(x\right) - I_i\left(\Phi_i\left(x\right)\right)\right)^2\right]. \tag{7}$$

3.2 Label Likelihood

We use the distance transform representation to encode the label prior information, cf. [19]. Let D_i^l denote the signed distance transform of label l in training subject i, assumed to be positive inside the structure of interest. We define the label likelihood as:

$$p_i(L(x) = l|L_i, \Phi_i(x)) \propto \exp(\rho D_i^l(\Phi_i(x))), \tag{8}$$

where $\rho > 0$ is the slope constant and $\sum_{l=1}^{\mathcal{L}} p_i(L(x) = l|L_i, \Phi_i(x)) = 1$, where $\mathcal{L}$ is the total number of labels including a background label. $p_i(L(x) = l|L_i, \Phi_i(x))$ encodes the conditional probability of observing label l at voxel $x \in \Omega$ of the test image, given that it was generated from training image i.

3.3 Membership Prior

The latent random field $M : \Omega \mapsto \{1, \ldots, N\}$ encodes the local association between the test image and training data. We place a Markov Random Field (MRF) prior on M:

$$p(M) = \frac{1}{Z_\beta} \prod_{x \in \Omega} \exp\left(\beta \sum_{y \in \mathcal{N}_x} \delta(M(x), M(y))\right), \tag{9}$$

where $\beta \geq 0$ is a scalar parameter, $\mathcal{N}_x$ is a spatial neighborhood of voxel x, Z_β is the partition function that only depends on β, and $\delta(M(x), M(y)) = 1$, if $M(x) = M(y)$ and zero otherwise. In our implementation, $\mathcal{N}_x$ includes the immediate 8 neighbors of each voxel. Similar models have been used in the segmentation literature, e.g. [5, 9], mainly as priors on label maps to encourage spatially contiguous segmentations. In contrast, we use the MRF prior to pool local intensity information in determining the association between the test subject and training data.

The parameter β influences the average size of the local patches of the test subject that are generated from a particular training subject. In this work, we consider three settings of the parameter β. For $\beta = 0$, the model effectively assumes that each test image voxel is independently generated from a training subject, drawn with a uniform prior. $\beta \to +\infty$ forces the membership of all voxels to be the same and corresponds to assuming that the whole test subject is generated from a single unknown training subject, drawn from a uniform prior. A positive, finite β favors local patches of voxels to have the same membership.

The $\beta \to +\infty$ case reduces to a model similar to the one we presented in [16], except now we make the simplifying assumption that the training data is apriori mapped to the test subject's coordinate frame as a preprocessing step. Due to this simplification, the warp cost in registration plays no role in the segmentation algorithms we present in this paper. Without this simplification, however, inference for finite values of β becomes intractable. As demonstrated in the next section, the resulting inference algorithms allow us to determine the association between the training data and test subject using local intensity information.

4 Algorithms

4.1 Efficient Pairwise Registration

To perform pairwise registration, we employ an efficient algorithm [20, 21] that uses a one-parameter subgroup of diffeomorphisms, where a warp Φ is parameterized with a smooth, stationary velocity field $v : \mathbb{R}^3 \mapsto \mathbb{R}^3$ via an ODE [22]: $\frac{\partial \Phi(x,t)}{\partial t} = v(\Phi(x,t))$ and initial condition $\Phi(x,0) = x$. The warp $\Phi(x) = \exp(v)(x)$ can be computed efficiently using scaling and squaring and inverted by using the negative of the velocity field: $\Phi^{-1} = \exp(-v)$ [22].

We impose an elastic-like regularization on the stationary velocity field:

$$p(\Phi = \exp(v)) = \frac{1}{Z_\lambda} \exp\left[-\lambda \sum_{y \in \Omega} \sum_{j,k=1,2,3} \left(\frac{\partial^2}{\partial x_j^2} v_k(x)\Big|_{x=y}\right)^2\right], \qquad (10)$$

where $\lambda > 0$ is the warp stiffness parameter, Z_λ is a partition function that depends only on λ, and x_j and v_k denote the j'th and k'th component (dimension) of position x and velocity v, respectively. A higher warp stiffness parameter λ yields more rigid warps.

