Bjoern H. Menze Georg Langs
Le Lu Albert Montillo Zhuowen Tu
Antonio Criminisi (Eds.)

Medical Computer Vision

Recognition Techniques and Applications in Medical Imaging

LNCS 7766

Second International MICCAI Workshop, MCV 2012
Nice, France, October 2012
Revised Selected Papers

Springer

Lecture Notes in Computer Science 7766

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Bjoern H. Menze Georg Langs
Le Lu Albert Montillo Zhuowen Tu
Antonio Criminisi (Eds.)

Medical Computer Vision

Recognition Techniques and Applications in Medical Imaging

Second International MICCAI Workshop, MCV 2012
Nice, France, October 5, 2012
Revised Selected Papers

Springer

Volume Editors

Bjoern H. Menze
ETH Zurich, Sternwartstrasse 7, 8092 Zürich, Switzerland
E-mail: bjoern@ethz.ch

Georg Langs
Medical University of Vienna, Währinger Gürtel 18-20, 1090 Wien, Austria
E-mail: georg.langs@meduniwien.ac.at

Le Lu
Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA
E-mail: le-lu@siemens.com

Albert Montillo
GE Global Research, 1 Research Circle, Niskayuna, NY 12309, USA
E-mail: montillo@ge.com

Zhuowen Tu
University of California, 635 Charles E. Young Drive South
Los Angeles, CA 90095-7334, USA
E-mail: zhuowen.tu@loni.ucla.edu

Antonio Criminisi
Microsoft Research, 7 JJ Thomson Avenue, Cambridge, CB3 0FB, UK
E-mail: antcrim@microsoft.com

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-36619-2 e-ISBN 978-3-642-36620-8
DOI 10.1007/978-3-642-36620-8
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013931266

CR Subject Classification (1998): I.4.6-7, I.4.9, I.4.3, I.2.10, I.5.2-4, J.3

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

Preface

The Second MICCAI Workshop on Medical Computer Vision (MICCAI-MCV 2012) was held in conjunction with the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) on October 5, 2012 in Nice, France. It succeeded the First Workshop on Medical Computer Vision that was held in September 2010 in conjunction with MICCAI 2010 in Beijing.

The workshop aimed at exploring the use of modern computer vision technology in tasks such as automatic segmentation and registration, localization of anatomical features and detection of anomalies, as well as 3D reconstruction and biophysical model personalization. In this it focuses on principled approaches that go beyond the limits of current model-driven image analysis, which are provably efficient and scalable, and which generalize well to previously unseen images.

The goal of the workshop was to foster discussions among researchers working on novel computational approaches at the interface of computer vision, machine learning, and medical image analysis, and who are interested in pushing the boundaries of what current medical software applications can deliver in both clinical and medical research settings. To this end we invited Nikos Paragios from INRIA and Ecole Centrale Paris and Nassir Navab from TU Munich to discuss challenges and opportunities lying at the interface of medical computer vision and "classic" computer vision. The following panel discussion with the invited speakers – in which Nicholas Ayache, INRIA Sophia-Antipolis, and Simon Mercer, Microsoft Research, joined in – dealt with the following questions:

- How do we turn research into clinical use? How do we turn research into products?
- How do we make data available to the broader research community?
- What makes medical imaging data special compared to classic computer vision data and problems?
- How would we set up large data sets for training efficient computer vision like algorithms? And is this a good idea at all?
- How do we solve the annotation problem? What are perspectives in times of mechanical turk? What are effective incentives for clinical collaborators to share knowledge and to annotate data image?

Central to the workshop were the contributions of the participants. Our call for papers resulted in 42 submissions of up to 12 pages. Each paper received at least three reviews. Based on these peer reviews, we selected 24 submissions for presentation out of which 12 were presented as a poster and 12 as a poster together with a plenary talk. Three talks were awarded the "MCV Best Paper Award" based on the popular vote of the workshop attendees: Herve Lombaert

et al. "Groupwise Spectral Log-Demons Framework for Atlas Construction," Tobias Gass et al. "Semi-supervised Segmentation Using Multiple Segmentation Hypotheses from a Single Atlas," and Rene Donner et al. "Fast Anatomical Structure Localization Using Top-down Image Patch Regression."

The present volume contains the reworked papers of the MICCAI-MCV 2012 workshop. It also features four selected papers by Zhong et al., Li et al., Song et al., and Wu et al. that were presented at the previous CVPR Medical Computer Vision Workshop, which was co-organized by L. Lu, B. Menze, G. Langs, Y. Zhan, and Z. Tu and was held in conjunction with the International Conference on Computer Vision and Pattern Recognition on June 21, 2012, in Providence, Rhode Island, USA.

December 2012

Bjoern H. Menze
Georg Langs
Le Lu
Albert Montillo
Zhuowen Tu
Antonio Criminisi

Organization

Workshop Chairs

Bjoern Menze	ETHZurich, INRIA, Switzerland/France
Georg Langs	MU Vienna, MIT, Austria/USA
Albert Montillo	GE, USA
Zhuowen Tu	UCLA, USA
Antonio Criminisi	Microsoft Research, UK

Invited Speakers

Nassir Navab	TU Munich, Germany
Nikos Paragios	Ecole Centrale Paris, France

Program Committee

Alison Noble	Oxford, UK
Ben Glocker	Microsoft Research, UK
Cagatay Demiralp	Brown University, USA
Christian Barillot	IRISA Rennes, France
Christos Davatzikos	University of Pennsylvania, USA
Daniel Rueckert	Imperial College London, UK
Darko Zikic	Microsoft Research, UK
Ender Konukoglu	Microsoft Research, UK
Hayit Greenspan	Tel Aviv University, Israel
Helmut Grabner	ETH Zurich, Switzerland
Horst Bischof	TU Graz, Austria
Jan Margeta	INRIA, France
Juan Eugenio Iglesias	Harvard MGH, USA
Juergen Gall	Max-Planck Gesellschaft Tübingen, Germany
Kayhan Batmanghelich	MIT, USA
Kilian Pohl	University of Pennsylvania, USA
Koen Van Leemput	Harvard MGH, DTU, USA
Leo Grady	Siemens Corporate Research, USA
Lin Yang	University of Kentucky, USA
Marleen de Bruijne	EMC Rotterdam, University of Copenhagen, The Netherlands/Denmark
Matthew Blaschko	Ecole Centrale Paris, France

Michael Kelm	Siemens Corporate Research, Germany
Michael Wels	Siemens Corporate Research, Germany
Milan Sonka	University of Iowa, USA
Paul Suetens	KU Leuven, Belgium
Rachid Deriche	INRIA, France
Ron Kikinis	Harvard BWH, USA
Sebastian Ourselin	University College London, UK
Tammy Riklin Raviv	Harvard BWH, USA
Tom Vercauteren	Mauna Kea Technology, France
Victor Lempitsky	Yandex, Russia
Yefeng Zheng	Siemens Corporate Research, USA

Table of Contents

Registration

Segmentation

Detection, Localization, Tracking

3D Reconstruction

Biophysical Model Personalization

Real-Time 2D/3D Deformable Registration Using Metric Learning

Chen-Rui Chou[1] and Stephen Pizer[1,2]

[1] Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
cchou@cs.unc.edu
[2] Department of Radiation Oncology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Abstract. We present a novel 2D/3D deformable registration method, called Registration Efficiency and Accuracy through Learning Metric on Shape (*REALMS*), that can support real-time Image-Guided Radiation Therapy (*IGRT*). The method consists of two stages: planning-time learning and registration. In the planning-time learning, it firstly models the patient's 3D deformation space from the patient's time-varying 3D planning images using a low-dimensional parametrization. Secondly, it samples deformation parameters within the deformation space and generates corresponding simulated projection images from the deformed 3D image. Finally, it learns a Riemannian metric in the projection space for each deformation parameter. The learned distance metric forms a Gaussian kernel of a kernel regression that minimizes the leave-one-out regression residual of the corresponding deformation parameter. In the registration, REALMS interpolates the patient's 3D deformation parameters using the kernel regression with the learned distance metrics. Our test results showed that REALMS can localize the tumor in 10.89 ms (91.82 fps) with 2.56 ± 1.11 mm errors using a single projection image. These promising results show REALMS's high potential to support real-time, accurate, and low-dose IGRT.

1 Introduction

Tumor localization in 3D is the main goal of Image-guided Radiation Therapy (*IGRT*). It is usually accomplished by computing the patient's treatment-time 3D deformations based on an on-board imaging system, usually x-ray. The treatment-time 3D deformations can be computed by doing image registration between the treatment-time reconstructed 3D image and the treatment-planning 3D image (3D/3D registration) or between the treatment-time on-board projection images and the treatment-planning 3D image (2D/3D registration). Recent advances of the IGRT registration methods emphasize real-time computation and low-dose image acquisition. Russakoff et al. [1,2], Khamene et al. [3], Munbodh et al. [4], Li et al. [5,6] rejected the time-consuming 3D/3D registration and performed 2D/3D registration by optimizing similarity functions defined in

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 1–10, 2013.

the projection domain. Other than the optimization-based methods, Chou et al. [7,8] recently introduced a faster and low-dose 2D/3D image registration by using a linear operator that approximates the deformation parameters. However, all of the above registration methods involve computationally demanding production of Digitally-Reconstructed Radiographs (*DRRs*) in each registration iteration (e.g., 15ms on a modern GPU to produce a 256×256 DRR from a $256 \times 256 \times 256$ volume [9]), which makes them difficult to be extended to support real-time (> 30 fps) image registration.

We present a novel real-time 2D/3D registration method, called Registration Efficiency and Accuracy through Learning Metric on Shape (*REALMS*), that does not require DRR production in the registration. It calculates the patient's treatment-time 3D deformations by kernel regression. Specifically, each of the patient's deformation parameters is interpolated using a weighting Gaussian kernel on that parameter's training case values. In each training case, its parameter value is associated with a corresponding training projection image. The Gaussian kernel is formed from distances between training projection images. This distance for the parameter in question involves a Riemannian metric on projection image differences. At planning time, REALMS learns the parameter-specific metrics from the set of training projection images using a Leave-One-Out (*LOO*) training.

To the best of our knowledge, REALMS is the first 2D/3D deformable registration method that achieves real-time (> 30 fps) performance. REALMS uses the metric learning idea firstly introduced in Weinberger and Tesauro [10] to tackle the 2D/3D image registration problem. Particularly, in order to make the metric learning work for the high dimensional ($D \gg 10^3$) projection space, REALMS uses a specially-designed initialization approximated by linear regression. The results have led to substantial error reduction when the special initialization is applied.

The rest of the paper is organized as follows: In section 2, we describe REALMS's novel registration scheme that uses kernel regression. In section 3, we describe its deformation space modeling approach for generating training samples in the deformation space. In section 4, we describe the metric learning scheme and the specialized initialization in REALMS. We show our synthetic and real results in section 5. Finally, we discuss the results and conclude in section 6.

2 2D/3D Registration Framework

In this section, we describe REALMS's 2D/3D registration framework. REALMS uses kernel regression (eq. 1) to interpolate the patient's n 3D deformation parameters $\mathbf{c} = (c^1, c^2, \cdots, c^n)$ separately from the on-board projection image $\mathbf{\Psi}(\theta)$ where θ is the projection angle. It uses a Gaussian kernel $K_{\mathbf{M}^i,\sigma^i}$ with the width σ^i and a metric tensor $\mathbf{M}^i$ on projection intensity differences to interpolate the patient's i^{th} deformation parameter c^i from a set of N training projection images $\{\mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta) \mid \kappa = 1, 2, \cdots, N\}$ simulated at planning time. Specifically, the training projection image, $\mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta)$, is the DRR of a 3D image deformed

from the patient's planning-time 3D mean image I with sampled deformation parameters $\mathbf{c}_\kappa = (c_\kappa^1, c_\kappa^2, \cdots, c_\kappa^n)$. T and $\mathbf{P}$ are the warping and the DRR operators, respectively. $\mathbf{P}$ simulates the DRRs according to the treatment-time imaging geometry, e.g., the projection angle θ.

In the treatment-time registration, each deformation parameter c^i in $\mathbf{c}$ can be estimated with the following kernel regression:

$$c^i = \frac{\sum_{\kappa=1}^{N} c_\kappa^i \cdot K_{\mathbf{M}^i,\sigma^i}(\mathbf{\Psi}(\theta), \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta))}{\sum_{\kappa=1}^{N} K_{\mathbf{M}^i,\sigma^i}(\mathbf{\Psi}(\theta), \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta))}, \tag{1}$$

$$K_{\mathbf{M}^i,\sigma^i}(\mathbf{\Psi}(\theta), \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta)) = \frac{1}{\sqrt{2\pi}\sigma^i} e^{-\frac{d^2_{\mathbf{M}^i}(\mathbf{\Psi}(\theta), \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta))}{2(\sigma^i)^2}}, \tag{2}$$

$$d^2_{\mathbf{M}^i}(\mathbf{\Psi}(\theta), \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta)) = (\mathbf{\Psi}(\theta) - \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta))^\intercal \mathbf{M}^i (\mathbf{\Psi}(\theta) - \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta)), \tag{3}$$

where $K_{\mathbf{M}^i,\sigma^i}$ is a Gaussian kernel (kernel width$= \sigma^i$) that uses a Riemannian metric $\mathbf{M}^i$ in the squared distance $d^2_{\mathbf{M}^i}$ and gives the weights for the parameter interpolation in the regression. The minus signs in eq. 3 denote pixel-by-pixel intensity subtraction.

We describe in section 3 how REALMS, at planning time, parameterizes the deformation space and describe in section 4 how it learns the metric tensor $\mathbf{M}^i$ and decides the kernel width σ^i.

3 Deformation Modeling at Planning Time

REALMS limits the deformation to a shape space. It models deformations as a linear combination of a set of basis deformations calculated through PCA analysis. In our target problem – lung IGRT, a set of Respiratory-Correlated CTs (*RCCTs*, dimension: $512 \times 512 \times 120$) $\{J_\tau \mid \tau = 1, 2, \cdots, 10\}$ are available at planning time. From these a mean image $I = \overline{J}$ and a set of deformations ϕ_τ between J_τ and $\overline{J}$ can be computed. The basis deformations can then be chosen to be the primary eigenmodes of a PCA analysis on the ϕ_τ.

3.1 Deformation Shape Space and Mean Image Generation

REALMS computes a respiratory Fréchet mean image $\overline{J}$ from the RCCT dataset via an *LDDMM* (Large Deformation Diffeomorphic Metric Mapping) framework described in Lorenzen et al. [11]. The Fréchet mean $\overline{J}$, as well as the diffeomorphic deformations ϕ from the mean $\overline{J}$ to each image J_τ, are computed using a fluid-flow distance metric:

$$\overline{J} = \underset{J}{argmin} \sum_{\tau=1}^{10} \int_0^1 \int_\Omega ||v_{\tau,\gamma}(x)||^2 dx d\gamma + \frac{1}{s^2} \int_\Omega ||J(\phi_\tau^{-1}(x)) - J_\tau(x)||^2 dx, \tag{4}$$

where $J_\tau(x)$ is the intensity of the pixel at position x in the image J_τ, $v_{\tau,\gamma}$ is the fluid-flow velocity field for the image J_τ in flow time γ , s is the weighting variable on the image dissimilarity, and $\phi_\tau(x)$ describes the deformation at the pixel location x: $\phi_\tau(x) = x + \int_0^1 v_{\tau,\gamma}(x)d\gamma$.

3.2 Statistical Analysis

With the diffeomorphic deformation set $\{\phi_\tau \mid \tau = 1, 2, \cdots, 10\}$ calculated, our method finds a set of linear deformation basis vectors ϕ^i_{pc} by PCA analysis. The scores λ^i_τ on each ϕ^i_{pc} yield ϕ_τ in terms of these basis vectors.

$$\phi_\tau = \overline{\phi} + \sum_{i=1}^{10} \lambda^i_\tau \cdot \phi^i_{pc} \tag{5}$$

We choose a subset of n eigenmodes that captures more than 95% of the total variation. Then we let the n scores form the the n-dimensional parametrization $\mathbf{c}$.

$$\mathbf{c} = (c^1, c^2, \cdots, c^n) = (\lambda^1, \lambda^2, \cdots, \lambda^n) \tag{6}$$

For most of our target problems, $n = 3$ satisfies the requirement.

4 Metric Learning at Planning Time

4.1 Metric Learning and Kernel Width Selection

REALMS learns a metric tensor $\mathbf{M}^i$ with a corresponding kernel width σ^i for the patient's i^{th} deformation parameter c^i using a Leave-One-Out (LOO) training strategy. At planning time, it samples a set of N deformation parameter tuples $\left\{\mathbf{c}_\kappa = (c^1_\kappa, c^2_\kappa, \cdots, c^n_\kappa) \mid \kappa = 1, 2, \cdots N\right\}$ to generate training projection images $\{\mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta) \mid \kappa = 1, 2, \cdots, N\}$ where their associated deformation parameters are sampled uniformly within three standard deviations of the scores λ observed in the RCCTs. For each deformation parameter c^i in $\mathbf{c}$, REALMS finds the best pair of the metric tensor $\mathbf{M}^{i\dagger}$ and the kernel width $\sigma^{i\dagger}$ that minimizes the sum of squared LOO regression residuals $\mathcal{L}_{c^i}$ among the set of N training projection images:

$$\mathbf{M}^{i\dagger}, \sigma^{i\dagger} = \underset{\mathbf{M}^i, \sigma^i}{arg\ min} \mathcal{L}_{c^i}(\mathbf{M}^i, \sigma^i), \tag{7}$$

$$\mathcal{L}_{c^i}(\mathbf{M}^i, \sigma^i) = \sum_{\kappa=1}^{N} \left(c^i_\kappa - \hat{c^i_\kappa}(\mathbf{M}^i, \sigma^i)\right)^2, \tag{8}$$

$$\hat{c^i_\kappa}(\mathbf{M}^i, \sigma^i) = \frac{\sum\limits_{\chi \neq \kappa} c^i_\chi \cdot K_{\mathbf{M}^i, \sigma^i}(\mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta), \mathbf{P}(I \circ T(\mathbf{c}_\chi); \theta))}{\sum\limits_{\chi \neq \kappa} K_{\mathbf{M}^i, \sigma^i}(\mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta), \mathbf{P}(I \circ T(\mathbf{c}_\chi); \theta))}, \tag{9}$$

where $\hat{c^i_\kappa}(\mathbf{M}^i, \sigma^i)$ is the estimated value for parameter c^i_κ interpolated by the metric tensor $\mathbf{M}^i$ and the kernel width σ^i from the training projection images χ other than κ; $\mathbf{M}^i$ needs to be a positive semi-definite ($p.s.d$) matrix to fulfill the pseudo-metric constraint; and the kernel width σ^i needs to be a positive real number.

To avoid high-dimensional optimization over the constrained matrix $\mathbf{M}^i$, we structure the metric tensor $\mathbf{M}^i$ as a rank-1 matrix formed by a basis vector $\mathbf{a}^i$: $\mathbf{M}^i = \mathbf{a}^i\mathbf{a}^{i\intercal}$. Therefore, we can transform eq. 7 into a optimization over the unit vector $\mathbf{a}^i$ where $\left\|\mathbf{a}^i\right\|_2 = 1$:

$$\mathbf{a}^{i\dagger}, \sigma^{i\dagger} = \underset{\mathbf{a}^i, \sigma^i}{arg\,min} \mathcal{L}_{c^i}(\mathbf{a}^i\mathbf{a}^{i\intercal}, \sigma^i) \tag{10}$$

Then we can rewrite the squared distance $d^2_{\mathbf{M}^i} = d^2_{\mathbf{a}^i\mathbf{a}^{i\intercal}}$ used in the Gaussian kernel $K_{\mathbf{M}^i,\sigma^i}$ as follows:

$$d^2_{\mathbf{a}^i\mathbf{a}^{i\intercal}}(\mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta), \mathbf{P}(I \circ T(\mathbf{c}_\chi); \theta)) = (\mathbf{a}^{i\intercal} \cdot \mathbf{r}_{\kappa,\chi})^\intercal(\mathbf{a}^{i\intercal} \cdot \mathbf{r}_{\kappa,\chi}), \tag{11}$$

$$\mathbf{r}_{\kappa,\chi} = \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta) - \mathbf{P}(I \circ T(\mathbf{c}_\chi); \theta), \tag{12}$$

where $\mathbf{r}_{\kappa,\chi}$ is a vector of intensity differences between projection images generated by parameters $\mathbf{c}_\kappa$ and $\mathbf{c}_\chi$; and $\mathbf{a}^i$ is a metric basis vector where the magnitude of the inner product of $\mathbf{a}^i$ and the intensity difference vector $\mathbf{r}_{\kappa,\chi}$, $\mathbf{a}^{i\intercal} \cdot \mathbf{r}_{\kappa,\chi}$ gives the Riemannian distance for the parameter c^i (eq. 11).

The learned metric basis vector $\mathbf{a}^{i\dagger}$ and the selected kernel width $\sigma^{i\dagger}$ form a weighting kernel $K_{\mathbf{a}^{i\dagger}\mathbf{a}^{i\dagger\intercal},\sigma^{i\dagger}}$ to interpolate the parameter c^i in the registration (see eq. 1).

4.2 Linear-Regression Implied Initial Metric

Since the residual functional $\mathcal{L}$ (see eq. 7) that we want to minimize is non-convex, a good initial guess of the metric basis vector $\mathbf{a}$ is essential. Therefore, REALMS uses a vector $\mathbf{w}^i$ as an initial guess of the metric basis vector $\mathbf{a}^i$ for the parameter c^i. Let $\mathbf{W} = \begin{pmatrix}\mathbf{w}^1 & \mathbf{w}^2 & \cdots & \mathbf{w}^n\end{pmatrix}$ list these initial guesses. The matrix $\mathbf{W}$ is approximated by a multivariate linear regression (eq. 13 and eq. 14) between the projection difference matrix $\mathbf{R} = (\mathbf{r}_1\mathbf{r}_2 \cdots \mathbf{r}_N)^\intercal$ and the parameter differences matrix $\Delta\mathbf{C}$. In particular, the projection difference vector $\mathbf{r}_\kappa = \mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta) - \mathbf{P}(I; \theta)$ is the intensity differences between the DRRs calculated from the deformed image $I \circ T(\mathbf{c}_\kappa)$ and the DRRs calculated from the mean image I (where $\mathbf{c} = \mathbf{0}$).

$$\Delta\mathbf{C} = \begin{pmatrix} c^1_1 & c^2_1 & \cdots & c^n_1 \\ c^1_2 & c^2_2 & \cdots & c^n_2 \\ \vdots & \vdots & \ddots & \vdots \\ c^1_N & c^2_N & \cdots & c^n_N \end{pmatrix} - \mathbf{0} \approx \begin{pmatrix} \mathbf{r}^\intercal_1 \\ \mathbf{r}^\intercal_2 \\ \vdots \\ \mathbf{r}^\intercal_N \end{pmatrix} \cdot \begin{pmatrix}\mathbf{w}^1 & \mathbf{w}^2 & \cdots & \mathbf{w}^n\end{pmatrix} \tag{13}$$

$$\mathbf{W} = (\mathbf{R}^\intercal \mathbf{R})^{-1} \mathbf{R}^\intercal \Delta \mathbf{C} \tag{14}$$

The inner product of the matrix $\mathbf{W}$, calculated by the pseudo-inverse in eq. 14, and the projection intensity difference matrix $\mathbf{R}$, $\mathbf{W}^\intercal \mathbf{R}$, gives the best linear approximation of the parameter differences $\Delta \mathbf{C}$. Therefore, we use $\mathbf{w}^i$ as the initial guess of the metric basis vector $\mathbf{a}^i$ for the parameter c^i.

4.3 Optimization Scheme

REALMS uses a two-step scheme to optimize the metric basis vector $\mathbf{a}^i$ and the kernel width σ^i in eq. 10.

First, for each candidate kernel width σ^i, it optimizes the metric basis vector $\mathbf{a}^i$ using the quasi-Newton method (specifically, the BFGS method) with the vector $\mathbf{w}^i$ as the initialization. The gradient of the function $\mathcal{L}_{c^i}$ with respect to $\mathbf{a}^i$ can be stated as

$$\frac{\partial \mathcal{L}_{c^i}}{\partial \mathbf{a}^i} = \frac{2\sqrt{2}}{\sigma^i} \mathbf{a}^i \sum_{\kappa=1}^{N} (\hat{c_\kappa^i} - c_\kappa^i) \sum_{\chi=1}^{N} (\hat{c_\chi^i} - c_\chi^i) K_{\mathbf{a}^i \mathbf{a}^{i\intercal}, \sigma^i} (\mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta), \mathbf{P}(I \circ T(\mathbf{c}_\chi); \theta)) \mathbf{r}_{\kappa,\chi} \mathbf{r}_{\kappa,\chi}^\intercal \tag{15}$$

Second, REALMS selects a kernel width $\sigma^{i\dagger}$ among the candidate kernel widths where its learned metric basis vector $\mathbf{a}^{i\dagger}$ yields minimum LOO regression residuals $\mathcal{L}_{c^i}$ for parameter c^i.

4.4 Projection Normalization

To account for variations caused by x-ray scatter that produces inconsistent projection intensities, REALMS normalizes both the training projection images $\mathbf{P}(I \circ T(\mathbf{c}_\kappa); \theta)$ and the on-board projection image $\mathbf{\Psi}(\theta)$. In particular, it uses the localized Gaussian normalization introduced in Chou et al. [8], which has shown promise in removing the undesired scattering artifacts.

5 Results

5.1 Synthetic Tests

We used coronal DRRs (dimension: 64×48) of the target CTs as synthetic on-board cone-beam projection images. The target CTs were deformed from the patient's Fréchet mean CT by normally distributed random samples of the first three deformation parameters.[1] We generated 600 synthetic test cases from 6 lung datasets and measured the registration quality by the average *mTRE* (mean Target Registration Error) over all cases and all voxels at tumor sites.

[1] In our lung datasets, the first three deformation parameters captured more than 95% lung variation observed in their RCCTs.

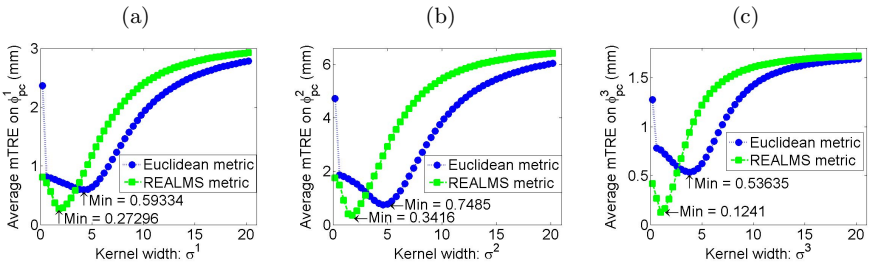


Fig. 1. Average mTREs over 600 test cases projected onto the (a) first, (b) second, and (c) third deformation basis vector versus the candidate kernel widths using $N = 125$ training projection images

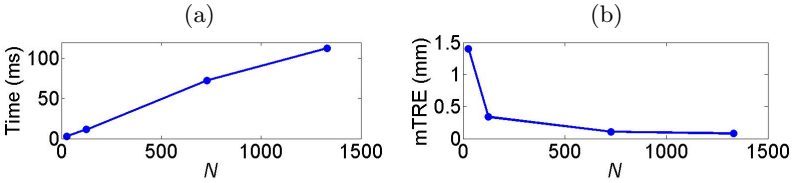


Fig. 2. (a) Time and (b) accuracy v.s. the number of training projection images N

With REALMS's registrations, the average mTRE and its standard deviation are down from 6.89 ± 3.53 mm to 0.34 ± 0.24 mm using $N = 125$ training projection images. The computation time for each registration is 11.39 ± 0.73 ms (87.79 fps) on Intel Core2 Quad CPU Q6700. As shown in figure 1, REALMS reduces the minimum errors produced by kernel regressions that use the Euclidean metric ($\mathbf{M}^i = \mathbf{I}$).

Figure 2 shows the computation time and registration accuracy tradeoff in REALMS.

5.2 Real Tests

We tested REALMS on 6 lung datasets with an on-board CBCT system where a *single* coronal on-board CB projection (dimension downsampled to 64×48 for efficient computation) at both *EE* (End-Expiration) and *EI* (End-Inspiration) phases were used for the testing. See the top image of figure 4(b) for illustration. For each dataset, we generated $N = 125$ training DRRs to learn the metrics and select optimal interpolation kernel widths. The learned metrics and the selected kernel widths were used to estimate deformation parameters for the testing EE and EI on-board projections. The estimated CTs were deformed from the Fréchet mean CT with the estimated deformation parameters. The results were validated with reconstructed CBCTs at target phases.[2] Table 1 shows the

[2] The CBCTs were reconstructed by the retrospectively-sorted CB projections at target breathing phases.

Table 1. Tumor Centroid Differences (*TCD*) after REALMS's registration at EE and EI phases of 6 lung datasets. Numbers inside the parentheses are the initial TCDs.

dataset#	TCD at EE phase (mm)	TCD at EI phase (mm)	Time (ms)
1	2.42 (9.70)	4.06 (7.45)	10.40
2	3.60 (4.85)	3.60 (4.89)	10.92
3	2.30 (8.71)	3.60 (4.03)	10.91
4	1.27 (2.69)	2.80 (2.29)	10.91
5	0.70 (9.89)	3.28 (8.71)	11.15
6	1.98 (2.03)	1.12 (1.72)	11.08

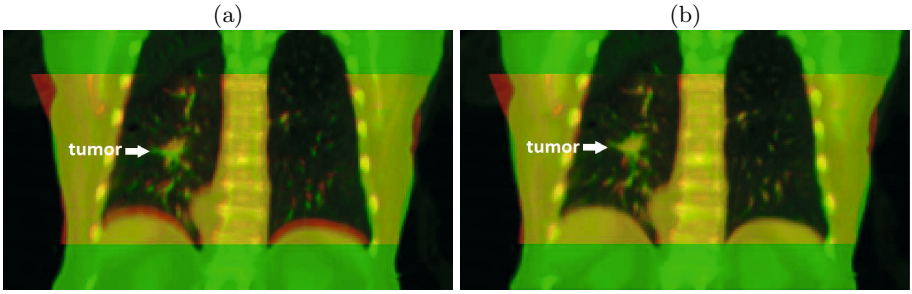


Fig. 3. (a) Image overlay of the reconstructed CBCT at EE phase (red) and the Fréchet mean CT (green) (b) Image overlay of the reconstructed CBCT at EE phase (red) and the REALMS-estimated CT (green) calculated from an on-board cone-beam projection image at EE phase. The yellow areas are the overlapped region.

3D Tumor Centroid Differences (*TCD*s) between REALMS-estimated CTs and the reconstructed CBCTs at the same respiratory phases. Tumor centroids were computed via Snake active segmentations. As shown in table 1, REALMS reduces the TCD from 5.58 ± 3.14 mm to 2.56 ± 1.11 mm in 10.89 ± 0.26 ms (91.82 fps).

Figure 3 illustrates an example REALMS registration on a lung dataset where the tumor, the diaphragm, and most of the soft tissues are correctly aligned.

5.3 The Learned Metric Basis Vector

The learned metric basis vector $\mathbf{a}^{i\dagger}$ will emphasize projection pixels that are significant for the distance calculation of the deformation parameter c^i (e.g. give high positive or high negative values). As shown in figure 4(a), the learned metric basis vector $\mathbf{a}^{1\dagger}$ emphasized the diaphragm locations and the lung boundaries as its corresponding deformation basis vector ϕ^1_{pc} covers the expansion and contraction motion of the lung. See the bottom image of figure 4(b) for illustration.

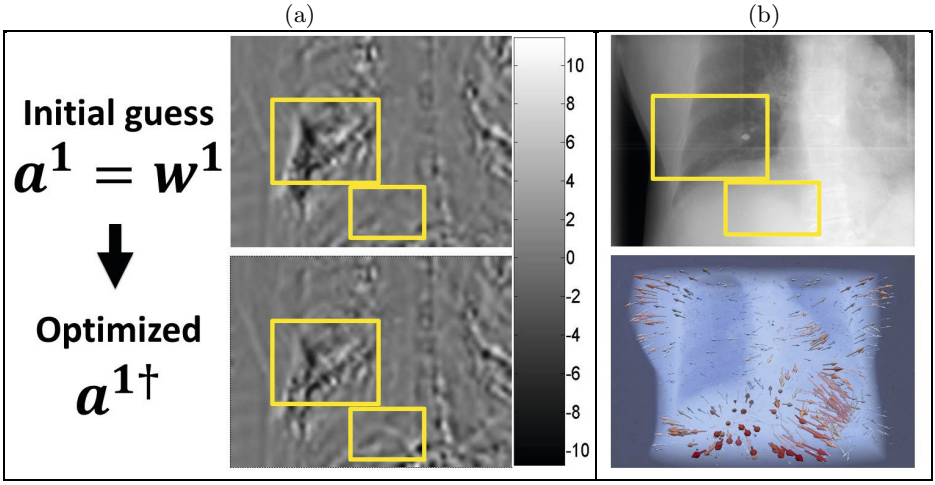


Fig. 4. (a) Initial guess of the metric basis vector $\mathbf{a}^1 = \mathbf{w}^1$ (top) and the optimized metric basis vector $\mathbf{a}^{1\dagger}$ (bottom) of a lung dataset. They are re-shaped into projection image domain for visualization. As shown in the figure, the diaphragm locations and the lung boundaries (yellow boxes) were emphasized after metric learning. (b) Top: a coronal on-board CB projection at EE phase of the lung dataset used in (a). The yellow boxes in (a) and (b) correspond to the same 2D locations. Bottom: the first deformation basis vector ϕ^1_{pc} (the color arrows indicate heat maps of the deformation magnitudes) overlaid with the volume rendering of the Fréchet mean CT of the lung dataset used in (a). For this dataset, ϕ^1_{pc} covers the expansion and contraction motion of of the lung.

6 Conclusion and Discussion

This paper presents an accurate and real-time 2D/3D registration method, REALMS, that estimates 3D deformation parameters from a single projection image using kernel regressions with learned rank-1 projection distance metrics. The learned distance metrics are optimized with an initialization approximated by linear regression that we found, is essential to the success of this high dimensional metric learning. Without this special initialization, the optimization would have easily converged to local minimum and thus produce wrong distance metrics. With this special initialization, the regression estimation on both synthetic and real test cases showed its good promise in supporting real-time and low-dose IGRT by using a single projection image. In this paper, we use highly down-sampled projection images for efficient learning at planning time. To support efficient learning for projection images of higher dimensions, the future work of REALMS will incorporate neighborhood approximation methods in the leave-one-out training such that the computation complexity will be reduced from $O(N^2)$ to $O(kN)$ if only k nearest training neighbors are considered for the regression estimation.

References

1. Russakoff, D.B., Rohlfing, T., Maurer, C.: Fast intensity-based 2D-3D image registration of clinical data using light fields. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, vol. 1, pp. 416–422 (2003)
2. Russakoff, D.B., Rohlfing, T., Mori, K., Rueckert, D., Ho, A., Adler, J.R., Maurer, C.R.: Fast generation of digitally reconstructed radiographs using attenuation fields with application to 2d-3d image registration. IEEE Transactions on Medical Imaging 24, 1441–1454 (2005)
3. Khamene, A., Bloch, P., Wein, W., Svatos, M., Sauer, F.: Automatic registration of portal images and volumetric ct for patient positioning in radiation therapy. Medical Image Analysis 10, 96–112 (2006)
4. Munbodh, R., Jaffray, D.A., Moseley, D.J., Chen, Z., Knisely, J.P.S., Cathier, P., Duncan, J.S.: Automated 2d-3d registration of a radiograph and a cone beam ct using line-segment enhancement. Medical Physics 33, 1398–1411 (2006)
5. Li, R., Jia, X., Lewis, J.H., Gu, X., Folkerts, M., Men, C., Jiang, S.B.: Real-time volumetric image reconstruction and 3d tumor localization based on a single x-ray projection image for lung cancer radiotherapy. Medical Physics 37, 2822–2826 (2010)
6. Li, R., Lewis, J.H., Jia, X., Gu, X., Folkerts, M., Men, C., Song, W.Y., Jiang, S.B.: 3d tumor localization through real-time volumetric x-ray imaging for lung cancer radiotherapy. Medical Physics 38, 2783–2794 (2011)
7. Chou, C.R., Frederick, B., Chang, S., Pizer, S.: A Learning-Based patient repositioning method from Limited-Angle projections. In: Angeles, J., Boulet, B., Clark, J.J., Kövecses, J., Siddiqi, K. (eds.) Brain, Body and Machine. AISC, vol. 83, pp. 83–94. Springer, Heidelberg (2010)
8. Chou, C.R., Frederick, B., Liu, X., Mageras, G., Chang, S., Pizer, S.: Claret: A fast deformable registration method applied to lung radiation therapy. In: Fourth International (MICCAI) Workshop on Pulmonary Image Analysis, pp. 113–124 (2011)
9. Miao, S., Liao, R., Zheng, Y.: A hybrid method for 2-d/3-d registration between 3-d volumes and 2-d angiography for trans-catheter aortic valve implantation (tavi). In: ISBI, pp. 1215–1218 (2011)
10. Weinberger, K., Tesauro, G.: Metric learning for kernel regression. In: Eleventh International Conference on Artificial Intelligence and Statistics, pp. 608–615 (2007)
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(3), 440–451 (2006)

Groupwise Spectral Log-Demons Framework for Atlas Construction

Herve Lombaert[1,2], Leo Grady[3], Xavier Pennec[2], Jean-Marc Peyrat[4], Nicholas Ayache[2], and Farida Cheriet[1]

[1] Ecole Polytechnique de Montreal, Canada
[2] INRIA Sophia Antipolis, France
[3] Siemens Corporate Research, Princeton, NJ
[4] Siemens Molecular Imaging, Oxford, UK

Abstract. We introduce a new framework to construct atlases from images with very large and complex deformations. The atlas is build in parallel with groupwise registrations by extending the symmetric Log-Demons algorithm. We describe and evaluate two forms of our framework: the *Groupwise Log-Demons* (GL-Demons) is faster but is limited to local nonrigid deformations, and the *Groupwise Spectral Log-Demons* (GSL-Demons) is slower but, due to isometry-invariant representations of images, can construct atlases of organs with high shape variability. We demonstrate our framework by constructing atlases from hearts with high shape variability.

1 Introduction

Statistics on complex characteristics with high anatomical and functional variability require the normalization of measurements across subjects to establish a population average and deviations from that average. The process of shape averaging [22,5,27] becomes particularly complex, and still remains unsolved, with organs undergoing large shape disparities. In the present state-of-the-art, the concept of geodesic shape averaging allows unbiased constructions of atlases through diffeomorphic methods [12,2,17], i.e., the transformation of a reference shape toward an average (the geometry of the atlas) follows a geodesic path on a Riemannian manifold (the space of diffeomorphic transformations). While the LDDMM [4,3,6] or forward scheme approaches [1,8] provide elegant mathematical frameworks for averaging shapes, these methods could be slow and find their limitations with high shape variability. Guimond *et al.* [10] proposed a fast and efficient algorithm [19,16,26] with sequential (pairwise) registrations to a reference image. A new simultaneous (groupwise) registration approach would enable the construction of an atlas in parallel, during the registration process (rather than with a series of pairwise registrations). To do so, *firstly*, we extend the symmetric Demons algorithm [25] to perform a groupwise registration of a set of images in order to construct their atlas. However, as in most registration methods, transformation updates based on the image gradients are inherently limited by their local scope. *Secondly*, we introduce a new update scheme for groupwise

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 11–19, 2013.

registration based on the spectral decomposition of graph Laplacians [7,23,13], that is invariant to shape isometry and is capable of capturing large deformations during the construction of the atlas. We provide *two forms* of our groupwise registration framework that we name the *Groupwise Log-Demons* (**GL-Demons**, faster and suited for local nonrigid deformations), and the *Groupwise Spectral Log-Demons* (**GSL-Demons**, slower but capable of capturing very large deformations). We evaluate the two forms of our new framework by constructing atlases of images with very large deformations.

2 Method

The atlas is defined as the set of N images $\{I_i\}_{i=1..N}$ nonrigidly aligned to their average shape $\tilde{I}$. Our new shape averaging framework extends the symmetric Log-Demons algorithm [25] and can use classical gradient-based updates (*GL-Demons*) or an improved spectral matching for groupwise registration (*GSL-Demons*). We begin by briefly reviewing each component.

2.1 Diffeomorphic Registration

A diffeomorphic transformation ϕ between two images (such that $F(\cdot) \mapsto M(\phi(\cdot))$ or simply $F \mapsto M \circ \phi$) guarantees a smooth one-to-one mapping (i.e., differentiable and invertible, without creating foldings in space). From the theory of Lie groups, the exponential map of a stationary velocity field v generates a diffeomorphic transformation $\phi = \exp(v)$ (approximated with the scaling-and-squaring method [24]). The Log-Demons algorithm alternates the optimization of a similarity term and a regularization term by decoupling them with a hidden variable (the correspondence c). The algorithm is slightly modified from [25] to converge toward an average shape by minimizing the following energy (controlled with $\alpha_i, \alpha_x, \alpha_T$):

$$E(F, M, c, v) = \alpha_i^2 \mathrm{Sim}(F', M') + \alpha_x^2 \mathrm{dist}(c, v)^2 + \alpha_T^2 \mathrm{Reg}(v), \text{ where} \quad (1)$$
$$\mathrm{Sim}(F', M') = (F' - M')^2, \ \mathrm{dist}(c, v) = \|c - v\|, \text{ and } \mathrm{Reg}(v) = \|\nabla v\|^2$$

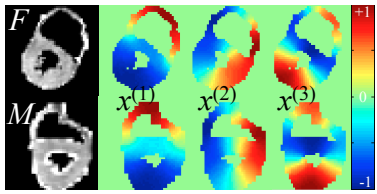
The similarity term incorporates diffeomorphism and symmetry with $F' = F \circ \exp(-c)$ and $M' = M \circ \exp(+c)$. Both images F' and M' effectively converge toward an average shape $\tilde{I} = F \circ \phi^{-1} + M \circ \phi$ (similar to the approaches in [2,6]).

2.2 Spectral Correspondence

The computation of the velocity field updates in the Log-Demons is inherently limited by the local scope of the update forces derived from the image gradient, i.e., it requires texture data which is generally local information. We now describe a new update scheme based on spectral correspondence [21,11,18,14,13] that will enable the construction of atlases with large deformations. Let us first consider I_Ω, the portion of an image I bounded by a contour Ω. We build a connected

graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ where the vertices $\mathscr{V}$ represent the pixels of I_Ω and the edges $\mathscr{E}$ define the neighborhood structure within I_Ω. The corresponding adjacency matrix W [9] represents the edge weights ($W_{ij} = w_{ij}$ if pixels (i, j) are neighbors, 0 otherwise), such that pixels with similar intensity and close in space would have strong links in $\mathscr{G}$ (e.g., $w_{ij} = \exp(-\beta(I(i) - I(j))^2) \,/\, \|\mathbf{x}(i) - \mathbf{x}(j)\|^2$ where $\mathbf{x}$ are Euclidean coordinates and β a parameter). The Laplacian operator on a graph [9] is formulated as a $|\mathscr{V}| \times |\mathscr{V}|$ matrix with the form $\mathcal{L} = D^{-1}(D - W)$, where D is the (diagonal) degree matrix containing the node degrees $D_{ii} = \sum_j W_{ij}$.

Spectral Coordinates. The decomposition of the Laplacian matrix $\mathcal{L} = \mathscr{X}^T \Lambda \mathscr{X}$ reveals the graph spectrum [7] which comprises the eigenvalues $\Lambda = \text{diag}(\lambda_0, \lambda_1, ..., \lambda_{|\mathscr{V}|})$ (in increasing order) and their associated eigenmodes $\mathscr{X} = \left(\mathscr{x}^{(0)}, \mathscr{x}^{(1)}, ..., \mathscr{x}^{(|\mathscr{V}|)}\right)$ (a $|\mathscr{V}| \times |\mathscr{V}|$ matrix where columns $\mathscr{x}^{(\cdot)}$ are eigenmodes). The first eigenmode is trivial ($\lambda_0 = 0$) and the following non-trivial eigenmodes are the fundamental modes of vibrations of a shape depicted by I_Ω. The eigenmodes associated with the first k smallest non-zero eigenvalues (the lower frequencies) represent the k-dimensional *spectral coordinates* (each point $i \in I_\Omega$ has the coordinates $\mathscr{x}(i) = \left(\mathscr{x}^{(1)}(i), \mathscr{x}^{(2)}(i), ..., \mathscr{x}^{(k)}(i)\right)$ defined in a spectral domain). These lowest modes of vibration have the strong property of being smooth and invariant to shape isometry (i.e., shapes in different poses would share the same spectral coordinates at each point, see *below*).



Three lowest frequency eigenmodes of two images

However, the eigenmodes need to be rearranged as a result of sign ambiguity ($\mathscr{x}^{(\cdot)}$ and $-\mathscr{x}^{(\cdot)}$ are both valid eigenmodes), algebraic multiplicity (many eigenmodes can share the same eigenvalue), and imperfection in isometry (changing the multiplicity and ordering of the eigenvalues). Firstly, their values are scaled to fit the range $[-1; +1]$, i.e., for negative values: $\mathscr{x}^{(\cdot)-} \leftarrow \mathscr{x}^{(\cdot)-} / \min\{\mathscr{x}^{(\cdot)-}\}$ and for positive values: $\mathscr{x}^{(\cdot)+} \leftarrow \mathscr{x}^{(\cdot)+} / \max\{\mathscr{x}^{(\cdot)+}\}$. Secondly, the eigenmodes of two images, $\mathscr{x}_F$ and $\mathscr{x}_M$, are reordered with the optimal permutation π (where $\mathscr{x}_F^{(\cdot)} \mapsto \mathscr{x}_M^{\pi\circ(\cdot)}$) which may be found with the Hungarian algorithm that minimizes the following dissimilarity matrix:

$$C(u,v) = \sqrt{\frac{1}{|I_\Omega|} \sum_{i \in I_\Omega} \left(\mathscr{x}_F^{(u)}(i) - \mathscr{x}_M^{(v)}(i)\right)^2} + \sqrt{\sum_{i,j} \left(h_F^{\mathscr{x}_F^{(u)}}(i,j) - h_M^{\mathscr{x}_M^{(v)}}(i,j)\right)^2} \tag{2}$$

The first term is the difference in spectral coordinates between the images. The second term measures the dissimilarities between the joint histograms $h(i, j)$ (a 2D matrix where the element (i, j) is the joint probability of having at the same time the intensity i and the eigenmodal value $\mathscr{x}^{(\cdot)} = j$). The sign ambiguity can be removed by optimizing, instead, the dissimilarity matrix $Q(u, v) = \min\{C(u, v), C(u, -v)\}$. To keep the notation simple in the next sec-

Algorithm 1. Spectral Correspondence

Input: Images F, M.
Output: Correspondence c mapping F to M

- Compute general Laplacians $\mathcal{L}_F$, $\mathcal{L}_M$.
  $\mathcal{L} = D^{-1}(D - W)$, where
  $W_{ij} = \exp(-\beta(I(i) - I(j))^2)/\|\mathbf{x}(i) - \mathbf{x}(j)\|^2$
  $D_{ii} = \sum_j W_{ij}$,
- Compute first k eigenmodes of Laplacians
- Reorder $\mathscr{X}_M$ with respect to $\mathscr{X}_F$ (Eq. (2))
- Build embeddings:
  $\mathbf{F} = (I_F, \mathbf{x}_F, \mathscr{X}_F)$; $\mathbf{M} = (I_M, \mathbf{x}_M, \mathscr{X}_M)$
- Find c mapping nearest points $\mathbf{F} \mapsto \mathbf{M}$

Algorithm 2. Groupwise Demons Framework

Input: N images with initial reference (e.g., $\tilde{I} = I_1$)
Output: Transformations $\phi_i = \exp(v_i)$ mapping $\tilde{I}$ to I_i

Average shape is $\tilde{I} = \frac{1}{N}\sum_{i=1}^{N} I_i \circ \exp(v_i)$
repeat
for $i = 1 \rightarrow N$ **do**

- Find updates $u_i \leftarrow \text{mapping}(\tilde{I}, I_i \circ \exp(v_i))$.
  (mapping() differs in GL and GSL-Demons)
- Smooth updates: $u_i \leftarrow K_{\text{fluid}} \star u_i$.
  (convolution of a Gaussian kernel on u_i)
- Update velocity fields: $v_i \leftarrow \log(\exp(v_i) \circ \exp(u_i))$
  (approximated with $v_i \leftarrow v_i + u_i$).
- Smooth velocity fields: $v_i \leftarrow K_{\text{diff}} \star v_i$.

end for

- Get reference update: $u_{\text{ref}} = -\frac{1}{N}\sum_{i=1}^{N} v_i$
- Update velocity fields: $v_i \leftarrow v_i + u_{\text{ref}}$.
- Update reference: $\tilde{I} \leftarrow \frac{1}{N}\sum_{i=1}^{N} I_i \circ \exp(v_i)$.

until convergence

tions, we assume the spectral coordinates have been appropriately signed, scaled and reordered using this method.

Spectral Matching. The correspondence between two images F and M is established (Alg. (1)) by finding the nearest neighbors in the spectral domain (e.g., with fast k-d trees). Put differently, if $\mathscr{X}_F(i)$ is the closest point to $\mathscr{X}_M(j)$ then the pixel i corresponds with j. This simple nearest-neighbor scheme is extended to add similarity constraints on intensity and space by adding image intensities and Euclidean coordinates to the spectral embedding: $\mathbf{X} = (\alpha_{\text{i}} I, \alpha_{\text{s}}\mathbf{x}, \alpha_{\text{g}}\mathscr{X})$. Nearest points between $\mathbf{X}_F$ and $\mathbf{X}_M$ actually locate the best compromise among three strong properties: points with similar isometric (or geometric) properties, similar image intensities, and similar location (each weighted with $\alpha_{\text{g,i,s}}$). To be more precise, this corresponds to minimizing the energy $E(F, M, \phi) = \text{Sim}(F, M)$ where the regularization (similarly to [14]) is enforced with the smoothness of the spectral and spatial components:

$$\text{Sim}(F, M) = (F - M \circ \phi)^2 + \frac{\alpha_{\text{s}}^2}{\alpha_{\text{i}}^2}(\mathbf{x}_F - \mathbf{x}_{M\circ\phi})^2 + \frac{\alpha_{\text{g}}^2}{\alpha_{\text{i}}^2}(\mathscr{X}_F - \mathscr{X}_{M\circ\phi})^2, \qquad (3)$$

where $\mathscr{X}_F$ and $\mathscr{X}_{M\circ\phi}$ are the spectral coordinates of corresponding points. This matching technique that is invariant to isometry will enable the capture of large deformations for our atlas construction.

2.3 Groupwise Demons Framework

Our framework is based on Guimond's *et al.* approach [10] where they construct the average image $\tilde{I}$ *sequentially* by alternating between pairwise registrations

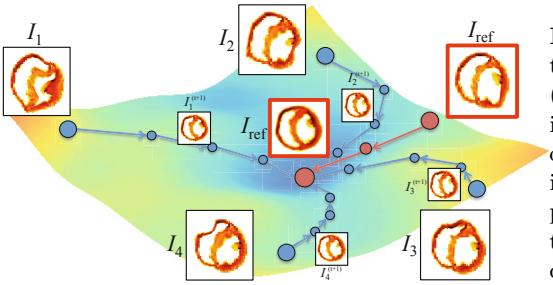


Fig. 1. Groupwise Demons: Simultaneous registration of 4 images (blue circles) toward a reference image that evolves in the space of diffeomorphisms (colored manifold). The reference image is computed in parallel and converges to the average shape (middle red circle).

(fixing a reference image) and updates of the average image (transforming the reference image). Our novelty is to directly compute $\tilde{I}$ *in parallel* with simultaneous (groupwise) registrations (illustrated in Fig. 1). To do so, Eq. (1) is extended to incorporate N velocity fields that warp all images $\{I_i \circ \exp(c_i)\}$ toward the average image $\tilde{I}$. The new groupwise framework is summarized in Alg. (2) and the underlying energy is:

$$E(\tilde{I}, \{I_i, c_i, v_i\}) = \frac{1}{N} \sum_{i=1}^{N} \left(\alpha_i^2 \text{Sim}(\tilde{I}, I_i \circ \exp(c_i)) + \alpha_x^2 \text{dist}(c_i, v_i)^2 + \alpha_T^2 \text{Reg}(v_i) \right) \tag{4}$$

The reference image can be optionally generated with weighted contributions from all images (e.g., weights different than $1/N$ in order to remove outliers). The minimization of all similarity terms, $\{\text{Sim}(\tilde{I}, I_i')\}$, causes all warped images to become similar to the reference image and the sum of all velocity fields is brought to a minimal value at convergence. Similar to the convergence of [10], the Groupwise Demons framework effectively brings the reference image toward the barycenter of all images. The average image is simply generated with $\tilde{I} = \frac{1}{N} \sum_{i=1}^{N} I_i \circ \exp(c_i)$.

Groupwise Spectral Log-Demons. The update schemes based on image gradients and on spectral correspondence can be used in the Groupwise Demons framework. The *Groupwise Log-Demons* (GL-Demons) algorithm uses update forces derived from the image gradient and is well suited for images with local nonrigid deformations, while the *Groupwise Spectral Log-Demons* (GSL-Demons) algorithm uses spectral correspondences as update forces (i.e., u is found with Alg. (1)) and is better suited for large and highly non-local deformations. GSL-Demons enables large jumps during the construction of the atlas where points move toward their isometric equivalents even if they are far away in space. The atlas construction can handle very large deformations and convergences in fewer iterations (typically 5 iterations are sufficient). The energy has the same form of Eq. (4) and uses the similarity term of Eq. (3).

Multilevel Scheme. Moreover, large and complex deformations can be captured in a low resolution level with *GSL-Demons*, improving thus the processing time, while the remaining small and local deformations can be recovered with

GL-Demons in higher resolutions. This multilevel approach keeps the computation of the eigenmodes tractable.

3 Results

GL-Demons and *GSL-Demons* are evaluated by constructing atlases of images with large deformations. In the synthetic experiment, we verify convergence toward an average shape, and the handling of highly complex deformations (parameters: $\sigma_{\text{fluid,diff}} = 1, \alpha_x = 1, k = 5, \alpha_g = 0.1, \alpha_s = 0.2, \alpha_i = 0.7$ in 2D). In a second experiment, we use both algorithms with real cardiac images that exhibit high shape variability (parameters: $\sigma_{\text{fluid,diff}} = 0.75, \alpha_x = 1, k = 5, \alpha_g = 0.25, \alpha_s = 0.35, \alpha_i = 0.4$ in 3D).

Synthetic Deformations. Convergence and capture of large deformations are now evaluated. $N/2$ velocity fields v are generated randomly using 15 control points with random locations in the image and random displacements of at most 15 pixels (20% of the image size) that are diffused over the image. Their forward and background transformations ($\exp(v)$ and $\exp(-v)$) are applied to an initial image I_0, holding thus the average shape to I_0 (establishing our ground truth). Since we compare the convergence and its rate, and not the final performance, the multi-level scheme (which should be used in real applications) is not applied. Fig. 2 shows the groupwise registrations of 10 random hearts (2D 75×75 images) through 100 trials (a total of 1000 hearts). The average Dice metric (measuring the overlap) between all computed average shapes and I_0 as well as the intensity errors (MSE) reveal that the reference shape (defined arbitrarily as one of the 10 images) evolves toward the ground truth (i.e., Dice increases and MSE decreases). Moveover, the N deformation fields become closer to the

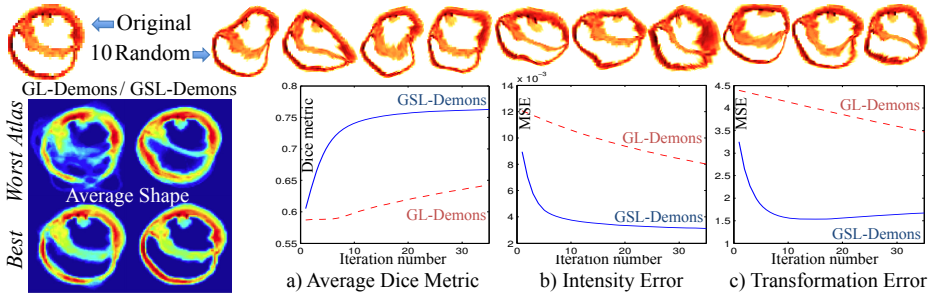


Fig. 2. Groupwise registration of 10 images deformed randomly (100 trials, 1 sample on top row, with known ground truth) using *GL-Demons* and *GSL-Demons*, *Left)* Best and worst atlases (based on Dice metric among 100 trials) demonstrating the capability of the *GSL-Demons* to handle large deformations, *a)* Average Dice metric with ground truth, *b)* Intensity difference between average shape and ground truth, *c)* transformation error with ground truth. *GSL-Demons* converges faster toward the average shape.

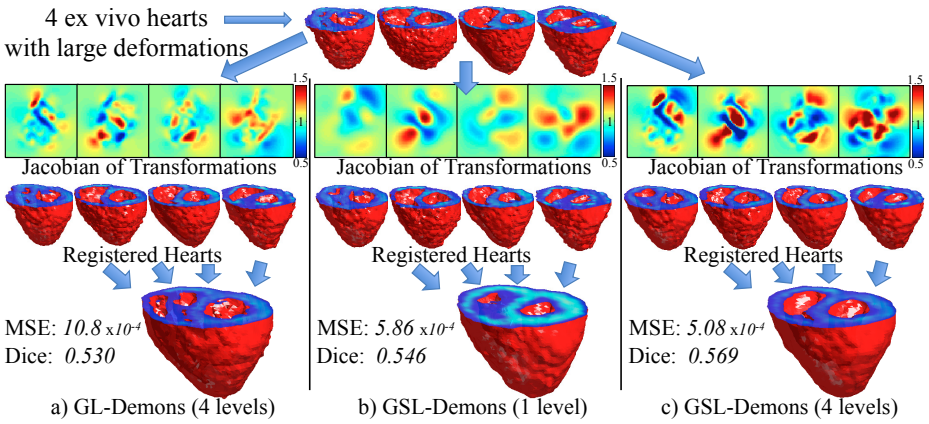


Fig. 3. Atlas of *ex vivo* hearts (isosurfaces are shown) using *a) GL-Demons* (4 levels, showing failure in the right ventricle), *b) GSL-Demons* (1 level), *c)* and *GSL-Demons* (4 levels, with correct right ventricle). *GSL-Demons* capture successfully large deformations. Jacobian determinants (axial planes) show that spectral matching capture smooth and large deformations while gradient-based updates capture local deformations.

ground truth during registration. The striking difference in the convergence rates shows the full power of GSL-Demons (less than 5 iterations are required) while *GL-Demons* might not converge with such large deformations (we stopped the algorithms after 200 iterations). Time-wise, 35 iterations takes 194 seconds with GSL-Demons, and 53 seconds with *GL-Demons* (using unoptimized Matlab code on a 2.53GHz Core 2 Duo). GSL-Demons shows a better performance with high deformations than *GL-Demons*.

Cardiac Atlases. We now evaluate the construction of atlases with organs of high shape variability. *Ex vivo* hearts are particularly challenging to register as they present a high variability in fixture poses due to flabby ventricular walls. The human *ex vivo* DTMRI dataset [20,16,15] provides good candidates to evaluate our algorithms. We use four hearts ($b = 0$ images of size 64^3) that were excluded in the construction of the human atlas [15] due to their hypertrophy and highly deformed shapes (see Fig. 3). GL-Demons (with 4 resolution levels) fail in recovering the shapes of the right ventricles, while GSL-Demons successfully constructs the atlas even with 1 level of resolution (downsampled images at size 28^3). As a comparison, 35 iterations takes 40 minutes in Matlab with *GSL-Demons* and 9 minutes with *GL-Demons*. Using *GSL-Demons* with 4 resolution levels reduce the intensity error (MSE) by half (from 10.8 to 5.08). Moreover, the Jacobian determinants of the transformation fields show that the large and highly non-local deformations are successfully captured with the spectral-based update scheme (high and smooth Jacobian in Fig. 3 b) while local deformations are captured with the gradient-based update scheme in the higher levels of *GSL-Demons* (Fig. 3 c).

4 Conclusion

We addressed the problem of atlas construction that is limited by large deformations between images. We proposed a new framework with two forms to construct an atlas in parallel with groupwise registrations: *GL-Demons* is faster but is limited by its gradient-based forces, while *GSL-Demons* is slower but can capture very large deformations due to its spectral components. We evaluated our framework by constructing atlases from images with complex deformations. Results showed convergence to an average shape and atlases were successfully created under large deformations of 20% of the image size using 1000 random hearts. We additionally showed that *GSL-Demons* can construct an atlas for a challenging dataset of *ex vivo* hearts with high shape variability. Future work will focus on implementation (converting the Matlab code, also, the groupwise nature of our framework could highly benefit from parallel computing, e.g., GPU) and improving the computation time of the spectral decomposition (e.g., reuse of pre-computations, approximations). Nevertheless, our current framework enables the construction of atlases from images with very large and complex deformations.

Acknowledgements. The authors wish to thank Pierre Croisille for *ex vivo* herts as well as Hervé Delingette for helpful comments. The project was supported financially by the National Science and Engineering Research Council of Canada (NSERC).

References

1. Allassonnière, S., Amit, Y., Trouvé, A.: Towards a coherent statistical framework for dense deformable template estimation. J. Royal Stat. Soc. 69, 3–29 (2007)
2. Avants, B., Gee, J.C.: Geodesic estimation for large deformation anatomical shape averaging and interpolation. NeuroImage 23, 139–150 (2004)
3. Beg, M.F., Khan, A.: Computing an average anatomical atlas using LDDMM and geodesic shooting. In: ISBI, pp. 1116–1119 (2006)
4. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. IJCV 61, 139–157 (2005)
5. Bhatia, K.K., Hajnal, J.V., Puri, B.K., Edwards, A.D., Rueckert, D.: Consistent groupwise non-rigid registration for atlas construction. In: ISBI, pp. 908–911 (2004)
6. Bossa, M., Hernandez, M., Olmos, S.: Contributions to 3D Diffeomorphic Atlas Estimation: Application to Brain Images. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 667–674. Springer, Heidelberg (2007)
7. Chung, F.: Spectral Graph Theory. AMS (1997)
8. Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., Ayache, N.: Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. NeuroImage 55, 1073–1090 (2011)
9. Grady, L., Polimeni, J.R.: Discrete Calculus: Applied Analysis on Graphs for Computational Science. Springer (2010)
10. Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: a convergence study. In: Computer Vision and Image Understanding, pp. 192–210 (2000)
11. Jain, V., Zhang, H.: Robust 3D shape correspondence in the spectral domain. In: Int. Conf. on Shape Modeling and App., p. 19 (2006)

12. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)
13. Lombaert, H., Grady, L., Pennec, X., Ayache, N., Cheriet, F.: Spectral Demons – Image Registration via Global Spectral Correspondence. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 30–44. Springer, Heidelberg (2012)
14. Lombaert, H., Grady, L., Polimeni, J.R., Cheriet, F.: Spectral correspondence for brain matching. In: IPMI, pp. 660–670 (2011)
15. Lombaert, H., Peyrat, J.-M., Croisille, P., Rapacchi, S., Fanton, L., Cheriet, F., Clarysse, P., Magnin, I., Delingette, H., Ayache, N.: Human atlas of the cardiac fiber architecture: Study on a healthy population. IEEE Trans. on Med. Imaging 31, 1436–1447 (2012)
16. Lombaert, H., Peyrat, J.-M., Croisille, P., Rapacchi, S., Fanton, L., Clarysse, P., Delingette, H., Ayache, N.: Statistical Analysis of the Human Cardiac Fiber Architecture from DT-MRI. In: Metaxas, D.N., Axel, L. (eds.) FIMH 2011. LNCS, vol. 6666, pp. 171–179. Springer, Heidelberg (2011)
17. Marsland, S., Twining, C.J., Taylor, C.J.: Groupwise Non-rigid Registration Using Polyharmonic Clamped-Plate Splines. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2879, pp. 771–779. Springer, Heidelberg (2003)
18. Mateus, D., Horaud, R., Knossow, D., Cuzzolin, F., Boyer, E.: Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration. In: CVPR, pp. 1–8 (2008)
19. Peyrat, J.-M., Sermesant, M., Pennec, X., Delingette, H., Xu, C., McVeigh, E.R., Ayache, N.: A computational framework for the statistical analysis of cardiac diffusion tensors: application to a small database of canine hearts. IEEE Trans. on Med. Imaging 26(11), 1500–1514 (2007)
20. Rapacchi, S., Croisille, P., Pai, V., Grenier, D., Viallon, M., Kellman, P., Mewton, N., Wen, H.: Reducing motion sensitivity in free breathing DWI of the heart with localized Principal Component Analysis. In: ISMRM (2010)
21. Shapiro, L.S., Brady, J.M.: Feature-based correspondence: an eigenvector approach. Image and Vision Computing 10, 283–288 (1992)
22. Studholme, C., Cardenas, V.: A template free approach to volumetric spatial normalization of brain anatomy. Pattern Recogn. Lett. 25, 1191–1202 (2004)
23. van Kaick, O., Zhang, H., Hamarneh, G., Cohen-Or, D.: A survey on shape correspondence. Eurographics 30(6), 1681–1707 (2011)
24. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Non-parametric Diffeomorphic Image Registration with the Demons Algorithm. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 319–326. Springer, Heidelberg (2007)
25. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 754–761. Springer, Heidelberg (2008)
26. Wu, G., Jia, H., Wang, Q., Shen, D.: SharpMean: groupwise registration guided by sharp mean image and tree-based registration. NeuroImage 56(4), 1968–1981 (2011)
27. Zollei, L., Miller, L.E., Grimson, W.E.L., Wells, W.M.: Efficient population registration of 3D data. In: ICCV 2005, Computer Vision for Biomedical Image Applications (2005)

Robust Anatomical Correspondence Detection by Graph Matching with Sparsity Constraint

Yanrong Guo[1,2], Guorong Wu[2], Yakang Dai[2], Jianguo Jiang[1], and Dinggang Shen[2,*]

[1] School of Computer and Information, Hefei University of Technology, Hefei, China
[2] IDEA Lab, Department of Radiology and BRIC, University of North Carolina at Chapel Hill
dgshen@med.unc.edu

Abstract. Graph matching is a robust correspondence detection approach which considers potential correspondences as graph nodes and uses graph links to measure the pairwise agreement between potential correspondences. In this paper, we propose a novel graph matching method to augment its power in establishing anatomical correspondences in medical images, especially for the cases with large inter-subject variations. Our contributions have twofold. First, we propose a robust measurement to characterize the pairwise agreement of appearance information on each graph link. In this way, our method is more robust to ambiguous matches than the conventional graph matching methods that generally consider only the simple geometric information. Second, although multiple correspondences are allowed for robust correspondence, we further introduce the sparsity constraint upon the possibilities of correspondences to suppress the distraction from misleading matches, which is very important for achieving accurate one-to-one correspondences in the end of the matching procedure. We finally incorporate these two improvements into a new objective function and solve it by quadratic programming. The proposed graph matching method has been evaluated in the public hand X-ray images with comparison to a conventional graph matching method. In all experiments, our method achieves the best matching performance in terms of matching accuracy and robustness.

1 Introduction

Robust anatomical correspondence detection is very important in many medical image applications, such as deformable image registration [1] and organ motion correction [2]. Although a lot of local image descriptors have been proposed with great success in computer vision area in the last decade, it remains a big challenge in establishing correspondences between subjects with large anatomical differences.

Recently, graph matching has emerged as a robust correspondence detection approach by modeling not only the point-to-point correspondence [3] but also the pair-to-pair matching consistency in a graph [4]. Specifically, each possible correspondence is considered as a node in the graph and the pairwise agreement between any two possible correspondences is described as a link in the graph. Then,

* Corresponding author.

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 20–28, 2013.

the problem of correspondence matching becomes an optimization problem for finding a cluster of these nodes that can produce the maximal pairwise agreement.

In general, the advantages of graph matching over other pointwise correspondence detection methods lie in two aspects: (1) the matching coherence is explicitly modeled in the graph to leverage the problem of ambiguous matches; (2) multiple correspondences are allowed in correspondence detection while the final one-to-one correspondences are simultaneously solved on all correspondences by the spectral-based optimization method [5]. However, there are two major issues in the current graph matching methods: (1) only simple geometric information is generally used for constructing the graph links; (2) its solution is usually suboptimal due to the lack of effective mechanism to control the quality of each possible correspondence established.

To alleviate these two issues, we present a novel graph matching method to augment its power in establishing anatomical correspondences, especially for the cases with large inter-subject shape variations in the medical images. Our contributions have twofold. **First**, we propose a robust appearance measurement to characterize the pairwise agreement on each graph link. Specifically, for any two possible matches (with the two starting points in the template image and the two ending points in the subject image), a sequence of local intensity profiles (called *line patch*) along the line connecting two starting points in template image, or two ending points in the subject image, is constructed. Then the appearance discrepancy between these two line patches is computed to measure their pairwise agreement. Using this novel measurement, our method is more robust to ambiguous matches than the conventional graph matching methods that generally use only the simple geometric compatibility. **Second**, inspired by the discriminative power of sparse representation in machine learning and pattern recognition [6, 7], we apply a sparsity constraint on the possibilities of multiple correspondences, which requires to seek for only a small number of qualified correspondences for each feature point. Thus, the risk of ambiguous matches can be significantly avoided when determining one-to-one correspondences in the end of matching procedure. We finally construct a new objective function by integrating these two improvements. An efficient solution is further provided, via quadratic programming, to jointly estimate correspondences for all feature points. Our graph matching method has been evaluated in the public hand X-ray images and compared with the state-of-the-art graph matching method, namely Spectral Matching with Affine Constraint (SMAC) [4], which was reported with one of the best matching performances among the conventional graph matching methods. In all experiments, our method outperforms SMAC in terms of both matching accuracy and robustness.

2 Methods

Considering a feature point set $T = \{t_i | i = 1, \dots, n\}$ in the template image and another feature point set $S = \{s_{i'} | i' = 1, \dots, n'\}$ in the subject image, our goal is to find an assignment matrix $\boldsymbol{X} = \left[X_{i,i'}\right]_{n \times n'} \left(X_{i,i'} \in \{0,1\}\right)$ between these two point

sets, where each assignment $X_{i,i'}$ indicates whether a feature point t_i in the template is matched to a feature point $s_{i'}$ in the subject with '1' denoting correspondence and '0' denoting non-correspondence. Fig. 1 schematically illustrates the main idea of our method by using the hand X-ray images as example. Given the template feature point set T (Fig. 1(a)) and the subject feature point set S (Fig. 1(b)), all possible correspondences between T and S are established as shown by the white lines in Fig. 1(c). Then, the $nn' \times nn'$ affinity matrix $\boldsymbol{M}$ (Fig. 1(d)) can be constructed to describe the confidence of all established correspondences as well as the pairwise agreement between any two possible matches. Specifically, each diagonal element (shown with boxes in Fig. 1(d)) in the affinity matrix $\boldsymbol{M}$ represents the pointwise similarity between two feature points $t_i \in T$ and $s_{i'} \in S$. Each off-diagonal element (shown with pink triangle in Fig. 1(d)) measures the pairwise agreement between two possible matches $((i,i')$ indicated by the red box and (j,j') indicated by the blue box in Fig. 1(d)), where we propose to use appearance-based line patch, combined with simple geometric relationship [4], to robustly characterize their coherence. The continuous relaxed assignment matrix $\boldsymbol{X} = [X_{i,i'}]_{n\times n'}(X_{i,i'} \in [0,1])$ can be optimized by finding the cluster of correspondences among the diagonal elements of $\boldsymbol{M}$ while maximizing its pairwise agreements. To alleviate the potential ambiguity in determining one-to-one correspondences in Fig. 1(e) directly from the one-to-many assignment matrix $\boldsymbol{X}$, the sparsity constraint is applied to $\boldsymbol{X}$ to suppress the distraction of ambiguous matches during the correspondence detection procedure.

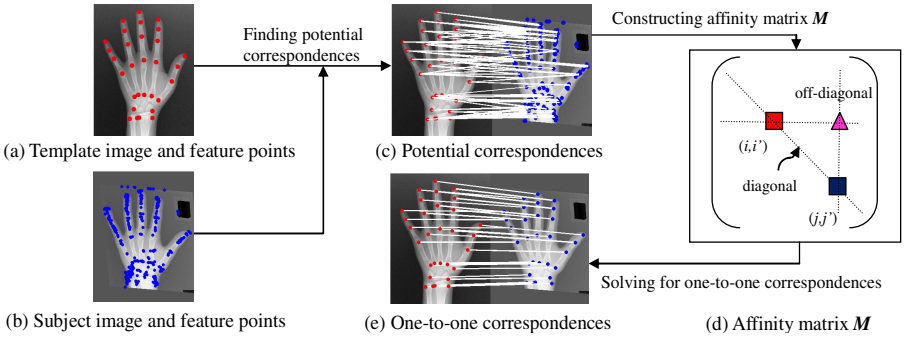


Fig. 1. The scheme of the proposed anatomical correspondence detection by graph matching

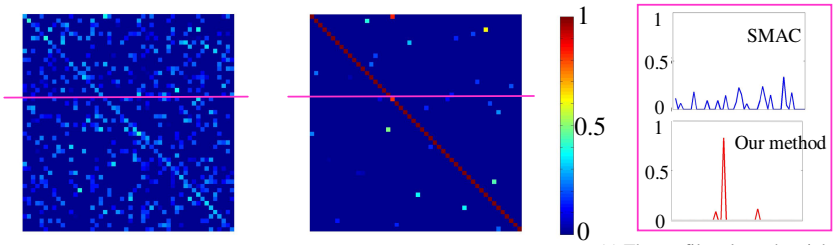
2.1 Limitation of Conventional Graph Matching Method

In the conventional graph matching method, such as the SMAC method, the coherence between possible matches (i,i') and (j,j') is usually measured by the geometric distance $\frac{|d(i,j)-d(i',j')|}{\min(d(i,j),d(i',j'))}$, and the angle between two matches/ correspondences (i,i') and (j,j'). Here, $d(.,.)$ denotes the Euclidian distance of two points. Then, the energy function is defined to maximize the following quadratic score function of $\boldsymbol{x}$:

$$J(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} \quad s.t.\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{1}^{\boldsymbol{T}} \text{ and } \{0 \le x_m \le 1 | m = 1, \dots, nn'\} \quad (1)$$

where assignment vector $\boldsymbol{x}$ is a nn' column vector after concatenating each row of $\boldsymbol{X}$. Thus, each element x_m $(m = 1, \dots, nn')$ in the vector $\boldsymbol{x}$ is associated with a particular correspondence (i, i') in the assignment matrix $\boldsymbol{X}$, i.e., $x_m = X_{i,i'}$. Since the optimization of $J(\boldsymbol{x})$ is NP-hard, each element in $\boldsymbol{x}$ is relaxed to be a continuous value between 0 and 1. Thus, the objective function $J(\boldsymbol{x})$ is subject to the affine constraint $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{1}^{\boldsymbol{T}}$ (as in [3]) to enforce the one-to-one correspondences. $\boldsymbol{A}$ is a $(n + n') \times (nn')$ selection matrix applied to vector $\boldsymbol{x}$ (vectorization of $\boldsymbol{X}^{\boldsymbol{T}}$) to represent the summation of each column or each row of $\boldsymbol{X}$ equals to 1, i.e., $\sum_{i=1}^{n} X_{i,i'} = 1$ or $\sum_{i'=1}^{n'} X_{i,i'} = 1$. Spectral relaxation technique can be used to maximize the energy function in Eq. 1.

Fig. 2(a) shows the optimized assignment matrix $\boldsymbol{X}$ by the SMAC method. It can be observed that the distribution of assignment in most rows (or columns) of $\boldsymbol{X}$ is not sharp (with an example of $X_{i,i'}$ values along the pink line shown in the top of Fig. 2 (c)), indicating that it is still very difficult to determine the one-to-one correspondence for each feature point based on the one-to-many correspondences (each with similar likelihood). Thus, a good solution is to keep the large assignments only for the good matches while suppress the distractions from ambiguous matches. To achieve this, we propose to (1) utilize the appearance-based line patch to exclude the in-correct matches when constructing the affinity matrix $\boldsymbol{M}$ and (2) further apply sparsity to the assignment matrix $\boldsymbol{X}$ during the optimization procedure to suppress the influence from ambiguous matches, as will be presented below.



(a) Assignment matrix $\boldsymbol{X}$ of SMAC (b) Assignment matrix $\boldsymbol{X}$ of our method (c) The profiles along the pink lines

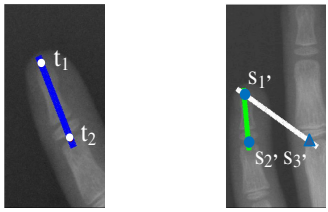
Fig. 2. The assignment matrix $\boldsymbol{X}$ optimized from the same affinity matrix by SMAC method (without sparsity constraint) and our method (with sparsity constraint)

2.2 Improved Graph Matching Method

Construction of Robust Affinity Matrix with Line Patch: It is clear that the matching performance is largely dependent on the established affinity matrix $\boldsymbol{M}$, especially the off-diagonal elements which characterize the pairwise agreement between two possible correspondences (i, i') and (j, j'). However, the conventional graph matching methods only consider the geometric coherence between (i, j) and (i', j'). Although local image descriptor can be used to measure the appearance similarities between feature point t_i and $s_{i'}$, as well as between t_j and $s_{j'}$, it still

fails to discriminate the unreasonable matches as shown in Fig. 3. In this example, there are two template feature points and three subject feature points. Subject feature points $s_{1'}$ and $s_{2'}$ (blue circles) are the correct matches of template feature points t_1 and t_2 (white circles) in the template image, while $s_{3'}$ (blue triangle) is the incorrect match to t_2. However, neither the geometric coherence nor local descriptor based measurement is able to distinguish the incorrect correspondence $(2, 3')$ from the correct one $(2, 2')$ in the affinity matrix $\boldsymbol{M}$, which affects the optimization of assignment matrix $\boldsymbol{X}$ in Eq. 1.

To solve this problem, we define the *line patch* by utilizing a sequence of intensity profiles along the line connecting the two feature points in the template or subject image. In Fig. 3, the image intensity profiles along the lines $\overline{t_1 t_2}$, $\overline{s_{1'} s_{2'}}$ and $\overline{s_{1'} s_{3'}}$ are displayed as blue, green, and white stripes, respectively. Thus, a collection of intensity profiles along the underlying stripe can be captured, which is referred to as the *line patch* in our method, to measure the pairwise agreement of two possible matches. Specifically, normalized cross correlation is used to measure the similarity between line patches. As shown in the right part of Fig. 3, the pairwise agreement measured by the line patches is able to distinguish between the correct and incorrect matches. Here we note that the radius of intensity profile is set to 5 pixel and we uniformly sample 60 local intensity profiles for each line patch. Thus, the number of intensity values included in the line patch of our method is 11×60.



(a) Line patch on template T (b) Line patches on subject S

Normalized cross correlation between line patches

	($s_{1'}$, $s_{2'}$) (correct)	($s_{1'}$, $s_{3'}$) (incorrect)
(t_1, t_2)	0.7616	0.1750

(c) Similarity between line patches

Fig. 3. Demonstration of using line patches in removing incorrect matches. Three possible correspondences are shown, i.e., $(t_1, s_{1'})$ (correct), $(t_2, s_{2'})$ (correct), $(t_2, s_{3'})$ (incorrect). The pairwise agreement between correct matches $(t_1, s_{1'})$ and $(t_2, s_{2'})$ is measured by the similarity of blue and green line patches, while another pairwise agreement is measured by blue and white line patches for incorrect match. Since each line patch utilizes the intensity profiles between two feature points, it is able to suppress the incorrect matches in the affinity matrix, as quantitatively measured by the normalized cross correlation listed in the right part of this figure.

Sparse Constraint on Assignment Vector: Although the one-to-many correspondence strategy ensures detection of all possible matches for each feature point, it also introduces many ambiguous matches, which could affect the final determination of one-to-one correspondences as shown in Fig. 2(a). Inspired by the discriminative power of sparse representation, we apply the l_1-norm on the assignment vector $\boldsymbol{x}$ to require the number of non-zero elements in $\boldsymbol{x}$ to be as small as possible. Since the affine constraint $\boldsymbol{Ax} = \boldsymbol{1}^T$ in Eq. 1 specifies each feature point to have at least one correspondence, the l_1-norm regularization term on the entire

vector $\boldsymbol{x}$ eventually leads to the sparsity on the possible matches for each feature point.

The advantage of using l_1-norm regularization $\|\boldsymbol{x}\|_1$ is demonstrated in Fig. 2(b). Compared with the assignment matrix obtained by SMAC without l_1 -norm constraint, the distribution of assignments along each row and each column of matrix $\boldsymbol{X}$ is much sharper by our method. Thus, it is easier to finally apply the Hungarian algorithm to binarize $\boldsymbol{X}$ and obtain the one-to-one correspondences. It is worth noting that both methods are performed on the same affinity matrix, in order to evaluate only the effectiveness of including l_1-norm regularization in correspondence detection.

New Energy Function for Graph Matching: Incorporating the two improvements described above, our energy function for graph matching is given as:

$$F(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} - \gamma \cdot \|\boldsymbol{x}\|_1 \quad s.t.\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{1}^{\boldsymbol{T}} \text{ and } \{0 \le x_m \le 1 | m = 1, \dots, nn'\} \quad (2)$$

Apparently, the first term is similar to SMAC method, except that the affinity matrix $\boldsymbol{M}$ is constructed by adding our newly-defined line patch to measure the pairwise agreement (i.e., off-diagonal elements in $\boldsymbol{M}$). The second term $\|\boldsymbol{x}\|_1$ is called as sparsity constraint term, with its strength being controlled by the parameter γ.

2.3 Optimization for the Improved Graph Matching

We can incorporate the affine constraint $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{1}^{\boldsymbol{T}}$ into the energy function in Eq. 2 as:

$$F'(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} - \gamma \cdot \|\boldsymbol{x}\|_1 - \lambda \cdot \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{1}^{\boldsymbol{T}}\|_2^2 \quad s.t.\ \{0 \le x_m \le 1 | m = 1, \dots, nn'\} \quad (3)$$

Since each element x_m in $\boldsymbol{x}$ is non-negative, we can simplify $\|\boldsymbol{x}\|_1$ as $\sum_{m=1}^{nn'} x_m = \boldsymbol{1}\boldsymbol{x}$. Then the energy function $F'(\boldsymbol{x})$ becomes the quadratic function of $\boldsymbol{x}$. Finally the maximization of $F'(\boldsymbol{x})$ falls into the constrained indefinite quadratic programming problem and can be efficiently solved by the trust region reflective algorithm [8].

3 Experiments

A publicly available USC hand atlas[1] is used for evaluation of our method. The resolution for each image is $0.1mm \times 0.1mm$ [9]. Thirty landmarks were manually placed for each of 43 left hand radiographs, randomly selected from the images of 11-year-old children, and these manual landmarks are used as ground-truth in this paper. Correspondence results are evaluated by the matching errors computed as the Euclidean distances between the automatically detected correspondences and the ground-truth.

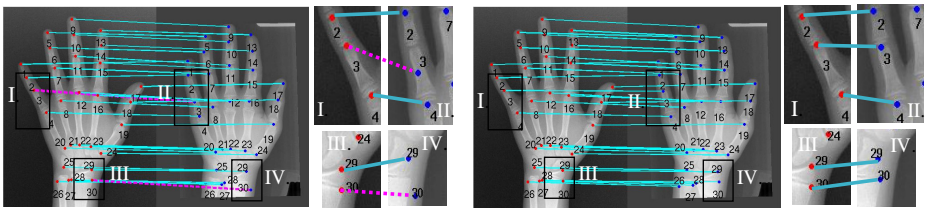
In order to demonstrate the advantages of line patch and sparsity constraint separately, we compare the following four methods: (1) SMAC, (2) SMAC with line patch, (3) our method without line patch, and (4) our full method (equipped with both

[1] http://www.ipilab.org/BAAweb/

line patch and sparsity constraint). For the four methods, the normalized cross correlation between local intensity patches are used to measure the pointwise similarities in establishing possible correspondences (i.e., diagonal elements in affinity matrix $\boldsymbol{M}$). The experiments are conducted by randomly selecting one image as the template image and the rest 42 images as the subject images. In each round of cross validation case, affine registration is performed for each subject image before detecting its correspondence with the selected template image. For the template image, 30 manually placed landmarks are used as its feature points, while, for each subject image, we follow the automatic landmark detection method in [10] to select around 450 feature points.

Typical correspondence matching results by SMAC and our full method are shown in Fig. 4, where correct matches are displayed by solid cyan lines and incorrect matches are displayed by dashed pink lines. By visual inspection, it can be concluded that our full method is able to correctly identify all 30 correspondences, while SMAC method failed at two landmarks (#3 and #30). For the better illustration, we also zoom in the regions enclosing the landmarks #3 and #30 (see black rectangle in the original images), and show them in the right of Fig. 4(a) and Fig. 4(b), respectively. For matching two images of size about 1500×2000, the average runtimes for SMAC and our full method are 451seconds and 537 seconds by a Matlab implementation.

Table 1 shows the mean and standard deviation of matching errors between the ground-truth and the estimated correspondences by the four methods, with respect to the two different templates. It can be seen that (1) our full method achieves the highest matching accuracy; and (2) each improvement strategy proposed in our method has significant effect in enhancing the performance of correspondence detection.