To derive the registration objective function, we assume a simple additive Gaussian noise model, consistent with the image likelihood term described in Section 3.1. This model leads to the following optimization problem for registering the i-th training image to the test subject:

$$\hat{v}^i = \underset{v}{\operatorname{argmin}} \sum_{y \in \Omega} \left[(I(y) - I_i(\exp(v)(y)))^2 + 2\lambda\sigma^2 \sum_{j,k=1,2,3} \left(\frac{\partial^2}{\partial x_j^2} v_k(x) \Big|_{x=y} \right)^2 \right], \quad (11)$$

where σ^2 is the stationary image noise variance, and $\Phi_i \triangleq \exp(\hat{v}^i)$. To solve Eq. (11), we use the bidirectional log-domain Demons framework [20], which decouples the optimization of the first and second terms by introducing an auxiliary transformation. The update warp is first computed using the Gauss-Newton method. The regularization is achieved by smoothing the updated warp field. It can be shown that the smoothing kernel corresponding to Eq. (10) can be approximated with a Gaussian; $K(x) \propto \exp(-\alpha \sum_{i=1,2,3} x_i^2)$, where $\alpha = \frac{\gamma}{8\lambda\sigma^2}$ and $\gamma > 0$ controls the size of the Gauss-Newton step.

4.2 Segmentation Algorithms

Here, we present algorithms to solve the optimization problem of Eq. (6) for the three cases of β in the model presented in Section 3.

4.3 Global Mixture

First, we consider $\beta \to +\infty$, which is equivalent to a global mixture model, where the test subject is assumed to be generated from a single, unknown training subject. In this case, the segmentation problem in Eq. (6) reduces to

$$\hat{L} = \underset{L}{\operatorname{argmax}} \sum_{i=1}^{N} \prod_{x \in \Omega} p_i(L(x)|L_i, \Phi_i(x)) p_i(I(x)|I_i, \Phi_i(x)). \quad (12)$$

Eq. (12) cannot be solved in closed form. However, an efficient solution to this MAP formulation can be obtained via Expectation Maximization (EM). Here, we present a summary.

The E-step updates the posterior of the membership associated with each training image:

$$m_i^{(n)} \propto \prod_{x \in \Omega} p_i(I(x)|I_i, \Phi_i(x)) p_i(\hat{L}^{(n-1)}(x)|L_i, \Phi_i(x)), \quad (13)$$

where $\hat{L}^{(n-1)}(x)$ is the segmentation estimate of the test image from the previous iteration and $\sum_i m_i^{(n)} = 1$. The M-step updates the segmentation estimate:

$$\hat{L}^{(n)}(x) = \underset{l \in \{1,\ldots,\mathcal{L}\}}{\operatorname{argmax}} \sum_{i=1}^{N} m_i^{(n)} \log \left(p_i(L(x) = l|L_i, \Phi_i(x)) \right). \quad (14)$$

The E-step in Eq. (13) determines a single membership index for the entire training image, based on all the voxels. The M-step in Eq. (14) performs an independent optimization at each voxel $x \in \Omega$; it determines the mode of a length-$\mathcal{L}$ vector, where $\mathcal{L}$ is the number of labels. The EM algorithm is initialized with $m_i^{(1)} \propto \prod_{x\in\Omega} p_i(I(x)|I_i, \Phi_i(x))$ and iterates between Equations (14) and (13), until convergence.

4.4 Local Mixture: Independent Prior

The second case we consider is $\beta = 0$, which corresponds to assuming a voxel-wise independent mixture model with a uniform prior on M:

$$p(M) = \frac{1}{N^{|\Omega|}}, \tag{15}$$

where $|\Omega|$ is the cardinality of the image domain, i.e., the number of voxels. It is easy to show that the segmentation problem reduces to

$$\hat{L}(x) = \underset{l\in\{1,\ldots,\mathcal{L}\}}{\operatorname{argmax}} \sum_{i=1}^{N} p_i(L(x) = l|L_i, \Phi_i(x)) p_i(I(x)|I_i, \Phi_i(x)), \tag{16}$$

where the image and label likelihood terms in the summation can be computed using Eqs. (7) and (8). The optimization problem can be solved by simply comparing $\mathcal{L}$ numbers at each voxel.