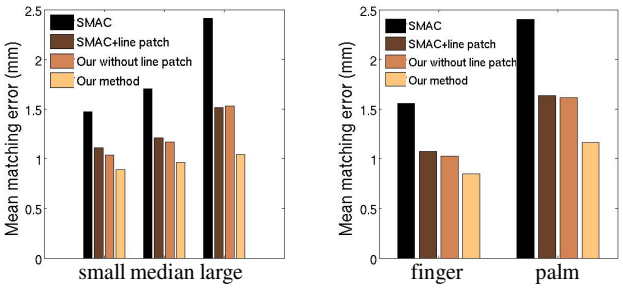


(a) SMAC (b) Our method

Fig. 4. Matching results by (a) SMAC and (b) our full method, with solid cyan lines showing correct matches and dashed pink lines showing incorrect matches. In the right of (a) and (b), regions I and III represent the enlarged views at landmarks 3 and 30 of the template image, and regions II and IV represent the corresponding enlarged views of the subject image.

Table 1. Mean and standard deviation of matching errors between the manual ground-truth and the estimated correspondences by the four methods (mm)

	SMAC	SMAC + line patch	Our method (without line patch)	Our method (line patch + sparsity)
Template 1	1.78 ± 2.54	1.33 ± 1.79	1.20 ± 1.50	0.98 ± 1.08
Template 2	2.12 ± 4.57	1.78 ± 3.96	1.36 ± 1.89	1.07 ± 1.28



(a) Different shape differences to template (b) Different parts

Fig. 5. Mean matching errors of four methods (a) under different amounts of shape difference from the selected template, and (b) at different parts of hand images

Furthermore, we demonstrate the robustness of the four methods under two different cases: (a) shape variation (such as difference between subject image and the template image), and (b) image contrast (such as in different parts of hand images). Specifically, for the first case, we classify all subject images into three groups (i.e., small, median, large) according to the total shape distance of their 30 manually labeled landmarks to the selected template image after affine alignment. Fig. 5(a) shows the mean distance errors by the four methods, which indicate that our full method achieves the lowest distance error. It is worth noting that, for the group with large shape difference, the matching error by our full method is almost 50% lower than that by SMAC method. This shows the great performance of our method in dealing with large inter-subject variations. For the second case, we separate landmarks into two different groups, i.e., located in the finger areas with high image contrast or in the palm areas with poor contrast. According to the results in Fig. 5(b), our method in difficult areas (i.e., palm areas) achieves even lower distance error than the SMAC method in the easy areas (i.e., finger areas).

4 Conclusion

In this paper, we have proposed a new graph matching method to improve the accuracy in establishing anatomical correspondences between two images. Our contributions have twofold: (1) a new concept of line patch is proposed to robustly characterize the pairwise agreement of two possible matches/correspondences; and (2) sparsity constraint is further introduced for the correspondence assignment, to suppress the influence from ambiguous matches. Promising results have been achieved by our method on the hand X-ray images, outperforming the state-of-the-art SMAC graph matching method. In the future, we will extend our method to other medical applications, e.g., deformable registration and motion correction for lung 4D-CT images by extending our method to deal with large number of feature points under the framework of hierarchical correspondence matching.

References

1. Chui, H., Rangarajan, A.: A New Point Matching Algorithm for Non-Rigid Registration. Comput. Vis. Image Underst. 89(2-3), 114–141 (2003)
2. Castillo, E., Castillo, R., Martinez, J., et al.: Four-Dimensional Deformable Image Registration Using Trajectory Modeling. Phys. Med. Biol. 55(1), 305–327 (2010)
3. Maciel, J., Costeira, J.P.: A Global Solution to Sparse Correspondence Problems. IEEE Trans. on Pattern Anal. Mach. Intell. 25(2), 187–199 (2003)
4. Cour, T., Srinivasan, P., Shi, J.: Balanced Graph Matching. In: Advances in Neural Information Processing Systems 19, pp. 313–320. MIT Press (2006)
5. Leordeanu, M., Hebert, M.: A Spectral Technique for Correspondence Problems Using Pairwise Constraints. In: International Conference of Computer Vision, vol. 2, pp. 1482–1489. IEEE Computer Society (2005)
6. Tibshirani, R.: Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society Series B 58(1), 267–288 (1996)
7. Wright, J., Yang, A.Y., Ganesh, A., et al.: Robust Face Recognition Via Sparse Representation. IEEE Trans. on Pattern Anal. Mach. Intell. 31(2), 210–227 (2009)
8. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2006)
9. Cao, F., Huang, H.K., Pietka, E., et al.: An Image Database for Digital Hand Atlas. In: Proceedings of SPIE Medical Imaging: PACS and Integrated Medical Information Systems: Design and Evaluation, vol. 5033, pp. 461–470 (2003)
10. Zhang, P., Cootes, T.: Automatic Construction of Parts+Geometry Models for Initialising Groupwise Registration. IEEE Transactions on Medical Imaging 31(2), 341–358 (2012)

Semi-supervised Segmentation Using Multiple Segmentation Hypotheses from a Single Atlas

Tobias Gass, Gábor Székely, and Orcun Goksel

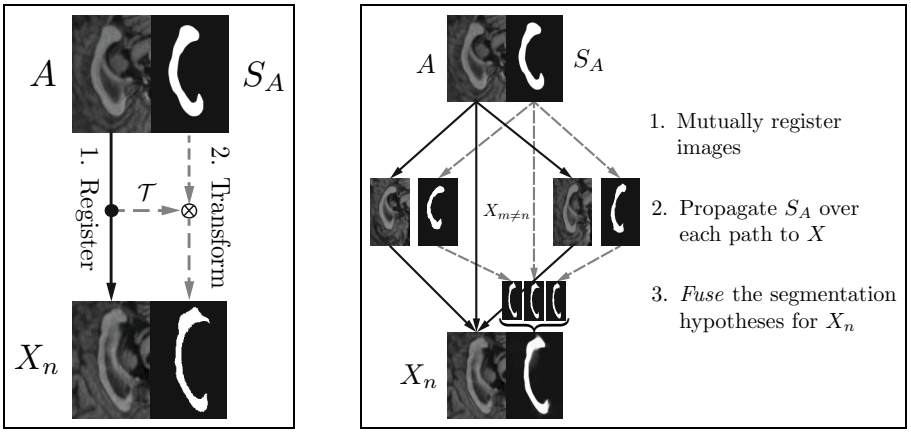
Computer Vision Lab, Dep. of Electrical Engineering, ETH Zurich, Switzerland
{gasst,gszekely,ogoeksel}@vision.ee.ethz.ch

Abstract. A semi-supervised segmentation method using a single atlas is presented in this paper. Traditional atlas-based segmentation suffers from either a strong bias towards the selected atlas or the need for manual effort to create multiple atlas images. Similar to semi-supervised learning in computer vision, we study a method which exploits information contained in a *set* of unlabelled images by mutually registering them non-rigidly and propagating the single atlas segmentation over multiple such registration paths to each target. These multiple segmentation hypotheses are then fused by local weighting based on registration similarity. Our results on two datasets of different anatomies and image modalities, corpus callosum MR and mandible CT images, show a significant improvement in segmentation accuracy compared to traditional single atlas based segmentation. We also show that the bias towards the selected atlas is minimized using our method. Additionally, we devise a method for the selection of intermediate targets used for propagation, in order to reduce the number of necessary inter-target registrations without loss of final segmentation accuracy.

1 Introduction

Image segmentation is an essential problem in medical image processing. Among automatic segmentation methods, the amount of prior information needed is a major distinguishing characteristic of different approaches. It is desirable to limit the constraints posed by prior knowledge both to retain generalizability and to reduce the effort required to acquire the information needed. Arguably, intensity-based approaches are among the methods that require the least amount of prior information. However, these methods are often susceptible to imaging artifacts, ambiguous intensities, and low contrast low signal-to-noise ratio imaging modalities, since no prior knowledge of shape or pose is assumed. Examples of such methods are active contours [1] and MRF-based segmentation [2, 3]. A straight-forward method for incorporating anatomical knowledge into automatic segmentation is the registration of an *atlas* image with known segmentation to a target image, namely *atlas based segmentation* [4, 5]. Then, the resulting transformation can be applied to the labelled atlas, yielding a segmentation of the target image. As the registration is ill-posed due to ambiguous and non-convex criteria [6], only approximate solutions can be achieved and these are influenced

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 29–37, 2013.



(a) Atlas based segmentation (b) The proposed method

Fig. 1. Illustration of standard atlas based segmentation (left) and our proposed method (right). In the latter, additional segmentation hypotheses are created by deforming the atlas segmentation multiple times along each path to each target. In a last step, these hypotheses are then fused to create the final segmentations.

strongly by the choice of the atlas image [7]. To remedy this, the use of multiple atlases is a common approach [7–12]: The information contained in such set of atlases can be used to create an average atlas [11]; to train statistical models of shape [8] or deformations [12]; or to register them individually to the target and fuse these segmentations afterward [13]. While all these methods improve segmentation accuracy compared to single atlas based segmentation, several studies indicate that the latter fusion of multiple deformed atlas segmentations is superior to registering an average atlas [7, 9].

In contrast to increasing the amount of manual annotation, harnessing information from unlabelled data has been a major research focus in computer vision. To this end, semi-supervised learning [14] is an established framework which enables tasks like image classification in large and diverse databases [15, 16]. For medical image segmentation, the situation is similar: large amounts of raw data are readily available while annotated data (atlases) are scarce.

In this paper, we study a method that propagates atlas segmentation labels via a graph of inter-target registrations. This is inspired by label propagation in semi-supervised learning [17] and similar to a method presented in [13], which used *indirect propagation* of atlas labels to validate multi-atlas segmentation results. In our framework, the traditional single atlas segmentation can be seen as direct, or *zero-hop* propagation. Multiple segmentation hypotheses per target can be obtained by allowing more *hops* via other target images, generating a different segmentation hypothesis for each path from the atlas to the target (c.f. Fig. 1b), which are then fused. In this paper, different strategies are studied for fusing such propagated labels for the segmentation of each target. Experiments with different anatomies and image modalities show significant improvement

over the traditional atlas-based segmentation. In order for the method to scale successfully to larger datasets, we also investigate methods to reduce the number of propagation connections in the graph, which in turn reduces the computational complexity by removing registrations that would otherwise be necessary.

2 Label Propagation

In this section, we present our method of generating and fusing multiple segmentation hypotheses for each target image X_n in a set of N unlabelled images using a single atlas A. An image is a function $\Omega \to \mathbb{R}$, where $\Omega \in \mathbb{N}^D$ is the discrete coordinate domain and D is the dimensionality of the image. We define a binary segmentation as $S_{(\cdot)} = \Omega \to \{0, 1\}$. A transformation is denoted with $\mathcal{T}$, e.g. $\mathcal{T}(A)$ is the transformed atlas A. While $\mathcal{T}$ can be an arbitrary transformation, we will assume non-rigid deformations represented by dense displacement fields throughout this paper. Let $\mathcal{T}_{\text{source,target}}$ denote a registration from X_m to X_n. We first mutually register all images $\{A, X_1, \ldots, X_N\}$, finding all possible combinations of $\mathcal{T}$, which are also connections in the graph. The traditional atlas based segmentation for a target X_n is then given by a *zero-hop* segmentation $S_n^0 = \mathcal{T}_{A,n}(S_A)$. Such segmentations can then be *propagated* to a target X_n over other target images $X_{m \neq n}$ as secondary (*one-hop*) segmentation hypotheses:

$$S_{m,n}^1 = \mathcal{T}_{m,n}(S_m^0). \tag{1}$$

These hypotheses must then be *fused*: We use a function F to first generate a spatial segmentation probability map $\hat{S}_n = \Omega \to [0, 1]$ and subsequently binarize this using thresholding. The said probability map is generated as a weighted average of the zero- and one-hop segmentation hypotheses:

$$\hat{S}_n = F\left(S_n^0, \{S_{m,n}^1 | m \neq n\}\right) = \frac{1}{\sum \lambda}\left(\lambda_n^0 S_n^0 + \sum_{m \neq n} \lambda_{m,n}^1 S_{m,n}^1\right). \tag{2}$$

We propose and evaluate two different strategies for the choice of weights λ:

Global Similarity Weighting (GSW): Assuming correlation between segmentation accuracy and a normalized post-registration similarity f_{G}, the latter can be used as a scalar weight. Using the zero-hop deformed atlas image $A_n^0 = \mathcal{T}_{A,n}(A)$, the zero-hop weight is then $\lambda_n^0 = f_{\text{G}}(X_n, A_n^0)$. The atlas image is propagated analogously to the atlas segmentation along each path, i.e. $A_{m,n}^1 = \mathcal{T}_{m,n}(A_m^0)$, which leads to one-hop weights as follows:

$$\lambda_{m,n}^1 = f_{\text{G}}\left(X_n, A_{m,n}^1\right). \tag{3}$$

Locally Adaptive Weighting (LAW): In contrast to the constant weights per hypothesis in GSW, a spatially-varying *local* weighting scheme is used:

$$\lambda_{m,n}^1(p) = f_{\text{L}}\left(X_n(p), A_{m,n}^1(p)\right), \forall p \in \Omega\ . \tag{4}$$

In contrast to GSW, such locally adaptive weighting is expected to leverage useful information even from partially mis-registered images.

3 Compact Graphs

To use all one-hop segmentation hypotheses, all graph connections should be computed requiring N^2 registrations. With large datasets this may easily become computationally challenging. Below, different methods are proposed for selecting a target image subset $\mathbb{X}$, called *support-samples*, that will act as intermediate nodes via which the atlas segmentation is propagated using (1). Reducing the size $K{=}|\mathbb{X}|$ then decrease the number of edges in $\mathcal{G}$ and hence the number of necessary inter-target registrations. A natural choice is to sort nodes by their GSW-based zero-hop weights λ_n^0, as this corresponds to image similarity to deformed atlas, and to use the highest ranked images as support samples. We call this GSW-based ranking. This scheme makes two assumptions: first, that it is utmost important to propagate 'good' zero-hop segmentations; and second, that the quality of these segmentations can be assessed reliably using the similarity function f. Using our label propagation framework, we propose the following two additional ranking criteria for selecting support samples.

In segmenting a target X_n, we wish to quantify how reliable a one-hop segmentation hypotesis $S^1_{m,n}$ via an intermediate node X_m is. As we only know the segmentation of the atlas, we define an *atlas reconstruction error* (ARE) for such quantification. Exploiting the fact that most non-rigid registration algorithms are not symmetric, we compute deformations $\mathcal{T}_{m,A}$ to obtain back-propagated one-hop atlas segmentation hypotheses $S^1_{m,A}{=}\mathcal{T}_{m,A}(S^0_m)$ via each graph node. Then, ARE for each node is defined based on Dice's similarity coefficient:

$$\text{ARE}\,(m) = 1 - \text{Dice}\left(S_A, S^1_{m,A}\right), \tag{5}$$

and support samples are selected from the smallest error nodes. While such ranking is expected to perform superior to GSW-based ranking, it cannot ensure that complementary information is contained in the set. For example, a subset $\mathbb{X}$ might have individually low AREs, however, their one-hop hypotheses may all contain similar errors which are then amplified when they are fused to create a target segmentation. It is thus desirable to find a *complementary* basis, where each support sample is likely to contain information that other samples do not provide. We therefore propose a *groupwise* error criterion (ARE-G) to score a *set* of K graph nodes:

$$\text{ARE-G}(S_A, \{m_1, \ldots, m_K\}) = 1 - \text{Dice}\left(S_A, F\left(S^1_{m_1,A}, \ldots, S^1_{m_K,A}\right)\right), \tag{6}$$

where F is the fusion function in (2). As it is not feasible to evaluate all $\binom{N}{K}$ K-sized sets of support samples, we rely on a greedy scheme where we pick the first support sample based on its individual ARE, and iteratively add support samples that reduce ARE-G the most.

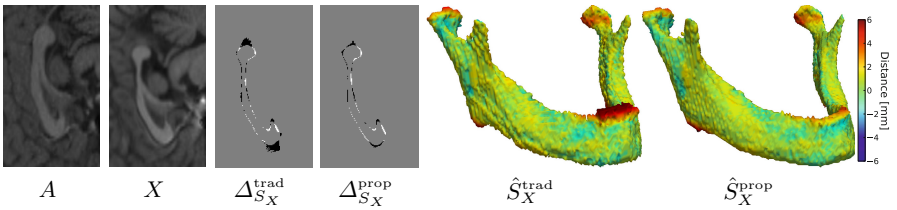


Fig. 2. Sample results for both datasets. For the MR dataset, the atlas, a target image and the difference of both the traditional single-atlas based segmentation $\Delta_{S_X}^{\text{trad}}$ and our proposed method $\Delta_{S_X}^{\text{prop}}$ from the ground-truth are shown. False positives and false negatives are shown in black and white, respectively. For the CT dataset, the traditional atlas based segmentation and our method are shown for the same atlas/target, and the surface is colored with the distance error to the ground-truth.

4 Results and Discussion

We evaluated our method using a set of 70 mid-saggital slices of MR brain scans containing the corpus callosum and a set of 15 3D CT scans of the head. Both datasets were rigidly pre-aligned. In the MR dataset a fixed region of interest containing the corpus callosum was cropped out in all images. The images are 120x200 pixels with 0.3 mm spacing in the MR dataset and 160x160x129 voxels with 1 mm spacing in the CT dataset. We used the Dice coefficient to measure volume overlap, the Hausdorff distance (HD) for estimating maximum surface-to-surface distance, and the mean surface distance (MSD) as an additional distance based metric. In both datasets we performed a leave-one-out (l1o) evaluation scheme, using each image as atlas in turn to segment all remaining images. We used our own implementation of the MRF-based registration in [4] as the registration method of choice in our experiments. We used normalized cross correlation (NCC) as the registration similarity criterion throughout our experiments, and accordingly defined GSW weights using $f_{\text{G}}=\frac{1-\text{NCC}}{2}$. Since NCC is not suitable as a point-wise metric for LAW, we used the following intensity difference based radial basis function $\lambda_{m,n}^1(p) = \exp\left(-\frac{|X_n(p)-A_{m,n}^1(p)|}{\sigma^2}\right)$ where σ^2 is the intensity variance over all images. Sample results are given in Fig. 2.

Quantitative results can be found in Tab. 1. For both datasets, the improvement over traditional atlas based segmentation is significant, especially considering the distance based metrics which are outperformed by $\approx 35\%$(HD) and $\approx 60\%$(MSD). The CT images also show a strong improvement in Dice similarity metric, and LAW expectedly performed superior to GSW. GSW only slightly outperforms an uniform weighting, which results in a simple max-voting scheme. We also compared our method to a recent group-wise registration approach (ABSORB [11]) a 3D implementation of which is publicly available. This method uses the atlas as a reference image, on which all test images are aligned iteratively to improve a mean image. Similarly to ours, this method also utilizes

Table 1. Mean segmentation accuracy from leave-one-out evaluation on the MR(2D) and CT(3D) datasets

	Corpus Callosum in MR			Mandibles in CT		
	DICE	HD	MSD	DICE	HD	MSD
single-atlas [4]	0.926	2.45	0.088	0.828	12.37	0.50
Propagation-maxvote	0.944	1.50	0.027	0.852	9.24	0.32
Propagation-GSW	0.944	1.49	0.027	0.859	9.32	0.27
Propagation-LAW	**0.946**	**1.46**	**0.025**	**0.886**	**8.42**	**0.18**
ABSORB [11]	-	-	-	0.698	14.74	0.86
multi-atlas-LAW	0.965	1.06	0.014	0.917	5.77	0.13

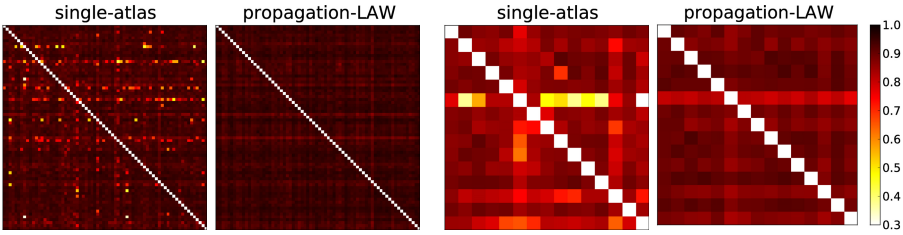


Fig. 3. Segmentation accuracy (Dice) of traditional single atlas based segmentation and our proposed method with LAW fusion for each atlas (x-axis) and target (y-axis) image combination. (The diagonal values were not computed.)

image information contained in the entire dataset. However, unlike the probablistic map in our method, ABSORB generates a single binary segmentation per target and thus suffers from a bias towards the atlas similarly to the traditional atlas based segmentation. This is seen in the results as it is outperformed by our proposed method. In order to estimate an upper bound for the expected performance of our method, we also computed multi-atlas segmentation of each target by using all remaining images and their ground-truth segmentations as multiple atlases. We used the same registration method and LAW weighting to achieve comparable results. As seen in Tab. 1, even though our method did not reach the performance of such multi-atlas segmentation, it performed remarkably close to it while using orders of magnitude less prior knowledge.

In Fig. 3, the Dice measure is plotted for each atlas/target pair of our l1o-experiments. For both datasets, it is seen that using traditional atlas based segmentation some atlases lead to sub-optimal segmentations. Using our method, however, these sub-optimal pairs seen as 'speckles' disappear, indicating an improved performance for arbitrary atlas selection.

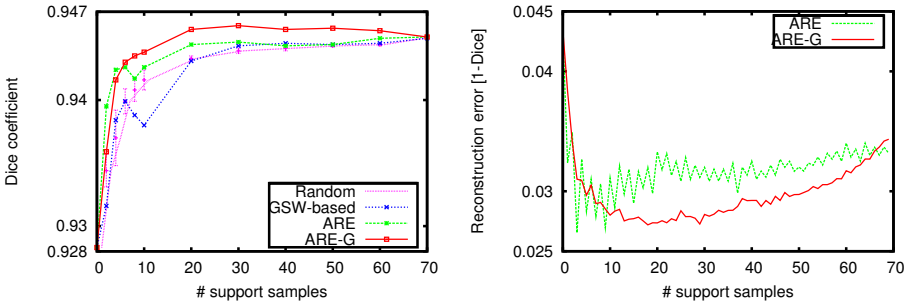


Fig. 4. Dice metric for support sample selection for a single atlas using random selection, GSW-based, individual ARE, and ARE-G ranking criteria (left). The progression of error as more support samples are added.

Support Sample Reduction: We first evaluated a random selection of K support samples for each target and repeated this 10 times each. As shown in Fig. 4(left), Dice metric improves rapidly until reaching 20 support samples, which is also the point at which the standard deviation becomes negligible. We then used the GSW-based ranking, which did not provide any improvement over random draws. Individual ARE based ranking expectedly outperformed random draws, and the group-wise ARE ranking was superior to all other support sample selection methods. Interestingly, the results indicate the presense of support sample subsets that can perform *better* in comparison to using all the samples (the full graph). We also analyzed the progression of error during the expansion of the support-sample subset. As seen in Fig. 4(right), error using ARE-G based ranking starts deteriorating beyond a certain number of subset size. This number is also near the optimal Dice metric shown in Fig. 4(left) as the posterior target segmentation accuracy. Based on this observation, we propose to use the deflecting point (minimum value) of ARE-G to determine optimal graph size. Note that individual ARE does not exhibit such a behaviour. Accordingly, we have repeated the l1o-experiments on the MR dataset using only 20 support samples selected by ARE-G ranking. This yielded mean Dice of 94 % and mean HD of 1.56 mm, which are nearly identical to the results using all target images while requiring less than three times the number of intermediate target images and their registrations.

Discussion: Our method was shown to increase segmentation accuracy substantially compared to the standard atlas based segmentation. This can be attributed to the boosting nature of the approach, which can be seen as creating and fusing multiple weak classifiers in a semi-supervised manner. Our results being close to that of multi-atlas segmentation, we conclude that a substantial amount of the information contained in a set of *atlas* images is indeed available in the support (*target*) images, and that is the information leveraged by our method. Note that this contradicts partially with the findings of [13], which concludes that the major benefit of multi-atlas segmentation is due to the increase in anatomical

variation in the available ground-truth. We believe that this difference in findings might be due to different registration algorithms and datasets, which will be important to explore in the future work. Additionally, it will be interesting to explore whether a principled, probabilistic approach can be employed to also take into account the improved segmentations in an iterative manner. A similar method was proposed in [18] for images aligned to a single template, whereas we aim to include information from inter-target registrations as well. For the CT images, we used the probabilistic segmentation output as shape prior in an MRF-based segmentation [2]. This led to a considerable improvement in results with a mean Dice of 93 %. As this relies on strong edges to find bone boundaries, it does not yield a significant improvement in the MR images.

5 Conclusions

We presented a novel method that augments single-atlas based segmentation using multiple segmentation hypotheses for each target obtained by propagating atlas segmentation along different paths. This outperforms both the traditional single-atlas based registration and group-wise registration. We also demonstrated that a smaller set of support samples providing complementary information can be found automatically. Using such reduced set of support samples both decreases the computational complexity of the method and improves the results as redundant and possibly detrimental information is then discarded.

Acknowledgements. This work has been funded by the Swiss National Center of Competence in Research on Computer Aided and Image Guided Medical Interventions (NCCR Co-Me) supported by the Swiss National Science Foundation.

References

1. Kass, M., Witkin, A.: Snakes: Active contour models. International Journal of Computer Vision 331, 321–331 (1988)
2. Furnstahl, P., Fuchs, T., Schweizer, A., Nagy, L., Székely, G., Harders, M.: Automatic and robust forearm segmentation using graph cuts. In: ISBI, pp. 77–80 (2008)
3. Boykov, Y., Funke-Lea, G.: Graph Cuts and Efficient N-D Image Segmentation. International Journal of Computer Vision 70(2), 109–131 (2006)
4. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Medical Image Analysis 12(6), 731–741 (2008)
5. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic Registration Using B-Splines. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 702–709. Springer, Heidelberg (2006)
6. Fischer, B., Modersitzki, J.: Ill-posed medicine - an introduction to image registration. Inverse Problems 24(3), 1–19 (2008)
7. Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., Maurer, C.: Quo vadis, atlas-based segmentation? In: Handbook of Biomedical Image Analysis, pp. 435–486 (2005)

8. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis 13(4), 543–563 (2009)
9. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion–application to cardiac and aortic segmentation in CT scans. IEEE T. Med. Imaging 28(7), 1000–1010 (2009)
10. van Rikxoort, E.M., Isgum, I., Arzhaeva, Y., Staring, M., Klein, S., Viergever, M.A., Pluim, J.P.W., van Ginneken, B.: Adaptive local multi-atlas segmentation: application to the heart and the caudate nucleus. Medical Image Analysis 14(1), 39–49 (2010)
11. Jia, H., Wu, G., Wang, Q., Shen, D.: NeuroImage ABSORB: Atlas building by self-organized registration and bundling. NeuroImage 51(3), 1057–1070 (2010)
12. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic Construction of 3D Statistical Deformation Models Using Non-rigid Registration. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 77–84. Springer, Heidelberg (2001)
13. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage 33(1), 115–126 (2006)
14. Chapelle, O., Schölkopf, B., Zien, A. (eds.): Semi-supervised learning, vol. 2. MIT Press, Cambridge (2006)
15. Wang, F., Wang, J., Zhang, C., Shen, H.C., Bay, C.W., Kong, H.: Semi-Supervised Classification Using Linear Neighborhood Propagation. Sci. & Tech. (2006)
16. Guillaumin, M., Verbeek, J., Schmid, C.: Multimodal semi-supervised learning for image classification. In: IEEE CVPR, pp. 902–909 (June 2010)
17. Zhu, X., Ghahramani, Z.: Learning from labeled and unlabeled data with label propagation. Technical report, School Comput. Sci. Carnegie Mellon Univ. (2002)
18. Riklin-Raviv, T., Van Leemput, K., Menze, B.H., Wells, W.M., Golland, P.: Segmentation of image ensembles via latent atlases. Medical Image Analysis 14(5), 654–665 (2010)

Carotid Artery Wall Segmentation by Coupled Surface Graph Cuts

Andres Arias[1], Jens Petersen[2], Arna van Engelen[1], Hui Tang[1,3], Mariana Selwaness[4,5,6], Jacqueline C.M. Witteman[5], Aad van der Lugt[4], Wiro Niessen[1,3], and Marleen de Bruijne[1,2]

[1] Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
[2] Image Group, Department of Computer Science, University of Copenhagen, Denmark
[3] Faculty of Applied Sciences, Department of Imaging Science and Technology, Delft University of Technology, The Netherlands
[4] Department of Radiology, Erasmus MC, Rotterdam, The Netherlands
[5] Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
[6] Department of Biomedical Engineering, Erasmus MC, Rotterdam, The Netherlands

Abstract. We present a three-dimensional coupled surface graph cut algorithm for carotid wall segmentation from Magnetic Resonance Imaging (MRI). Using cost functions that highlight both inner and outer vessel wall borders, the method combines the search for both borders into a single graph cut optimization. Our approach requires little user interaction and can robustly segment the carotid artery bifurcation. Experiments on 32 carotid arteries from 16 patients show good agreement between manual segmentation performed by an expert and our method. The mean relative area of overlap is more than 85% for both lumen and outer vessel wall. In addition, differences in measured wall thickness with respect to the manual annotations were smaller than the in-plane pixel size.

Keywords: Carotid artery, flow lines, graph, segmentation.

1 Introduction

Atherosclerosis is one of the primary causes of death in the world [11]. Atherosclerotic plaques in the carotid arteries cause lumen narrowing. This may lead to plaque rupture, which can cause a stroke or Transient Ischemic Attack (TIA). Therefore, the early detection of plaque and accurate quantification of lumen narrowing and plaque volume are important. In order to determine these parameters, segmentation of both the vessel lumen and the outer vessel wall are required. As manual segmentation is highly time consuming and subject to observer variability, automated techniques are needed.

Although most work on automated segmentation of blood vessels has focused on segmenting the vessel lumen only, several automatic and semi-automatic methods have been proposed in the past for segmenting the outer vessel wall.

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 38–47, 2013.

Active Shape Models (ASMs) have been used for detecting the outer vessel wall of the abdominal aorta in CTA scans [3]. These ASMs utilize a statistical model of shape and boundary grey level appearance to restrict the search space to anatomically reasonable solutions. To segment the carotid arteries in MRI, gradient-based ellipse fitting combined with fuzzy clustering [1] and Closed Contour Snakes (CCS) [14] have been proposed. A drawback of both these methods is that user interaction is required for each image slice. More recently, van 't Klooster et al. [8] proposed a three-dimensional (3D) deformable vessel model, in which a vessel is modeled using a 3D cylindrical surface that can be modified by moving control points located on the model surface. Good results were achieved on Black-Blood MRI images of the carotids. This method can however segment only a single, non-bifurcating vessel and will therefore not give reliable results in the bifurcation region. Furthermore, it uses a local optimization procedure with the lumen segmentation as initialization, which may get stuck in a local optimum in diseased vessels where the distance between the inner and outer wall can be large. Better segmentation results may be achieved if both walls are estimated jointly across the bifurcation and if local image information is combined with a globally optimal solution.

Global optimality can be guaranteed with graph based methods, and recently these have been used for vessel segmentation with promising results [9,5,13]. Surface based graph methods such as [9,10,13] as opposed to voxel based [5] make it possible to enforce topology constraints as well to encourage smoothness without biasing the solution towards smaller surfaces. To use these methods the problem has to be transformed from image space to a discretized graph space defined by a set of columns. Each column is associated with a point on the sought surface and represents the set of possible positions it can take. The suitability of the graph space depends on how well the graph columns cross the sought surface [10]. Xu et al. [13] oriented the graph columns in the normal direction of the centerline of the vessels, but this leads to long columns and thus in-efficiency if the sought surface is far from the centerline. Moreover, straight columns intersect in regions with curvature leading to possibly self-intersecting surfaces [10].

We propose to use a 3D coupled surface graph cut algorithm for carotid artery wall segmentation from MRI images. Similar to Petersen et al. [10] who applied such a technique for segmenting airway trees, we define the graph columns based on flow lines traced from a coarse initial segmentation. As such flow lines are non-intersecting this enables accurate segmentation across high curvature areas such as the carotid bifurcation. Moreover, as the inner and outer surfaces are estimated jointly, the proposed method can use information from both surfaces locally and globally to reach an optimal solution.

2 Method

2.1 Initial Segmentation

An initial segmentation of the lumen was obtained using the method proposed by Tang et al. [12]. In this method first the lumen centerlines are determined

by finding a minimum cost path between three user-defined seed points in the common, internal, and external carotid arteries. To improve accuracy in high curvature regions, this path is refined iteratively by computing a new minimum cost path in a curved multi-planar reformatting based on the current center line estimate. Subsequently, the lumen is segmented using a levelset method, which is initialized by the extracted centerlines and steered by the MR intensities.

2.2 Graph Construction

First, to obtain the graph columns the initial segmentation is converted to a mesh. We located graph vertices at the center of each surface face. This set of vertices is denoted by V_B.

Flow Lines. The graph columns are traced from V_B, and follow the direction of flow lines of the gradient vector field of the smoothed segmentation. An example of columns traced along flow lines is depicted in figure 1. If this gradient vector field is defined in terms of a scalar potential field ϕ, the flow lines will follow the direction of largest change of this potential. We define the scalar field ϕ by the convolution of the initial segmentation with a Gaussian kernel G_σ as:

$$\phi(\boldsymbol{x}) = \int Q\left(\hat{\boldsymbol{x}}\right) G_\sigma\left(\hat{\boldsymbol{x}} - \boldsymbol{x}\right) d\hat{\boldsymbol{x}}, \tag{1}$$

where $Q\colon \mathbb{R}^3 \to \mathbb{Z}$ is the initial lumen segmentation represented by a binary scalar field. Flow lines traced along the gradient of ϕ are smooth and non-intersecting and the surfaces are thus non-self-intersecting [10], see figure 1.

The parametric flow lines $\boldsymbol{f}\colon \mathbb{R} \to \mathbb{R}^3$ that cross each vertex of the initial surface mesh $\boldsymbol{i}_0 \in V_B$ can be computed by solving the following differential equation:

$$\frac{\partial \boldsymbol{f}}{\partial t}(t) = \nabla\phi(\boldsymbol{f}(t)), \tag{2}$$

with initial value given by $\boldsymbol{f}(0) = \boldsymbol{i}_0$. Solving equation (2) for all vertices on the initial surface mesh V_B leads to all graph columns, where inner and outer graph columns are represented by the same flow lines. We use the Runge-Kutta-Fehlberg method to approximate the solution of these differential equations [4]. The solution of $\boldsymbol{f}(t)$ is approximated at regular intervals δ along the flow line. This defines the positions of the other graph vertices. The columns vary in length depending on the point where the gradient of the scalar field ϕ flattens.

Graph Construction and Optimization. To construct the coupled surface graph $G = (V, E)$ with vertices V and edges E, we define the set of vertices in a column by $V_{\boldsymbol{i}}$ with $\boldsymbol{i} \in V_B$. Therefore, the complete set of vertices V is defined by:

$$V = \bigcup_{\boldsymbol{i} \in V_B, m \in M} V_{\boldsymbol{i}}^m \cup \{s, t\}. \tag{3}$$

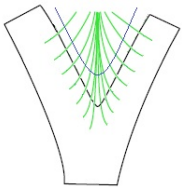


Fig. 1. Graph columns based on flow lines (green) traced from an initial segmentation (black), which are crossing the sought surface (blue)

Here M represents the surfaces to find and s and t denote the source and sink vertices respectively. In our case there are two surfaces, lumen and outer vessel wall surface. Moreover, given that the inner and outer columns are the same, we have $\forall_{m \in M} V_{\boldsymbol{i}}^m = V_{\boldsymbol{i}}$.

The edge set E of the coupled surface graph G consists of intra-column edges E_{intra} and edges between columns E_{inter}. For the intra-column edges E_{intra}, we define directed edges connecting each vertex to the next vertex in outward direction in the same column. We assign edges from the source vertex s to all innermost vertices in the graph, and from the outermost vertices to the sink vertex t. Topology preserving edges in the opposite direction with infinite capacity ensure that a minimum cut can cut each column only once. In addition, we assign a cost function $w^m\,(i_k^m) > 0$ to these edges, mapping a vertex with index k in column $V_{\boldsymbol{i}}$ to the inverse likelihood representation that it is part of surface m. An example of the intra-column edges and their respective costs is shown in figure 2(a) (for simplicity we do not show the infinity capacity edges).

Selecting a vertex for each column indicates a possible solution for all M surfaces. Therefore, a cut that separates the graph in two parts: sink and source, represents a solution to the segmentation problem. The main aim is then to find a cut that minimizes the cost of the edges that are being cut as depicted in figure 2(b). There are several approaches to solve this optimization problem. We used a min-cut/max-flow algorithm described in [2] to find the minimum cut.

Computing the minimum cut without considering any interaction between columns may lead to irregular surfaces and/or un-realistic relations between surfaces such as borders that are too far from each other, or an outer surface that is inside the inner surface. In order to deal with these problems, we include smoothing penalty edges connecting vertices in columns belonging to the same surface, separation penalty edges and separation constraint edges that connect vertices from columns of different surfaces. These represent the edges between columns E_{inter}.

To ensure smooth surfaces, we linearly penalize the distance in a cut between consecutive columns of the same surface. To do this, we assign edges with the same capacity p between vertices at the same column level. When the length of two consecutive columns are different, the remaining vertices at the inner most part of the column are connected to the source vertex, and the remaining vertices at the outer most part of the column are connected to the sink. If these edges

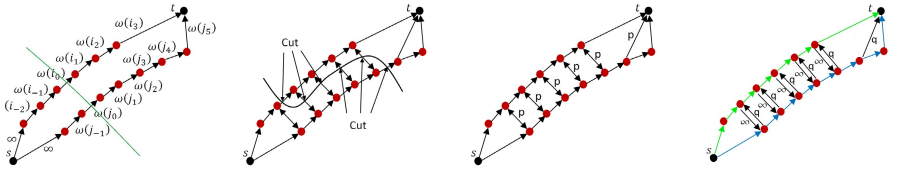


Fig. 2. Examples of intra and inter column edges and a graph cut. In figure 2(a) the intra-column edges, the initial segmentation (green), and the associated cost to each edge are depicted. An example of a graph cut is depicted in figure 2(b), indicating which edges are part of the cut. Figure 2(c) shows the smooth penalty edges which connect vertices from neighbor columns of the same surface. Finally, the separation penalty edges and separation constraint edges are depicted in figure 2(d). These edges connect vertices from columns of different surfaces lying at the same flow line (green: inner column, blue: outer column).

coincide with the intra-column edges, only one edge is assigned and the capacities are added. An example of these smoothing penalty edges is shown in figure 2(c). Using these edges we obtain a linear penalty function of the form $\psi(x) = px$, where x represents the vertex index difference. In a similar way, the separation between surfaces is penalized by assigning capacity q to edges between vertices of columns lying at the same flow line but belonging to different surfaces. In addition, to avoid solutions where parts of the outer surfaces are inside the inner surface, we assign constraint edges with infinite capacity at the same location of the separation edges but pointing from the inner column to the outer column. An example of these separation penalty edges and separation constraint edges is shown in figure 2(d).

2.3 Cost Functions

For the intra-column edges, we define a cost function $w^m\left(i_k^m\right) > 0$, which represents the inverse likelihood that the vertex i_k^m is associated to the edge $i_k^m \rightarrow i_{k+1}^m$ is part of surface m. In the case of the carotid walls in our MR images, the graph columns will start inside the lumen area, which looks dark in the image, move through the carotid wall where the voxels are normally brighter, and finally end up out of the carotid wall where the image is darker compared to the wall intensity. We therefore define a cost function for the inner wall w^i which is low for strong dark-to-bright edges, and a cost function for the outer wall w^o which is low for strong bright-to-dark edges. We use a similar approach to Petersen et al. [10] to define the cost functions. First, we define the functions $C^i : \mathbb{R} \rightarrow \mathbb{R}$ and $C^o : \mathbb{R} \rightarrow \mathbb{R}$ that highlight the inner and outer walls. These use a linear combination of the first and second order derivatives of the intensity along the columns:

$$C^i(t) = \gamma^i \frac{\partial P}{\partial t}(t) + \left(1 - \left|\gamma^i\right|\right) P(t), \tag{4}$$

$$C^o(t) = \gamma^o \frac{\partial N}{\partial t}(t) + \left(1 - \left|\gamma^o\right|\right) (-N(t)), \tag{5}$$

where $\gamma^i, \gamma^o \in [-1, 1]$ are weighting parameters that can be tuned to adjust the position of the edge slightly inwards or outwards, and P and N the positive and negative parts of the first order derivative respectively. These derivatives are computed using central differences from cubic interpolated values. Subsequently, we invert and normalize C^i and C^o in order to get a representation of the wall inverse likelihood given by w^i and w^o.

3 Experiments and Results

3.1 Data

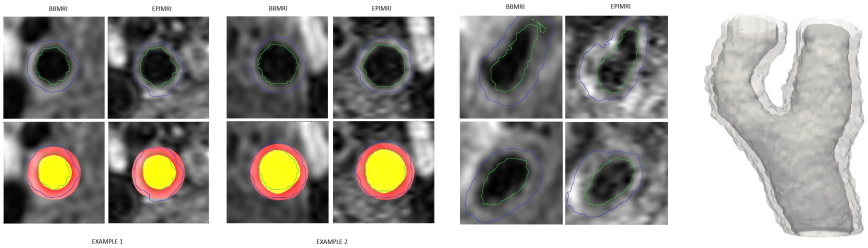
Proton Density Weighted Black-Blood MRI (BBMRI) and Proton Density Weighted Echo Planar MRI (EPIMRI) images were obtained from 26 subjects that were randomly selected from the Rotterdam study [6]. BBMRI images were acquired using an in-plane pixel size of 1.105mm × 0.8125mm, and 0.9 mm slice thickness. The EPIMRI images have an in-plane pixel size of 0.43mm × 0.8125mm, and a slice thickness of 1.2 mm. BBMRI and EPIMRI images were interpolated on the scanner to a pixel size of 0.507mm × 0.507mm. B-spline registration from EPIMRI to BBMRI using mutual information was performed using with Elastix [7]. To train and evaluate our method, we used manually annotated cross-sectional images with a resolution of 0.05mm×0.05mm extracted at random positions perpendicular to center-lines of both carotid arteries. The manual annotations of the inner and outer carotid walls were drawn by an expert on the BBMRI images. Six manually annotated cross sections were extracted from each carotid artery.

3.2 Graph Parameters Tuning

The proposed method has several parameters: inner and outer smoothness penalties p^i and p^o, separation penalties q, inner and outer cost function derivative weightings γ^i and γ^o, the intervals for sampling the flow lines to define the positions of the vertices δ, and the standard deviation of the Gaussian kernel σ. We used the carotid arteries of ten patients randomly selected to search for the optimal values for these parameters on each image sequence (BBMRI and EPIMRI). The optimal values were obtained by searching the parameter space on the training data-set using an iterative binary search algorithm [10]. In this algorithm, manually annotated cross-sections and automatically segmented cross-sections are compared based on the relative area of overlap. The set of parameters that generated the highest overlap was selected. To reduce the searching time of the parameter optimization algorithm, we fixed the column sampling interval δ to 0.35 mm.

Table 1. Relative area of overlap, WTD, LAD, and OVAD for both sequences (mean absolute and P-values in parentheses)

	BBMRI	EPIMRI
φ^i	$85.2\% \pm 1.6\%$	$83.9\% \pm 5\%$
φ^o	$85.6\% \pm 2.7\%$	$84.1\% \pm 6\%$
WTD(mm)	$-0.25 \pm 0.24(0.28 \pm 0.19; p < 0.001)$	$0.04 \pm 0.23(0.17 \pm 0.15; p = 0.5)$
LAD(mm^2)	$2.7 \pm 2.1(3.0 \pm 1.7; p < 0.001)$	$-0.07 \pm 3.29(2.3 \pm 2.21; p = 0.9)$
OVAD(mm^2)	$-0.3 \pm 1.1(0.92 \pm 0.8; p = 0.3)$	$-0.3 \pm 1.05(0.76 \pm 0.75; p = 0.25)$



(a) (b) (c)

Fig. 3. Automatic segmentation results using the proposed method. In figure 3(a), two example of automated segmented cross-section in BBMRI and EPIMRI are depicted (top). The automatic segmentation is represented by green (inner wall) and blue (outer wall) lines. The overlay of the automatic segmentation to the manual annotations (yellow: lumen, red: vessel wall) is also depicted in figure 3(a) (bottom). Figure 3(b) shows two examples of automatic segmentations obtained in the bifurcation section. Finally, figure 3(c) shows a 3D representation of the automatic segmentation of the complete carotid artery in an image (darker gray: lumen, bright gray: outer wall).

3.3 Segmentation Results

Thirty two carotid arteries of 16 patients not included in the training set were used for the evaluation. Table 1 gives the average relative area of overlap (Dice coefficient) for inner φ^i and outer vessel surface φ^o on this testing data set. In addition, table 1 describes the mean signed and mean absolute difference between wall thickness (WTD) measured by the manual annotation and by the automatic segmentations in BBMRI and EPIMRI. Notice that these values are smaller than the image in-plane pixel size (0.51 mm). We observed good segmentation overlap for both sequences with a slightly higher overlap for BBMRI images. Using EPIMRI images we obtained lower WTD. The table shows also the mean cross-sectional lumen area difference (LAD), and mean cross-sectional outer vessel area difference (OVAD). P-values of the paired t-test including 95% of confidence intervals are also given in the table. Figure 3(a) shows examples of the automatic segmentation results using BBMRI and EPIMRI images together with the overlay to the manual annotations. Results in the bifurcation section, for which no manual annotations were available, are depicted in figure 3(b). Figure 3(c) shows a 3D representation of the automatic segmentation.

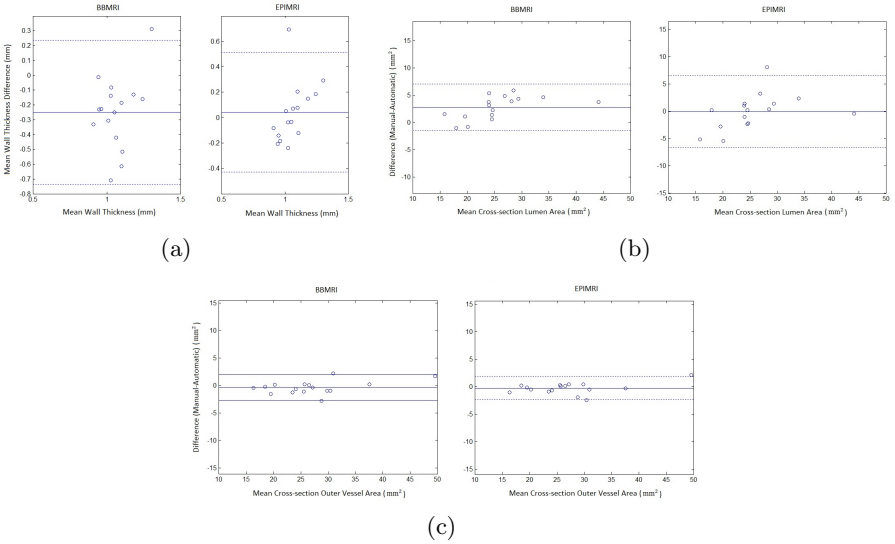


Fig. 4. Bland-Altman plots comparing manual annotations and automatic segmentation for both sequences BBMRI and EPIMRI. Figure 4(a) depicts the comparison of the mean wall thickness. Figure 4(b) and figure 4(c) show a comparison of the mean lumen area and outer vessel area respectively.

Bland-Altman analyses for the mean wall thickness, mean cross-section lumen area, and mean cross-section outer vessel area for the 16 patient data sets are shown in figure 4. From the figure a good agreement between automatic and manual area measurements for lumen and outer vessel wall is observed. Pearson correlation coefficients were 0.95 and 0.98 respectively for BBMRI, and 0.87 and 0.99 for EPIMRI.

4 Discussion and Conclusion

In this paper, we presented a new 3D method for carotid wall segmentation in MRI. Results show a good agreement between manual segmentation performed by an expert and our method. The mean relative area of overlap was about 84% and 85% for EPIMRI and BBMRI respectively. Our results are comparable to or slightly better than those reported in the literature. Van 't Klooster et al. [8] reported a WTD of $0.12\text{mm} \pm 0.21\text{mm}$. Their method only analyzes the common carotid artery and not the bifurcation. This section may represent the most difficult section to segment. In contrast, we analyze the complete carotid artery. We found a somewhat lower mean WTD with a similar variance ($0.04\text{mm} \pm 0.23\text{mm}$) using the EPIMRI sequence.

Adame et al. [1] use similar in-plane resolution images of 17 patients. They reported a LAD of $-2.19\text{mm}^2 \pm 5.21\text{mm}^2$ and an OVAD of $-5.56\text{mm}^2 \pm 19.55\text{mm}^2$

with correlation coefficients of 0.92 and 0.91 respectively. Yuan et al. [14] reported a LAD of $1.05\text{mm}^2 \pm 2.26\text{mm}^2$ and an OVAD of $1.36\text{mm}^2 \pm 3.46\text{mm}^2$ on five patients, and focus on the internal carotid artery. We reported in general better results on our data compared to these two methods (see table 1). Furthermore, these two methods require a large amount of user interaction. In contrast, our method only requires the location of three seed points for obtaining the initial segmentation.

A potential drawback of our approach is that it relies on an initial lumen segmentation. Although this segmentation does not need to be very accurate, the smoothness constraints are most effective if the shape of the initial segmentation is similar to shape of the true vessel surfaces. Another potential source of errors in our method is related to registration errors of the EPIMRI images. Overall, segmentation results were best for the sequence in which the manual annotations were performed, the BBMRI. However, the EPIMRI images have better wall contrast, which generates better results in some images compared to the results obtained by BBMRI. Therefore, we expect that combining information from both image types in the cost function can still improve upon the results presented here.

To conclude, we propose a graph-based method for segmenting the carotid artery wall that shows good agreement with manual segmentations. In contrast to previous approaches, our method jointly optimizes both surfaces, finds a globally optimal solution, and can reliably segment the bifurcation section which may represent the most clinically relevant area to assess.

References

1. Adame, I.M., van der Geest, R.J., Wasserman, B.A., Mohamed, M.A., Reiber, J.H.C., Lelieveldt, B.P.F.: Automatic segmentation and plaque characterization in atherosclerotic carotid artery MR images. Magnetic Resonance Materials in Physics, Biology and Medicine 16, 227–234 (2004)
2. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(9), 1124–1137 (2004)
3. de Bruijne, M., van Ginneken, B., Viergever, M.A., Niessen, W.J.: Adapting Active Shape Models for 3D Segmentation of Tubular Structures in Medical Images. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 136–147. Springer, Heidelberg (2003)
4. Butcher, J.: Numerical Methods for Ordinary Differential Equations. Wiley (2008)
5. Freiman, M., Frank, J., Weizman, L., Nammer, E., Shilon, O., Joskowicz, L., Sosna, J.: Nearly automatic vessels segmentation using graph-based energy minimization. The MIDAS Journal (2009)
6. Hofman, A., van Duijn, C., Franco, O., Ikram, M., Janssen, H., Klaver, C., Kuipers, E., Nijsten, T., Stricker, B., Tiemeier, H., Uitterlinden, A., Vernooij, M., Witteman, J.: The rotterdam study: 2012 objectives and design update. European Journal of Epidemiology, 1–30 (August 2011)
7. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging 29(1), 196–205 (2010)

8. van't Klooster, R., de Koning, P.J., Dehnavi, R.A., Tamsma, J.T., de Roos, A., Reiber, J.H., van der Geest, R.J.: Automatic lumen and outer wall segmentation of the carotid artery using deformable three-dimensional models in MR angiography and vessel wall images. Journal of Magnetic Resonance Imaging 35(1), 156–165 (2012)
9. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images – a graph-theoretic approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(1), 119–134 (2006)
10. Petersen, J., Nielsen, M., Lo, P., Saghir, Z., Dirksen, A., de Bruijne, M.: Optimal Graph Based Segmentation Using Flow Lines with Application to Airway Wall Segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 49–60. Springer, Heidelberg (2011)
11. Ross, R.: Atherosclerosis — an inflammatory disease. New England Journal of Medicine 340(2), 115–126 (1999)
12. Tang, H., van Walsum, T., van Onkelen, R.S., Hameeteman, K., Klein, S., Schaap, M., Bouwhuijsen, Q.J.B., Witteman, J., van der Lugt, A., van Vliet, L.J., Niessen, W.: Semiautomatic carotid lumen segmentation for quantification of lumen geometry in multispectral MRI. Medical Image Analysis (May 2012)
13. Xu, X., Niemeijer, M., Song, Q., Garvin, M., Reinhardt, J., Abramoff, M.: Retinal vessel width measurements based on a graph-theoretic method. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 641–644 (April 2011)
14. Yuan, C., Lin, E., Millard, J., Hwang, J.: Closed contour edge detection of blood vessel lumen and outer wall boundaries in black-blood MR images. Magnetic Resonance Imaging 17(2), 257–266 (1999)

Graph Cut Segmentation Using a Constrained Statistical Model with Non-linear and Sparse Shape Optimization

Tahir Majeed, Ketut Fundana, Silja Kiriyanthan,
Jörg Beinemann, and Philippe Cattin

Medical Image Analysis Center (MIAC), University of Basel, Switzerland
{tahir.majeed,ketut.fundana,philippe.cattin}@unibas.ch

Abstract. This paper proposes a novel segmentation method combining shape knowledge obtained from a constrained Statistical Model (SM) into the well known Markov Random Field (MRF) segmentation framework. The employed SM based on Probabilistic Principal Component Analysis (PPCA) allows to compute local information about the remaining variance *i.e.* uncertainty about the correct segmentation boundary. This knowledge about the local segmentation uncertainty is then used to construct a prior with a non-linear shape update mechanism, where a high cost is incurred in locations with little uncertainty and a low cost for shifting the segmentation boundary in locations with high uncertainty.

Experimental results for segmenting the masseter muscle from CT data are presented showing the advantage of including the knowledge about local segmentation uncertainties into the segmentation framework.

Keywords: Graph-Cut, MRF, Statistical Model, Shape Prior, PCA, Segmentaion, Facial, Muscles, Medical Image.

1 Introduction

Most of human's sense organs are located in the head and face area, which makes it one of the most important parts of the human body. The shape and the unique features of a face are largely determined by the musculoskeletal system underneath the facial skin. The importance of the face for socio-ecological interaction increases the demand on any surgical intervention on the facial musculoskeletal system. This explains the widespread need for pre-operative planning and simulations based on segmented patient specific data.

The goal of image segmentation is to partition the imaging data into multiple segments that will then be used for example patient specific simulations. At a lower level, the objective is to assign a label to each pixel in an image in a way that pixels with the same label share certain visual characteristics. Image segmentation is, however, an ill-posed problem that has been often casted in the Markov Random Field (MRF) framework in the literature. The recent advances in Graph-cut theory [2,3], that guarantee to globally optimize the MRF for a certain class of energy functions, made this approach even more attractive.

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 48–58, 2013.

Graph-cuts are very successful at segmenting objects that can be distinguished from their background producing globally optimal results, but they fail when the object is similar in appearance to its adjacent structures [8]. To alleviate the problem, people suggested incorporating prior shape knowledge into the graph-cuts. Since shape knowledge is used as a prior, this knowledge is incorporated in the smoothness term of the energy function by Veksler [15], Freedman and Zhang [8] and Das *et al.* [5]. Others do not consider the shape knowledge as a prior but more as likelihood information, therefore, they proposed to incorporate it in the data term of the energy function such as El-Zehiry and Elmaghraby [7], Freiman *et al.* [9], Ali *et al.* [1], Slabaugh and Unal [14] and Malcolm *et al.* [13]. Slabaugh and Unal [14] describe a class of representable shapes and add a constant factor to the data term, while El-Zehiry and Elmaghraby [7], Freiman *et al.* [9] and Ali *et al.* [1] proposed a more elaborated factor in the data term. Common to all solutions is the creation of a probability map by registering the shapes in the training datasets. They all propose an iterative scheme to refine their initial estimates and shape probabilities. These methods are prone to generating invalid shapes as there is no statistical dependence between the shapes. Malcolm *et al.* [13] use non-linear shape priors learned through Kernel PCA which does not suffer from statistical non-dependence. They then iteratively refine the shape prior and the segmentation by fitting the shape prior in the high dimensional space to the segmentation. The pre-image of the fitted shape prior in the input space is computed and then the updated shape prior is used to obtain better segmentation in the next iteration.

This paper bases on the earlier work of Majeed *et al.* [12] but extends it in several ways. In particular we introduce a non-linear cost function together with L^1 regularization [17] to provide bolder and more accurate shape update than that of [12] which uses a linear cost function. The shape knowledge is provided by the variability constrained SM as explained in the earlier work [12]. The main advantages of the proposed method is that non-linear cost function and L^1 regularization provides better shape update and guard against the SM from degenerating and collapsing onto itself.

The paper is organized as follows: Sec. 2 lays out the segmentation framework and how shape knowledge is extracted from the variability constrained SM. The creation of the non-linear cost function over which the SM is optimized to get a better shape fitting to the segmentation is detailed in Sec. 3. The complete algorithm is given in Sec. 4. Sec. 5 provides the results of applying the proposed method to segment masseter muscle and finally Sec. 6 provides the conclusion.

2 Segmentation Framework

The segmentation problem is cast as a binary labeling problem in the MRF framework. Let $\mathcal{L} = \{0, 1\}$ be the set of binary labels, "1" for object and "0" for background, $\mathcal{P}$ be the set of voxels of the volume dataset and $\mathbf{z} = \{z_p : p \in \mathcal{P}, z_p \in \mathcal{L}\}$ be the set of labeling which defines the segmentation. The goal of our segmentation is to find a labeling z, which is a mapping from $\mathcal{P} \longmapsto \mathcal{L}$ by minimizing the energy functional

$$E(\mathbf{z}|\mathbf{I},\mathbf{x}^*) = \sum_{p\in\mathcal{P}} \Big\{ V_p(z_p|\mathbf{I}) + \mu V_p(z_p|\mathbf{x}^*) \Big\} + \lambda \sum_{p\in\mathcal{P}} \sum_{q\in\mathcal{N}_p} V_{p,q}(z_p, z_q|\mathbf{I}) \,, \qquad (1)$$

where $\mathbb{N} = \{N_p|\forall p \in \mathcal{P}\}$ is an unordered 26 neighborhood system over $\mathcal{P}$, $\mathbf{I}$ is the observed intensity data, $\mathbf{x}^*$ is the shape prior, λ is the smoothness parameter and μ is the shape parameter. $V_p(z_p|\mathbf{I})$ and $V_{p,q}(z_p, z_q|\mathbf{I})$ are the data and the smoothness terms respectively, based on the image intensity information. The data term encodes how likely a voxel is to belong to object and background given its intensity while the smoothness term encodes our prior assumption about the target object that it consists of a homogeneous region, therefore, the smoothness term assigns a penalty whenever adjacent voxels p, q are assigned different labels z_p and z_q. The data and the smoothness terms are based on the traditional graph-cut intensity based energy functional of Boykov and Jolly [2].

$V_p(z_p|\mathbf{x}^*)$ is the shape data term which encodes how likely a particular voxel p is to belong to the object "1" and the background "0", given the shape prior $\mathbf{x}^*$ obtained from a SM explained below. Shape knowledge is encoded by creating a probability map both for the object and the background from the unsigned distance map of the shape prior's contour and it is similar to that of Majeed *et al.* [12]. Based on the closeness to the shape's contour; the object probability map is created for the voxels enclosed by the contour while the background probability map is created for the voxels not enclosed by the contour.

2.1 Statistical Model

This section summarizes the method of Lüthi *et al.* [11]. The same anatomical structures show considerable shape variability among the population which cannot be represented by a fixed shape template. Statistical shape models have been extensively used as a mathematical framework to capture this shape variability [4,10,16]. A set of shapes in the training dataset are used to capture the shape variability of the particular structure. The shapes in the training dataset are assumed to be Independent and Identically Distributed (i.i.d) having an underlying unknown multivariate Gaussian distribution with probability density function $p \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{\Sigma})$ with mean $\bar{\mathbf{x}}$ and covariance $\mathbf{\Sigma}$. The shapes in the training dataset $\{\mathbf{x}^i \in \mathbb{R}^{3m} | i = 1, \ldots, n\}$, where n represents the number of training shapes each having m number of vertices, are brought into correspondence using the method of Dedner *et al.* [6], which results in all shapes having the same number of vertices. Singular Value Decomposition (SVD) is then applied to decompose $\mathbf{\Sigma} = \mathbf{U}\mathbf{D}^2\mathbf{U}^T$, where $\mathbf{U}$ are the eigenvectors while $\mathbf{D}^2$ represents the eigenvalues of $\mathbf{\Sigma}$.

Reconstruction from Partial Information. The shape is represented by a surface mesh $\mathbf{x}$ which can be partitioned into $\mathbf{x} := (\mathbf{x}_a, \mathbf{x}_b)^T$, based on the available l-landmark information $\mathbf{x}_b \in \mathbb{R}^{3l}$ and unknown $\mathbf{x}_a \in \mathbb{R}^{3m-3l}$. The landmarks ($l = 6$), which are manually labelled, provide the location of the

muscle attachments at the facial bones. $\mathbf{x}_b$ is then used to estimate $\mathbf{x}_a$. Using the PPCA based approach of Lüthi *et al.* [11] a probability distribution over the shape $\mathbf{x}$ can be defined as

$$p(\mathbf{x}) = p(\mathbf{x}_a, \mathbf{x}_b) = \mathcal{N}\left(\begin{bmatrix} \bar{\mathbf{x}}_a \\ \bar{\mathbf{x}}_b \end{bmatrix}, \begin{bmatrix} \mathbf{W}_a\mathbf{W}_a^T & \mathbf{W}_a\mathbf{W}_b^T \\ \mathbf{W}_b\mathbf{W}_a^T & \mathbf{W}_b\mathbf{W}_b^T \end{bmatrix} + \sigma_m^2 \mathcal{I}_{3l} \right), \tag{2}$$

where $\mathcal{I}_{3l}$ is a $3l \times 3l$ identity matrix, $\mathbf{W} = \mathbf{UD} = [\mathbf{W}_a\mathbf{W}_b]^T \in \mathbb{R}^{3m\times d}$ is the d-largest scaled eigenvectors and σ_m^2 is a parameter that controls the remaining variance of the SM. If $\sigma_m > 0$ then $\mathbf{x}_b$ is allowed to move. Since $\mathbf{x}$ has a multivariate normal distribution, the conditional distribution $p(\mathbf{x}_a|\mathbf{x}_b) \sim \mathcal{N}(\bar{\mathbf{x}}_{\mathbf{x}_a|\mathbf{x}_b}, \mathbf{\Sigma}_{\mathbf{x}_a|\mathbf{x}_b})$ is also a multivariate normal distribution with mean $\bar{\mathbf{x}}_{\mathbf{x}_a|\mathbf{x}_b}$ and covariance $\mathbf{\Sigma}_{\mathbf{x}_a|\mathbf{x}_b}$. $p(\mathbf{x}_a|\mathbf{x}_b)$ needs to be computed to reconstruct the shape $\bar{\mathbf{x}}_{\mathbf{x}_a|\mathbf{x}_b}$ from partial information $\mathbf{x}_b$. Since coefficients of the modes of variation $\boldsymbol{\alpha} = \mathcal{N}(0, I_n)$ of the SM defines a shape $\mathbf{x} = \mathbf{W}\boldsymbol{\alpha} + \bar{\mathbf{x}}$, therefore, first the coefficients are determined from the partial information $\mathbf{x}_b$ as $p(\boldsymbol{\alpha}|\mathbf{x}_b)$ and then $\bar{\mathbf{x}}_{\mathbf{x}_a|\mathbf{x}_b}$ can be reconstructed using

$$\bar{\mathbf{x}}_{\mathbf{x}_a|\mathbf{x}_b} = \arg\max_x p(\mathbf{x}|\boldsymbol{\alpha}) = \mathbf{W}\boldsymbol{\alpha} + \bar{\mathbf{x}}\,. \tag{3}$$

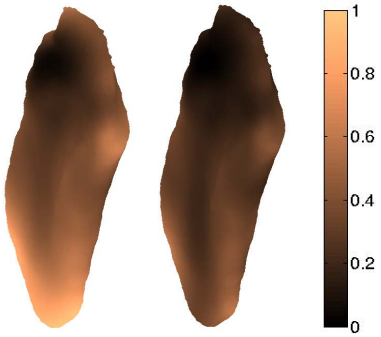


Fig. 1. Normalized variance of the SM. (a) Original variance, (b) Remaining variance (color online)

Remaining Variance. The known l-landmark information can be further utilized to constrain the variability of the SM. Since $\mathbf{x}_b$ provides additional information, therefore, in a probabilistic setting it is natural to assume that the uncertainty of the SM will reduce as further evidence is obtained. The covariance matrix $\mathbf{\Sigma}_{\mathbf{x}_a|\mathbf{x}_b}$ can be decomposed by applying SVD into its eigenvectors $\mathbf{U}_{\mathbf{x}_a|\mathbf{x}_b}$ and eigenvalues $\mathbf{D}^2_{\mathbf{x}_a|\mathbf{x}_b}$. $\mathbf{U}_{\mathbf{x}_a|\mathbf{x}_b}$, $\mathbf{D}^2_{\mathbf{x}_a|\mathbf{x}_b}$ and $\bar{\mathbf{x}}_{\mathbf{x}_a|\mathbf{x}_b}$ can now be used to generate an optimal shape $\mathbf{x}^*$ with the remaining flexibility of the model using

$$\mathbf{x}^* = \mathbf{U}_{\mathbf{x}_a|\mathbf{x}_b}\mathbf{D}_{\mathbf{x}_a|\mathbf{x}_b}\boldsymbol{\alpha} + \bar{\mathbf{x}}_{\mathbf{x}_a|\mathbf{x}_b}. \tag{4}$$

As an illustration of the concept of remaining variability, we show in Fig. 1 the original variance of the model (a) and the remaining variance of the model after being fit to the muscle attachments (b). It is however, not possible to compute $\mathbf{\Sigma}_{\mathbf{x}_a|\mathbf{x}_b}$ directly since it is potentially huge. For an in-depth analysis of the reconstruction of the shape given partial information and calculating the remaining variance see Lüthi *et al.* [11].

3 Adaptive Shape Optimization

For updating the shape prior with respect to the segmentation, we propose to use a shape optimization based on adaptive weights with respect to the remaining variances of the SM and sparse shape optimization.

3.1 Shape Cost Function

Creating the cost function for shape optimization is the second major step of our algorithm after the segmentation corresponding to the energy function E Eq. 1. Since the surface mesh is very dense consisting of 39156 vertices, therefore, adjacent vertices are close enough to occupy adjacent voxels. We have used the morphable model of Blanz and Vetter [16] which is very dense (around 76000 vertices) as compared to the active shape model of Cootes *et al.* [4] which are not dense (only 72 vertices). On average for all datasets there are 1.75 vertices per voxel with a density of 7.66 vertices per mm^2. Once a vertex is in a voxel, the cost of the vertex can be directly read out of the voxel, defined by

$$C(v) = \beta C_{obj}(v) + \eta C_{edge}(v) + C_{seg}(v), \tag{5}$$

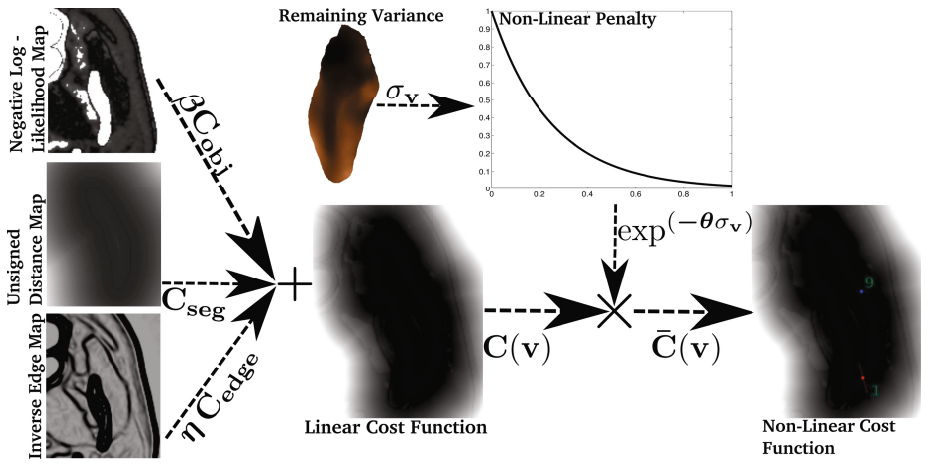
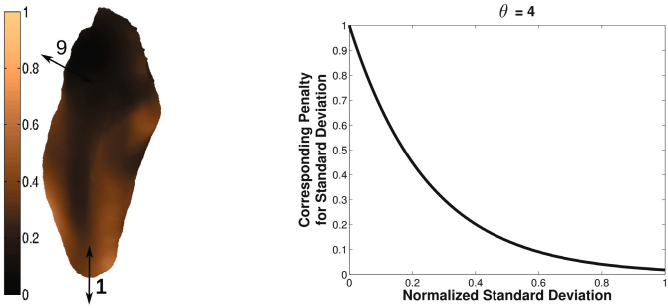


Fig. 2. Generating non-linear cost function $\bar{C}(v)$



(a) Illustration of the adaptive cost. (b) Penalty $\exp^{(-\theta\sigma_v)}$ for the normalized standard deviation.

Fig. 3. Normalized standard deviations and their corresponding penalty

where η and β are weighting parameters, $v = (v_x, v_y, v_z)$ represents the x, y, z coordinates of a vertex of $\mathbf{x}$. The object intensity negative log-likelihood map (C_{obj}) is calculated using the parzen window estimation that has already been estimated during the graph-cut segmentation. An inverse edge map C_{edge} provides low values where there is an edge. The third term in Eq. 5 is the unsigned distance map (C_{seg}) which is calculated from the segmentation boundary. All the maps are then linearly combined as in Eq. 5 and shown in Fig. 2 to generate the linear cost function $C(v)$ which is then weighted with non-linear variance penalties $\exp^{(-\theta\sigma_v)}$ to generate non-linear cost function $\bar{C}(v)$ as shown in Fig. 2 over which the SM is optimized.

3.2 Shape Optimization

Once the non-linear cost function has been created, the next step is to optimize the SM over the generated cost function. We propose to use an adaptive cost instead of linear cost employed by Majeed *et al.* [12] in order to make the SM robust to local minima encountered during the shape update. The cost is adapted with respect to the remaining variance of the vertex σ_v. The sum of the cost of all the vertices for a particular setting of shape coefficients $\boldsymbol{\alpha}$ gives the cost of the shape as follows

$$\bar{C}(\mathbf{x}) = C(\mathbf{UD}\boldsymbol{\alpha} + \bar{\mathbf{x}}) \exp^{(-\theta\sigma_\mathbf{v})} = \sum_v C(v) \exp^{(-\theta\sigma_v)}, \tag{6}$$

where σ_v is the remaining variance of vertex v and θ is a weight parameter.

Here vertices with higher variance incur lower cost in comparison to vertices with lower variance. As a consequence, a vertex with higher remaining variance (color coded in light golden in Fig. 3(a)) as given by the SM is allowed to move further with less cost (cost 1) while a vertex with lower remaining variance (color coded in black) incurs higher cost (cost 9) when it moves the same distance. With linear cost, vertices irrespective of their remaining variance in the SM would incur

equal cost when they move equal distances. The adaptive weights with respect to the variances are shown by the graph in Fig. 3(b).

The SM is optimized by minimizing the sum of the cost of vertices over the non-linear cost function $\bar{C}(v)$. The coefficients $\boldsymbol{\alpha}$ corresponding to the main modes of variation are obtained by solving the minimization problem

$$\min_{\boldsymbol{\alpha}} \left\{ \bar{C}(\mathbf{x}) + \xi |\boldsymbol{\alpha}|_{L^1} \right\}, \tag{7}$$

where ξ is a weight parameter. Since the adaptive cost is used, the SM has more flexibility, therefore, it is required that the model be regularized to constrain the solutions space and generate smoother shape priors. Note that we use L^1 regularization [17] to constrain the solution space and generate sparse and more accurate solutions. Once the optimal $\boldsymbol{\alpha}$ are found, the optimized shape is then constructed using Eq. 4 and used as a shape prior for the next iteration.

4 Algorithm

Figure 4 outlines the algorithm. The algorithm starts with the initial shape prior obtained from the shape reconstruction from partial information (see Sec. 2.1), therefore, $\bar{\mathbf{x}}_{\mathbf{x}_a|\mathbf{x}_b}$ which is the mean shape of the constrained variability SM is the initial shape prior used for the first iteration. The shape prior is used to generate the probability maps for the object and the background which is similar to the one used by Majeed *et al.* [12] and encodes the shape knowledge. These maps

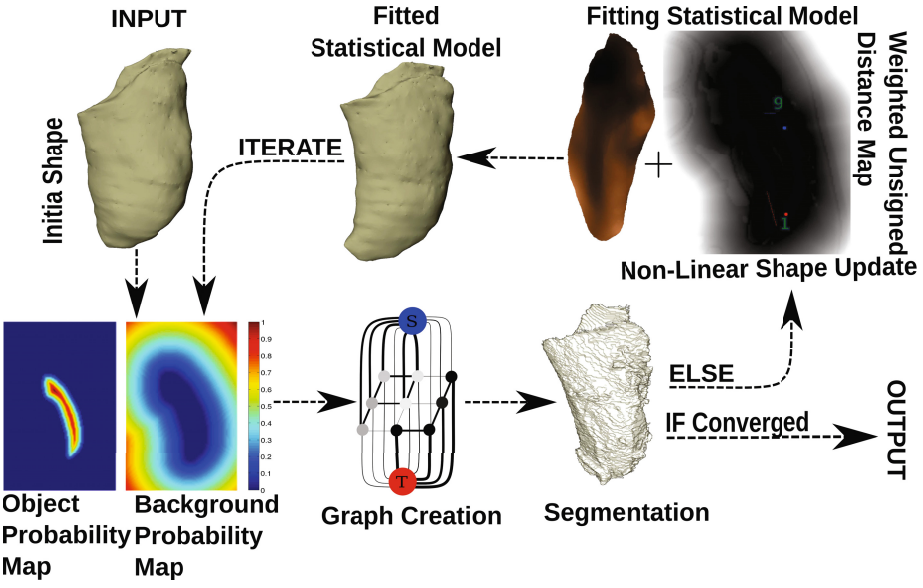


Fig. 4. Segmentation Process (color online)

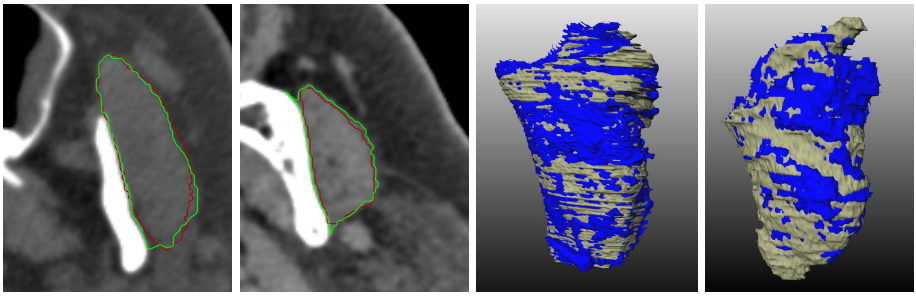
are then used to create graph corresponding to the energy function E given by Eq. 1 and then the graph-cut algorithm of Boykov and Kolmogorov [3] is used to obtain the muscle segmentation. If the segmentation has not converged then a non-linear cost function $\bar{C}(\mathbf{x})$ outlined in Sec. 3.1 is created over which the shape prior is updated. Once the SM has been fitted to the current segmentation (see Sec. 3.2), the fitted SM provides better and more accurate shape knowledge for the next iteration. This process is repeated until segmentation converges.

The update of the shape prior is required as the initial estimate of the shape is not perfect, therefore, previous segmentation is used to update shape knowledge and get a better fitting of the SM to the specific patients muscle anatomy.

5 Experimental Results

The proposed segmentation method was tested on 20 CT datasets - the ground truth was provided by a medical expert - using a Leave-One-Out approach. The dataset dimensions were 79-156×148-214×125-384 voxels and spacing 0.3-0.5×0.3-0.5×0.3-1 mm^3. All datasets possessed high-density artifacts caused by dental fillings and dental implants. The parameters $\sigma_m = 10$, $\lambda = 0.016$ and $\mu = 0.0037$ were optimized on three different datasets and used throughout the entire segmentation experiments. The parameters $\beta = 0.01$, $\eta = 0.07$, $\theta = 4$ were used to generate the non-linear cost function while $\xi = 600000$ was used for sparse shape optimization. The dice coefficient, sensitivity and specificity of the segmentation were calculated as similarity measures to ascertain the accuracy of the proposed method.

Shape convergence was achieved within $5 - 11$ iterations. The algorithm is computationally quite fast; it takes on average 4.1 ± 1.5 minutes. 4 min. is not real time but on the other hand it takes around an hour and a medical expert to segment the muscle. It should be noted that although the mesh employed is very dense, the algorithm itself is independent of the density of the mesh.



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

Fig. 5. (a,b) Qualitative segmentation result in 2D where red is the ground truth and green is the segmentation boundary. (c,d) Qualitative segmentation result in 3D where ground truth is in gray and segmentation is in blue (color online).

The cost function is evaluated where the vertices end up and that gives the total cost of the shape. The algorithm will work equally well with a less dense mesh.

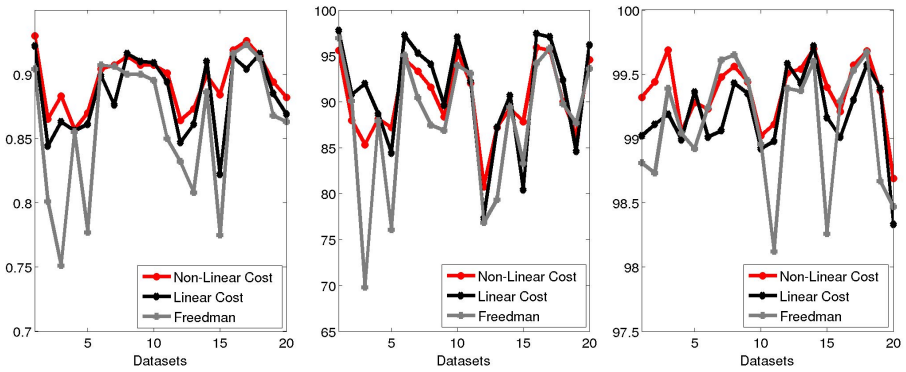
Figure 5(a+b) shows qualitative results of our technique in 2D, while the qualitative results in 3D are shown in Fig. 5(c+d). The experimental results obtained using the proposed method is clinically acceptable as validated by the medical expert.

The graphs in Fig. 6 show the results of the method of [12] using a linear cost function (black curve) and the proposed method with non-linear cost function (red curve). The gray curve shows the results of using the method of Freedman *et al.* [8]. The proposed method is statistically significantly better than both the methods [8] with p-value $p < 0.01$ and [12] with p-value $p < 0.01$. The improvemnt over [12] is mainly due to the use of the non-linear cost function using L^1 regularization. The dice coefficient (see Fig. 6(a)) and specificity (see Fig. 6(c)) for the proposed method is better for all datasets except for a few. Table 1 lists the mean, median, standard deviation and the smallest and the largest dice coefficient values for the methods.

We show that our novel approach shows a further improvement in the segmentation accuracy. In this paper we showed that SM models can not only be used to restrict the shape variability during segmentation but also how to make use of the remaining shape variability in the SMs to even further improve the segmentation.

Table 1. The table list the mean, median and the standard deviation of the dice coefficient of the proposed method, method with linear cost [12] and Freedman [8]

	DC (Mean ± Std)	DC (Median)	DC (Smallest - Largest)
Proposed	0.895 ± 0.022	0.900	(0.857 - 0.930)
Linear [12]	0.884 ± 0.029	0.890	(0.822 - 0.922)
Freedman [8]	0.861 ± 0.054	0.877	(0.751 - 0.923)



(a) Dice Coefficient (b) Sensitivity (c) Specificity

Fig. 6. Quantitative segmentation result: (a) Dice coefficient. (b) Sensitivity. (c) Specificity; for all the methods (color online).

6 Conclusion

In this paper we have proposed an improved segmentation approach that combines a constrained SM with an MRF-based segmentation approach. As compared to the state-of-the-art methods we employ a non-linear cost function when fitting the SM. This new cost function has shown to be superior as it generates more consistent shape updates. The method's performance has been evaluated on 20 masseter CT dataset and quantitatively compared to state-of-the-art segmentation approaches. Although the method has been shown and evaluated on the masseter muscle it is of general use and can be applied whenever SM is available.

Acknowledgment. We would like to thank Marcel Lüthi for providing the registered PCA based statistical model of the masseter muscle. This work has been supported by the NCCR/CO-ME research network of the Swiss National Science Foundation.

References

1. Ali, A.M., Farag, A.A., El-Baz, A.S.: Graph Cuts Framework for Kidney Segmentation with Prior Shape Constraints. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 384–392. Springer, Heidelberg (2007)
2. Boykov, Y., Jolly, M.P.: Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images. In: ICCV, vol. 1, pp. 105–112 (2001)
3. Boykov, Y., Kolmogorov, V.: An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. PAMI 26(9), 1124–1137 (2004)
4. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active Shape Models; Their Training and Application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
5. Das, P., Veksler, O., Zavadsky, V., Boykov, Y.: Semi-Automatic Segmentation with Compact Shape Prior. Image and Vision Computing 27(1), 206–219 (2009)
6. Dedner, A., Lüthi, M., Albrecht, T., Vetter, T.: Curvature Guided Level Set Registration using Adaptive Finite Elements. In: Pattern Recognition, pp. 527–536 (2007)
7. El-Zehiry, N., Elmaghraby, A.: Graph Cut Based Deformable Model with Statistical Shape Priors. In: ICPR, pp. 1–4 (2008)
8. Freedman, D., Zhang, T.: Interactive Graph Cut Based Segmentation with Shape Priors. In: CVPR, pp. 755–762 (2005)
9. Freiman, M., Kronman, A., Esses, S.J., Joskowicz, L., Sosna, J.: Non-parametric Iterative Model Constraint Graph min-cut for Automatic Kidney Segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 73–80. Springer, Heidelberg (2010)
10. Leventon, M.E., Grimson, W.E.L., Faugeras, O.: Statistical Shape Influence in Geodesic Active Contours. In: CVPR, p. 1316 (2000)
11. Lüthi, M., Albrecht, T., Vetter, T.: Probabilistic Modeling and Visualization of the Flexibility in Morphable Models. In: Mathematics of Surfaces, pp. 251–264 (2009)

12. Majeed, T., Fundana, K., Lüthi, M., Kiriyanthan, S., Beinemann, J., Cattin, P.C.: Using a Flexibility Constrained 3D Statistical Shape Model for Robust MRF-Based Segmentation. In: MMBIA, pp. 57–64 (2012)
13. Malcolm, J., Rathi, Y., Tannenbaum, A.: Graph Cut Segmentation with Nonlinear Shape Priors. In: ICIP, vol. 4, pp. 365–368 (2007)
14. Slabaugh, G.G., Unal, G.: Graph Cuts Segmentation Using an Elliptical Shape Prior. In: ICIP, pp. 1222–1225 (2005)
15. Veksler, O.: Star Shape Prior for Graph-Cut Image Segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 454–467. Springer, Heidelberg (2008)
16. Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces. In: Computer Graphics and Interactive Techniques, pp. 187–194 (1999)
17. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Sparse Shape Composition: A New Framework for Shape Prior Modeling. In: CVPR, pp. 1025–1032 (2011)

Novel Context Rich *LoCo* and *GloCo* Features with Local and Global Shape Constraints for Segmentation of 3D Echocardiograms with Random Forests

Kiryl Chykeyuk, Mohammad Yaqub, and J. Alison Noble

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

Abstract. This work addresses the challenging problem of segmenting the myocardium in 3D LV echocardiograms by Random Forests (RF). While the RF framework has proven to be a good discriminative classifier for segmentation of 3D echocardiography [1], our hypothesis is that richer features than those traditionally used (Haar etc) need to be employed for accurate segmentation to tackle artifacts in ultrasound images such as missing anatomical boundaries. To address this, we propose two new context rich and shape invariant features, called *LoCo* and *GloCo*. The new features impose a local and global constraint on the coupled endocardial and epicardial shape of the left ventricle and use barycentric coordinates to uniquely identify the position of a voxel with respect to a number of landmarks on the epicardial and endocardial border. The landmarks are found using a new measure (COFA) to separate the two boundaries. Experimental results show that the new features provide a smoother segmentation and improve the accuracy compared with a classic RF implementation.

Keywords: Random forests, 3D echocardiographic segmentation, the monogenic signal, feature asymmetry, barycentric coordinates.

1 Introduction

Accurate automatic segmentation of the myocardium of the left ventricle (LV) provides quantitative data from 3D echocardiographic images that helps in the assessment of heart abnormalities and diseases. However, automatic segmentation of 3D echocardiography is challenging due to ultrasound artifacts such as shadowing, attenuation, signal drop-out, speckle, missing boundaries and similarity in appearance of different tissues, for instance of the left and right ventricles.

The accuracy of image-based classification techniques reported in the literature to segment and/or detect structures in medical images varies depending on, for example, the classification model, chosen feature set or the complexity of the structure of interest. Segmentation of the LV is challenging. Therefore, the classification model and features need to be chosen and developed carefully. Random Forests (RF) have

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 59–69, 2013.

proven to show good performance in recent publications [1-6] and was adopted in this work on 3D echocardiography segmentation. In the RF framework, segmentation is formalized as a voxel classification problem. Although the choice of the classifier is important, the accuracy of the model is pre-dominantly determined by the features used within the classifier. In our approach, first a novel image alignment method is proposed to make the widely used position features stronger. Second, a novel set of context rich features is introduced to improve automated segmentation of 3D echocardiography.

The contributions of this work are twofold. Firstly, as a pre-segmentation step, we propose a new method for left ventricular long axis detection in 3D echocardiography. The robust detection of the mid line is an important step in alignment of echocardiography volumes which is done prior to segmentation. Our method is based on a Feature Asymmetry measure (FA), Local Orientation (LO) and a modified Hough transform. Secondly, we introduce a set of new context rich features that utilize the relevant position of a voxel with respect to a specific landmark or landmarks at the epicardial and/or endocardial boundaries of the left ventricle. The landmarks are detected using the new Centrally Oriented Feature Asymmetry measure (COFA) to highlight and separate the epicardial and endocardial boundaries in an image.

1.1 Random Forests

Random Forests [2] is a learning-based technique gaining popularity in medical imaging, in which training using a gold standard segmentation is done by building multiple decision trees. Each node in the tree, except the leaves, is a decision node and contains a feature and a threshold. Each leaf node contains a class distribution for the voxels that reached the node. Testing is performed by traversing voxels over the trees starting from the root of each tree to a leaf node. The voxels are split at a given node based on the learned feature and the threshold value at that node. The mean class distribution from all trees is considered the final probabilistic class distribution of the test case. For more information see [1-6].

1.2 The Monogenic Signal and Feature Asymmetry Images

In echocardiography, low-level feature extraction is an important step before the segmentation of the LV is performed. For LV segmentation, the goal is to detect endocardial and epicardial boundaries. It is usually assumed that they have step-like edge characteristics. It has been shown that intensity based methods do not perform well due to the low-contrast nature of echocardiographic images, whereas local-phase based techniques have been shown to be intensity-invariant and less sensitive to speckle [7]. In [7], the original Phase Congruency measure [8] was adapted to the Feature Asymmetry (FA) measure, and outperformed the intensity based methods for detecting step-like edges in echocardiographic images. It has since been modified in [9] using the monogenic signal [10].

The monogenic signal is a high dimensional generalization of the analytic signal. It is based on the Riesz transform, which is used instead of the Hilbert transform. The monogenic signal is formed by combining the original band-pass image with its Riesz components:

$$\mathrm{I_M}(x,y,z) = [\mathrm{I}(x,y,z) * \mathrm{g}(x,y,z), \mathrm{I}(x,y,z) * \mathrm{g}(x,y,z) * \mathrm{h_x}(x,y,z), \mathrm{I}(x,y,z) * \mathrm{g}(x,y,z) * \mathrm{h_y}(x,y,z), \mathrm{I}(x,y,z) * \mathrm{g}(x,y,z) * \mathrm{h_z}(x,y,z)], \quad (1)$$

where $g(x,y,z)$ is the spatial domain representation of a bandpass filter and $h_i(x,y,z)$ are the Riesz filter components.

The odd and even filter responses are then defined as follows:

$$\mathrm{even}(x,y,z) = \mathrm{I_{M,1}}(x,y,z) \quad (2)$$
$$\mathrm{odd}(x,y,z) = \sqrt{\mathrm{I_{M,x}}(x,y,z)^2 + \mathrm{I_{M,y}}(x,y,z)^2 + \mathrm{I_{M,z}}(x,y,z)^2}$$

In computing the monogenic signal, aside from choosing how to combine the results from different scales, the selection of a bandpass filter has to be made. In our case we used the log-Gabor filter though other filters could have been used.

The feature asymmetry (FA) measure is defined as:

$$\mathrm{FA_{3D}}(x,y,z) = \sum_s \frac{\lfloor \|\mathrm{odd_s}(x,y,z)\| - \|\mathrm{even_s}(x,y,z)\| - \mathrm{T_s} \rfloor}{\sqrt{(\mathrm{even_s}(x,y,z))^2 + (\mathrm{odd_s}(x,y,z))^2} + \varepsilon} \quad (3)$$

where ε is a small constant to avoid division by zero and T_s is a scale specific threshold that suppresses any response due to noise or symmetric points of the image:

$$\mathrm{T_s} = \exp\left[\mathrm{mean}\left(\log\left(\sqrt{(\mathrm{even_s}(x,y,z))^2 + (\mathrm{odd_s}(x,y,z))^2}\right)\right)\right] \quad (4)$$

2 Method

In this section, we describe the procedure for myocardium segmentation using the RF with the new *LoCo* and *GloCo* features, see Fig. 1. We first describe the pre-segmentation steps for the *LoCo* and *GloCo* features extraction.

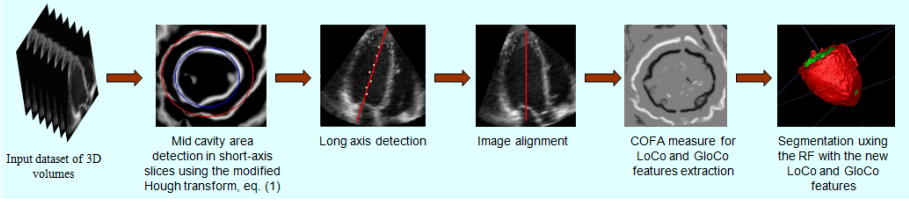
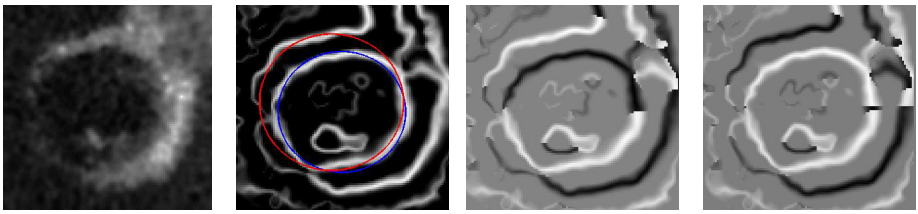


Fig. 1. Algorithm flowchart



(a) Original image (b) Circle detection on FA (c) FA·sign($I_{M,y}$) (d) FA·sign($I_{M,y}$) · sign$(y - c_y)$

Fig. 2. Circle detection on a short-axis slice from the 3D echocardiogram. (a) The original image. (b) The FA image and the detected circles by the original Hough transform for circles (in red, failed) and by the proposed modified Hough transform for circles (in blue) defined by (5). (c) and (d) The intermediate measures used by the modified Hough transform (5). In (d) the clear separation of epicardial and endocardial borders makes it feasible for the modified Hough transform not to confuse between the two borders.

2.1 Detection of the Long Axis

We propose a novel method for long axis detection from 3D echocardiographic images utilizing a modified Hough transform for circles, the 3D monogenic signal and the FA images as edge maps.

Specifically, a local-phased based version of the Hough transform for circle detection can be defined as follow:

$$H_z(c_x, c_y, r) = \sum_s \sum_\theta \text{sign}(x - c_x) \cdot \text{sign}\left(I^s_{M,x}\right) \cdot FA^s_z(x,y) + \text{sign}(y - c_y) \cdot \text{sign}\left(I^s_{M,y}\right) \cdot FA^s_z(x,y); \quad (5)$$

$$\text{where} \quad \begin{cases} x = c_x + r\cos\theta \\ y = c_y + r\sin\theta \end{cases} \quad (6)$$

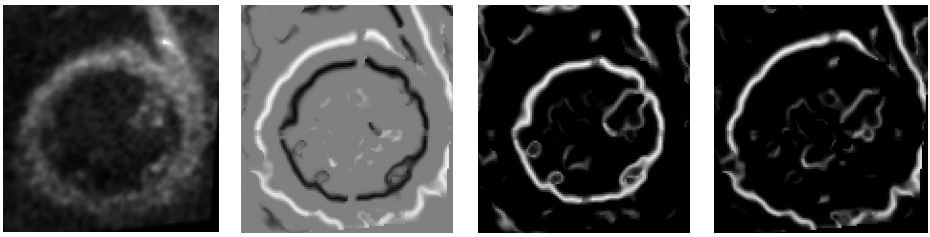
c_x, c_y, r parameterize the circle, and $I^s_{M,x}$ and $I^s_{M,y}$ are the *x* and *y* monogenic signal components respectively, as defined in (1). See also Fig. 2.

In our implementation the summation is performed over multiple scales *s* of the band-pass filter. Notice that the Hough transform can be fairly accurately computed using only one of the summands in (5). However, we have found in our application that employing both terms leads to better accuracy and robustness.

In our application we use the local-phase based Hough transform on each short-axis slice z, to give the endocardial and epicardial center points, $(x^z_{c,endc}, y^z_{c,endc})$ and $(x^z_{c,epic}, y^z_{c,epic})$, and the endocardial and epicardial radii, r^z_{endc} and r^z_{epic}, as:

$$\begin{aligned} x^z_{c,epic}, y^z_{c,epic}, r^z_{epic} &= \text{argmax}_{c_x,c_y,r}\, H_z(c_x, c_y, r); \\ x^z_{c,endc}, y^z_{c,endc}, r^z_{endc} &= \text{argmin}_{c_x,c_y,r}\, H_z(c_x, c_y, r); \end{aligned} \quad (7)$$

Having detected the center points, the long axis is then fitted as the first principal component of the detected endocardial centers. The long axis is utilized to align the 3D LV images, to facilitate the extraction of a new set of features as described next.



(a) Original Image (b) $COFA$ (c) $COFA_{endc}$ (d) $COFA_{epic}$

Fig. 3. Visualisation of the COFA measure. (a-b) The original image and COFA measure defined in (9). (c-d) The COFA measures for the endocardial and epicardial borders defined in (10) and (11).

2.2 Centrally Oriented Feature Asymmetry (COFA) Measure for Separating and Highlighting the Epicardial and Endocardial Boundary

The Centrally Oriented Feature Asymmetry (COFA) measure is defined in terms of the detected epicardial and endocardial points for each of the short-axis slices z, as follows:

$$\mathrm{COFA_z}(x,y) = \sum_s \sum_{i=epic,endc} \mathrm{sign}\left(x - x_{c,i}^z\right) \cdot \mathrm{sign}\left(I_{M,x}^s\right) \cdot \mathrm{FA_z^s}(x,y) + \mathrm{sign}\left(y - y_{c,i}^z\right) \cdot \mathrm{sign}\left(I_{M,y}^s\right) \cdot \mathrm{FA_z^s}(x,y) \quad (8)$$

See also Fig. 3 (b). The epicardial and endocardial boundaries can be separated in the following way:

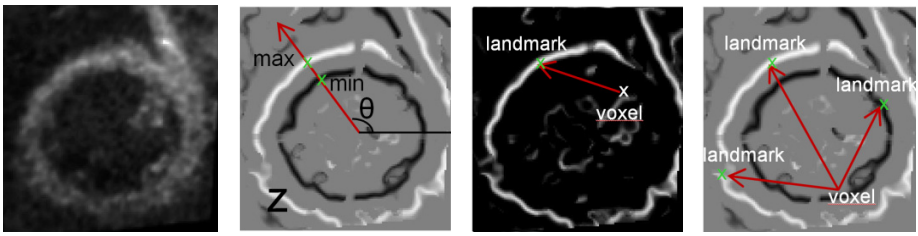
$$\mathrm{COFA_{z,epic}}(x,y) = \sum_s \sum_{i=epic,endc} \left\lfloor \mathrm{sign}\left(x - x_{c,i}^z\right) \cdot \mathrm{sign}\left(I_{M,x}^s\right) \cdot \mathrm{FA_z^s}(x,y) \right\rfloor + \left\lfloor \mathrm{sign}\left(y - y_{c,i}^z\right) \cdot \mathrm{sign}\left(I_{M,y}^s\right) \cdot \mathrm{FA_z^s}(x,y) \right\rfloor \quad (9)$$

$$\mathrm{COFA_{z,endc}}(x,y) = \sum_s \sum_{i=epic,endc} \left\lceil \mathrm{sign}\left(x - x_{c,i}^z\right) \cdot \mathrm{sign}\left(I_{M,y}^s\right) \cdot \mathrm{FA_z^s}(x,y) \right\rceil + \left\lceil \mathrm{sign}\left(y - y_{c,i}^z\right) \cdot \mathrm{sign}\left(I_{M,x}^s\right) \cdot \mathrm{FA_z^s}(x,y) \right\rceil \quad (10)$$

where $(x_{c,i}, y_{c,i})$ is either epicardial or endocardial center point in the slice *z,* determined by (7), $\lfloor\rfloor$ and $\lceil\rceil$ are the operators that zero the negative and positive values correspondingly. See also Fig. 3 (c) and (d).

2.3 Feature Sets

Having estimated the center lines, each volume is rotated to a common co-ordinate system so that the long axis is positioned vertically and centered in the aligned image. This corresponds to two translations and two rotations. The remaining five parameters (1 translation, 1 rotation and 3 scaling) are found using a standard rigid registration technique (in our work we used the FLIRT registration tool, http://www.fmrib.ox.ac.uk/fsl/flirt/), fixing the determined four parameters. The process of long axis detection and image alignment is repeated in an iterative manner until no further improvement is achieved. By aligning the images, features extracted from different images correspond, which improves the testing accuracy.



(a) Original Image (b) Detection of landmarks (c) *LoCo* feature (d) *GloCo* feature

Fig. 4. Visualisation of the *LoCo* and *GloCo* features. The landmark detection is illustrated in (b): randomly chosen z and θ define the slice and the direction of the search (red line) for the endocardial landmark (minimum along the line) or the epicardial landmark (maximum along the line). (c) An example of the *LoCo* feature using the epicardial landmark. (d) An example of the *GloCo* feature (shown in 2D for illustration purposes). Four randomly selected landmarks form a simplex in 3D, the *GloCo* features are Barycentric coordinates of this voxel with respect to the simplex.

Conventional Local Appearance Features. In this work we adopted several classic low level features. We used rectangle3D, Haar3D and Difference3D features [4-6]. These features capture local appearance information, but in practice can be sub-optimal for analysis of low quality images.

Absolute Voxel Position Features. Following the work of [1], we employ a position3D feature to capture the absolute position of the voxels in the image. The alignment procedure boosts the strength of such features. However, due to the global and local geometric variability of the myocardium, position features only provide a weak geometric constraint.

Local Shape Constraint Contextual Features (LoCo). Here, we introduce a novel set of context rich features $\mathbf{F}(z,\theta)$ that capture geometric variability of the myocardium and can be regarded as a shape constraint. This type of feature is determined by the relative position of a considered voxel $\mathbf{v}^k = (v_x^k, v_y^k, v_z^k)$ to a pre-chosen landmark $\mathbf{p}^k = (p_x^k, p_y^k, p_z^k)$ on the endocardial or epicardial boundary shape in image k. The *LoCo* feature set is thus defined as:

$$f_{LoCo}^{x}(\mathbf{v}^k;\mathbf{p}^k) = v_x^k - p_x^k \quad f_{LoCo}^{y}(\mathbf{v}^k;\mathbf{p}^k) = v_y^k - p_y^k \qquad f_{LoCo}^{z}(\mathbf{v}^k;\mathbf{p}^k) = v_z^k - p_z^k \qquad (11)$$
$$f_{LoCo}^{xy}(\mathbf{v}^k;\mathbf{p}^k) = \sqrt{(v_x^k - p_x^k)^2 + (v_y^k - p_y^k)^2} \qquad f_{LoCo}^{dist}(\mathbf{v}^k;\mathbf{p}^k) = |\mathbf{v}^k - \mathbf{p}^k|$$

Landmark selection and detection. With all the images globally aligned endo- and epicardial points can be corresponded in different images. Two randomly chosen parameters, the short-axis slice z and the polar angle θ, define in which slice and in what direction from the center of the cavity area the (aligned) epicardial or endocardial landmark will be detected, Fig. 4 (b). To find the epicardial and endocardial landmarks along the chosen direction, COFA images, described in sect. 3.2, are utilized. The maximum COFA detected feature along the θ direction defines the epicardial landmark, whereas the minimum defines the endocardial landmark, Fig. 4 (b).

Thus, a detected landmark is defined by the two parameters z and θ, and is detected in each image independently as follows:

$$\begin{cases} p_x^{k,q}, p_y^{k,q} = \mathrm{argmax}_{x,y} \left| \mathrm{COFA}_{z,q}(x,y) \right| \\ p_z^{k,q} = z \end{cases} \text{subject to} \begin{cases} x^q = x_{c,q}^z + r\cos\theta \\ y^q = y_{c,q}^z + r\sin\theta \end{cases} \quad \text{where } r_q^z - \text{shift} < r < r_q^z + \text{shift} \tag{12}$$

Here the endocardial/epicardial centers $(x_{c,q}^z, y_{c,q}^z)$ and radii r_q^z in the slice z are found by (7).

Global Shape Constraint Contextual Features (GloCo). Unlike *LoCo* features, *GloCo* features use the relative position of a voxel to a number of structures on the epicardial and endocardial boundary shapes. Barycentric coordinates are utilized to uniquely specify the location of the voxel with respect to the specified structures on the boundary shapes. To calculate the barycentric coordinates in 3D space, the 3D simplex, a tetrahedron, is constructed by randomly selecting four structure points on endo- and epicardial boundaries. Mathematically, the scalars u_1, u_2, u_3, u_4 are the barycentric coordinates of an arbitrary voxel $\mathbf{v} = (v_1\ v_2\ v_3)$ with respect to the four nonplanar structure points $\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_3}, \mathbf{p_4}$ if

$$\mathbf{v} = u_1\mathbf{p_1} + u_2\mathbf{p_2} + u_3\mathbf{p_3} + u_4\mathbf{p_4} \tag{13}$$

$$\text{subject to } u_1 + u_2 + u_3 + u_4 = 1 \tag{14}$$

where $\mathbf{v}$ and $\mathbf{p_i}$ denote the Euclidean coordinates. The four structure points are defined by randomly selecting a short-axis slice z and a directional angle θ for each of the points. The four landmarks are further detected separately in each image using (12). To ensure invariance to arbitrary pose of the epicardial and endocardial shape the coefficients are constrained to sum to one (14).

The proposed *GloCo* features are barycentric coordinates and computed as follows:

$$\begin{aligned} f_{GloCo}^1(\mathbf{v}^k; \mathbf{p}_1^k, \mathbf{p}_2^k, \mathbf{p}_3^k, \mathbf{p}_4^k) &= u_1 & f_{GloCo}^2(\mathbf{v}^k; \mathbf{p}_1^k, \mathbf{p}_2^k, \mathbf{p}_3^k, \mathbf{p}_4^k) &= u_2 \\ f_{GloCo}^3(\mathbf{v}^k; \mathbf{p}_1^k, \mathbf{p}_2^k, \mathbf{p}_3^k, \mathbf{p}_4^k) &= u_3 & f_{GloCo}^4(\mathbf{v}^k; \mathbf{p}_1^k, \mathbf{p}_2^k, \mathbf{p}_3^k, \mathbf{p}_4^k) &= u_4 \end{aligned} \tag{15}$$

where

$$(u_1\ u_2\ u_3)^{\mathrm{T}} = \mathbf{T}^{-1}(\mathbf{v} - \mathbf{p_4}), \quad u_4 = 1 - u_1 - u_2 - u_3 \tag{16}$$

$\mathbf{T} = \begin{pmatrix} x_1 - x_4 & x_2 - x_4 & x_3 - x_4 \\ y_1 - y_4 & y_2 - y_4 & y_3 - y_4 \\ z_1 - z_4 & z_2 - z_4 & z_3 - z_4 \end{pmatrix}$, x_i, y_i, z_i are the coordinates of landmark $\mathbf{p_i}$

Comparison of position and LoCo and GloCo features. A position feature does not exploit image intensity information and considers only the absolute position of the voxel in the image. Thus, it provides a weak constraint on geometric variability of the myocardium. On the contrary, the *GloCo* feature quantifies the relative position of the voxel with respect to four landmarks detected using intensity information from the COFA images. The detected landmarks are the tetrahedron vertices used in calculation of the barycentric coordinates and have different spatial locations across the images. Thus, the tetrahedron captures the coupled endo- and epicardial shape variability enforcing a strong constraint on the geometry of the myocardium.

3 Experimental Results

3.1 Dataset

25 3D end-diastolic echocardiograms from health subjects were used in this study. A Philips iE33 ultrasound system was used to acquire the images. Volume dimensions are (224×208×208) with an average of 0.88mm^3 spatial resolution. The myocardium and the blood pool of all volumes were manually segmented by an expert.

3.2 Validation Methodology

20 volumes were chosen randomly to train the RFs and the remaining 5 volumes were used in testing to report the results. To understand the impact of the new *LoCo* and *GloCo* features, two RF classifiers were trained: 1) using the previously reported local appearance and position features on the original images and 2) using the conventional local appearance, position features and also the new *LoCo* and *GloCo* features on the aligned images. The two RF were trained with 15 trees. The stopping criteria for growing the tree were the maximum depth - 16, no information gain of splitting and the minimum number of points at a node - 50. For the classic RF, 100 conventional features were randomly chosen at a node and further investigated to find the one that gives the highest information gain. For the RF with the new *LoCo* and *GloCo* features, 100 randomly chosen conventional features and 150 randomly chosen *LoCo* and *GloCo* features were used.

The RF was implemented in C++. Testing volumes were segmented in 20 seconds per volume using Intel Xeon 2.8GHz computer with 12 cores and 48GB RAM running Win7. With a parallel tree implementation, training required about 3 hours and only needed to be done once.

Fig. 5 shows a visual comparison of the segmentation from the classic RF and from the RF using the new features.

For quantitative analysis, the mean and standard deviation of the Dice and Jaccard similarity coefficients for both the myocardium and the blood pool are reported in Table 1. The Dice similarity coefficient is defined as $\text{Dice} = \frac{2\times|\text{GT} \cap \text{Auto}|}{|\text{GT}|+|\text{Auto}|}$, while the Jaccard is defined as $\text{Jaccard} = \frac{|\text{GT} \cap \text{Auto}|}{|\text{GT} \cup \text{Auto}|}$, where GT is the ground truth represented as manual segmentation and Auto is the automatic segmentation, |.| is the cardinality of a set.

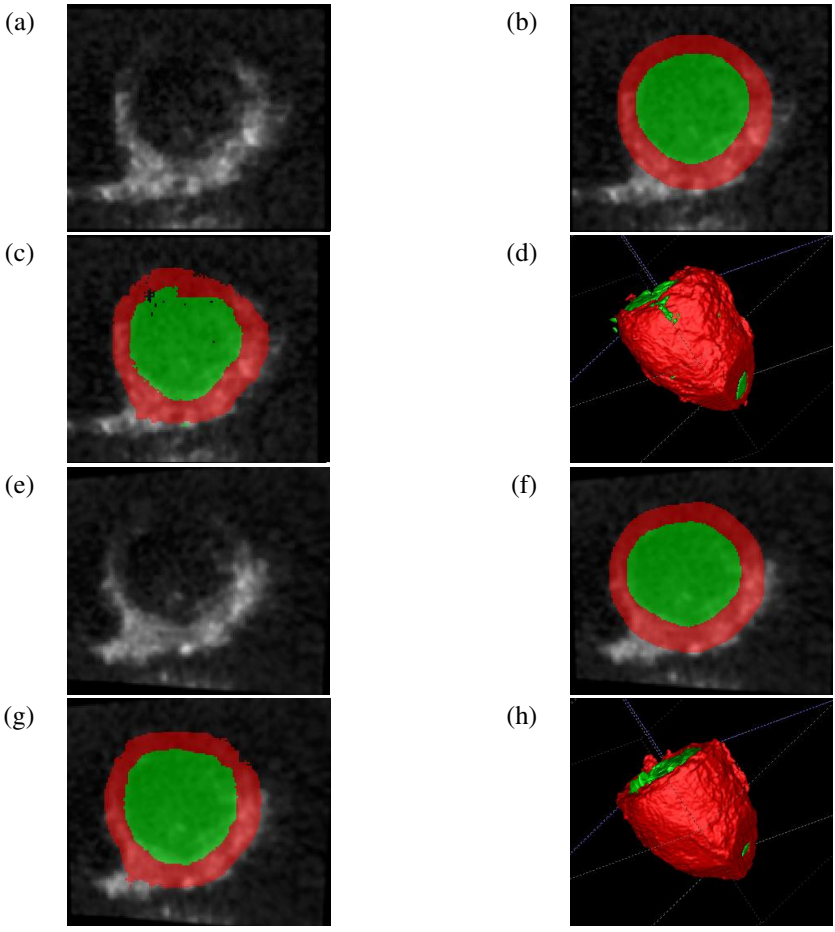
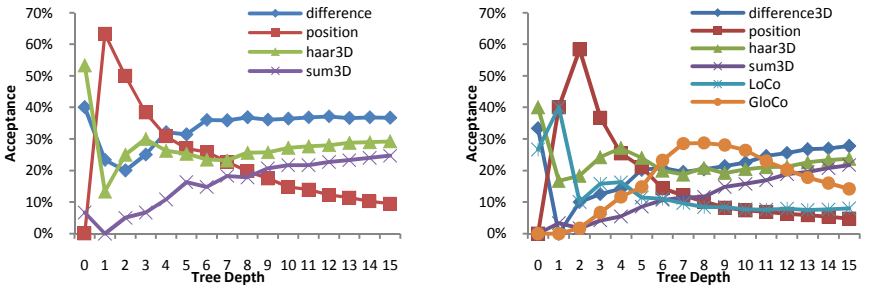


Fig. 5. (a) and (e) Short-axis slice of the original images ((e) is vertically aligned). (b) and (f) Manual segmentations of (a) and (e). (c) Automatic segmentation by the RF with the classic features (g) Automatic segmentation by the RF with the new LoCo and GloCo features. (d) and (h) 3D mesh of the automatically segmented volumes (a) and (e).

Table 1. Mean (μ) ± standard deviation (σ) of the dice and Jaccard coefficients for the myocardium and the blood pool

		Classic RF	RF with *LoCo* and *GloCo*
Myocardium	Dice	0.77 ± 0.06	0.77 ± 0.05
	Jaccard	0.63 ± 0.08	0.63 ± 0.06
Blood pool	Dice	0.88 ± 0.04	0.90 ± 0.02
	Jaccard	0.79 ± 0.06	0.82 ± 0.04



(a) Acceptance of conventional features within the RF

(b) Acceptance of conventional features and the new *LoCo and GloCo* features within the RF

Fig. 6. Feature acceptance at different depth with (a) the RF using the classic features and (b) the RF using the classic and the new *LoCo and GloCo* features

3.3 Discussion on the Conventional and the New *LoCo* and *GloCo* Features

The frequency of each type of feature selected during the training stage for different tree depths is shown in Fig. 6.

Position and appearance features. Recall that a position feature looks only at the absolute position of a voxel. The position features are pre-dominantly selected by both RFs at depths 1-4 when the complexity of the data is the highest, see Fig. 6. At this stage, the appearance features poorly distinguish the classes well due to the similar appearance of different tissues or the different appearance of the same tissue.

LoCo and GloCo features. Both the position and the *LoCo* and *GloCo* features divide the image domain spatially. In the case of the classic RF, the position features tend to split the data spatially until the depth 4, after which, as seen in Fig. 6 (a), the appearance features prove to separate the classes better within each of the spatial regions in the image. In the case of the RF with the new features the behavior is different. After depth 4, the *GloCo* features are shown to dominate. They continue splitting the image spatially into regions until the depth 11. This suggests that the *GloCo* features carry more detailed contextual information than the position or the *LoCo* features. Fig. 6 (b) suggests that after the depth 11, the separation is mainly done using the appearance information within each of the regions in the image. Note that by splitting the image spatially, the *LoCo* and *GloCo* features also encode a constraint on local and global shape variability.

We caveat our findings with two comments related to the dataset we used. The current training set consists of 20 3D volumes which is relatively small to capture the complexity of coupled variability of the epicardial and endocardial shapes. Thus firstly one would hypothesise an improvement of the Jaccard/Dice measures for the myocardial segmentation with an increase of training dataset size. Secondly one cannot guarantee that the small testing dataset used fairly captures the variability seen in the training dataset. Having said this, the results are encouraging and convincingly demonstrate the usefulness of the new proposed features for 3D echocardiography segmentation.

4 Conclusions

This paper proposes new context rich *LoCo* and *GloCo* features for medical imaging machine learning based segmentation that embed local and global shape constraints respectively. Our first contribution is the detection of the ventricular long axis based on image intensity information. This enables a) accurate alignment of all the volumes, which decreases the variability of LV position and thus, improves the quality of classic features; and b) detection of corresponding landmarks throughout the images, which are needed for the proposed features. Our second contribution is the new *LoCo* and *GloCo* features that encode local and global variability of coupled endo- and epicardial shapes.

We demonstrated improvement of 3D echocardiography segmentation over classic RF segmentation. The proposed features need to be tested on a larger training set to draw strong conclusions on the accuracy for myocardial segmentation which will be the subject of future work.

Acknowledgement. This research was funded by the UK EPSRC on grant EP/G030693/1.

References

1. Lempitsky, V., Verhoek, M., Noble, J.A., Blake, A.: Random Forest Classification for Automatic Delineation of Myocardium in Real-Time 3D Echocardiography. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 447–456. Springer, Heidelberg (2009)
2. Breiman, L.: Random Forests. Machine Learning 45(1), 5–32 (2001)
3. Schroff, F., Criminisi, A., Zisserman, A.: Object Class Segmentation using Random Forests. In: BMVC (2008)
4. Shotton, J., Johnson, M., Cipolla, R.: Semantic Texton Forests for Image Categorization and Segmentation. In: CVPR (2008)
5. Yaqub, M., et al.: Efficient Volumetric Segmentation using 3D Fast-Weighted Random Forests. In: MICCAI-MLMI, Toronto, Canada (2011)
6. Geremia, E., Menze, B.H., Clatz, O., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel MR Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 111–118. Springer, Heidelberg (2010)
7. Mulet-Parada, M., Noble, J.A.: 2D+T Acoustic Boundary Detection in Echocardiography. Medical Image Analysis 4, 21–33 (2000)
8. Kovesi, P.: Image Features from Phase Congruency. Journal of Computer Vision Research 1 (1999)
9. Rajpoot, K., Grau, V., Noble, J.A.: Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3-D echocardio images. In: ISBI (2009)
10. Felsberg, M., Sommer, G.: The Monogenic Signal. IEEE Transactions on Signal Processing 49, 3136–3144 (2001)

Novel Vector-Valued Approach to Automatic Brain Tissue Classification

Nataliya Portman and Alan Evans

McConnell Brain Imaging Centre, Montreal Neurological Institute
3801 University H3A 2B4, Montreal, Quebec, Canada
{nataliya,alan}@bic.mni.mcgill.ca
http://www.bic.mni.mcgill.ca

Abstract. In this work we propose a novel SSIM (Structural Similarity Index Measure)-guided brain tissue classification approach, implementing Kernel Fisher Discriminant Analysis (KFDA). In Computer Vision, KFDA has been shown to be competitive with other state-of-the-art techniques. In the KFDA-based framework, we exploit the complex structure of grey matter, white matter and cerebro-spinal fluid intensity clusters to find an optimal classification. We illustrate our novel technique using a dataset of early normal brain development in the age range from 10 days to 4.5 years. The SSIM metric, an objective measure of an image quality as perceived by the Human Visual System, is used to evaluate the quality of brain segmentation. SSIM comparison of the quality of classification obtained by the KFDA-based and the Expectation-Maximization algorithms shows the superior performance of the proposed technique.

Keywords: Kernel Fisher Discriminant Analysis, classification, testing set, partial volume effect, feature space, brain tissue classes.

1 Introduction

Motivation. This paper addresses the problem of automatic classification of brain tissue into white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) for an MR pediatric dataset of early brain development from birth through 4.5 years of age [1].
This dataset exhibits dramatic qualitative changes in GM/WM contrast during early brain maturation. The MRI signal is affected by myelinated axons of the major pathways (white matter and the corpus callosum). As a result of poor and highly variable GM/WM contrast and tight sulcal packing, automatic classification via INSECT [2] was difficult to implement.
We set out to achieve high quality classification of this dataset as this is fundamental for the accuracy of cortical surface extraction and the following assessment of normal surface variability in children before 4.5 years of age.

Classification Techniques in NeuroImaging. Many brain tissue classification techniques have been proposed, e.g. a k Nearest Neighbour classifier [3], an

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 70–81, 2013.

Artificial Neural Network classifier [4], an Expectation-Maximization (EM) algorithm [5], a modified EM-based algorithm using a Markov Random Field model [6] and a watershed-based segmentation [7], being among the most popular.

Different methods for the evaluation of classifier performance show that existing automatic classification algorithms do not fully capture expert tracings [8]. The major drawback is incorrect classification of the CSF into either background or GM.

The MR brain tissue labeling process is complicated by the presence of a partial volume (PV) effect due to the limited spatial resolution of the scanner, which leads to the presence of multiple tissue types within a single voxel. PV estimation (PVE) or computation of the mixing proportions of tissue classes per voxel is essential for an accurate quantification of tissue volumes and cortical surface extraction [9]. Among the proposed PVE techniques, a Trimmed Minimum Covariance Determinant (TMCD) approach [9] provides a generalized segmentation framework since it uses Gaussian distributions with different covariance matrices for modeling tissue intensity histograms and it can be applied to multi-channel data.

The disadvantage of the TMCD method is its computational complexity. We seek a simpler approach that would classify brain image data with high accuracy.

Why KFDA? We introduce *the first of its kind* KFDA-based brain tissue classification algorithm to the NeuroImaging field that explores the structure of GM, WM and CSF clusters, reveals their non-linearity in the original space and exploits this non-linearity for improved classification [10]. KFDA [11] is particularly useful for the separation of input data into classes when their histogram distributions overlap as is the case with the MR pediatric dataset. KFDA attempts to make the image data more separable by non-linearly mapping them from the original space to an abstract feature space and classify them via optimal discriminant hyperplanes.

The KFDA-based approach is natural for the identification of PV voxels that lie near the boundaries between tissue types. Since KFDA finds complex decision surfaces that best separate the data into GM, WM and CSF in the original space then overlapping subsets (e.g., WM voxels trapped in the intensity range of GM) and class cluster outliers located near these separating surfaces identify the voxels with significant PV effect. KFDA is related to kernel-based classifiers such as the Support Vector Machine (SVM) approach [13]. The superior performance of KFDA over SVM as shown in [11] can be explained by the fact that KFDA uses all training samples to compute the discriminant function, not only the ones that lie closest to the decision surface, i.e. the Support Vectors. The appealing features of KFDA include:

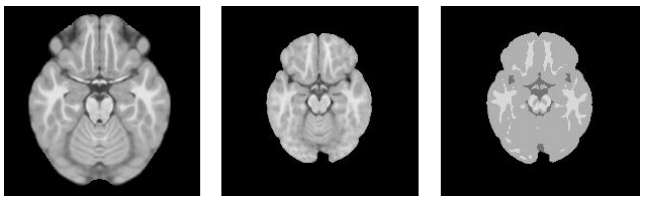
- KFDA is vector-valued, i.e. applied to multi-channel data
- Non-linear generalization of LDA (Linear Discriminant Analysis) that implies a higher prediction accuracy.
- Precision; algebraic formulation of the maximization of the discriminant criterion provides an exact solution.

– Minimal dependency on parameters (unlike SVM whose performance depends also on the number of support vectors and training samples). The parameters that are used in KFDA define kernel functions.

Dataset. The pediatric dataset (NIHPD) collected by the National Institutes of Health (NIH) [15] consists of 72 healthy subjects aged 10 days to 4.5 years scanned repeatedly at quarterly intervals. Imaging data includes structural MRI (T1w, T2w, PDw). Data were acquired on a 1.5 T Siemens Sonata scanner with a $1 \times 1 \times 3$ mm spatial resolution. MR brain scans were corrected for image intensity non-uniformity [16] and registered to the MNI stereotaxic space using spatial normalization [17]. The data were resampled to 1 mm^3 grid using tricubic interpolation. T1w, T2w and PDw average atlases have been created for important developmental age ranges for the NIHPD data[1] [12].

2 KFDA-Based Algorithm

1. ***Initialization.*** We started with a template for the oldest age range (44-60 months) and transferred GM, WM and CSF probability maps known for the age range 4.5 to 8.5 years [12] [2] onto the oldest pediatric template via registration *mni_autoreg* [17] of the T1w template (4.5 to 8.5 years) (see Fig. 1.a) with the T1w template (44-60 months) (see Fig. 1.b). The *mni_autoreg* procedure estimates a 3D non-linear deformation field iteratively in a multiscale hierarchy, i.e. by matching blurred template volumes and subsequent refining of a resulting displacement field. Hard labeling of the template for the age range of 44 to 60 months is then used as the best guess for initialization (see Fig. 1.c).



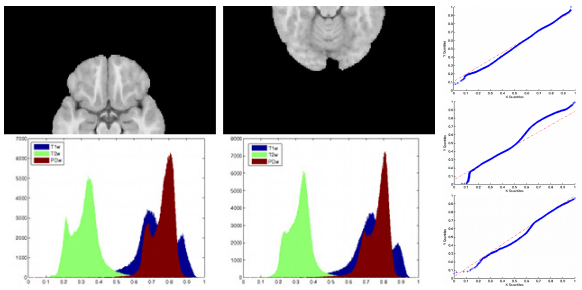
(a) (b) (c)

Fig. 1. (a) T1w pediatric template (4.5-8.5 years), (b) T1w pediatric template (44-60 months), (c) Hard labels of the template (b) obtained from tissue probability maps of older brains registered with (b).

2. ***Preliminary Quantile Analysis.*** In the posterior brain, the MRI signal in WM tends to weaken towards the occipital lobe introducing more uncertainty to

[1] The age-dependent pediatric atlas is available for download at `http://www.bic.mni.mcgill.ca/ServicesAtlases/NIHPD-obj2`

[2] This probabilistic brain atlas was obtained from 82 normal subjects within the age range 4.5 to 8.5 years using an unsupervised genetic tissue classification algorithm.



(a) (b) (c)

Fig. 2. T1w, T2w and PDw intensity histograms in (a) the anterior and (b) posterior parts of the 3D template (44-60 months), (c) quantile-quantile plots of the anterior (X-axis) versus posterior (Y-axis) intensity samples in 3D (from top to bottom) T1w, T2w and PDw templates

GM/WM boundary location. This results in a greater overlap of GM and WM intensity histograms as shown in Fig. 2.a-b. Scatter plots of quantiles computed from samples of grey levels in the anterior and posterior parts of the 3D template brain suggest that they come from different probability distributions (see Fig. 2.c). Therefore, we explore tissue cluster structures in each of the brain halves.

3. ***3D Brain partitioning.*** To proceed with KFDA implementation in MATLAB, namely, to solve an eigen-value problem in a high-dimensional feature space the brain partitioning into subvolumes is needed. Due to MATLAB limitations on the maximum matrix size and available system memory it is not possible to carry out computations for each interior brain half (containing $\approx 300,000$ voxels). Therefore, given a vector-valued image function

$$\boldsymbol{I}(i,j,k) = (T1w(i,j,k), T2w(i,j,k), PDw(i,j,k)),$$

where $1 \leqslant i \leqslant M,\ 1 \leqslant j \leqslant N,\ 1 \leqslant k \leqslant L$ are voxel coordinates of the interior brain, we partition each brain half into subsets of K slices in the axial direction. Due to insufficient system memory we have chosen $K = 3$. That is, we have $\left[\frac{L}{2}\right]$ subvolumes

$$\{\boldsymbol{I}\}_k = \{(T1w(i,j,k+l-1), T2w(i,j,k+l-1), PDw(i,j,k+l-1))\}_{l=1}^{3},$$

where $k = 1, 3, 5, ..., L-2$.

4. ***Kernel transformation of the data and optimal projection in the feature space.*** We partition the non-binary brain tissue classification problem into two two-class problems, namely, separation of the image data into G+W matter and CSF and then separation of G+W matter into GM and WM. We consider each vector-valued intensity at the interior brain voxel as a training sample. Given the brain subvolume $\{\boldsymbol{I}\}_k$ with M_1 labeled training samples we implicitly transform them to the M_1-dimensional feature space $\mathcal{F}$ with a non-linear map

Φ. We then calculate the direction $\boldsymbol{w}$ of maximal information discrimination in $\mathcal{F}$ [11] and project the mapped data $\boldsymbol{\Phi}(\boldsymbol{I})$ onto the vector $\boldsymbol{w}$

$$\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{I}) = \sum_{m=1}^{M_1} \alpha_m k(\boldsymbol{I}_m, \boldsymbol{I}) + \beta, \tag{1}$$

where β is an offset and $k(\boldsymbol{I}_m, \boldsymbol{I}) = \boldsymbol{\Phi}(\boldsymbol{I}_m) \cdot \boldsymbol{\Phi}(\boldsymbol{I})$ is the kernel function that computes a dot product in $\mathcal{F}$. Experimental work with various kernel functions shows that a sigmoid kernel function $k(\boldsymbol{I}_m, \boldsymbol{I}) = \tanh(a({\boldsymbol{I}_m}^T \cdot \boldsymbol{I}) + b)$ yields the best separation into G+W matter and CSF, and a polynomial of a degree 3 $k(\boldsymbol{I}_m, \boldsymbol{I}) = ({\boldsymbol{I}_m}^T \cdot \boldsymbol{I} + b)^3$ or higher best separates GM and WM.
Figures 3.b and 4.b show optimal projections of G+W matter (in red) and CSF (in blue) classes and GM (in blue) and WM (in red) classes correspondingly (according to their initial classifications). In both Figures, X-axis represents column-wise enumeration of the interior brain voxels from 1 to M_1 and Y-axis represents the projected values $\boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{I}_i)$, $1 \leqslant i \leqslant M_1$. When calculated with the offset β they are positive for one class and negative for another.

Classification into G+W Matter and CSF. For the anterior subvolume displayed in Fig. 3 KFDA identified 22 CSF and 384 G+W matter outliers shown in cyan and green, respectively in Fig. 3.c. Their spatial positions in stereotaxic space (see Fig. 3.a) suggest that they are likely to be PV voxels. To determine the dominant tissue type for each of these outliers we split the projected data into testing and training sets. Namely, the outliers form the testing set and the rest of the projected data forms the training set.

Using Mahalanobis distance, KFDA predicted CSF membership for all 384 G+W matter outliers from the classified training set (see Fig. 3.c). As a result, a new classification detects more CSF (see Fig. 3.d).

A separating surface corresponding to this new classification is shown in Fig. 4.e with G+W matter and CSF intensities depicted in red and blue, respectively. The template decision surface is used for the subject classification into G+W matter and CSF shown in cyan and magenta, respectively in Fig. 4.e. Notice the complexity of the separating surfaces displayed in Fig. 4.e-f due to the fact that all training samples are used to compute them.

Classification into GM and WM. Unlike the case with G+W matter and CSF clusters, both GM and WM distributions of the projected data contain only a few if any outliers. There is a significant number of GM voxels trapped in the negative range of WM distribution as seen from Fig. 4.b. More precisely, KFDA has identified 1647 overlapping GM voxels and 74 overlapping WM voxels shown in cyan and green correspondingly in Fig. 4.b. Fig. 4.a suggests that these voxels are likely to contain a significant PV effect. Treating them as testing samples we predict their labels via a kNN classifier ($k = 8$ neighbours) from a training set comprised of the rest of the projected subvolume in $\mathcal{F}$ (see Fig. 4.c).

The separating polynomial surface corresponding to the KFDA classification of the posterior subject subvolume into GM (in magenta) and WM (in cyan) is shown in Fig. 4.g.

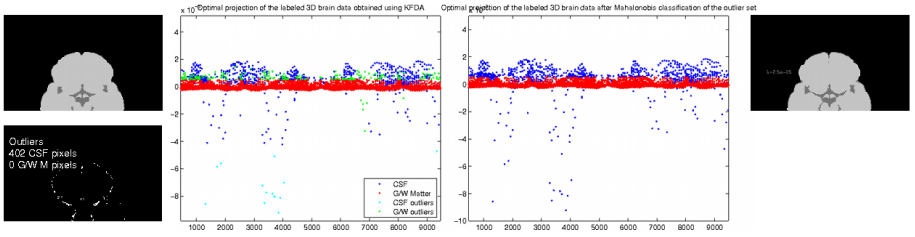


Fig. 3. (a) top: Labeled input data (one of the three anterior slices is shown) from the template (44-60 months), bottom: spatial location of outliers in stereotaxic space, (b) data projection onto $\boldsymbol{w}$ in $\mathcal{F}$: G+W (in red) and CSF (in blue), (c) Mahalanobis classification of the outliers, (d) KFDA classification into G+W matter and CSF

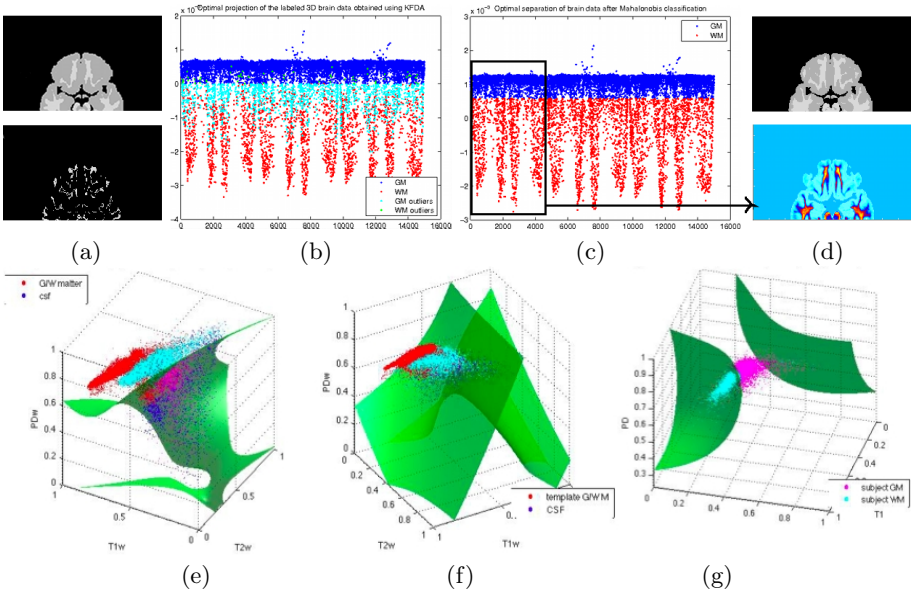


Fig. 4. (a) top: Initial GM and WM labels of G+W matter (one of the three anterior slices is shown) in the template subvolume (44-60 months), bottom: overlapping voxels in the stereotaxic space, (b) optimal projection in the feature space, (c) kNN classification of GM and WM overlapping voxels, WM (in red), GM (in blue), (d) top: KFDA classification into GM and WM, bottom: An image display of projected anterior subvolume (one slice out of $K = 3$ is shown) in $\mathcal{F}$ contained in a rectangular area (c). Red peaks in (c) correspond to interior WM voxels. Optimal decision surfaces for the CSF and G+W matter (e) in the anterior part, (f) in the posterior part, (g) optimal separation of the posterior subject data into GM and WM.

3 Spatial Regularization

To increase robustness to misclassification, we introduce a spatial regularization term that penalizes local kernel projected intensity differences in $\mathcal{F}$. We define matrix H that describes local relationships between the interior brain voxels as follows

$$H_{ij} = \begin{cases} 1, & \text{if voxels } i \text{ and } j \text{ are neighbours } ((i,j) \text{ is an edge}); \\ -d_{ij}, & \text{if } i = j, \text{ the degree of vertex (voxel) } i; \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{Then for } \forall \boldsymbol{V} \in R^{M_1} \quad \boldsymbol{V}^T H \boldsymbol{V} = -\sum_{(i,j)\in E} (V_i - V_j)^2, \tag{2}$$

where E is an edge set comprised of edges $\{V_i, V_j\}$.
Let $\boldsymbol{V} = \boldsymbol{w} \cdot \boldsymbol{\Phi}(\boldsymbol{I})$ be the kernel projection of the input data $\boldsymbol{I}$ onto the optimal direction $\boldsymbol{w}$ in $\mathcal{F}$. $\boldsymbol{V}$ can be rewritten as $\boldsymbol{V} = \sum_{i=1}^{M_1} \alpha_i k(\boldsymbol{I_i}, \boldsymbol{I})$ due to the expansion of $\boldsymbol{w} = \sum_{i=1}^{M_1} \alpha_i \boldsymbol{\Phi}(\boldsymbol{I}_i)$ in $\mathcal{F}$ spanned by the mapped training samples $\boldsymbol{\Phi}(\boldsymbol{I}_i)$. We modify the KFDA optimality criterion by adding the penalty term of the form $V^T H V = \alpha^T K H K^T \alpha$, where K is the kernel matrix of size $M_1 \times M_1$

$$\hat{\alpha} = \arg\max_{\alpha} \left(\frac{\alpha^T M \alpha + \lambda \alpha^T K H K^T \alpha}{\alpha^T N \alpha} \right). \tag{3}$$

Here, M is a between-class covariance matrix and N is a within-class covariance matrix in $\mathcal{F}$ (see [11], [14] for details). In this setup the penalty function forces misclassified voxels closer to another class cluster centroid. The problem (3) can be solved by computing a leading eigen-vector of $N^{-1}(M + \lambda K H K^T)$.

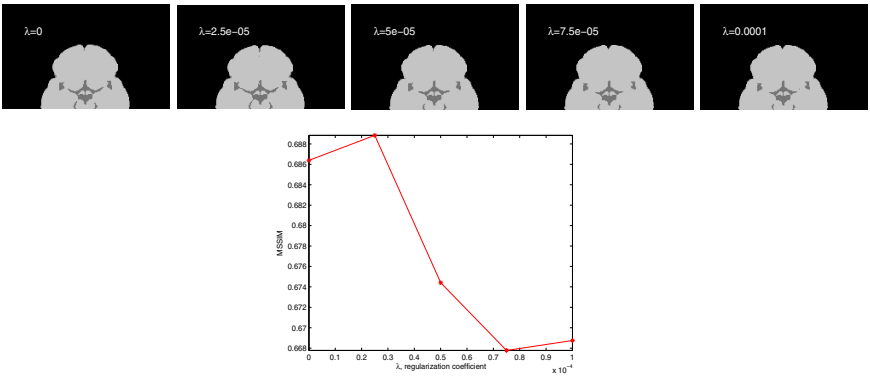


Fig. 5. Upper panel: anterior template brain classified into G+W M and CSF for $\lambda_i = 0.000025 \cdot i,\ i = 0, ..., 4$; Lower panel: MSSIM between T1w template and each of the classified anterior brains shown in the upper panel

A modified version of the KFDA criterion (3) depends on the value of the regularization coefficient λ. The upper panel in Fig. 5 shows the influence of λ-coefficient on the quality of segmentation into G+W M and CSF as it increases by an increment of 0.000025 from 0 to 0.0001. By a visual inspection, the classification corresponding to $\lambda = 0.000025$ is most plausible as the CSF pattern appears to be most connected compared to that with other λ-values. This choice was operationalized with the SSIM metric described below.

4 Objective Quality Evaluation via SSIM

For the automatic control of λ-parameter we need a quantitative assessment of segmentation quality. Such measures to MR brain segmentation as the Jaccard coefficient and the Dice coefficient are commonly used, however, they rely on knowledge of a reference segmentation. We seek a similarity measure to evaluate λ-dependent classifications in the absence of a ground truth and use the Structural Similarity Index Measure (SSIM) [18], [19]. The SSIM is an objective similarity metric that quantifies the degree of structural similarity between ideal and distorted images. It is based on the assumption that the Human Visual System (HVS) is an optimal extractor of structural information from images.

We evaluated the performance of our classification algorithm relying on the Computer Vision hypothesis that the HVS focuses on image components with high information content [20]. In our case, these image components are WM/GM and G+W matter and CSF boundaries. We evaluated how well these boundaries are captured by our algorithm versus the boundaries that we can visually extract from T1w data (we use one imaging modality for simplicity).

We created classified brain subvolumes in the form of mean T1w intensity values for the two tissue types. We computed the SSIM between each classified and T1w brain slices and the mean SSIM (MSSIM) defined by

$$MSSIM = \frac{1}{M_1}\sum_{i=1}^{M_1} SSIM(x_i, y_i),\ \ SSIM(x_i, y_i) = l(x_i, y_i)\cdot c(x_i, y_i)\cdot s(x_i, y_i),$$

x_i and y_i are local image patches[3] and $l(x_i, y_i)$, $c(x_i, y_i)$, $s(x_i, y_i)$ are the luminance, contrast and structure comparison measures defined by

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1};\ \ c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}; s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}.$$

Here, $\mu_x(\mu_y)$, $\sigma_x(\sigma_y)$ and σ_{xy} represent the local mean, standard deviation and cross-correlation estimates, respectively, and C_1, C_2, C_3 are small constants [18].

Shown in the lower panel of Fig. 5 is the MSSIM computed between the T1w and classified anterior template subvolumes and plotted against the values of λ. It achieves its maximum at $\lambda = 0.000025$ as expected. Thus, by choosing λ-value corresponding to the largest MSSIM we are able to automatically guide the classification procedure.

[3] a sliding window that moves across the entire brain slice pixel by pixel. For the MSSIM the background patches have been excluded.

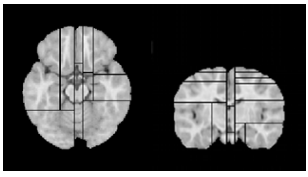
5 Results

We compared classification results obtained by prior hard labeling, KFDA-based and EM approaches using MSSIM. Figures 7 and 8 show single slices from the classified template (44-60 months) and subject brain subvolumes obtained by joining overlapping anterior and posterior parts. Namely, if $\boldsymbol{I_1}(i,j,k),\ i \in \{1,2,3,...,\left[\frac{N}{2}\right]+2\}$ is an anterior brain, and $\boldsymbol{I_2}(i,j,k),\ i \in \{\left[\frac{N}{2}\right]-1,...,N\}$ is a posterior brain, where $y = \left[\frac{N}{2}\right]$ is the middle plane, then the whole brain image function $\boldsymbol{I}(i,j,k)$ is defined as follows

$$\boldsymbol{I}(i,j,k) = \begin{cases} \boldsymbol{I_1}(i,j,k), & \text{for } i \in \{1,2,3,...,\left[\frac{N}{2}\right]\}, \\ \boldsymbol{I_2}(i,j,k), & \text{for } i \in \{\left[\frac{N}{2}\right]+1,...,N\}. \end{cases}$$

Note that in order to preserve neighbourhood relations of the middle plane voxels in anterior and posterior parts, we defined I_1 and I_2 as subsets with an overlapping region $R(i,j,k),\ i \in \{\left[\frac{N}{2}\right]-1,..,\left[\frac{N}{2}\right]+2\}$.

Remark. In order to better accommodate grey level intensity inhomogeneities present in brain tissues, we intend to optimally partition the brain into regions that differ significantly in average intensity values. Shown in Fig. 6 are transverse and coronal slices of the template brain subdivided into parallelepipeds using a binary space partition. Having created and classified overlapping parallelepipeds, 3D image stitching can then be performed via simulated annealing.



(a) (b)

Fig. 6. 3D template brain partitioning based on mutual information maximization between the intensity histogram bins and the regions of the subdivided image: (a) transverse slice 37 (out of 105), (b) coronal slice 110 (out of 235). The total number of the brain regions is 32.

Fig. 7.a-c show that the CSF structure captured by KFDA is more similar to the one seen in T1w. The comparison of MSSIMs given in Fig. 7.b-c suggests that our proposed algorithm improves CSF detection.

The comparison of the classified WM patterns (see Fig. 7.d-e) with the one seen in T1w (see Fig. 7.a) and of their respective MSSIMs for the template shows that the proposed algorithm also improves classification into GM and WM. EM algorithm with a prior seen in Fig. 7.f yields a reasonable estimate of CSF and a significant underestimate of WM. MSSIM comparison of the GM/WM classified results demonstrates a superior performance of KFDA over EM algorithm. Seen

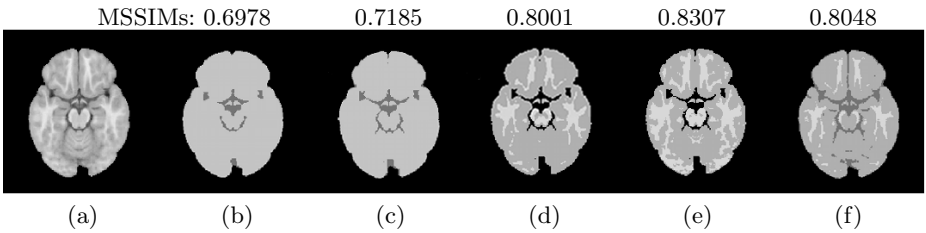


Fig. 7. (a) T1w template (44-60 months) and its classification into G+W matter and CSF: (b) prior, (c) KFDA; into GM and WM: (d) prior, (e) KFDA, (f) EM classification into GM, WM and CSF with a prior

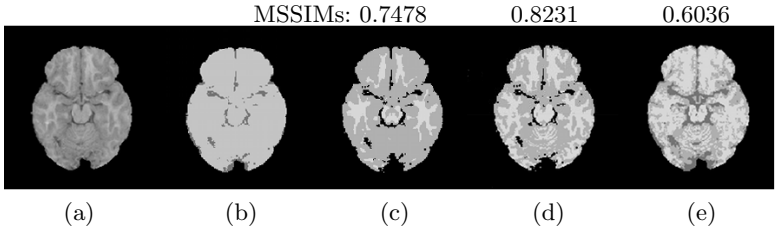


Fig. 8. (a) T1w image of a 4.5 year old subject and its classification into G+W matter and CSF: (b) KFDA; into GM and WM: (c) prior, (d) KFDA, (e) EM with a prior.

in Fig. 8.a is a T1w scan of a 4.5 year old subject. We generated classification into G+W matter and CSF by treating the entire subject data as a testing set and by its kernel projection onto $\boldsymbol{w}$ in $\mathcal{F}$ spanned by the template samples $\Phi(\boldsymbol{I}_i)$. The resulting classification is displayed in Fig. 8.b.
The comparison of a WM pattern seen in Fig. 8.a with the classified WM seen in Fig. 8.c-e demonstrates the remarkable capability of KFDA to reveal a complex WM structure given a poor GM/WM contrast. The EM algorithm with a prior applied to the vector-valued subject subvolume tends to overestimate WM and CSF (see Fig. 8.e). The comparison of MSSIMs given in Fig. 8.c-e shows that KFDA yields the most similar WM pattern to the one extracted by our visual system.

6 Conclusion

We have developed a novel and elegant KFDA-based algorithm for automatic brain tissue classification. It is a vector-valued non-parametric approach that relies on prior hard labels of tissue types for initialization. The proposed algorithm takes into account spatial correlations between interior brain intensities, identifies voxels with PV effect, predicts the dominating tissue types in the PV set and constructs complex optimal decision surfaces precisely.

In this work, we classified the oldest template (44-60 months) of this dataset and showed how a subject chosen from the same age range can be labeled us-

ing KFDA-classified template. SSIM comparison of GM, WM and CSF patterns detected by KFDA and EM approaches showed a superior performance of the latter. Our next incentive is to apply the KFDA-based technique for the classification of age-dependent templates for earlier age ranges.

Acknowledgement. This research was supported by the Montreal Neurological Institute in the form of Jeanne Timmins Costello postdoctoral fellowship. The authors would like to thank their colleagues from the University of Waterloo (Canada), Dr. Zhou Wang, a Prof. in the ECE Department, and Dr. Edward Vrscay, a Prof. in the Department of Applied Mathematics, for the discussion and insightful comments on the proposed KFDA-based methodology.

References

1. Evans, A.C., Brain Development Cooperative Group: The NIH MRI study of normal brain development. NeuroImage 30, 184–202 (2006)
2. Zijdenbos, A.P., Forghani, R., Evans, A.: Automatic Quantification of MS Lesions in 3D MRI Brain Data Sets: Validation of INSECT. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 439–448. Springer, Heidelberg (1998)
3. Warfield, S.: Fast kNN classification for multichannel image data. Pattern Recogn. Lett. 17(7), 713–721 (1996)
4. Zijdenbos, A.P., Dawant, B.M., et al.: Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. IEEE Trans. Med. Imag 21(10), 1280–1291 (1994)
5. Wells, W.M., Kikinis, R., Grimson, W.E.L., Jolesz, F.: Adaptive Segmentation of MRI Data. IEEE Trans. Med. Imag. 23, 429–442 (1996)
6. Pohl, K.M., Bouix, S., Kikinis, R., Grimson, W.E.L.: Anatomical Guided Segmentation with non-Stationary Tissue Class Distributions in an Expectation-Maximization Framework. In: IEEE Int. Symposium on Biomed. Imag., Arlington, VA, pp. 81–84 (2004)
7. Grau, V., Mewes, A.U.J., et al.: Improved Watershed Transform for Medical Image Segmentation Using Prior Information. IEEE Trans. Med. Imag. 23(4), 447–458 (2004)
8. Bouix, S., Martin-Fernandez, M., Ungar, L., et al.: On Evaluating Brain Tissue Classifiers without a Ground Truth. NeuroImage 36, 1207–1224 (2007)
9. Tohka, J., Zijdenbos, A., Evans, A.: Fast and Robust Estimation for Statistical Partial Volume Models in Brain MRI. NeuroImage 23, 84–97 (2004)
10. Portman, N., Evans, A.: Novel Vector-Valued Approach to Automatic Brain Tissue Classification. Poster 6488, 18th Annual Meeting of the OHBM, Beijing (2012)
11. Mika, S., Ratsch, G., Weston, J., et al.: Fisher Discriminant Analysis with Kernels. In: Neural Networks for Signal Processing IX: Proc. of the 1999 IEEE Signal Proc. Soc. Workshop, pp. 41–48 (1999)
12. Fonov, V., Evans, A.C., Botteron, K., et al.: Unbiased Average Age-Appropriate Atlases for Pediatric Studies. Neuroimage 54(1), 313–327 (2011)
13. Vapnik, V.N.: Statistical learning theory. John Wiley & Sons (1998) (manuscript)
14. Baudat, G., Anouar, F.: Generalized Discriminant Analysis Using a Kernel Approach. Neural Computation 12(10), 2385–2404 (2000)

15. Almli, C.R., Rivkin, M.J., McKinstry, R.C., Brain Development Cooperative Group: The NIH MRI study of normal brain development (Objective-2): Newborns, infants, toddlers, and preschoolers. NeuroImage 35(1), 308–325 (2007)
16. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A Non-Parametric Method for Automatic Correction of Intensity Non-Uniformity in MRI Data. IEEE Trans. Med. Imag. 17, 87–97 (1998)
17. Collins, D.L., Neelin, P., Peters, P.M., Evans, A.C.: Automatic 3D Intersubject Registration of MR Volumetric Data in Standardized Talairach Space. J. Comput. Assist. Tomogr. 18(2), 192–205 (1994)
18. Wang, Z., Bovik, A.C.: A Universal Image Quality Index. IEEE Signal Processing Letters 9, 81–84 (2002)
19. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multi-scale Structural Similarity for Image Quality Assessment. In: IEEE Proc. Asilomar Conf. Signals, Syst.,Comput., pp. 1398–1402 (2003)
20. Wang, Z., Bovik, A.C., Sheikh, H.R.: Image Quality Asessment: From Error Visibility to Structural Similarity. IEEE Trans. Image Proc. 13(4), 600–612 (2004)

Atlas-Based Whole-Body PET-CT Segmentation Using a Passive Contour Distance

Fabian Gigengack[1,2], Lars Ruthotto[3], Xiaoyi Jiang[2], Jan Modersitzki[3], Martin Burger[4], Sven Hermann[1], and Klaus P. Schäfers[1]

[1] European Institute for Molecular Imaging (EIMI), University of Münster, Germany
[2] Department of Mathematics and Computer Science, University of Münster, Germany
[3] Institute of Mathematics and Image Computing, University of Lübeck, Germany
[4] Institute for Computational and Applied Mathematics, University of Münster, Germany

Abstract. In positron emission tomography (PET) imaging, the segmentation of organs is necessary for many quantitative image analysis tasks, e.g., estimation of individual organ concentration or partial volume correction. To this end we present a fully automated approach for whole-body segmentation which enables large-scale and reproducible studies. The approach is based on joint segmentation and atlas registration. The classical active contour approach by Chan and Vese is modified to a novel *passive contour* energy term with implicitly incorporated information about shape and location of the organs. This new energy is added to a registration functional which is based on both functional (PET) and morphological (CT) data. The proposed method is applied to medical data, given by 13 PET-CT data sets of mice, and quantitatively compared to manually drawn VOIs. An average Dice coefficient of 0.73 ± 0.10 for the left ventricle, 0.88 ± 0.05 for the bladder, and 0.76 ± 0.07 for the kidneys shows the high accuracy of our method.

Keywords: Segmentation, Active Contour, Passive Contour, Registration, Atlas, PET-CT, Whole-Body.

1 Introduction

Positron emission tomography (PET) is widely used in medical imaging to assess functional information in the body. However, quantitative evaluation of PET images is challenging due to the rather limited spatial resolution and low signal-to-noise ratio which makes the segmentation of organs necessary for various applications. Estimating organ concentration in biodistribution studies [6], [9], or analyzing organ specific diseases such as myocardial infarction demands for an adequate whole-body segmentation. In addition, organ segmentation is mandatory for many partial volume correction techniques [15]. To this end we developed a general approach for whole-body segmentation based on joint segmentation and registration.

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 82–92, 2013.

1.1 Related Work

Many approaches originating from computer vision are transferred to medical imaging as they are well understood and, at the same time, also efficiently applicable to volumetric (3D) medical images. A popular approach in computer vision for automatic segmentation is active contours as introduced by Chan and Vese [2]. The method was successfully applied to medical imaging based on brain MRI data [3]. The main idea of active contours is also exploited in our work, but in a reversed interpretation, cf. Sec. 2.

There is a large demand for automatic segmentation in medical imaging as the manual segmentation of organs is time-consuming for 3D data sets. Further, inter- and intra-observer variability can have a high impact. This is why manual segmentation is inapplicable for large-scale and reproducible studies. We restrict the following discussion to related literature on segmentation of PET and CT and joint registration and segmentation.

An automated method for whole-body segmentation in Micro-CT data of mice was introduced by Baiker et al. [1]. The approach consists of a model-based registration with a subsequent intensity-based registration. They achieved high accuracies for skin and skeleton. However, they did not report results for inner organs which are the focus of this work. This might be due to the low soft tissue contrast of the CT images which makes the localization of inner organs challenging. We overcome this limitation (inter alia) by using functional information in terms of PET images (and additional CT images).

Wang et al. presented a registration approach based on a statistical shape model for small-animal PET segmentation [13]. High uptake organs guide the registration using a conditional Gaussian model and allow good estimates for low uptake organs as well. However, for the labeling of organs the method requires user interaction.

Recently various techniques were published combining registration and segmentation. A taxonomy on this topic is given in [8]. A method, which is basically similar to our proceeding, was presented by Yezzi et al. [14]. They propose a variational framework that uses active contours for segmentation with a simultaneous registration of features. The level-set based segmentation separates only one object from the background which makes this method inapplicable for multiple organ segmentation tasks. Further, only rigid and affine transformations were practically explored.

2 Methods

In this paper we present a novel atlas-based segmentation approach. The general scheme is illustrated in Fig. 1. Given a pair of spatially aligned PET and CT images (real data on the left of Fig. 1) of the same subject, we follow a two-step strategy. After aligning the atlas (atlas data on the right of Fig. 1) and the real data with an affine transformation, a tailored registration functional with joint segmentation is minimized. Three distance terms drive the registration: 1. Distance of the atlas CT and real CT, 2. Distance of the atlas PET and real PET,

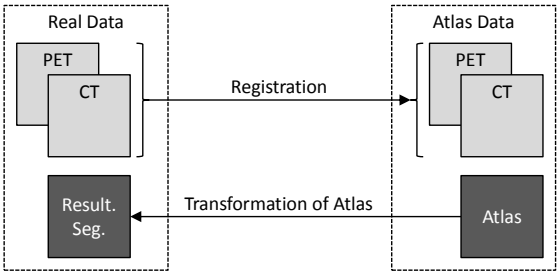


Fig. 1. General scheme: The inverse of the estimated transformation is applied to the atlas to segment the real data

3. Segmentation distance motivated by Chan and Vese [2]. Instead of matching a contour to the data, the novel segmentation distance is used to optimize for the transformation that aligns the data best to the (passive) contours. This turns around the interpretation of standard active contours models. Finally, the atlas organ definitions are transformed with the inverse transformation and yield the resulting segmentation of the real data, cf. bottom of Fig. 1.

In particular, we address the following points:

1. Transition of 2D active contours to 3D passive contours for medical image segmentation
2. Fully automation to make large-scale studies possible (user interaction is time-consuming)
3. Non-rigidity of atlas-based whole-body segmentation
4. Multimodality treatment (function and morphology)
5. Handling of multiple organs for joint registration and segmentation

2.1 Joint Passive Contour Segmentation and Registration

As a technical preprocessing step, a rough alignment of the atlas dataset and the real dataset is performed by matching the atlas CT to the real CT with an affine transformation to overcome differences in the orientation, scaling, and translation. As both images are of the same modality we choose the sum of squared differences (SSD) distance measure.

To overcome anatomical variations of organs, the information of the PET and the CT images is used simultaneously in a joint registration functional. Hence, anatomical and functional information is exploited at the same time. In addition, we include a novel segmentation distance term into the functional, inspired by Chan and Vese [2]. The Chan-Vese distance measures the in-class variance according to the atlas organ definitions. We derive the complete registration model by first looking at standard image registration for the CT images.

For the alignment of the CT images, the real data $\mathcal{T}_{CT} : \Omega \to \mathbb{R}$ (template image) is registered to the atlas CT image $\mathcal{R}_{CT} : \Omega \to \mathbb{R}$ (reference image), where

$\Omega \subset \mathbb{R}^3$ is the image domain. The output of the registration is a transformation $y : \mathbb{R}^3 \to \mathbb{R}^3$ representing point-to-point correspondences between $\mathcal{T}_{CT}$ and $\mathcal{R}_{CT}$. To find y, the following functional has to be minimized

$$\min_y \left\{ \mathcal{D}^{\text{SSD}}(\mathcal{T}_{CT} \circ y,\ \mathcal{R}_{CT}) + \alpha_{\mathcal{S}} \cdot \mathcal{S}(y) \right\} . \tag{1}$$

$\mathcal{D}^{\text{SSD}}$ is the SSD distance functional and $\alpha_{\mathcal{S}} \in \mathbb{R}^+$ is a weighting factor of the regularization functional $\mathcal{S}$. By using regularized spline image interpolation we reduce artifacts in the PET images which justifies the usage of the SSD measure.

We assume that the PET and corresponding CT measurement approximately share the same geometry and hence y can be used to align both modalities. In practice the images provide complementary information which motivates the exploration of both modalities in a joint registration functional. The CT images guide the registration whereas the PET images provide important information in soft tissue regions. As the scanned mice are anesthetized the spatial variations are kept to a minimum. However, changes due to, e.g., bladder filling, are possible.

Our joint registration functional is an extension of (1) by adding a term for the PET data and an additional passive contour term $\mathcal{D}_{PC}$

$$\begin{aligned} \min_y \{\ \alpha_{CT} \cdot \mathcal{D}^{\text{SSD}}(\mathcal{T}_{CT} \circ y,\ \mathcal{R}_{CT}) + \alpha_{PET} \cdot \mathcal{D}^{\text{SSD}}(\mathcal{T}_{PET} \circ y,\ \mathcal{R}_{PET}) \\ + \alpha_{PET}^{PC} \cdot \mathcal{D}_{PC}(\mathcal{T}_{PET} \circ y,\ A) + \alpha_{\mathcal{S}} \cdot \mathcal{S}(y)\ \} , \end{aligned} \tag{2}$$

where $\mathcal{T}_{PET}, \mathcal{R}_{PET} : \Omega \to \mathbb{R}$ are the real PET image and the atlas PET image. $\alpha_{CT}, \alpha_{PET}, \alpha_{PET}^{PC}, \alpha_S \in \mathbb{R}^+$ are weighting factors for the individual distance functionals and are discussed later. $\mathcal{D}_{PC}$ is the passive contour distance and A denotes the delineation of the atlas organs.

Passive Contour Distance. Let us now derive the passive contour term $\mathcal{D}_{PC}$. The classical Chan-Vese functional [2] is defined as follows

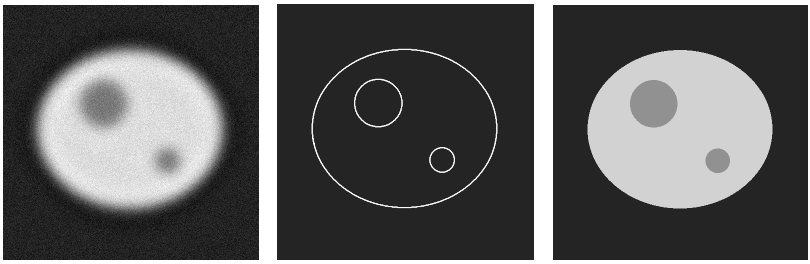
$$\mathcal{CV}(C) = \int_{C^{in}} (\mathcal{T}(x) - \mu(\mathcal{T}, C^{in}))^2\ dx + \int_{C^{ex}} (\mathcal{T}(x) - \mu(\mathcal{T}, C^{ex}))^2\ dx\ . \tag{3}$$

The function μ computes an average value of $\mathcal{T}$ (we omit the subscript for simplicity) according to the interior (C^{in}) respectively the exterior (C^{ex}) of the contour C. The aim is to find the (active) contour C that minimizes the energy $\mathcal{CV}(C)$. We can rewrite this formulation as a functional of the transformation y

$$\mathcal{CV}(y) = \int_{y(\Omega)} (\mathcal{T}(x) - \mu(\mathcal{T},\ A \circ y;\ x))^2\ dx\ . \tag{4}$$

$\mu(\mathcal{T},\ A \circ y;\ \cdot\ )$ is constant inside each organ containing the average intensity of $\mathcal{T}$ over the respective segment. A simple 2D example to illustrate the function μ is given in Fig. 2.

The atlas definitions A in Fig. 2(b) exactly match the contours of the blurred and noisy input image (a). By applying the segmentation function μ we result in a recovered image without noise and blur (c).



(a) Input image $\mathcal{T}$ (b) Atlas definitions A (c) $\mu(\mathcal{T},\ A;\ \cdot\,)$

Fig. 2. Illustration of μ (2D). Given the image $\mathcal{T}$ (left) and the atlas definitions A (middle) we can apply the segmentation function $\mu(\mathcal{T},\ A;\ \cdot\,)$ (right).

By substitution $x \to y(x)$ in (4) we receive

$$\mathcal{CV}(y) = \int_{\Omega} (\mathcal{T}(y(x)) - \mu(\mathcal{T} \circ y,\ A;\ x))^2 \cdot |\det(\nabla y(x))|\, dx\ . \tag{5}$$

The term $\mathcal{D}_{PC}$ is then defined as:

$$\mathcal{D}_{PC}(\mathcal{T} \circ y,\ A) := \frac{1}{2} \int_{\Omega} (\mathcal{T}(y(x)) - \mu(\mathcal{T} \circ y, A; x))^2 \cdot \det(\nabla y(x)) dx\ . \tag{6}$$

Thus, by finding an adequate transformation y the in-class variance of $\mathcal{T} \circ y$ according to the atlas A is minimized. Note that we can drop the absolute value bars for the Jacobian determinant, if the transformation is diffeomorphic, cf. Sec. 2.2.

Instead of adjusting the contour to the data (active contour, analogously deformable templates), the data is adjusted to the contour (passive contour) in our case. Hence we have an optimization problem in the transformation y and not in the contour. This allows us directly to treat multiple segments at once and not only to separate one foreground object from the background (note that there exist also active contour approaches for multiple segments [12]). A further advantage of passive contours compared to active contours is the implicitly incorporated information about shape and location of the organs. In contrast to active contours, contours can not split in multiple objects. Further, active contour approaches require proper initialization. In our case the initialization of the passive contours is directly given by the atlas definitions. Furthermore, the fixed integration domain for segmentation simplifies computations compared to exiting atlas-based segmentation methods.

2.2 Regularization

The non-rigid nature of whole-body segmentation poses challenges to the estimation of the transformation y. To guarantee diffeomorphic transformations

and to be highly robust against noise, we utilize hyperelastic regularization [5]. The regularization functional $\mathcal{S}$ controls changes in length and volume of the transformation y. The weighting factor $\alpha_{\mathcal{S}}$ in (2) is thus a compact notation for the weighting of two regularization terms.

Local adaptive regularization prevents unphysiological contraction or expansion of organs. The organ definitions are given by our atlas organ delineations A. The areas inside organs get a higher volume regularization value $(2 \cdot 10^5)$ compared to normal body tissue $(1 \cdot 10^5)$ which keeps volumetric changes inside organs to a minimum.

2.3 Evaluation

The resulting segmentations are compared to manually drawn VOIs. The Dice coefficient is used to quantitatively compare our segmentation to the ground-truth. For two sets X and Y the Dice coefficient is defined as $D(X,Y) = \frac{2|X \cap Y|}{|X|+|Y|}$.

To assess whether the registration algorithm performs successful or not we analyze the Jacobian determinant. It specifies the volumetric change due to the transformation. A value of 1 represents no volumetric change and a value smaller (greater) than 1 indicates compression (expansion). For positive values the transformations are diffeomorphic. Fig. 4 shows a distribution of the Jacobian for all results.

2.4 Implementation

The implementation is based on the FAIR registration toolbox [10] in MATLAB®. In a first step the images are brought to the same resolution (voxel size of 0.35 mm). We use a multi-level strategy with a scaling of 0.5 between two adjacent levels, starting with a resolution of $16 \times 10 \times 40$ (voxel size of 2.77 mm) and going to a final resolution of $64 \times 40 \times 160$ (voxel size of 0.69 mm). Optimization is performed with a Gauss-Newton scheme in combination with a PCG solver for the linear system of equations, cf. [10]. Spline interpolation is used along with a regularization of the moments. The parameter controlling the amount of regularization is chosen to be 1 for the affine pre-registration and 0.5 for the joint registration. The regularization for the affine pre-registration is higher to reduce the amount of details in the images for the rough alignment.

3 Experimental Results

3.1 Data

This work is based on ^{18}F-FDG-PET/CT data of 13 healthy adult C57/Bl6 mice (without any intervention), representing the most widely used radiotracer and mouse strain in preclinical PET studies.

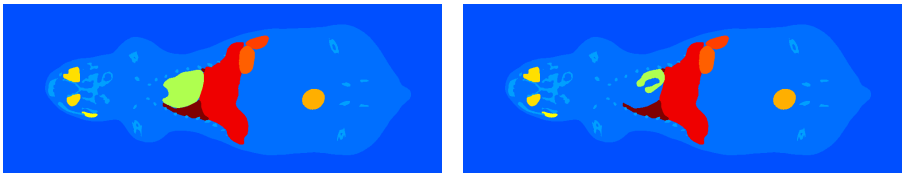
PET experiments were carried out using a high resolution (0.7 mm full width at half maximum) small animal scanner (32 module quadHIDAC, Oxford Positron Systems Ltd., Oxford, UK) with uniform spatial resolution over a large

cylindrical field-of-view (165 mm diameter, 280 mm axial length). Mice were anesthetized with oxygen/isoflurane inhalation (2% isoflurane, 0.4 l/min oxygen) and body temperature was maintained at physiological values by a heating pad. One hour after intravenous injection of 10 MBq ^{18}F-FDG in 100 µl 0.9% saline list-mode data were acquired for 15 min. Subsequently, the scanning bed was transferred to the CT scanner (Inveon, Siemens Medical Solutions, USA) and a CT acquisition with a spatial resolution of ∼80 µm was performed for each mouse after intravenous injection of a contrast agent. The reconstructed image data sets were aligned with a rigid transformation based on extrinsic markers attached to the scanning bed and the image analysis software (Inveon Research Workplace 3.0, Siemens Medical Solutions, USA).

3.2 Atlas

The Digimouse software phantom [4] serves as an atlas. The organ delineations of the pixel atlas are filled with realistic values according to our scanning protocol to construct a pseudo-PET and pseudo-CT phantom image. This has to be done only once in advance. The resulting images are spatially aligned phantom images with a known ground-truth segmentation. No blurring or noise is added to the images.

For the heart, the used ^{18}F-FDG tracer accumulates mainly in the left ventricle. As Digimouse provides only a combined segment for the whole heart (including left and right ventricle and the blood pool) we apply some minor modifications, see Fig. 3. The heart region of the atlas is replaced by a manual threshold segmentation of the left ventricle using the accompanied Digimouse PET data. In addition, the bladder is slightly moved in posterior direction to better fit our real data (this stabilizes the transformation estimation by minimizing the local average deformation). The original image is shown in Fig. 3(a) and the modified version in (b).



(a) Original atlas (b) Modified atlas

Fig. 3. The heart's segmentation (green area) and the bladder (orange area) in the original version of the Digimouse phantom (a) is replaced in the modified atlas (b) to better match the real data

3.3 Results

For the non-parametric registration, the following approach is used to provide meaningful values for the various parameters in (2). An exhaustive parameter

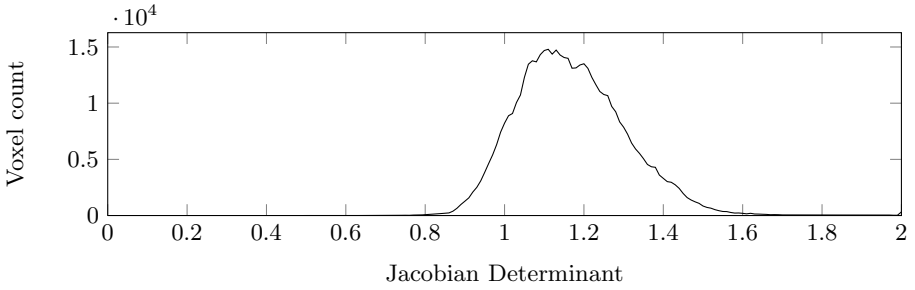


Fig. 4. Summed histograms of the Jacobian determinant of all data set

Table 1. Dice coefficients of the 13 mice for the heart (left ventricle), bladder and kidneys

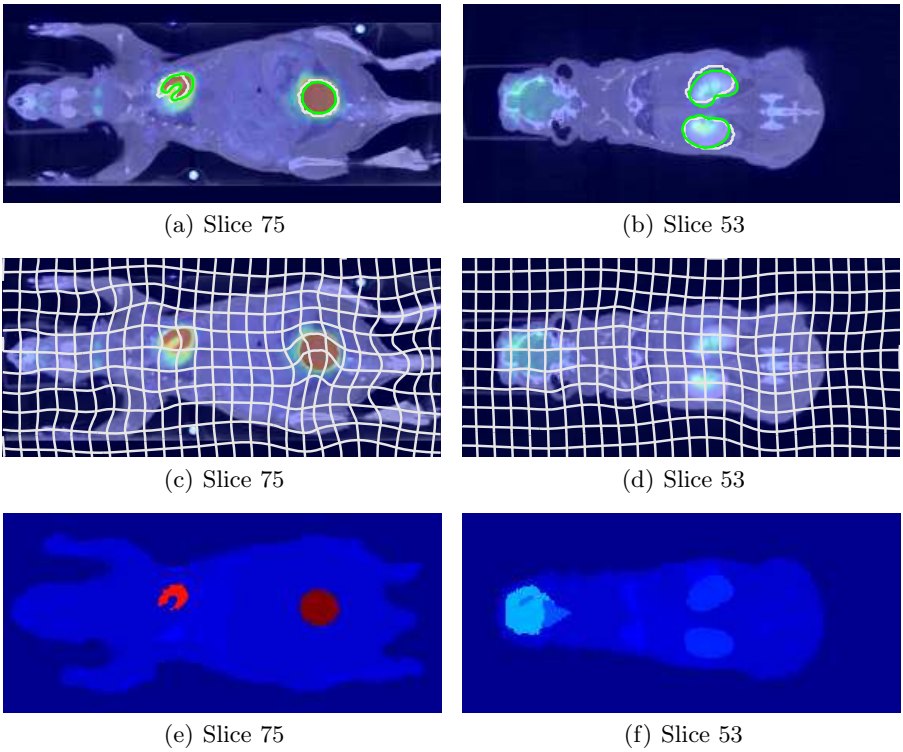
Mouse	1	2	3	4	5	6	7	8	9	10	11	12	13	Avg.	Std.
Heart	0.85	0.84	0.84	0.85	0.79	0.62	0.77	0.72	0.60	0.60	0.68	0.60	0.75	0.73	0.10
Bladder	0.90	0.88	0.78	0.91	0.92	0.88	0.93	0.92	0.93	0.79	0.86	0.82	0.87	0.88	0.05
Kidneys	0.83	0.63	0.80	0.82	0.73	0.65	0.66	0.83	0.80	0.73	0.80	0.78	0.84	0.76	0.07
Avg.	0.86	0.78	0.81	0.86	0.81	0.72	0.79	0.82	0.77	0.71	0.78	0.74	0.82		
Std.	0.04	0.13	0.03	0.05	0.10	0.14	0.14	0.10	0.17	0.10	0.09	0.11	0.06		

search is performed for a randomly selected mouse. For each parameter combination the estimated segmentation is compared to the manual segmentation. The estimation giving the best fit is declared as the optimal parameter set for all experiments as they follow all the same protocol. We found the following optimal parameter set: $\alpha_{CT} = 10$, $\alpha_{PET} = 10$, $\alpha_{PET}^{PC} = 100$. For the hyperelastic regularization we found an optimal weighting for the length term of 1000 and for the volumetric regularization we refer to the regularization paragraph in Sec. 2.2.

For all transformations, the Jacobian determinant is everywhere positive and centered around 1, see Fig. 4. The global minimum is 0.26 and the global maximum is 2.66 which implicates diffeomorphisms. Note that the small shift of the maximum peak to a value greater than 1 in Fig. 4 is due to the affine component of the transformations indicating that the atlas is on average a little bit bigger than the real mice.

For all datasets an average Dice coefficient of 0.73 ± 0.10 could be achieved for the left ventricle, 0.88 ± 0.05 for the bladder, and 0.76 ± 0.07 for the kidneys. The estimated segmentation for one mouse is exemplified in Fig. 5. The Dice coefficients for all analyzed organs and mice can be found in Table 1.