4.5 Semi-local mixture: MRF Prior

Finally, we consider a finite, positive β. This leads to an MRF prior, which couples neighboring voxels and thus the exact marginalization of Eq. (6) becomes computationally intractable. An efficient approximate solution can be obtained using variational mean field [23]. The main idea of variational mean field is to approximate the posterior distribution of the membership $p(M|I, L, \{I_i, L_i, \Phi_i\})$, with a simple distribution q that is fully factorized:

$$q(M) = \prod_{x\in\Omega} q_x(M(x)). \tag{17}$$

The objective function of Eq. (6) can then be approximated by an easier-to-optimize lower bound, which is a function of q. This approximate problem can be solved via coordinate-ascent, where the segmentation L and approximate posterior q are updated sequentially, by solving the optimization for each variable while fixing the other. One particular formulation leads to a straightforward update rule for L:

$$\hat{L}^{(n)}(x) = \underset{l\in\{1,\ldots\mathcal{L}\}}{\operatorname{argmax}} \sum_{i=1}^{N} q_x^{(n-1)}(M(x) = i) \log p_i(L(x) = l|L_i, \Phi_i(x)), \tag{18}$$

where $q^{(n-1)}$ is an estimate of the posterior at the $(n-1)$'th iteration. Eq. (18) is independent for each voxel and entails determining the mode of a length-$\mathcal{L}$ vector. For a fixed segmentation estimate $\hat{L}^{(n)}(x)$ the optimal q is the solution of the following fixed-point equation:

$$q_x^{(n)}(M(x)) \propto p_{M(x)}(I(x)|I_{M(x)}, \Phi_{M(x)}(x)) \times p_{M(x)}(\hat{L}^{(n)}(x)|L_{M(x)}, \Phi_{M(x)}(x)) \exp\left(\beta \sum_{y \in \mathcal{N}_x} q_y^{(n)}(M(x))\right), \quad (19)$$

and $\sum_i q_x^{(n)}(M(x)=i) = 1$. We solve Eq. (19) iteratively. The variational mean field algorithm alternates between Eqs. (19) and (18), until convergence.

5 Experiments

We validate the proposed framework on 39 T1-weighted brain MRI scans of dimensions $256 \times 256 \times 256$, 1mm isotropic. Each MRI volume is an average of 3-4 scans and was gain-field corrected and skull-stripped. These volumes were then manually delineated by an expert anatomist into left and right White Matter (WM), Cerebral Cortex (CT), Lateral Ventricle (LV), Hippocampus (HP), Thalamus (TH), Caudate (CA), Putamen (PU), Pallidum (PA) and Amygdala (AM). We use volume overlap with manual labels, as measured by the Dice score [24], to quantify segmentation quality. The Dice score ranges from 0 to 1, with higher values indicating improved segmentation.

5.1 Setting Parameters Through Training

The proposed nonparametric parcellation framework has two stages with several input parameters. The registration stage has two independent parameters: γ that controls the step size in the Gauss-Newton optimization and α that determines the smoothness of the final warp. The segmentation stage has two additional input parameters: σ^2, which is the intensity variance of the image likelihood in Eq. (7) and the slope ρ of the distance transform uses to compute the label prior in Eq. (8). Furthermore, the semi-local model of Section 4.5 has a non-zero, finite β parameter.

Nine subjects were used to determine the optimal values of these parameters. First, 20 random pairs of these nine subjects were registered for a range of values of γ and α. Registration quality was assessed by the amount of pairwise label overlap and used to select the optimal (γ^*, α^*) pair.

We used the optimal (γ^*, α^*) pair to register all 72 ordered pairs of the nine training subjects. We performed nine leave-one-out segmentations using both the global and local models to determine the corresponding optimal pairs of σ^2 and ρ. The optimal pair for the local model was then used to determine the optimal value for β in the semi-local model. The optimal parameter values were finally used to segment the remaining 30 subjects.

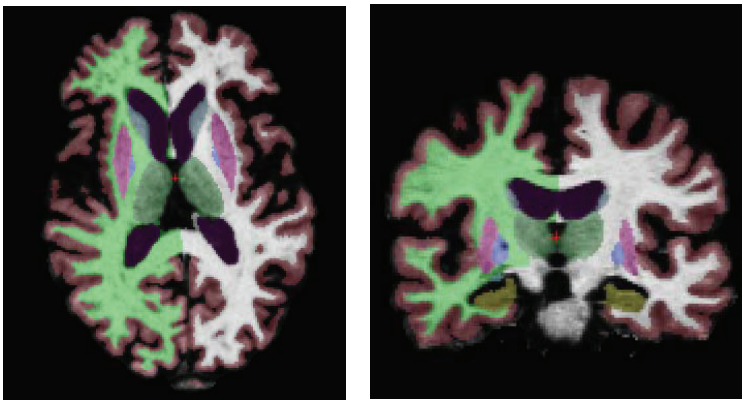


Fig. 2. A typical segmentation obtained with the local mixture model.