The improvement due to our new passive contour distance can be assessed by setting $\alpha_{PET}^{PC} = 0$ and thus disabling the segmentation input. The objective is to analyze whether the additional passive contour distance can even improve the high accuracy of our multimodal PET-CT registration functional alone. For $\alpha_{PET}^{PC} = 0$, the Dice coefficient for the left ventricle was 0.61 ± 0.12, for the bladder 0.80 ± 0.07, and for the kidneys 0.76 ± 0.08. This means an improvement



(a) Slice 75 (b) Slice 53

(c) Slice 75 (d) Slice 53

(e) Slice 75 (f) Slice 53

Fig. 5. Visualization of 3D registration results for whole-body segmentation. Overlay of 2D projections of PET, CT and contours ((a) heart and bladder, (b) kidneys) and transformation grid y_{opt} ((c), (d)). The estimated segmentations are plotted with white contours and the ground-truth segmentation is shown in green. The estimated contour of the body is plotted for additional visual assessment of the registration accuracy. Slices of the piecewise constant approximations $\mu(\mathcal{T}_{PET} \circ y_{opt}, A)$ are shown in (e) and (f).

of 16% for the left ventricle and 9% for the bladder. We found no improvement for organs with relatively low uptake like the kidneys.

4 Conclusion and Future Work

A novel fully automated approach for whole-body segmentation of PET data is presented in this work. The centerpiece of the proposed joint segmentation and registration method is the introduction of a novel segmentation distance for registration inspired by Chan and Vese [2]. As the interpretation is reversed to active contour models, we denote this as passive contours. Further, the registration is performed based on functional and morphological data simultaneously.

A validation based on the Dice coefficient and the Jacobian determinant demonstrates the high accuracy of our method. Further, the benefit of the

additional Chan-Vese distance, in contrast to multimodal PET-CT registration alone, was shown.

Compared to existing atlas-based segmentation methods the novelty of our passive contours approach is given by implicitly incorporated information about shape and location of the organs. The general shape of the contour can not degrade (e.g. split in multiple objects) as we control the spatial regularity of the guaranteed diffeomorphic transformation by using hyperelastic regularization. Local adaptive volume regularization additionally prevents unnatural contraction or expansion of organs.

We overcome the limitation of low soft tissue contrast in CT by using additional PET images. Although the spatial resolution of PET is magnitudes lower compared to CT, the function information does not perturb the CT registration, but provides important complementary information in some soft tissue regions.

The primary goal is to apply our method to human data in future work. In addition, we will extend this work by analyzing a larger number of data sets with a larger number of VOIs. In this context it is also planned to analyze the applicability of the proposed method to subjects with tumors. It is planned to extend our method to dynamic PET data as activity over time carries important information for segmentation. An integration of our passive contour distance into the intensity-based registration of [1] is particularly promising. In future work we further plan to extend the data term to handle Poisson statistics and inhomogeneous areas as in [11].

Acknowledgments. The authors would like to thank Thomas Kösters for providing his reconstruction software EMrecon (`http://emrecon.uni-muenster.de`, [7]) used for the reconstruction of the PET datasets. This work was partly funded by the Deutsche Forschungsgemeinschaft, SFB 656 MoBil (projects B2 and B3).

References

1. Baiker, M., Staring, M., Löwik, C.W.G.M., Reiber, J.H.C., Lelieveldt, B.P.F.: Automated Registration of Whole-Body Follow-Up MicroCT Data of Mice. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 516–523. Springer, Heidelberg (2011)
2. Chan, T., Vese, L.: Active contours without edges. IEEE Trans Image Process 10(2), 266–277 (2001)
3. Chan, T., Vese, L.: Active contour and segmentation models using geometric PDE's for medical imaging. In: Malladi, R. (ed.) Geometric Methods in Bio-Medical Image Processing: Mathematics and Visualization, pp. 63–75. Springer (2002)
4. Dogdas, B., Stout, D., Chatziioannou, A., Leahy, R.: Digimouse: a 3D whole body mouse atlas from CT and cryosection data. Physics Med. Biol. 52(3), 577 (2007)
5. Gigengack, F., Ruthotto, L., Burger, M., Wolters, C., Jiang, X., Schäfers, K.: Motion correction in dual gated cardiac PET using mass-preserving image registration. IEEE Trans. Med. Imag. 31(3), 698–712 (2012)

6. Hugenberg, V., Breyholz, H.J., Riemann, B., Hermann, S., Schober, O., Schäfers, M., Gangadharmath, U., Mocharla, V., Kolb, H., Walsh, J., Zhang, W., Kopka, K., Wagner, S.: A new class of highly potent matrix metalloproteinase inhibitors based on triazole-substituted hydroxamates (radio)synthesis, *in vitro* and first *in vivo* evaluation. J. Med. Chem. 55(10), 4714–4727 (2012)
7. Kösters, T., Schäfers, K., Wübbeling, F.: EMrecon: An expectation maximization based image reconstruction framework for emission tomography data. In: NSS/MIC Conference Record. IEEE (2011)
8. Erdt, M., Steger, S., Sakas, G.: Regmentation: A new view of image segmentation and registration. Journal of Radiation Oncology Informatics, 1–23 (2012)
9. Massoud, T., Gambhir, S.: Molecular imaging in living subjects: seeing fundamental biological processes in a new light. Genes Dev. 17(5), 545–580 (2003)
10. Modersitzki, J.: FAIR: Flexible Algorithms for Image Registration. SIAM, Philadelphia (2009)
11. Sawatzky, A., Tenbrinck, D., Jiang, X., Burger, M.: A variational framework for region-based segmentation incorporating physical noise models. CAM Report 11-81, UCLA (December 2011)
12. Vese, L., Chan, T.: A multiphase level set framework for image segmentation using the Mumford and Shah model. International Journal of Computer Vision 50, 271–293 (2002)
13. Wang, H., Olafsen, T., Stout, D., Chatziioannou, A.: Quantification of organ uptake from small animal PET images via registration with a statistical mouse atlas. In: MICCAI Workshop Proceedings (2011)
14. Yezzi, A., Zöllei, L., Kapur, T.: A variational framework for integrating segmentation and registration through active contours. Med. Image Anal. 7(2), 171–185 (2003)
15. Zaidi, H., Ruest, T., Schoenahl, F., Montandon, M.: Comparative assessment of statistical brain MR image segmentation algorithms and their impact on partial volume correction in PET. Neuroimage 32(4), 1591–1607 (2006)

Spatially Aware Patch-Based Segmentation (SAPS): An Alternative Patch-Based Segmentation Framework

Zehan Wang, Robin Wolz, Tong Tong, and Daniel Rueckert

Department of Computing, Imperial College London, UK
zehan.wang06@imperial.ac.uk

Abstract. Patch-based segmentation has been shown to be successful in a range of label propagation applications. Performing patch-based segmentation can be seen as a k-nearest neighbour problem as the labelling of each voxel is determined according to the distances to its most similar patches. However, the reliance on a good affine registration given the use of limited search windows is a potential weakness. This paper presents a novel alternative framework which combines the use of kNN search structures such as ball trees and a spatially weighted label fusion scheme to search patches in large regional areas to overcome the problem of limited search windows. Our proposed framework (SAPS) provides an improvement in the Dice metric of the results compared to that of existing patch-based segmentation frameworks.

Keywords: patch-based segmentation, label propagation, multi-atlas, nearest neighbour search, spatial.

1 Introduction

Accurate segmentations in medical imaging form a crucial role in many applications from patient diagnosis to clinical research. The amount of data generated from medical images can take a substantial amount of time for clinicians to manually segment, often becoming prohibitive as a regular task. Consequently, automatic methods for performing these tasks are becoming more important for image analysis. However, obtaining accurate results is highly important and still poses a challenge in many medical imaging applications.

Patch-based approaches for label propagation [1], [2] have been shown to be a robust and effective solution for applications in medical images. These methods label each voxel of a target image by comparing the image patch centred on the voxel with patches from an atlas library and choosing the most probable label according to the closest matches.

In this paper, we propose an alternative framework for patch-based segmentation which uses efficient k nearest neighbour structures, such as ball trees and a spatially weighted label fusion method which is loosely based on a non-local means approach [3] to allow segmentation of data with greater variability in alignment after affine registration. We validate this approach on 202 images from the ADNI database and compare the results with an existing method.

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 93–103, 2013.

2 Methods

2.1 Pre-processing

Atlases are all registered to a common template space using affine registration and intensities are normalised using the method proposed by Nyúl and Udupa [4]. A general mask is then created for each label of interest in the atlas by taking the union of the labels from all the training data and dilating the result. This mask is used to narrow the search space and restrict search to valid areas where a label might appear. The mask needs to be large enough to allow for possible variations in anatomical variability, but not too large as this would make the search process less efficient.

The training data is also denoised to improve robustness. We used Total Variation denoising as a quick and easy to apply method which is effective in regularizing images without smoothing boundaries and edges [5].

2.2 *k*NN Data Structure Construction

Performing patch-based segmentation can be seen as a k-nearest neighbour problem as the labelling of each voxel is determined according to the distances to its most similar patches. An exhaustive search would have a computational complexity that is linearly proportional to the size of the dataset and can be quite prohibitive in large datasets, especially given the number of voxels that require this process in an image. This is one reason why existing methods use a small search volume size, such as in the region of $11 \times 11 \times 11 = 1331$ voxels, and why a good alignment of images is required.

To increase the search volume size without a detrimental impact to the search speed, an efficient kNN search data structure is required. Any exact kNN data structure could be used in this framework, but in our implementation, a ball tree [6] was used. Ball trees provide much better search performance than kd trees or brute force searches for high dimensional data [7]. Ball trees are metric trees which use a given distance metric to partition the data so that only a small part of the data need to be queried. The distance metric used must obey the triangle inequality for metric trees to work correctly. Since Euclidean distances are used in both patch based comparisons and atlas selection, and this obeys the triangle inequality, ball trees can then be used to provide the results to kNN queries.

In principle all patches could be stored in a single tree, however, the memory requirements would grow prohibitively large as the number of atlases increases as well as giving decremental search performances. So instead, a ball tree is constructed offline for each label in each atlas region of every atlas in the training set. Each patch stored in the ball tree also has its spatial coordinates within the template space stored with it. This information is used in a soft-weighting scheme when performing patch selection as spatial correspondence can help distinguish between patches with homogeneous intensities which provide very little structural information. This is particularly the case in brain images where patches from different structures of the brain can be very similar when only voxel intensities are compared.

Fig. 1. Example: ball tree construction from patches. Split the brain into 2 regions centred around the left and right hippocampus and create a tree in each region for each label, including the background label.

2.3 Search Strategy

Target images undergo the same pre-processing steps as the training images prior to segmentation as some degree of spatial correspondence is required for an effective segmentation.

Atlas Selection by Region. For each of the regions of interest that requires labelling, the nearest N atlases are found for each region by comparing their Euclidean distance. Using a limited selection of the best subjects from the atlas library has been shown to provide more effective segmentation results [8]. Another kNN data structure such as the ball tree can be built offline to allow fast atlas selection in the case of a large atlas library. The corresponding kNN data structure for those atlas regions are then used for segmentation. By performing atlas selection on the regional level, more appropriate atlases can be chosen for each region rather than selecting a single set of atlases to use for the whole image. This can be improve the accuracy of segmentations in cases where images differ in their similarity from region to region. For example, for performing a hippocampus segmentations, the set of atlases selected for the left hippocampus can differ to those selected for the right hippocampus.

Patch Search and Label Fusion. The corresponding kNN data structures for the nearest N regions are then used for finding the nearest k patches for each voxel location i in target image x. The Euclidean distance between the patch, $P(x_i)$, in the target image and the nearest k patches, $\{P(y_j)\}$, from the atlas library are weighted with the Euclidean distance on their spatial location to provide an overall weighting for each label. An additional weighting, α, can be

applied to control the influence of spatial correspondence. The resulting weighting for label l at voxel i is then determined by the sum of weights between patch $P(x_i)$ and the k nearest patches, $\{P(y_j)\}$ as follows:

$$w_{l_i}(x) = \sum_{j=1}^{k} w_l(x_i, y_j) \tag{1}$$

coordinate where

$$w_l(x_i, y_j) = e\frac{-\{||P(x_i) - P(y_j)||_2^2 + \alpha||x_i - y_j||_2^2\}}{h^2} \tag{2}$$

h is a decay parameter which controls the level of influence of patches as the distance increases, an automatic estimation of this parameter is used for each voxel based on the minimum distance between patch $P(x_i)$ and the nearest k patches, weighted by their spatial coordinates:

$$h^2(x_i) = \min\{||P(x_i) - P(y_j)||_2^2 + \alpha||x_i - y_j||_2^2\} \tag{3}$$

An overall weight for each label at each voxel i is then calculated from the sum of the distances of these patches and the resulting label is decided based on majority voting of the labels according to these weights:

$$L(x_i) = \arg\max_l w_{l_i}(x) \tag{4}$$

3 Experiments and Results

3.1 Dataset

Images from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI) were used for validation. These images consists of 202 subjects (68 normal, 93 with mild cognitive impairment, 41 with Alzheimer's disease) imaged using different scanners. Reference segmentations were obtained semi-automatically using a commercially available high dimensional brain mapping tool (Medtronic Surgical Navitgation Technologies, Louisville, CO) by propagating 60 manually labelled images. Images were pre-processed by the ADNI pipeline [9] and were linearly registered to the MNI152 template space using affine registration.

To test the proposed framework, a leave-one-out validation strategy was applied where each image was segmented in turn, using the remaining dataset as the atlas database. A patch size of $7 \times 7 \times 7$ was used whilst we experimented with the number of atlases used, N, the spatial weights, α, and the number of nearest neighbours for each patch, k. All image intensities were normalised and scaled to the same range and TV denoising [5] was applied to the training data.

3.2 Implementation

The main framework was implemented in Python using open source modules such as NumPy, SciPy and SciKit-learn. The computation time is around 10 minutes for each image using 8 cores clocked at 2.67GHz each when using 20 atlases and using the 100 nearst neighbours. Given that Python is an interpreted language, further speed ups can be achieved if the framework was implemented in C/C++.

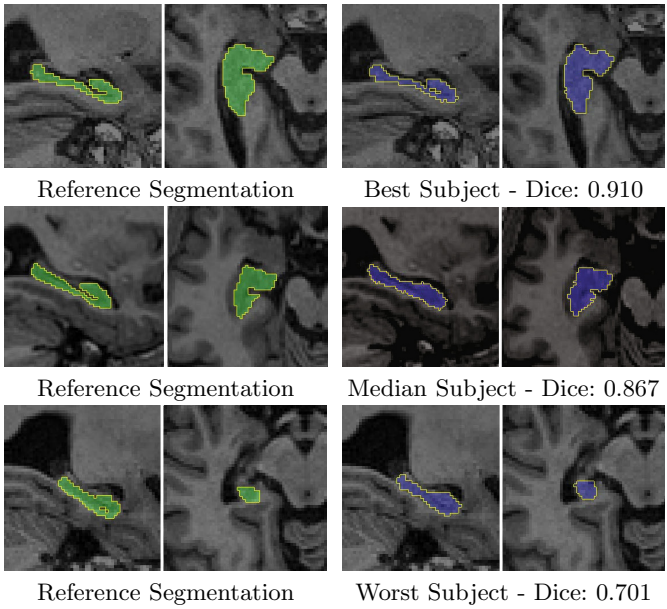


Fig. 2. Segmentations of the right hippocampus with parameters $N = 40$, $k = 79$, $\alpha = 13$

3.3 Effect of the Number of Nearest Patches and Atlases Used

With the spatial weight fixed at $\alpha = 13$, we experimented using a range of values for the number of patches, k, as well as the number of atlases, N. k is dependent on N as using more atlases would present a bigger selection of patches to choose from and we see in figure 3 that the optimal k value differs for the different N values.

Generally, we find that accuracy increases as k increases, but reaches a limit after $k > 60$. There is an increase in computational cost as k increases as more comparisons must be made in the kNN data structures, so it is most computationally optimal to select the lowest k value that provides the desired segmentation accuracy.

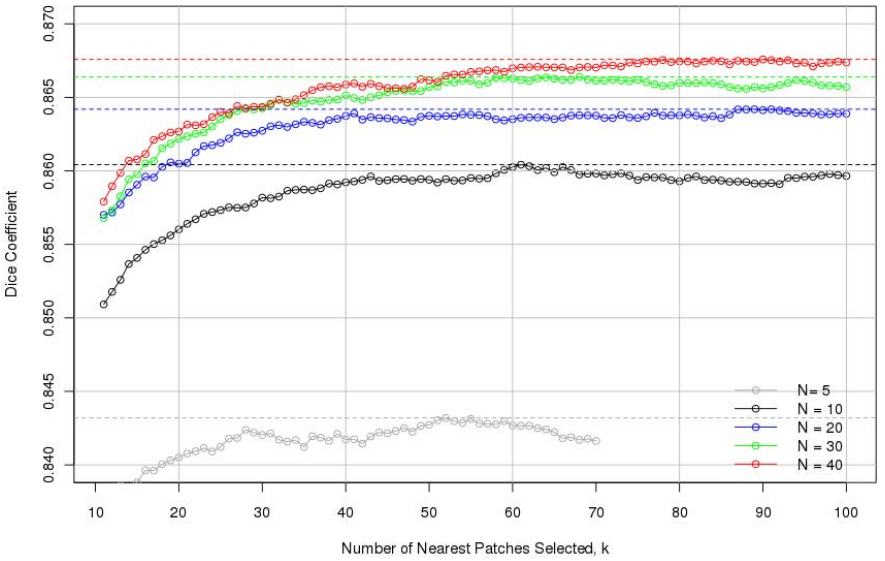


Fig. 3. Median Dice coefficients for the whole hippocampus whilst using a range of k values with different N values

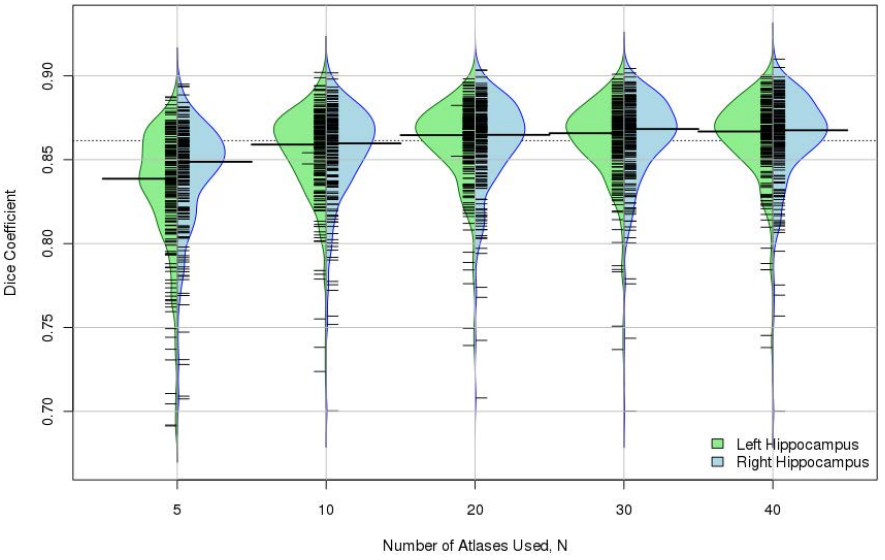


Fig. 4. Beanplot showing overall Dice coefficients distributions for a range of N values with $k = 64$. Large thick lines indicate medians, dotted line indicates median across all k values. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean.

Table 1. Dice Coefficients for the hippocampus (HC) when using different number of atlases, N, with $k = 64$. Highest values are show in bold.

N	Left HC			Right HC			Overall		
	Best	Worst	Median	Best	Worst	Median	Best	Worst	Median
5	0.887	0.691	0.839	0.895	0.707	0.849	0.886	0.719	0.842
10	**0.902**	0.724	0.860	0.902	0.700	0.859	0.898	0.719	0.860
20	0.898	**0.740**	0.864	0.904	**0.708**	0.865	0.899	**0.724**	0.864
30	0.901	0.737	0.866	0.904	0.700	**0.868**	0.899	0.719	0.866
40	0.900	0.738	**0.867**	**0.910**	0.700	0.868	**0.902**	0.719	**0.867**

An increase in the number of atlases used generally increases segmentation accuracy, but the gain accuracy after $N > 20$ is marginal. Given that the computational cost increases linearly with the number atlases used, this suggests that using more than 30 atlases would not provide a sufficient trade-off between the extra time spent and the accuracy gained. Our findings on here agree with those presented in [1] on the number of training subjects used, with proportional gains in accuracy as N increases.

3.4 Effect of the Spatial Weight, α

Experiments using several values for spatial weights, α, showed that using spatial information to provide a soft-weighting has significant impact on the

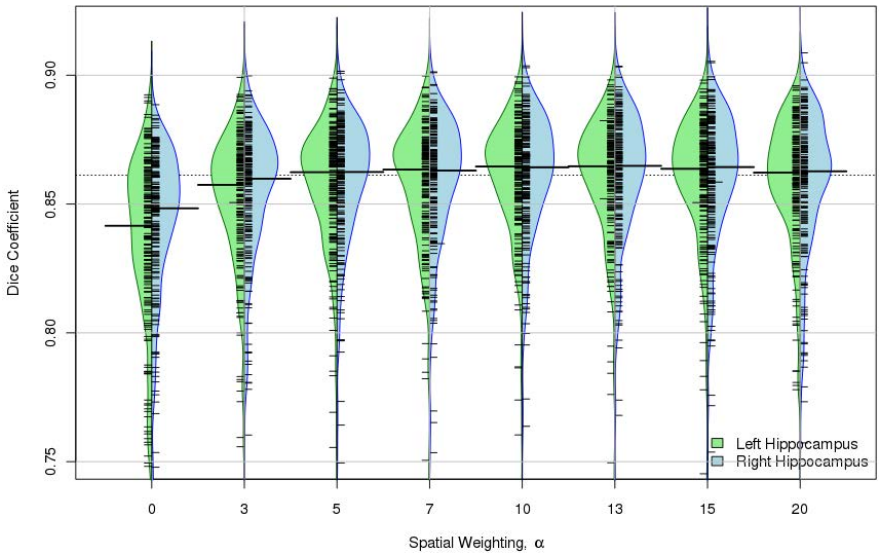


Fig. 5. Beanplot showing Dice coefficients distributions for a range of spatial weighting values, α with $N = 20$, $k = 64$. Large thick lines indicate medians, dotted line indicates median across all α values. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean.

Table 2. Dice Coefficients for the hippocampus (HC) when using different spatial weights, α, with $k = 64$ and $N = 20$. Highest values are show in bold.

α	Left HC			Right HC			Overall		
	Best	Worst	Median	Best	Worst	Median	Best	Worst	Median
0	0.892	0.669	0.842	0.889	0.674	0.848	0.884	0.702	0.844
5	0.899	0.736	0.862	0.902	0.700	0.862	0.897	0.718	0.862
10	**0.900**	**0.744**	**0.865**	0.906	0.692	0.864	0.899	0.723	0.863
13	0.898	0.740	0.864	0.904	0.708	**0.865**	0.899	**0.724**	**0.864**
15	0.898	0.736	0.864	0.905	0.704	0.864	**0.900**	0.720	0.863
20	0.860	0.729	0.862	**0.909**	**0.710**	0.863	**0.900**	**0.724**	0.862

segmentation accuracy (see figure 5 and table 2). The distribution of the results as seen in the beanplots shows that the consistency of the results increases significantly when we use spatial information. The values attempted suggests that segmentation accuracy peaks between $\alpha = 12$ and $\alpha = 13$. If the spatial weighting is too high, there is a detrimental effect on the segmentation accuracy as this soft-weighting becomes too restrictive when comparing patch intensities.

3.5 Effect of Denoising

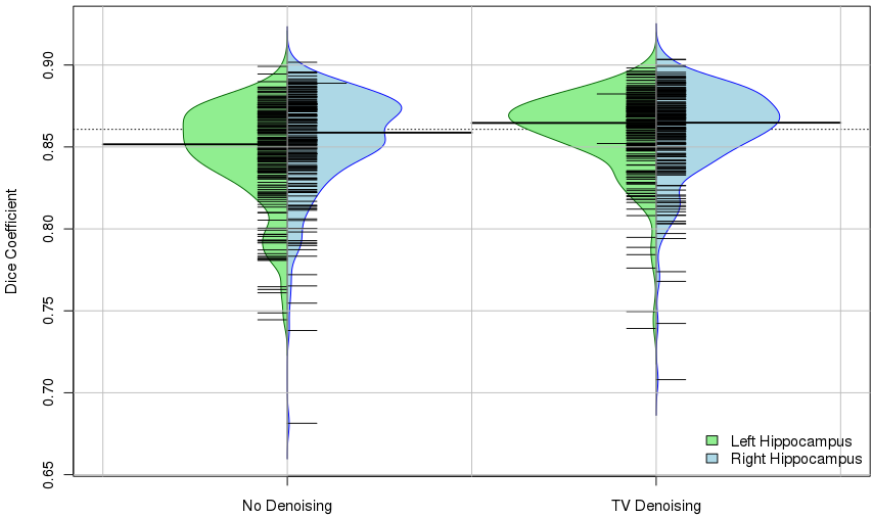


Fig. 6. Dice coefficients distributions for results using denoised and non-denoised training data with $N = 20$, $k = 64$, $\alpha = 13$. Large thick lines indicate medians, dotted line indicates median across both datasets. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean.

Comparing results from using non-denoised training data to those from using denoised training data, it can be seen that using denoised training data provides an improvement to the median segmentation accuracy (see figure 6). Further to this, the range of the results is significantly smaller with a more favourable distribution when using denoised training data, suggesting that this does indeed improve the generality and robustness of the framework.

3.6 Comparison of Results to an Existing Method

Finally, with the same dataset of ADNI images, we compared the results obtained by our proposed method to that using the method described in [1] (see figure 7 and table 3), with 10 training atlases in both cases. It can be seen that our

Table 3. Median Dice coefficients for the hippocampus (HC) comparing with the existing method in [1] with the number of atlases, $N = 10$ (*and* $N = 40$ *for reference*). Proposed method uses $k = 64$, $\alpha = 13$ as its other parameters.

Method	Left HC			Right HC			Overall		
	Best	Worst	Median	Best	Worst	Median	Best	Worst	Median
Existing[1]	0.894	0.696	0.842	**0.910**	0.644	0.848	**0.901**	0.709	0.844
Proposed, $N = 10$	**0.902**	**0.724**	**0.860**	0.902	**0.700**	**0.859**	0.898	**0.719**	**0.860**
Proposed, $N = 40$	*0.900*	*0.738*	*0.867*	*0.910*	*0.700*	*0.868*	*0.902*	*0.719*	*0.867*

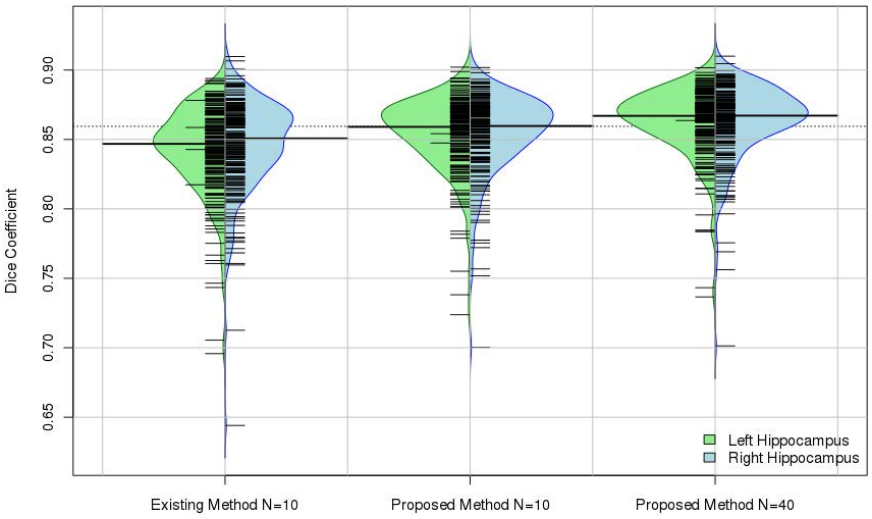


Fig. 7. Dice coefficients distributions for results comparing SAPS with an existing method [1]. Other parameters for SAPS are $k = 64$, $\alpha = 13$. Large thick lines indicate medians, dotted line indicates median across both datasets. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean.

method generally outperforms the existing method and is more robust. The two methods performs quite similarly when no spatial information is used (see table 2). This is because the label fusion would be equivalent to the non-local means method if the spatial weight, α, is 0.

Employing Welch's two sample t-test on these results gave p-values of 0.00003, 0.007 and 0.004 for the left, right and overall hippocampus respectively. Additionally, our proposed method has a 0.05 decrease in the standard deviation of the results.

4 Conclusion

We have presented a new generalized framework for applying patch based segmentation which is able to robustly segment data in conditions where images can have large variations in alignment by looking at a much larger search windows in addition to applying a spatial location weighting to each patch. We validated the proposed framework against 202 ADNI images of patients at various stages of Alzheimer's disease and achieved an overall median dice coefficient of 0.867 using patches from the 40 most similar atlases. The framework allows a trade-off between segmentation accuracy and speed. If we use patches from half as many atlases, we can complete the segmentation in half as much time and are still able to attain a median dice coefficient of 0.864. At the lowest limit tested, using 5 atlases is still able to yield a median Dice coefficient of 0.842 for the whole hippocampus whilst taking around 2-3 minutes on a machine with 8 cores.

In future work, we plan on further validating our proposed framework using a multi-scale extension against different anatomical structures and image types. We are currently working on a multi-scale extension to speed up segmentation of large structures such as bones in knee images or when performing brain extraction.

References

1. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage 54(2), 940–954 (2011)
2. Rousseau, F., Habas, P., Studholme, C.: A supervised patch-based approach for human brain labeling. IEEE Transactions on Medical Imaging 30(10), 1852–1862 (2011)
3. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE Transactions on Medical Imaging 27(4), 425–41 (2008)
4. Nyúl, L.G., Udupa, J.K.: On standardizing the MR image intensity scale. Magnetic Resonance in Medicine 42(6), 1072–1081 (1999)
5. Chambolle, A.: An Algorithm for Total Variation Minimization and Applications. Journal of Mathematical Imaging and Vision 20(1), 89–97 (2004)
6. Omohundro, S.M.: Five Balltree Construction Algorithms. Technical Report 1, International Computer Science Institute (1989)

7. Kumar, N., Zhang, L., Nayar, S.K.: What Is a Good Nearest Neighbors Algorithm for Finding Similar Patches in Images? In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 364–378. Springer, Heidelberg (2008)
8. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46(3), 726–738 (2009)
9. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L.G., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging 27(4), 685–691 (2008)

Efficient Geometrical Potential Force Computation for Deformable Model Segmentation

Igor Sazonov[1], Xianghua Xie[2], and Perumal Nithiarasu[1]

[1] College of Engineering, Swansea University, Singleton Park, Swansea, UK SA2 8PP
[2] Department of Computer Science, Swansea University, Singleton Park, Swansea, UK SA2 8PP
{i.sazonov,x.xie,p.nithiarasu}@swansea.ac.uk

Abstract. Segmentation in high dimensional space, e.g. 4D, often requires decomposition of the space and sequential data process, for instance space followed by time. In [1], the authors presented a deformable model that can be generalized into arbitrary dimensions. However, its direct implementation is computationally prohibitive. The more efficient method proposed by the same authors has significant overhead on computer memory, which is not desirable for high dimensional data processing. In this work, we propose a novel approach to formulate the computation to achieve memory efficiency, as well as improving computational efficiency. Numerical studies on synthetic data and preliminary results on real world data suggest that the proposed method has a great potential in biomedical applications where data is often inherently high dimensional.

1 Introduction

Among many others, deformable modeling is a popular approach to image segmentation, e.g. [2–4]. Conventional techniques suffer from weak edge, image noise and convergence issues. For instance, in [2] a constant pressure force is necessary in order to improve its capture range, resulting in monotonic expanding or shrinking of the mode that is problematic. There have been numerous work reported in the literature to improve the performance of both image gradient based methods, such as [5–7], and region based approaches, e.g. [8]. In [1], Yeo *et al.* proposed a 3D deformable model that is based on a hypothesised geometrical interactions between image gradient vectors and embedding level set surface normal vectors. It is shown that the geometrical potential force (GPF) is robust towards noise interference, weak edges, and exhibits invariant convergence capabilities such that the model can be initialized across object boundary and converge to deep concavities and propagate through narrow passages to recover complex geometries, that are conventionally difficult for image gradient based deformable modeling techniques. The authors also showed its theoretical relationship to the 2D Magnetostatic Active Contour (MAC) model [7] , which

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 104–113, 2013.

is inspired by a physical analogy. The MAC model can be considered a special case of GPF in 2D, whereas GPF can be more conveniently extended to higher dimensional applications.

The computation of the GPF comprises two stages. At the first stage, the so-called geometrical potential (GP) $G(x, y, z)$ is computed through the convolution of the image gradient and the kernel $\mathbf{K}$:

$$G(\mathbf{x}) = \sum_{\mathbf{x}' \in \Omega} \nabla I(\mathbf{x}') \cdot \mathbf{K}(\mathbf{x} - \mathbf{x}'), \qquad \mathbf{K}(\mathbf{x}) = \begin{cases} \mathbf{x} / \|\mathbf{x}\|^{n+1}, & \mathbf{x} \neq \mathbf{0} \\ \mathbf{0}, & \mathbf{x} = \mathbf{0} \end{cases} \tag{1}$$

where $\mathbf{x} = [x, y, z]^T$ is the vector of coordinates of the image grid-points (voxel centres), $I(\mathbf{x})$ is the greyscale image, ∇I is its gradient, Ω is the image domain, dot denotes the scalar product of two vector functions (∇I and kernel $\mathbf{K}(\mathbf{x})$), and n is the image dimension ($n = 3$ for 3D images).

At the second stage, the derived geometrical potential is then integrated into the deformable surface evolution under the level set framework. The active surface, $S(t)$, is embedded in the level set function, $\Phi(t, \mathbf{x})$: $S(t) = \{\mathbf{x}, \Phi(t, \mathbf{x}) = 0\}$, and its deformation is achieved by solving the following PDE proposed and developed in [9–11] and related to the energy minimization approach:

$$\partial\Phi/\partial t = \alpha\, g\kappa \|\nabla\Phi\| - (1-\alpha)\, \mathbf{F}\nabla\Phi \tag{2}$$

where α is a weighting parameter, $g(\mathbf{x}) = 1/(1+\|\nabla I\|^2)$ is the stopping function, $\kappa(t, \mathbf{x}) = \nabla\hat{\mathbf{n}}$ denotes the curvature of isosurfaces of Φ, $\hat{\mathbf{n}}(t, \mathbf{x})$ is the unit vector normal to isosurfaces of Φ, and $\mathbf{F}(t, \mathbf{x}) = G\,\hat{\mathbf{n}}$ is the GPF that acts as the external force.

Direct calculation of the geometrical potential G is computationally expensive, particularly in 3D. However, Eq. (1) can be computed as a convolution of two functions. Hence a natural approach is to apply the fast Fourier transform (FFT) to compute the convolution, which is described in [1].

However, a significant drawback of using the FFT based computation as proposed in [1] is that it requires lots of computer memory for a large number of intermediate arrays of the same size as the initial image I. That is, it needs to compute and store 3 components of the image gradient $\nabla I = [I_x, I_y, I_z]^T$ and twice more for the real and imaginary part of their Fourier image, also 3 components of the kernel $\mathbf{K}$ and twice more for the Fourier image. Thus, it requires about 20 times more than the direct method, which can be problematic when dealing with volumetric data or extending this method to 4D, i.e. dynamic volumetric data. Dedicated memory management may become necessary and even crucial. Memory economic and computationally efficient method to evaluate the GP is thus desirable.

In this paper, we propose to compute spectrum of the kernel by an analytical formula so that there is no need to store components of the vector kernel and the real or imaginary part of its spectrum. We also change the vector form of the integrand into a scalar form to achieve further efficiency. The proposed methods are valuated on both numerical examples and real world 3D data.

2 Analytical Formula for Kernel's Spectrum

One of the possible approaches to reduce memory usage is to use an analytical formula for the kernel spatial spectrum rather than kernel's formula (1) in the x-space. To derive an analytical formula for the kernel Fourier image, it is useful to consider the computation of G in the continuous infinite 3D Euclidian space. In this case the kernel should be described by a generalized function (distribution) (see, e.g. [12]):

$$G(\mathbf{x}) = \int_{\mathbf{x}' \in \mathbb{R}^3} \nabla I(\mathbf{x}') \cdot \mathbf{K}(\mathbf{x} - \mathbf{x}') \, \mathrm{d}^3\mathbf{x}, \qquad \mathbf{K}(\mathbf{x}) = P.V. \frac{\mathbf{x}}{\|\mathbf{x}\|^{n+1}} \tag{3}$$

where *P.V.* denotes *principal value*, i.e. integral in (3) diverging when $\mathbf{x}' \to \mathbf{x}$, should be treated as the limit

$$G(\mathbf{x}) = \lim_{\varepsilon \to 0+} \int_{\|\mathbf{x}' - \mathbf{x}\| > \varepsilon} \nabla I(\mathbf{x}') \cdot \frac{\mathbf{x} - \mathbf{x}'}{\|\mathbf{x} - \mathbf{x}'\|^{n+1}} \, \mathrm{d}^3\mathbf{x}' \tag{4}$$

Performing the Fourier transform

$$\tilde{\mathbf{K}}(\mathbf{k}) = \mathcal{F}[\mathbf{K}](\mathbf{k}) = \int \mathbf{K}(\mathbf{x}) e^{\mathrm{i}\mathbf{k}\mathbf{x}} \, \mathrm{d}^3\mathbf{x}, \qquad \mathrm{i} = \sqrt{-1} \tag{5}$$

we can show that that the spectrum depends only on direction of wavevector $\mathbf{k}$ and is independent of its magnitude

$$\tilde{\mathbf{K}}(\mathbf{k}) = -\mathrm{i}\pi^2 \frac{\mathbf{k}}{\|\mathbf{k}\|}. \tag{6}$$

Comparing spectrum $\tilde{\mathbf{K}}(\mathbf{k})$ computed analytically via Eq. (6) and that computed by performing FFT for the kernel calculated in the x-space by (1) (see Figure 1(left)), we see that near the origin they have close values. However, the spectrum computed via the FFT decays when any component of the wavevector grows. Moreover, it vanishes when any component of the wavevector reaches its maximum value which is determined by the grid size in the correspondent direction: $k_{i,\max} = \pi/h_i$ where h_1, h_2, h_3 are voxel sizes in x,y,z direction, respectively. Therefore, to obtain the G-function close to that computed by FFT based method, spectrum (6) should be multiplied by a function $f(\mathbf{k})$ which equals 1 in the origin and smoothly decays when $k_i \to k_{i,\max}$. As numerical computation shown later, a good approximation of a 3D spectrum can be formulated as

$$\tilde{\mathbf{K}}(\mathbf{k}) = \mathrm{i}\pi^2 \frac{\mathbf{k}}{\|\mathbf{k}\|} f(\mathbf{k}), \qquad f(\mathbf{k}) = (1 - \|\mathbf{k}'\| + V(\mathbf{k})) \tag{7}$$

where

$$V = \frac{(\xi \|\mathbf{k}'\| - 1)^2}{(\xi + \xi \|\mathbf{k}'\| - 2)\,\xi}, \quad \mathbf{k}' = \left[\frac{k_1}{k_{1,\max}}, \frac{k_2}{k_{2,\max}}, \frac{k_3}{k_{3,\max}}\right]^T, \quad \xi = \max_{i=1,2,3} |k_i'|.$$

This makes the computation much more memory economic; however, we still have to compute the FFT for components of ∇I and then multiply every element of the arrays of the kernel spectrum computed directly for every element.

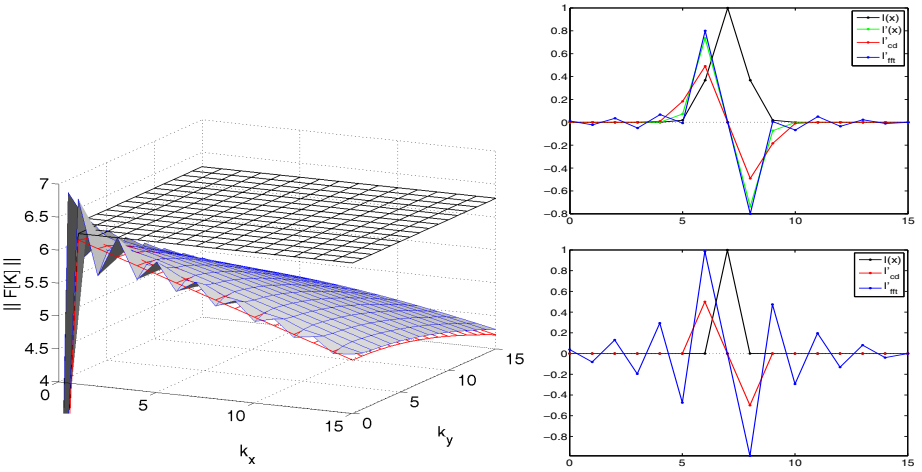


Fig. 1. Left: Absolute value of spectrum $\|\mathrm{Im}\,\tilde{\mathbf{K}}\|$ in the 128×128 2D domain (only first 16 positive wave components are shown) computed via Eq. (6) (black), computed by FFT from kernel evaluated in the x-space (blue) and approximated by Eq. (7) (red). Right: Function $I(x)$ (black), its exact derivative (green, on the top only), its derivative computed by central-difference (red), the same—through FFT (blue). Top: $I(x) = \exp\{-(x-7)^2\}$, bottom: $I(x) = \delta(x-7)$.

3 Use of a Scalar Kernel

Alternatively, we may rearrange the intergrand shown in Eqn. (3) as a product of scalar function and a scalar kernel, instead of a dot product between vectors. To derive the correspondent formula in x-space, we again temporally consider continuous infinite space in which initial integral takes the form given in (3). We then reforumate (3) as

$$G(\mathbf{x}) = \int_{\mathbf{x}' \in \mathbb{R}^3} I(\mathbf{x}') \cdot \nabla \mathbf{K}(\mathbf{x} - \mathbf{x}')\, \mathrm{d}^3\mathbf{x} \tag{8}$$

Thus, we only have to deal with the scalar kernel which is the divergence of the vector kernel $\mathbf{K}$. In the discretized finite domain Eqn. (8) can be approximated as

$$G(\mathbf{x}) = \sum_{\mathbf{x}' \in \Omega} I(\mathbf{x}') \cdot K(\mathbf{x} - \mathbf{x}'), \qquad K(\mathbf{x}) = \nabla \mathbf{K}(\mathbf{x}) \tag{9}$$

where the best way to calculate $\nabla \mathbf{K}$ is to compute vector kernel $\mathbf{K}$ and compute the spatial derivatives by central differences.

4 Combined Approach

However, we may combine the above two methods together to achieve even more efficient computation. In the k-space, the calculation of the geometrical potential spectrum, $\tilde{G}(\mathbf{k})$, is read as

$$\tilde{G} = \left(\mathrm{i}\mathbf{k}\tilde{I}\right) \cdot \tilde{\mathbf{K}} = \tilde{I}\left(\mathrm{i}\mathbf{k} \cdot \tilde{\mathbf{K}}\right) = \tilde{I}\tilde{K} \tag{10}$$

where $\tilde{K}(\mathbf{k})$ is spectrum of the scalar kernel ($K = \nabla\mathbf{K}$), factor $\mathrm{i}\mathbf{k}$ in the k-space corresponds to the nabla (∇) operator in the infinite continuous x-space. Because we are dealing with discretized images with noise, the computation of the gradient through multiplication by $\mathrm{i}\mathbf{k}$ in the k-space can result in undesired sensitivity to noise. Derivative of a function on a finite uniform grid can be approximated by forward, backward or central differences, but also can be computed through the direct and inverse FFT. The latter method gives very high accuracy for smooth functions (periodic or decaying fast toward the grid borders).

For example, for a 1D function $I(x) = \exp\{-(x-7)^2\}$ set on $x = \{0, 1, \ldots, 15\}$ the error of derivative computed by the central differences is 0.24 whereas the error of derivative computed trough FFT is only 0.08 as seen in Figure 1(right-top). But if the function is not smooth (for example, contains delta-correlated noise) the situation is quite opposite. Consider, as an example, a discrete implementation of Dirac's delta $\delta(x-7)$. Then the derivative computed by the central differences gives a reasonable approximation of $\delta'(x-7)$ with a three point support, whereas the FFT method gives an oscillating result, as depicted in Figure 1(bottom right).

Thus, for image segmentation when noise is common in presence it is more appropriate to use central differences approximation than the FFT method. Fortunately though, the Fourier transform can be used to compute the central differences as well. Recall that in a continuous infinite space the derivative can be expressed as a convolution with $\delta'(x)$

$$\partial I/\partial x = I * (\delta'(x)) = \int_{-\infty}^{+\infty} I(x')\,\delta'(x-x')\,\mathrm{d}x', \tag{11}$$

Computing this derivative by use of the Fourier transform, we should recall its spectrum $\mathcal{F}[\delta'(x)] = \mathrm{i}k$. The central differences can be computed analogously as a convolution with the function $\frac{1}{2h}\big(\delta(x+h) - \delta(x-h)\big)$ having spectrum

$$\mathcal{F}\big[\frac{1}{2h}\big(\delta(x+h) - \delta(x-h)\big)\big] = \frac{\mathrm{i}}{h}\sin(kh). \tag{12}$$

which tends to $\mathrm{i}k$ when $h \to 0$.

In 3D case, spectrum of the gradient operator, $\mathrm{i}\mathbf{k}$, should be substituted by vector

$$\mathbf{g}(\mathbf{k}, \mathbf{h}) = \left[\frac{\mathrm{i}\sin k_1 h_1}{h_1}, \frac{\mathrm{i}\sin k_2 h_2}{h_2}, \frac{\mathrm{i}\sin k_3 h_3}{h_3}\right]^T \tag{13}$$

Then Eqn. (10) should be transformed to

$$\tilde{G} = \tilde{I} \times \left(\mathbf{g}(\mathbf{k}, \mathbf{h}) \cdot \tilde{\mathbf{K}}\right). \tag{14}$$

Thus, for this combined approach we perform FFT on the image $I(\mathbf{x})$; then for every element of the obtained arrays we calculate the scalar kernel spectrum by

employing Eqn. (7) for the vector kernel and Eqn. (13) for the modified nabla operator in the k-space; finally we carry out the inverse Fourier transform. It requires memory space 4 times less than that for the initial image $I(\mathbf{x})$: I, $\mathrm{Re}\,\tilde{I}$, $\mathrm{Im}\,\tilde{I}$, G.

5 3D Numerical Examples

To compare different methods for computation of the geometrical potential, an artificial 3D star-like gray-scale image is created shown in Figure 2(right). Its dimension is $64 \times 64 \times 32$ pixel: this relatively small size image is chosen for the sake of convenience in visualizing the results. To understand the noise interference, the 3D data is then added with 5% Gaussian noise.

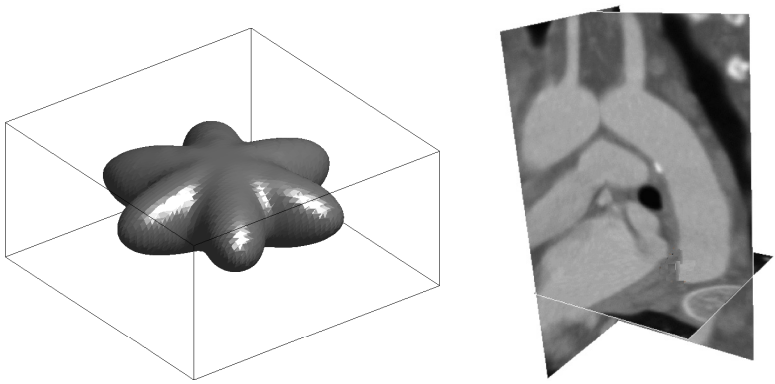


Fig. 2. Left: Isosurface of the 3D image (without added Gaussian noise). Right: an example of 3D scan of a human aorta.

Figure 3(left) shows the mid-slice along the z-axis. Note, the zero-crossings in the geometrical potential are in effect indicating the locations where the deformable model will converge, since on either side of the zero crossing the deformable model will converge towards zero-crossings. Hence, in the numerical studies, we examine the accuracy of the zero-crossings of different methods compared to the object boundary (groundtruth). The colored contours in Figure 3(left) indicate the results from different methods. Also the black curve shows the isoline for $I(x, y, z_m) = I_m$, i.e. result of segmentation performed by thresholding [13]: the middle value $I_m = \frac{1}{2}(I_{\max} + I_{\min})$ is used as the threshold.

All the lines are very close to each other, which suggests that the proposed methods are close approximation to the direct method. Plots of geometrical potential $G(x, y_m, z_m)$ along the x-coordinates is shown in Figure 3(right), where $y_m = \frac{1}{2}(y_{\min} + y_{\max})$. The curve $I_m - I(x, y_m)$ is plotted in black. It shows that the difference zero crossing is small. The rectangular region indicated by the dotted

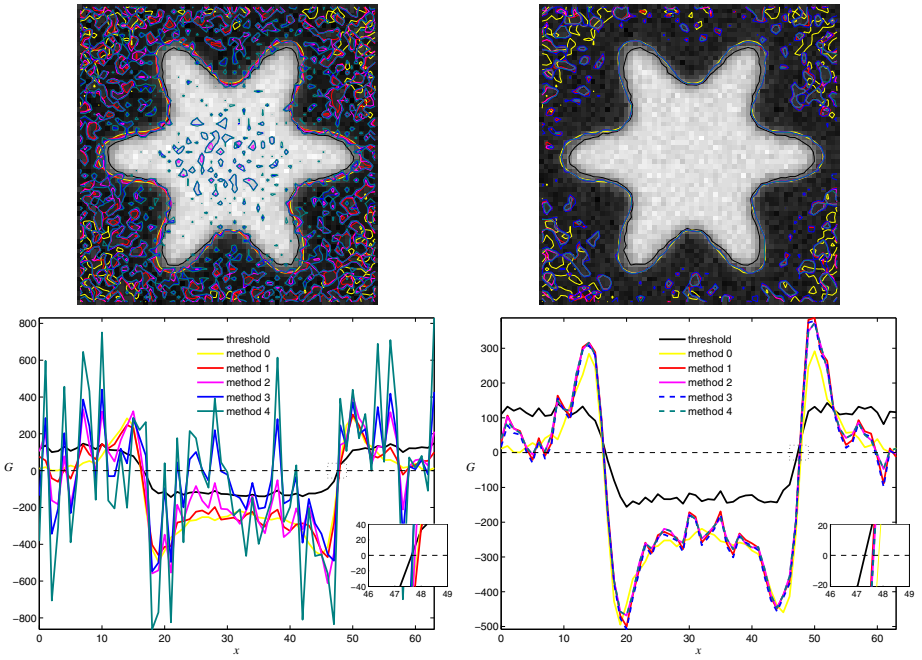


Fig. 3. Top: A slice of the 3D image; the colored curved indicate isolines of $G = 0$. Bottom: The G variation along the x direction through the center of the 3D image computed by the different methods explained in the legend. Method 0 is direct computation of the geometric potential; method 1 is the FFT based implementation of method 0; method 2 is using analytical formula for kernel's spectrum; method 3 using the scalar kernel alone; and method 4 is combining methods 2 and 3. Left: methods 2–4 without corrections, right: methods 2–4 with corrections (7) and (14).

line is zoomed and depicted at the right border of the plot. The difference is in sub-pixel level.

The direct computation is less susceptible to noise, but it is too slow to be practical. The proposed methods produce very similar result to that using FFT computation as proposed in [7, 1]. However, the proposed methods, particularly the combined approach, are far more memory efficient.

The CPU time of all the methods can be found in Table 1. Note, the combined approach (method 4) uses 4 times less memory than the FFT based computation used in [1]. The experiment was carried out on Linux, Intel(R) Xeon 3.00GHz, RAM 4G. A typical 3D scan of 512^3 voxels can only be pratically processed by method 4 and it requires 8 min of the CPU time and 3G of memory.

To demonstrate the effectiveness of the proposed combined approach, we show an example of segmenting a human aorta from a 3D CT dataset. The testing data and the results are shown in Figures 2(left) and 4. The initial surface is a sphere placed inside the lower part of the aorta and the model is able to propagate efficiently and converge accurately.

Table 1. CPU time and memory comparison for a 256^3 image. Method 0 is direct computation of the geometric potential; method 1 is the FFT based implementation of method 0; method 2 is using analytical formula for kernel's spectrum; method 3 using the scalar kernel alone; and method 4 is combining methods 2 and 3.

	method 0	method 1	method 2	method 3	method 4
CPU time	$\sim$ 7days	91s	55s	42s	30s
Memory required	0.6G	1.8G	1.0G	0.6G	0.4G

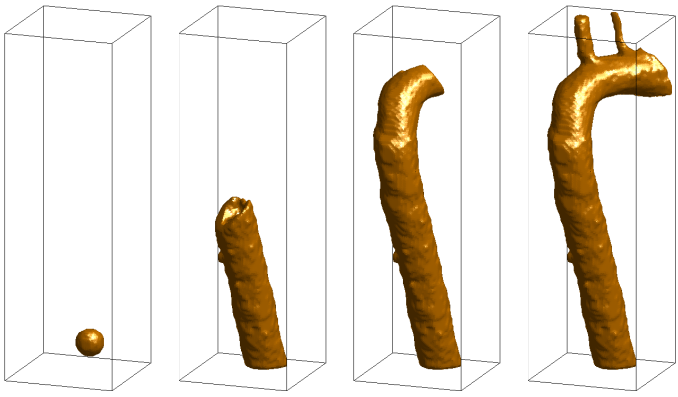


Fig. 4. An example of segmenting human aorta in 3D CT shown in Fig. 2(right) using the combined approach. From left: initial surface, intermediate stages, and final converged result.

6 4D Numerical Examples

Note that all the equations derived for the proposed methods can be readily generalised to 4D medical scans (dynamic volumetric data). We should treat the coordinate vector as $\mathbf{x} = \{x, y, z, t\}$, use the 4D wavenumber vector $\mathbf{k}$ with k_4-component treated as the frequency, and substitute $n = 4$ into the correspondent formulae for the kernel in Eqns. (1) and (3).

Here, we present a numerical study that is similar to that in the 3D case, but using a dynamic 3D shape. We vary the ray length shape parameters of the 3D star-like harmonic object periodically in time with the maximum near the middle of the cycle. The ray length parameter evolution is given as $[\frac{1}{2}(1 + \cos(2\pi(t - t_m - \frac{1}{3})/N_t))]^{1.5}$ where $N_t = 16, t_m = N_t/2$. The image dimension is $64 \times 64 \times 32 \times 16$. Thus the image contains 16 3D images, some of which are shown in Figure 5. Similarly, Gaussian noise is also added to the dynamic shape.

Figure 6(left) shows a slice of the image at instant $t = t_m = 7$ (the maximal length of the star-rays) and $z = z_m$. Here one can find colored contours $G(x, y, z_m, t_m) = 0$ with the geometrical potential computed by the different methods implemented in 4D. Spatial zero-crossings of geometrical potential:

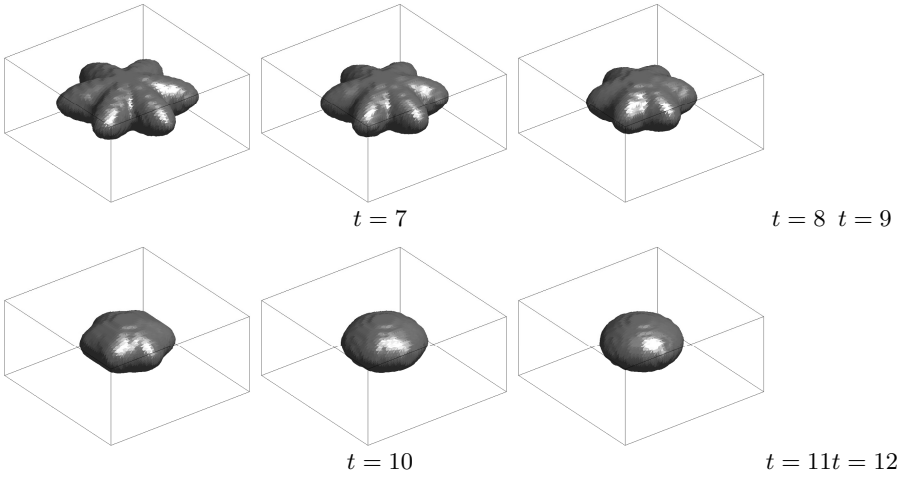


Fig. 5. Object shape at instances of 7 to12

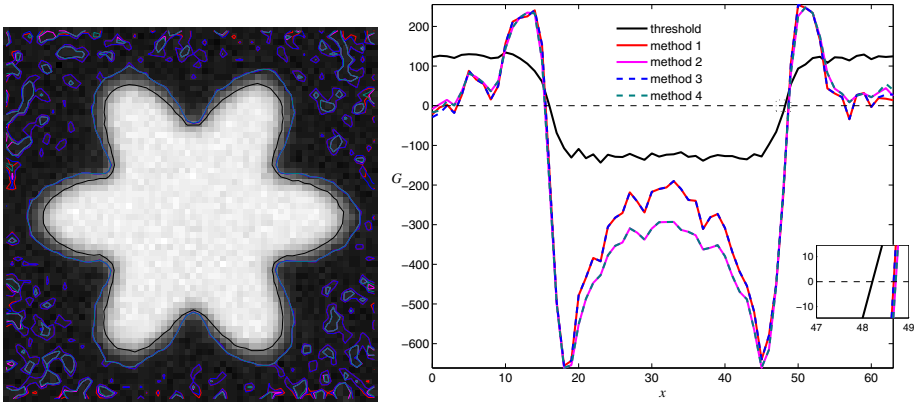


Fig. 6. Left: A slice of the 4D image; the colored curved indicate isolines of $G = 0$. Right: The G variation along the x direction through the center of the 3D image computed by the different methods explained in the legend. Method 1 is the FFT based implementation of direct computation; method 2 is using analytical formula for kernel's spectrum; method 3 using the scalar kernel alone; and method 4 is combining methods 2 and 3.

$G(x, y_m, z_m, t_m)$ where y_m, z_m, t_m are plotted in Figure 6(right). Note, the direct method is not shown as it takes prohibitive amount of time to compute the geometrical potential. There is no discernible difference among methods with improved computational efficiency. However, the proposed combined approach requires significantly less memory. This is particularly advantageous in dealing with 4D dataset.

7 Conclusion

We proposed several computationally efficient and memory economic methods to evaluate the geometrical potential in the GPF model [1]. The approach which combines analytical kernel spectrum and scalar kernel conversion provides most satisfactory results. The methods were evaluated on 3D and 4D synthetic datasets, as well as 3D real world data. This preliminary work provided promising results which suggest that the proposed method has a great potential in efficient deformable modelling in high dimensional space without decomposing the space into a sequential order.

References

1. Yeo, S.Y., Xie, X., Sazonov, I., Nithiarasu, P.: Geometrically induced force interaction for three-dimensional deformable models. IEEE T-IP 20(5), 1373–1387 (2011)
2. Malladi, R., Sethian, J.A., Vemuri, B.C.: Shape modelling with front propagation: A level set approach. IEEE T-PAMI 17(2), 158–175 (1995)
3. Whitaker, R.: Modeling deformable surfaces with level sets. IEEE Computer Graphics and App. 24(5), 6–9 (2004)
4. Xie, X.: Active contouring based on gradient vector interaction and constrained level set diffusion. IEEE T-IP 19(1), 154–164 (2010)
5. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE T-IP 7(3), 359–369 (1998)
6. Xiang, Y., Chung, A., Ye, J.: A new active contour method based on elastic interaction. In: IEEE CVPR, pp. 452–457 (2005)
7. Xie, X., Mirmehdi, M.: MAC: Magnetostatic active contour model. IEEE T-PAMI 30(4), 632–647 (2008)
8. Chan, T., Vese, L.: Active contours without edges. IEEE T-IP 10(2), 266–277 (2001)
9. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contour. IJCV 22(1), 61–79 (1997)
10. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. IJCV 46(3), 223–247 (2002)
11. Parigios, N., Mellina-Gottardo, O., Ramesh, V.: Gradient vector flow geometric active contours. IEEE T-PAMI 26(3), 402–407 (2004)
12. Vladimirov, V.S.: Methods of the Theory of Generalized Functions. Taylor & Francis (2002)
13. Smith, C.M., Smith, J., Williams, S.K., Rodriguez, J.J., Hoying, J.B.: Automatic thresholding of three-dimensional microvascular structures from confocal microscopy images. J. Microscopy 225(3), 244–257 (2007)

Shape Prior Model for Media-Adventitia Border Segmentation in IVUS Using Graph Cut

Ehab Essa[1], Xianghua Xie[1], Igor Sazonov[2], Perumal Nithiarasu[2], and Dave Smith[3]

[1] Department of Computer Science, Swansea University, Singleton Park, Swansea, UK SA2 8PP
[2] College of Engineering, Swansea University, Singleton Park, Swansea, UK SA2 8PP
[3] ABM University NHS Trust, Swansea, UK
{csehab,x.xie,i.sazonov,p.nithiarasu}@swansea.ac.uk

Abstract. We present a shape prior based graph cut method which does not require user initialisation. The shape prior is generalised from multiple training shapes, rather than using singular templates as priors. Weighted directed graph construction is used to impose geometrical and smooth constraints learned from priors. The proposed cost function is built upon combining selective feature extractors. A SVM classifier is used to determine an optimal combination of features in presence of calcification, fibrotic tissues, soft plaques, and metallic stent, each of which has its own characteristics in ultrasound images. Comparative analysis on manually labelled ground-truth shows superior performance of the proposed method compared to conventional graph cut methods.

Keywords: IVUS, graph cut, image segmentation, shape prior.

1 Introduction

Intra-vascular Ultrasound (IVUS) imaging is a catheter-based technology, which shows 2D cross-sectional images of the coronary artery. A typical IVUS image consists of lumen, vessel that includes intima and media layers, and adventitia that surrounds the vessel wall. The media-adventitia border represents the outer coronary arterial wall located between the media and adventitia. The media layer exhibits as a thin dark layer in ultrasound and has no distinctive feature. It is surrounded by fibrous connective tissues called adventitia. The appearance of the media-adventitia border in IVUS is affected by various forms of artifact, such as acoustic shadow which can be caused by catheter guide wire, dense fibrous tissue or calcification. Fig. 1 gives an example of IVUS image.

Segmentation in IVUS images has shown to be an intricate process and often requires user initialisation to achieve meaningful results. Among many others, graph based segmentation has shown to be a promising approach to IVUS segmentation. In [1], dynamic programming is used to search a minimum path in a cost function, which incorporates edge information with a simplistic prior based on echo pattern and border thickness. Manual initialisation is necessary. In [2],

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 114–123, 2013.

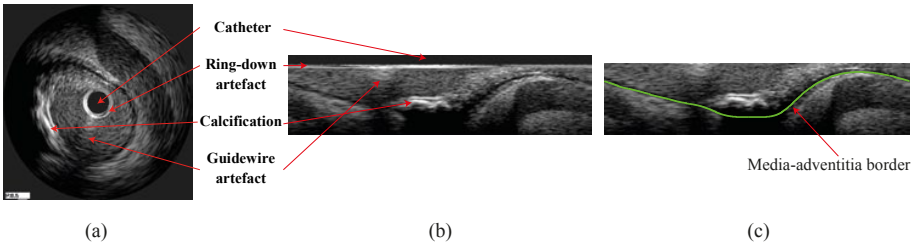


Fig. 1. Pre-Processing steps. (a) Original IVUS image. (b) Polar transformed image. (c) After removing the catheter region.

the authors applied spatio-temporal filters to enhance the lumen region based on the assumption that the blood speckles have higher spatial and temporal variations than arterial wall. However, image features introduced by acoustic shadow or metallic stent would seriously undermine this assumption when searching for media-adventitia border. The *s-t* cut method [3] is employed in [4] to segment 3D IVUS data. Intensity distribution in the radial directions from catheter origin and regional features based on piecewise constant assumption are used to design the cost function. However, intensity based features are susceptible to artifacts.

Learning *a priori* using a set of representative shapes is an effective approach to impose a general constraint in searching global minimum using graph cut. Freedman and Zhang [5] defined the shape template as a distance function and embedded the average distance between every pair of pixels into the neighbourhood edges in the graph. However, this method effectively requires the user to place landmarks to define the initial shape. In [6], the authors proposed an iterative graph cut method. Kernel PCA was used to build the shape model. The method ignores the affine transformation, and needs a rectangle window initialisation of the location of the objects. Iterative graph cut framework was also adopted in [7]. The method penalises the terminal edges of the graph according to the similarity between the previous segmentation and the shape template.

In this paper, we propose an efficient graph cut algorithm to segment media-adventitia border in IVUS images without user initialisation. Its objective functional consists of boundary based cost and shape penalties that are generalised from multiple training shapes. The boundary based features are dynamically selected to optimise the cost function based on trained classifier. The generalised shape prior is incorporated in the cost function, as well as embedded in graph construction. The method is evaluated on a large set of real data with groundtruth.

2 Proposed Method

The images are first transformed from Cartesian coordinates to polar coordinates and the catheter regions are removed (see Fig. 1). This transformation not only

facilitates our feature extraction and classification but also transfers a closed contour segmentation to a "height-field" segmentation (see Fig. 1(c)). The border to be extracted intersects once and once only with each column of pixels. This particular form of segmentation allows us to construct a node-weighted directed graph, on which a minimum path can be found without any user initialisation.

2.1 Graph Construction without Shape Prior

We first present our basic graph construction, following [8], which does not require user initialisation. Our extended version with incorporated shape prior will be discussed later in Section 2.5. Let $G = \langle V, E \rangle$ denote the graph, where each node $V(x, y)$ corresponds to a pixel in the transformed IVUS image $I(x, y)$ in polar coordinates. The graph G consists of two arc types: intra-column arcs and inter-column arcs. For intra-column, along each column every node $V(x, y)$, where $y > 0$, has a directed arc to the node $V(x, y-1)$ with $+\infty$ weight assigned to the arc to ensure that the desired interface intersects with each column exactly once. In the case of inter-column, for each node $V(x, y)$ a directed arc with $+\infty$ weight is established to link with node $V(x+1, y-\Delta_{p,q})$, where $\Delta_{p,q}$ is the maximum difference between two neighbouring columns p and q and acts as a smoothness constraint. Similarly, node $V(x+1, y)$ is connected to $V(x, y-\Delta_{p,q})$. For IVUS segmentation, the first and the last columns are connected by inter-column arcs to enforce connectivity. Finally, the nodes in the last row of the graph are connected to each other with $+\infty$ weight to maintain a closed graph. Inter-columns and intra-columns arcs are illustrated in Figure 2 (a).

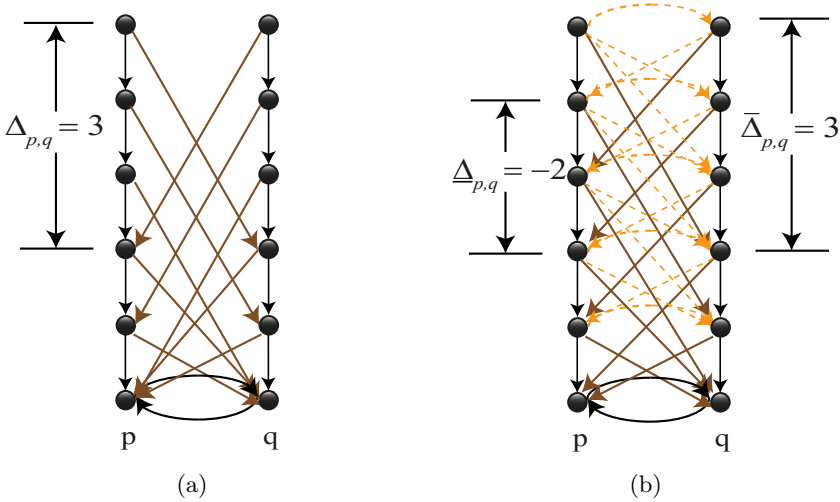


Fig. 2. Graph construction. (a) without shape prior where shape constraint is a global constant. (b) shape prior model (refer to Section 2.5 for details).

2.2 Feature Extraction and Classification

The media layer is usually thin and generally dark, and the adventitia layer tends to be brighter, see Fig. 1 as an example. Hence, edge based features are appropriate to extract the media-adventitia border. However, calcification and other interfering image features commonly exist above the media-adventitia border and cast acoustic shadows over the border, disrupting its continuity. Those imaging artifacts generally have large responses to image gradient based feature extraction. In this work, we propose to detect those artifacts and treat them differently when incorporating into the cost function.

To highlight the media-adventitia border, we use a combination of derivative of Gaussian (DoG) features and local phase features. A set of first and second order DoG filters are applied to capture the intensity difference between media and adventitia.

Local phase [9] has shown to be effective in suppressing speckles in ultrasound images. We use the dark symmetry feature [9] to highlight bar-like image patterns, which are useful to detect the thin media layer. This feature extraction operates at a coarser scale and complements to the edge features extracted using DoG filters.

For those parts of media-adventitia border that are beneath various forms of image artifacts, such as calcification, their image features are suppressed by those artifacts. Hence, it is desirable to detect those artifacts and treat those columns of pixels differently to others. However, instead of a usual attempt of localising those image artifacts based on intensity profile, e.g. [10,11], which is problematic, we classify entire columns of pixels that contain those image artifacts. The detection result will then have an influence on the formulation of the cost function. To this end, we train a SVM classifier to classify individual columns of pixels in the polar coordinates into one of the following five categories: calcification, fibrous plaque, stent, guide-wire artifact, and normal tissue or soft plaque. Each of those has their characteristics; however, the difference between some categories may be small, e.g. calcification and fibrous plaque. To achieve efficient classification, the matching pursuit algorithm is used to reduce the number of support vectors.

2.3 Boundary Based Cost Function

The boundary based energy term can be expressed as $E_B = \sum_{V \in S} \hat{c}_B(x, y)$, where $\hat{c}_B$ denotes the normalised cost function ($\hat{c}_B(x, y) \in [0, 1]$) and S is a path in the directed graph. The formulation of the pre-normalisation cost function, c_B, is determined by the SVM classification result as presented below.

For normal tissue (or soft plaque), the media layer has a good contrast to adventitia. Hence, c_B is defined as $c_B(x, y) = D_1(x, y) - D_2(x, y)$ where D_1 is a summation of raw filtering response of the first order DoG at four different orientations and D_2 denotes maximum response of second order DoG filtering from different orientations across three scales. That is D_1 measures total edge strength and D_2 is rotational invariant measurement of bar-like feature. Note, the media layer is generally darker than the lower layer, adventitia. The first

order DoG filters are designed so that the stronger the media-adventitia border the lower the raw filtering response, i.e. negative values.

Calcified plaque exhibits strong edge features and casts varying degree of acoustic shadow on the media-adventitia border.

Thus, we use the second order DoG responses to suppress calcification and enhance possible media layer. Fibrous tissue behaves similarly to calcification, except in majority cases media-adventitia border is still discernible. Hence, bar feature detection is more appropriate and to enhance the effect we combine it with phase symmetry feature, i.e. $c_B(x, y) = -D_2(x, y) - FS(x, y)$ where FS is the local phase feature.

The presence of stent causes scattering of ultrasound signals, leading to very bright pixels. Once stent is detected by SVM, it is straight forward to localise the stent region which should not be part of media or adventitia. The cost for stent region is assigned a positive constant. Second order DoG responses are used to assign cost value for non-stent region. As for guide-wire artifact, there are also very bright pixels but it casts complete shadow over entire column. Hence, we do not extract any feature and a positive constant is used as their cost value.

2.4 Shape Prior Based Cost Function

The shape prior is defined as a likelihood term of each node in the graph, which is based on the similarity between the initial shape (obtained through finding the minimum closed set of our basic graph) and a set of templates from the training set. The graph construction is then modified so that inter-column arcs change dynamically according to the prior. The energy term for shape prior can be expressed as:

$$E_S = \sum_{x,y \in S} c_P(x, y) + \sum_{(p,q) \in \mathcal{N}} f_{p,q}(S(p) - S(q)), \tag{1}$$

where c_P denotes the cost function associated to prior and f is a convex function penalising abrupt changes in S between neighbouring columns p and q in the set $\mathcal{N}$ of neighbouring columns in the graph. The second term is realised through graph construction, detailed in the following Section 2.5. Notably in [7] the authors also used multiple templates in the graph cut. The terminal edge connection is determined by comparing the initial labelling with the template, e.g. if the node is in the template but not in the initial labelling, it connects to the source.

Each shape in the training set is treated as a binary template, ψ where the area inside shape is one and the outside area is zero. The distance between two templates ψ^a and ψ^b is defined using a discrete version of Zhu and Chan distance [12]:

$$d^2(\psi^a, \psi^b) = \sum_P (\psi^a - \psi^b)^2. \tag{2}$$

where P denotes the image domain. This distance measure is a true metric and is not influenced by image size. Let $\Psi = \psi^1, ..., \psi^N$ denote the N number of aligned

shapes from the training set. Given a possible cut in the graph which produces an aligned binary shape f, its similarity to a shape template ψ^n in the training set is computed as $\alpha(f,\psi^n) = exp(-\frac{1}{2\sigma^2}d^2(f,\psi^n))$. Thus, the likelihood of this particular cut can be evaluated by taking into account of all training shapes:

$$c_{R_0} = \frac{\sum_{n=1}^{N}\alpha(f,\psi^n)\psi^n}{\sum_{n=1}^{N}\alpha(f,\psi^n)}. \tag{3}$$

In our case, an initial cut can be conveniently obtained by minimising the boundary based cost function alone. Note, it is fully automatic and there is no need for user initialisation. The labelling of the shape likelihood and initial cut needs to be compared in order to assign appropriate terminal arcs. The shape prior cost is defined as:

$$c_P(x,y) = \lambda_1 |c_{R_0}(x,y) - c_{R_1}(x,y)|, \tag{4}$$

where c_{R_0} and c_{R_1} denote the cost associated to prior for the inferior region (the region under the border) and superior region (the region above the border) respectively, and λ_1 is the weight for the shape prior cost. The normalised weighted templates c_{R_0} is in effect the inferior-region cost and is inversely proportional to the likelihood of a pixel belong to the region underneath the media-adventitia border. To define the superior-region prior cost c_{R_1}, we simply compute the complement of c_{R_0}, i.e. $c_{R_1} = \max_{x,y} c_{R_0}(x,y) - c_{R_0}(x,y)$. As shown in Section 2.6, the shape prior cost $c_P(x,y)$ is used to assign weights for each pixel according to its position from the border. By assigning the shape prior cost in this way, we eliminate the need to identify the terminal connection type.

2.5 Graph Construction Using Shape Prior

In non-prior graph construction the inter-column maximum distance Δ is set as a constant. For our prior model, inter-column change should be influenced by the derived shape prior. In calculating the shape prior cost function, the training shapes are aligned to our initial graph cut. The inter-column changes are then generalised using mean $m_{p,q}$ and standard deviation $\sigma_{p,q}$ at individual column. These statistics are then used in determining maximum and minimum distances when connecting neighbouring columns in graph construction, i.e. $\bar{\Delta}_{p,q} = m_{p,q} + c \cdot \sigma_{p,q}$, $\underline{\Delta}_{p,q} = m_{p,q} - c \cdot \sigma_{p,q}$, and c is a real constant. Note, these inter-column arcs alone will impose a hard constraint on shape regularisation.

Hence, additional inter-column arcs are necessary in order to allow smooth transition (see dashed arcs in Fig. 2 (b)), that is intermediate values, $h \in [\underline{\Delta}_{p,q}, \bar{\Delta}_{p,q}]$, are used to construct inter-column arcs. The direction of these arcs is based on the first order derivative of the function $f_{p,q}(h)$ as in (1). Here, we employ a quadratic function, $f_{p,q} = \lambda_2(x - m_{p,q})^2$ where λ_2 is a weighting factor for smoothness constraint. If $f'_{p,q}(h) \geq 0$ an arc from $V(x,y)$ to $V(x+1, y-h)$ is established; otherwise, the arc is connected from $V(x+1,y)$ to $V(x,y+h)$. The weight for these arcs is assigned as the second order derivative of $f_{p,q}$. Note,

when $f'_{p,q}(h) = 0$, only single arc is defined to reduce the shape prior influence in presence of strong boundary features, instead of using bi-directional arcs on the mean difference $m_{p,q}$.

2.6 Compute the Minimum Closed Set

The cost function $C(x, y) = c_B(x, y) + c_P(x, y)$ is inversely correlated to the likelihood that the border of interest passes through pixel (x, y). The weight for each node on the directed graph can be assigned as:

$$w(x, y) = \begin{cases} C(x, y) & \text{if } y = 0, \\ C(x, y) - C(x, y - 1) & \text{otherwise.} \end{cases} \tag{5}$$

For a feasible path $\mathcal{P}$ in the graph, the subset of nodes on or below $\mathcal{P}$ form a closed set and it can be shown that the cost of $\mathcal{P}$ is equivalent to the cost of nodes in the corresponding subset (differ by a constant) [8]. Hence, segmenting the media-adventitia is equivalent to finding the minimum closed set in the directed graph. The s-t cut algorithm [3] can then be used to find the minimum closed set, based on the fact that the weight can be used as the base for dividing the nodes into nonnegative and negative sets. The source s is connected to each negative node and every nonnegative node is connected to the sink t, both through a directed arc that carries the absolute value of the cost node itself.

The smoothing parameter in graph construction prevents sudden drastic changes in the extracted interfaces. However, the segmented media-adventitia may still contain local oscillations. Here, efficient 1D RBF interpolation using thin plate base function is used to obtain the final interface.

3 Experimental Results

A total of 1197 IVUS images of 240×1507 pixels in the polar coordinates from 4 sequences are used to evaluate the proposed method. These images contain various forms of fibrous plaque, calcification, stent, and acoustic shadow. Manual labelling was carried out on every 10 frames, i.e. 1197 frames in total, to establish groundtruth for quantitative analysis. The training set contains 278 images.

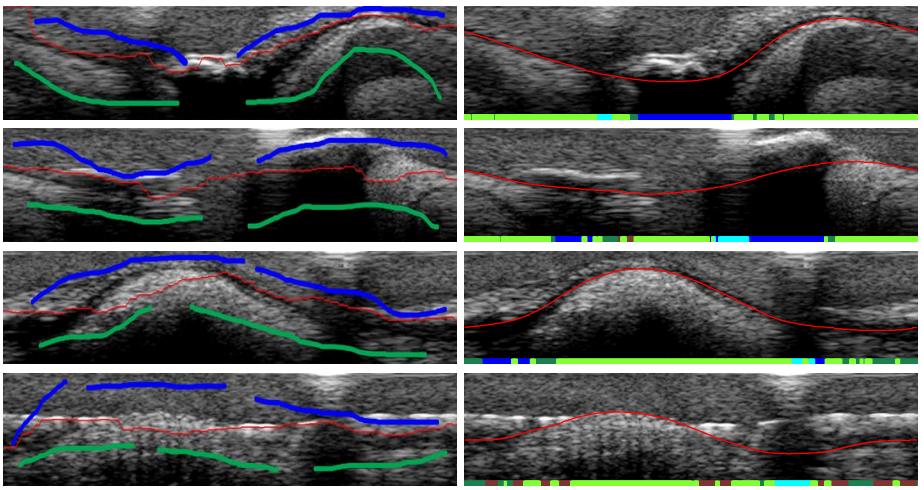
First, we compared our method against the $s - t$ cut algorithm [13]. The boundary cost function was kept the same, and careful manual initialisations were carried out for $s - t$ cut. The proposed method does not need user intervention. The first column in Fig. 3 shows typical results achieved using $s - t$ cut. Manual initialisations are shown in blue and green, and segmentation results are shown in red. Despite reasonable care in initialisation, the $s - t$ cut result was not satisfactory. The corresponding results of the proposed method are shown in the second column. The bottom of the each image shows the classification result of detecting different types of tissue. The proposed method achieved better accuracy and consistency. The quantitative comparison was carried out on a randomly selected subset of 50 images, since manual initialisation of 1197 images is too labour intensive. Table 1 shows that the proposed method clearly

outperformed $s-t$ cut in both area difference measure (AD) and absolute mean difference measure (AMD) based on groundtruth.

Next, the proposed method was tested on the full dataset (1197 images) and its performance based on 1197 labelled groundtruth can be summarised as: 9.00 % mean AD with standard deviation of 6.35 and 9.16 pixel mean AMD with standard deviation of 6.20. This is marginally better than the first subset. Fig. 4 shows example comparisons to groundtruth. It is evident that the proposed method can handle various forms of ultrasound artifacts. Overall, the quantitative results suggest that the proposed method is an effective method in segmentation media-adventitia border in IVUS.

Table 1. Quantitative comparison to $s-t$ cut. AD: area difference in percentage; AMD: absolute mean difference in pixel in comparison to groundtruth.

	$s-t$ cut		proposed method	
	AD	AMD	AD	AMD
Mean	22.54	23.91	9.286	10.05
Std.	8.87	7.49	5.03	5.41



(a) (b)

Fig. 3. (a) $s-t$ cut result (red) with user initialization (object: blue, background: green). (b) proposed method result; the bottom of each image also shows the classification result: calcified plaque (blue), fibrotic plaque (dark green), stent (dark red), guide-wire shadowing (cyan), and soft plaque/normal tissue (light green).

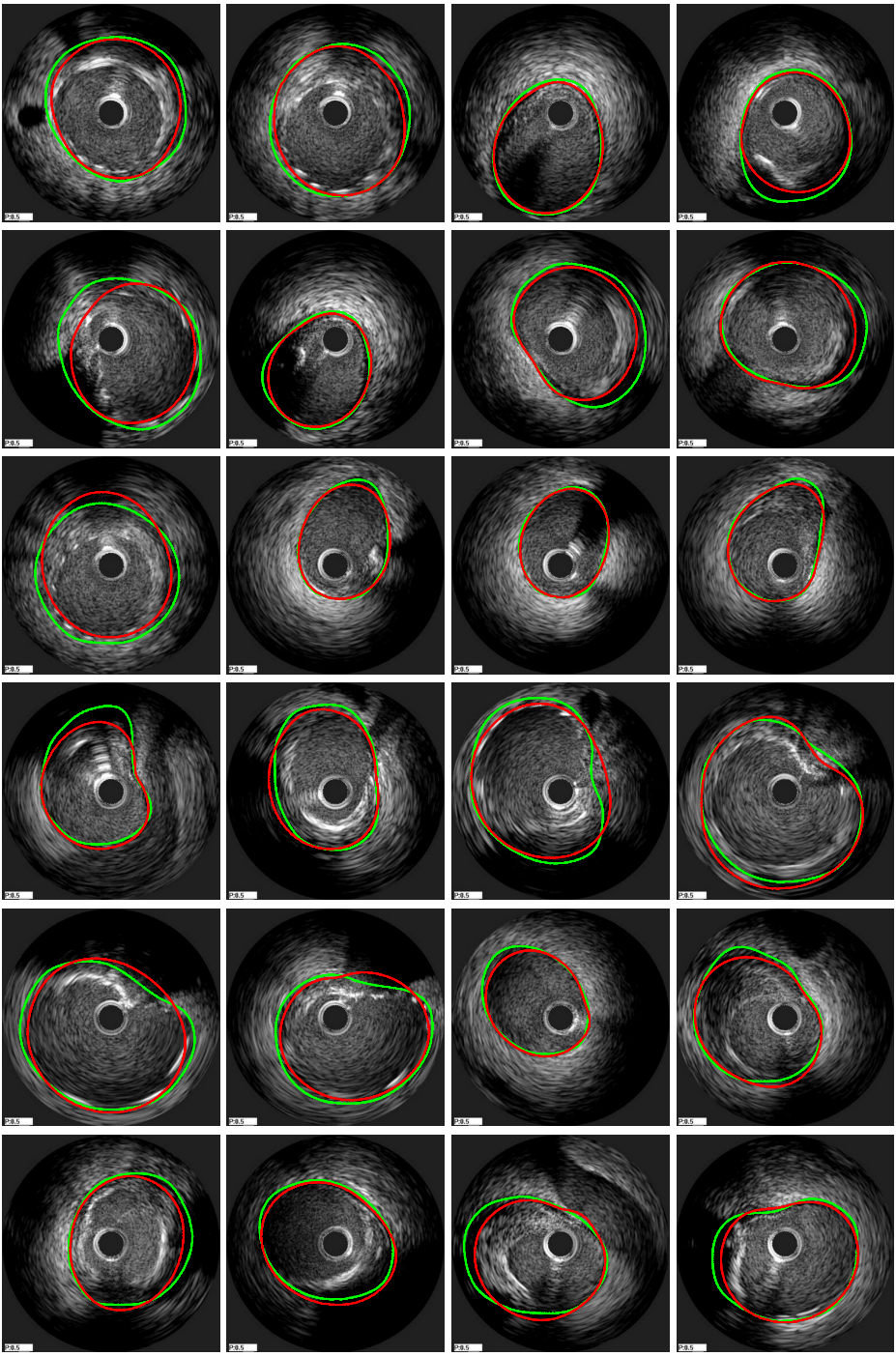


Fig. 4. Comparison between groundtruth (green) and the proposed method (red)

4 Conclusions

We presented an automatic graph based segmentation method for delineating the media-adventitia border in IVUS images. Boundary based features were dynamically selected to optimise the cost function. The use of multiple training shapes proved to be beneficial. The generalised shape prior was used in both incorporating the cost function but also graph construction. Smoothness constraint was intrinsically imposed in graph construction. Qualitative and quantitative results on a large number of IVUS images showed superior performance of the method.

Acknowledgement. We would like to thank Welsh Government NISCHR for funding this research work (Grant ID: HA09/035).

References

1. Sonka, M., et al.: Segmentation of intravascular ultrasound images: A knowledge-based approach. T-MI 14, 719–732 (1995)
2. Takagi, A., et al.: Automated contour detection for high frequency intravascular ultrasound imaging: A technique with blood noise reduction for edge enhancement. Ultrasound in Medicine and Biology 26(6), 1033–1041 (2000)
3. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. T-PAMI 26(9), 1124–1137 (2004)
4. Wahle, A., et al.: Plaque development, vessel curvature, and wall shear stress in coronary arteries assessed by x-ray angiography and intravascular ultrasound. MIA 10(1), 615–631 (2006)
5. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: CVPR, pp. 755–762 (2005)
6. Malcolm, J., Rathi, Y., Tannenbaum, A.: Graph cut segmentation with nonlinear shape priors. In: ICIP, pp. 365–368 (2007)
7. Vu, N., Manjunath, B.S.: Shape prior segmentation of multiple objects with graph cuts. In: CVPR, pp. 1–8 (2008)
8. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images-a graph-theoretic approach. T-PAMI 28(1), 119–134 (2006)
9. Mulet-Parada, M., Noble, J.: 2D + T acoustic boundary detection in echocardiography. MIA 4(1), 21–30 (2000)
10. Filho, E., et al.: Detection & quantification of calcifications in ivus by automatic thresholding. Ultrasound in Medicine and Biology 34(1), 160–165 (2008)
11. Unal, G., et al.: Shape-driven segmentation of the arterial wall in intravascular ultrasound images. IEEE Trans. Info. Tech. Biomed. 12(3), 335–347 (2008)
12. Chan, T., Zhu, W.: Level set based shape prior segmentation. In: CVPR (2005)
13. Boykov, Y., Funka-Lea, G.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. IJCV 70(2), 109–131 (2006)

Multiple Atlases-Based Joint Labeling of Human Cortical Sulcal Curves

Ilwoo Lyu[1], Gang Li[2], Minjeong Kim[2], and Dinggang Shen[2]

[1] Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599, USA
ilwoolyu@cs.unc.edu
[2] Department of Radiology and BRIC, University of North Carolina, Chapel Hill, NC 27599, USA
{gang_li,mjkim,dgshen}@med.unc.edu

Abstract. We present a spectral-based sulcal curve labeling method by considering geometrical information of neighboring curves in a multiple atlases-based framework. Compared to the conventional method, we propose to use neighboring curves for avoiding ambiguity in curve-by-curve labeling and to integrate the labeling results obtained from multiple atlases for consistent labeling. In particular, we compute a histogram of points on the neighboring curves as a new feature descriptor for each point on a sulcal curve under consideration. To better resolve ambiguity in the curve labeling, we also employ the neighboring curves that are parallel to major sulcal curves. Moreover, we further integrate all the results from multiple atlases into a linear system, by solving which our method ultimately gives accurate labels to the major curves in the subjects. Experimental results on evaluation of 12 major sulcal curves of 12 human cortical surfaces indicate that our method achieves higher labeling accuracy 7.87% compared to the conventional method, while reducing 4.41% of false positive labeling errors on average.

Keywords: sulcal curve labeling, multiple atlases, spectral matching.

1 Introduction

The sulcal folding patterns of human cortical fundic regions are used as key features for analyzing brain function, monitoring brain growth, and discovering diseases. Since sulcal curves can be defined along fundic regions, automatic labeling of sulcal curves is important for these studies. There have been recent studies on automatic extraction of sulcal curves on human cortical surfaces [1,2]. However, these methods extract not only major curves but also many extraneous minor curves, which should be further removed for sulcal curve labeling. Due to the extremely complicated and variable sulcal folding patterns and extraneous minor sulcal branches, even if sulcal curves can be perfectly extracted, it is still challenging to identify major curves among the automatically extracted ones.

Atlas(es)-based sulcal curve labeling methods have been proposed for automatic labeling of major curves [3,4,5]. Compared to the single atlas-based methods [3,4], the multiple atlases-based labeling method is thought to be able to

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 124–132, 2013.

give more accurate labels by considering individual sulcal variability. Recently, a spectral-based sulcal curve labeling method using multiple atlases has been reported [5]. In their method, they just picked the most matched sulcal curve from the multiple atlases to label the corresponding curve in the subject. The correspondence is established by solving an affinity matrix that stores all possible assignments based on the geometric features between two curves under consideration. However, there are two main drawbacks in their method. First, since only the best matched curve is considered as the candidate to label the subject, large false positive errors can be introduced if there is no similar curve in the atlases or the number of atlases is too small. Second, the labeling process is done independently for each major curve without considering its neighboring curves. This could reduce a chance for the major curves to be accurately labeled due to the ambiguity in the curve matching.

In this paper, we present a sulcal curve labeling method for cortical surfaces, which jointly exploits the geometric information of multiple atlases and neighboring curves in the subject space. We focus on "finding correct assignments", which can be formulated as a linear system similarly as in [6]. Specifically, for the feature description, each curve stores its neighboring curves' information (i.e., a histogram of position information of points on the neighboring curves), and in the curve matching, a major curve finds the most similar curves in the subject, guided by its neighboring curves. In addition, we incorporate all labeling results obtained from multiple atlases since it is likely that major curves in the atlases are only partially similar to those in the subject. To this end, we extend the affinity matrix in [6] to integrate labeling results into a linear system. Experimental results indicate that our method achieves 7.87% improvement of labeling accuracy as well as 4.41% reduction of false positive labeling errors on average for 12 major curves on 12 cortical surfaces, compared to the conventional method [5].

2 Method

Given a set of sulcal curves P in atlases and that of unlabeled sulcal curves Q in the subject, our goal is to label major curves in Q while discarding minor ones in Q. Note that the curves in P are pre-labeled major curves by following neuroanatomical conventions while Q contains (possibly disconnected) major curves and many minor ones. For curve labeling, we first automatically extract sulcal curves from the triangulated cortical surface using [1] and deform all curves in each atlas to the subject space using a diffeomporphic surface registration method [7]. It is worth noting that landmark-free surface registration methods can only roughly align the sulcal folding patterns [8], thus still leaving a certain amount of ambiguity in the curve labeling (see Fig. 1a). To better resolve ambiguity in the labeling, unlike the "hard" matching strategy in the conventional method, we use the geometric features of the major curve and its nearby curves for measuring curve similarity. Moreover, the final label is jointly determined by all atlases, which differs from the conventional method that directly retrieves the label from the most similar curve in a selected atlas.

2.1 Spectral-Based Curve Matching Using Neighboring Curves

To measure similarity for every possible pair of curves $p \subseteq P$ and $q \subseteq Q$, we basically measure the individual and pairwise affinities of an assignment $a = (p_i, q_j)$, where $p_i \in p$ and $q_j \in q$. For an assignment a, we denote $D(a)$ as the displacement vector between geometric features of p_i and q_j, each element of which is normalized with respect to its maximum value. Let w be a nonnegative weight vector that gives the importance of every element in $D(a)$. The individual affinity is then defined as follows:

$$A(a) = \exp(-\frac{\|D(a)\|_w^2}{2\sigma^2}) \ , \tag{1}$$

where $\|D\|_w$ denotes the weighted L_2-norm of D with respect to the weight vector w and σ is a user-provided regularization parameter. Similarly, for two distinct assignments a and b, the pairwise affinity is given by

$$A(a, b) = \exp(-\frac{\|D(a, b)\|_w^2}{2\sigma^2}) \ , \tag{2}$$

where $D(a, b) = D(a) - D(b)$.

Geometric Features Considering Neighboring Curves. Several geometric features are defined for each sulcal point, i.e., positions, curvatures, and unit tangent vectors from the major curve under consideration. Besides, we further incorporate the features from its neighboring curves. Basically, we calculate a histogram based on the position information of the neighboring curves in the Euclidean space. Given a major curve $p = \{p_1, \cdots, p_i, \cdots, p_N\}$ with N sulcal points for $p \subseteq P$, let S_p be a set of its neighboring curves. To compute a histogram of the neighboring sulcal points around a point $p_i \in p$, we first build a spherical kernel K centered at p_i with radius r. The size of r is automatically determined by the maximum Hausdorff distance between p and s for $\forall s \subseteq S_p$.

$$r = \max_{s \subseteq S_p} d_H(p, s) \ , \tag{3}$$

where $d_H(\cdot, \cdot)$ denotes the Hausdorff distance between two curves. The size of K is identical for any point on p. Let $F(\cdot)$ be the position-information vector of a sulcal point in the atlases, which stands for location information in the Euclidean space. Once the size of spherical kernel K is determined, an initial set of neighboring points L_{p_i} within K is obtained as follows:

$$L_{p_i} = \left\{ x \mid x \in s \subseteq S_p, \frac{\|F(x) - F(p_i)\|^2}{r^2} \leq 1 \right\} \ . \tag{4}$$

Our interest is to find sulcal points on the neighboring curves that are "parallel" to curve p, referring to those with similar global shapes and orientations to p. To emphasize such neighboring points in L_{p_i}, we apply the principal component

analysis (PCA) on L_{p_i} since the principal direction u_1 of L_{p_i} stands for the direction of the parallel curves. We then discard as many sulcal points on the neighboring curves as possible that are not parallel to curve p within K, by reducing spherical kernel K to an ellipsoid with its three axes aligned to the three eigenvectors of PCA, $u_n, n = 1, 2, 3$. The eigenvalue λ_1 is given along the first major axis. We then have the following final set of neighboring points L'_{p_i} by letting $l_1 = \sqrt{\lambda_1}$ and $l_2 = l_3 = r$:

$$L'_{p_i} = \left\{ x \mid x \in L_{p_i}, \sum_{n=1}^{3} \frac{((F(x) - F(p_i)) \cdot u_n)^2}{l_n^2} \leq 1 \right\} . \tag{5}$$

Now, we build a bounding cube centered at p_i that fully contains the neighboring sulcal points in L'_{p_i}. Then, we uniformly divide the cube into m subvolumes. Let h_k be a ratio of points in L'_{p_i} that belong to a subvolume b_k, $1 \leq k \leq m$. We finally have a histogram $H_{p_i} = [h_1, h_2, \cdots, h_m]^T$ by the following equation.

$$h_k = \frac{\sum_{x \in L'_{p_i}} I(x, b_k)}{\left| L'_{p_i} \right|} , \tag{6}$$

$$I(x, b_k) = \begin{cases} 1 & \text{if } \{x\} \cap b_k \nsubseteq \emptyset, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

For a sulcal point q_j in the subject, it is difficult to compute its actual spherical kernel because its neighboring major curves are unknown. Therefore, for an assignment $a = (p_i, q_j)$, we use the same kernel as p_i in the atlas for computing the histogram of q_j.

Synchronized Curve Matching. To account for sulcal shape variability, we generate the mean curve for each major curve [5]. We denote $\phi(\cdot)$ as the corresponding point on the mean curve to a given sulcal point in the atlas. For an assignment $a = (p_i, q_j)$, we now set a threshold of the distance between p_i and q_j with respect to the covariance of $\phi(p_i)$. Thus, the assignment a is rejected if

$$\sum_{n=1}^{3} \frac{((F(q_j) - F(p_i)) \cdot v_n)^2}{(3\tau_n)^2} > 1 , \tag{8}$$

where $\tau_n^2 (n = 1, 2, 3)$ are the covariances along the corresponding principal axes of the covariance matrix of $\phi(p_i)$. This constrains assignments statistically valid in terms of the sulcal shape variability.

Let s be a neighboring curve for a given major curve p as we defined above. We first measure affinities for p and s, respectively. To incorporate affinities of the neighboring curves into the affinity matrix M, we also measure all possible pairwise affinities between p and s. For $p_i \in p$ and $s_{i'} \in s$, suppose that assignments are given by $a = (p_i, q_j)$ and $b = (s_{i'}, q_{j'})$, where q_j, $q_{j'} \in q \subseteq Q$. Since a major curve is unable to share an identical label with its neighboring curves,

in such a undesirable case of the coexistence of a and b, the pairwise affinity between a and b is set to zero. Once M is built, we compute the principal eigenvector of M to find the highly confident assignments. Since s only helps find the correspondences between p and q, the possibly remaining assignments in s will be left out.

2.2 Joint Labeling Using Multiple Atlases

It is worth noting that major curves in the atlases could be only partially similar to those in the subject. For all major sulcal curves in P, once the highly confident assignments with the corresponding curves in Q are selected, we incorporate the assignments to determine final labels based on their correspondences. Let p^α and p^β be the distinct major sulcal curves in P with an identical label. For two distinct assignments $a = (p_i^\alpha, q_j)$ and $b = (p_{i'}^\beta, q_{j'})$, it is highly desirable that $q_j = q_{j'}$ if $\phi(p_i^\alpha) = \phi(p_{i'}^\beta)$. To implement that idea, we construct a new affinity matrix M that describes relationships of all possible assignments between p^α and p^β. The diagonal entries of M are filled with confidence values that are obtained from the principal eigenvector of the affinity matrix in Sect. 2.1. For two distinct assignments $a = (p_i^\alpha, q_j)$ and $b = (p_{i'}^\beta, q_{j'})$, $M(a, b)$ is set to $A(a, b)$ as defined in Eq. 2. Then, $M(a, b)$ is updated as follows by letting $c = (q_j, q_{j'})$ if $\phi(p_i^\alpha) = \phi(p_{i'}^\beta)$:

$$M(a, b) = A(a, b) \cdot A(c) \; . \tag{9}$$

Finally, we compute the principal eigenvector of M to select the highly confident assignments for the joint labeling.

3 Experimental Results

Since the dataset in [5] is not publicly available, we used the MRIs Surfaces Curves dataset [8] for validation (total 12 subjects). However, in this dataset, several major curves delineated by experts were still crossed gyral regions, which slightly differ from the automatically extracted curves we used in the experiment. Thus, we generated ground-truth curves by combining the manual delineation results with the automatically extracted sulcal curves.

Given an automatically labeled curve q and its corresponding ground-truth curve q_g, the labeling accuracy $acc(q, q_g)$ and false positive labeling error $err(q, q_g)$ were measured by the following equations:

$$acc(q, q_g) = \frac{l(q \cap q_g)}{l(q_g)} \text{ and } err(q, q_g) = \frac{l(q - q_g)}{l(q_g)} \; , \tag{10}$$

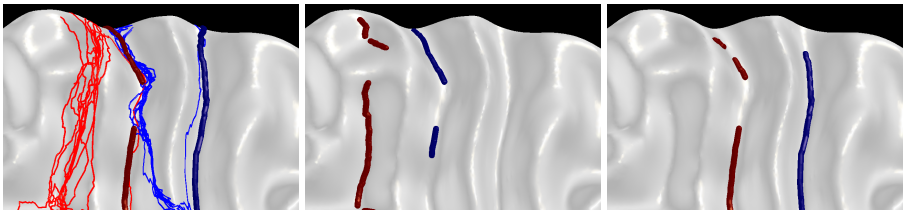
where $l(\cdot)$ denotes the length of a curve.

In our experiment, we adapted a jackknife technique to validate the accuracy and false positive errors: For each validation set, one subject was leaved out from the subject set to be labeled, and other subjects were regarded as the atlases.

Table 1. 12 Major curves and their neighboring curves

Curve	Neighbors	Curve	Neighbors	Curve	Neighbors	Curve	Neighbors
STS	ITS	ITS	STS, OTS	CS	preCS, postCS	preCS	CS
postCS	CS	SFS	IFS	IFS	SFS	CingS	-
CalcS	colS	OcPS	-	OTS	ITS, colS	colS	OTS, CalcS

12 out of major curves for both left and right hemispheres were used for validation: the superior temporal sulcus (STS), inferior temporal sulcus (ITS), central sulcus (CS), precentral sulcus (preCS), postcentral sulcus (postCS), superior frontal sulcus (SFS), inferior frontal sulcus (IFS), cingulate sulcus (CingS), calcarine sulcus (CalcS), occipito parietal sulcus (OcPS), occipito temporal sulcus (OTS), and collateral sulcus (colS). We selected the neighboring curves for each major sulcal curve based on neuroanatomical prior knowledge as summarized in Table 1. For fair comparison of different methods in all experiments, we used the same set of the deformed atlases obtained by the same registration method [7], even for the conventional method.



(a) (b) (c)

Fig. 1. Poorly deformed atlases and labeling results for the central sulcus (blue) and postcentral sulcus (red): (a) deformed atlases (thin curves) and the ground-truth curves (bold curves), (b) the labeling results by the conventional method, and (c) the labeling results with neighboring curves

3.1 Neighboring Curves

We employed neighboring curves and chose the most similar curve among multiple atlases for the final result. For the histogram computation, we subdivided the bounding cube into $4 \times 4 \times 4$ subvolumes, i.e., $m = 64$. For the affinity matrix computation, we set the weight vector $w = [0.75, 0.15, 0.05, 0.05]^T$ and the regularization parameter $\sigma = 0.3$. Each of the elements in w corresponds to weight of the position, curvature, tangent vector, and histogram of neighboring sulcal points, respectively. We rejected an assignment if the norm of the difference between the two histograms is greater than 0.1. Note that the parameters were empirically set according to [5] and by our experiment. In Fig. 1, the labeling results with neighboring curves are consistent although the atlases are poorly deformed. The results with neighboring curves exhibited better agreement

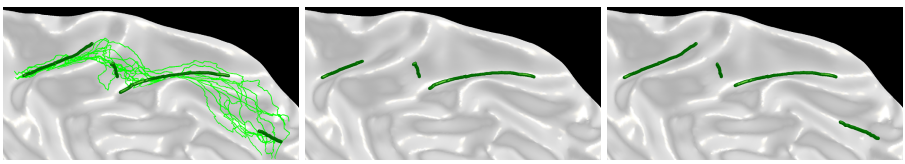
Table 2. Average labeling accuracy and false positive errors in the left (lh) and right hemispheres (rh) (unit: %):

	Conventional method		Neighboring curves (a)		Joint labeling (b)		Our method (a+b)	
	lh	rh	lh	rh	lh	rh	lh	rh
Accuracy	68.65	69.19	71.22	72.27	74.53	74.87	77.12	76.47
False positives	20.22	19.85	25.06	23.41	16.82	15.53	15.79	15.46

with the ground-truth than the conventional spectral-based method as summarized in Table 2. Interestingly, the average false positive errors also increased because several false positive assignments that had a low confidence value in the conventional method can gain a higher confidence, resulting from guidance of neighboring curves.

3.2 Joint Labeling Using Multiple Atlases

We applied the joint labeling without guidance of neighboring curves. The results obtained from 12 atlases were incorporated to determine the final label to each major sulcal curve. The same parameter setting as in Sect. 3.1 was used here. Figure 2 shows that the the joint labeling also gives labels to a part of major sulcal curves that is missed in the conventional spectral-based method. Compared to the conventional method, the labeling accuracy increased while the false positive errors decreased as summarized in Table 2.



(a) (b) (c)

Fig. 2. Comparison of results by the conventional spectral-based method and joint labeling for the superior frontal sulcus: (a) deformed atlases (thin curves) and the ground-truth curves (bold curves), (b) the labeling results by the conventional spectral-based method, and (c) the labeling results by the joint labeling

3.3 Overall Performance

By incorporating two aspects, i.e., synchronized matching with neighboring curves and joint labeling using multiple atlases, into the our framework, we obtained the overall labeling accuracy and false positive errors as summarized in Table 2.

The labeling performance by our method was highly achieved after incorporating the two aspects. Also, our labeling results were comparable to the corresponding ground-truth curves (see an example in Fig. 3). Figure 4 demonstrates the statistical comparison of the labeling results for 12 major sulcal curves. The results show the average accuracy and false positive errors across subjects. This indicates that our labeling results were consistent on most of the curves, compared to the conventional method.

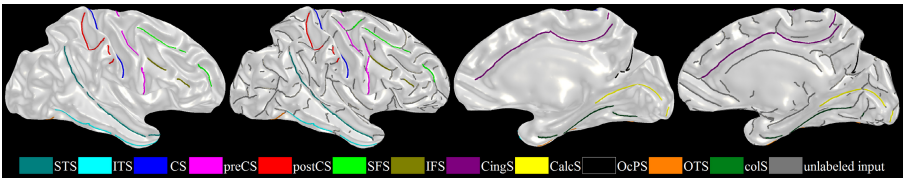
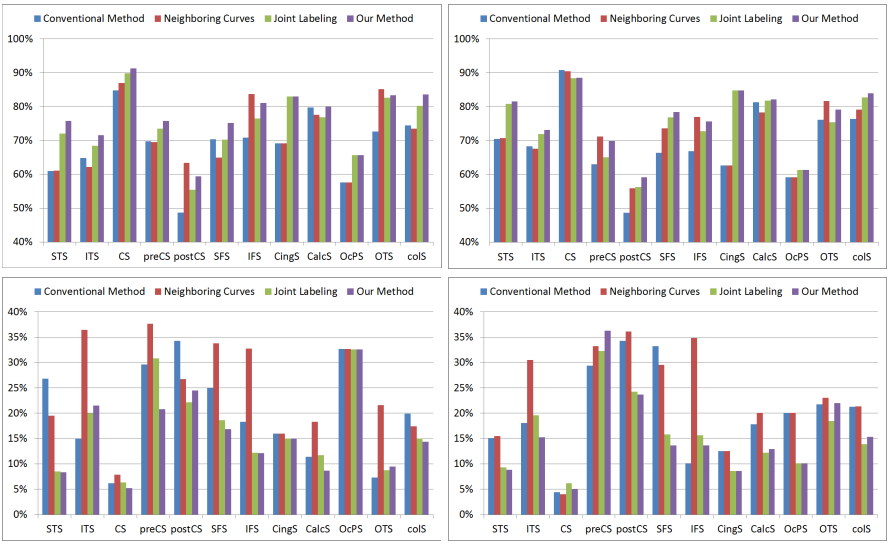


Fig. 3. A visual comparison of our automatic labeling results with the ground-truth for the right hemisphere: the lateral and medial views of ground-truth labeled curves (1st and 3rd columns) and the respective views of automatically labeled curves by our method (2nd and 4th columns). Note that there are many extraneous minor curves in the input (gray). For better visualization, a partially inflated surface model is used.



(a) left hemispheres (b) right hemispheres

Fig. 4. Performance comparisons: average labeling accuracy (top row) and false positive errors (bottom row) for major sulcal curves in the left and right hemispheres

4 Conclusion

We presented a method for multiple atlases-based labeling of major sulcal curves on the cortical surface. Specifically, to resolve ambiguity in the labeling, we proposed a histogram feature for each sulcal point and incorporated the geometric information of neighboring curves into the affinity matrix for the curve matching. Since major curves in the atlases are likely to be partially similar to those in the subject, we incorporated the results obtained from all atlases into the linear system for accurate labeling. We have shown in experiment that compared to the conventional method, the performances were improved for 7.87% labeling accuracy and reduced for 4.41% of false positive errors. In our future work, we will employ a learning technique for optimizing parameters used in the curve matching.

References

1. Li, G., Guo, L., Nie, J., Liu, T.: An automated pipeline for cortical sulcal fundi extraction. Medical Image Analysis 14, 343–359 (2010)
2. Seong, J., Im, K., Yoo, S., Seo, S., Na, D., Lee, J.: Automatic extraction of sulcal lines on cortical surfaces based on anisotropic geodesic distance. Neuroimage 49, 293–302 (2010)
3. Lohmann, G., Von Cramon, D.: Automatic labelling of the human cortical surface using sulcal basins. Medical Image Analysis 4, 179–188 (2000)
4. Tao, X., Prince, J., Davatzikos, C.: Using a statistical shape model to extract sulcal curves on the outer cortex of the human brain. IEEE Trans. on Medical Imaging 21, 513–524 (2002)
5. Lyu, I., Seong, J., Shin, S., Im, K., Roh, J., Kim, M., Kim, G., Kim, J., Evans, A., Na, D., et al.: Spectral-based automatic labeling and refining of human cortical sulcal curves using expert-provided examples. Neuroimage 52, 142–157 (2010)
6. Leordeanu, M., Hebert, M.: A spectral technique for correspondence problems using pairwise constraints. In: Computer Vision, ICCV 2005, vol. 2, pp. 1482–1489. IEEE (2005)
7. Yeo, B., Sabuncu, M., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical demons: Fast diffeomorphic landmark-free surface registration. IEEE Trans. on Medical Imaging 29, 650–668 (2010)
8. Pantazis, D., Joshi, A., Jiang, J., Shattuck, D., Bernstein, L., Damasio, H., Leahy, R.: Comparison of landmark-based and automatic methods for cortical surface registration. Neuroimage 49, 2479–2493 (2010)

Fast Anatomical Structure Localization Using Top-Down Image Patch Regression

René Donner[1,2,⋆], Bjoern H. Menze[3,4,5], Horst Bischof[2], and Georg Langs[1,3]

[1] Computational Image Analysis and Radiology Lab, Department of Radiology, Medical University Vienna, Austria
[2] Institute for Computer Graphics and Vision, Graz University of Technology, Austria
[3] CSAIL, MIT, Cambridge MA, USA
[4] Asclepios Project, INRIA Sophia-Antipolis, France
[5] Computer Vision Laboratory, ETH Zurich, Switzerland
rene.donner@meduniwien.ac.at

Abstract. Fully automatic localization of anatomical structures in 2D and 3D radiological data sets is important in both computer aided diagnosis, and the rapid automatic processing of large amounts of data. We present a simple, accurate and fast approach with low computational complexity to find anatomical landmarks, based on a multi-scale regression codebook of informative image patches and encoded landmark contexts.

From a set of annotated training volumes the method captures the appearance of landmarks over several scales together with relative positions of neighboring landmarks and a spatial distribution model. During multi-scale search in a target volume, starting from the coarsest level, each landmark model predicts all landmark positions it has encoded, with the median of all predictions yielding the final prediction for each scale.

We present results on two challenging data sets (hand radiographs and hand CTs), where our method achieves comparable accuracy to the state of the art with substantially improved run-time.

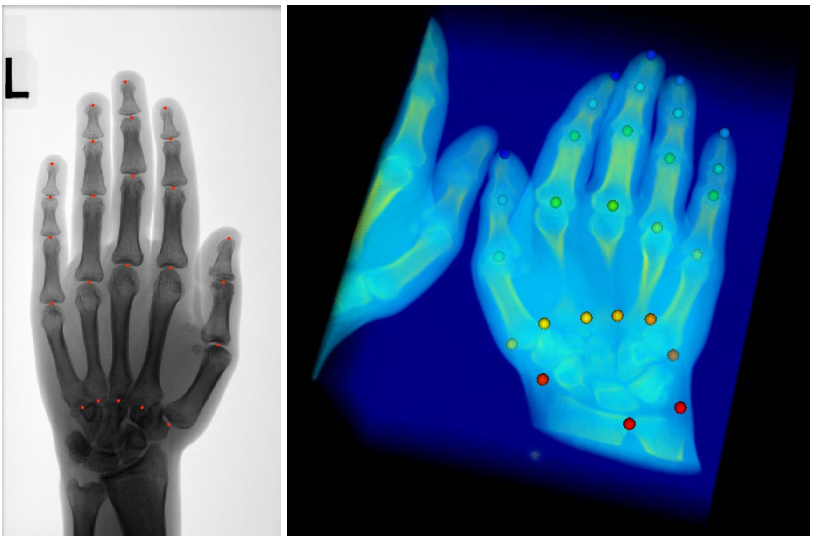
Keywords: Anatomical structure localization, nearest neighbor regression, image patch codebooks.

1 Introduction

The accurate localization of anatomical landmarks in medical imaging data is a challenging problem, due to rich variability and frequent ambiguity of their appearance. Among the reasons for the difficulties are noise (including local

⋆ This work was partly supported by the European Union FP7 Project KHRESMOI (FP7-257528), by the Austrian National Bank grants BIOBONE (13468) and AORTAMOTION (13497) and the Austrian Sciences Fund grant PULMARCH (P 22578-B19).

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 133–141, 2013.



(a) (b)

Fig. 1. Examples from the two data sets employed in this paper. a) Hand radiographs and b) high resolution hand CTs. The objective of the proposed method is to localize the depicted anatomical landmarks in an unseen target image or volume.

and global intensity changes), cluttered image data (overlapping structures in 2D projections, highly structured background in 3D organ segmentation), and anatomical structures that exhibit a high degree of similarity (e.g., fingers or vertebrae). We propose an algorithm that copes with these challenges and offers a general approach to accurately localize landmarks without initialization or subsequent refinement. The method constructs a multi-level regression codebook which associates image patches with the corresponding positions of anatomical landmarks depicted in the patch. During search the scale-pyramid is traversed, finding the most similar patch for each landmark using k-nearest neighbor search.

The localization of anatomical structures is crucial for several areas of medical imaging analysis: Segmentation approaches such as Level-Sets [4] and Appearance Models [3], typically require at least a coarse initial localization, while registration approaches can exploit spatial initialization to avoid local minima. The automatic localization of anatomical structures is fundamental for the field of Computer Aided Diagnosis [7] and for structuring image information in image retrieval, since it allows the algorithms to focus on target regions in the data and subsequently invoke more specialized analysis stages. Landmark localization can also be regarded as a form of semantic parsing [13] when point-wise rather than regional information is required.

State of the art. Several approaches to anatomical structure localization exist in recent literature. They mainly differ in the type of semantic representation that is obtained to describe the image data. We thus distinguish between approaches

that either 1) indicate the *positions* of individual landmarks, 2) provide *bounding boxes* for entire organs, 3) result in *model parameters* which describe the position and shape of the object or 4) provide *voxel-wise labels* for different organs.

Localizing anatomical landmarks using the *positions* of selected interest points has been the objective of [8,1]. The methods learn interest point detectors on training data, estimate positions of landmark candidates in the target volume and finally disambiguate these candidates through a model matching step. Both methods rely on the classification of the entire volume. [9] reduces this computational burden by performing a low-resolution step and a refinement step using Hough regressors. Reducing the complexity by working on axial slices, [13] parse whole body CT data in a hierarchical fashion, but are concerned with finding larger organs. While substantially speeding up the localization this only works for objects which are rather large in respect to the overall volume size, since the objects have to be visible in at least one of the three central orthogonal slices. Using Random Forests for the localization of organs in thorax CTs through *bounding boxes* has been been proposed in [5]. An extension using Hough ferns was presented in [12] to predict the bounding boxes of multiple organs at once in full-body MR data. Relying on stochastic optimization instead of ensemble classification or regression, Marginal Space Learning [15] tries to find the parameters of a bounding box or a parametric and data-driven *shape model* [2] to localize and segment anatomical structures. This allows for fast localization, but instead of representing a global search algorithm, iterative approaches have to be used to cope with repetitive structures [10]. The task of assigning *voxel-wise labels* to segment entire organs or organ structures has been approached by [6] and [11] using Random Forest classification.

Contribution. We present a simple, fast method for the global, accurate localization of anatomical structures in 2D/3D data based on an appearance codebook, and location predictors that capture sub-configurations of a landmark set. It demonstrates that a top-down nearest neighbor matching strategy of image patches drastically reduces the number of required feature computations and yields localization results comparable to the state of the art.

Paper structure. The paper is structured as follows: Sec. 2.1 details the construction of the codebook, with the localization on a target volume described in Sec. 2.2. Sec. 3 introduces the experiments, with the results presented in Sec. 3.3. A discussion and an outlook can be found in Sec. 3.4 and Sec. 4.

2 Methods

The approach is divided into a training phase and a localization phase as shown in Fig. 2 and Fig. 3. During localization a multi-scale codebook of image patches and landmark positions is constructed, which is traversed during the localization phase to obtain increasingly accurate landmark estimates at each scale.

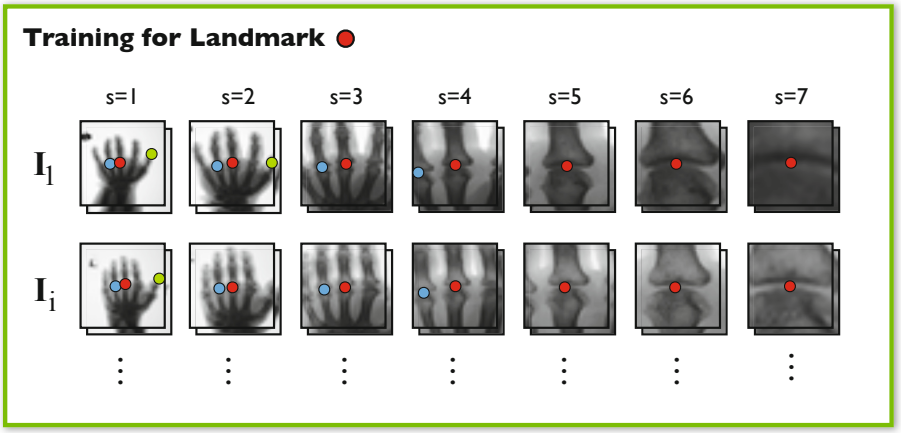


Fig. 2. Construction of the regression codebooks during training. For each landmark and scale patches at various offsets and the corresponding relative landmark positions are recorded, using all training images/volumes.

2.1 Training – Constructing the Landmark Regression Codebook

The training phase requires a set of N training images or volumes $\mathbf{I}_i$ with corresponding annotations. The annotations represent the coordinates $\mathbf{x}_x^i$ of the $x \in \{1, \ldots, L\}$ landmarks of the anatomical structure in question. Each landmark is present in each of the training volumes.

Codebook Construction to Connect Local Appearance and Landmark Information. Our aim is to build multi-scale regression codebooks $\mathcal{C}$ of image patches and corresponding relative landmark positions – one codebook per scale $s \in 1, \ldots, S$ and landmark x. The patches stored in the codebook are extracted around the landmarks with varying offsets and scaling, capturing the typical visual appearance around each landmark. For each patch the positions of all landmarks visible in the patch are recorded, relative to the patch's center. Each of the PN entries in the codebook $\mathcal{C}_{s,x}$ consists of the tuple $\langle \mathbf{P}^p, \mathbf{L}^p \rangle$ of the patch $\mathbf{P}^p$ and the corresponding relative $D \times L$ landmark coordinates $\mathbf{L}^p$ which are visible in the patch. $\mathbf{L}^p$ specifies the coordinates of the landmarks $x \in 1 \ldots L$ relative to the center of the given patch[1]. Landmarks which are outside of the patch are denoted as not visible.

The construction of the codebook proceeds as follows: At the top-most scale $s = 1$ each image or volume is represented by an an-isotropically downscaled miniature of size $m \times m \times m$ (similarly $m \times m$ for images). At each scale s the volume is considered to possess an edge length of $\sqrt{2}(s-1)m$. This re-sampling of the entire image is never actually computed, it simply forms the reference frame for each scale of the codebook generation.

[1] The necessary transformations between image coordinates and patch coordinates are omitted for clarity throughout the text.

At each scale s, patches $\mathbf{P}$ are extracted from the image or volume data using linear interpolation for each landmark x from all training volumes N. The patches are of size $m \times m \times m$, i. e. at scale $s = 1$ they correspond to the entire image, and for scales $s > 1$ the patches *zoom in* on the landmark, as illustrated in Fig. 2. Parts of patches which would be sampled from outside of the volume are set equal to the closest voxel on the volume's border. The gray values of each patch is normalized to zero mean and unit variance.

To explore the image information in the vicinity of a landmark the entries in the codebook $\mathcal{C}_{s,x}$ at a certain scale s and landmark x, are constructed by extracting several patches around the landmark with, empirically chosen, 7 offsets in the range of $[-6, 6]$ voxels for each dimension, along with scaling factors of $\{0.9, 1, 1.1\}$, resulting in $P = 1029$ patches for one landmark in one training volume at one scale ($P = 147$ for images). To considerably reduce the memory requirements and computational complexity for the codebook lookup, dimensionality reduction of each codebook is performed using PCA, retaining 90% of variance, resulting in PCA coefficients $\mathbf{P}_{PCA}$ and final codebook tuples $\langle \mathbf{P}^p_{PCA}, \mathbf{L}^p \rangle$. This training scheme results in the $S \times L$ regression codebooks $\mathcal{C}_{s,x}$.

Shape model to regularize the localization. To be able to regularize the intermediate solutions during the prediction phase, a model of the spatial distribution of the landmarks $\mathbf{s} = \langle \mathbf{x}^i_1, \ldots, \mathbf{x}^i_L \rangle$ in the training data is learned. We compute a point distribution model $\mathcal{S} = \langle \bar{\mathbf{s}}, \mathbf{S} \rangle$ using an eigen-decomposition of the covariance matrix of the training landmarks $\mathbf{x}_x$ as proposed in [2], retaining all eigenvectors and thus the entire shape variance observable in the training set, where the shapes $\mathbf{s}$ in the model can be constructed through a parameter vector $\mathbf{b}$ such that:

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{S}\mathbf{b}$$

2.2 Localization – Regularized Top-Down Matching

Similar to the training phase the localization is performed in a multi-scale fashion, shown in Fig. 3. The $D \times L$ landmark localization matrix $\mathbf{L}^*_{s=1}$ is initialized with all landmarks starting at the center of the test volume $\mathbf{I}_{target}$. Starting with scale $s = 1$, a patch $\mathbf{P}^x$ for each landmark x is extracted (without additional offsets or scaling variations). The patch is normalized and projected onto the patch PCA model of $\mathcal{C}_{s,x}$, resulting in $\mathbf{P}^x_{PCA}$. The most similar patch p^{x*} in the codebook is found using euclidean nearest neighbor search – leading to the tuple $\langle \mathbf{P}^{x*}_{PCA}, \mathbf{L}^{x*}_p \rangle$ and thus the landmark coordinate predictions $\mathbf{L}^{x*}_p$ as estimated by landmark x. Repeating this codebook lookup for all landmarks yields the $D \times L \times L$ prediction tensor $\mathbf{M_{d,i,j}}$ with position estimates from each landmark i to all landmarks that are visible in the same patch. The median over all predictions j which are not marked as not-visible yields the updated landmark localization matrix $\mathbf{L}^*_s$. This procedure is repeated through all scales, resulting in the final localization result $\mathbf{L}^*_S$.

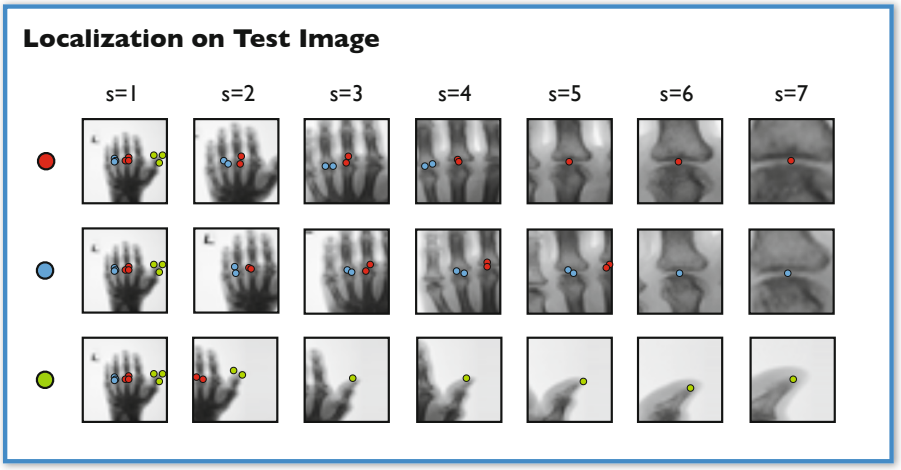


Fig. 3. The localization of three landmarks on a test image/volume descends the scale pyramid. At each level regression based on the image patch generates not only a position estimate for the primary landmak, but also for other landmarks visible in the patch. When progressing to a finer scale, for each landmark these estimates vote for the next estimate and center of the finer patch.

Shape regularization. The position estimates $\mathbf{L}_s^*$ are regularized by projecting them onto the shape PCA model $\mathcal{S}$ and reconstructing them again thereafter. This enforces landmark positions which can be modeled by a linear combination of the shapes observed in the training data. This regularization is performed for scales $s \leq S - 3$, to allow for landmark positions which can not be modeled though the shape model at scales $s > S - 3$.

3 Experiments

3.1 Data Sets

We evaluated the proposed approach on the two separate data sets shown in Fig. 1: 20 hand radiographs and 12 high resolution hand CTs.

Data set 1: Hand Radiographs $N = 20$ hand radiographs with an average size of 460×260 pixels with a resolution of 0.423mm/pixel were annotated with $L = 24$ landmarks. The landmarks include the five finger tips, as well as the distal interphalangeal (DIP), proximal interphalangeal (PIP), metacarpophalangeal (MCP) and carpometacarpal (CMC) joints for each finger.

Data set 2: Hand CTs The 3D hand CTs have a voxel size of $0.5mm \times 0.5mm \times 0.66mm$ resulting in an average size of $256 \times 384 \times 330$ voxels. They are annotated with the same 24 landmarks as the hand radiographs, with three additional landmarks placed around the carpus at the radiocarpal, radioulnar, and ulnocarpal joints, totaling in $L = 27$.

Table 1. Experimental results, localization accuracy in mm: Residual distances of the localization result to the ground truth annotation for the proposed method, in comparison with a state of the art approach

Residual in mm	MRF-based graph-matching			Proposed Patch-Regression Method		
	Median	Mean	Std	Median	Mean	Std
Hand Radiographs	0.80	0.99	0.82	0.63	0.77	0.64
Hand CTs	1.19	1.45	1.13	1.43	1.96	1.80

3.2 Setup

The experiments were run using four-fold cross validation, learning the landmark regression codebook on 75% of the N images / volumes and performing the localization on the remaining images / volumes. The main measure of interest for each landmark is the residual distance between the position of the predicted landmark position and the corresponding ground truth. The parameter settings are identical for the experiments on the two data sets, except for the size of the patches: 32×32 in the 2D case and $32 \times 32 \times 32$ for the 3D data. The results are compared with the recently proposed pre-filtered Hough regression Random forests [9], which in turn showed to outperform alternative approaches such as classification-based landmark candidate estimation with graph-based optimization [1] and classification + mean-shift based approaches [14].

3.3 Results

The results of the evaluation of the landmark localization are presented in Tab. 1, which shows the aggregated localization performance for the two data sets. The accuracy on the 2D radiograph data set is very high with a median residual of 0.63 mm and a mean/std of 0.77/0.64 mm. This result compares favorably with the results reported and methods tested on the same data in [9]. The result on the 3D hand CT data set show a median residual of 1.43 mm and a mean/std of 1.96/1.80 mm. It can be seen that despite a similar median residual, the proportion of localizations with higher error is slighty larger in this case. The run-times of the proposed approach were in the order of 0.6sec for the 2D data set and 4.5sec for the 3D data set on a single core of a 2009 Xeon MacPro. The method was entirely implemented in Matlab - we expect a potential speed-up by a factor of 10 to 100 through a more optimized implementation.

3.4 Discussion - Feature Computation Complexity

The main contribution of this work is the demonstration of a feature computation scheme which requires significantly less memory accesses then existing methods.

Voxel-wise classification / prediction approaches such as those proposed in [1,11] scale with the number of voxels, while pre-filtered Hough regression [9] reduces

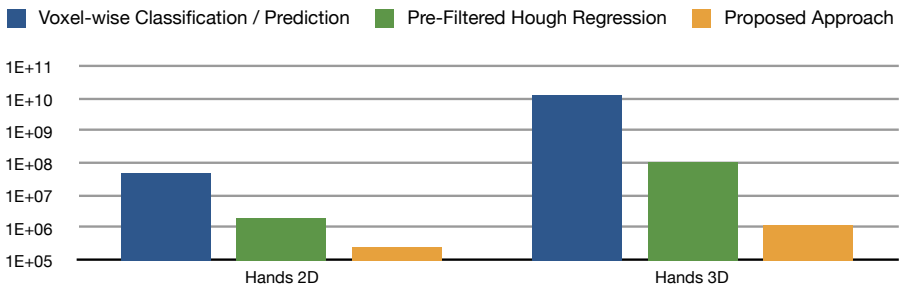


Fig. 4. Number of image/volume accesses necessary to compute the features required during the localization phase. Voxel-wise classification / prediction approaches [1,11] scale with the number of voxels, while pre-filtered Hough regression [9] works on strongly downsampled volumes. In constrast to this, the proposed approach is indepedent of the number of voxels and scales with the number of landmarks.

computational complexity by working on strongly down-sampled volumes. A typical number of 400 memory accesses to compute the classification for a single voxel was assumed in the calculation, corresponding to e. g. 20 individual features in an ensemble of 20 individual classifiers.

In contrast to this, the proposed approach is independent of the number of voxels and only depends on the number of landmarks, with $m \times m \times m$ voxels sampled for the patch at each landmark and scale. The proposed approach thus requires one to four orders of magnitude less image/volume accesses, allowing for fast localization even in unoptimized implementations or cheap commodity hardware.

4 Conclusion and Outlook

We present an approach for localizing complex, partly repetitive anatomical structures in 2D and 3D data. We demonstrate that a top-down nearest neighbor matching strategy of image patches drastically reduces the number of required feature computations and that the prediction of relative landmark positions using codebook regression is feasible.

The results on the two data sets clearly demonstrate the ability of the proposed approach to find the landmark positions in the target volume with accuracy comparable to the state of the art, with the consistent localization of detailed anatomical structures with a median residual of 1.7 to 2.7 pixels/voxels.

We consider the results to be very promising for such a simple method, and will focus on several topics in upcoming work: A detailed analysis of the parameters involved, namely the patch size and the perturbation strategy during codebook generation, as well as approximations of the nearest neighbor search through random subspaces.

References

1. Bergtholdt, M., Kappes, J., Schmidt, S., Schnörr, C.: A Study of Parts-Based Object Class Detection Using Complete Graphs. IJCV 87(1-2), 93–117 (2010)
2. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graha, J.: Active Shape Models - Their Training and Application. CVIU 61(1), 38–59 (1995)
3. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. TPAMI 23(6), 681–685 (2001)
4. Cremers, D., Rousson, M., Deriche, R.: A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. IJCV 72(2), 195–215 (2007)
5. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression forests for efficient anatomy detection and localization in ct studies. In: Medical Computer Vision 2010: Recognition Techniques and Applications in Medical Imaging, MICCAI Workshop (2010)
6. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. In: Proc. of MICCAI Workshop on Probabilistic Models for Medical Image Analysis, MICCAI-PMMIA (2009)
7. Doi, K.: Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized Medical Imaging and Graphics 31, 198–211 (2007)
8. Donner, R., Birngruber, E., Steiner, H., Bischof, H., Langs, G.: Localization of 3D Anatomical Structures Using Random Forests and Discrete Optimization. In: Proc. MICCAI Workshop on Medical Computer Vision (2010)
9. Donner, R., Menze, B.H., Bischof, H., Langs, G.: Global Localization of 3D Anatomical Structures by Pre-filtered Hough Forests and Discrete Optimization. Medical Image Analysis (accepted, 2013)
10. Kelm, B.M., Zhou, S.K., Suehling, M., Zheng, Y., Wels, M., Comaniciu, D.: Detection of 3D Spinal Geometry Using Iterated Marginal Space Learning. In: Proc. MICCAI Workshop on Medical Computer Vision (2010)
11. Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled Decision Forests and Their Application for Semantic Segmentation of CT Images. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 184–196. Springer, Heidelberg (2011)
12. Pauly, O., Glocker, B., Criminisi, A., Mateus, D., Möller, A.M., Nekolla, S., Navab, N.: Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 239–247. Springer, Heidelberg (2011)
13. Seifert, S., Barbu, A., Zhou, S., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical Parsing and Semantic Navigation of Full Body CT Data. In: SPIE Medical Imaging (2009)
14. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moorea, R., Kipman, A., Blake, A.: Real-Time Human Pose Recognition in Parts from a Single Depth Image. In: Proc. CVPR (2011)
15. Zheng, Y., Georgescu, B., Comaniciu, D.: Marginal Space Learning for Efficient Detection of 2D/3D Anatomical Structures in Medical Images. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 411–422. Springer, Heidelberg (2009)

Oblique Random Forests for 3-D Vessel Detection Using Steerable Filters and Orthogonal Subspace Filtering*

Matthias Schneider[1], Sven Hirsch[1], Gábor Székely[1], Bruno Weber[2], and Bjoern H. Menze[1]

[1] Computer Vision Laboratory, ETH Zurich, Switzerland
[2] Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

Abstract. We propose a machine learning-based framework using oblique random forests for 3-D vessel segmentation. Two different kinds of features are compared. One is based on orthogonal subspace filtering where we learn 3-D eigenspace filters from local image patches that return task optimal feature responses. The other uses a specific set of steerable filters that show, qualitatively, similarities to the learned eigenspace filters, but also allow for explicit parametrization of scale and orientation that we formally generalize to the 3-D spatial context. In this way, steerable filters allow to efficiently compute oriented features along arbitrary directions in 3-D. The segmentation performance is evaluated on four 3-D imaging datasets of the murine visual cortex at a spatial resolution of $0.7\,\mu\mathrm{m}$. Our experiments show that the learning-based approach is able to significantly improve the segmentation compared to conventional Hessian-based methods. Features computed based on steerable filters prove to be superior to eigenfilter-based features for the considered datasets. We further demonstrate that random forests using oblique split directions outperform decision tree ensembles with univariate orthogonal splits.

Keywords: vessel segmentation, orthogonal subspace filtering, steerable filters, oblique random forest.

1 Introduction

Blood vessel enhancement and segmentation play a crucial role for numerous medically oriented applications and has attracted a lot of attention in the field of medical image processing. The multiscale nature of vessels, image noise and contrast inhomogeneities make it a challenging task. In this context, a large variety of methods have been developed exploiting photometric and structural properties of tubular structures. Extensive reviews on various state-of-the-art vessel segmentation techniques can be found in the literature [14,15]. Rather simple methods, e.g., absolute or locally adaptive thresholding, are in fact regularly used in practice due to their conceptual simplicity and computational efficiency but they are a serious source of error and require careful parameter selection [20,22]. More sophisticated segmentation techniques such as optimal filtering and Hessian-based approaches commonly rely on idealized appearance

* Supplementary material for this article is available at `http://www.vision.ee.ethz.ch/ReCoVa`

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 142–154, 2013.

and noise models. The former includes optimal edge detection [2], and steerable filters providing an elegant theory for computationally efficient ridge detection at arbitrary orientations [12,9]. The latter is based on the eigenanalysis of the Hessian capturing the second order structure of local intensity variations [4,24]. The Hessian is commonly computed by convolving the image patch with the partial second order derivatives of a Gaussian kernel as the method of choice for noise reduction and to tune the filter response to a specific vessel scale. This basic principle has already been used by Canny for edge and line detection [2]. The differential operators involved in the computation of the Hessian are well-posed concepts of linear scale-space theory [16]. Modeling vessels as elongated elliptical structures, the eigendecomposition of the Hessian has a geometric interpretation, which can be used to define a "vesselness" measure as a function of the eigenvalues [4,24]. Due to the multi-scale nature of vascular structures, Hessian-based filters are commonly applied at different scales. Besides, the eigenvector corresponding to the largest eigenvalue of the Hessian computed at the most discriminative scale is a good estimate for the local vessel direction. In practice, vesselness filters tend to be prone to noise and have difficulty in detecting vessel parts such as bifurcations not complying with the intrinsic idealized appearance model. Vesselness filters have also been successfully applied for global vessel segmentation in X-ray angiography using ridge tracking [26] and graph cut theory [10].

In this paper, we devise a machine learning approach for vessel segmentation based on the 2-D filament detection framework proposed by Gonzalez et al. [9] using steerable filters [5,12]. In our application, we aim at efficient classification of 3-D high-resolution imaging datasets ($> 10^{10}$ voxels) of the murine visual cortex (see Figure 1), which is of great interest for the analysis of the cerebrovascular system [22,11]. Due to the considerable computational challenge that comes with our application, we focus on a fast classification approach using local linear filters rather than complex non-local spatial models incorporating prior knowledge and regularization [26,10]. We compare different features computed from, respectively, orthogonal subspace filtering [17,23] and steerable filters using Gaussian derivatives [5,8]. In contrast to the framework proposed by Gonzalez et al. [9,8], we use oblique random forests (RF) for efficient classification We test "elastic net" node models that combine ℓ_1 and ℓ_2 regularization leading to sparser node models than the ℓ_2 regularized oblique splits proposed in [18].

2 Methods

In this section, we first introduce two different sets of features based on (1) orthogonal subspace filtering and (2) steerable filters computed at different scales and orientations in order to achieve rotational invariance. These features are then used to train an oblique random forest (RF) classifier that is well adapted to correlated feature responses from local image filters [18]. Different from standard discriminative learning algorithms such as support vector machines, RF classifiers return continuous probabilities when predicting vessel locations, which allows to choose an operating point by adapting the decision threshold. Moreover, RF is capable of coping with high dimensional feature vectors and tolerate false training labels. It is fast to train with only very few parameters to be optimized and even faster to apply. Efficient prediction becomes particularly important in view of our specific application using high-resolution image data at μm resolution.

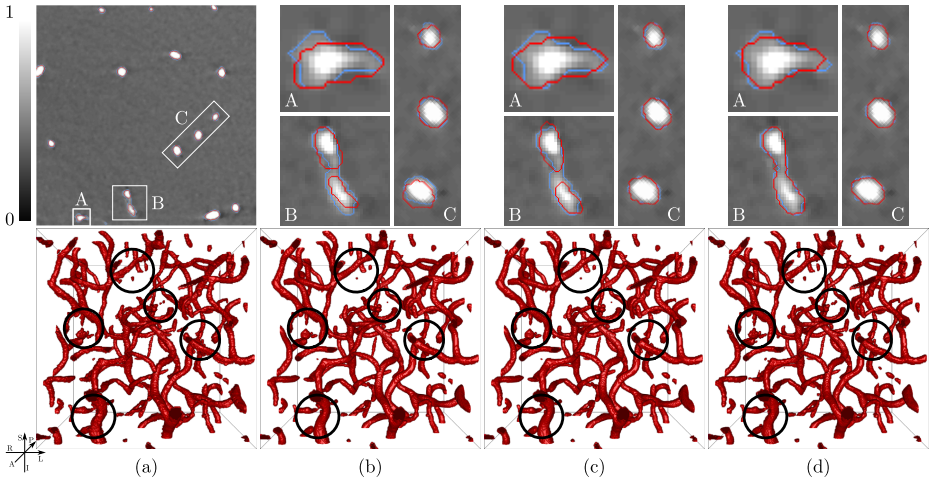


Fig. 1. Visualization of segmented cerebrovascular network for single axial slice (top) and whole 3-D test ROI (bottom) using different segmentation techniques. (a) Ground truth. (b) Frangi [4]. (c) RF-OSF ($d = 102$). (d) RF-SFT ($M = 4$). The binary segmentation maps are computed at the corresponding F_1-optimal operating points marked in Figure 4(b). The results are rendered in 3-D (bottom) and outlined in red (top) along with the ground-truth contours in blue for three subregions within the axial slice (A-C). Red contours in (a) mark the Otsu labels [20] used for RF training. Black circles in the 3-D plots highlight prominent differences in the segmentation. More results for the other datasets are provided in the supplementary material.

2.1 Orthogonal Subspace Filters (OSF)

Matched filters (MF) have widely been used in signal processing. They allow to detect a signal of known shape (template) by cross-correlation and perform provably optimal under additive Gaussian white noise conditions [19]. In terms of image processing, this corresponds to the convolution of the image with the MF. From a learning and classification perspective, matched filtering (signal detection) is closely related to linear regression for binary classification between background and pattern (vessel) [17]. Considering the image as a composition of local image patches with each pixel in the patch representing a feature, MF defines a 1-D linear subspace (regression coefficients) of this feature space which allows for separation of the pattern from background. Instead of an optimal 1-D subspace assuming linear separability in the feature space as implied by using a single matched filter, we use a less restrictive dimensionality reduction similar to [17], namely (linear) principal component analysis (PCA), in order to define a subspace of higher dimensionality. More formally, let $\boldsymbol{p_i} \in \mathbb{R}^{P^3}$ denote a (cubic) image patch of size $P \times P \times P$. A d-dimensional subspace ($d \leq P^3$) capturing the most important modes of variation in the image patches can then be defined using PCA [13]:

$$\forall 1 \leq k \leq d \leq P^3 : \ \boldsymbol{\alpha_k} \ = \ \underset{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{P^3},\ \|\boldsymbol{\alpha}\|=1, \\ \forall 1 \leq i < k:\ \mathrm{cov}(\boldsymbol{\alpha_i}, \boldsymbol{\alpha})=0}}{\arg\max} \ \mathrm{var}(\boldsymbol{\alpha}^{\mathrm{T}} P_{\mathrm{OSF}}) \quad , \tag{1}$$

where $P_{\mathrm{OSF}} = [\boldsymbol{p_i}]_{1 \leq i \leq N_{\mathrm{P}}} \in \mathbb{R}^{P^3 \times N_{\mathrm{P}}}$ is the data matrix assembling N_{P} patches labeled as vessel. The principal axes $\boldsymbol{\alpha_k}$ form an orthonormal basis of the d-dimensional subspace and are ordered according to their preserved variance. They can be computed efficiently as the d eigenvectors corresponding to the largest eigenvalues of the covariance matrix of P_{OSF} after mean centering using singular value decomposition. Projecting an arbitrary image patch $\boldsymbol{p} \in \mathbb{R}^{P^3}$ onto the PCA subspace yields its d principal components (PC). The PCs of the image patches centered at pixels $\boldsymbol{x}$ in image I can thus be computed by d independent convolution operations of the image with each (properly reshaped) principal axis $\tilde{\boldsymbol{\alpha}}_{\boldsymbol{k}} \in \mathbb{R}^{P \times P \times P}$:

$$\boldsymbol{f_{\mathrm{OSF}}}(I, \boldsymbol{x}) = \left[(\tilde{\boldsymbol{\alpha}}_{\boldsymbol{k}} * I)(\boldsymbol{x}) - \boldsymbol{\alpha}_{\boldsymbol{k}}^{\mathrm{T}} \frac{1}{N_{\mathrm{P}}} \sum_{i=1}^{N_{\mathrm{P}}} \boldsymbol{p_i} \right]_{1 \leq k \leq d} \in \mathbb{R}^d \quad . \tag{2}$$

The (reshaped) principal axes will also be referred to as orthogonal subspace filters (OSF). The PCs, i.e., the OSF response of an image patch, are used as features along with a nonlinear decision rule for vessel segmentation as described in Section 2.3.

2.2 Steerable Filter Templates (SFT)

The OSF eigenfilters learned from image patches as described in the previous section turn out to be highly structured (see Figure 2(a)). Instead of learning the structured filter kernels, we hence attempt to explicitly parametrize them. For this, we choose a steerable filter model based on Gaussian derivatives, which allows for efficient directional filtering at different scales and, most importantly, implicates rotational invariance [12]. Similar to [8], we define the filter templates as normalized derivatives of Gaussians up to order M [16]:

$$\forall\, m \geq 1 \ \wedge \ 0 \leq b \leq a \leq m \leq M : \ G^{\sigma}_{m,a,b}(\boldsymbol{x}) = \sigma^m \frac{\partial^{m-a} \partial^{a-b} \partial^{b}}{\partial_x^{m-a} \partial_y^{a-b} \partial_z^{b}} G^{\sigma}(\boldsymbol{x}) \quad , \tag{3}$$

where $G^{\sigma}(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^3} \exp(-\frac{\|\boldsymbol{x}\|}{2\sigma^2})$ denotes the 3-D symmetric Gaussian kernel with variance σ and zero mean. As in Equation (2), each template induces a single feature by convolution with image I. They can be assembled to a feature vector of dimension $d_{\mathrm{M}} = 1/6(M^3 + 6M^2 + 11M)$ at a fixed scale σ:

$$\boldsymbol{f^{\sigma}}(I, \boldsymbol{x}) = \left((G^{\sigma}_{1,0,0}, G^{\sigma}_{1,1,0}, G^{\sigma}_{1,1,1}, \ldots, G^{\sigma}_{M,M,M})^{\mathrm{T}} * I \right)(\boldsymbol{x}) \in \mathbb{R}^{d_{\mathrm{M}}} \quad . \tag{4}$$

We enhance the features by concatenating feature vectors at different scales $\sigma_1, \ldots, \sigma_{\mathrm{S}}$:

$$\boldsymbol{f}_{\mathrm{SFT}}(I, \boldsymbol{x}) = \left(\boldsymbol{f^{\sigma_1}}(I, \boldsymbol{x}), \ldots, \boldsymbol{f^{\sigma_{\mathrm{S}}}}(I, \boldsymbol{x}) \right)^{\mathrm{T}} \in \mathbb{R}^{d_{\mathrm{M}} S} \quad . \tag{5}$$

The steerability of Gaussian derivatives has been derived for the 2-D case in [12] and can readily be extended to 3-D [5,8]. Steerability refers to the property that the convolution of an image with a rotated version of the steerable filter template (SFT) can

be expressed by a linear combination of the filter response of the image with the SFT without rotation:

$$I * G^{\sigma}_{m,a,b}(R\boldsymbol{x}) = \sum_{i=0}^{m} \sum_{j=0}^{i} \omega^{i,j}_{m,a,b} \underbrace{\left(I * G^{\sigma}_{m,i,j}\right)(\boldsymbol{x})}_{\boldsymbol{f}^{\sigma}_{\boldsymbol{m},\boldsymbol{i},\boldsymbol{j}}(I,\boldsymbol{x})} \quad , \tag{6}$$

where $R \in SO(3)$ denotes a 3-D rotation matrix and $\omega^{i,j}_{m,a,b}$ the uniquely defined coefficients that can be computed in closed form [12].[1] This formalism allows to efficiently evaluate the feature vector $\boldsymbol{f}_{\mathbf{SFT}}$ for an arbitrary rotation without any additional costly convolution. We use a restricted set of rotations in our application considering the tubular structure of vessels. The local vessel direction $\boldsymbol{d} = (d_x, d_y, d_z)^{\mathrm{T}} \in \mathbb{R}^3$, $\|\boldsymbol{d}\| = 1$ can be parametrized using spherical coordinates (θ, ϕ) with unit radius, elevation $\theta = \arctan\left(d_z / \sqrt{d_x^2 + d_y^2}\right)$, and azimuth $\phi = \arctan(d_y/d_x)$ relative to the x-y plane ($z = 0$). It is sufficient to restrict the parametrization to the positive hemisphere ($z > 0$), i.e., $0 \leq \theta \leq \pi/2$ and $-\pi < \phi \leq \pi$. The vessel can then be transformed to the normalized pose $\boldsymbol{d_0} = (1, 0, 0)^{\mathrm{T}}$ by applying the rotation matrix

$$R_{\theta,\phi} = \begin{pmatrix} \cos\theta\cos\phi & \cos\theta\sin\phi & \sin\theta \\ -\sin\phi & \cos\phi & 0 \\ -\sin\theta\cos\phi & -\sin\theta\sin\phi & \cos\theta \end{pmatrix} \quad . \tag{7}$$

The SFT features evaluated for this rotation according to Equation (6) hence describe the intensity variation characteristics of different order along the vascular structure as well as in the orthogonal plane. Assuming a symmetric vessel (intensity) profile perpendicular to the local vessel direction $\boldsymbol{d}$, restricting the set of rotations is reasonable as the vessel appearance is (locally) invariant under rotation about $\boldsymbol{d}$.

2.3 Vessel Classification - Shape Learning and Prediction

The OSF and SFT features as defined in Equations (2) and (5), respectively, are each used along with a non-linear decision rule for vessel segmentation. We train separate classifiers for the different feature types as follows: A representative set $\mathcal{S}$ of $2N_{\mathrm{S}}$ tuples (image I_k, location $\boldsymbol{x_k}$, vessel orientation $\boldsymbol{d_k}$, class label y_k) is randomly sampled from a labeled set of images corresponding to N_{S} foreground ($y_k = 1$) and background ($y_k = -1$) samples, respectively: $\mathcal{S} = \{(I_k, \boldsymbol{x_k}, \boldsymbol{d_k}, y_k) \mid 1 \leq k \leq 2N_{\mathrm{S}}\}$. For these samples, the features $\boldsymbol{f}(I, \boldsymbol{x})$ can be extracted as defined in Equations (2) and (5). The SFT features are additionally rotated to the normalized orientation according to Equations (6) and (7) w.r.t. the local vessel direction $\boldsymbol{d}$. This defines the training set $\mathcal{T} = \left\{\left(\boldsymbol{f_k} = \boldsymbol{f}(I_k, \boldsymbol{x_k}), y_k\right) \mid 1 \leq k \leq 2N_{\mathrm{S}}\right\}$ that is ultimately used to train a random forest (RF) classifier [1]. RF consists of an ensemble of decision trees used to model the posterior probability of each class (vessel/background). During training, each tree is fully grown from bootstrapped datasets using stochastic discrimination. For this, the data is split at each tree node by a hyperplane in the feature (sub-)space. In contrast to traditional bagging, the split is based on a small number of randomly selected

[1] Further details are provided in the supplementary material.

features only. We investigated both "orthogonal" and "oblique" trees. As proposed in Breiman's original paper [1], the former is based on optimal thresholds for randomly selected single features in every split, i.e., mutually orthogonal 1-D hyperplanes. The latter uses multidimensional hyperplanes to separate the feature space, e.g., by choosing randomly oriented hyperplanes [1] or applying linear discriminative models [18]. For the oblique RFs in this work, we employ a linear regression with an elastic net penalty [6] in order to learn multivariate (optimal) split directions $\boldsymbol{w}$ at each node:

$$\hat{\boldsymbol{w}} = \underset{\boldsymbol{w}\in\mathbb{R}^{N_\mathrm{F}}}{\arg\min} \frac{1}{2|\mathcal{T}|} \sum_{k=1}^{|\mathcal{T}|} \left(y_k - \boldsymbol{w}^\mathrm{T} \tilde{\boldsymbol{f}}_{\boldsymbol{k}} \right)^2 + \lambda P_\alpha(\boldsymbol{w}) \quad , \tag{8}$$

where $\tilde{\boldsymbol{f}}_{\boldsymbol{k}} \in \mathbb{R}^{N_\mathrm{F}}$ are randomly selected (but fixed) features and $\lambda > 0$ is the regularization parameter for the elastic net penalty $P_\alpha(\boldsymbol{w}) = (1-\alpha)\frac{1}{2}\|\boldsymbol{w}\|_{\ell_2}^2 + \alpha\|\boldsymbol{w}\|_{\ell_1}$ as a compromise between the ridge regression ($\alpha = 0$) and the lasso penalty ($\alpha = 1$), where $\|\cdot\|_{\ell_1}$ and $\|\cdot\|_{\ell_2}$ denote the ℓ_1 and ℓ_2-norm, respectively. The advantage is joint regularization of the coefficients and sparsity — coefficients are both encouraged to be small, and to be zero if they are very small. The latter lasso property reduces the dimensionality of the split space, which is desirable for memory and robustness purposes. With $\alpha = 1$ (and $\lambda \gg 0$) we will get a single non-zero coefficient, i.e., RF with univariate splits, whereas choosing $\alpha = 0$ we have ridge regression as in [18].

The decision trees are grown separately as follows:

1. For each tree, a new set of samples is randomly drawn from the training data $\mathcal{T}$ with replacement, i.e., $\frac{2}{3}|\mathcal{T}|$ bootstrapped samples.
2. For every node, N_F features are randomly sampled without replacement from the feature pool of size $N_\mathrm{F}^0 = d$ for OSF features and $N_\mathrm{F}^0 = d_\mathrm{M} S$ for SFT features, respectively (see Equations (2) and (4)).
3. The selected features of the bootstrapped samples are normalized to zero mean and unit variance at every split in order to enhance the stability of the linear model.
4. Finding optimal split
   a) Orthogonal split ($N_\mathrm{F} = 1$): The feature values of all samples are tested as threshold to split the data w.r.t. the selected feature.
   b) Oblique split ($N_\mathrm{F} = \left\lceil \sqrt{N_\mathrm{F}^0} \right\rceil$): The optimal split direction is computed according to Equation (8) for $\alpha = 0.5$ using covariance updates [6].
5. Steps 2 – 4 are repeated $\left\lceil \sqrt{N_\mathrm{F}^0} \right\rceil$ times. The optimal split and threshold are ultimately selected w.r.t. the information gain as a result of the split. The samples are split accordingly and passed on to the child nodes.
6. For each of the N_T trees, steps 2 – 5 are repeated until (1) all samples in a (leaf) node belong to the same class, (2) the maximum tree depth has been reached, or (3) there are too few samples to further split the data (avoid excessive overfitting).
7. Each leaf node is assigned a class label according to the majority vote of the training samples ending up in the considered leaf.

Previously unseen samples (images) can be classified by pushing the extracted features down all N_T decision trees of the ensemble. Thus, each tree assigns a class label

$\hat{y}_i \in \{0, 1\}$ associated with the leaf node in which the tested sample ends up. The ensemble confidence can then be defined as $\frac{1}{N_\mathrm{T}} \sum_{i=1}^{N_\mathrm{T}} \hat{y}_i$ as an estimate of the posterior. The binary class label $\hat{y}$ can finally be assigned using a majority vote or any other decision threshold.

In the case of OSF features, a single RF is trained for all vessel orientations. Therefore, the intrinsic orientation-induced structure in the OSF feature space has to be captured in the training set both for RF training and learning the OSF eigenfilters. In contrast, SFT features allow for explicit parametrization of the orientation. The expected filter response for an arbitrary orientation can efficiently be computed from the set of stationary base features $\boldsymbol{f}_\mathrm{SFT}$ as defined in Equations (5) and (6). As the corresponding RF classifiers are trained on SFT features extracted from vessels with normalized orientation only, we sample the space of possible vessel orientations (half sphere) and compute the corresponding (rotated) SFT features in order to build an orientation independent predictor. The classification result with the maximum confidence is ultimately assigned as proposed in [9]. In contrast to OSF features, this allows to not only estimate the class posteriors but also a probability distribution on the vessel orientation.

3 Experiments

We have evaluated the performance of our method on four 3-D datasets $\mathcal{D}_{1-4}$ obtained from synchrotron radiation X-ray tomographic microscopy (srXTM) of cylindrical samples of the murine somatosensory cortex (volume size $2048\,\mathrm{px} \times 2048\,\mathrm{px} \times 4000\,\mathrm{px}$, isotropic voxel spacing $0.7\,\mu\mathrm{m}$, grayscale $16\,\mathrm{bit}$) [22]. In a preprocessing step we applied anisotropic diffusion filtering in order to reduce image noise while preserving edge contrast [21]. From each (preprocessed) dataset we extracted two disjoint regions of interest (ROI) of size $(256\,\mathrm{px})^3$ for training and testing, respectively. In the following, we will refer to these non-overlapping ROIs as test and train data/ROI, respectively (see Figure 1(a)). For each test ROI, ground truth labels were manually generated by an expert assisted by a semi-automatic segmentation tool [27] on 15 evenly distributed slices along each reference direction (axial, coronal, sagittal). Thus, 125 slices have been labeled containing 7.3×10^4 foreground and 2.7×10^6 background labels in average ($\pm 3.9 \times 10^4$) corresponding to a vascular volume fraction of $2.6 \pm 1.4\,\%$.

In a first baseline experiment, all ROIs were segmented using both Otsu's method [20] and multiscale vessel enhancement filtering [4,24]. For the latter, we have performed an exhaustive grid search to optimize the vesselness scale on the test ROIs w.r.t. maximum area under the ROC curve using the ground-truth labels of the test ROIs. In the majority of the cases five logarithmically spaced scales performed best for both Frangi's and Sato's vesselness: $\sigma \in \{2.00, 3.09, 4.76, 7.35, 11.33\}[\mathrm{px}]$.

In a next step, we computed the OSF eigenfilters introduced in Section 2.1 from 3000 randomly sampled patches centered at voxels labeled as vessel in the Otsu label map. In particular, background patches were not considered during OSF learning. Besides the original vessel patches, five randomly rotated versions of each patch have been added to the set of patches P_OSF used in Equation (1) in order to account for rotational symmetry of vessel structures while keeping the total number of patches at a moderate level ($N_\mathrm{P} = 1.8 \times 10^4$). As in [17], the OSF patch size P was assessed from the random forest feature importance and set to $P = 19$.

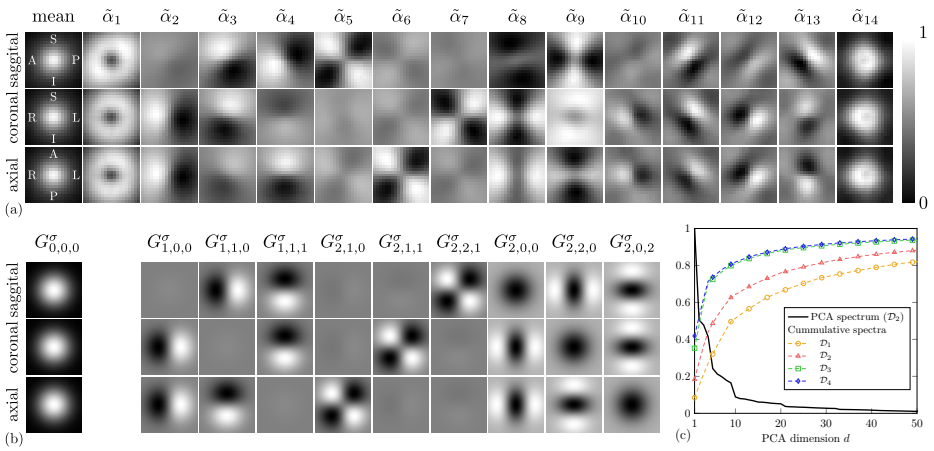


Fig. 2. (a) Visualization of the mean pattern and the most significant (reshaped) eigenfilters $\tilde{\boldsymbol{\alpha}}_{\boldsymbol{k}}$ along centered sagittal, coronal, and axial slices as learned from dataset $\mathcal{D}_2$ ($P = 19$). (b) Normalized Gaussian derivatives $G^{\sigma}_{m,a,b}$ at a fixed scale σ up to order $M = 2$ as defined in Equation (3). (c) Normalized PCA spectrum λ_k/λ_1 and variance preservation as measured by the cumulative spectrum $\sum_{k=1}^{d} \lambda_k / \sum_{k=1}^{P^3} \lambda_k$ for different datasets, where λ_k denotes the k-th eigenvalue of the data covariance matrix.

As for the SFT model, we performed a small parameter study to optimize the SFT scales similar to the multiscale vesselness parameters. In order to avoid overfitting, however, we used the train ROIs for the parameter optimization along with the Otsu labels considered as ground truth in this case. We ultimately select $S = 3$ logarithmically spaced scales $\sigma \in \{2.00, 3.65, 6.67\}$. The SFT model hence defines $d_{\mathrm{M}} S = 9$ (27, 57, 102) features for maximum Gaussian derivative order $M = 1$ (2, 3, 4), respectively (see Equations (4) and (5)). For a fair comparison of the SFT and OSF feature models, the PCA subspace dimension d of the OSF models, i.e., the number of OSF features, was chosen accordingly.

Different RF classifiers consisting of $N_{\mathrm{T}} = 256$ decision trees have been trained separately on the train ROI of a single dataset using OSF and SFT features along with orthogonal and oblique splits, respectively, as explained in Section 2.3. The training was repeated for each dataset using $N_{\mathrm{S}} = 4000$ foreground (vessel) and background samples, respectively, randomly drawn from the Otsu label map. The local vessel direction was estimated from the eigenanalysis of the Hessian computed at the most discriminative scale as defined by Frangi's multiscale vesselness [4]. Note that the training labels were computed fully automatically without any user interaction. The manually annotated ground-truth labels have been used for RF validation only.

Finally, the different RF models were applied to the test ROIs of each dataset. The classification performance was evaluated on the uniformly aligned slices with ground-truth labels available (see above). In this way, the generalization of the individual

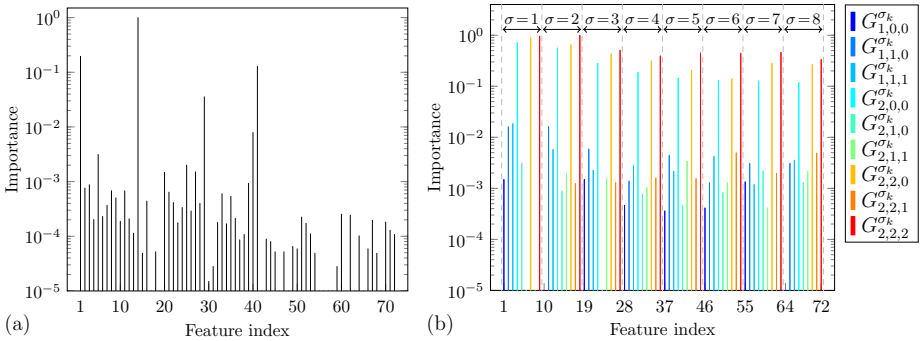


Fig. 3. Variable importance [1] of the (a) RF-OSF model ($P = 19$, $d = 57$) and (b) RF-SFT model ($M = 2$, $d_{\mathrm{M}} = 9$, $S = 8$ scales $\sigma \in \{1, \ldots, 8\}$) on a logarithmic scale (oblique splits). The prominent peaks in (b) correspond to the Gaussian derivatives $G^{\sigma}_{2,0,0}$, $G^{\sigma}_{2,2,0}$, and $G^{\sigma}_{2,2,2}$.

classifiers is investigated (test ROIs of datasets not used for training) as well as the prediction quality for unseen samples from the dataset used for training (train ROI) but from a different subvolume (test ROI).

4 Results and Discussion

The learned OSF filter templates are highly structured (see Figure 2(a)). The ball-shaped mean shows a Gaussian-like pattern. The most significant principal axis captures the average image intensity in the vicinity of the sample. Patches $\boldsymbol{\alpha_2}, \ldots, \boldsymbol{\alpha_4}$ capture first order derivatives along the right-left (R-L), superior-inferior (S-I), and anterior-posterior (A-P) direction, respectively. Similar first-order patterns at a smaller scale appear in $\boldsymbol{\alpha_{10}}, \ldots, \boldsymbol{\alpha_{13}}$. Differently oriented second order derivatives are described by $\boldsymbol{\alpha_5}, \ldots, \boldsymbol{\alpha_9}$. The corresponding PCA spectra show a sharp profile as indicated in Figure 2(c). These observations can be made for all OSF models regardless of the considered patch size. For comparison of the structural similarities, the parameterized Gaussian derivatives up to order $M = 2$ as used for the SFT feature extraction are shown in Figure 2(b).

The normalized RF feature relevance score, i.e., the permutation importance from [1], for the RF-OSF and RF-SFT model using oblique splits are shown in Figure 3. The OSF patches describing the average image intensity in the local neighborhood ($\boldsymbol{\alpha_1}$, $\boldsymbol{\alpha_{14}}$) show high variable importance as compared to the patches capturing higher order derivatives $\boldsymbol{\alpha_2}, \ldots, \boldsymbol{\alpha_{13}}$. It also becomes clear that the OSF feature importance (discrimination capability) is not correlated to the PCA spectrum (variance preservation). This makes it difficult to choose a proper cutoff for the PCA subspace dimension. The variable importance of the SFT features indicates that the second order derivatives parallel and orthogonal to the vessel direction ($G^{\sigma}_{2,0,0}$, $G^{\sigma}_{2,2,0}$, $G^{\sigma}_{2,2,2}$) are most significant for the classification. Note that the Hessian-based segmentation approaches also rely on these features [4,24]. For larger scales σ, the importance values tend to decline.

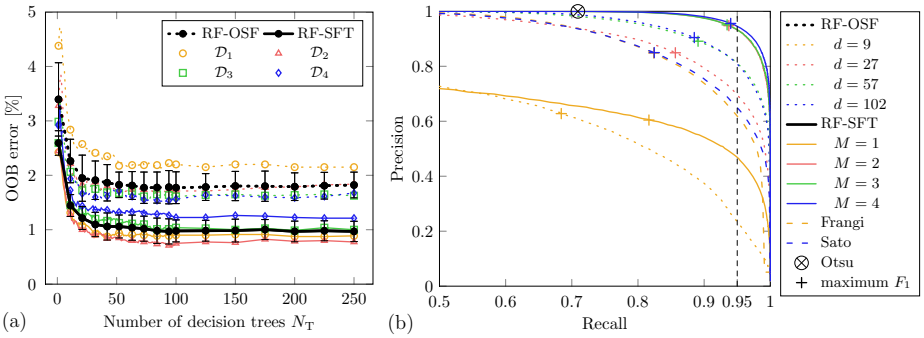


Fig. 4. Comparison of the classification performance. (a) Out of bag (OOB) error of the RF-OSF ($d = 102$) and RF-SFT ($M = 4$, $d_M = 34$) classifiers trained on dataset $\mathcal{D}_k$ for varying number of trees N_T (oblique splits). The average error is plotted in black with error bars indicating the standard deviation. (b) Precision-recall curves (PRC) and optimal operating points w.r.t. F_1 measure for RF-OSF and RF-SFT models ($N_T = 256$, oblique splits, trained on $\mathcal{D}_2$) with varying parameters d and M, respectively, in comparison to (optimized) Frangi's/Sato's vesselness, and Otsu thresholding [20] evaluated on test ROI of $\mathcal{D}_2$.

Figure 4(a) visualizes the out of bag (OOB) error of the RF-OSF and RF-SFT classifiers for different number of decision trees N_T. In both cases the OOB error declines rapidly for increasing N_T. The SFT model consistently shows smaller error rates compared to the OSF features. Moreover, the RF-SFT classifier is more robust across different datasets as indicated by the smaller standard deviation. Also note that the absolute values of the OOB error estimates may be somewhat overoptimistic due to the spatial correlation between the training samples.

Comparing the overall classification performance of the proposed learning-based approaches with different model parameters to standard segmentation approaches reveals the superior performance of the SFT features as indicated by the precision-recall curves (PRC) in Figure 4(b). The RF-based segmentation outperforms Frangi's/Sato's vesselness filters even for a small number of features ($d = 27$, $M = 2$). Note that the reported results for the vesselness-based segmentation have to be considered as upper bound as the scale parameters have been optimized on the test data (overfitting). The analysis also shows that for $M > 1$ the performance of the RF-SFT model hardly changes anymore, which is consistent with the observation of the second order derivatives being the most discriminative features (see Figure 3(d)).

A more detailed numerical analysis of the classification performance of the different approaches is summarized in Table 1 and confirms the superior performance of the RF-SFT model over the OSF features and the multiscale vesselness filters. Otsu's method [20] tends to underestimate the global threshold and hence results in an inaccurate segmentation of the vessel boundaries as indicated by the increased balanced error rate [3]. In order to assess the robustness of the learning-based segmentation approaches, we apply "intra-dataset" and "inter-dataset" cross-validation, i.e., choosing the (non-overlapping) train and test ROIs from the same (intra) or different (inter) datasets, respectively. The average segmentation performance for "totally" unseen data

Table 1. Detailed evaluation of classification performance of different RF-OSF ($d = 102$) and RF-SFT ($M = 4$, $d_{\mathrm{M}} = 34$) classifiers ($N_{\mathrm{T}} = 256$) using orthogonal and oblique splits, respectively. The performance is evaluated using "intra-dataset" and "inter-dataset" cross-validation (see text). The operating point was selected at the 95 % recall level (see Figure 4(b)). The partial area under the precision-recall curve (AUC-PR) has been computed on the recall interval $[0.5, 1]$.

	Method	Validation	Precision [%]	Specificity [%]	Error Rate [%]	AUC-PR [$\times 10^{-2}$]	OOB Error [%]	Tree Depth
orthogonal	RF-OSF	intra-data	74.32 ± 7.26	99.20 ± 0.13	2.92 ± 0.06	45.29 ± 1.74	2.02 ± 0.44	9.02 ± 0.45
		inter-data	70.99 ± 7.55	98.97 ± 0.62	3.06 ± 0.33	44.35 ± 2.09		
	RF-SFT	intra-data	89.43 ± 1.19	99.70 ± 0.14	2.70 ± 0.09	48.35 ± 0.20	1.35 ± 0.22	7.42 ± 0.38
		inter-data	88.25 ± 2.05	99.67 ± 0.14	2.70 ± 0.07	48.13 ± 0.41		
oblique	RF-OSF	intra-data	78.96 ± 5.81	99.37 ± 0.16	2.84 ± 0.07	46.38 ± 1.19	1.82 ± 0.25	6.26 ± 0.21
		inter-data	78.35 ± 4.39	99.33 ± 0.26	2.87 ± 0.15	45.95 ± 1.15		
	RF-SFT	intra-data	93.53 ± 1.47	99.83 ± 0.07	2.62 ± 0.05	48.96 ± 0.22	0.96 ± 0.19	5.85 ± 0.24
		inter-data	92.80 ± 1.94	99.82 ± 0.06	2.62 ± 0.03	48.84 ± 0.31		
	Sato	average	62.15 ± 2.71	98.46 ± 0.75	3.27 ± 0.37	42.34 ± 0.64	n/a	n/a
	Frangi	average	59.61 ± 2.09	98.26 ± 0.90	3.37 ± 0.45	41.66 ± 0.53	n/a	n/a
	Otsu	average	99.96 ± 0.03	100.00 ± 0.00	14.59 ± 1.38	n/a	n/a	n/a

(inter-dataset) slightly decreases compared to the (still unseen) test data in the case of intra-dataset validation. The figures also reveal that oblique splits, as compared to orthogonal splits, yield both better classification performance and smaller (average) tree depth. The advantage of oblique over orthogonal splits may result from the highly correlated features [18]. Further experiments would be required to investigate the influence of the elastic net penalty of Equation (8) in more detail.

Figure 1 compares the binary segmentation of the cerebrovascular networks for the different approaches applied to the test data $\mathcal{D}_2$ using the F_1-optimal operating points marked in Figure 4(b). Visually, the Frangi filter and partly also the RF-OSF model generate very smooth networks missing some of the details on the vessel surface. The ideal elliptical appearance model underlying the Hessian-based vesselness filters produces many false negatives at bifurcations, in particular, where the model assumptions do not hold. Here the classification approach is able to consider more complex geometries, that are in accordance with higher order filter responses in the training data. As already indicated by the precision-recall analysis, the axial views reveal that the Frangi segmentation varies significantly from the ground-truth labels in many cases, whereas the RF-OSF and especially the RF-SFT results are in much better agreement to the reference segmentation.