5.2 Benchmarks

The first benchmark we consider is the whole-brain parcellation tool available in the Freesurfer software package [25]. The Freesurfer parcellation tool uses a unified registration-segmentation procedure that models across-scanner intensity variation [2, 3]. We consider this as a state-of-the-art benchmark, since numerous imaging studies across multiple centers have shown Freesurfer's a robustness and accuracy as a segmentation tool.

As a second benchmark, we use our implementation of the Label Fusion algorithm [12, 14]. We employ the pairwise registrations obtained with (γ^*, α^*) to transfer the labels of the training subjects via the trilinear interpolation of the probability maps, obtained by assigning 1 to entries corresponding to the manual labels and zero elsewhere. Segmentation is then computed through majority voting at each voxel. We use trilinear interpolation instead of nearest neighbor interpolation because we find that trilinear interpolation yields better results.

5.3 Results

We report test results for the 30 subjects not included in the group used for setting the algorithm parameters $\gamma, \alpha, \sigma^2, \rho$, and β. For each test subject, we treated the remaining subjects as training data in a cross-validation evaluation.

Fig. 2 illustrates a typical automatic segmentation result obtained with the local mixture model and overlaid on the MRI volume. Fig. 3 shows box-plots of Dice scores for the two benchmarks and the proposed non-parametric parcellation algorithms. Table 1 provides the mean Dice scores averaged over all subjects and both hemispheres. Fig. 4 provides an overall comparison between the average dice scores achieved by the algorithms.

On average, the local and semi-local mixture models yield better segmentations than the global mixture model, mainly due to the large improvement in the white matter, cerebral cortex and lateral ventricles, the segmentation of which

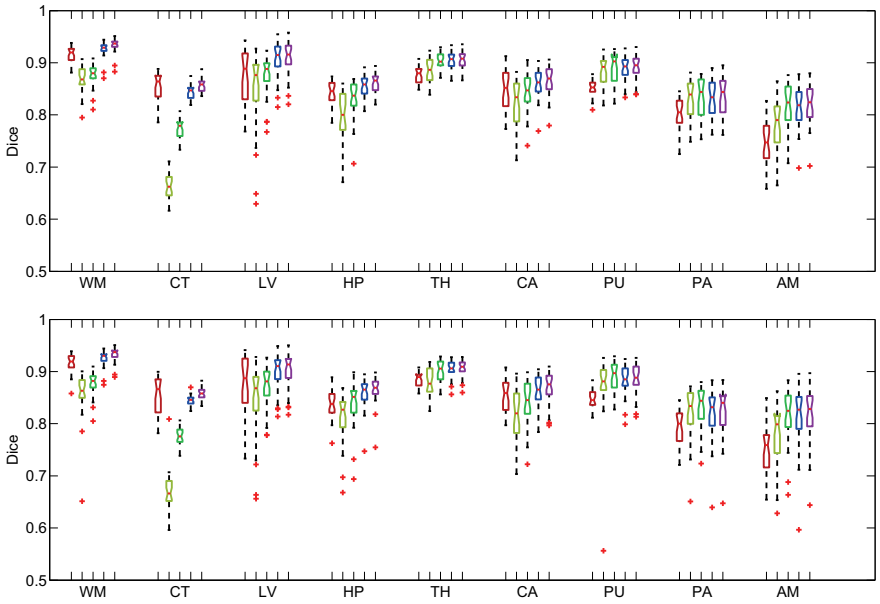


Fig. 3. Boxplots of Dice scores for Freesurfer (red), Label Fusion (yellow), the global mixture model (green), the local mixture model (blue) and the semi-local mixture model (purple). Top row is left hemisphere. Bottom row is right hemisphere. Medians are indicated by horizontal bars. Boxes indicate the lower and upper quartiles and vertical lines extend to 1.5 inter-quartile spacing. '+'s indicate outliers.

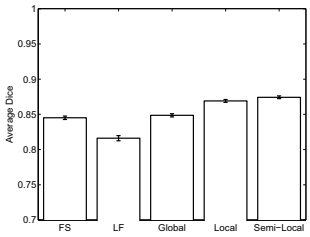


Fig. 4. Average Dice scores for each algorithm (FS: Freesurfer, LF: Label Fusion, Global: Global Mixture, Local: Local Mixture, and Semi-Local: MRF-based model). Error bars show standard error. Each subject and ROI was treated as an independent sample with an equal weight.

clearly benefits from the additional use of local intensity information. A paired t-test between the local and semi-local models reveals that a statistically significant improvement is achieved with the MRF model that pools local intensity information. Yet, this improvement is overall quite modest: about 1% per ROI.

As discussed earlier, the global mixture model is similar to that of [16], except that [16] incorporates registration into the model. Despite this, we find that both algorithms achieve similar segmentation accuracy (results not shown).

A paired sample t-test implies that the difference in accuracy between the proposed semi-local mixture model and Freesurfer is statistically significant ($p < 0.05$, Bonferroni corrected) for all ROIs, except the cerebral cortex and right

Table 1. Comparison of average dice scores. Boldface font indicates best scores for each structure. As a reference, the last row lists approximate average volumes.

	WM	CT	LV	HP	TH	CA	PU	PA	AM
Freesurfer	0.92	0.85	0.87	0.84	0.88	0.85	0.85	0.80	0.75
Label Fusion	0.85	0.66	0.84	0.77	0.86	0.80	0.86	0.81	0.75
Global Mixture	0.88	0.77	0.87	0.83	0.90	0.84	**0.89**	**0.83**	0.81
Local Mixture	**0.93**	0.84	0.90	0.86	0.90	0.86	0.88	0.82	**0.82**
Semi-local Mixture	**0.93**	**0.86**	**0.91**	**0.87**	**0.91**	**0.87**	**0.89**	**0.83**	**0.82**
Volumes ($\times 10^3 mm^3$)	450	448	25	7	14	7	10	3	3

Table 2. Approximate average run-time to segment one test subject (in CPU hours).

Freesurfer	Label Fusion	Global Mixture	Local Mixture	Semi-local Mixture
10	24	32	24	40

Caudate, where the two methods yield comparable results. The same results are obtained when comparing the local mixture model and Freesurfer.

Compared to the Label Fusion benchmark, the nonparametric parcellation algorithms (global, local and semi-local) yield significantly better segmentation (paired sample t-test, $p < 0.05$, Bonferroni corrected) in all regions, except Pallidum and Putamen, where the improvement over Label Fusion does not reach statistical significance. We note, however, that the results we report for our Label Fusion implementation are lower than the ones reported in [12]. This might be due to differences in the data and/or registration algorithm. Specifically, normalized mutual information (NMI) was used as the registration cost function in [12]. Entropy-based measures such as NMI are known to yield more robust alignment results. We leave a careful analysis of this issue to future work.

Table 2 lists the average run-times for all five algorithms. The parametric atlas-based Freesurfer algorithm is the fastest, mainly because it needs to compute only a single registration. The remaining algorithms take up more than 20 hours of CPU time on a modern machine, most of which is dedicated to the many registrations performed with the training data. The two iterative algorithms that solve the global mixture and semi-local mixture models (EM and variational mean field, respectively) require significantly longer run-times. The local-mixture model, on the other hand, requires minimal computation time once the registrations are complete, since it simply performs a voxelwise weighted averaging. Its run-time is similar to that required by Label Fusion.

6 Conclusion

This paper presents a novel, nonparametric mixture model of images and label maps, that yields accurate image segmentation algorithms. The resulting algorithms are conceptually similar to recent label fusion (or multi-atlas) methods that utilize the entire training data, rather than a summary of it, and register the test subject to each training subject separately. Segmentation is then achieved

by fusing the transferred manual labels. In the proposed framework, similarities between the test image and training data determine how the transferred labels are weighed during fusion. As we discuss in this paper, different settings of a model parameter yields various weighting strategies. Our experiments suggests that a semi-local strategy that is derived from an MRF model that encourages local image patches to be associated with the same training data provides the best segmentation results. We also show that a computationally less expensive local strategy that treats each voxel independently leads to accurate segmentations that are better than the current state-of-the-art.