5 Conclusions and Future Work

We have compared two kinds of features for 3-D vessel segmentation using a machine learning approach. Starting from orthogonal subspace filtering, we learn an orthogonal basis from vessel patches to describe the local vessel appearance in a low-dimensional feature space. In a second step, we parametrize and approximate the highly structured base filters by Gaussian derivatives, which allows to efficiently decompose the image into a multiscale rotational basis using steerable filter theory [12,8]. Both kinds of features are used to train random forest classifiers for vessel segmentation. The steerable filters in fact allow to train a single classifier on normalized (canonically oriented) vessel

samples as proposed in [9] for 2-D filament detection. Our experiments on 3-D high-resolution srXTM imaging data of the murine visual cortex demonstrate that the steerable filter features outperform the orthogonal subspace features. Moreover, the machine learning approach proves to be superior to Hessian-based segmentation approaches, especially for vessel structures, such as bifurcations, that cannot easily be modeled explicitly and violate the common cylindrical appearance assumption. The RF classifiers show excellent classification performance on the 3-D datasets even for imperfect and incomplete training data as obtained by Otsu's method in our experiments. The proposed segmentation framework hence allows to fully automatically learn RF models for 3-D vessel segmentation on new datasets.

The choice of the type of splits to be used in the decision tree ensembles of the RF classifier turned out to have a major impact on the classification performance. For our task, oblique splits using linear regression are clearly favorable over univariate orthogonal splits. Besides a more comprehensive study on the choice of the elastic net penalty, it would be interesting to investigate if more complex information such as vessel caliber or centerline can be learned and predicted in a general and computationally cheap fashion on different types of 3-D angiographic datasets by extending the framework using Hough forests [7]. These additional data on the vessel morphology and topology may allow to ultimately reconstruct physiologically consistent full-fledged cerebrovascular networks possibly in combination with proper methods to replace or extend missing or faulty regions by synthetic vasculatures [25] in order to overcome shortcomings of the reconstruction technique or limitations of the imaging modality.

Acknowledgements. This work has been funded by the Swiss National Center of Competence in Research on Computer Aided and Image Guided Medical Interventions (NCCR Co-Me) supported by the Swiss National Science Foundation.

References

1. Breiman, L.: Random forests. Mach. Learn. 45, 5–32 (2001)
2. Canny, J.: Finding edges and lines in images. Tech. rep., Massachusetts Institute of Technology, Cambridge, MA, USA (1983)
3. Chen, Y.W., Lin, C.J.: Combining SVMs with various feature selection strategies. In: Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L. (eds.) Feature Extraction. STUDFUZZ, vol. 207, pp. 315–324. Springer, Heidelberg (2006)
4. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale Vessel Enhancement Filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
5. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. IEEE Trans. Pattern Anal. Mach. Intell. 13(9), 891–906 (1991)
6. Friedman, J.H., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33(1), 1–22 (2010)
7. Gall, J., Yao, A., Razavi, N., Van Gool, L., Lempitsky, V.: Hough forests for object detection, tracking, and action recognition. IEEE Trans. Pattern Anal. Mach. Intell. 33(11), 2188–2202 (2011)
8. González, G., Aguet, F., Fleuret, F., Unser, M., Fua, P.: Steerable Features for Statistical 3D Dendrite Detection. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 625–632. Springer, Heidelberg (2009)

9. González, G., Fleurety, F., Fua, P.: Learning rotational features for filament detection. In: CVPR 2009, pp. 1582–1589 (June 2009)
10. Hernández-Vela, A., Gatta, C., Escalera, S., Igual, L., Martin-Yuste, V., Radeva, P.: Accurate and Robust Fully-Automatic QCA: Method and Numerical Validation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 496–503. Springer, Heidelberg (2011)
11. Hirsch, S., Reichold, J., Schneider, M., Székely, G., Weber, B.: Topology and hemodynamics of the cortical cerebrovascular system. J. Cereb. Blood Flow Metab (April 2012)
12. Jacob, M., Unser, M.: Design of steerable filters for feature detection using canny-like criteria. IEEE Trans. Pattern Anal. Mach. Intell. 26(8), 1007–1019 (2004)
13. Jolliffe, I.T.: Principal Component Analysis, 2nd edn. Springer (2002)
14. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms. ACM Comput. Surv. 36, 81–121 (2004)
15. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. Med. Image Anal. 13(6), 819–845 (2009)
16. Lindeberg, T.: Edge detection and ridge detection with automatic scale selection. Int.J. Comput. Vis. 30, 465–470 (1996)
17. Menze, B.H., Kelm, B.M., Hamprecht, F.A.: From eigenspots to fisherspots - latent spaces in the nonlinear detection of spot patterns in a highly varying background. In: Decker, R., Lenz, H.J. (eds.) Advances in Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization., vol. 33, pp. 255–262. Springer (2006)
18. Menze, B.H., Kelm, B.M., Splitthoff, D.N., Koethe, U., Hamprecht, F.A.: On Oblique Random Forests. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) ECML PKDD 2011, Part II. LNCS, vol. 6912, pp. 453–469. Springer, Heidelberg (2011)
19. Moon, T., Stirling, W.: Mathematical methods and algorithms for signal processing. Prentice Hall (2000)
20. Otsu, N.: A threshold selection method from gray-level histograms. IEEE T. Syst. Man Cyb. 9(1), 62–66 (1979)
21. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. 12, 629–639 (1990)
22. Reichold, J., Stampanoni, M., Keller, A.L., Buck, A., Jenny, P., Weber, B.: Vascular graph model to simulate the cerebral blood flow in realistic vascular networks. J. Cereb. Blood Flow Metab. 29(8), 1429–1443 (2009)
23. Rigamonti, R., Türetken, E., González Serrano, G., Fua, P., Lepetit, V.: Filter learning for linear structure segmentation. Tech. rep., Swiss Federal Institute of Technology, Lausanne (EPFL) (2011)
24. Sato, Y., Nakajima, S., Atsumi, H., Koller, T., Gerig, G., Yoshida, S., Kikinis, R.: 3D Multi-Scale Line Filter for Segmentation and Visualization of Curvilinear Structures in Medical Images. In: Troccaz, J., Mösges, R., Grimson, W.E.L. (eds.) CVRMed-MRCAS 1997. LNCS, vol. 1205, pp. 213–222. Springer, Heidelberg (1997)
25. Schneider, M., Hirsch, S., Weber, B., Székely, G.: Physiologically Based Construction of Optimized 3-D Arterial Tree Models. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 404–411. Springer, Heidelberg (2011)
26. Schneider, M., Sundar, H.: Automatic global vessel segmentation and catheter removal using local geometry information and vector field integration. In: ISBI 2010, pp. 45–48 (April 2010)
27. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. NeuroImage 31(3), 1116–1128 (2006), `http://www.itksnap.org`

Pipeline for Tracking Neural Progenitor Cells

Jacob S. Vestergaard[1], Anders L. Dahl[1], Peter Holm[2], and Rasmus Larsen[1]

[1] Department of Informatics and Mathematical Modelling,
Technical University of Denmark
[2] Department of Basic Animal and Veterinary Sciences, Faculty of Life Sciences,
Copenhagen University

Abstract. Automated methods for neural stem cell lineage construction become increasingly important due to the large amount of data produced from time lapse imagery of *in vitro* cell growth experiments. Segmentation algorithms with the ability to adapt to the problem at hand and robust tracking methods play a key role in constructing these lineages. We present here a tracking pipeline based on learning a dictionary of discriminative image patches for segmentation and a graph formulation of the cell matching problem incorporating topology changes and acknowledging the fact that segmentation errors do occur. A matched filter for detection of mitotic candidates is constructed to ensure that cell division is only allowed in the model when relevant. Potentially the combination of these robust methods can simplify the initiation of cell lineage construction and extraction of statistics.

1 Introduction

Tracking of neural stem cells (NSCs) is fundamental in understanding the causes for cell fate outcomes in *in vitro* cell growth experiments. Previous studies of stem cells have used manually constructed cell lineages of a limited population to analyze, e.g., the developmental potential [7] or the morphological properties during cell division [5] and clearly show the benefit and importance of cell lineage construction. The development of automated methods for cell lineage construction is a key ingredient in processing large amounts of time lapse imagery and extracting meaningful statistics, previously not possible due to the need for extensive manual interaction.

We present a data driven pipeline for tracking pig neural progenitor cells in phase microscopy time lapse imagery using a supervised segmentation method, accommodating for small imprecisions in the manual annotation, and a completely data driven approach to tracking cells between time points. This pipeline enables for segmentation and tracking thousands of cells from a manual annotation of only 288 cells. The contribution includes novel approaches to mitosis detection, automatic correction of segmentation errors and data driven parameter estimation for the cell matching cost function.

The proposed mitosis detector is based on the observation that a NSC about to undergo mitosis becomes circular and moves out of focus of the imaging device,

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 155–164, 2013.

making it easily detectable. A similar behavior is observed by [5] for human neural progenitor cells.

Previously, systems aiming to accomplish the same have been proposed, including LEVER [8] incorporating published methods for segmentation and lineaging [1,2]. A limitation by this and other systems is the sensitivity to image data with a slightly different appearance. We explore the possibilities of overcoming this limitation by driving the analysis by simple manual annotation of the image data. This allows the segmentation algorithm to adapt to the problem at hand.

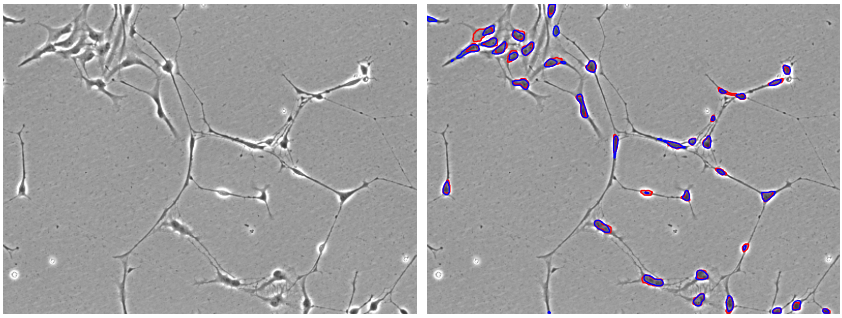
Manual annotation of neural progenitor cells are tedious and difficult even for an expert. A single image cannot be annotated without preceding and following images from the time series. The inherent inaccuracy in these annotations are accommodated for by the choice of segmentation algorithm, namely dictionary learning from image patches. This method exploits the property that the textural appearance of neural progenitor cells can be condensed to a number of typical image patches.

Tracking of the cells during the time lapse image sequence is reduced to match the cells between two time frames. This is accomplished by a modification of the bipartite graph formulation of the matching problem proposed by [6]. The modifications introduced are 1) restricting topology changes to ensure cell division occur during cell mitosis and 2) acknowledging that segmentation errors *are* present and minimizing their disruptions to the cell lineage construction.

The methods applied have been chosen based on the problems arising from analysis of an approximately 83 hours time lapse image sequence with 5 minutes between acquisitions. This sequence consists of 1000 phase contrast microscopy images of neural progenitor cells with very irregular shapes and movement patterns. An example of such an image can be seen in Figure 1a. In the following sections the methodology embedded in the proposed pipeline is outlined and results are reported.

2 Dictionary Learning for Robust Segmentation

The cell segmentation is based on a trained dictionary of image and label patches [3]. Each intensity patch in the dictionary has a corresponding label patch. The dictionary is build from manually annotated image exemplars by randomly sampling a set of intensity patches with corresponding label patches. In the training phase the aim is to find a dictionary that well represents the image texture and simultaneously have unique label patches. The label patches have the same spatial resolution as the intensity patches and in each pixel they store the probability of the labels in the training set. A label patch that has high probability for one class and low for other classes in each pixel is considered unique. To optimize the dictionary a weighted k-means procedure is employed where weights are updated in each step based on the uniqueness criterion.



(a) (b)

Fig. 1. a) Phase contrast microscopy image approximately 17 hours into the timeseries. b) Manual annotation (red) together with learned dictionary segmentation (blue) overlaid image.

Segmentation of an unknown image is computed using the trained dictionary. For each pixel an image patch is extracted and the label patch corresponding to the closest match in the intensity dictionary is assigned. The patches are overlapping, so the obtained probabilities are averaged.

In this experiment we chose the parameters for the segmentation based on a training and a test set. We had 15 manually annotated image where 8 were used for training and 7 were used for test. Our initial experiments suggested that we needed relatively large image patches, so we chose to downscale the images to half the size giving a spatial resolution of 300×400. The results of our experiments are shown in Table 1. Dice's coefficient denotes how well the segmentation captures the area and the ratio reported is the number of cells detected versus the number manually annotated. Thus a value above one is over segmentation and below one is under segmentation. The performance improves slightly going from a patch size of 7 to 9 but only little improvement is obtained by going from 9 to 11. We chose 9 as a good tradeoff between segmentation performance and computation time. It should be noted that it is a difficult task to manually annotate these images, so the results should be seen together with visual inspection of the segmentations as shown in Figure 1.

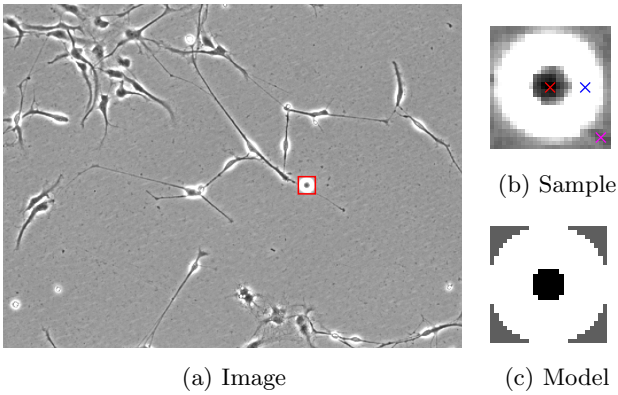
Table 1. Segmentation results obtained by varying patch sizes

	Training			Test		
Patch size	7	9	11	7	9	11
Dice's coefficient	0.80	0.81	0.81	0.78	0.79	0.80
$N_{\text{detected}}/N_{\text{true}}$	1.13	1.05	0.96	1.25	1.12	1.02

3 Mitosis Detection

Visual inspection of the time lapse imagery revealed that when a neural progenitor cell is about to undergo mitosis, it separates itself from the gel in the petri dish, floating up a bit and out of focus. When a cell is out of focus it has a very distinct pattern due to the imaging process.

This pattern can be derived analytically [10], but requires knowledge of the internal microscope parameters, which have not been available in this case. Therefore a model has been constructed from an image containing the pattern of interest. The image and extracted sample is shown in Figures 2a–b.



(a) Image (b) Sample (c) Model

Fig. 2. a) Phase contrast microscopy image with cell in pre-mitotic stage. b) Sample of out-of-focus cell extracted from phase contrast image. Intensity values are extracted from the marked points to transfer the contrast between the halo, center and background to the model. c) Constructed model.

The model is constructed by initializing an image of size 27×27, which is approximately the same size as the sample. The donut-like center of the sample is modeled by a 27×27 disk filter, where the hole in the donut is a 7×7 disk. The intensity values in these three regions are extracted from the sample as marked in Figure 2b. The resulting filter h, shown in Figure 2c is normalized by the maximum response from a convolution of the constructed model $\hat{h}$ with the sample S , such that $h_i = \frac{\hat{h}_i}{\max\{\hat{h} * S\}}$, $i \in \{1, \ldots, 27^2\}$ whereby subsequent filtering can be interpreted as "percentage of perfect response".

The constructed model is used for matched filtering of every phase contrast image. Connected components with a response above 0.9 and an eccentricity below 0.6 is marked as a mitotic candidate. The eccentricity is here the ratio of the distance between the foci of the ellipse and its major axis length. Examples of these detections can be seen in Figure 3.

This detector enables the tracking pipeline to detect and handle cell mitosis, which will be described in Section 4.

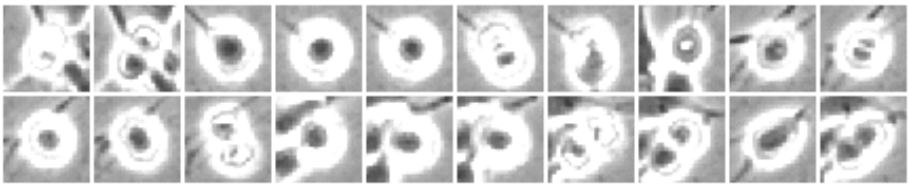


Fig. 3. 20 examples of detected mitotic candidates using the constructed model

4 Tracking

Finding the best match for each cell between two time points is needed to construct cell lineages. Here we employ a tracking method based on initially segmenting the cells, as described above, and subsequently matching the cells between two frames. This is opposed to integrating segmentation and tracking in a single scheme, such as the model evolution approach [4,9,11] where level sets are leveraged as a framework.

The goal is to match N cells $\{\mathcal{C}_i^{t-1}\}_{i=1}^{N}$ detected at the $t-1$'th time point to the M cells $\{\mathcal{C}_j^{t}\}_{j=1}^{M}$ at the t'th time point. Each of these cells are described with a feature vector $\mathbf{f}$ of length K, such that the feature vector for the j'th cell at time t will be denoted $\mathbf{f}_j^t$. Specifically we choose to describe each cell with its x- and y-coordinates and area, whereby $K = 3$.

To match the cells in a way that accounts for all cell features, the matching of cells between two time points can be formulated as a minimum cost problem. We adopt the formulation of the matching problem suggested by [6] where a bipartite graph with coupled edges is set up to accommodate for topology changes.

Tracking by acknowledging segmentation errors. Given the difficulty of the segmentation problem the tracking algorithm needs to accommodate for segmentation errors. The possible four types of segmentation errors are:

1. Undetected cell (false negative).
2. Two cells are mistakenly segmented as a single cell.
3. One cell is mistakenly segmented as two cells.
4. Cell detected where none is present (false positive).

It is assumed that any of these segmentation errors are only temporary, i.e., a cell is only undetected or mistakenly segmented for one or a few consecutive time points.

The model by [6] is modified to honor only the biologically possible topology changes, namely that a cell can only split into two if it is undergoing mitosis. A cell is marked as a mitotic candidate if it in the near-past (15 time points = 75 minutes) has been detected as in the pre-mitotic stage using the detector described in Section 3. The graph illustrating the possible topology changes between two time points is shown in Figure 4.

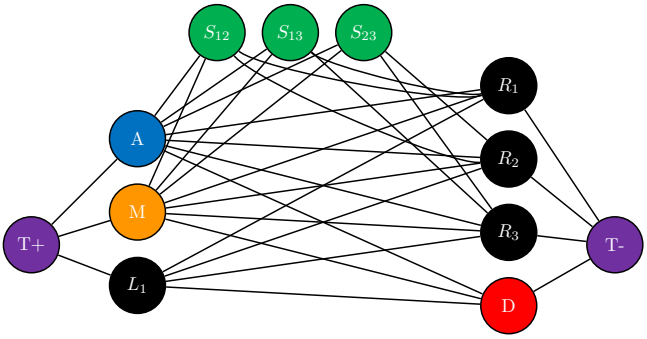


Fig. 4. Graph formulation of the matching problem. In this example there is one cell M detected as undergoing mitosis, wherefore allowed to split. The other cell L_1 can only move to cells R_1, R_2, R_3 or disappear.

Cell merging has not been included in the model as it is not possible for neural progenitor cells to merge with each other. Thereby the possibilities remaining for a cell – not marked as a mitotic candidate – are to move, appear or disappear between frames. The "appear" and "disappear" events include the cases where a cell enters or leaves the image frame, as suggested originally, but also covers the option for a cell to disappear or appear anywhere in the image. This is necessary to accommodate for the segmentation errors listed above.

In the case where a cell from time point $t-1$ is found to disappear, without being near the image border, a phantom (a copy) of the cell is included in the set of cells at time point t and these are coupled as an ordinary "move" event. If no match is found for the phantom for a few time steps (here we choose 2 as the limit), the cell is finally marked as disappeared. For the first two cases listed above, the effect is obviously that the gap between detections is filled with the phantom. For the third and fourth cases, the spurious detection of a new cell in a few images will result in a very short cell track which can easily be detected during post-processing of the lineages. This approach effectively accommodates for the segmentation errors and allows for a robust tracking.

Edge costs. Calculation of the edge costs in the graph problem is inspired by [1]. The assigned cost $a(\mathcal{C}_i^{t-1}, \mathcal{C}_j^t)$ for matching the i'th cell at time point $t-1$ to the j'th cell at time point t is the Mahalanobis distance

$$a(\mathcal{C}_i^{t-1}, \mathcal{C}_j^t) = \sqrt{(\mathbf{d}_{ij} - \boldsymbol{\mu})^T \boldsymbol{\Sigma} (\mathbf{d}_{ij} - \boldsymbol{\mu})} \quad \text{where} \quad \mathbf{d}_{ij} = \mathbf{f}_j^t - \mathbf{f}_i^{t-1} \tag{1}$$

from the proposed change feature difference vector $\mathbf{d}_{ij}$ to a reference distribution described by the mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

In [1] this reference distribution is estimated by manually supervising and correcting the tracks from a number of time points. Here we employ a purely data driven approach for estimating this distribution. Specifically, cells are matched over a period of 100 time points with a matching criterion specified as the nearest

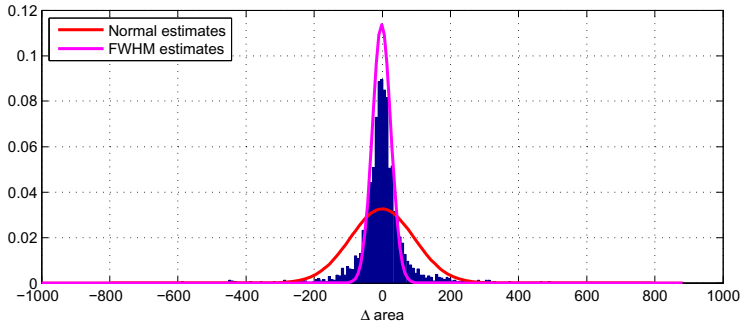


Fig. 5. Illustration of the FWHM principle when estimating parameters for the distribution describing change in area from a one-to-one nearest neighbor tracking

neighbor, with the constraint that only one-to-one correspondences are accepted, i.e., the nearest neighbor for cell i at time point $t-1$ must also have this cell as its own nearest neighbor. Thereby only the very most obvious matches are included. The feature differences for these P matches are extracted and collected in the $P \times K$ matrix from which the mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ of the reference distribution are estimated.

However, with this approach a few erroneous matches are inevitably still included, whereby the parameter estimates are corrupted. To ensure that this does not happen, the principle of full-width-at-half-maximum (FWHM) is used to extract the dominant distribution for each feature difference. Figure 5 illustrates the difference in parameter estimates for the reference Gaussian distribution. It is clear that this is approach is necessary in order to extract viable parameters in a data driven context.

The cost for a cell to appear or disappear is set as the cost of moving from or to a cell with 1) a position of 10 standard deviations of the change in coordinates away and 2) its area set to the mode of the manually annotated cells' area. This forces the model to only let a cell appear or disappear if no suitable match is found in its proximity.

Splitting a cell into two has the cost of moving the cell to the convex hull of the resulting two cells. While there exist $\binom{M}{2}$ pairs of potential split candidates, this number can be heavily reduced by selecting only the top β percent pairs sorted according to mutual distance as proposed by [6].

5 Results

The methodology outlined above is applied to the entire time lapse image sequence of 1000 phase contrast microscopy images with 5 minutes between acquisitions. The images were captured using a Nikon BioStation IMQ with a magnification level of 10, an exposure time of 1/125s and a resolution of 600×800 pixels. The microscope acquires images in a 4×4 grid, but the images analyzed

here are only from a single data point and therefore only 1/16 of the available scene. Thus this analysis should be seen as a proof-of-concept rather than a full analysis.

The mitosis detector described in Section 3 only allows cell division when cells undergo mitosis. Examples of the automatically detected mitotic cells can be seen in Figure 6. From these sequences it is seen that the detector successfully detects the out-of-focus shape characterizing the mitotic candidates, allowing the cell to divide within the near-future (chosen as 15 time points). The examples illustrate how the mitoses can be detected even in highly confluent areas. A total of 29 mitoses were detected during the entire time lapse sequence using this method. For completeness it should be mentioned that only 62% were true positive detections.

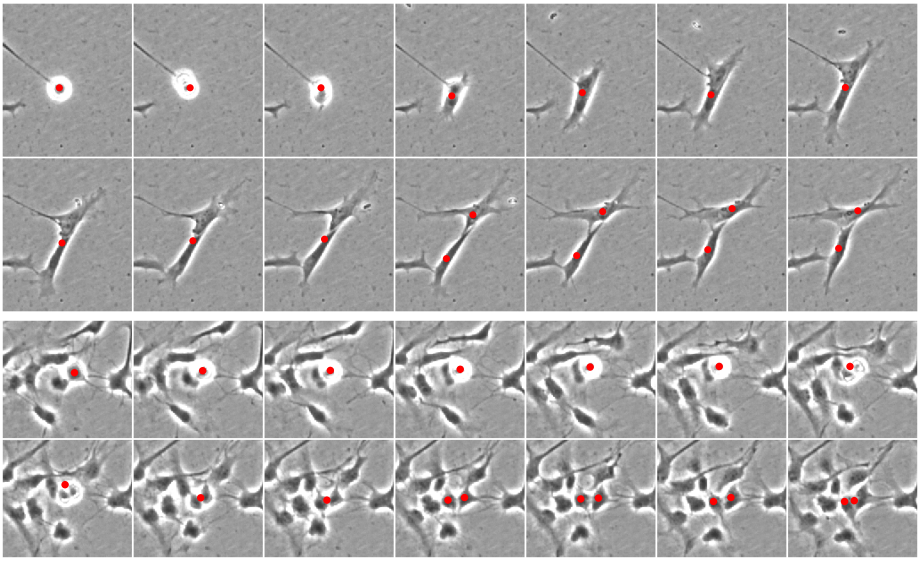
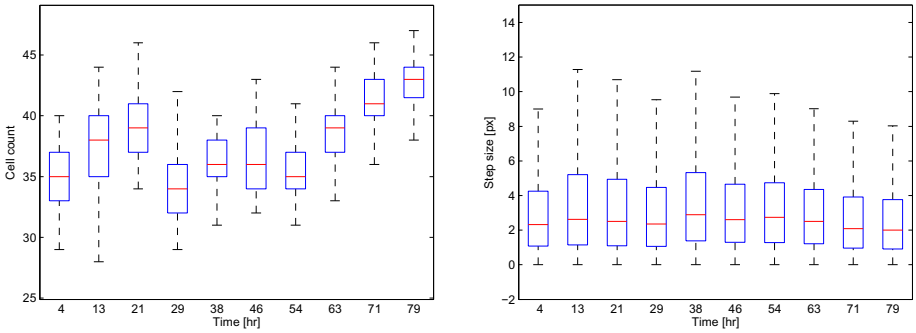


Fig. 6. Two examples of detected mitotic events. Each sequence shows the area of interest from −10 to +3 time points around the detected cell division. The dots indicate the centroids of the detected cells.

While the pipeline enables us to follow cells over time, direct interpretation of the cell trajectories is of limited value compared to statistics derived from these. In Figure 7 cell count and step lengths are documented as a function of time. The time series has been divided into ten equally sized periods, wherefore each statistic can be visualized as ten boxes. It is seen that the number of cells increase slightly in the beginning of the period followed by a decrease. Over the entire time period a definite increase in cell count is seen, which is expected given that mitosis occur.



(a) Cell count (b) Step lengths

Fig. 7. Simple statistics for the segmented and tracked time series. Each box represents 1/10 of the 83 hour period. The red line indicates the median, the edges of the box the 25th and 75th percentiles and the whiskers extend to the most extreme data points not considered outliers. Outliers are omitted. a) Cell count as a function of time. b) Cell step length between frames as a function of time.

The step length is reasonably constant over the period, except for a reduction in the last 16 hours. Visual inspection of the time lapse imagery confirms that the cells are less mobile towards the end of the time series.

Statistics concerning individual cells' tendency to undergo mitosis, i.e., whether it is the daughters of the same cell that continuously divides, are interesting, but the number of of mitotic events are too few to state anything with regards to this. To answer this question and similar, a large scale study using the proposed pipeline will be carried out in the near future.

6 Conclusions

A tracking pipeline based on a few manual annotations has been proposed. The pipeline accommodates for imprecisions in the manual annotation, by choice of segmentation method, and segmentation errors in the tracking model. Parameters for the tracking model were chosen using the principle of full-width-at-half-maximum to ensure meaningful extraction of parameters in a data driven context. A detector for mitotic candidate cells enables the model to restrict topology changes to those valid for a neural progenitor cell.

Validation of the segmentation algorithm was performed using a division into training and test set of 8 and 7 fully annotated images respectively. It was shown that a Dice's coefficient of 0.79 could be achieved while preserving a slight over-segmentation using a dictionary atom size of 9×9 in an image down sampled to 50% of the original size.

The pipeline was applied to a sequence of 1000 phase contrast microscopy images of moving and proliferating neural progenitor cells of very irregular shapes and movement patterns and varying confluence. A total of 30 mitotic events were detected and simple statistics were extracted from the cell lineages. While leaving

room for improvements, this work shows that dictionary learning of discriminative image patches combined with a topology change enabling graph formulation is a flexible pipeline that can be applied even to very difficult tracking problems.

Acknowledgments. The authors would like to thank laboratory technician Jytte Nielsen from Department of Basic Animal and Veterinary Sciences, Faculty of Life Sciences, Copenhagen University, for manual annotation of cells and pleasant collaboration.

References

1. Al-Kofahi, O., Radke, R., Goderie, S., Shen, Q.: Automated Cell Lineage Construction. Cell Cycle 5(3), 327–335 (2006)
2. Cohen, A.R., Gomes, F.L.F., Roysam, B., Cayouette, M.: Computational prediction of neural progenitor cell fates. Nature Methods 7(3), 213–218 (2010)
3. Dahl, A., Larsen, R.: Learning dictionaries of discriminative image patches. In: Proceedings of the British Machine Vision Conference, BMVA (2011)
4. Dzyubachyk, O., van Cappellen, W.A., Essers, J., Niessen, W.J., Meijering, E.: Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. IEEE Transactions on Medical Imaging 29(3), 852–867 (2010)
5. Keenan, T.M., Nelson, A.D., Grinager, J.R., Thelen, J.C., Svendsen, C.N.: Real time imaging of human progenitor neurogenesis. PloS One 5(10), e13187 (2010)
6. Padfield, D., Rittscher, J., Roysam, B.: Coupled minimum-cost flow cell tracking for high-throughput quantitative analysis. Medical Image Analysis 15(4), 650–668 (2011)
7. Ravin, R., Hoeppner, D., Munno, D., Carmel, L.: Potency and fate specification in CNS stem cell populations in vitro. Cell Stem Cell 3(6), 670–680 (2008)
8. Winter, M., Wait, E., Roysam, B., Goderie, S.K., Ali, R.A.N., Kokovay, E., Temple, S., Cohen, A.R.: Vertebrate neural stem cell segmentation, tracking and lineaging with validation and editing. Nature Protocols 6(12), 1942–1952 (2011)
9. Yang, F., Mackey, M.A., Ianzini, F., Gallardo, G., Sonka, M.: Cell Segmentation, Tracking, and Mitosis Detection Using Temporal Context. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 302–309. Springer, Heidelberg (2005)
10. Yin, Z., Li, K., Kanade, T., Chen, M.: Understanding the Optics to Aid Microscopy Image Segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 209–217. Springer, Heidelberg (2010)
11. Zimmer, C., Labruyère, E., Meas-Yedid, V., Guillén, N., Olivo-Marin, J.-C.: Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cell-based drug testing. IEEE Transactions on Medical Imaging 21(10), 1212–1221 (2002)

Automatic Heart Isolation in 3D CT Images

Hua Zhong[1], Yefeng Zheng[1], Gareth Funka-Lea[1], and Fernando Vega-Higuera[2]

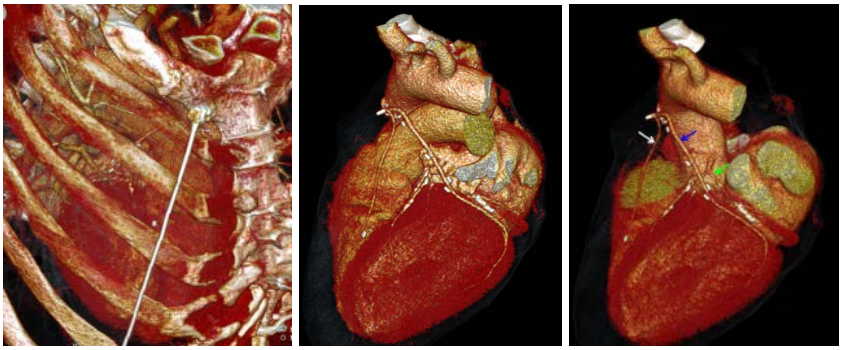
[1] Imaging and Computer Vision, Siemens Corporate Technology, Princeton, NJ, USA
[2] Healthcare Sector, Siemens AG, Forchheim, Germany
hua.zhong@gmail.com, yefeng.zheng@siemens.com

Abstract. In this chapter, we present an automatic heart segmentation algorithm for the diagnosis of coronary artery diseases (CAD). The goal is to visualize the heart from a cardiac CT image with irrelevant tissues such as the lungs, rib cage, pulmonary veins, pulmonary arteries and left atrial appendage hidden so that doctors can clearly see the major coronary artery trees, aorta and bypass arteries if they exist. The algorithm combines a model-based detection framework with data-driven post-refinements to create a mask for a given cardiac CT image that contains only the relevant part of the heart. The marginal space learning [1] technique is used to localize mesh model or landmark points of different cardiovascular structures in the CT volume. Guided by such detected models, local data-driven voxel-based refinements are employed to produce precise boundaries of the heart mask. The algorithm is fully automatic and can process a 3D cardiac CT volume within a few seconds.

1 Introduction

Coronary Artery Disease (CAD) or Coronary Heart Disease (CHD) is the leading cause of death in the world [2]. Computed tomography (CT) is often used for diagnosis and treatment planing of CAD/CHD. Usually, 2D images from the stack of acquired axial images are used for diagnosis. However, only a small portion of a coronary artery is visible in a single 2D axial image. A 3D visualization provides a global and intuitive view for physicians to identify suspicious coronary segments (which are then verified on 2D slices). However, in cardiac CT images, the whole chest is imaged: both the heart and surrounding anatomical structures, which usually block the direct view of the heart in a 3D visualization. Figure 1 (a) shows an image from a CT scan. Ribs, sternum, and other structures block any direct view of the heart. Manual segmentation of the heart is tedious and error-prone. Here, we introduce a fully automatic system based on machine learning algorithms to reliably isolate the heart from 3D CT images. Figure 1 (b) presents a 3D visualization of the heart after segmenting it from the surrounding non-cardiac tissues. With this result, physicians can easily see detailed heart structures. However, the left atrial appendage (LAA), the pulmonary arteries (PA) and pulmonary veins (PV) still block the left coronary artery (LCA) tree. Figure 1 (c) is the improved result with the LAA, PA and PV being removed. Now, the left coronary artery tree can be clearly seen without any occlusion. Such a 3D view can greatly help physicians to perform CAD/CHD diagnosis.

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 165–180, 2013.



(a) (b) (c)

Fig. 1. Heart isolation visualization. (a) The original CT scan. Note that bones blocked any direct view of the heart. (b) The result of the pericardial isolation which only isolates the whole heart. Still, the pulmonary artery (PA), the pulmonary veins (PV) and the left atrial appendage (LAA) occlude the left coronary artery (LCA). (c) The result of final heart isolation. The PA, PV and LAA are removed automatically and the LCA (green arrow) is easily seen. After heart isolation, the plaques that block the left coronary artery tree can be easily identified. Note that in this case, there are two bypass arteries (white and blue arrows). The algorithm reliably keeps them intact.

There are a couple of other applications of heart isolation, e.g., radiotherapy planning and calcium scoring. In radiotherapy planning for the treatment of lung or liver tumors, the heart needs to be identified as part of the effort to reduce the radiation to it. Normally, a non-contrasted volume, as shown in the bottom row of Figure 2, is used for radiotherapy planning. A non-contrasted scan is used for calcium scoring as well. A calcium score is a well-established biomarker to predict future cardiac events [3]. To calculate a calcium score, the calcified coronary artery plaques (appearing as bright voxels in CT) need to be segmented. However, other bright tissues (e.g., the rib cage and sternum) need to be excluded. Heart isolation can provide a region of interest for detecting coronary calcifications.

Heart isolation is a hard problem due to the following challenges.

1. The boundary between the heart and some of the neighboring tissues (*e.g.*, liver and diaphragm) is quite weak in a CT volume.
2. The heart is connected to other organs by several major vessel trunks (*e.g.* aorta, vena cava, pulmonary veins, and pulmonary arteries). We must cut those trunks somewhere (normally at the position where the vessels connect to the heart), though there is no visible boundary.
3. The deformation of the whole heart in a cardiac cycle is more complicated than each individual chamber. This brings a large variation in the heart shape. Furthermore, there are quite a few scans with a part of the heart missing in the captured volume, especially at the top or bottom of the heart, which introduces extra shape variation.

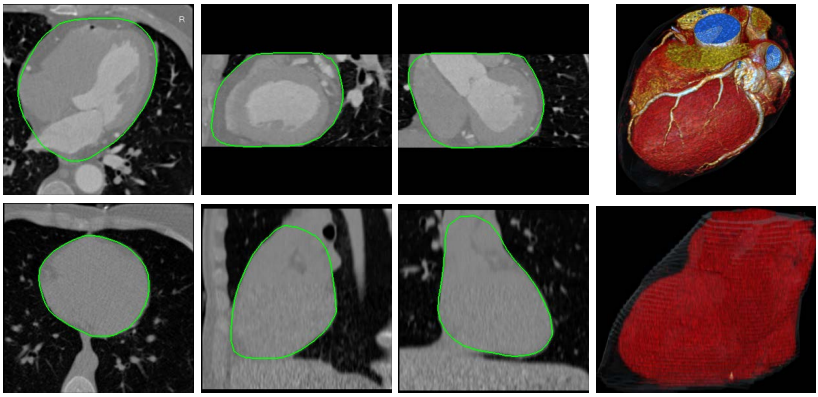


Fig. 2. Cardiac pericardium segmentation for a contrasted scan (top row) and a non-contrasted scan (bottom row). The first three columns show orthogonal cuts of the volume with green contours showing the automatically segmented heart surface mesh. The last column is 3D visualization of the segmented heart.

4. We are targeting both contrasted and non-contrasted data (as shown in Figure 2), instead of just one homogeneous set (*e.g.*, [4] for contrasted data and [5] for non-contrasted data). This presents an additional challenge.

While most previous work on heart segmentation focuses on segmenting heart chambers [1], there are only a limited number of papers on heart isolation. Atlas based methods are often used to segment the heart. For example, Rikxoort *et al.* [6] presented an adaptive, local multi-atlas based approach. It took about 30 minutes to segment a scan. Lelieveldt *et al.* [7] proposed another atlas based approach, segmenting several organs (*e.g.*, lung, heart, and liver) in a thoracic scan using a hierarchical organ model. Their approach only provided a rough segmentation and an error as large as 10 mm was regarded as a correct segmentation. It took 5 to 20 minutes to process one volume. Gregson *et al.* [8] proposed to segment the lungs first and the heart was approximated as a sphere between the left and right lungs. Moreno *et al.* [5] presented a more thorough model for the geometric relationship between lungs and the heart. Funka-Lea *et al.* [4] proposed an automatic approach based on a graph cut segmentation. They used the volumetric barycenter weighted by intensity as an initial estimate of the heart center. A small ellipsoid was put at the estimated heart center and progressively grown until it touched the transition between heart and lung (which was easy to detect in a CT volume). Graph cut was then applied to achieve the final detailed boundary delineation. It took about 20 seconds to process one volume, which was still slow for a clinical application.

In this chapter, we present an efficient and fully automatic approach for heart isolation. It contains two steps:

1. Pericardial Isolation. In this step, the whole heart (including the PA, PV and LAA) is isolated from surrounding structures. Most of the right coronary artery (RCA) tree can be seen after this step.
2. Removal of the PA, PV and LAA. Based on the pericardial isolation result, the PA, PV and LAA are automatically segmented and removed from the image. After that, the left coronary artery (LCA) tree is clearly visible. This step is not necessary for the radiotherapy planning and calcium scoring using non-contrasted scans.

We describe the two steps in details in the following sections. However, it is worth noting that all the algorithms used in both steps heavily rely on machine learning algorithms to reliably estimate the heart or the heart components' location, orientation and size. Because of this, the machine learning based component segmentation algorithm is briefly described first, followed by detailed description of the two steps of the heart isolation algorithm.

2 Marginal Space Learning Based Object Segmentation

Marginal space learning (MSL) [1] has been proposed as an efficient and robust method for 3D anatomical structure detection/segmentation in medical images. In MSL, object detection or localization is formulated as a binary classification problem: whether an image block contains the target object or not. For detection, an object can be found by testing exhaustively all possible combinations of locations, orientations, and scales using a trained classifier. However, exhaustive searching is very time consuming. The idea of MSL is not to learn a monolithic classifier, but split the estimation into three steps: position estimation, position-orientation estimation, and position-orientation-scale estimation. Each step can significantly prune the search space, therefore resulting in an efficient object detection algorithm. Please refer to [1] for more details of MSL.

After MSL based object pose estimation, a mean shape (which is trained on a set of example shapes of the object to be segmented) is aligned with the estimated translation, rotation, and scale as an initial shape. The mean shape is generally calculated as the average of the normalized shapes in an object-centered coordinate system. Therefore, the mean shape depends on the definition of the object-centered coordinate system, which is often set heuristically [1]. After initialization, we deform the shape for more accurate boundary delineation under the guidance of shape prior and a learning based boundary detector.

The MSL based object segmentation algorithm is extensively used by the heart isolation algorithm to segment the pericardial surface, the left atrium, the pulmonary artery trunk, the aortic root, and many landmark points as described in the following sections. Each object has its own trained pose detector, mean shape and boundary detector. With a reasonable amount of training data, these objects can be reliably segmented within a fraction of a second. In the following sections we will describe the two-step heart isolation algorithm in details.

3 Cardiac Pericardium Segmentation

The segmentation of the pericardial surface is based on the MSL algorithm. However, the mean shape generation process is modified to better fit the requirement for an accurate initialization for hearts. The MSL segmentation result is a smooth 3D mesh. However, such smooth meshes cannot capture the details of the heart surface around the rib cage or sternum. We introduce a post-processing step to fix this.

3.1 Optimal Mean Shape for Accurate Shape Initialization

In MSL segmentation, after the initial pose of the object is estimated, a mean shape based on training data is fit to the image as an initial shape for later boundary delineation. In [1], the orientation of a heart chamber is defined by its long axis; the position and scale are determined by the bounding box of the chamber surface mesh. Although working well in applications with relatively small shape variations, the mean shape derived using the previous methods is not optimal for this application.

Here, we present an approach to searching for an optimal mean shape $\bar{m}$ that represent the whole population well. A group of shapes, $M_1, M_2, \ldots, M_N$ are supposed to be given and each shape is represented by J points $M_i^j, j = 1, \ldots, J$. The optimal mean shape $\bar{m}$ should minimizes the residual errors after alignment,

$$\bar{m} = arg \min_{m} \sum_{i=1}^{N} \|\mathcal{T}_i(m) - M_i\|^2 . \tag{1}$$

Here, $\mathcal{T}_i$ is the corresponding transformation from the mean shape $\bar{m}$ to each individual shape M_i. This procedure is called generalized Procrustes analysis [9] in the literature. An iterative approach can be used to search for the optimal solution. We first randomly pick an example shape as a mean shape. We then align each shape to the current mean shape. The average of the aligned shapes (the simple average of the corresponding points) is calculated as a new mean shape. The iterative procedure converges to an optimal solution after a few iterations.

Previously, the similarity transformation (with isotropic scaling) has often used as the transformation $\mathcal{T}$. MSL can estimate the anisotropic scales of an object efficiently. By removing more deformations, the shape space after alignment is more compact and the mean shape can represent the whole population more accurately. Therefore, we use an anisotropic similarity transformation to represent the transformation between two shapes,

$$\hat{T}, \hat{R}, \hat{S} = arg \min_{T,R,S} \sum_{j=1}^{J} \left\| \left(R \begin{bmatrix} S_x & 0 & 0 \\ 0 & S_y & 0 \\ 0 & 0 & S_z \end{bmatrix} M_1^j + T \right) - M_2^j \right\|^2 . \tag{2}$$

To the best of our knowledge, there are no closed-form solutions for estimating the anisotropic similarity transformation. In this work, we propose a two-step iterative approach to searching for the optimal transformation. Suppose there is

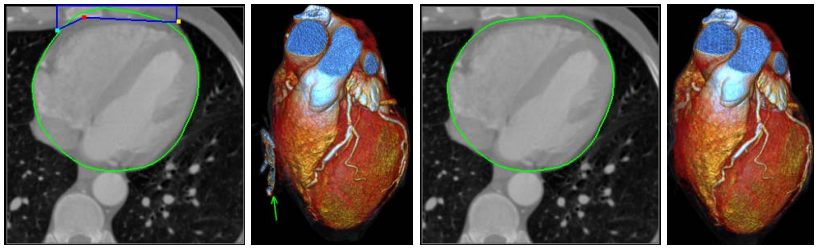


Fig. 3. Post-processing to exclude the rib cage from the heart mask. **Left**: Cross-section and 3D visualization of the result before post-processing. **Right**: After post-processing.

a common scale $s = (S_x + S_y + S_z)/3$, let $S'_x = S_x/s$, $S'_y = S_y/s$, and $S'_z = S_z/s$. Equation (2) can be re-written as

$$\hat{T}, \hat{R}, \hat{S} = arg \min_{T,R,S} \sum_{j=1}^{J} \left\| \left(Rs \begin{bmatrix} S'_x & 0 & 0 \\ 0 & S'_y & 0 \\ 0 & 0 & S'_z \end{bmatrix} M_1^j + T \right) - M_2^j \right\|^2 . \tag{3}$$

In the first step, suppose the anisotropic scales S'_x, S'_y, and S'_z are known. (At the beginning, we can assume the scaling is isotropic, $S'_x = 1$, $S'_y = 1$, and $S'_z = 1$.) We can calculate the isotropic similarity transformation using a closed-form solution [9]. In the second step, assuming that the isotropic similarity transformation (T, R, s) is given, we estimate the optimal anisotropic scales S'_x, S'_y, and S'_z. Simple mathematic derivation gives us the following closed-form solution,

$$\hat{S}'_x = \frac{\sum_{j=1}^{J} M_1^j(x) P_2^j(x)}{\sum_{j=1}^{J} M_1^j(x)^2} \quad \hat{S}'_y = \frac{\sum_{j=1}^{J} M_1^j(y) P_2^j(y)}{\sum_{j=1}^{J} M_1^j(y)^2} \quad \hat{S}'_z = \frac{\sum_{j=1}^{J} M_1^j(z) P_2^j(z)}{\sum_{j=1}^{J} M_1^j(z)^2}, \tag{4}$$

where

$$P_2^j = \frac{1}{s} R^{-1} (M_2^j - T). \tag{5}$$

The above two steps iterate a few times until they converge.

With a module solving the anisotropic similarity transformation between two shapes, we can plug it into the generalized Procrustes analysis method to search for the optimal mean shape $\bar{m}$. Besides the optimal mean shape, the optimal alignment $\mathcal{T}_i$ from the mean shape to each example shape is also obtained as a by-product. The transformation parameters of the optimal alignment provide the pose ground truth that MSL can learn to estimate.

3.2 Excluding Rib Cage from Heart Mask

For most cases, good segmentation results can be achieved after 3D heart pose detection and boundary delineation. However, for a few cases, a part of the rib cage (sternum and ribs) may be included in the heart mask (left columns of Figure 3) since the heart boundary is quite weak around that region. A post-processing step is further applied to explicitly segment the sternum and ribs based on adaptive thresholding and connected component analysis. We first detect three landmarks, namely the sternum (red dot), the left (yellow dot) and

right (cyan dot) lung tips on each slice, as shown in the left columns of Figure 3. These landmarks determine a region of interest (ROI) (indicated by a blue polygon in Figure 3). A machine learning based technique is used to detect the landmarks on each slice. To be specific, 2D Haar wavelet features and the probabilistic boosting tree (PBT) [10] are used to train a detector for each landmark.

After landmark detection, we extract the ROI on each slice. Stacking the ROIs on all slices, we get a volume of interest (VOI). Normally, bones are brighter than the soft tissues in a CT volume, therefore, we can use intensity thresholding to extract the rib cage. However, due to the variations in the scanners, patients, and scanning protocols, a predefined threshold does not work for all cases. An adaptive optimal threshold [11] is automatically determined by analyzing the intensity histogram of the VOI. For some cases, a part of a chamber may be included in the VOI, however this is rare. Three dimensional connected component analysis of the bright voxels is performed and only the large components are preserved as the rib cage. We then adjust the heart mesh to make sure the rib cage is completely excluded from the mask (see the right columns of Figure 3).

3.3 Pericardium Segmentation Results

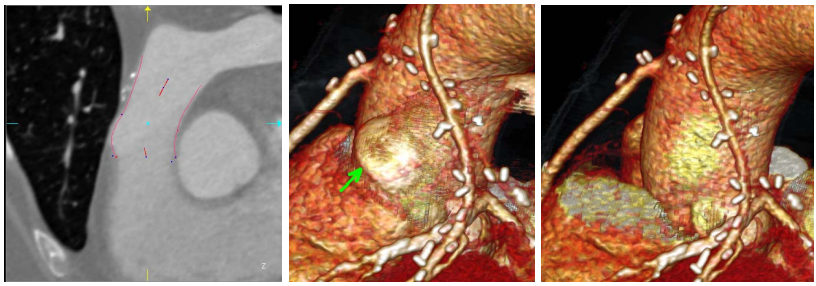
The method has been tested on 589 volumes (including both contrasted and non-contrasted scans) from 288 patients. The scanning protocols are heterogeneous with different capture ranges and resolutions. Each volume contains 80 to 350 slices and the slice size is 512×512 pixels. The resolution inside a slice is isotropic and varied from 0.28 mm to 0.74 mm, while slice thickness is generally larger than the in-slice resolution and varied from 0.4 mm to 2.0 mm.

For training and evaluation purposes, the pericardium surface of the heart was annotated, using a semi-automatic tool, with a triangulated mesh of 514 points and 1024 triangles. The cross-volume point correspondence was established using the rotation-axis based resampling method [1]. The point-to-mesh error, E_{p2m}, was used to evaluate the segmentation accuracy. For each point in a mesh, we search for the closest point in the other mesh to calculate the minimum distance. We calculate the point-to-mesh distance from the detected mesh to the ground-truth mesh and vice versa to make the measurement symmetric. A four-fold cross-validation was used to evaluate the performance of the algorithm.

First, we evaluate the shape initialization error of the optimal mean shape and the heuristic bounding-box based mean shape [1]. After MSL based heart pose estimation, we align the mean shape with the estimated position, orientation, and anisotropic scales. We then calculate the error E_{p2m} of the aliged mean shape w.r.t. the ground truth mesh. As shown in Table 1, the optimal mean shape is more accurate than the heuristic bounding-box based mean shape. It reduces the mean initialization error from 4.35 mm to 3.60 mm (a 17% reduction). After shape initialization, we deform the mesh under the guidance of a learning based boundary detector, which further improves the boundary delineation accuracy. As shown in Table 1, the mean error is 2.12 mm if we start from the bounding-box based mean shape. Using the proposed optimal shape initialization, we can reduce the final mean error

Table 1. Comparison of the proposed optimal mean shape and the heuristic bounding-box based mean shape [1] on shape initialization and final heart isolation errors. The point-to-mesh error (in millimeters) is used to measure the accuracy in the boundary delineation.

	Shape Initialization		Final Segmentation	
	Bounding-Box Mean Shape	Optimal Mean Shape	Bounding-Box Mean Shape	Optimal Mean Shape
Mean Error	4.35	**3.60**	2.12	**1.91**
Std Deviation	1.43	1.05	0.89	0.71
Median Error	4.11	**3.52**	1.89	**1.77**



(a) (b) (c)

Fig. 4. Directly applying the model-based machine-learning algorithm from [1] on the PA root usually result in a thin layer of the PA (green arrow) remaining in the image (b), even though the mesh looks accurate in (a). That's because the mesh model's resolution cannot capture the voxel-level details of the shape. For comparison, our algorithm can create a clean mask with the PA removed (c).

further to 1.91 mm (a 10% reduction). Our method works well on both contrasted and non-contrasted scans. The mean and median errors on the contrasted data are 1.85 mm and 1.71 mm, respectively. The corresponding errors increase moderately on the non-contrasted data to 2.22 mm and 2.11 mm, respectively.

We also compared our approach with the graph cut based approach proposed by Funka-Lea *et el.* [4], which was used for 3D visualization of the heart. Tissues darker than the myocardium (e.g., lung) included in the heart mask does not effect the visualization since the intensity window can be tuned to hide these extra tissues. Consequently the outputs of the two methods are not likely to be identical and whe should expect more accuracy in our proposed method. Keeping this in mind, the mean and median errors achieved by the graph-cut based method are 4.60 mm and 4.00 mm, respectively (our method is 1.85 mm and 1.71 mm). Furthermore, our method is about 10-20 times faster than the graph-cut based method.

4 Removal of the PA, PV and LAA

The pericardial isolation can separate the heart from surrounding non-cardiac structures such as the lungs and ribs. However, it is not sufficient for the 3D

visualization of coronary arteries since some heart structures such as the PA, PV and LAA still remain and cover the proximal left coronary artery (LCA) tree in many cases. In this step, we segment and remove these heart structures to reveal the LCA.

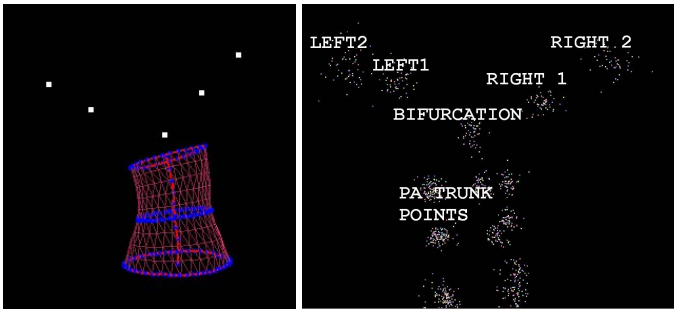
Since the PA, PV and LAA are very close to the coronary artery tree and they are all connected to heart chambers we want to keep, pure data-driven algorithms such as region growing cannot segment them cleanly without leaking into nearby chambers or the aorta. Though the model-based segmentation algorithm[1] can reliably detect the anatomies based on mesh models, it has some limitations. First, it works well for anatomies with relatively smaller variations, like the four chambers, but not highly variable structures like the LAA and PV, which can hardly be represented by a single-part mesh model. Second, note that standard local refinements based on a statistical shape model [12] and mesh smoothing algorithms [13] are used by the algorithm to generate a smoothed mesh. However, such a smoothed mesh, when converted to a voxel mask, may not cover all the voxels of the detected anatomy and consequently will generate visible artifacts (Figure 4).

To overcome these problems, we combine a local region growing algorithm with the global shape model to solve the PA, PV and LAA segmentation problems. We use slightly different segmentation algorithms for each of the PA, PV and LAA. However, the frameworks of all these algorithms are similar: a global shape-model detected by the MSL [1] and local refinement based on the statistics of voxel intensities. After global shape model (either mesh based or fiducial control point based) based detection/segmentation, we use constrained local intensity based region growing algorithms to refine the shapes and generate a detailed voxel mask of the objects. In order to avoid any removal of the aorta and LA which we want to keep for context, we also use a model-based algorithm to explicitly segment them and the segmented mask is used as a "protection" zone where no removal is allowed. Using the proposed method, we can achieve a fully-automatic, efficient and clean removal of the PA, PV, and LAA for 3D visualization of the LCA.

4.1 Globe Shape Segmentation

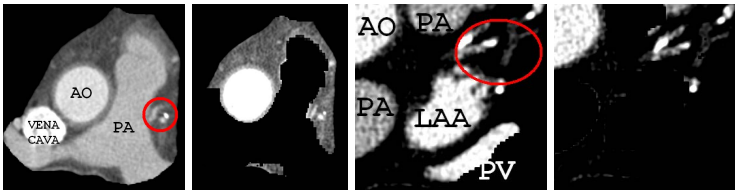
Pulmonary Artery Model: The PA trunk root, the portion of the PA from the pulmonary valve to the bifurcation, is modeled as a tubular mesh. From the bifurcation, it is difficult to approximate the shape with a tube. In this case, we use five fiducial control points: one at the bifurcation, two at the left PA branch and two at the right PA branch as shown in Figure 5 (a). We first describe how the PA trunk mesh is detected.

For the PA trunk mesh, we use the MSL algorithm [1]. The shape model, the bounding box detector and the boundary detector were trained with 320 manually annotated volume data. After the the PA trunk is detected, the detection of the five fiducial points from the PA bifurcation is a mixture of a statistical shape model and individual fiducial point detectors using the MSL algorithm [1] trained on 120 manually labeled volumes. The reason for this mixture is that in



(a) (b)

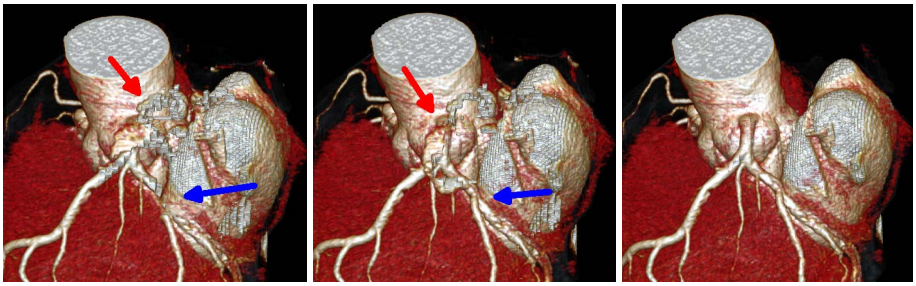
Fig. 5. PA model: (a) the mesh and five fiducial point model. (b) the statistical shape model for detecting the fiducial points (bifurcation, left 1 and 2, right 1 and 2). Based on 120 manually labeled data, we select nine points from the PA trunk mesh and combine them together with the five PA fiducial points to create a statistical shape model.



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

Fig. 6. Voxel-based refinement for the PA, PV, and LAA. (a) Before removal, the bypass arteries are highlighted by the red circle. (b) the PA and the vena cava are removed by region growing while bypass right adjacent to PA is kept intact. (c) before removal, we can see the small isolated chambers of the LAA very close to the coronary arteries highlighted by the red circle. (d) the LAA, PA and PV are removed cleanly while the coronary arteries are intact.

many cases, the PA fiducial points are not inside the image, or are very close to the image borders. Thus, the MSL-based bounding box detector may fail since it relies on image features which are not available outside an image. However, the statistical shape model method can handle this out-of-boundary situations well. In our method, we build a statistical shape model [12] containing nine PA trunk points selected from the PA trunk mesh and the five PA fiducial points: bifurcation, left 1 and 2, right 1 and 2 as shown in Figure 5 (b). When the PA trunk is detected, we extract the nine PA trunk points from the detected mesh. We then use the statistical shape model to estimate the optimal location of the five PA fiducial points given the nine PA trunk points' locations. The statistical shape model can estimate the location of a fiducial point even if it is outside the volume. We select only nine PA trunk points instead of all the mesh points because we want the statistical shape model to capture variations for both the PA trunk and the left/right PA branches in a balanced way. If all the PA trunk



(a) (b) (c)

Fig. 7. LAA removal: (a) With LAA mesh model only: some LCA is cut (blue arrow) while some LAA is not removed (red arrow). (b) First pass of connected component analysis: only the largest connected region in the bounding box of LAA mesh is removed. LCA is intact (blue arrow) however still some small isolated regions of LAA remain (red arrow). (c) Second pass of connected component analysis: run in the whole image and any isolated pieces that are entirely within the LAA bounding box are removed. The result is a clean removal of all LAA voxels.

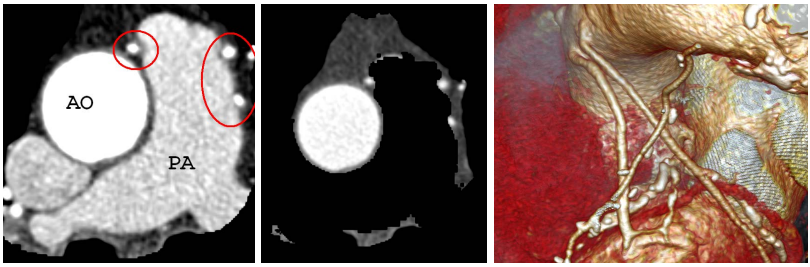
mesh points are included, the statistical shape model will be dominated by the shape variations of the PA trunk, and makes the estimation of the left and right PA less accurate. We found that with nine PA trunk mesh points the algorithm works very well for our purpose. Next, we use the learned MSL detectors to refine each of the five estimated PA fiducial points. The MSL detectors will only search a small neighborhood around the current estimated locations thus it is reliable and fast. If MSL detectors failed because a fiducial point is close to or out of the image border, the statistical shape model result will be used as the final detection result. Otherwise the MSL detector's result will be used.

Pulmonary Vein Model: The PV's shape varies too much to be represented by a single mesh model. Instead, we use two fiducial points defined on the detected left atrium (LA) mesh model to locate the root of the left and the right pulmonary veins. In practice, they are defined as two specified vertices on the LA mesh. The detailed mask for the PV is handled by a region growing method described in the next section.

Left Atrial Appendage Model: We model the LAA using the same mesh model as a heart chamber. The mesh is designed to capture the outer boundary of LAA. However, the LAA's shape varies much more than any heart chambers both for its topology and size. This mesh model usually cannot capture the exact boundary of the LAA. Instead, we only use this model to locate the LAA's bounding box so that the exact boundary can be segmented using the intensity based refinement described in the next section.

4.2 Local Voxel-Based Refinement

As we have stated before, the global shape model usually cannot generate the exact voxel mask for the PA, PV and LAA. A local refinement is necessary



(a) (b) (c)

Fig. 8. Protection of vessels while removing the PA. (a) One case where bypass is deeply embedded in the PA as shown with the red circles. (b) Region growing constrained by vesselness classification can reliably remove any voxels belong to the PA while keep the bypass arteries untouched. Also the aorta is protected by segmentation. (c) 3D visualization of the case, the bypass arteries are intact and clearly visible.

for our heart isolation application. For the PA, PV and LAA, we use different refinement strategies. However, the goals are the same: to find clear boundaries without cutting into any of the CA or bypass.

Pulmonary Artery: For the PA, the global shape model contains two parts: the PA trunk mesh and the five fiducial points. For the mesh, we first close its openings and then mask out any voxels inside the mesh. As shown in Figure 4, usually a thin layer of the PA trunk still remains due to the mesh smoothing. We then use the region growing algorithm to dilate the mask out-ward for 2-3 millimeters. The region growing algorithm's threshold is determined by the mean and standard deviation of the voxels which are already in the mask. With an adaptive threshold, region growing can work for images with or without contrast agent in the PA to successfully remove the thin layer left by the PA trunk mesh. For the left PA, right PA and the PA bifurcation regions, we start region growing from each of the five fiducial points. In this step, the range for region growing is limited to 15 mm since the PA fiducial points are defined as less than 15 mm apart from each other. The region growing from the fiducial points thus can cover all the voxels of the PA bifurcation and the two branches. However, it tends to leak into surrounding objects, especially to nearby bypass arteries or the LCA as it only relies on local information. To prevent such "leaks" and protect the coronary arteries, we use a learning-based vesselness measurement to create a forbidden zone, which will be described later.

Pulmonary Vein: For the PV, we apply the same region growing algorithm from the two root fiducial points of left and right PV as for the PA fiducial points. The intensity threshold is based on the statistics of voxel intensities within the detected LA mesh model. To prevent leakage into nearby structures, we limit the growing range to 25 mm.

Left Atrial Appendage: The LAA is more complex. First, the LAA mesh model only gives an approximate boundary: it may not cover the whole LAA

and it may include some LCA segment or other structures. This is due to the high variation of the LAA's shape. Second, there usually are many small chambers in the LAA which make the LAA not look like a single connected region in the image. To deal with these challenges, we design an algorithm composed with model-based mesh detection and connected component analysis (CCA). The algorithm consists of three steps (as illustrated in Figure 7):

1. The LAA mesh is detected. It gives us an initial estimation of the LAA's location and shape. We then create a bounding box slightly larger than the mesh to make sure we cover the whole LAA regions as the LAA mesh may be smaller than the exact LAA region.
2. The first CCA pass is run *within the bounding box* and the largest connected region is removed. We assume it is the largest chamber of the LAA. However, smaller isolated chambers still remain and they are difficult to be separated from LCA pieces within the bounding box.
3. The second CCA pass is run on the *whole image*. The LCA pieces in the LAA bounding box in this pass should be connected to the whole LCA tree and eventually to the aorta and LV. Thus, they should form a large connected region spanning across the LAA bounding box. The remaining LAA pieces form smaller isolated regions that are *entirely within* the bounding box. We remove all such small regions.

4.3 Chamber and Vessel Protection

Sometimes pieces of the important structures such as the aorta, LA or CA are removed by the leakage of the region growing process because of similar voxel intensities of them to the PA, PV or LAA. To prevent this, we introduce several measures to protect these structures. First, we use the segmentation results of the aorta and the LA to mask them as "not possible to grow." The region growing algorithms for the PV and PA and the connected component analysis for the LAA then will ignore any such regions.

It is more difficult to protect the CA and the bypass arteries since they are small and usually very close to the PA, PV and LAA. Furthermore, we do not have a clean mask of the CA tree as we have for the aorta and the LA. Here, we use a machine-learning based vesselness protection algorithm. As described in [14], the idea is to train a voxel classifier based on image context to tell the probability of the voxel being in a vessel. This algorithm is capable of quickly classifying a voxel to be vessel or not by applying a threshold to the returned vesselness probability. We found that a threshold equal to 75% works well for our purpose. However, there would be a waste of computation power if we classify all voxels in an image. Instead, we confine the classification to only those voxels around the PA trunk, LAA and PV where cutting of the CA or bypass arteries by the region-growing or CCA algorithms could happen.

For arteries around the PA trunk, any voxels within 3 mm to the PA trunk mesh will be classified for vesselness. For regions around the LAA and PV, usually only the LCA may be cut. In order to efficiently identify the LCA region

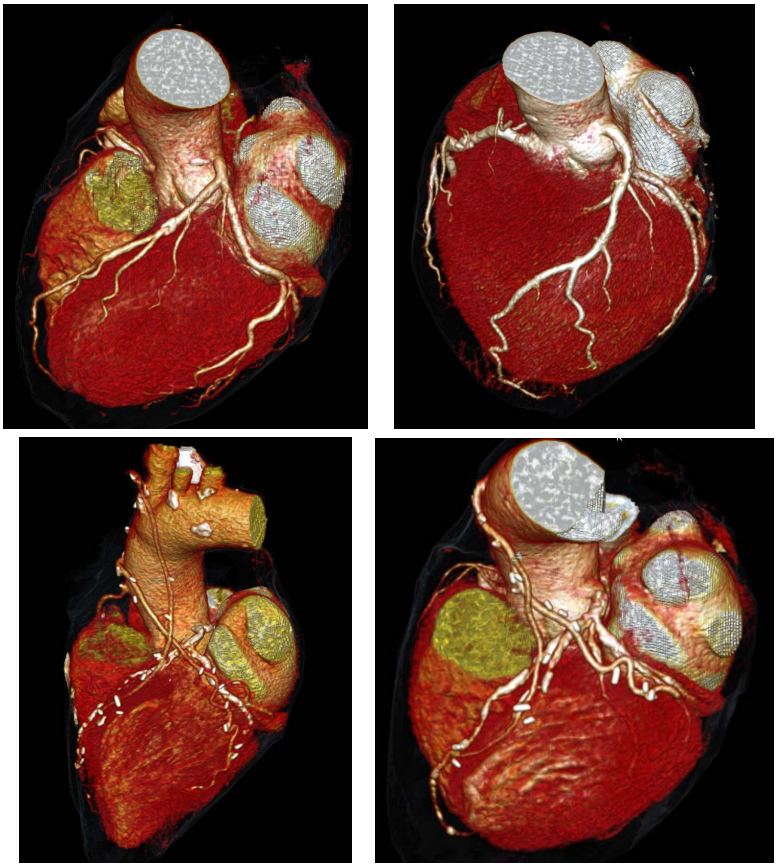


Fig. 9. Heart isolation results on normal (top row) and bypass (bottom row) cases. It shows that our algorithm can reliably remove the PA, PV and LAA while keeping coronary arteries and bypass intact. Such 3D visualization provides a global and intuitive view for physicians to identify suspicious coronary segments.

around LAA and PV, we build a similar fiducial point model as the PA trunk: it contains the left coronary ostium point, the point where left main (LM) coronary artery bifurcated into the left anterior descending artery (LAD) and left circumflex artery (LCX), 20 control points along LCX and 20 selected points from the LA mesh. We then train a statistical shape model for these 42 points based on a manually labeled training database. During the detection, the 20 LA points from the detected LA mesh, the detected left coronary ostium and the bifurcation point (using [15]) are used to estimate the positions of the 20 LCX points based on the learned statistical model. We then run vesselness classification around the region of this estimated LCA control points.

In our test, the vesselness classification in the regions described above takes only 0.02 seconds on a 4-core Xeon 2.53 GHz CPU. After the vesselness classification, we can create a vessel protection mask. This vesselness protection method

Table 2. Subjective score of the heart isolation quality on the test dataset. Score 1 is failed, 2 is acceptable, 3 is good, 4 is very good and 5 is perfect. Our algorithm achieves an average score of 3.73.

Score	1	2	3	4	5
Percentage	0.00%	13.33%	13.33%	60.00%	13.33%

can preserve LCA, RCA and bypass very well in our application, as shown in Figure 8.

5 Evaluation

The goal of the algorithm is to remove most of the LAA, PA and PV so that the coronary arteries and bypass can be clearly seen in 3D visualization. The removal should not touch any coronary arteries or bypass arteries. The algorithm is tested on a database containing 120 cardiac CT images and most of them are bypass cases. The result is then visually examined by experienced testers and a score of 1-5 is given for each case:

1. **Fail:** Major CA cut or bypass cut, important structures removed.
2. **Acceptable:** Large pieces of un-wanted structures are not removed, some minor shave or cut on the CA or the bypass arteries.
3. **Good:** Un-wanted structures are largely removed, no cut on CA or bypass.
4. **Very good:** Only very little of un-wanted structures remained, no cut on CA or bypass.
5. **Perfect:** Clean mask of the heart, no CA or bypass cut.

A score of 3 is thought to be useful, a score of 4 is very good and 5 is perfect. Score 1 is not useful and regarded as failed. Our algorithm's average score is 3.73 and there are no failed cases. The distribution of scores is shown in Table 2. Some examples of our result images are shown in Figure 9. We tested the speed on 80 cardiac CT scans. The size of the scans varies from $512 \times 512 \times 419$ to $512 \times 512 \times 667$ voxels. Resolution of the scans is around $0.4mm \times 0.4mm \times 0.4mm$. The longest processing time is less than 5 seconds on a 2.53 GHz Xeon E5630 CPU.

6 Conclusion

In this chapter, we presented an algorithm that can reliably remove both non-cardiac structures and pulmonary artery, the pulmonary veins and the left atrial appendage for 3D visualization of the coronary arteries from CT. The approach combines global shape models recovered through machine learning techniques with local intensity-based region growing to segment the important anatomical structures. The approach also provides important structural preservation mechanisms to ensure that native or bypass coronary arteries are not cut. The test results demonstrate that this approach can achieve the goal well and this is useful for an efficient CAD/CHD daignosis and treatment planning.

References

1. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Medical Imaging 27(11), 1668–1681 (2008)
2. Lloyd-Jones, D., Adams, R., Carnethon, M., et al.: Heart disease and stroke statistics. Circulation 119(3), 21–181 (2009)
3. Blaha, M., Budoff, M., DeFilippis, A., Blankstein, R., Rivera, J., Agatston, A., O'Leary, D., Lima, J., Blumenthal, R., Nasir, K.: Associations between C-reactive protein, coronary artery calcium, and cardiovascular events: implications for the JUPITER population from MESA, a population-based cohort study. The Lancet 378(9792), 684–692 (2011)
4. Funka-Lea, G., Boykov, Y., Florin, C., Jolly, M.P., Moreau-Gobard, R., Ramaraj, R., Rinck, D.: Automatic heart isolation for CT coronary visualization using graph-cuts. In: Proc. IEEE Int'l Sym. Biomedical Imaging, pp. 614–617 (2006)
5. Moreno, A., Takemura, C.M., Colliot, O., Camara, O., Bloch, I.: Using anatomical knowledge expressed as fuzzy constraints to segment the heart in CT images. Pattern Recognition 41(8), 2525–2540 (2008)
6. van Rikxoort, E.M., Isgum, I., Staring, M., Klein, S., van Ginneken, B.: Adaptive local multi-atlas segmentation: Application to heart segmentation in chest CT scans. In: Proc. of SPIE Medical Imaging (2008)
7. Lelieveldt, B.P.F., van der Geest, R.J., Rezaee, M.R., Bosch, J.G., Reiber, J.H.C.: Anatomical model matching with fuzzy implicit surfaces for segmentation of thoracic volume scans. IEEE Trans. Medical Imaging 18(3), 218–230 (1999)
8. Gregson, P.H.: Automatic segmentation of the heart in 3D MR images. In: Canadian Conf. Electrical and Computer Engineering, pp. 584–587 (1994)
9. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. John Wiley, Chichester (1998)
10. Tu, Z.: Probabilistic boosting-tree: Learning discriminative methods for classification, recognition, and clustering. In: Proc. Int'l Conf. Computer Vision, pp. 1589–1596 (2005)
11. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Sys., Man., Cyber. 9(1), 62–66 (1979)
12. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
13. Taubin, G.: Curve and surface smoothing without shrinkage. In: Proc. Int'l Conf. Computer Vision, pp. 852–857 (1995)
14. Zheng, Y., Loziczonek, M., Georgescu, B., Zhou, S.K., Vega-Higuera, F., Comaniciu., D.: Machine learning based vesselness measurement for coronary artery segmentation in cardiac CT volumes. In: Proc. of SPIE Medical Imaging, pp. 1–12 (2011)
15. Zheng, Y., John, M., Liao, R., Boese, J., Kirschstein, U., Georgescu, B., Zhou, S.K., Kempfert, J., Walther, T., Brockmann, G., Comaniciu, D.: Automatic aorta segmentation and valve landmark detection in C-arm CT: Application to aortic valve implantation. In: Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention, pp. 1–8 (2010)

Randomness and Sparsity Induced Codebook Learning with Application to Cancer Image Classification

Quannan Li[1,2], Cong Yao[2,3], Liwei Wang[2,4], and Zhuowen Tu[1,2]

[1] Lab of Neuro Imaging, University of California, Los Angeles
[2] Microsoft Research Asia
[3] Huazhong University of Science and Technology
[4] The Chinese University of Hong Kong
{quannan.li,yaocong2010,wlwsjtu1989,zhuowen.tu}@gmail.com

Abstract. Codebook learning is one of the central research topics in computer vision and machine learning. In this paper, we propose a new codebook learning algorithm, Randomized Forest Sparse Coding (RFSC), by harvesting the following three concepts: (1) ensemble learning, (2) divide-and-conquer, and (3) sparse coding. Given a set of training data, a randomized tree can be used to perform data partition (divide-and-conquer); after a tree is built, a number of bases are learned from the data within each leaf node for a sparse representation (subspace learning via sparse coding); multiple trees with diversities are trained (ensemble), and the collection of bases of these trees constitute the codebook. These three concepts in our codebook learning algorithm have the same target but with different emphasis: subspace learning via sparse coding makes a compact representation, and reduces the information loss; the divide-and-conquer process efficiently obtains the local data clusters; an ensemble of diverse trees provides additional robustness. We have conducted classification experiments on cancer images as well as a variety of natural image datasets and the experiment results demonstrate the efficiency and effectiveness of the proposed method.

Keywords: Sparsity, Randomness, Codebook Learning, Cancer Image Classification.

1 Introduction

A large number of applications in machine learning, medical image classification, and computer vision deals with the fundamental representation problem where the data are high-dimensional and live in complex manifolds. With their intrinsic and mathematical properties gradually unfolded, research in three general directions has led to significant progress on classification, recognition, and compression: (1) ensemble learning, (2) divide-and-conquer, and (3) sparse coding. More specifically, four concepts have emerged as being essential to the three directions: (1) voting, (2) randomizing, (3) partitioning, and (4) sparsity.

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 181–193, 2013.

Ensemble learning approaches such as bagging [2], boosting [11], and random forests [3] have shown to be among the best choices for classifiers [6,5]. The randomness in the data and feature selection stage leads to robustness in classification, as shown in the random forests [3] where trees are learned from randomly drawn subsets with the splitting criterion being locally optimal on some features randomly chosen. In Extremely Randomized Trees [14] and Random Projection Trees [7], the full data sets are used as the randomization in both feature/basis and threshold selection can provide sufficient diversities.

As real data are of high dimension and they typically do not live in a well-regularized space, the Gaussian type distribution leads to limited representational power [26] and a divide-and-conquer strategy is more appropriate. In machine learning, decision tree [23] is a standard approach where training data are recursively partitioned into subsets. The random projection tree [7] also has recursive data partition based on randomly generated bases.

More recently, sparse representations such as compressed sensing [4] and LASSO [25] have gained a great deal of popularity. One message emerging from sparse representation is that high-dimensional data within intrinsic lower dimension can be well represented by sparse samples of high dimension. The robustness of the sparse representation often assumes a subspace of certain regularity, e.g. well-aligned data [29].

In this paper, we tackle the problem of codebook learning for high dimensional visual data. Inspired by the above observations, we propose a randomized forest sparse coding (RFSC) method. Given a large set of visual data, we train an ensemble of random splitting/projection trees (when we are not sure about the form of the whole data population, it is desirable to perform random partition with certain local optimality); for each leaf node in the tree, we learn a set of bases to best represent the data with sparse coefficients. The overall codebook is a collection of all the bases from all the tree leaves. RFSC carries the ideas of voting, randomizing, partitioning, and sparse coding in a natural way. Its applicable to applications such as Modern cancer diagnosis, which largely benefits from high resolution histopathology images providing distinctive and reliable cues for discriminating abnormal tissues from normal ones.

2 Related Works

As we have discussed, our approach is inspired by the literature in ensemble learning [2,11,3], divide-and conquer approaches [23,14,7], and sparse representation [4,25,29,19]. Two types of work are particularly related to our approach: tree based splitting/projection methods, e.g., Extremely Randomized Trees [14] and Random Projection Trees [7], and sparse coding based codebook learning techniques [30,15,13].

Extremely Randomized Tree (ERT) [14] is a variant of random forest. ERTs randomize both the feature selection and the quantization threshold searching process, making the trees less correlated. When used for visual codebook learning (ERC-Forest) in [20], the generated trees are not treated as an ensemble of decision trees, instead, they are referred to as an ensemble of hierarchical spatial

partitioners. The samples (image patches) in each leaf node are assumed to form a small cluster in the feature space. The leaves in the forest are uniquely indexed and serve as the codes for the codebook. When a query sample reaches a leaf node, the index of that leaf is assigned to the query sample. A histogram is formed by accumulating the indices of the leaf nodes, which serves as a Bag of Words (BOA) representation. Similar to ERC-Forest, [24] introduces a semantic texton forest using ERT to perform image classification and segmentation.

Random Projection Tree [7] is a variant of k-d tree. The k-d tree splits the data set along one coordinate at the median and recursively builds the tree. Though widely used for spatial partitioning, it suffers from the curse of dimensionality problem. Based on the realization that, high dimensional data often lies on low-dimensional manifold, RPT splits the samples into two roughly balanced sets according to a randomly generated direction. This randomly generated direction approximates the principal component direction, and can adapt to the low dimensional manifold. The RPT naturally leads to tree-based vector quantization and an ensemble of RPTrees can be used as a codebook.

We use Extremely Randomized Trees/Random Projection Trees to partition the samples. But instead of splitting the samples till we cannot split any more, we stop early according to certain criterion and find some bases that can best reconstruct all the samples in that node. These bases serve as codes of the codebook.

There are already some methods using sparse coding for codebook learning. In [30], the authors generalize vector quantization to sparse coding, and construct the histogram using multi-scale spatial max pooling. Each patch can be assigned to several (sparse) codes, and thus the reconstruction error can be reduced. Also, this method extends the Spatial Pyramid Matching method [15] to a linear SPM kernel. In [13], Laplace sparse coding preserves the consistency in the sparse representation and alleviates the problem in [30] that similar patches may be assigned to different codes. In [28], a locality-constrained linear coding scheme is proposed that utilizes the locality constraints to project descriptors to their local-coordinate system. This scheme can preserve the property of local smooth sparsity. Compared with these methods, the advantages of RFSC is obvious. One advantage is the efficiency. Utilizing techniques such as ERT and RPT, the sparse coding is performed only in subspaces and the computational burden is greatly reduced. The second advantage is the potential promotion of the discriminative ability. The label information can easily be used into the tree splitting process (ERT) and the codebook created could have more discriminative power.

3 Randomized Forest Sparse Coding

3.1 Problem Formulation

Suppose we are given a set of training data $S = \{\mathbf{x}_i\}_{i=1}^n$ and $\mathbf{x}_i \in \mathbb{R}^D$ (in a supervised setting, each $\mathbf{x}_i$ is also associate with a label $y_i \in \mathcal{Y} = \{0, ..., K\}$ and thus $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$), our goal is to learn a codebook (a set of bases) $\mathbf{B} = \{\mathbf{b}_i\}_{i=1}^m$ and $\mathbf{b}_i \in \mathbb{R}^D$ such that

$$\min_{\mathbf{B},\mathbf{w}} \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{j=1}^{m} w_{ij}\mathbf{b}_j \right\|_2^2$$
$$s.t.\ \ \forall i, \sum_j |w_{ij}| \leq \tau \tag{1}$$

The first term in Eqn. (1) minimizes the reconstruction error and the second term gives the sparsity constraints on the reconstruction coefficients. Eqn. (1) actually includes two coupled optimization problems: (1) given $\mathbf{w}$, find the optimal codebook $\mathbf{B}$; (2) given a codebook $\mathbf{B}$, find the best reconstruction coefficients $\mathbf{w}$. A similar formulation appears in [30]. After an optimal basis set $\mathbf{B}^*$ is found, for a new sample $\mathbf{x}$, we can compute its reconstruction coefficients $\mathbf{w}$ via:

$$\min_{\mathbf{w}} \left\| \mathbf{x} - \sum_{j=1}^{m} w_j\mathbf{b}_j \right\|_2^2$$
$$s.t.\ \ \sum_j |w_j| \leq \tau \tag{2}$$

The vector $\mathbf{w}$ can be used to characterize the sample $\mathbf{x}$. In codebook learning, each $\mathbf{b}_j$ serves as a code, and the reconstruction coefficients with respect to the codes are pooled to form a histogram.

In Eqn. (1), the norm of $\mathbf{b}_j$ can be arbitrarily large, making w_{ij} arbitrarily small. Further constraints should be made on $\mathbf{b}_j$. In our paper, we make a reasonable constraint that all the bases in the codebook should be from the training set S, i.e., $\mathbf{B} \subset S$. With this constraint, Eqn. (1) can be transformed into

$$\min_{\mathbf{v},\mathbf{w}} \sum_{i=1}^{n} \left\| \mathbf{x}_i - \sum_{j=1}^{n} w_{ij}v_j\mathbf{x}_j \right\|_2^2 \tag{3}$$
$$s.t.\ \ \sum_j v_j \leq m, v_j \in \{0,1\}$$
$$\forall i,\ \sum_j |w_{ij}| \leq \tau \tag{4}$$

Here, v_j serves as an indicator value $\in \{0,1\}$ and $\mathbf{B} = \{\mathbf{x}_j : \mathbf{x}_j \in S,\ v_j = 1\}$. Eqn. (3) is seemingly more complex than Eqn. (1) with the introduction of $\mathbf{v}$. However, it can be solved more efficiently since the search space for the bases is greatly reduced.

Learning a codebook of size greater than e.g. $5,000$ from tens of thousands of samples is computationally demanding. As motivated before, we could perform a divide-and-conquer strategy to partition the data space with complex manifolds into local subspaces. Within a subspace, it is then much more efficient to learn bases for a sparse representation.

3.2 Randomized Forest Data Partition

In this section, we take the Extremely Randomized Tree (ERT) [14] and Random Projection Trees (RPT) as examples to illustrate the data projection process. Both ERT and RPT partition the samples recursively in a top-down manner. ERT adopts the label information and uses normalized Shannon entropy as the criterion to select features. RPT is unsupervised and it does not need any label information; it splits the data via hyper-planes normal to the randomly generated projection bases.

Discriminative Partition via Extremely Randomized Tree. Given a labeled sample set $S = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, ERT proceeds by randomly selecting a subset of features from the feature pool $\{f_i, 1 \leq i \leq D\}$. For each selected feature f_i, a threshold θ_i is sampled according to a uniform distribution. Based on the features selected and thresholds sampled, boolean tests $\{T_i : \mathbf{x}(i) < \theta_i\}$ can be used to split the set S. If T_i = true, $\mathbf{x}$ goes to the left branch S_1; otherwise, $\mathbf{x}$ goes to the right branch S_2.

To select the best boolean test for splitting, the normalized Shannon entropy was used:

$$Score(S, T_i) = \frac{2 \cdot I_{\mathcal{Y},T_i}(S)}{H_{\mathcal{Y}}(S) + H_{T_i}(S)} \tag{5}$$

where, $I_{\mathcal{Y},T_i}(S) = H_{\mathcal{Y}}(S) - \sum_{p=1}^{2} \frac{n_p}{n} H_{\mathcal{Y}}(S_p)$. $I_{\mathcal{Y},T_i}(S)$ is the information gain; $H_{\mathcal{Y}}(S) = -\sum_{y \in \mathcal{Y}} \frac{n_y}{n} \log_2(\frac{n_y}{n})$ denoting the entropy of class distribution of the original set S. $H_{T_i}(S) = -\sum_{p=1}^{2} \frac{n_p}{n} \log_2(\frac{n_p}{n})$ denotes the entropy for the test T_i that splits the data into two branches. The T_i with the largest $Score(S, T_i)$ is selected.

The use of $H_{T_i}(S)$ as a normalization term in Eqn. (5) will favor uneven splitting, making the forest more unbalanced. In our randomized forest sparse coding scheme (RFSC), it is desirable to have balanced trees, so we use a slightly modified form of Eqn. (5):

$$Score(S, T_i) = \frac{2 \cdot I_{\mathcal{Y},T_i}(S)}{H_{\mathcal{Y}}(S) + 1 - H_{T_i}(S)} \tag{6}$$

Since $H_{T_i}(S)$ is a concave function and it achieves the maximum value 1 when the numbers of samples in S_1 and S_2 are the same, this criterion can make the trees more balanced.

Unsupervised Splitting via Random Projection Tree (RPT). At each node, RPT chooses a random unit projection direction $\mathbf{b} \in \mathbb{R}^D$, and splits the samples into two roughly equal-sized sets. The random projection and thresholding also serve as a type of boolean test. We use the splitting criterion as

$$T := \mathbf{x}^T \mathbf{b} \leq \left(\text{median}(\mathbf{z}^T \mathbf{b} : \mathbf{z} \in S) + \delta\right).$$

Here δ is a random perturbation that adapts to the structure of S. Splitting around the median value makes the splitting balanced while the perturbation δ introduces certain randomness [7].

Since RPTs can automatically adapt to the low dimensional manifold of the dataset S, the samples in the leaf nodes observe local subspaces. The local structures of all the leaf nodes thus collectively comprise the global structure of the data set S (Fig. 1 (b)).

Basis Pursuit at the Leaf Nodes. Both ERT and RPT build the trees to the fine scale and use the leaf nodes as the codes. Instead of building the trees of very deep level, RFSC stops at some relatively higher level (e.g., when the number of samples is less than M). At such nodes, the local manifold structure is assumed

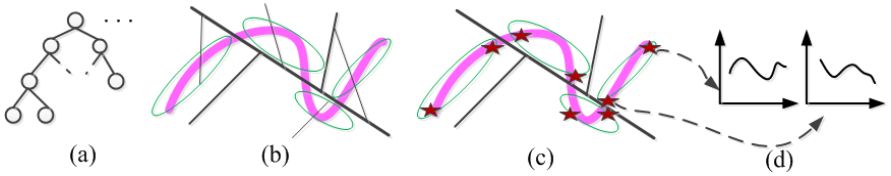


Fig. 1. Illustration of the idea of RFSC using Random Projection Tree (best viewed in color). (a) The forest consists of ensemble of random projection trees; (b) The spatial partition of the dataset by one tree (A copy from [10]). A cell stands for a leaf node. The width of the separation line indicates the level of the tree. (c) For RFSC, it does not build the tree to fine level. At certain level when local manifold structures are found, bases (indicated by the red stars) are learned for the local structure in each cell. (d) For the samples in each cell, their reconstruction coefficients with respect to the bases are different.

to be relatively simple and regularized. RFSC seeks a set of bases to sparsely represent the subspaces at those nodes. This process can be illustrated using Random Projection Tree in Fig. 1 in which a visualization is displayed and RPT tends to split the data along the principal component direction (Fig. 1 (b)). For RFSC, when the local structure is relatively regularized, it seeks some bases (the red stars) to sparsely represent the local subspace. Different from RPT or ERT that use the mean of the local subspace or a single index to represent the cell, the information conveyed via the reconstruction coefficients with respect to each basis (Fig. 1 (d)) is richer and more informative. Note that the bases in different clusters could be spatially close to each other. As an illustration, see the two bases on the bottom right in Fig. 1(c). From this point of view, the number of bases and the redundancy would increase. However, according to Theorem 1 in the justification part, the total number of bases in all the leaf nodes is bounded. Since at each node when the splitting process stops, there are generally $80 \sim 200$ samples (depending on the codebook size) and $3 \sim 10$ bases, the computational overhead of subspace learning is not significant compared with directly pursuing bases from the entire sample set.

3.3 Optimization Scheme

The constraint that $v_j \in \{0, 1\}$ makes Eqn. (3) a hard problem. In this subsection, we present two schemes to solve this optimization problem. The first one is to relax v_j to a real value and use coordinate descent algorithm to optimize on $\mathbf{w}$ and $\mathbf{v}$ iteratively. The second one is a greedy pursuit approach that selects the bases one by one.

Convex Relaxation. The first optimization scheme is to relax the values of v_j to real numbers and use ℓ^1 constraint $\sum_j |v_j| \leq m$ instead of ℓ^0 like constraint in Eqn. (3). Putting this constraint as a regularization term, we can transform this problem into an equivalent form:

$$\frac{1}{2}\sum_{i=1}^{n}\left\|\mathbf{x}_i-\sum_{j=1}^{n}w_{ij}v_j\mathbf{x}_j\right\|_2^2+\lambda_1\sum_{i,j}|w_{ij}|+\lambda_2\sum_j|v_j| \tag{7}$$

Here, $v_j \in \mathbb{R}$. λ_1 and λ_2 are regularization parameters that make the trade-offs between the residue and the norms of the weight vectors.

There are two sets of variables $\mathbf{w}$ and $\mathbf{v}$ in Eqn. (7). To optimize Eqn. (7), we adopt an EM-like algorithm that iterates by fixing one set of variables and optimize on the other set using coordinate descent algorithm [12].

Greedy Pursuit Approach. Starting from an empty basis collection, the greedy pursuit approach selects the bases one by one. Suppose some l samples $\mathbf{B}_l = \{\mathbf{x}_{s_i}, 0 \le i \le l, 1 \le s_i \le n\}$ have been selected from the n samples, i.e., $v_{s_i} = 1$. To select the $(l+1)$th basis, we optimize the following function:

$$\begin{aligned} s_{l+1} = \min_{k\notin\{s_i\}}\frac{1}{2}\sum_{i=1}^{n}\left\|\mathbf{x}_i-\sum_{j\in\{s_i\}}w_{ij}\mathbf{x}_j-w_{ik}\mathbf{x}_k\right\|_2^2 \\ +\lambda_1\sum_{i=1}^{n}\sum_{j\in\{s_i\}}|w_{ij}|+\lambda_1\sum_{i=1}^{n}|w_{ik}| \end{aligned} \tag{8}$$

According to Eqn. (8), the sample that reconstructs all the n patches together with the first l selected bases is selected as the $(l+1)$th basis.

The greedy approach finds suboptimal solution to Eqn. (3). But it's more efficient than the convex relaxation approach, and in practice, we find that its performance is comparable with the convex relaxed solution. Thus in some of our experiments, we only use this greedy approach.

3.4 Theoretical Justification

In this section, we give some theoretical justification to our approach. Our intuition is to show that the three steps in randomized forest sparse coding: (1) ensemble of trees, (2) randomized projection tree, and (3) sparse coding leads to the same complexity level in the number of bases as to the original data.

Given $S = \{\mathbf{x}_i, i = 1..n\}$ with $\mathbf{x}_i \in \mathbb{R}^D$, assume that $\mathbf{x}_i$ lives in the intrinsic lower dimension $d \ll D$. It can be seen that the number of bases needed to reconstruct S is bounded. Following the definition of Assouad dimension [1] [7]:

Definition: For any point $\mathbf{x} \in \mathbb{R}^D$ and $r > 0$, let $B(\mathbf{x}, r) = \{||\mathbf{x}-\mathbf{z}|| \le r\}$ denote the closed ball of radius r centered at $\mathbf{x}$. The Assouad dimension of $S \in \mathbb{R}^D$ is the smallest integer d such that for any ball $B(\mathbf{x}, r) \in \mathbb{R}^D$, the set $B(\mathbf{x}, r) \cap S$ can be covered by 2^d balls of radius $r/2$.

Theorem 1. *The number of bases needed to reconstruct S by Randomized Forest Sparse Coding (RFSC) is $O(2^{d\log d})$.*

Proof:

Fixing radius r, suppose we want to create a codebook such that each basis function covers $r/2$, a size of $O(2^d)$ codebook is required to cover the entire dataset S, according to the definition of Assouad dimension.

The main result in [7] shows that $O(d \log d)$ levels of a random projection/partition tree would reach cells with radius $r/2$. Therefore, the number of cells is $O(2^{d \log d})$. Suppose there are k trees in the forest, and in each leaf node, l bases are found, then the number of the bases becomes $O(kl2^{d \log d})$. As k and l are generally small and can be kept constant, the bound still reduces to:

$$O(2^{d \log d}).$$

Although RFSC slightly increases the size of the codebook compared to $O(2^d)$, since d is generally small ($d \ll D$), this is reasonably bounded.

4 Experiments

To evaluate the effectiveness of the proposed codebook learning algorithm, we conducted extensive classification experiments on a collection of cancer images and a variety of natural image datasets: Graz-02 image set, and the PASCAL 2005 image set.

As the baselines, we obtained the source code for ERC-Forest from the authors of [20] and implemented the RPTs according to [7]. In our experiments, the feature vectors are used without any normalization, which is sometimes done in subspace learning and sparse coding (we found that performing normalization does not affect the overall performance in the experiments reported here). For each leaf node, 5 bases are learned. For the Graz-02 image set, $\lambda_1 = 2$ and $\lambda_2 = 6$, while for the PASCAL 2005 image set, $\lambda_1 = 15$ and $\lambda_2 = 6$. To solve the subspace learning problem via sparse coding defined in Eqn. (3), 10 iterations between $\mathbf{w}$ and $\mathbf{v}$ are enough to find a good sparse solution.

In the following, we use RFSC to denote subspace learning via sparse coding under Extremely Randomized Trees; RPT-SC denotes subspace learning on Random Projection Trees. For RFSC and RPT-SC, the postfix "-Cvx" refers to using the convex relaxation version and "-Gdy" regards to using the greedy basis pursuit version. For the classification task of Cancer Images, the performance is measured using the Area under the curve of the ROC curve, while for natural image classification, the performance is measured using the classification accuracy at the equal error rate and the reported accuracies are the averages of 10 rounds of execution.

4.1 Experiments on Cancer Images

Dataset: We used a histopathology image data set with 60 colon images (30 cancer images and 30 non-cancer images). Sample images are shown in Fig. 2. The images are labeled as cancer or non-cancer by two pathologists independently. If disagreement happens for a certain image, these two pathologists together with a third senior pathologist will carefully examine and discuss until final agreement.

Experimental Setup: Before feature extraction, the original images are downsampled with a factor of 2. As no obvious spatial regularities are observed from the images (Fig. 2), instead of using the Bag-of-Features (BOF) vectors, we

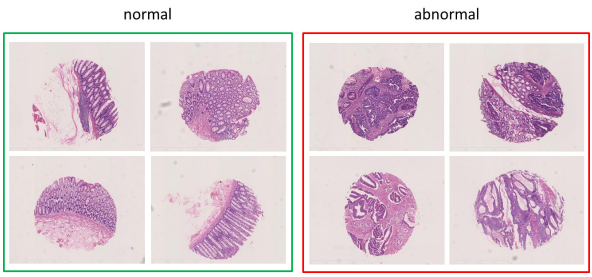


Fig. 2. Cancer image examples. The images in the green box are normal samples. i.e. there are no cancerous cells. The images in the red box are abnormal samples, i.e. there are cancerous cells.

randomly sample $N = 200$ local patches (32×32) for each image. Each patch is represented by Lab color histogram, Local Binary Pattern [21], and SIFT [18]. For the proposed method, each patch is encoded by the proposed coding schemes RFSC or RPT-SC; for the baseline, we use the raw feature. Random Forests [3] is adopted as the strong classifier for its simplicity and high performance. The overall classification score of an image is the mean of the scores of all the patches. Half of the images in the dataset are chosen randomly for training and the rest for testing. We run the experiments 5 times for each method and report the averaged performance. For RFSC and RPT-SC, the convex relaxation versions are used. The Area Under Curve (AUC) for the methods are RPT-SC 0.98, RFSC 0.987, RPT 0.927, ERC-Forest 0.95, and raw feature 0.967 respectively; our method performs better than the alternatives.

4.2 Experiments on Natural Images

The reconstruction coefficients are pooled in the natural image classification task. To pool the reconstruction coefficients, unless otherwise stated, max-pooling is used as in [30]. The reconstruction coefficients of the trees are pooled and concatenated to form a histogram leaving the voting process till the classification step; Linear SVM is used in the image classification stage. In all the following image classification experiments, we do not include the adaptive saliency map process. This makes the image classification performance of ERC-Forest slightly worse than that reported in [20]. However, this performance degeneration is reasonable and in accordance with the case illustration in [20].

GRAZ-02 Dataset [22]. GRAZ-02 image set consists of three object categories and one counter-category. For each category, the categorization task is to distinguish the object category from the counter-category, None. Similar to [20], we also pick the two hardest cases: Cars vs. None and Bikes vs. None.

To make a direct comparison with [20] and [22], we conduct the experiment according to the setting in [20] using the first 300 images of each category for training. We use the greedy version of RFSC and vary the codebook size from

5000 to 9000. From Table 1 and Table 2, we observe that, RFSC-Gdy performs better than ERC-Forest and the method in [22]. Though without the adaptive saliency map process, the accuracy (83.9%) of RFSC-Gdy on the case of Bikes vs. None approaches that reported in [20] (84.4%).

Table 1. Comparison of the accuracy on the case of Cars vs. None in the GRAZ-02 images [22]

size of codebook	5000	6000	7000	8000	9000
[22]	70.5%				
ERC-Forest	71.3%	73.5%	74.5%	74.7%	74.8%
RPT	66.5%	66.6%	65.3%	67.7%	66.9%
RFSC-Gdy	73.4%	74.3%	**75.7%**	74.9%	74.3%
RPT-SC-Gdy	68%	69.8%	69%	69.5%	68.2%

Table 2. Comparison of the accuracy for Bikes vs. None in the GRAZ-02 images [22]

size of codebook	5000	6000	7000	8000	9000
[22]	77.8%				
ERC-Forest	78.8%	78%	78.5%	78.5%	78.5%
RPT	73.3%	74.3%	74.1%	75.1%	74.4%
RFSC-Gdy	80.7%	**83.9%**	80.8%	81.3%	80%
RPT-SC-Gdy	76.5%	76.8%	76.1%	76.7%	76%

We also conduct the experiments using all the images instead of the first 300 images. Average-pooling is adopted here and the results are reported in Table 3 and Table 4. The performance of the two optimization schemes is similar. RFSC-Cvx-1tree refers to using one randomized tree instead of the forest, an ensemble of trees. It performs worse than RFSC. This justifies the benefit of using ensembles.

Table 3. Comparison of the accuracy using all the images for Cars vs. None in the GRAZ-02 images [22]

size of codebook	5000	6000	7000	8000	9000
ERC-Forest	67.2%	67%	68.6%	68.8%	71.3%
Leaf-Kmeans	68.2%	70.9%	73%	72.6%	73.2%
RFSC-Cvx-1tree	72.6%	72.2%	71.4%	75%	75%
RFSC-Cvx	75%	75%	73.7%	73.1%	75.2%
RFSC-Gdy	74.3%	75.5%	74.5%	74.8%	**75.5%**

We do not compare RFSC and RPT-SC with directly performing dictionary learning on the image classification task since solving Eqn. (1) directly when $m = 5,000$ or $9,000$ is time consuming. However, benefiting from the divide-and-conquer process, it takes less than 1 hour for RFSC and RPT-SC to build a

Table 4. Comparison of the accuracy using all the images for Bikes vs. None in the GRAZ-02 images [22]

size of codebook	5000	6000	7000	8000	9000
ERC-Forest	77.8%	78.3%	78.3%	79.1%	78.8%
Leaf-Kmeans	75.1%	74.4%	79.7%	78.7%	79.5%
RFSC-Cvx-1tree	77.8%	78.2%	78.6%	79.5%	79.5%
RFSC-Cvx	80%	82.2%	**82.6%**	81.4%	81.8%
RFSC-Gdy	81.5%	80.3%	81.5%	80.8%	80.9%

forest with 5 trees and $9,000$ codes. This improvement in efficiency stems from seeking a small amount of bases from hundreds of patches instead of seeking thousands of bases from tens of thousands of training patches. Other efficient algorithms such as [16] can be used to solve Eqn. (1), but the conclusion of the improvement in efficiency induced by the divide-and-conquer process still holds. RFSC and RPT-SC are also very efficient at the testing stage. It takes about 0.5 second to process an image and pooling the reconstruction coefficients. As a comparison, it would take around 30 seconds for K-Means to assign patches to the codes for an image when the feature vector is of dimension 768 and the codebook size K is $5,000$.

PASCAL 2005 Image Set [8]. We also compare our method with ERC-Forest on PASCAL 2005 image set. The results are shown in Table 5. RFSC-Gdy achieves better results on all of the 4 categories than ERC-Forest.

Table 5. Comparison of the accuracy on PASCAL 2005 image set [8]

method	motobikes	cars	bikes	person
ERC-Forest	96%	95%	89%	90.9%
RFSC-Gdy	**96.4%**	**95.3%**	**90.6%**	**91.4%**

5 Conclusion

In this paper, we have introduced a codebook learning method called randomized forest sparse coding that integrates three concepts: ensemble, divide-and-conquer and sparse coding. Justifications for the effectiveness and efficiency of our method are also provided. The proposed scheme is applied to both the Cancer Image Classification and natural image classification and observes significant improvement in performance.

Acknowledgment. This work is supported by Office of Naval Research Award N000140910099 and NSF CAREER award IIS-0844566.

References

1. Assouad, P.: Plongements lipschitziens dans rn. Bull. Soc. Math. France (4), 429–448 (1983)
2. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)
3. Breiman, L.: Random forests. Machine Learning 45(1), 5–32 (2001)
4. Candes, E., Tao, T.: Near-optimal signal recovery from random projections: universal encoding strategies. IEEE Trans. Inform. Theory 52(2), 5406–5425 (2005)
5. Caruana, R., Karampatziakis, N., Yessenalina, A.: An empirical evaluation of supervised learning in high dimensions. In: ICML, pp. 96–103 (2008)
6. Caruana, R., Niculescu-Mizil, A.: An empirical comparison of supervised learning algorithms. In: ICML, pp. 161–168 (2006)
7. Dasgupta, S., Freund, Y.: Random projection trees and low dimensional manifolds. In: STOC, pp. 537–546 (2008)
8. Everingham, M., Zisserman, A., Williams, C.K.I., Van Gool, L., Allan, M., Bishop, C.M., Chapelle, O., Dalal, N., Deselaers, T., Dorkó, G., Duffner, S., Eichhorn, J., Farquhar, J.D.R., Fritz, M., Garcia, C., Griffiths, T., Jurie, F., Keysers, D., Koskela, M., Laaksonen, J., Larlus, D., Leibe, B., Meng, H., Ney, H., Schiele, B., Schmid, C., Seemann, E., Shawe-Taylor, J., Storkey, A.J., Szedmak, S., Triggs, B., Ulusoy, I., Viitaniemi, V., Zhang, J.: The 2005 PASCAL Visual Object Classes Challenge. In: Quiñonero-Candela, J., Dagan, I., Magnini, B., d'Alché-Buc, F. (eds.) MLCW 2005. LNCS (LNAI), vol. 3944, pp. 117–176. Springer, Heidelberg (2006)
9. Ferrari, V., Jurie, F., Schmid, C.: Accurate Object Detection with Deformable Shape Models Learnt from Images. In: CVPR (2007)
10. Freund, Y., Dasgupta, S., Kabra, M., Verma, N.: Learning the structure of manifolds using random projections. In: NIPS, vol. 20 (2007)
11. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. J. of Comp. and Sys. Sci. 55(1) (1997)
12. Friedma, J., Hastie, T., Hofling, H., Tibshirani, R.: Pathwise Coordinate Optimization. The Annals of Applied Stat. (2007)
13. Gao, S., Tsang, I.W.H., Chia, L.T., Zhao, P.: Local features are not lonely - laplacian sparse coding for image classification. In: CVPR (2010)
14. June, P.G., Ernst, D., Wehenkel, L.: Extremely Randomized Trees. In: Machine Learning, vol. 36 (2003)
15. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: CVPR (2006)
16. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: NIPS (2007)
17. Li, Y., Osher, S.: Coordinate descent optimization for ℓ^1 minimization with application to compressed sensing; a greedy algorithm. CAM Report (2009)
18. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
19. Mairal, J., Bach, F., Ponce, J.: Task-driven dictionary learning. IEEE Trans. on PAMI (to appear)
20. Moosmann, F., Nowak, E., Jurie, F.: Randomized clustering forests for image classification. IEEE Trans. on PAMI 30(9), 1632–1646 (2008)
21. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24(7), 971–987 (2002)

22. Opelt, A., Pinz, A., Fussenegger, M., Auer, P.: Generic Object Recognition with Boosting. IEEE Trans. on PAMI 28(3), 416–431 (2006)
23. Quinlan, J.R.: Induction of decision trees. Machine Learning 1 (1986)
24. Shotton, J., Johnson, M., Cipolla, R.: Semantic texton forests for image categorization and segmentation. In: CVPR (2008)
25. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Royal. Statist. Soc B. 56(1), 267–288 (1996)
26. Turk, M.: Eigenface for recognition. Journal of Cognitive Neuroscience (1991)
27. Vedaldi, A., Fulkerson, B.: Vlfeat: an open and portable library of computer vision algorithms. In: ACM Multimedia, pp. 1469–1472 (2010)
28. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: CVPR (2010)
29. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. on PAMI 31(2) (2009)
30. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: CVPR (2009)

Context Enhanced Graphical Model for Object Localization in Medical Images

Yang Song[1], Weidong Cai[1], Heng Huang[2], Yue Wang[3], and David Dagan Feng[1]

[1] BMIT Research Group, School of IT, University of Sydney, Australia
[2] Computer Science and Engineering, University of Texas at Arlington
[3] Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University

Abstract. Object localization is an important step common to many different medical applications. In this Chapter, we will review the challenges and recent approaches tackling this problem, and focus on the work by Song et.al. [20]. In [20], a new graphical model with additional contrast and interest-region potentials is designed, encoding the higher-order contextual information between regions, on the global and structural levels. A discriminative sparse-coding based interest-region detector is also integrated as one of the context prior in the graphical model. This object localization method is generally applicable to different medical imaging applications, in which the objects can be distinguished from the background mainly based on feature differences. Successful applications on two different medical imaging applications – lesion dissimilarity on thoracic PET-CT images and cell segmentation on microscopic images – are demonstrated in the experimental results.

1 Introduction

A wide variety of medical applications comprise object localization as an important step for discovering the anatomical or pathological information from images. For example, region-of-interest (ROI) detection is helpful for early screening of diseases; and lesion segmentation is useful for treatment planning. We consider object localization as a generalization of both detection and segmentation, with both automatic identification of ROI, and a good delineation of its boundary.

We focus on medical imaging problems in which objects can be localized based on local-level features and feature differences between the objects and background. For example, in positron emission tomography – computed tomography (PET-CT) images, abnormalities typically show higher uptakes than normal tissues. In fluorescence microscopic images, the cell nuclei normally depict darker colors then the other cell structures and the background. In brain magnetic resonance imaging (MRI), the white and gray matter display quite different intensities and spatial patterns.

Local features are usually not sufficient for a good localization, because of large inter-subject variations causing same anatomical structures appearing quite

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 194–205, 2013.

differently across images. The problem is further complicated due to low feature differences between different tissue types and especially for the boundary areas between the objects and background. In addition, pathologies often lead to larger imaging variations, and an accurate object localization is thus more challenging.

For such imaging problems, while lots of work have been reported [25,16,19,5,18,4,21,15], they are mostly designed to be domain specific; and often rely on sophisticated feature sets, which can be computational-intensive and difficult to adapt to other imaging problems. Furthermore, because such features are usually designed based on domain knowledge and empirical studies, their effectiveness may be restricted to the limited scenarios available in the datasets.

Therefore, in [20], we proposed an object localization method that can be generally applicable, requires simpler feature sets, and addresses low feature differences and large inter-subject variations. With region-based labeling, each image region is classified as the object or background to produce the localization output. In summary, our main contributions are the following: (i) the discriminative capability of the basic conditional random field (CRF) is enhanced with two contextual priors, namely the contrast and interest-region potentials, to encode the global contrast information and region-based feature similarities, for improving the boundary delineations; (ii) a sparse-coding classification method is proposed for interest-region detection, with improved discriminative power of the learned dictionaries; and (iii) the design is kept general with simple feature sets configurable for the specific application, and has been successfully applied to both lesion dissimilarity on thoracic PET-CT images and cell segmentation on microscopic images.

Related Work. We focus our review on CRF-based localization methods in both medical and general imaging domains. As an undirected graphical model, CRF is now one of the most successful trends in object class image segmentation [6]. The basic and most commonly used formulation is to have local features represented as graph nodes and consistency constraints between neighboring regions as edge connections [17]. However, comparing to the non-graphical discriminative approach, generally such models add advantages little more than spatial smoothing of labelings [25].

Higher-level features, i.e. contexts in images, are often acknowledged as important discriminative factors [6,4]. In particular, relationship information on a larger scale, such as those across image slices [8], relating to reference objects [2], or between distant image regions [7], can be modeled as pairwise connections to encourage labeling consistency or enhance the discriminative power of local features. Such ideas of connecting beyond immediate neighbors are inspiring; however, choosing the related pairs and describing their interactions are rather application specific. To explore multi-scale region interactions, hierarchical models have been proposed [11,3]; however, they are normally created based on region clustering, without considering the actual object structures. At a more structural level, object detectors with bounding box outputs have been incorporated into CRFs as consistency constraints [12,6]. Although the idea is sound,

such methods are normally built based on well-established object detectors and thus require only simple interaction modeling; but both assumptions are not suitable for our problem domain.

2 Object Localization

Given an image I, we first oversegment it into a set of regions $\{r_p\}$, using quick-shift clustering [24], to incorporate superpixel-level information around the pixels. The objective of object localization is then to derive a binary mask $L = \{l_p\}$, with each $l_p \in \{0, 1\}$ indicating whether the region r_p belongs to the object.

2.1 The Proposed CRF Model

We formulate the object localization problem as a binary labeling task using a new CRF model, with the following energy function:

$$E(L|I) = \underbrace{\sum_{p} \phi_L(l_p)}_{\text{local}} + \underbrace{\sum_{(p,q)\in N_S} \psi_S(l_p, l_q)}_{\text{smooth}} + \underbrace{\sum_{(p,c)\in N_C} \psi_C(l_p, l_c)}_{\text{contrast}} + \underbrace{\sum_{(p,i)\in N_R} \psi_R(l_p, l_i)}_{\text{interest-region}} \tag{1}$$

where the set of random variables or nodes of the graph is denoted by $L = \{\{l_p\} \cup \{l_c\} \cup \{l_i\}\}$, including the new auxiliary nodes from the contrast (l_c) and interest-region (l_i) potentials. The probability of a certain configuration is a conditional distribution on the energy function $E(L|I)$, and the optimal labeling is derived by minimizing the total energy using the graph cut [10].

The local potential $\phi_L(l_p)$ describes the cost of r_p labeled as 0 or 1:

$$\phi_L(l_p) = 1 - P(r_p = l_p|f_p) \tag{2}$$

where f_p is the local feature vector of r_p, and $P(.)$ is the probability estimate of labeling obtained using a binary support vector machine (SVM).

The smooth potential $\psi_S(r_p, r_q)$ penalizes the differences in labeling of the neighboring regions r_p and r_q based on their feature distances with a Potts model:

$$\psi_S(l_p, l_q) = \exp(-\frac{\parallel f_p - f_q \parallel^2}{2\beta_S})\mathbf{1}(l_p \neq l_q) \tag{3}$$

where β_S is the normalization factor as the average of all L2 distances between neighboring feature vectors in I. The regions r_p and r_q are considered neighbors if they share some common border in I, and the set of all neighboring pairs is denoted by N_S.

While the first two potentials follow the standard CRF constructs (Figure 1a), we describe the contrast and interest-region potentials (ψ_C, ψ_R, N_C and N_R) in the following.

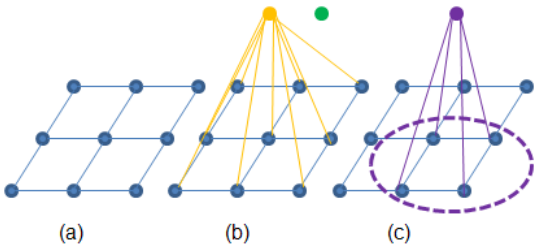


Fig. 1. The proposed CRF model. (a) The standard CRF construct, with nodes representing the image regions and edges linking the neighboring regions. (b) Introducing two auxiliary nodes (object and background) for the contrast potential, with edges linking the image regions and the auxiliary nodes (showing only one set of edges for easier viewing). (c) Based on the detected interest region (purple circle), an auxiliary node for the interest-region potential is added, with edges linking all image regions in the interest-region and the added node.

2.2 Contrast Potential

To improve the labeling accuracy, we want to explore the contrast information in the image I, with the following motivations. Across different images, there are often large inter-subject variations, causing overlaps between the feature ranges and hence misclassifications. Nevertheless, within one image, there is always a certain degree of contrast between the objects and background; and the contrast information helps to discriminate between the two types. To encode the contrast information, two additional nodes corresponding to the object and background, namely the contrast nodes l_c^o and l_c^b, are then added to the graph. A pairwise connection between the image region l_p and each of the two nodes is also established (Figure 1b), and N_C denotes the set of all such pairwise connections. With such a construct, we expect to encourage the same labelings between the image region and contrast nodes if they exhibit similar features, and also different labelings otherwise.

To do this, we first define the unary potentials of the two contrast nodes:

$$\phi_C(l_c^{o/b}) = \left\{ \begin{array}{ll} 0 & \text{if } l_c^{o/b} = 1/0 \\ C & \text{otherwise} \end{array} \right\} \tag{4}$$

where C is a large constant, so that large costs are assigned to $l_c^o \neq 1$ and $l_c^b \neq 0$ and 0 costs otherwise, to effectively fix the labelings of the two nodes in the inference results.

We then define the pairwise potentials for the edges (l_p, l_c) with the following. First, based on the labeling outputs with local features only (Eq. (2)), we obtain the initial estimation of the objects and background areas, and two feature vectors f_c^o and f_c^b are then derived for the estimated objects and background (details of feature derivation in Sec 4). Next, we compute the contrast features between r_p and the objects and background as $g_p = \{f_p, f_p/f_c^o, f_p/f_c^b\}$, and

classify the feature g_p to two classes – likely or unlikely to represent the object, denoted as $likely(1)$ and $unlikely(1)$ – using a binary SVM. Then, based on the probability estimates γ_p of class $likely(1)$, the pairwise costs are computed as:

$$\psi_C(l_p, l_c^o) = \left\{ \begin{array}{ll} 0 & \text{if } l_p = 1, \text{ and } likely(1) \\ 1-\gamma_p & \text{if } l_p = 1, \text{ and } unlikely(1) \\ \gamma_p & \text{if } l_p = 0 \end{array} \right\} \tag{5}$$

$$\psi_C(l_p, l_c^b) = \left\{ \begin{array}{ll} 0 & \text{if } l_p = 0, \text{ and } unlikely(1) \\ \gamma_p & \text{if } l_p = 0, \text{ and } likely(1) \\ 1-\gamma_p & \text{if } l_p = 1 \end{array} \right\} \tag{6}$$

Note that because of the *likely* and *unlikely* terms, the above pairwise potentials no longer follow the Potts model, and penalize labeling consistency if the features of the image regions and the contrast nodes are actually dissimilar. The total energy of the contrast potential can however, be rewritten in the following format, to keep it submodular (binary and with pairwise term encouraging consistency) for efficient graph-cut energy minimization:

$$\begin{array}{l} \sum_{(p,c)\in N_C} \psi_C(l_p, l_c) = \sum_c \phi_C(l_c) + \\ \sum_p \alpha_p \mathbf{1}(unlikely(l_p)) + \sum_{(p,c)\in N_C} \alpha_p \mathbf{1}(l_p \neq l_c) \end{array} \tag{7}$$

where $\alpha_p = \gamma_p$ if $l_p = 0$, and $\alpha_p = 1 - \gamma_p$ otherwise.

2.3 Interest Region Potential

Although the contrast nodes represent the object and background regions of an image I on a global scale, the structural information between image regions is not explored. An obviously important structural information is that, regions that are likely parts of the same anatomical or pathological structure should take the same labelings. In our formulation, the hypothesis is that, if we can detect a set of structures, i.e. interest regions R_i, the comprising regions $r_p \in R_i$ should preferably be assigned to the same category, but also depending on their individual suitability of such an labeling. The advantage of such an approach is that, we can employ a totally different method to detect the interest regions (e.g. non-CRF and different features), so the generated regions can serve as a second opinion to refine the object localization.

Assume a set of interest regions R_i are detected from an image I (details in Sec 3), and each interest region is characterized by its feature f_i, most probable label $l_i^* \in \{0, 1\}$ and a set of image regions r_p covered. Note that r_p might partially overlap with R_i especially around the boundary areas of R_i, and hence not all r_p covered by R_i should have the same label as l_i^*. To determine the the probability of $l_p = l_i^*$, we first compute the following feature vector:

$$v_p = \{\cap(r_p, R_i)/r_p, \parallel f_p - f_i \parallel, f_{i-p}/f_i\} \tag{8}$$

which represents the degrees of area overlap and feature homogeneity between r_p and R_i, with f_{i-p} denoting the feature of R_i excluding r_p. Then a binary SVM is trained to classify v_p into *same* or $diff$ categories, indicating if $l_p = l_i^*$ or otherwise, and the probability estimate of $l_p = l_i^*$ is denoted by $\theta_{p,i}$.

Next, to integrate the interest-region detection hypothesis into the CRF formulation, for each R_i detected, a node l_i is added to the graph, with the unary potential $\phi_R(l_i)$ defined similarly to Eq. (4). An edge is then connected between each pair of (l_p, l_i) for all $r_p \in R_i$ (Figure 1c) with N_R denoting all such edges for image I, and we define the pairwise potential as:

$$\psi_R(l_p, l_i) = \theta_{p,i}\mathbf{1}(l_p \neq l_i) \tag{9}$$

Since $r_p \in R_i$ is quite likely to exhibit the same labeling as R_i, we choose to use the Potts model to encourage such consistency. The cost of different labelings is directly related to the probability of $l_p = l_i^*$, and hence we use $\theta_{p,i}$ as the pairwise cost. If r_p is less likely to be labeled as l_i^*, the use of $\theta_{p,i}$ is also able to lessen the consistency constraint.

With the above definitions, the total energy term of the interest-region potential is thus rewritten as the following:

$$\sum_{(p,i)\in N_R} \psi_R(l_p, l_i) = \sum_i \phi_R(l_i) + \sum_{(p,i)\in N_R} \theta_{p,i}\mathbf{1}(l_p \neq l_i) \tag{10}$$

2.4 Graph Inference

All energy terms are given equal weights (based on our empirical evaluation), and piecewise learnings of the probability estimates used in the local, contrast and interest-region potentials are conducted first. The binary inference problem $L^* = \text{argmin } E(L|I)$ is then solved efficiently using the graph cut.

3 Detection for Interest Region Potential

Due to our motivation of detecting the interest regions in a totally different way from the graph-based approach to support the interest-region potential (Sec 2.3), we choose to design a sparse-coding based classification method for interest-region detection. Besides its popularity and widely demonstrated effectiveness [14], we believe sparse coding can be particularly suitable for our problem because of the large variations in the dataset.

3.1 Sparse Coding for Classification

Let Y be a set of n-dimensional data samples $Y = \{y_j : j = 1, ..., J\}$ and $Y \in R^{n\times J}$. A representative dictionary for Y with K atoms is denoted as $D = \{d_k : k = 1, ..., K\} \in R^{n\times K}$. Each y_j can then be represented as a linear combination of a few (i.e. $\leq T$) atoms in D with minimum reconstruction error, and the

corresponding coefficient vector x_j is the sparse code. Denoting the set of sparse codes of the data samples Y as $X = \{x_j : j = 1, ..., J\} \in R^{K \times J}$, both the dictionary D and sparse coding X can be learned with K-SVD [1] by solving the following problem:

$$\langle D, X \rangle = \underset{D,X}{\operatorname{argmin}} \ \|Y - DX\|_2^2 \quad s.t. \forall j, \|x_j\|_0 \leq T \tag{11}$$

where $\|Y - DX\|_2^2$ represents the reconstruction error.

Once the dictionary D is learned, a given data sample y can then be represented as a sparse code x by solving the following using the OMP algorithm [23]:

$$x = \underset{x}{\operatorname{argmin}} \ \|y - Dx\|_2^2 \quad s.t. \ \|x\|_0 \leq T \tag{12}$$

A classifier (e.g. SVM) can then be trained based on a set of such sparse codes, so that x and hence y can be classified.

In our context, an image I is divided into grid-based patches, and each image patch is represented by its feature descriptor y. The dictionary D is generated with a training set Y, and each image patch is then classified as interest region or otherwise ($h \in \{1, 0\}$) based on its sparse code x.

3.2 Discriminative Sparse Learning

A shortcoming with the above approach is the separation of the dictionary learning and classifier training, hence the learned dictionary might not produce discriminative sparse codes for the classification. Several approaches have thus been proposed to integrate the two steps of learning [9]. However, it is observed that such an integrated approach is still largely optimized for the reconstruction term, which may affect the discriminative power of W. Therefore, we suggest that the integrated learning for dictionary D should not totally replace the separate classifier training, and propose a different method as follows.

First, for the data samples $Y \in R^{n \times J}$, we create a corresponding labeling vector $H = \{h_j\} \in \{-1, 1\}^{1 \times J}$, with 1 for interest region. Based on linear-kernel SVM, the optimization objective of the weight vector $w \in R^{1 \times K}$ is:

$$\begin{aligned} &\operatorname{argmin}_{w,\xi,b} \ \tfrac{1}{2}\|w\|^2 + C \textstyle\sum_j \xi_j \\ &\qquad s.t. \ \forall j, h_j(w * x_j + b) \geq 1 - \xi_j, \ \xi_j \geq 0 \end{aligned} \tag{13}$$

Combining Eq. (11) and (13), and by simplifying the complexities caused by the inequality constraints on ξ_j and the signed h_j, we relax the formulation based on least squares SVM (LS-SVM) [22] as:

$$\begin{aligned} &\langle D, X, w \rangle = \operatorname{argmin}_{D,X,w} \ \|Y - DX\|_2^2 + \|w\|^2 + \textstyle\sum_j \xi_j^2 \\ &s.t. \ \forall j, \|x_j\|_0 \leq T, \ h_j(w * x_j + b) = 1 - \xi_j \end{aligned} \tag{14}$$

By combining w and b, and substituting ξ_j, the problem is then equivalent to the following:

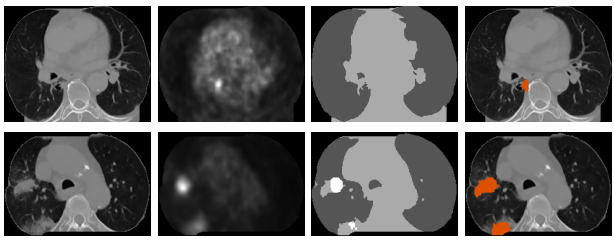
$$\langle D', X', w' \rangle = \text{argmin}_{D',X',w'} \ \|Y - D'X'\|_2^2 + \|w'\|^2 + \|H - w'X'\|_2^2 \qquad s.t.\ \forall j, \|x'_j\|_0 \leq T \tag{15}$$

where $w' = [w, b] \in R^{1\times(K+1)}$ and $X' \in R^{(K+1)\times J}$ appended an addition dimension with constant value 1 to absorb b, and $D' \in R^{n\times(K+1)}$ with an additional atom to be dimensionally compatible with X'. To solve Eq. (15), an alternative approach is used, as detailed in [20].

4 Experimental Results

4.1 Results on Lesion Dissimilarity

Measuring lesion similarity is important in many medical applications, such as content-based image retrieval for diagnosis referencing. In our approach, first, lesions (i.e. lung tumors and abnormal lymph nodes) in thoracic PET-CT images are localized in each image slice with the proposed method. Second, their textural and spatial features are extracted in 3D. Lastly, a weighted histogram-intersection is used to compute the feature distance. The actual implementation details are referred to [20]. The datasets comprise of 40 thoracic PET-CT 3D image sets from non-small cell lung cancer studies. A total of 64 lesions including lung tumors and abnormal lymph nodes are annotated, and the similarity/dissimilarity relationships between each pair of 3D image sets are marked as the ground truth. Three image sets showing typical thoracic characteristics are selected for training, and testing is performed on all image sets.



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

Fig. 2. Two example localization outputs. (a) Transaxial CT image slices (showing the thorax after preprocessing). (b) Co-registered PET image slices. (c) The labeling outputs using standard CRF, with dark gray for lung field, light gray for mediastinum and white for lesion. (d) Our localization outputs with the two additional potentials, with lesions highlighted in orange.

Figure 2 shows examples of the lesion localization. The first example illustrates the benefits of the contrast potential, in which the lesion is initially not detected with standard CRF, due to the relatively low PET intensities. The interest-region potential is particularly useful in refining the lesion boundaries, which tend to be underestimated with the standard CRF, as shown in the second example. It is observed that, the standard CRF tends to produce a large number of either totally undetected or underestimated lesions. Based on the measured 3D object-level localization results, we summarize the localization recall, precision and F-score in Table 1.

The localized lesions are then used to retrieval images with similar lesions. The retrieval tests are performed by using each 3D image set as a query image, and the remaining 39 images are ranked accordingly. We compare the retrieval performance with three other approaches: (i) state-of-the-art of thoracic PET-CT image retrieval [18]; (ii) spatial pyramid matching (SPM) with local intensity features extracted from grid-based image patches; and (iii) bag-of-words with SIFT descriptor. As shown in Figure 3, our proposed method exhibits the highest retrieval precisions for all recall levels.

Table 1. The localization performances comparing our proposed method with standard CRF

	Recall (%)	Precision (%)	F-score (%)
Ours	97.0	95.4	96.2
CRF	76.6	94.2	84.5

4.2 Results on Cell Segmentation

Cell nucleus segmentation is one of the most important tasks in analyzing and quantifying fluorescence microscopic images. In our approach, the cell nucleus is localized using the proposed method; and since the localization results also tend to delineate the nucleus boundaries closely, such an approach can be directly used for segmentation. The actual implementation details are referred to [20]. The serous database [13] is used to evaluate the cell segmentation. The database contains 10 microscopic images. A total of 254 cell nuclei are present in the images, with ground truth of cell nuclei segmentation provided. Same as [4], half of the images are used for training and the others for testing.

To evaluate the segmentation performance, we compute the PASCAL VOC criteria of pixel- and object-level accuracies, both as TP/(TP+FN+FP). We also compare our results with three approaches: (i) L+S, the standard CRF with local and smooth potentials; (ii) L+S+C, with additional contrast potential; (iii) L+S+R, with additional interest-region potential; and (iv) the state-of-the-art discriminative labeling method [4] reported for the same database. As listed in Table 2, our method achieves the highest pixel- and object-level accuracies.

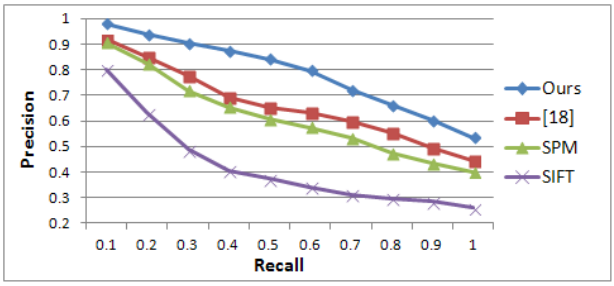


Fig. 3. The retrieval precision and recall

Table 2. The segmentation results comparing various methods

	Ours	L+S	L+S+C	L+S+R	[4]
Pixel Acc (%)	85.6	82.0	83.1	84.6	85.1
Obj Acc (%)	89.3	84.5	86.2	88.7	84.0

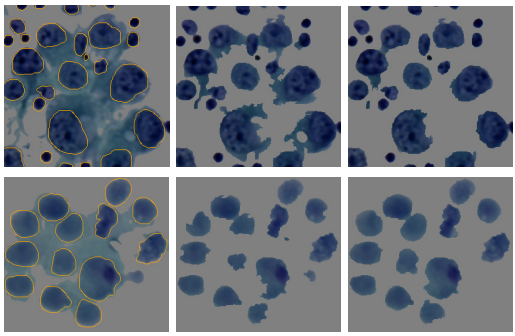


Fig. 4. Two example segmentation results. (a) Cropped microscopic images, with orange circles delineating the segmentation ground truth. (b) The segmentation results with L+S. (c) The segmentation results of our proposed method.

The improvements of having the contrast and potential terms are evident. The performance difference between L+S and [4] suggests that if we incorporate the feature set of [4], the segmentation accuracies would be further improved. By replacing the interest-region detection with standard sparse-coding classification, it is found that our proposed method exhibits on average 1.1% improvement for both pixel- and object-level measurements with the new approach.

The first example shown in Figure 4 indicates that our method is quite effective in removing the cytoplasm areas that connect the cell nuclei. As shown in the second example, lighter intensities of the cell nuclei cause many false

negatives with the standard CRF approach; and our result shows more accurate delineations of the actual contours.

5 Summary

In this Chapter, we describe a new method for object localization in medical images [20]. A new CRF model with additional contrast and interest-region potentials is proposed for effective object localization, addressing large inter-subject variations and low feature differences between the objects and background. A new sparse-coding classification approach is also designed for the interest-region detection, with enhanced discriminative power of the learned dictionaries. The proposed method is applied to lesion dissimilarity on thoracic PET-CT images, and cell segmentation on microscopic images, and shows higher performance compared to the state-of-the-art techniques.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. 54(1), 4311–4322 (2006)
2. Ben Ayed, I., Punithakumar, K., Garvin, G., Romano, W., Li, S.: Graph Cuts with Invariant Object-Interaction Priors: Application to Intervertebral Disc Segmentation. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 221–232. Springer, Heidelberg (2011)
3. Bauer, S., Nolte, L.-P., Reyes, M.: Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 354–361. Springer, Heidelberg (2011)
4. Cheng, L., Ye, N., Yu, W., Cheah, A.: Discriminative Segmentation of Microscopic Cellular Images. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 637–644. Springer, Heidelberg (2011)
5. Feuerstein, M., Glocker, B., Kitasaka, T., Nakamura, Y., Iwano, S., Mori, K.: Mediastinal atlas creation from 3-d chest computed tomography images: application to automated detection and station mapping of lymph nodes. Med. Image Anal. 16(1), 63–74 (2011)
6. Gonfaus, J., Boix, X.: Harmony potentials for joint classification and segmentation. In: CVPR, pp. 3280–3287 (2010)
7. Guo, R., Dai, Q., Hoiem, D.: Single-image shadow detection and removal using paired regions. In: CVPR, pp. 2033–2040 (2011)
8. Jagadeesh, V., Vu, N., Manjunath, B.S.: Multiple Structure Tracing in 3D Electron Micrographs. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 613–620. Springer, Heidelberg (2011)
9. Jiang, Z., Lin, Z., Davis, L.: Learning a discriminative dictionary for sparse coding via label consistent K-SVD. In: CVPR, pp. 1697–1704 (2011)
10. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. 26(2), 147–159 (2004)

11. Ladicky, L., Russell, C., Kohli, P., Torr, P.H.S.: Associative hierarchical CRFs for object class image segmentation. In: ICCV, pp. 739–746 (2009)
12. Ladický, Ľ., Sturgess, P., Alahari, K., Russell, C., Torr, P.H.S.: What, Where and How Many? Combining Object Detectors and CRFs. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 424–437. Springer, Heidelberg (2010)
13. Lezoray, O., Cardot, H.: Cooperation of color pixel classification schemes and color watershed: a study for microscopical images. IEEE Trans. Image Process. 11(7), 783–789 (2002)
14. Liu, M., Lu, L., Ye, X., Yu, S., Salganicoff, M.: Sparse Classification for Computer Aided Diagnosis Using Learned Dictionaries. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 41–48. Springer, Heidelberg (2011)
15. Lu, C., Chelikani, S., Jaffray, D.A., Milosevic, M.F., Staib, L.H., Juncan, J.S.: Simultaneous nonrigid registration, segmentation, and tumor detection in MRI guided cervical cancer radiation therapy. IEEE Trans. Med. Imag. 31(6), 1213–1227 (2012)
16. van Ravesteijin, V.F., van Wijk, C., Vos, F.M., Truyen, R., Peters, J.F., Stoker, J., van Vliet, L.J.: Computer-aided detection of polyps in CT colonography using logistic regression. IEEE Trans. Med. Imag. 29(1), 120–131 (2010)
17. Shotton, J., Winn, J.M., Rother, C., Criminisi, A.: *TextonBoost*: Joint Appearance, Shape and Context Modeling for Multi-class Object Recognition and Segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 1–15. Springer, Heidelberg (2006)
18. Song, Y., Cai, W., Eberl, S., Fulham, M.J., Feng, D.: Discriminative Pathological Context Detection in Thoracic Images Based on Multi-level Inference. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 191–198. Springer, Heidelberg (2011)
19. Song, Y., Cai, W., Eberl, S., Fulham, M., Feng, D.: Thoracic image case retrieval with spatial and contextual information. In: ISBI, pp. 1885–1888 (2011)
20. Song, Y., Cai, W., Huang, H., Wang, Y., Feng, D.D.: Object localization in medical images based on graphical model with contrast and interest-region terms. In: CVPR Workshop, pp. 1–7 (2012)
21. Song, Y., Cai, W., Kim, J., Feng, D.D.: A multistage discriminative model for tumor and lymph node detection in thoracic images. IEEE Trans. Med. Imag. 31(5), 1061–1075 (2012)
22. Suykens, J., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Letters 9(3), 293–300 (1999)
23. Tropp, J.: Greed is good: algorithmic results for sparse approximation. IEEE Trans. Inf. Theory 50(10), 2231–2242 (2004)
24. Vedaldi, A., Soatto, S.: Quick Shift and Kernel Methods for Mode Seeking. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 705–718. Springer, Heidelberg (2008)
25. Wu, D., Lu, L., Bi, J., Shinagawa, Y., Boyer, K., Krishnan, A., Salganicoff, M.: Stratified learning of local anatomical context for lung nodules in CT images. In: CVPR, pp. 2791–2798 (2010)

A Cascade Learning Method for Liver Lesion Detection in CT Images

Dijia Wu[1], David Liu[1], Michael Suehling[1], Kevin S. Zhou[1], and Christian Tietjen[2]

[1] Siemens Corporate Research, Princeton NJ 08544, USA
[2] Siemens Healthcare, Siemensstr. 1, Forchheim 91301, Germany

Abstract. The automatic detection and segmentation of liver lesion is useful in many clinical application, whereas it remains a challenging task due to the largely varied shape, size and texture of the diseased masses. In this paper, we present a cascade learning approach comprising multiple classifiers for the detection of two different types of solid liver lesions, hypodense and hyperdense lesions. In particular, we propose an efficient gradient based locally adaptive segmentation method for the solid lesions, where the segmentation results are used to extract shape features to boost up the detection performance. The proposed method is validated on a total of 660 volumes with 1,302 hypodense lesions, and 234 volumes with 328 hyperdense lesions. The experimental results show a resulting 90% detection rate at 1.01 false positives per volume for hypodense lesion and 1.58 false positives per volume for hyperdense lesion, respectively, using three fold cross validation.

1 Introduction

Detection and segmentation of abnormal hepatic masses is important to liver disease diagnosis, treatment planning and follow-up monitoring. As a significant part of clinical practice in radiology, liver tumors are usually examined and tracked every several weeks or months to assess the cancer staging and therapy response based on 3D Computed Tomography (CT) data. However, manually finding these lesions is tedious and time consuming, and highly dependent on the observer's experiences. Hence, a system of automatic lesion detection and measurement is desirable.

There is a limited amount of previous work directed to automatic liver lesion detection, compared with lesion segmentation. Ye et al. proposed the use of gray-level statistical features and temporal enhancement pattern of different contrast phases to classify the liver tissues with SVM, however, it required experienced radiologists to draw region of interest in advance and multi-phase enhanced CT images [1]. Moltz et al. presented a simple threshold filtering method followed by circular structure detection with Hugh transform to locate matching lesions in follow-up CT examinations [2]. It assumes that the lesion mask of the baseline scan is available, hence the tumor location is known coarsely and detection can be restricted to a local area in the follow up scans. A multi-level Otsu's method with

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 206–214, 2013.

level set algorithm was used in [3] to segment complicated liver lesion but the user has to first manually select several points covering the whole lesion inside the liver area. Other lesion segmentation methods dependent on user interactions include random walker [4], graph cut and watershed [5], and seeded region growing [6].

In this work, we present a fully automated method to detect two most common types of hepatic lesions, *hypodense* (darker) and *hyperdense* (brighter) lesions, from single 3D CT image of any contrast phase. It generates lesion candidates with a learning based approach as opposed to simplistic thresholding or painstaking local minimum point clustering [7,8]. Because lesion detection is usually a highly unbalanced classification problem where negative samples, i.e., healthy tissues or other structures such as vessels, are many more than positive samples, a cascade learning framework [9] is employed to speed up the detection and improve the classification result for unbalanced data problems [10]. This differentiates our method from other learning based liver lesion segmentation approaches such as ensemble segmentation using AdaBoost [11] or iterative Bayesian approach [12]. In addition, we propose a new gradient based locally adaptive lesion segmentation method. The aim of the segmentation is not to perfectly locate the lesion boundaries, but provide fast and reasonable segmentation results from which geometric and statistical features can be extracted to improve detections. The idea of coupling segmentation and detection was proposed in [13] and later applied in lymph node detection problem [14]. Our work uses a much simpler segmentation method than the Gaussian MRF and gradient descent in [14] and extracts different segmentation based features.

The rest of the paper is organized as follows. Section 2 describes this cascade learning system for liver lesion detection and outlines the gradient based locally adaptive segmentation method. It will be explained how the unsupervised constructed segmentation can be used to improve the supervised detection performance. In Section 3, experimental validation results on two particular types of liver lesions, hypodense and hyperdense lesions, are presented. We conclude with a review of our contribution and potential extensions in Section 4.

2 Liver Lesion Detection and Segmentation

2.1 Liver Segmentation and Liver Lesion Annotation

To constrain the search, the liver is first automatically segmented using a hierarchical learning based method described in [15]; the liver subvolume is then cropped and resampled to 1.5 mm isotropic resolution. Each liver lesion of size at least 10 mm in the dataset is annotated by placing a bounding box around it as shown in Fig.1. The voxels within a predefined distance from the bounding box centers are used as positive samples and voxels out of the boxes as negative samples in training. The lesions are labeled as hypodense (darker) or hyperdense (brighter) depending on the enhancement pattern difference between normal liver parenchyma and lesions.

2.2 Detection and Segmentation

Like many other Computer Aided Diagnosis (CAD) problems, lesion detection data sets are large and extremely unbalanced between positive and negative classes, given the fact that liver lesions are generally a few and small compared with the whole liver volume. Therefore, we use a cascade classification framework for lesion detection as shown in Fig.2. The coarse-to-fine cascade structure has been shown effective in speeding up the detection process by discarding many negative samples with fewer simple features before more complex classifiers are called upon to further reduce the false positives [9]. The cascade framework can also be used to simplify the difficult unbalanced classification problem into a sequence of linear programs, each of which separates only a subset of negative samples from the positives [10].

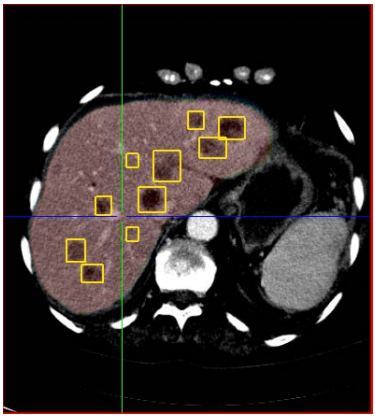


Fig. 1. The liver lesions are annotated with bounding boxes. The segmented liver is overlaid on the original volume in light red.

As shown in Fig.2, the proposed liver lesion detection system comprises four classifiers from simple to complex. First, we use a Haar based detector to generate lesion candidates followed by bootstrapping to prune these candidates also with Haar features. Then lesion segmentation is performed and the resulting segmentation is used to obtain geometric and statistical features to verify the candidates and reject the negative ones. Finally, a more informative set of steerable features [16] are extracted from the segmentation to further reduce false positives.

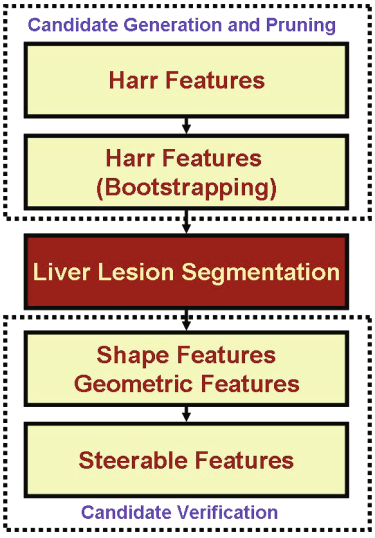


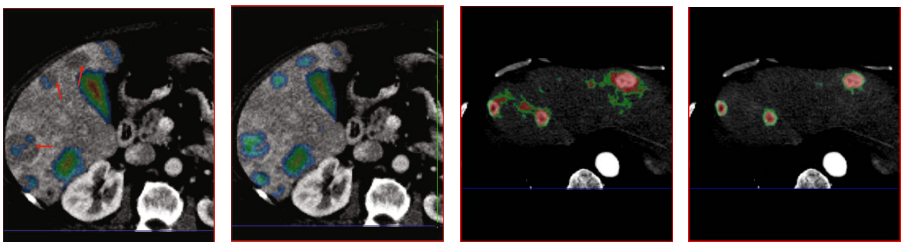
Fig. 2. Liver lesion detection system

Lesion Candidate Generation. Liver lesion center candidates are detected from all voxels in the liver subvolume in two stages. The initial set of candidates are generated using a fast Haar-based detector. It is a cascade of two AdaBoost classifiers [17], the first classifier has 100 weak learners and the second has 200. They are trained using

138,464 3D Haar features [18] with all positive voxels and 1% negative voxels randomly sampled. This stage achieves 100% detection rate with an average false positive rate about 0.4% on training data.

A second Haar detector is trained for bootstrapping, using the same set of features and classifier configuration but with all positives and all negatives passing through the first stage. This stage achieves about 25% false positive rate on the average at 100% detection rate on training data. The significantly increased false positive rate suggests that Haar features are not enough to further reduce the false detections. More complicated features such as texture features or shape features are often used to distinguish various lesions, e.g., gray-level co-occurrence matrix [1], local binary pattern [19] or Hessian eigen-system based filters [20]. However, these features are computationally expensive. In this work, we employ geometric and statistical features which embed the shape and texture information and can be efficiently computed from the lesion segmentation. From the perspective of marginal space learning [16], segmentation provides extra information about object size and orientation that improves the detection performance.

Lesion Segmentation. Many liver lesion segmentation methods have been proposed as mentioned in Section 1. We choose adaptive thresholding because it is fully automatic, very simple and precise segmentation is not our target. However, all previous works use single threshold for filtering the lesions, which is selected with different methods such as histogram analysis [2,21], or cross-entropy minimization [22]. However, single threshold is subject to the constraint of inhomogeneous lesions in one liver. To solve this problem, we present a multi-thresholding method based on local surface gradients as given in Algorithm 1. This method presents a connected component tree structure which is similar to maximally stable extremal regions (MSER) [23] approach used in stereo matching and object recognition. But we select the optimal threshold based on maximum gradient response rather than area stability criterion in [23]. The example segmentation results of the proposed multi-threshold method are given and compared with single threshold [2] in Fig.3. It is clear that the proposed



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

Fig. 3. (a) (c) single threshold segmentation; (b) (d) proposed multi-threshold segmentation. Note that for hyperdense lesion, the original volume should be inverted before applying algorithm 1.

Algorithm 1. Gradient based locally adaptive segmentation method

Input: Liver volume $I(x, y, z)$ and number of thresholds n.
Output: Threshold $\Omega(x, y, z)$, binary segmentation $S(x, y, z) = 1$ if $I(x, y, z) < \Omega(x, y, z)$ and 0 otherwise.

1. Run liver intensity histogram analysis and obtain the peak value τ and standard deviation σ. Set $\tau_{min} = \tau - n\sigma$,$\tau_{max} = \tau$ and $\Delta\tau = \sigma$.
2. Start from single threshold $\omega = \tau_{max}$ and obtain initial binary segmentation S.
3. Run 3D connected component labeling on S. Each connected component is denoted as C_i and its surface as R_i.
4. Calculate the mean surface gradient norm of C_i: $G_i = \Sigma_{(x,y,z)\in R_i}|\nabla I(x, y, z)|/|R_i|$.
5. For C_i, use threshold $\omega' = \omega - \Delta\tau$ to obtain new segmentation and connected components C_i', R_i' and G_i'.
6. $\Omega(x, y, z) = \omega$ if $G_i \geq G_i'$ and ω' otherwise, $\forall(x, y, z) \in C_i$, then update S.
7. Set $\omega = \omega'$ and $\omega' = \omega' - \Delta\tau$ and repeat step 3 to 7 until $\omega = \tau_{min}$.

method achieves better segmentation results. Finally, watershed transform [24] is performed on the segmentation results to separate closely connected lesions.

Lesion Candidate Verification. The segmentation is used to derive more informative features for further evaluation of the liver lesion candidates. The candidate verification consists of two coarse-to-fine detectors as shown in Fig.2. The first detector calculates 28 geometric features and 6 statistical features of each connected component obtained in the segmentation. The geometric features include diameter, volume size, surface area, compactness, rectangularity, elongation, central moments and so on; the statistical features comprise min, max, mean, variance, skewness and kurtosis of intensities. Because some structures in the liver show similar intensities to the lesions, for instance, vessels and hyperdense lesions are both enhanced in the arterial phase, many segmented objects are not lesions. Therefore, we use these shape and statistical descriptors to identify and reject the obvious non-lesion segmentations.

The second candidate verification detector uses much more dense steerable features [14,16] computed from the segmentation to further remove difficult false positives. The features are calculated by casting rays in 162 directions in 3D space from each candidate location as shown in Fig.4. In each direction, the following features are calculated at the boundary of the segmentation:

Intensity Based Features: Assume the intensity and gradient at boundary (x, y, z) is I and $g = (g_x, g_y, g_z)$, respectively. For each of the 162 directions, we compute 24 feature types, including I, $\sqrt{(I)}$, I^2, I^3, $\log I$, g_x, g_y, g_z, $||g||$, $\sqrt{(||g||)}$, $||g||^2$, $||g||^3$, $\log||g||$. To incorporate invariance into the detector, the 162 values for each feature type are sorted by value. This not only ensures rotational invariance, but invariance

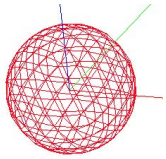


Fig. 4. Triangulation of a sphere using 162 vertices and 320 triangles

to all permutations, including mirroring. Additionally, for each of the 24 feature types, the 81 sums of feature values at the pairs of opposite vertices are computed and sorted by value.

Geometry Features: The 81 diameters (distances between opposite vertices relative to the segmentation center) are sorted. For each diameter the following features are computed: (a) The value of each diameter. (b) Asymmetry of each diameter, i.e. the ratio of the larger radius over the smaller radius. (c) The ratio of the i-th sorted diameter and the j-th diameter for all $1 \leq i < j \leq 81$. (d) For each of the 24 feature types, the max or min of the feature values at the two diameter ends. (e) For each of the 24 feature types, the max or min of the feature values half way to the diameter ends.

In total there are about 17,000 features. Using these features, a cascade of two AdaBoost classifiers with 70 and 140 weak learners each is trained. Because multiple candidates can be detected in a single lesion, all the remaining candidates at the final stage are clustered using non-maximal suppression [14]. To accommodate to lesions of vastly different sizes, the above process is repeated with different resolutions in a pyramid manner.

3 Experimental Results and Discussion

Data Collection. In the experiment, we collected 661 liver CT subvolumes from 564 subjects with 1,302 hypodense lesions, and 234 volumes from 198 subjects with 328 hyperdense lesions. The annotations were obtained as described in Section 2.1 by two radiologists based on visual assessment and consensus review. In this work, we target tumors of moderate size with diameter between 10 mm to 100 mm, therefore all the annotated lesions are of size in this range. Data were collected from multiple hospitals.

Evaluation Methodology. A lesion is considered as detected if there exists a detection with its center inside this lesion bounding box, whereas a detection is considered as false positive if its center outside any annotated lesion bounding boxes.

Results. The liver subvolumes are split into training and testing data via three-fold cross validation which is repeated for 5 times and all results presented here are the averages over 5 runs. The volumes of the same patient are always put into the same folder. The resulting ROC curves are given in Fig.6. The proposed detection system reaches 1.01 false positives per volume and 1.58 false positives per volume at 90% sensitivity for hypodense and hyperdense lesion, respectively. The detection performance based on single threshold segmentation method [2] and without watershed transform is also compared in Fig.6.

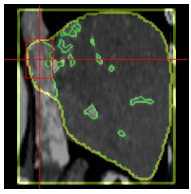


Fig. 5. False positive of hyperdense lesion detection

The hypodense lesion detection is better than hyperdense detection because we have less annotations of hyperdense lesions and more importantly, hyperdense lesions detection is more easily confused with other bright structures especially the vessels as shown in Fig.5. In this example, the aorta is falsely truncated and segmented as a part of liver, which is then misclassified as a hyperdense lesion.

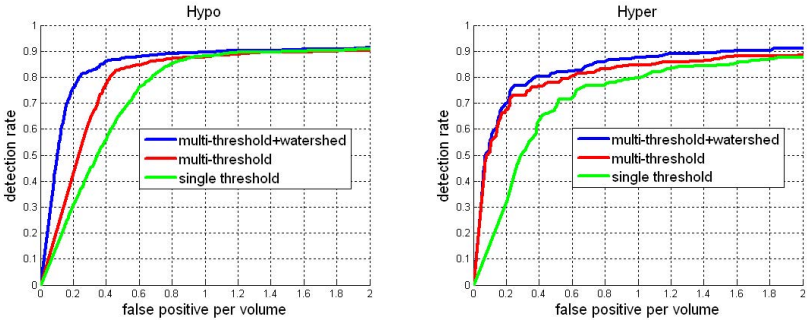


Fig. 6. Lesion detection ROC curves

Examples of detected true positives are shown in Fig.7. The bounding box of a detected lesion is obtained from the segmentation. Note that not all segmented objects are lesions. For hyperdense lesion, the liver segmentation is also given. As shown in Fig.7, the proposed system can detect lesions of highly different sizes, shapes, intensities and positions in the liver. The average running time is 20-30 seconds per volume.

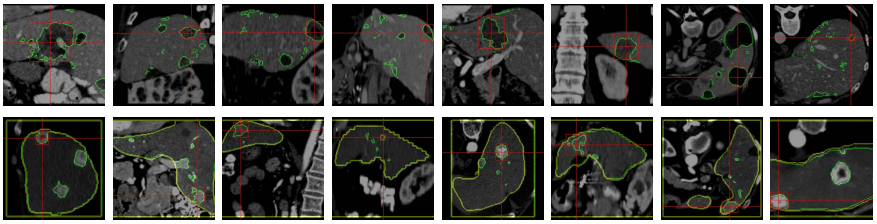


Fig. 7. Example detection and segmentation results. Top row: hypodense lesions. Bottom row: hyperdense lesion.

4 Conclusion

In this paper, we presented a cascade learning system for automated liver lesion detection based on a novel multi-threshold segmentation method. We discussed how the segmentation results can be used to extract shape and intensity features to improve the detection performance. Ongoing work will include improvement of hyperdense lesion detection, particularly separation from the vessels. As liver

lesions exhibit various appearances in different contrast phases, the prior knowledge of the contrast phase can also potentially benefit the detection performance. Also, because both hyperdense and hypodense lesions possess similar shapes and structures, they might be used in training together to improve each other's detection with techniques such as transfer learning.

References

1. Ye, J., Sun, Y., Wang, S., Gu, L., Qian, L., Xu, J.: Multi-Phase CT Image Based Hepatic Lesion Diagnosis by SVM. In: Proc. iCBBE, vol. 1, pp. 1–5 (2009)
2. Moltz, J.H., Schwier, M., Peitgen, H.O.: A General Framework for Automatic Detection of Matching Lesions in Followup CT. In: Proc. ISBI, vol. 1, pp. 843–846 (2009)
3. Luo, Z., Wu, X., Cen, R., Ou, S.: Segmentation of Complicated Liver Lesion Based on Local Multiphase Level Set. In: Proc. iCBBE, vol. 1, pp. 1–4 (2009)
4. Grady, L., Jolly, M.P.: 3D General Lesion Segmentation in CT. In: Proc. ISBI (2008)
5. Stawiaski, J., Decencire, E., Bidault, F.: Interactive Liver Tumor Segmentation Using Graph-cuts and Watershed. In: Proc. MICCAI Workshop (2008)
6. Wong, D., Liu, J., Yin, F., Tian, Q., Xiong, W.: A Semi-Automated Method for Liver Tumor Segmentation Based on 2D Region Growing with Knowledge-Based Constraints. In: Proc. MICCAI Workshop (2008)
7. Hame, Y.: Liver Tumor Segmentation Using Implicit Surface Evolution. In: Proc. MICCAI Workshop (2008)
8. Kubota, T.: Efficient Automated Detection and Segmentation of Medium and Large Liver Tumors: CAD Approach. In: Proc. MICCAI Workshop (2008)
9. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. Int J. Comput. Vision 57, 137–154 (2004)
10. Bi, J., Periaswamy, S., Okada, K., Kubota, T., Fung, G., Salganicoff, M., Rao, B.: Computer Aided Detection via Asymmetric Cascade of Sparse Hyperplane Classifiers. In: Proc. SIGKDD (2006)
11. Shimizu, A., Narihira, T., Furukawa, D., Kobatake, H., Nawano, S., Shinozaki, K.: Ensemble Segmentation Using AdaBoost with Application to Liver Lesion Extraction from a CT Volume. In: Proc. MICCAI Workshop (2008)
12. Taieb, Y., Eliassaf, O., Freiman, et al.: An Iterative Bayesian Approach for Liver Analysis: Tumors Validation Study. In: Proc. MICCAI Workshop (2008)
13. Leibe, B., Leonardis, A., Schiele, B.: Robust Object Detection with Interleaved Categorization and Segmentation. Int J. Comput. Vision 77, 259–289 (2008)
14. Barbu, A., Suehling, M., Xu, X., Liu, D., Zhou, S.K., Comaniciu, D.: Automatic Detection and Segmentation of Axillary Lymph Nodes. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 28–36. Springer, Heidelberg (2010)
15. Ling, H., Zheng, Y., Georgescu, B., Zhou, S.K., Suehling, M.: Hierarchical Learning-Based Automatic Liver Segmentation. In: Proc. CVPR (2008)
16. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-Chamber Heart Modeling and Automatic Segmentation for 3-D Cardiac CT Volumes Using Marginal Space Learning and Steerable Features. IEEE Trans. Med. Imag. 27(11), 1668–1681 (2008)

17. Tu, Z.: Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering. In: Proc. ICCV, vol. 2, pp. 1589–1596 (2005)
18. Tu, Z., Zhou, X.S., Barbu, A., Bogoni, L., Comaniciu, D.: Probabilistic 3D Polyp Detection in CT Images: The Role of Sample Alignment. In: Proc. CVPR (2006)
19. Zhou, J., Chang, S., Metaxas, D., Zhao, B., Schwartz, L.H., Ginsberg, M.S.: Automatic Detection and Segmentation of Ground Glass Opacity Nodules. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 784–791. Springer, Heidelberg (2006)
20. Wu, D., Lu, L., Bi, J., Shinagawa, Y., Boyer, K., et al.: Stratified Learning of Local Anatomical Context for Lung Nodules in CT Images. In: Proc. CVPR (2010)
21. Moltz, J.H., Bornemann, L., Dicken, V., Peitgen, H.O.: Segmentation of Liver Metastases in CT Scans by Adaptive Thresholding and Morphological Processing. In: Proc. MICCAI Workshop (2008)
22. Choudhary, A., Moretto, N., Ferrarese, F.P., Zamboni, G.A.: An Entropy Based Multi-Thresholding Method for Semi-Automatic Segmentation of Liver Tumors. In: Proc. MICCAI Workshop (2008)
23. Matas, J., Chum, O., Urba, M., Pajdla, T.: Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In: Proc. BMVC, pp. 384–396 (2002)
24. Vincent, L., Soille, P.: Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. IEEE Trans. Pattern Anal. Mach. Intell. 13(6), 583–598 (1991)

Automatic Event Detection within Thrombus Formation Based on Integer Programming

Loic Peter[1], Olivier Pauly[1,2], Sjoert B.G. Jansen[3,4], Peter A. Smethurst[3,4], Willem H. Ouwehand[3,4,5], Nassir Navab[1]

[1] Computer Aided Medical Procedures, Technische Universitaet Muenchen, Munich, Germany
[2] Institute of Biomathematics and Biometry, Helmholtz Zentrum Muenchen, Munich, Germany
[3] Department of Haematology, University of Cambridge, Cambridge, United Kingdom
[4] National Health Service Blood and Transplant, Cambridge, United Kingdom
[5] The Wellcome Trust Sanger Institute, Hinxton, United Kingdom

Abstract. After a blood vessel injury, blood platelets progressively aggregate on the damaged site to stop the resulting blood loss. This natural mechanism called thrombosis can however be prone to malfunctions and lead to the complete obstruction of the blood vessel. Thrombosis disorders play a crucial role in coronary artery diseases and the identification of genetic risk predispositions would therefore considerably help their diagnosis and therapy. *In vitro* experiments are conducted in this purpose by perfusing blood from several donors over a surface of collagen fibres, which results in the progressive attachment of platelets. Based on the segmentation over time of these aggregates called thrombi, we propose in this paper an automatic method combining tracking and event detection which allows the extraction of characteristics of interest for each thrombus growth individually, in order to find a potential correlation between these growth features and blood donors genetic disorders. We demonstrate the benefits of our approach and the accuracy of its results through an experimental validation.

Keywords: Microscopy image analysis, thrombus segmentation, multi-target tracking, event detection.

1 Introduction

Thrombosis denotes the abnormal coagulation of platelets that may occur after a blood vessel injury and eventually leads to the complete obstruction of the blood circulation. In addition to the identification of environmental risk factors such as smoking, obesity, and physical inactivity, the study of possible genetic predispositions to thrombosis is becoming increasingly important to improve the diagnosis and the therapy of coronary artery diseases. Genome-wide association studies identified some potential novel genes as being correlated with thrombosis. Experimental analyses are conducted to confirm their involvement in thrombosis

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 215–224, 2013.

malfunction, either *in vivo* on zebrafish larvae [1] or *in vitro* by perfusing freshly collected human blood through a chamber filled with collagen fibres. In the latter case, the progressive attachment of platelets leads to the formation of individual thrombi observed through a phase-contrast microscopic system. The growth rate of each individual thrombus over time is a measure of interest as well as the time to attachment of their first platelet. The high number of required experiments raises the need of a tool able to automatically segment and track the different thrombus areas over time to derive their individual growth characteristics.

The tracking of multiple objects within microscopic videos is a challenging problem for which many methods exist in the litterature. It often follows a first step where objects of interest are detected. Based on these detections, objects are matched from frame to frame using optimisation methods as for example the Hungarian algorithm for linear assignment [2] or the branch and bound algorithm for binary integer programming [3]. Other approaches are based on model evolution like particle filtering generalised for the tracking of several objects [4]. Most of the tracking methods follow one of these two kinds of approaches, or try to combine them [5]. A first attempt has been presented in [6] for the segmentation and tracking of thrombi in a similar experimental setup. Authors introduced three complementary gradient-based features that were learned on a video of reference. Then, by feeding these features into a decision tree, they could demonstrate promising segmentation results. Each thrombus was defined by the first platelet and tracked over the whole video to ultimately obtain the growth curve of each thrombus over time. However, the fact that thrombi grow at many different locations often results in merging events. The blood flow regularly causes the detachment of platelets, which can also result in splitting of thrombi. Several tracking methods able to follow objects along videos despite split and merge conditions emerged from the computer vision community [7,8,9]. They assume however that each object is well defined all along the scene. In our case, the integrity of each thrombus vanishes as soon as it exchanges platelets with other thrombi, e.g during merging or splitting. Moreover, such events disrupt the measured growth by inducing artificial changes of size without any biological meaning. Being able to identify such events is a major requirement since it permits to compute reliable growth rates to perform a meaningful comparison between blood donors. We introduce in this paper a new method which is able, from a segmented video, to match objects between two consecutive frames to identify appearance, disappearance, splitting and merging events. Thereby we can extract the relevant information to our application.

2 Methods

In this section, we first introduce our method for the thrombus segmentation (Section 2.1). We describe afterwards our automatic method to identify special events by assigning objects between two consecutive frames (Section 2.2).

2.1 Thrombus Segmentation

We propose to formulate the thrombus segmentation problem as a binary classification task in which each pixel of a given frame is assigned to one of these two classes of objects: background (B) or thrombus (T). The segmentation is performed independently within each frame. More formally, let us consider a frame represented by the intensity function $\mathbf{I} : \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{N}^2$ represents the pixel lattice in the image domain. We denote by $\mathbf{x} = (x, y)$ a pixel of coordinates $(x, y) \in \Omega$ in the frame $\mathbf{I}$. Our goal is to assign a label $\mathbf{c} \in \{\text{background}(B), \text{thrombus}(T)\}$ to each pixel $\mathbf{x}$ of the frame $\mathbf{I}$. In a probabilistic fashion, this could be done by modeling the posterior distribution $P(\mathbf{c}|\mathbf{x}, \mathbf{I})$ and using maximum a posteriori:

$$\hat{\mathbf{c}} = \underset{\mathbf{c}\in\{B,T\}}{\operatorname{argmax}} P(\mathbf{c}|\mathbf{x}, \mathbf{I}) \tag{1}$$

The posterior $P(\mathbf{c}|\mathbf{x}, \mathbf{I})$ quantifies the probability of observing the class $\mathbf{c}$ at the pixel $\mathbf{x}$ given the information available over the frame. To model this posterior, we propose to use a classification forest as described in [10]. Therefore, we generate a training set from a set of pixels extracted at different frames. Each training instance is a pair $(\mathbf{X}^{(n)}, \mathbf{c}^{(n)})$, where $n = \{1, \cdots, N_{\text{train}}\}$, that represents a pixel $\mathbf{x}^{(n)}$ from a given frame described by a feature vector $\mathbf{X}^{(n)}$ and its corresponding class label $\mathbf{c}^{(n)}$. To characterize the visual context of a pixel, we extract at different scales a set of gradient-based features [6]. Following a "divide and conquer" strategy, each tree of a forest $\{\mathbf{T}_i\}_{i=1}^{N_{\text{trees}}}$ provides a piece-wise approximation $P_i(\mathbf{c}|\mathbf{x}, \mathbf{I})$ by: (1) creating a partition over the feature space using simple decision functions and (2), estimating the posterior in each cell of this partition. Tree posteriors can be aggregated over the whole forest using averaging:

$$P(\mathbf{c}|\mathbf{x}, \mathbf{I}) = \frac{1}{N_{\text{trees}}} \sum_{i=1}^{N_{\text{trees}}} P_i(\mathbf{c}|\mathbf{x}, \mathbf{I}) \tag{2}$$

The final segmentation is obtained by thresholding this posterior, with a threshold chosen to maximise the performance with respect to manually delineated videos. Each connected component of the segmented image is called object. We also apply some post-processing operations (morphological opening to discard objects whose shape is too elongated and we remove the objects or holes smaller than the size of a platelet).

2.2 Event Detections

Given the segmentation at frames t and $t + 1$, we want to identify the events that occur between these two frames. The segmentation at frame t is a collection of binary objects $(\mathcal{O}_{t,k})_{1\leq k\leq N_t}$. Between t and $t + 1$, the possible events are the following:

- **Appearance**: A new object (generally a single platelet) appears in the frame $t+1$. Such appearances can occur anywhere within the field of view.
- **Disappearance**: An object of the frame t cannot be seen anymore in the frame $t+1$. It often corresponds to isolated platelets that detach.
- **Merge**: Several objects of the frame t merge into an object of the frame $t+1$.
- **Split**: An object of the frame t splits in several objects in the frame $t+1$.
- **Normal growth**: A thrombus grows undisrupted between the two frames.

The identification of these events is seen as an **assignment problem**. Similarly as for a multi-target tracking problem, each object in a frame must be associated to one, several or no objects in the next (or previous) one. The possibility to assign an object to several objects (resp. none) allows us to identify splitting and merging (resp. appearance and disappearance). Assignments are found globally through the resolution of a binary integer programming problem minimising a cost function especially designed for this application. We start the description of our method by the definition of two types of distances we will use. We then identify candidate regions for splitting and merging events by clustering thrombi that are close to each other. Finally, we explain in details our formalism with the help of a concrete example and the associated optimisation problem.

Distances between Two Objects. Distances between objects are an essential quantity to estimate the cost of each association of objects. In the following, we define two types of distances : the **static** and **dynamic** distances between two objects.

Static distance: Within a same frame, we define the static distance between two objects $\mathcal{O}_{t,i}$ and $\mathcal{O}_{t,j}$ by

$$d_S(\mathcal{O}_{t,i}, \mathcal{O}_{t,j}) = \min_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{O}_{t,i} \times \mathcal{O}_{t,j}} d(\mathbf{x}_i, \mathbf{x}_j) \tag{3}$$

where d is the classical Euclidean distance between two points.

Dynamic distance: To define a distance between an object $\mathcal{O}_{t,i}$ at frame t and an object $\mathcal{O}_{t+1,j}$ at frame $t+1$, we could use the distance we have just defined. However, we propose to introduce some additional knowledge. In our controlled experiment, we assume the blood flow to be laminar and constant through the chamber, i.e horizontally from left to right in our images. Therefore, objects can physically only move towards the right of the field of view with a significantly low vertical component. Anticipating the fact that the distance between objects $\mathcal{O}_{t,i}$ and $\mathcal{O}_{t+1,j}$ will be used as a cost to link $\mathcal{O}_{t,i}$ and $\mathcal{O}_{t+1,j}$, we propose to forbide physically impossible motions by assigning an infinite distance between the two objects if $\mathcal{O}_{t+1,j}$ is located "behind" $\mathcal{O}_{t,i}$. More precisely, the (asymmetric) dynamic distance is defined as

$$d_D(\mathcal{O}_{t,i}, \mathcal{O}_{t+1,j}) = \min_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{O}_{t,i} \times \mathcal{O}_{t+1,j}} d'_\theta(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

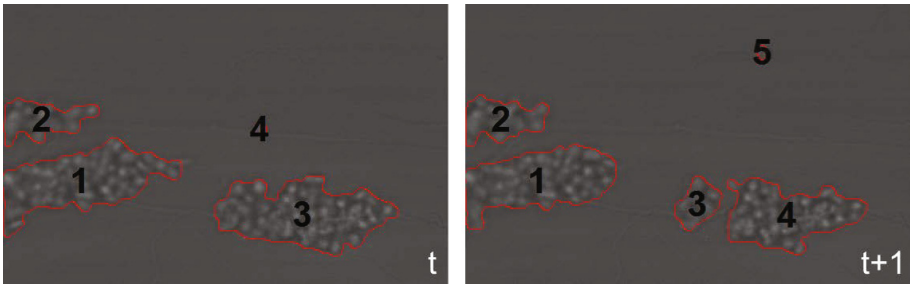


Fig. 1. Two consecutive frames from a video. The frames have been spatially cropped to reduce the number of objects and facilitate the visualisation. Please note that the labels of the objects are arbitrarily generated within each frame and do not symbolise any tracking.

where $d'_\theta(\mathbf{x}_i, \mathbf{x}_j)$ is the classical Euclidean distance between two points if $\mathbf{x}_j$ is located in the cone whose apex is $\mathbf{x}_i$, whose aperture is θ and horizontally oriented towards the right. If not, $d'_\theta(\mathbf{x}_i, \mathbf{x}_j)$ is set to infinity (a very high number in practice).

Potential Splitting and Merging Regions. Let us assume that N_t objects are segmented in the frame t, where N_t rarely goes above 30 in such experiments. We start our analysis by clustering objects that are close to each other in order to identify candidate regions where merging or splitting might occur. We define a maximum distance $d^{\text{SM}}_{\max}$ stating the maximum possible static distance between two objects at frame t that are susceptible to merge. For each object $\mathcal{O}_{t,i}$, we define its candidate objects for splitting and merging as the objects $\mathcal{O}_{t,j}$ verifying $d_S(\mathcal{O}_{t,i}, \mathcal{O}_{t,j}) \leq d^{\text{SM}}_{\max}$. The identification of these candidate regions reduces thereby the number of events we consider in our assignment problem and makes it more tractable (the theoretical number of all possible combinations would be exponential with respect to N_t).