We leave an investigation of various registration algorithms within the proposed framework to future work. It is clear that alternative strategies can be used to improve the alignment between the training data and test subject. For example, one could use a richer representation of diffeomorphic warps, cf. [4], or a more sophisticated registration cost function, cf. [12]. Since any multi-atlas segmentation algorithm will be robust against occasional registration failures, whether a better alignment algorithm will lead to more accurate segmentation remains an open question.

Acknowledgements: Support for this research is provided in part by: NAMIC (NIH NIBIB NAMIC U54-EB005149), the NAC (NIH NCRR NAC P41-RR13218), the mBIRN (NIH NCRR mBIRN U24-RR021382), the NIH NINDS R01-NS051826 grant, the NSF CAREER 0642971 grant, NCRR (P41-RR14075, R01 RR16594-01A1), the NIBIB (R01 EB001550, R01EB006758), the NINDS (R01 NS052585-01), the MIND Institute, and the Autism & Dyslexia Project funded by the Ellison Medical Foundation. B.T. Thomas Yeo is funded by the A*STAR, Singapore.

References

1. Ashburner, J., Friston, K.: Unified segmentation. Neuroimage **26** (2005) 839–851
2. Fischl, B., et al.: Sequence-independent segmentation of magnetic resonance images. Neuroimage **23** (2004) 69–84
3. Han, X., Fischl, B.: Atlas renormalization for improved brain MR image segmentation across scanner platforms. IEEE TMI **26**(4) (2007) 479–486
4. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphism atlas construction for computational anatomy. Neuroimage **23** (2004) 151–160
5. Van Leemput, K., et al.: Automated model-based bias field correction of MR images of the brain. **18**(10) (1999) 885–896
6. Pohl, K., et al.: A bayesian model for joint segmentation and registration. Neuroimage **31** (2006) 228–239
7. Twining, C., et al.: A unified information-theoretic approach to groupwise non-rigid registration and model building. In Proc. IPMI **LNCS 3565** (2005) 1–14
8. Tu, Z., et al.: Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE TMI **27**(4) (2008) 495–508
9. Yeo, B.T.T., et al.: Effects of registration regularization and atlas sharpness on segmentation accuracy. Medical Image Analysis **12**(5) (2008) 603–615
10. Aljabar, P., et al.: Classifier selection strategies for label fusion using large atlas databases. In Proc. of MICCAI **LNCS 4791** (2007) 523–531

11. Artaechevarria, X., et al.: Combination strategies in multi-atlas image segmentation: Application to brain MR data. IEEE TMI **28**(8) (2009) 1266 – 1277
12. Heckemann, R., et al.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage **33**(1) (2006) 115–126
13. Isgum, I., et al.: Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. IEEE TMI **28**(7) (2009)
14. Rohlfing, T., et al.: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. NeuroImage **21**(4) (2004) 1428–1442
15. Rohlfing, T., et al.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE TMI **23**(8) (2004) 983–994
16. Sabuncu, M., et al.: Supervised nonparameteric image parcellation. *In Proc.* of MICCAI **LNCS 5762** (2009)
17. Davis, B., et al.: Population shape regression from random design data. In Proc. of ICCV (2007) 1–7
18. Warfield, S., et al.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for validation of image segmentation. TMI **23**(7) (2004) 903–921
19. Pohl, K., et al.: Logarithm odds maps for shape representation. MICCAI 2006 **LNCS 4191** (2006) 955–963
20. Vercauteren, T., et al.: Symmetric log-domain diffeomorphic registration: A demons-based approach. In Proc of MICCAI **LNCS 5241** (2008) 754–761
21. Sabuncu, M., et al.: Asymmetric image template registration. *In Proc.* of MICCAI **LNCS 5761** (2009)
22. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-Euclidean framework for statistics on diffeomorphisms. In Proc of MICCAI **LNCS 4190** (2006) 924–931
23. Jaakkola, T.: Tutorial on variational approximation methods. Advanced Mean Field Methods: Theory and Practice (2000)
24. Dice, L.: Measures of the amount of ecologic association between species. Ecology **26**(3) (1945) 297–302
25. Freesurfer: http://surfer.nmr.mgh.harvard.edu