Assignment Using Binary Representation. Let us denote N_m and N_s respectively the number of possible merges and the number of possible splits computed as we just described. Our goal is to assign to each object the type of event it is involved in and to link it to the right object(s) in the other frame. We represent the assignment of all objects by a binary matrix X of size $K_t \times K_{t+1}$ with $K_t = N_t + N_m + 1$ and $K_{t+1} = N_{t+1} + N_s + 1$. The optimal assignment matrix X is found as the solution of a minimisation binary integer programming problem under equality constraint.

Before going in details into the construction of this minimisation problem, we illustrate on an example how a matrix X encodes the assignments. Let us consider the situation shown in Figure 1, taken from a real video but cropped to reduce the number of objects and increase the readability. 4 objects are seen

in the frame t and 5 objects in the frame $t+1$. By constructing the equivalent classes of thrombi in the frame t, we find the merge $M\{1,2\}$ as the only possible one. Similarly, looking at the equivalence classes in the frame $t+1$ informs us about the possible splits: only $S\{1,2\}$ and $S\{3,4\}$. The expected assignment matrix is the following:

	$\mathcal{O}_{t+1,1}$	$\mathcal{O}_{t+1,2}$	$\mathcal{O}_{t+1,3}$	$\mathcal{O}_{t+1,4}$	$\mathcal{O}_{t+1,5}$	$S\{1,2\}$	$S\{3,4\}$	Disappearance
$\mathcal{O}_{t,1}$	1	0	0	0	0	0	0	0
$\mathcal{O}_{t,2}$	0	1	0	0	0	0	0	0
$\mathcal{O}_{t,3}$	0	0	0	0	0	0	1	0
$\mathcal{O}_{t,4}$	0	0	0	0	0	0	0	1
$M\{1,2\}$	0	0	0	0	0	0	0	0
Appearance	0	0	0	0	1	0	0	0

We can reformulate the information included in this matrix as follows. The objects 1 and 2 of the frame t are normally growing without being involved in any particular event. They are respectively linked to the objects 1 and 2 in the frame $t+1$. Although the labelling number of these objects remains the same between the two frames in this case, please note that labels are independently assigned at each frame. The object 3 in the frame t splits into two objects labelled 3 and 4 in the frame $t+1$. Finally, the object 4 of the frame t is disappearing, while an object labelled 5 appears in the frame $t+1$. These two objects are not linked to each other since the trajectory which this link would form is unrealistic from a physical point of view (the vertical component is too high).

We can see with this example how X summarises in the general case all the assignments. The rows correspond to the frame t and the columns to the frame $t+1$. More precisely, the N_t first rows correspond to the objects of the frame t, the N_m following rows correspond to the possible merge events and the last one corresponds to appearances. Similarly, the columns correspond to objects in the frame $t+1$, splits and disapperances. We propose to estimate the assignment matrix X as the solution of this optimisation problem:

$$\min_{X \in \mathcal{M}_{K_t, K_{t+1}}(\{0,1\})} \|C.X\|_1 \tag{5}$$

where C is a cost matrix and . denotes the pointwise product. C summarizes the cost associated to each possible assignment and the way it is computed is described later in the paper. We add a linear equality constraint on X stating that there is one and only one positive assignment for every object. This can be formalised as

$$\forall k \in \{1, \ldots, N_t\} \sum_{i \in \phi(k)} \sum_{j=1}^{K_{t+1}} X(i,j) = 1 \tag{6}$$

and

$$\forall k \in \{1, \ldots, N_{t+1}\} \sum_{j \in \psi(k)} \sum_{i=1}^{K_t} X(i,j) = 1 \tag{7}$$

where $\phi(k)$ (resp. $\psi(k)$) denotes the set of the indices of the rows (resp. columns) where the object $\mathcal{O}_{t,k}$ (resp. $\mathcal{O}_{t+1,k}$) is involved. Our minimisation problem belongs to the class of binary integer programming problems. We classically propose to solve it with the branch and bound algorithm [11].

Cost Matrix. We define the cost of each assignment as follows:

- The cost of disappearance of an object $\mathcal{O}_{t,i}$ is set to γ size$(\mathcal{O}_{t,i})$
- The cost of appearance of an object $\mathcal{O}_{t+1,j}$ is set to γ size$(\mathcal{O}_{t+1,j})$
- The cost of linking an object $\mathcal{O}_{t,i}$ to an object $\mathcal{O}_{t+1,j}$ is set to $d_D(\mathcal{O}_{t,i}, \mathcal{O}_{t+1,j}) + \alpha \ |\text{size}(\mathcal{O}_{t,i}) - \text{size}(\mathcal{O}_{t+1,j})|$
- The cost of merging several objects $(\mathcal{O}_{t,i_k})_k$ into an object $\mathcal{O}_{t+1,j}$ is set to $\beta \ \max_k d_D(\mathcal{O}_{t,i_k}, \mathcal{O}_{t+1,j}) + \alpha \ |\text{size}(\sum_k \mathcal{O}_{t,i_k}) - \text{size}(\mathcal{O}_{t+1,j})|$
- The cost of splitting an object $\mathcal{O}_{t,i}$ into several objects $(\mathcal{O}_{t+1,j_k})_k$ is set to $\beta \ \max_k d_D(\mathcal{O}_{t,i}, \mathcal{O}_{t+1,j_k}) + \alpha \ |\text{size}(\sum_k \mathcal{O}_{t+1,j_k}) - \text{size}(\mathcal{O}_{t,i})|$

All the other costs of the matrix are set to infinity. Let us give some intuitions about these choices of costs. Appearance and disappearance events concern only small objects (mostly single platelets) since bigger objects are more robustly attached. We thus set the cost of appearance or disappearance as proportional to the size of the object. The cost of matching two objects, in the case of normal growth, is set as the dynamic distance between them (which takes into account the direction of the flow). Since the growth between two frames is always low, an additional cost comparing the sizes of the two objects is added to prevent unrealistic matchings. To compute the cost of several objects to merge into a single one, we compute for each of them the dynamic distance to this object and take the biggest of them as cost. If an object within a set of merging objects is unrelevant, the whole merging event is thus penalised and ultimately a merging set involving only relevant objects will be preferred. We also assume that there is not any motion of the objects involved in a merging event. The coefficient β is therefore taken high to penalise the distance term. Finally, the consistency in size is also checked. The cost for splitting events is similar than for merging. α, β and γ are coefficients balancing the relative weight of each term.

3 Experiments and Results

Random Forest Training. 7 videos were available to both learning and evaluating steps. One frame every 10 seconds, giving approximately 15 test frames for each video, was manually delineated to provide a reference for the learning and the evaluation. We trained the random forest using a "leave-one-video-out" approach: each video is segmented by learning the random forest on the 6 others. For each frame, we draw randomly points within background and thrombus. A subset of the points representing the background is purposely constrained to the neighbourhood of thrombi to perform a better robustness around thrombus edges. The number of trees in the forest is fixed at 50 and the optimal depth 20 has been tuned experimentally. The gradient-based features are computed at 13 different scales ($r \in \{8, \ldots, 20\}$).

Table 1. Evaluation of the segmentation performance

	Video 1	Video 2	Video 3	Video 4	Video 5	Video 6	Video 7
F-measure	0.89	0.90	0.89	0.895	0.89	0.86	0.85

Table 2. Evaluation of the performance of our event detection method

	Normal growth	Merging	Splitting	Disappearance	Appearance
TP	4159	164	138	188	253
FP	2	7	11	5	4
FN	12	0	0	3	9
Precision	0.9995	0.96	0.93	0.97	0.98
Recall	0.997	1.00	1.00	0.98	0.97
F-measure	0.998	0.98	0.96	0.98	0.975

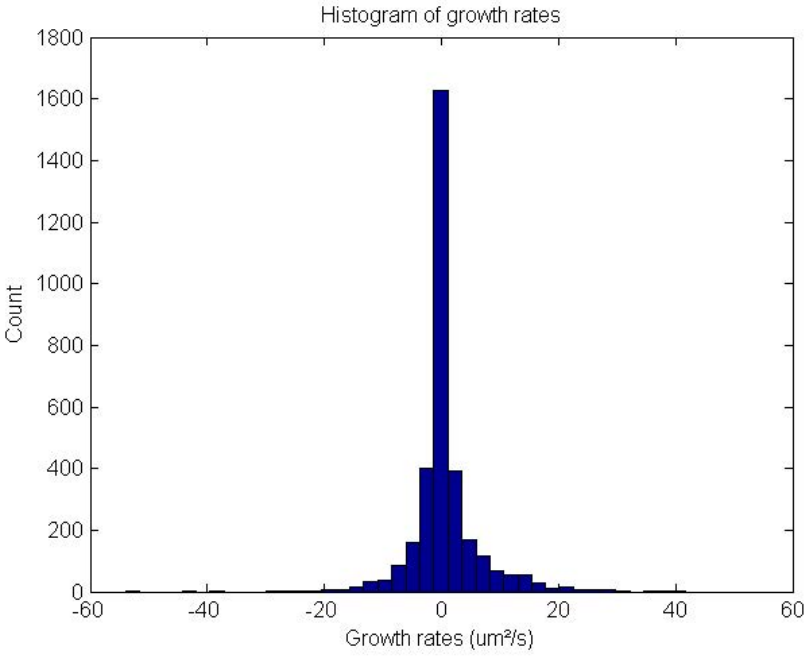


Fig. 2. The identification of normal growth events, i.e where no splitting and merging occurs, permits a reliable measurement of growth rates. We can then plot them as an histogram for each video. It is expected that the histograms are correlated with the genetic disorders of the blood donors.

Segmentation. We choose the F-measure to quantify the performance of the thrombus segmentation. The results for each video are briefly summarised in Table 1. These results show that our segmentation method is accurate enough to build our event analysis upon it.

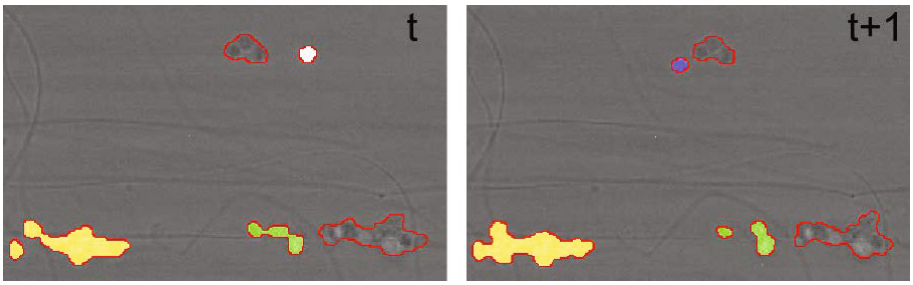


Fig. 3. Example of results of event detection between two consecutive frames. A code based on colour has been chosen for a better visualisation. A thrombus in yellow (resp. green) merges (resp. splits) between the two frames. A thrombus in white (resp. blue) disappears (resp. appears) between t and $t+1$. Thrombi that do not have any colour are evolving between the two frames without any interaction with other thrombi.

Event Detection. The parameters in the cost function are experimentally set to $\alpha = 0.1$, $\beta = 3$ and $\gamma = 0.2$. We test our event detection method on 2 videos entirely labeled. Please note that although the number of videos on which we test is very low, this represents within each video at least 100 pairs of frames for which the assignments are independant. We count for each kind of event how many times this event has been correctly identified (TP), how many times it has been by mistake detected (FP) and how many times it has not been identified (FN). We also compute for each event the precision, recall and F-measure. The results are summarised in Table 2 and demonstrate the accuracy of our method. In particular, the precision is extremely high for the normal growth events. This allows us to reliably measure growth rates by excluding splitting and merging events. Resulting growth rates can be plotted as histograms (Figure 2) to visualise the thrombotic characteristics of a given blood donor.

4 Conclusion

In this paper, we tackled the problem of measuring growth rates and time to attachment of thrombi under splitting, merging, appearance and disappearance conditions. We proposed a matching method between each pair of consecutive frames which is able to recognise such undesired events in order to measure growth rates only in normal conditions of growth. We modeled this situation as a binary integer programming problem which is tractable and solvable with the branch and bound algorithm. We showed through a quantification of performance the efficiency of the approach. The extracted characteristics of growth could be compared between blood donors and potentially allow to identify possible genetic risk predispositions of thrombosis.

References

1. Brieu, N., Navab, N., Serbanovic-Canic, J., Ouwehand, W.H., Stemple, D.L., Cvejic, A., Groher, M.: Image-based characterization of thrombus formation in time-lapse dic microscopy. Medical Image Analysis 16(4), 915–931 (2012)
2. Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L., Danuser, G.: Robust single-particle tracking in live-cell time-lapse sequences. Nature Methods 5(8), 695–702 (2008)
3. Li, F., Zhou, X., Ma, J., Wong, S.: Multiple nuclei tracking using integer programming for quantitative cancer cell cycle analysis. IEEE Transactions on Medical Imaging 29(1), 96–105 (2010)
4. Smal, I., Draegestein, K., Galjart, N., Niessen, W., Meijering, E.: Particle filtering for multiple object tracking in dynamic fluorescence microscopy images: Application to microtubule growth analysis. IEEE Transactions on Medical Imaging 27(6), 789–804 (2008)
5. Li, K., Miller, E.D., Chen, M., Kanade, T., Weiss, L.E., Campbell, P.G.: Cell population tracking and lineage construction with spatiotemporal context. Medical Image Analysis 12(5), 546–566 (2008); Special issue on the 10th international conference on medical imaging and computer assisted intervention - MICCAI 2007
6. Peter, L., Brieu, N., Jansen, S., Smethurst, P.A., Ouwehand, W.H., Navab, N.: Automatic segmentation and tracking of thrombus formation within in vitro microscopic video sequences. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2012)
7. Nillius, P., Sullivan, J., Carlsson, S.: Multi-target tracking - linking identities using bayesian network inference. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2187–2194 (2006)
8. Bose, B., Wang, X., Grimson, E.: Multi-class object tracking algorithm that handles fragmentation and grouping. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (June 2007)
9. Khan, Z., Balch, T., Dellaert, F.: Mcmc-based particle filtering for tracking a variable number of interacting targets. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(11), 1805–1819 (2005)
10. Criminisi, A., Shotton, J., Konukoglu, E.: Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. In: Foundations and Trends in Computer Graphics and Vision, vol. 7 (2012)
11. Wolsey, L.A.: Integer Programming. John Wiley and Sons, Inc. (1998)

Automatic Extraction of the Curved Midsagittal Brain Surface on MR Images

Hugo J. Kuijf, Max A. Viergever, and Koen L. Vincken

Image Sciences Institute, University Medical Center Utrecht, The Netherlands

Abstract. Many methods exist for the automatic extraction of the midsagittal plane from neuroimages, assuming bilateral symmetry. However, this assumption is incorrect owing to brain torque and the possible presence of pathology. In this paper, a method for extracting the curved midsagittal surface from brain images is presented.

First, the method localizes the interhemispheric fissure with an existing technique for midsagittal plane extraction. Next, the plane is modelled as a bicubic spline and the configuration of the control points is optimized to obtain the midsagittal surface.

The midsagittal surface results in a better segmentation of the cerebral hemispheres. Not only is the result visually more appealing, the absolute volume of misclassified tissue decreases significantly.

1 Introduction

Bilateral symmetry is an important concept in biology and many animal species, including humans. Our appearance exhibits bilateral symmetry and some organs in our body come in symmetrical pairs, for example our brain. The cerebrum is divided into two hemispheres, separated by the interhemispheric fissure (IF). Comparison of the two hemispheres and detection of differences has been a topic of discussion for many years. Besides the lateralization of brain function, anatomical differences can suggest the presence of pathology (like a brain mass or tumour), indicate schizophrenia [1], or various other diseases.

The midsagittal plane is a geometric plane that separates the two hemispheres and coincides with the IF. In the past years, multiple methods have been published to extract the midsagittal plane from neuroimages. Assuming bilateral symmetry, most of the methods work by optimizing a symmetry metric between the neuroimage and a reflected version of itself.

However, the human brain has no perfect bilateral symmetry. The left occipital and right frontal lobe are larger than their counterparts in the other hemisphere are, which is known as brain torque. Besides brain torque, the presence of brain masses could induce asymmetries in the cerebrum. Existing techniques to extract the midsagittal plane from neuroimages do not take these asymmetries into account and might therefore fail to correctly segment the two hemispheres.

The midsagittal surface is a curved surface following the IF. In the presence of asymmetries, either owing to natural variation or pathology, the midsagittal surface will correctly segment the two hemispheres, whereas a midsagittal plane

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 225–232, 2013.

would intersect or misclassify some brain tissue. It is therefore likely that the midsagittal surface will result in more accurate analysis of interhemispherical differences.

In the present study, a novel method for extracting the curved midsagittal surface will be presented, based on an existing method for extracting the midsagittal plane.

2 Methods and Materials

2.1 Participants and MRI

A total of 50 consecutive participants (mean age: 59 years, sd: 10 years) from the SMART study [2] have been included for evaluation of the method. The SMART study was approved by the Medical Ethics Committee and written informed consent was given by all participants.

MRI acquisition was performed on a 1.5T whole-body system (Gyroscan ACS-NT, Philips Medical Systems, Best, the Netherlands). The protocol included, among others, a transversal T1-weighted gradient-echo sequence (repetition time (TR)/echo time (TE): 235/2 ms); a transversal T2-weighted fluid-attenuated inversion recovery (FLAIR) (TR/TE/inversion time (TI): 6000/100/2000 ms), and a transversal inversion recovery (IR) (TR/TE/TI: 2900/22/410 ms), all with a reconstructed voxel size of $0.9 \times 0.9 \times 4.0$ mm.

For extraction of the midsagittal plane and surface, the T1-weighted sequence was used.

2.2 Midsagittal Plane

Many methods exist for the automatic extraction of the midsagittal plane, which can be roughly divided into two categories: symmetry-based methods (e.g. [3–10]) and fissure-based methods ([11–13]). Symmetry-based methods work with the implicit assumption that the brain possesses bilateral symmetry. These methods try to align the image with a reflected version of itself, while optimizing a symmetry-metric. However, due to the asymmetric nature of the brain, these techniques sometimes fail.

Fissure-based methods try to detect the IF, based on its distinctive characteristics visible in the image. With imaging modalities as CT and MRI, the cerebrospinal fluid (CSF) located in the IF gives a high visual contrast with the surrounding gray and white matter of both hemispheres. This contrast is clearly visible in Figures 1(a), (d), and (g), and can be used to extract the midsagittal plane and surface. A fissure-based method for extracting the midsagittal plane is described by Volkau *et al.* [12] and Nowinski *et al.* [13] and was used in the present study. The approach of this method, as will be explained below, allows for extension to extract the midsagittal surface and thus formed an ideal candidate for the present study.

First, two reference planes were taken 2 cm apart from the central sagittal slice of the image. As the method assumes that the brain is approximately located in

the centre of the image, these reference planes consist mostly of gray and white matter. A single probability distribution of the gray values present in the two references slices was created.

Next, all slices in-between the two reference slices were inspected. For each slice, a probability distribution of the gray values was created. The Kullback-Leibler (KL) divergence was computed using the reference probability distribution and the probability distribution of the current slice. This resulted in a measure of the difference between the two probability distributions (the KL-value). As the reference slices contains mostly gray and white matter and the IF contains mostly CSF, the slice containing (a large part of) the IF would result in a relatively large KL-value.

The sagittal slice with the largest KL-value was taken as an initial guess for the MSP. As the brain can be rotated, the IF will not always perfectly align with a sagittal slice in the image. Therefore, three random corner points of the MSP were taken and shifted along the left/right axis of the scan. The location of these three corner points could be optimized in terms of the KL-value. For each new location, the KL-value was computed and the rotated slice with the largest difference to the reference distribution was taken as the final MSP. This process is summarized in Figure 1.

2.3 Midsagittal Surface

The midsagittal plane computed in the previous section was used to initialize the method for extracting the midsagittal surface. The surface was represented as a bicubic spline, as implemented in ALGLIB [14]. Control points for the spline were placed in a regular $m \times n$ grid on the computed MSP, having m be the number of control points in the anterior-posterior direction and n in the head-feet direction. The values of m and n were user-defined. An example is shown in Figure 2(a)

An optimization method was used to determine the optimal configuration of the control points. The control points could only be moved along the left/right axis of the scan during optimization. The Kullback-Leibler's divergence was used as a cost function that needed to be maximized. It used the previously computed reference probability distribution and generated a probability distribution of the bicubic spline during optimization.

A limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-newton method (L-BFGS), as implemented in the dlib C++ library [15], was used to determine the direction of the search. This method required gradient information of the cost function to be optimized, which was numerically approximated. The step size of each control point in each iteration was scaled with the gradient at each control point, allowing subvoxel accuracy in the configuration. The optimization method was terminated when the cost function converged: two consecutive optimization steps had a difference in KL-value of 1×10^{-5} or less. An example of this procedure can be seen in Figure 2.

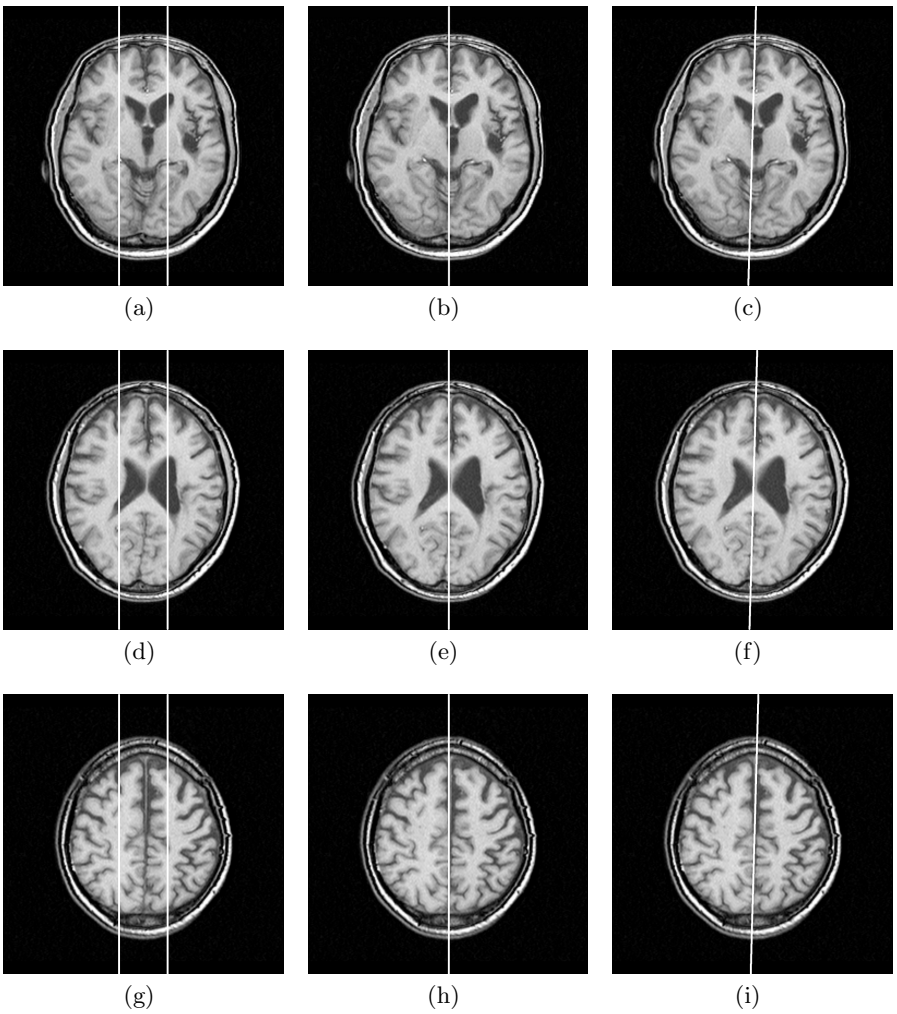
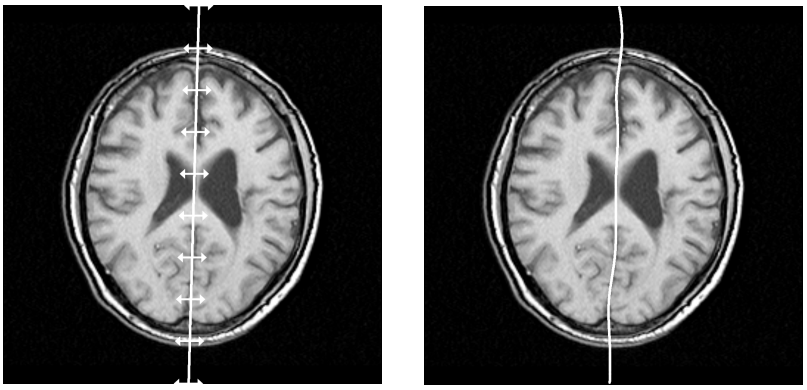


Fig. 1. Extraction of the midsagittal plane shown on three slices taken from one scan. Left: lines indicate the two initial reference planes. Middle: line indicates the sagittal slice with the largest KL-value. Right: line indicates the midsagittal plane.

2.4 Experiments and Validation

The quality of the midsagittal plane and surface extraction was evaluated visually and quantitatively in the cerebrum, ignoring the cerebellum. The cerebellum was ignored, since a left/right segmentation is ambiguous and ill-defined. This is commonly done in segmentation algorithms. [16] For the quantitative validation, the brain tissue volume in the cerebrum that was classified as either left or right was assessed automatically and compared to a ground truth. First, reasonable settings for m and n were determined heuristically on a smaller subset of



(a) (b)

Fig. 2. Left: Figure 1(f) is shown with the control points for the optimization. The arrows indicate the direction in which the control points could move. Right: Optimal configuration according to the KL-divergence is shown. The error that the midsagittal plane made at the left occipital lobe was corrected by the computed surface.

participants. The influence of these parameters was assessed visually and fixed for the quantitative analysis.

Second, a gray and white matter segmentation was obtained with a probabilistic k-Nearest Neighbour classification segmentation method. This method used the T1-weighted, IR, and FLAIR sequences, as described by Anbeek *et al.* [17]

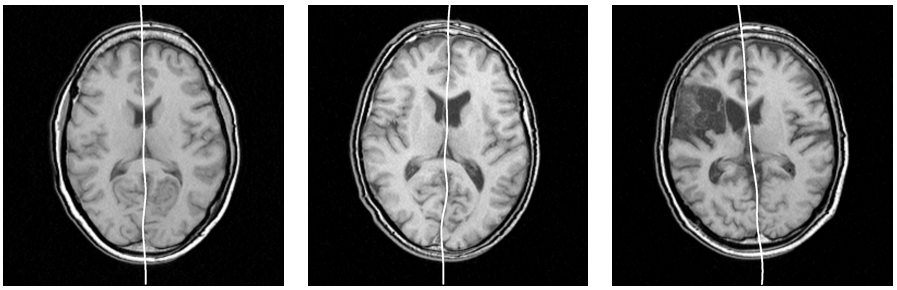
Using the MNI152 template [18, 19], a ground truth left/right segmentation was created. For this template, a true left/right atlas of the cerebrum is available. By computing a deformable registration of the MNI152 template to the T1-weighted scan, the left/right atlas could be propagated to the gray and white matter segmentation. Registrations were computed with `elastix` [20], with registration parameters taken from Van der Lijn *et al.* [21] The quality of the registration was assessed visually by an experienced observer and all registrations were considered accurate.

3 Results

Values for m and n were set at 10 and 5. Lower values were unable to capture the curvature of the IF and higher values resulted in overfitting of the spline.

The results of the extraction of the midsagittal plane were visually inspected for correctness. In all cases, the midsagittal plane was found correctly and aligned with the IF, as was also previously reported by Volkau and Nowinski. [12, 13] However, small errors were made by the method, mostly at the left occipital lobe (as visible in Figure 1) and the right frontal lobe. This was to be expected, owing to the possible presence of brain torque.

The midsagittal surface showed a visually more appealing result than the midsagittal plane. The surface followed the interhemispheric fissure at locations



(a) (b) (c)

Fig. 3. Example results of the method. (b) The midsagittal surface was sometimes fitted through the lateral ventricles. (c) Asymmetries in the brain do not influence the results, as would be the case for symmetry-based methods.

where the midsagittal plane would cut through tissue. An example of this was shown in Figure 2(b) and more results are shown in Figure 3.

For the quantitative validation, the absolute volume of tissue in the cerebrum that was misclassified (i.e. classified as left where it should be right, or vice versa) was assessed automatically. In the ideal case, this error would be zero. The average error (mean ± sd) of the midsagittal plane was 1.06 ± 0.89 ml and the average error of the midsagittal surface was 0.59 ± 0.63 ml. The difference between the midsagittal plane and the midsagittal surface was statistically significant, using a one-sided, paired, Student's t-test, with a p-value of 1.0×10^{-6}.

Computation time of the midsagittal plane was approximately 2 seconds. Depending on the number of iterations required, the computation of the midsagittal surface was 2 to 10 seconds.

4 Discussion

The implementation of the midsagittal surface shows a clear improvement over the midsagittal plane. Not only does the midsagittal surface show a visually more correct and appealing result, the improvement in the absolute volume of misclassified tissue is statistically significant. Besides the statistical significance, the absolute reduction in the error with 0.5 ml on average is relevant in many applications. Although a volume 0.5 ml is relatively small compared to the whole brain volume, it is a considerable amount of tissue in the vicinity of the IF. Next to the average reduction in error, the standard deviation of the error also decreases. This indicates that the midsagittal surface gives a more robust estimate of the left-right segmentation than the midsagittal plane does.

An error sometimes made by the midsagittal surface is found at the location of the lateral ventricles. During optimization, the cost function will try to avoid the septum pellucidum, the membrane separating the lateral ventricles, and fit the spline through the CSF-filled ventricles. This will give a more optimal solution in terms of KL-value, although it is not the most logical separation at that location.

However, it has no influence on the left/right segmentation of the tissue in the cerebrum. An example was shown in Figure 3(b).

The quantitative validation of the method required an atlas-registration of the MNI152 template to each individual image. By doing this, the quality of the ground truth depends on the quality of the registration. Thorough inspection of the registration results by an experienced observer did not reveal any errors in the left-right segmentation of the cerebrum generated by the registration.

Of course, one could argue that having a left-right segmentation available by means of registration with the MNI152 template already solves the problem of extracting the midsagittal surface. However, the registration of the MNI152 template to the scans required 7 minutes per scan. The computation of the midsagittal surface required, at most, 12 seconds, making an expensive atlas registration superfluous.

The method works without adaptation on other image contrasts, such as FLAIR and IR, and higher field strengths, such as 3.0T or 7.0T. The only prerequisite for the method is a visible contrast between the interhemispheric fissure and surrounding tissue. The method can be applied to other imaging modalities, such as CT, as well. [22]

Besides segmentation of the cerebrum into the left and right hemispheres, there is a possibility to use this method for the detection of midline shift. Techniques for this application have been published before [23], using the midsagittal plane and a Bézier curve. The midsagittal surface could be used instead, without the limited degrees of freedom of a Bézier curve.

References

1. Crow, T.: Schizophrenia as an anomaly of cerebral asymmetriy. In: Imaging of the Brain in Psychiatry and Related Fields, pp. 1–17 (1993)
2. Simons, P., Algra, A., van de Laak, M., Grobbee, D., van der Graaf, Y.: Second manifestations of arterial disease (smart) study: Rationale and design. European Journal of Epidemiology 15, 773–781 (1999)
3. Junck, L., Moen, J.G., Hutchins, G.D., Brown, M.B., Kuhl, D.E.: Correlation methods for the centering, rotation, and alignment of functional brain images. Journal of Nuclear Medicine 31(7), 1220–1226 (1990)
4. Minoshima, S., Berger, K.L., Lee, K.S., Mintun, M.A.: An automated method for rotational correction and centering of three-dimensional functional brain images. Journal of Nuclear Medicine 33(8), 1579–1585 (1992)
5. Ardekani, B.A., Kershaw, J., Braun, M., Kanno, I.: Automatic detection of the midsagittal plane in 3-d brain images. IEEE Transactions on Medical Imaging 16(6), 947–952 (1997)
6. Smith, S., Jenkinson, M.: Accurate Robust Symmetry Estimation. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 308–317. Springer, Heidelberg (1999)
7. Liu, Y., Collins, R., Rothfus, W.E.: Robust midsagittal plane extraction from normal and pathological 3d neuroradiology images. IEEE Transactions on Medical Imaging 20(1), 175–192 (2001)
8. Prima, S., Ourselin, S., Ayache, N.: Computation of the mid-sagittal plane in 3-d brain images. IEEE Transactions on Medical Imaging 21(2), 122–138 (2002)

9. Hu, Q., Nowinski, W.L.: A rapid algorithm for robust and automatic extraction of the midsagittal plane of the human cerebrum from neuroimages based on local symmetry and outlier removal. NeuroImage 20(4), 2153–2165 (2003)
10. Tuzikov, A.V., Colliot, O., Bloch, I.: Evaluation of the symmetry plane in 3d mr brain images. Pattern Recognition Letters 24(14), 2219–2233 (2003)
11. Brummer, M.E.: Hough transform detection of the longitudinal fissure in tomographic head images. IEEE Transactions on Medical Imaging 10(1), 74–81 (1991)
12. Volkau, I., Prakash, K.B., Ananthasubramaniam, A., Aziz, A., Nowinski, W.L.: Extraction of the midsagittal plane from morphological neuroimages using the kullback-leibler's measure. Medical Image Analysis 10(6), 863–874 (2006)
13. Nowinski, W.L., Prakash, B., Volkau, I., Ananthasubramaniam, A., Beauchamp Jr, N.J.: Rapid and automatic calculation of the midsagittal plane in magnetic resonance diffusion and perfusion images. Academic Radiology 13(5), 652–663 (2006)
14. Bochkanov, S., Bystritsky, V.: Alglib, `www.alglib.net`
15. King, D.E.: Dlib c++ library, `www.dlib.net`
16. Liang, L., Rehm, K., Woods, R.P., Rottenberg, D.A.: Automatic segmentation of left and right cerebral hemispheres from mri brain volumes using the graph cuts algorithm. NeuroImage 34(3), 1160–1170 (2007)
17. Anbeek, P., Vincken, K.L., van Bochove, G.S., van Osch, M.J., van der Grond, J.: Probabilistic segmentation of brain tissue in mr imaging. NeuroImage 27(4), 795–804 (2005)
18. Fonov, V., Evans, A., McKinstry, R., Almli, C., Collins, D.: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage 47(suppl.1) (2009); S102 Organization for Human Brain Mapping 2009 Annual Meeting.
19. Fonov, V., Evans, A.C., Botteron, K., Almli, C.R., McKinstry, R.C., Collins, D.L.: Unbiased average age-appropriate atlases for pediatric studies. NeuroImage 54(1), 313–327 (2011)
20. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: elastix: A toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging 29(1), 196–205 (2010)
21. van der Lijn, F., de Bruijne, M., Hoogendam, Y., Klein, S., Hameeteman, R., Breteler, M., Niessen, W.: Cerebellum segmentation in mri using atlas registration and local multi-scale image descriptors. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, pp. 221–224 (2009)
22. Puspitasari, F., Volkau, I., Ambrosius, W., Nowinski, W.: Robust calculation of the midsagittal plane in ct scans using the kullbackleibler measure. International Journal of Computer Assisted Radiology and Surgery 4, 535–547 (2009)
23. Liao, C.C., Xiao, F., Wong, J.M., Chiang, I.J.: Automatic recognition of midline shift on brain ct images. Computers in Biology and Medicine 40(3), 331–339 (2010)

Identification of Malignant Breast Tumors Based on Acoustic Attenuation Mapping of Conventional Ultrasound Images

Sivan Harary and Eugene Walach

IBM Research - Haifa, Israel

Abstract. Although breast cancer imaging techniques continue to improve rapidly, about 75% of all breast biopsies turn out to be benign. These unnecessary biopsies are expensive and very stressful for the patients. In this paper we propose a new method for reducing the number of unnecessary biopsies. Our approach consists of transforming conventional ultrasonic images into corresponding attenuation maps. These maps are then analyzed, yielding automatic classification of malignant tumors. We provide a proof of concept for this approach by testing it on a benchmark of clinical images from three different image acquisition systems. Our tests show excellent sensitivity and specificity, indicating that up to four-fold reduction in the number of unnecessary biopsies may be possible. Moreover, we demonstrate the system robustness by working on all the images without any system-specific tuning.

Keywords: Acoustic Attenuation, Breast Cancer, Computer-Aided Diagnosis, Ultrasound Imaging.

1 Introduction

Worldwide, breast cancer comprises just under 30% of all diagnosed cancers in women. Mammography is currently the most common modality for screening and detecting breast cancer. However, a large portion of the breast lesions found in mammograms are benign. In order to improve the specificity, doctors often examine the suspicious lesions using ultrasound (US) imaging. Nevertheless, even when using both mammography and US, about 80% of the biopsies turn out to be benign. Clearly, the unnecessary biopsies cause both physical pain and emotional stress. They also result in a significant waste of health-care resources.

Accordingly, a great deal of effort has been devoted to improving breast cancer diagnostic tools. A technological review of commonly used methods and new experimental techniques was conducted in [1]. Many of the newly developed techniques are based on US due to its non-ionizing nature, low cost, and mobility. US is often used for guidance during the biopsy itself, so it is only natural to use it as a final diagnostic tool before inserting the needle. A review of US techniques was compiled in [2].

The improvements to diagnostic tools can be divided into two categories: enhancements to the imaging equipment and the introduction of computerized

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 233–243, 2013.

image analytics. The first category includes solutions such as Elastography [3], which produces images of the breast stiffness or strain by applying compression or vibration using US waves. Another approach is to introduce tomographic 3D US images, which provide a more comprehensive view of the tumor in question [4]. In the second category, there are several computer-aided diagnosis (CAD) systems. A survey of CAD systems for breast US was conducted in [5]. These systems typically compute a variety of breast image features and use a variety of classification techniques to distinguish between malignant and benign tumors. These features include the shape of the tumor, its texture, and sometimes acoustic properties. Unfortunately, in many cases, the efficacy of US CAD systems tends to be limited due to high dependency on the specific image acquisition system.

Our work focuses on a specific acoustic feature, namely the acoustic tissue attenuation. Studies, such as [6,7], show that acoustic attenuation measurements can distinguish between malignant and benign tissues, and can therefore be used as an effective basis for a CAD system. Tissue attenuation can be calculated using transmission US in a tomographic manner [4]. However, clinical US systems produce B-scans, which are based on backscattering rather than transmission. Consequently, the authors of [8] developed a system for attenuation estimation that uses B-scans but with an additional metal plate. Other methods that use only B-scans with no alteration of the hardware setting are available [7, 9, 10]. These techniques produce one global attenuation measure either for the entire breast or for a pre-specified region of interest (ROI).

In this paper we propose a new technique for the estimation of local acoustic attenuation. This method uses conventional B-scan images so there is no need to modify the image acquisition process or hardware. Moreover, we have no dependency on a specific US system. Rather than computing the average attenuation for the entire ROI, as described in [7,9], we calculate the local attenuation of each pixel in the region and create an attenuation map. This map is more informative than a global measurement and is therefore more effective for differentiating between malignant and benign lesions. The attenuation map can be presented as is to the doctor, similar to what is usually done with Elastography images. However, to reduce the burden on the examining doctor, we also introduce an automatic analysis of the attenuation map to classify the breast tumors. This analysis can be used either as a stand-alone CAD system or in combination with other CAD systems.

Testing the algorithm on a benchmark of clinical images showed excellent sensitivity and specificity. Moreover, it demonstrated the system robustness by working on images from three deferent image acquisition systems without any system-specific tuning. Although promising, these are only preliminary results on a moderate benchmark that provide only a proof of concept for our approach. Our future work will focus on more extensive testing and clinical trials of this method.

Our method for attenuation estimation draws on earlier techniques for attenuation mapping of the liver [11]. However, we introduce significant changes to the method in order to enable reliable estimation of attenuation in breast

tissue, which is highly inhomogeneous. In Section 2, we present both the attenuation estimation algorithm and the subsequent module for tumor classification. Section 3 presents the results of our tests on clinical data. Finally, Section 4 concludes this work.

2 The Method

In this section we describe our proposed system for classifying breast tumors based on local acoustic attenuation estimation. We apply our processing on the same B-scan images that are displayed on the physician's screen. This renders our technique transparent to the image acquisition hardware, making it very convenient for the users. Unfortunately, this also means that we work with inherently distorted images. One of the major sources of such distortions is the Time Gain Compensation (TGC). All state-of-the-art US systems use TGC to compensate for the loss of echo amplitude with depth [12]. Images produced using TGC are more uniform and have an effective attenuation of zero. Although these images are easier for doctors to interpret, the TGC distorts our attenuation maps and prevents reliable quantitative analysis. Nevertheless, the attenuation maps are still useful as they show the relative attenuation difference between healthy and malignant tissue areas.

Since the proposed method is aimed at assisting doctors in the evaluation of suspicious lesions, the segmentation of the lesion is preformed manually by the doctor. This approach is very common in US CAD systems; see for instance [13].

Accordingly, we propose the following process:

1. The doctor marks the suspicious area (ROI) on the image.
2. The attenuation map is estimated inside and around the ROI.
3. The attenuation map is analyzed by the CAD algorithm to determine whether the marked tissue is benign or malignant.
4. The results are superimposed on the original US image.

To accommodate the above process, our system consists of two main complementary modules, which are described in the following sections. The first module computes the attenuation map and the second one provides classification of the attenuation results.

2.1 Acoustic Attenuation Map Estimation

Our proposed algorithm for attenuation mapping is a modification of the algorithm proposed in [11] for attenuation mapping of the liver. In [11], the attenuation is estimated for each pixel by looking at a small surrounding block. The attenuation in this block is assumed to be uniform, except for outliers (i.e., pixels that significantly differ from the central one), which are removed. Accordingly, the block average attenuation is computed using the least squares method. This approach works quite well for relatively uniform tissues such as liver. For highly heterogeneous tissues, such as breast, the uniformity assumption no longer holds.

Therefore, our modified algorithm identifies for each block a uniform subregion (mask) with properties similar to that of the central pixel. The modified algorithm is presented in Table 1. Below is an elaboration of its steps.

Table 1. Creating the attenuation map

For each pixel in and around the ROI:
1. Define a block of size $L \times P$ around the pixel.
2. Define a mask (a subregion) of the block.
3. Estimate the attenuation in the block using only the pixels in the mask.
4. Assign the block's attenuation value to the central pixel.

Following [12], we express the intensity of the pixel in the nth column and mth row of the US image as follows:

$$E_{m,n} = E_0 \sigma_{m,n} \exp\Big(- 2\Delta \sum_{k=1}^{m-1} \alpha_{k,n} \Big), \tag{1}$$

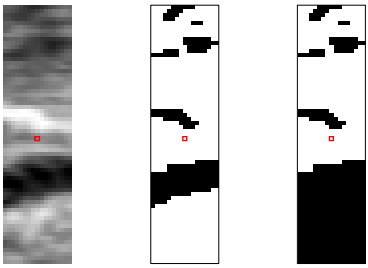
where E_0 is the initial amplitude, Δ is the size of the pixels, and $\sigma_{m,n}, \alpha_{k,n}$ are the backscattering and attenuation coefficients, respectively. Without loss of generality we assume that $\Delta = 1$. This is equivalent to simply changing the unit of measure of the attenuation.

We are interested in estimating $\alpha_{k,n}$ for each pixel in and around the ROI. To that end we assume that in a small vicinity of each pixel the attenuation and backscatter coefficients are constant. In order to define this vicinity, first we define a small block of constant size $L \times P$ around each pixel. For example, we use 65×17 pixels. If the attenuation was uniform in this block, then (1) would reduce to:

$$E_{j,i} = E_i \exp\Big(- 2 \sum_{k=1}^{j-1} \alpha \Big) = E_i e^{-2(j-1)\alpha}, \tag{2}$$

where $E_{j,i}$ is the intensity in the j,i pixel in the block, and α is the block's constant attenuation, which we are looking for. However, since the breast tissue is not homogeneous enough, the block's attenuation is not necessarily uniform. Therefore, instead of using all of its pixels, in the second step of the algorithm we identify a subregion (mask) of the block where the constant attenuation assumption is reasonable. Then, equation (2) is applied only to pixels on this mask.

We identify this mask in two stages. First, we find the mask for all in-range pixels, i.e., pixels in the block with intensity within 3dB proximity to the central pixel. Then, we remove peripheral blobs (connected nonzero pixels) from the mask. That is, any blob whose distance from the main blob exceeds 4 is removed. Note that we refer to 2D blobs in contrast to [11], where each column is processed separately. In case the remaining mask contains too few pixels, we can not rely on it for the estimation. Thus, we assign an attenuation value of zero, and move on to the next pixel. Fig. 1 presents an example for a block and its mask. As can be seen the mask indicates all pixels of the same tissue type as the central pixel.



(a) (b) (c)

Fig. 1. Block masking. (a) is the block defined for the central pixel (marked in red). (b) is the initial in-range mask. (c) is the final mask, after removing peripheral blobs.

Since the mask indicates a subregion with homogeneous attenuation, the model in (2) holds for all pixels indicated by the mask. Therefore, we define our cost function to be the sum over all those pixels of the squared differences between the actual and the theoretical intensities. That is:

$$C(E_i, \alpha) = \sum_{i=1}^{P} \sum_{j \in \Omega_i} \left(E_{j,i} - E_i e^{-2(j-1)\alpha} \right)^2, \tag{3}$$

where Ω_i is the ith column of the mask. The goal is to find E_i and α that minimize this cost function. Algorithms for solving similar least-squares problems exist in the literature, for example [14,15]. However, since this is a nonlinear problem, those algorithms tend to be iterative. To avoid the computational complexity involved, instead of using those methods we limit the solution to its first order approximation. This way the problem becomes linear and its solution can be expressed in a closed form. This limitation is appropriate under the assumption that $\alpha << 1$ such that the exponential in (3) is small. Since the attenuation value of most biological tissues is rarely above 0.01 nepers/pixel, in our case this assumption is quite valid. The first order approximation of (3) is:

$$C(E_i, \alpha) = \sum_{i=1}^{P} \sum_{j \in \Omega_i} \left(E_{j,i} - E_i(1 - 2(j-1)\alpha) \right)^2. \tag{4}$$

Let k_i denote the number of nonzero pixels on the ith column of the mask, and define:

$$a_i = \frac{2}{k_i} \sum_{j \in \Omega_i} (j-1) \quad , \quad b_i = \frac{1}{k_i} \sum_{j \in \Omega_i} E_{j,i} \quad , \quad c_i = \frac{1}{k_i} \sum_{j \in \Omega_i} (j-1) E_{j,i} - a_i b_i$$

Using this notation it is easy to see that setting to zero the derivative of $C(E_i, \alpha)$ according to E_i, and using fist order approximation yields:

$$\hat{E}_i = b_i - 2c_i\alpha. \tag{5}$$

(a) (b) (c)

Fig. 2. Mapping example. (a) is the original US image. (b) presents the doctor's annotation of the ROI, as a green line. (c) is the attenuation map of the ROI and its vicinity.

Substituting (5) into (4) and keeping only first order of α yields a simple square function of α, whose single global minimum is in:

$$\hat{\alpha} = \frac{\sum_{i=1}^{P} D_i^T B_i}{\sum_{i=1}^{P} B_i^T B_i}, \tag{6}$$

where $\Omega_{j,i}$ is the j,i pixel in the mask and

$$D_i = (b_i - E_{1,i}\,, \ldots,\, b_i - E_{L,i})^T \quad , \quad B_i = 2(\Omega_{1,i} c_i\,, \ldots,\, \Omega_{L,i}(c_i - (L-1)b_i)\ldots)^T$$

The estimated attenuation value $\hat{\alpha}$ is assigned only to the central pixel and not to the entire block. Performing this calculation for each pixel in and around the ROI yields the attenuation map. Fig. 2 presents an example of an US image and its associated attenuation map, estimated by the proposed method.

2.2 Attenuation Analysis

As discussed above, due to the TGC, attenuation maps are relative rather than absolute. In general, we expect zero attenuation for healthy tissue and higher attenuation values for malignant tumors. Fig. 3 presents some examples of attenuation maps of benign and malignant tumors. As can be observed in Fig. 3, malignant tumors have relatively large patches of high attenuation, while the overall structure is inhomogeneous. This fact fits well with our expectations based on the known morphology of cancerous tumors. Based on this insight we developed several features for classification between malignant and benign tumors, which are described in Table 2.

In order to quantify features number 1 and 2 in Table 2 we first smooth the attenuation map using the H-maxima transform, which suppresses mild maxima. For feature number 1, the regions of relatively high attenuation are identified by applying a fixed threshold (say of 10^{-3} nepers/pixel) on the smoothed map, while dismissing blobs with too small area. For feature number 2 we look for two regions with uniform attention in the smoothed map. The first region consists of all the pixels whose value equals the median value of the smoothed attenuation map. The second region is defined similarly as all the pixels whose value equals the median of the pixels that are not in the first region. Figure 4 presents an example for such uniform attenuation regions.

Table 2. Features of the attenuation map

	The Feature	Description
1	The portion of the tumor that is covered by regions of relatively high attenuation.	A small portion may indicate a benign tumor.
2	The portion of the tumor that is covered by uniform attenuation regions.	A large portion may indicate a benign tumor, since malignant tumors usually have inhomogeneous structure.
3	The maximal attenuation in the tumor.	Malignant tumors tend to have higher attenuation.
4	The portion of the tumor with attenuation close to the maximum from feature 3.	A small portion may indicate a benign tumor.
5	The portion of the tumor with negative attenuation.	A large portion may indicate a benign tumor.

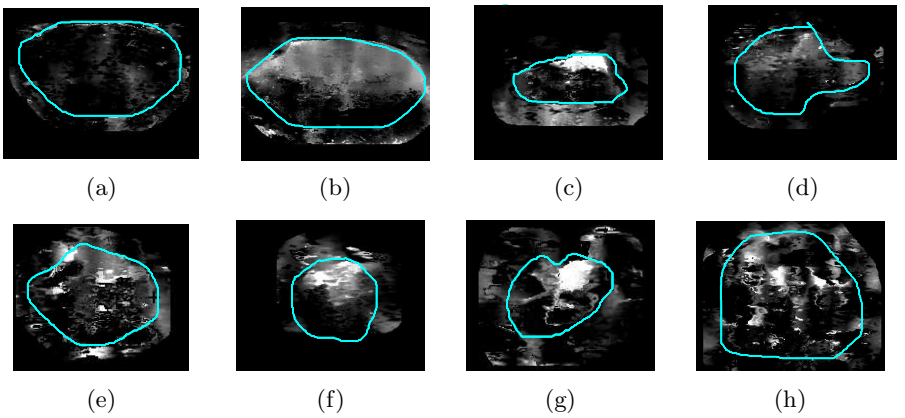
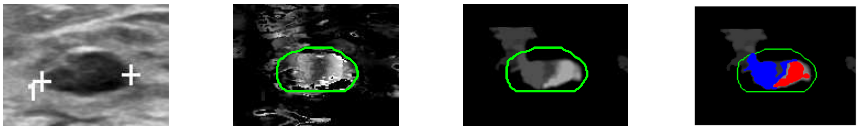


Fig. 3. Examples of attenuation maps. (a)-(d) are maps of benign tumors, (e)-(g) are maps of malignant ones. The gray-levels in the images correspond to the attenuation values. ROI marked in cyan.

In some benign cases the maximal attenuation value (feature number 3) is high due to artifacts. However, in those cases, this maximum value occurs only for small number of isolated pixels. Accordingly, we have introduced feature number 4 which examines the area where the attenuation is close to the maximum.

Another important feature is the size of the tumor, which can influence the assessment of the remaining features. For instance, our tests showed that the minimal acceptable intensity for feature number 3 and the parameter of the H-maxima transform should be smaller when dealing with smaller tumors. This is due the fact that for small tumors the attenuation estimation accuracy is lower.

Since the purpose of this work is to provide a proof of concept, in order to use these features for classification we had manually set thresholds and relation between them. Our future work will focus on using these features in a more sophisticated machine learning method, such as support vector machine classifier, on a much larger benchmark.



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

Fig. 4. Example for uniform attenuation regions. (a) is an US image of a benign tumor. (b) is the attenuation map. The green line marks the tumor's boundaries. (c) and (d) are the smoothed attenuation map, where the colored regions on (d) mark two uniform attenuation regions.

Special care should be taken of dark regions in the US image. Some doctors set the dynamic range such that those dark regions are saturated (i.e. the intensity level equals zero). Clearly, such regions yield false zero attenuation estimate. In order to mitigate this artifact we estimate the tumor attenuation and preform the classification based on non-saturated pixels only.

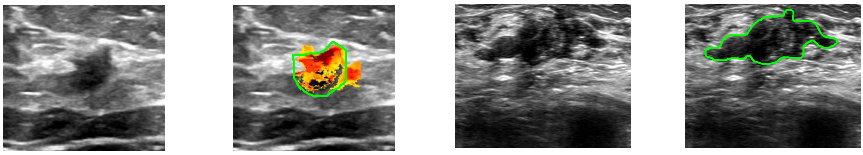
To facilitate the diagnostic process, we adopted a display combining the attenuation map and the original US image. If the given tumor is classified as benign, its attenuation map is discarded. Otherwise, the attenuation map is superimposed on the original US image with attenuation being proportional to the yellow-red color intensity. Fig. 5 presents display examples for both malignant and benign cases.

3 Results

To evaluate the efficacy of the above approach, we applied it to a diverse clinical benchmark including a total of 233 images of 80 different lesions. In all cases, the examining physician performed standard diagnostic procedure and decided to send the patient for a biopsy. We used the results of the biopsy as a gold standard to classify the benchmark into 46 benign and 34 malignant tumors. Most of the malignant tumors in our benchmark are IDC, while few are DCIS and ILC. Most of the benign lesions are fibroadnoma or fibrotic tissue, while the rest are: fat necrosis, PSH, cyst, tubular adenoma, hematoma, and abscess. The size of the tumors in the benchmark ranges from 3 to 40mm.

In should be noted that our database does not include mucinous (colloid) carcinoma which is a rare type of tumor that is, by nature, softer and has lower attenuation. Therefore, mucinous tumors are not likely to be detected by the proposed method. In order to detect such tumors our method should be applied together with a CAD system that processes the original US image, as apposed to the attenuation map.

We tested the robustness of our approach by acquiring the images using three different US systems, with no system-dependent tuning. All the images were processed using the aforementioned technique. Clearly, attenuation mapping provided additional information that, presently, is invisible to the examining physician. However, for the sake of objective quantitative testing, we limited our



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

Fig. 5. Display examples. (a) is an US image of a malignant tumor. (b) is the output image. The green line marks the ROI. The color indicates attenuation intensity ranging from yellow (mild) to dark red (high). (c) is an US image of a benign tumor. (d) is the output image; no attenuation map is visible, only the ROI is marked.

evaluation to the correctness of our malignant/benign classification. We performed our analysis on a per-case basis, which is crucial from the clinical point of view. In other words, the tumor was deemed to be benign only if the algorithm classified all its images as benign.

The summary of the results is presented in Table 3. As indicated by the results in the table, our algorithm identified all the malignant tumors - without exception. This corresponds to 100% sensitivity. Moreover, only 12 benign tumors were misdiagnosed yielding specificity of 74%. Thus, our results indicate that it may be possible to drastically cut down the number of unnecessary biopsies by a factor of 4 without any significant deterioration in the system sensitivity.

Table 3. Performance evaluation

	Images	Tumors	Marked as Malignant	Marked as Benign
Malignant	113	34	34	0
Benign	120	46	12	34

4 Discussion and Conclusions

We presented an algorithm for transforming conventional B-scan images into their corresponding attenuation maps. The algorithm is valid for inhomogeneous tissue such as breast and does not require any modifications in the US image acquisition hardware or software. The attenuation map can help physicians distinguish between benign and malignant tumors and thus reduce the number of redundant biopsies currently being carried out. The proposed scheme also includes an automatic classification algorithm, whose preliminary results indicate that the number of unnecessary biopsies may be rapidly reduced.

As future work, it is important to substantiate the results by significantly increasing the tested benchmark. A larger benchmark will allow us to replace the current CAD algorithm with a machine learning algorithm (for instance support vector machines), that will automatically find the best classification using our features. Moreover, we believe that our results can be further enhanced by

additional CAD features. One such example could be features aimed at detecting fibrotic cases, which are benign but highly attenuating. Accordingly, fibrotic cases cause quite a few false detections in our benchmark. Additional improvement would be the introduction of an automatic or semi-automatic tumor segmentation. This addition would make attenuation CAD fully automatic.

5 Patent Disclosure

The method described in this paper is the subject of two pending patent applications: 13/151303 and 13/558372.

Acknowledgments. The images used for this work were produced by Dr. Scott Fields from Hadassah Medical Organization in Jerusalem, and Dr. Ora Moskovitz from Or Breast Center in Haifa.

References

1. Nover, A.B., Jagtap, S., Anjum, W., Yegingil, H., Shih, W., Shih, W., Brooks, A.D.: Modern Breast Cancer Detection: A Technological Review. International Journal of Biomedical Imaging (2009)
2. Sehgal, C.M., Weinstein, S.P., Arger, P.H., Conant, E.F.: A Review of Breast Ultrasound. Journal of Mammary Gland Biology and Neoplasia 11(2) (2006)
3. Hall, T.: AAPM/RSNA Physics Tutorial for Residents: Topics in US: Beyond The Basics: Elasticity Imaging With US. Radiographics, 1657–1671 (2003)
4. Li, C., Duric, N., Huang, L.: Breast Imaging Using Transmission Ultrasound: Reconstructing Tissue Parameters of Sound Speed and Attenuation. In: BMEI (2008)
5. Chenga, H.D., Shana, J., Jua, W., Guo, Y., Zhang, L.: Automated Breast Cancer Detection and Classification Using Ultrasound Images: A Survey. Pattern Recognition 43(1), 299–317 (2010)
6. Goss, S.A., Johnston, R.L., Dunn, F.: Compilation of empirical ultrasonic properties of mammalian tissues II. Journal of The Acoustical Society of America 68(1), 93–108 (1980)
7. Chang, C.H., Huang, S.W., Li, P.C.: Attenuation Measurements for Ultrasonic Breast Imaging: Comparisons of Three Approaches. In: IEEE Ultrasonics Symposium, pp. 1306–1309 (2008)
8. Chang, C.H., Huang, S.W., Yang, H.C., Chou, Y.H., Li, P.C.: Reconstruction of Ultrasonic Sound Velocity and Attenuation Coefficient Using Linear Arrays: Clinical Assessment. Elsevier Ultrasound. Med. BioL. 33(11), 1681–1687 (2007)
9. Berger, G., Laugier, P., Thalabard, J.C., Perrin, J.: Global Breast Attenuation: Control Group and Benign Breast Diseases. Ultrasonic Imaging 12(1), 47–57 (1990)
10. Zheng, Y., Greenleaf, J.F., Gisvold, J.J.: Reduction of Breast Biopsies With a Modified Self Organizing Map. IEEE TNN 8(6), 46–56 (1997)
11. Walach, E., Shmulewitz, A., Itzchak, Y., Heyman, Z.: Local Tissue Attenuation Images Based on Pulsed-Echo Ultrasound Scans. IEEE Trans. Biomed. Eng. 36(2), 211–221 (1989)

12. Hughes, D.I., Duck, F.A.: Automatic Attenuation Compensation for Ultrasonic Imaging. Ultrasound in Medicine and Biology 23(5), 651–664 (1997)
13. Chen, D., Hsiao, Y.: Computer-aided Diagnosis in Breast Ultrasound. Elsevier Journal of Medical Ultrasound 16(1), 46–56 (2008)
14. McDonough, R., Huggins, W.: Best Least-Squares Representation of Signals by Exponentials. IEEE Trans. Autom. Control 13(4), 408–412 (1968)
15. Xie, W., Cai, C., Wang, Y.: Best Least Squares Solution for Prony Model. In: ISPACS, pp. 292–295 (2007)

What Genes Tell about Iris Appearance

Stine Harder[1], Susanne R. Christoffersen[1], Peter Johansen[2], Claus Børsting[2], Niels Morling[2], Jeppe D. Andersen[2], Anders L. Dahl[1], and Rasmus R. Paulsen[1]

[1] Technical University of Denmark
DTU Informatics - Informatics and Mathematical Modelling
2800 Lyngby, Denmark
[2] University of Copenhagen
Section of Forensic Genetics - Department of Forensic Medicine
Faculty of Health and Medical Sciences
2100 Copenhagen O, Denmark

Abstract. Predicting phenotypes based on genotypes is generally hard, but has shown good results for prediction of iris color. We propose to correlate the appearance of iris with DNA. Six single-nucleotide polymorphisms (SNPs) have previously been shown to correlate with human iris color, and we demonstrate that especially one of the six SNPs are correlated with iris appearance. To perform this analysis we need a method to model the iris appearance, and we suggest an iris characterization based on a bag of visual words, which gives us a similarity measure between images of eyes. We have a dataset of 215 eye images with corresponding SNP types, where the image of the iris has been segmented. We perform two experiments based on the iris characterization. An agglomerative clustering is performed and the result is that one SNP – rs12913832 (HERC2) is highly correlated with the image clustering. Furthermore subspace projections are performed supporting that this SNP is very important for eye color expression. With the suggested image characterizations we are able to investigate the correlation between the phenotypic iris appearance and specific SNPs. This has potential for further investigation of the relation between DNA and iris appearance, especially with focus on iris texture.

Keywords: Iris color, Iris texture, Image analysis, Image clustering, Canonical discriminant analysis, DNA.

1 Introduction

Predicting complex human phenotypes from genotypes has great potentials in application areas like personalized medicine [2, 6] or forensic genetics [7]. Personalized medicine does already exist for monogenetic disorders such as Huntington disease [6], but finding the etiology of more complex diseases is not an easy task. Liu et al. [11] did however demonstrate that genetic prediction of complex phenotypes is possible. Liu et al. [11] investigated 37 SNPs, representing all currently known genetic variants with statistically significant eye color association, and

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 244–253, 2013.

found that six SNPs were the major predictors. The six SNPs are rs12913832 (HERC2), rs1800407 (OCA2), rs12896399 (SLC24A4), rs16891982 (SLC45A2), rs1393350 (TYR), and rs12203592 (IRF4). Prediction of human phenotypes from genotypes is of large interest in forensic genetics, where externally visible characteristics (EVCs) could be used as a "biological witness" in forensic cases. The cases of interest are e.g. when a DNA profile from a crime scene does not match either the possible suspects or DNA profiles in the criminal database. Then it would be a great advantage to be able to predict the appearance of the suspect.

The six SNPs found by Liu et al. [11] have also been used by Walsh et al. [15] to build a tool, called IrisPlex, for iris color prediction. Their investigations also revealed that rs12913832 is the main determinant for blue or brown colored eyes. This SNP is however not a vey precise predictor for the human iris color, because it varies continuously from the darkest brown to the lightest blue and is not clearly separable into the discrete expressions of the investigated SNPs. Our work is a step towards a more precise prediction of the human eye appearance based on DNA by investigating the correlation between our proposed image characterisations and the six SNPs found by Liu et al. [11]. This approach avoids subjective evaluation of iris color, partitioning of iris color into classes and it enables us to investigate overall iris appearance including iris structures.

2 Data

The study was approved by the Danish Ethical Committee of the Capital Region (H-4-2009-125) for samples conducted at Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, or as part of the Danish Blood Donor Study for samples conducted at the Blood Bank, Glostrup Hospital. The data consist of 215 high resolution eye images and corresponding DNA types. The images were subsampled to a spatial resolution of 639×426 pixels in RGB color.

The camera was equipped with a Twin flash, which ensured precise and repeatable acquisition and uniform illumination. The Twin flash gave two overexposed square regions in the iris and pupil area, which were removed in our segmentation procedure. Image examples are shown in Fig. 1.

From the DNA sample the SNPs [11]: rs12913832, rs1800407, rs12896399, rs16891982, rs1393350 and 12203592 were typed. The SNPs have three expressions or layers, i.e. there are three types of each SNP, however not all combinations were identified.

Fig. 1. Eye image examples

3 Iris Characterization

To construct an image characterization based on the iris we first need to segment the image. After the segmentation we preform a radial image transformation giving us the iris as a square image. We represent this image both by its color as a histogram of RGB values and as a combined histogram of color and image descriptors – a bag of visual words (BOW), which also contains information about the iris texture.

Iris segmentation. The iris segmentation is performed by fitting a circle to the inner and outer boundaries of the iris and fitting a spline to the upper and lower eyelid boundaries, similar to the method proposed by Daugman [4].

The eye images have a large gradient from the pupil to the iris and also from the iris to the sclera. Utilizing this we propose an optimization scheme where we look for a circle with maximum radial gradient. For a circle with center at $\mathbf{x} = [x, y]^T$ and radius r, the image intensity of a point is given by $f_\theta(\mathbf{x}, r) = I([x + r\cos(\theta), y + r\sin(\theta)]^T)$, where I is the image intensity and θ is an angle. We estimate the radial gradient $\frac{df_\theta}{dr}$ using finite differences. We wish to find the parameters of the circle that maximizes the gradient along the circle

$$\arg\max_{\mathbf{x},r} \int_0^{2\pi} \left| \frac{df_\theta}{dr} \right| d\theta \approx \arg\max_{\mathbf{x},r} \sum_{i=0}^{n-1} \left| \frac{df_{i\Delta\theta}}{dr} \right|, \tag{1}$$

where $\Delta\theta = \frac{2\pi}{n}$. The solution to Eq. 1 is found using a coarse to fine sampling strategy. The search for the center coordinate is performed by sampling in a regular grid and choosing a finer sample grid around the position with the highest gradient sum. This is repeated until single pixel accuracy is obtained. The radial search is performed by calculating the gradient sum for a number of equally spaced radii and then limiting the search area to a region around the radius with the highest gradient sum. The search for radius and center coordinate is performed simultaneously and the process is continued until single pixel accuracy is obtained. To avoid detection of fine structures such as eyelashes we initially smooth the image using a Gaussian kernel with a standard deviation of $\sigma = 3$ and we use $n = 36$ sample points.

While the optimization works well for the pupil, it has problems with the iris since the eyelids often covers part of the iris. To avoid this we choose only to sum the gradient in the intervals $\theta \in [-\pi/4, \pi/4] \cup [3\pi/4, 5\pi/4]$, which is similar to the approach in [3]. Therefore $n = 20$ for the iris. The small square spots from the flash are removed using a simple threshold.

Eyelid Boundaries. The eyelid boundaries are located using a Markov Random Field (MRF) based segmentation [10]. The segmentation is performed on the second HSV component of the eye images, since this color component shows the

largest difference in pixel value between skin and inner eye regions. The chosen labels are sclera, eyelashes, skin and iris, and the iris label is divided into blue and brown. The statistics used in the MRF segmentation is calculated from manually annotated exemplars.

Let g be a label configuration and I the image. We estimate the posterior energy, $E(g|I)$, as

$$E(g|I) = \sum_i \left(\sum_{j \in \mathcal{N}_i} \delta(g_i, g_j) + \frac{1}{2} \log(\sigma_l^2) + \frac{(I_i - \mu_l)^2}{2\sigma_l^2} \right) (-log(Q(g_i)), \qquad (2)$$

where i is a site (pixel position) in the image. The neighborhood $\mathcal{N}_i$ is the four nearest sites. We assume the pixel intensities of the different labels to be normally distributed with $N(\mu_l, \sigma_l)$. Q is a probability matrix modeling the prior knowledge of position of the different classes based on 50 manually annotated images. The prior, $\delta(g_i, g_j)$, is modeled by

$$\delta(g_i, g_j) = \begin{cases} \beta, & \text{if } g_i \neq g_j \\ 0, & \text{if } g_i = g_j \end{cases}. \qquad (3)$$

The segmentation problem using MRF is solved using Graph Cut with α-expansion [8]. The three parameters in Eq. 2 are chosen experimentally. The final eyelid boundaries are found by fitting splines to the segmentation of the upper and lower eyelid boundaries.

The final iris image (the iris map) is obtained by radially sampling the segmented iris along lines going through the center of the pupil. The samples are chosen equidistantly along these lines from the circle fitted to the inner boundary of the iris to the circle on the outer boundary. The lines are sampled tangentially at equal angle steps. The number of angular steps is 720 and the number of radial steps is 120. The resulting image is therefore 120×720. We employ a mask to avoid the eyelid and highlight regions in our analysis. A result of the entire iris extraction procedure can be seen in Fig. 2. A few eye images was not precisely segmented, so for these samples we adjusted the segmentation manually.

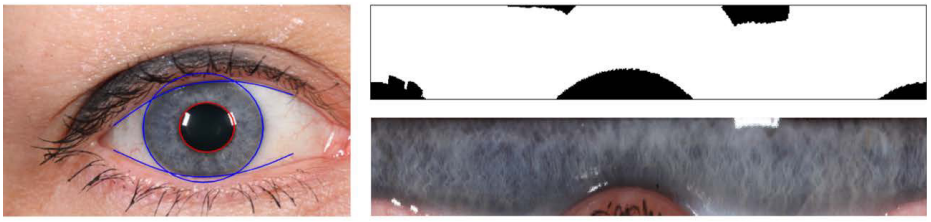


Fig. 2. Left: Final result for the detection of the eyelid boundaries, Right top: mask for iris map, Right bottom: iris map

Color characterization. The first image characterization is purely based on color. From a representative set of training images containing both blue and brown irises, we partition the three dimensional RGB color space into 852 color bins. This is done using a hierarchical binary separation of color channels based on the same principle as building a balanced kd-tree [1]. Hereby we ensure that each color bin contains approximately the same number of samples. The 852 color bins are used for constructing iris histograms, which are normalized using the L_1 norm.

BOW iris characterization. The second iris characterization is based on a bag of visual words (BOW) [13] in addition to color. We chose to use DAISY features [14] to represent the texture appearance because these features are computed in all pixels very efficiently and have shown similar performance as SIFT [12]. The resulting DAISY feature is a 100 dimensional vector $\mathbf{d}_d = [d_1, ..., d_{100}]^T$ and we represent the RGB value as the vector $\mathbf{d}_c = [R, G, B]^T$. These two vectors are L_2 normalized and concatenated to obtain a 103 dimensional descriptor vector $\overline{\mathbf{d}} = \sqrt{\frac{1}{2}}[\overline{\mathbf{d}}_d, \overline{\mathbf{d}}_c]^T$. We obtain a dictionary of visual words using k-means clustering into 400 clusters with cluster centers as visual words from 40000 randomly chosen training samples. Each image feature is labeled by assigning it to the nearest visual word using the L_2 norm. An overview of the process can be seen in Fig. 3. To include spatial information in the image characterization we perform a spatial weighting of the visual words using Gaussians distributed at 12 positions as illustrated in Fig. 4. Based on this representation we can estimate the similarity of iris maps as histogram differences.

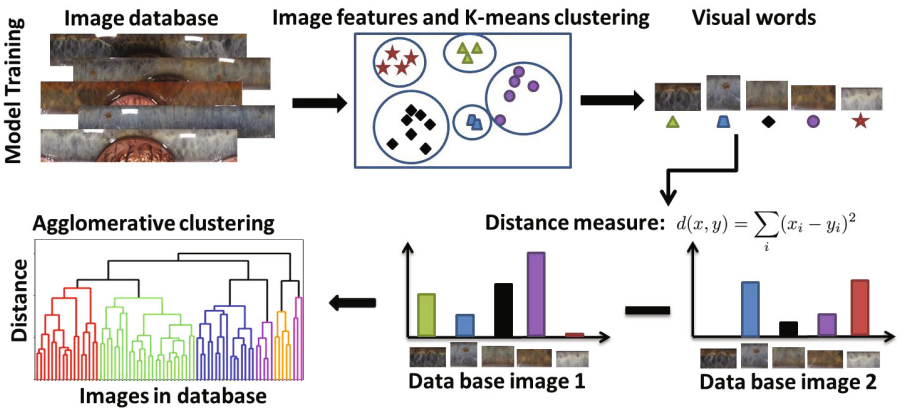


Fig. 3. Illustration of the bag of words model for images along with the images clustering procedure

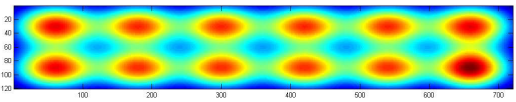


Fig. 4. Spatial Gaussian weighting used for building the explanatory histograms

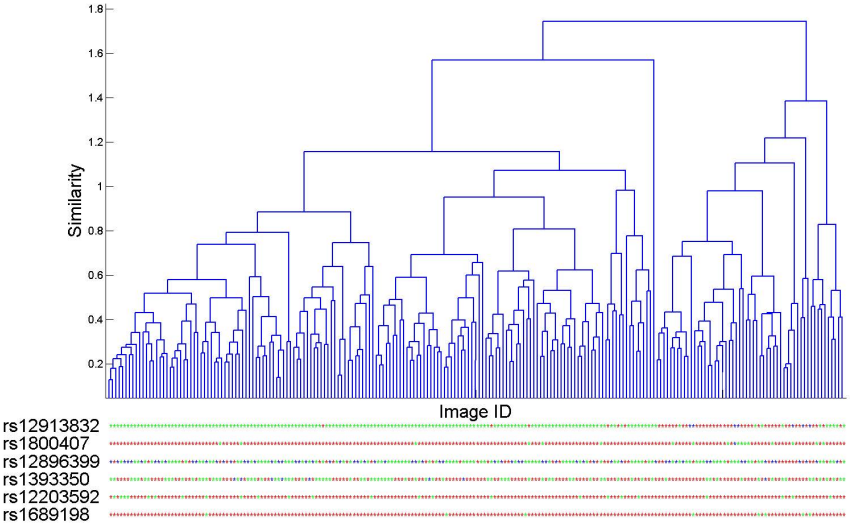
4 Analysis and Results

Based on the color and combined color and texture characterization our aim is to compare the genotypes with the visual appearance. We perform two explorative experiments – the first based on hierarchical agglomerative clustering [9] of the image characterizations and compare this to the genetic expression, and the second is subspace projection based on canonical discriminant analysis [5].

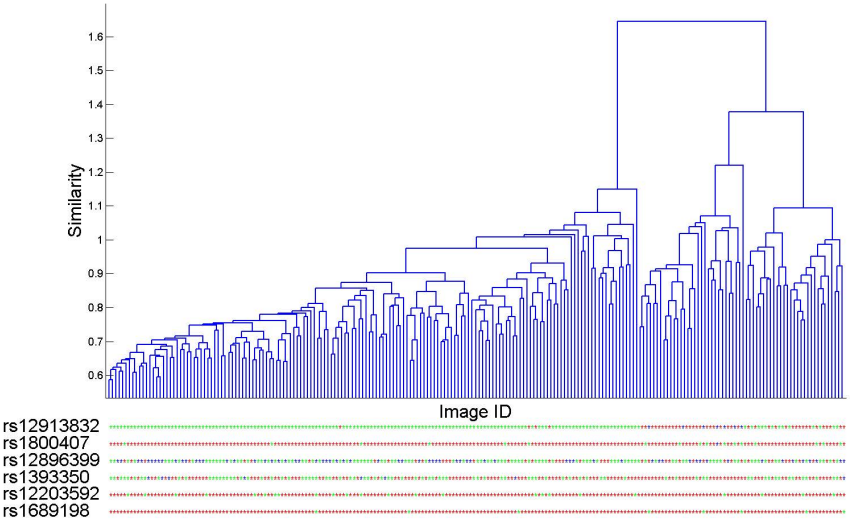
Agglomerative Clustering. The image characterizations based on respectively color and combined color and texture are used to generate an agglomerative clustering. The agglomerative clustering is based on the L_2 distance between the histograms, and similar histograms are clustered together. The result is a dendrogram for respectively color and combined color and texture, as seen in Fig. 5. The lower part of the figures show the six SNPs defined in Sec. 2. Each bin in a dendrogram, corresponding to an eye image, has the six SNPs represented with color below. The different layers of the SNPs are colored with respectively red, green or blue. It is clear that the image clustering results for both color and a combination of color and texture corresponds very well with the genotypes of rs12913832.

Subspace projections. We have performed a subspace projection of the iris descriptor histograms shown in Fig. 6 using principal component analysis (PCA) and canonical discriminant analysis (CDA). The first principal or canonical direction is horizontal and the second is vertical. We treat the three genetic expressions of rs12913832 as classes. To account for the rank deficiency of the estimated covariance matrices we initially perform a data projection using PCA where we keep the first 100 principal components. This corresponds to 98.8% of the variance of the color descriptor and 86.2% of the variance of the combined color and texture descriptor.

The iris maps are overlapping and in the overlapping regions their color is averaged. This gives a blurring effect where the iris maps are overlapping. Using PCA the iris maps are mainly sorted according to color, whereas the CDA clearly separates the iris maps into the three distinct groups according to the three expressions of the rs12913832 SNP.

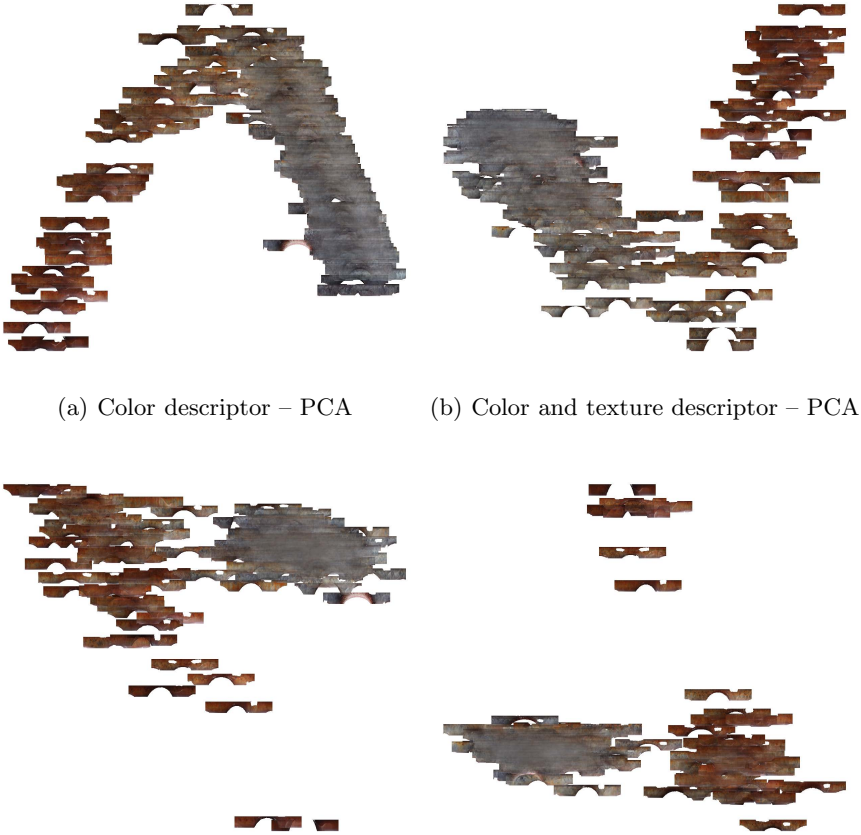


(a) Color descriptor



(b) Color and texture descriptor

Fig. 5. Top part shows a dendrogram obtained using agglomerative clustering based on (a) the color descriptor and (b) the combined color and texture descriptor histograms. The dendrograms are based on the L_2 distance between the histograms explained in Sec. 3. Below each bin in a dendrogram the expressions of the six SNPs are represented by a color.



(a) Color descriptor – PCA (b) Color and texture descriptor – PCA

(c) Color and texture descriptor – CDA (d) Color descriptor – CDA

Fig. 6. Iris maps projected to the first two (a,b) principal components and (c,d) the first two canonical dimensions based on (a,c) the color descriptor histograms and (b,d) the combined color and texture descriptor histograms. Classes in the canonical discriminant analysis are based on the genetic expression of rs12913832.

5 Discussion

The agglomerative clustering based on the image characterization for respectively color and a combination of color and texture showes a large correlation with rs12913832. rs12913832 seems to be the major explanation for iris appearance and the contribution from the remaining five SNPs seem to be minor. The dendrograms obtained for color and combined color and texture reveal very similar results. This indicates that iris color is the main contributor to the image characterization and that the texture is not expressed in the analysed SNPs.

We only present the subspace projection analysis based on rs12913832 as class labels. The other SNPs also separate nicely using CDA, but with PCA only rs12913832 was sorted according to its expression. This further underlines the observation from the clustering experiment, that rs12913832 is very important for the iris color, but there is only little influence on the iris texture. Our investigations so far do not support that any of the included SNPs influence iris texture, but the texture is an important element in the iris appearance. However with the suggested image descriptors we are able to analyze this further.

The radial transformation performed after the iris extraction consist of a radial sampling, where the inner part of the iris is sampled more densely than the outer part. The sampling procedure entail that the features close to the pupil will have a greater impact than features located in the periphery of the iris. This property is similar for all eye images and the distance measure is therefore not affected. A great advantage with the sampling method is that the new coordinate system becomes invariant to the size of the iris and to pupil dilation as explained by Daugman [4].

6 Conclusion

We have analyzed the genetic expression of six SNPs in relation to iris appearance. To perform this investigation, we have suggested a representation of the iris appearance based on color and texture from a radial warped eye image. The image representation is a histogram of image features. We perform an explorative analysis in the form of an image based clustering and a subspace projection. Our investigations show that especially rs12913832 is closely correlated with the iris color, whereas the other SNPs show a less clear pattern. We do not see a relation between iris texture and the investigated SNPs, but our descriptors clearly show that texture is an important part of the iris appearance. The proposed methodology enables us to investigate this further.

Acknowledgments. We thank Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen and the Blood Bank, Glostrup Hospital for collecting data.

References

[1] Bentley, J.L.: Multidimensional binary search trees used for associative searching. Communications of the ACM 18(9), 509–517 (1975)
[2] Brand, A., Brand, H., et al.: The impact of genetics and genomics on public health. European Journal of Human Genetics 16(1), 5–13 (2007)
[3] Daugman, J.G.: High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(11), 1148–1161 (1993)
[4] Daugman, J.G.: How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology 14(1), 21–30 (2004)

[5] Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Human Genetics 7(2), 179–188 (1936)
[6] Janssens, A.C.J.W., Van Duijn, C.M.: Genome-based prediction of common diseases: advances and prospects. Human molecular genetics 17(R2), R166–R173 (2008)
[7] Kayser, M., Schneider, P.M.: Dna-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations. Forensic Science International: Genetics 3(3), 154–161 (2009)
[8] Kolmogorov, V., Zabin, R.: What energy functions can be minimized via graph cuts? IEEE Transactions on Pattern Analysis and Machine Intelligence 26(2), 147–159 (2004)
[9] Krishnamachari, S., Abdel-Mottaleb, M.: Image browsing using hierarchical clustering. In: Proceedings of IEEE International Symposium on Computers and Communications, pp. 301–307. IEEE (1999)
[10] Li, S.Z.: Markov random field modeling in image analysis. Springer-Verlag New York Inc. (2009)
[11] Liu, F., van Duijn, K., Vingerling, J.R., Hofman, A., Uitterlinden, A.G., Janssens, A., Kayser, M.: Eye color and the prediction of complex phenotypes from genotypes. Current Biology 19(5), R192–R193 (2009)
[12] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
[13] Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: Proceedings of the Ninth IEEE International Conference on Computer Vision, pp. 1470–1477. IEEE (2003)
[14] Tola, E., Lepetit, V., Fua, P.: Daisy: An efficient dense descriptor applied to wide-baseline stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 32(5), 815–830 (2010)
[15] Walsh, S., Lindenbergh, A., Zuniga, S.B., Sijen, T., de Knijff, P., Kayser, M., Ballantyne, K.N.: Developmental validation of the irisplex system: determination of blue and brown iris colour for forensic intelligence. Forensic Science International: Genetics 5(5), 464–471 (2011)

Robust Dense Endoscopic Stereo Reconstruction for Minimally Invasive Surgery

Sylvain Bernhardt[1], Julien Abi-Nahed[2], and Rafeef Abugharbieh[1]

[1] Biomedical Signal and Image Computing Lab, University of British Columbia, Vancouver, Canada
[2] Qatar Robotic Surgery Centre, Qatar Science and Technology Park, Qatar
sylvainb@ece.ubc.ca

Abstract. Robotic assistance in minimally invasive surgical interventions has gained substantial popularity over the past decade. Surgeons perform such operations by remotely manipulating laparoscopic tools whose motion is executed by the surgical robot. One of the main tools deployed is an endoscopic binocular camera that provides stereoscopic vision of the operated scene. Such surgeries have notably garnered wide interest in renal surgeries such as partial nephrectomy, which is the focus of our work. This operation consists of the localization and removal of tumorous tissue in the kidney. During this procedure, the surgeon would greatly benefit from an augmented reality view that would display additional information from the different imaging modalities available, such as pre-operational CT and intra-operational ultrasound. In order to fuse and visualize these complementary data inputs in a pertinent way, they need to be accurately registered to a 3D reconstruction of the imaged surgical scene topology captured by the binocular camera. In this paper we propose a simple yet powerful approach for dense matching between the two stereoscopic camera views and for reconstruction of the 3D scene. Our method adaptively and accurately finds the optimal correspondence between each pair of images according to three strict confidence criteria that efficiently discard the majority of outliers. Using experiments on clinical in-vivo stereo data, including comparisons to two state-of-the-art 3D reconstruction techniques in minimally invasive surgery, our results illustrate superior robustness and better suitability of our approach to realistic surgical applications.

Keywords: stereovision, rectification, dense matching, 3D reconstruction, stereo camera, stereo vision, partial nephrectomy, augmented reality, robotic assisted surgery.

1 Introduction

The past decade witnessed an ever-increasing number of reports on robot-assisted surgical interventions where the surgeon remotely controls a robot that reproduces the motion of his/her hands on laparoscopic tools. Medical robots, such as the *da Vinci Surgical System* (Intuitive Surgical, Inc., Sunnyvale, CA, USA),

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 254–262, 2013.

have for example been widely used in renal surgery due to similar or even better clinical outcomes than those of standard procedures [1].

The fact that the surgeon's view of the operated scene is digitized via a stereo camera has made augmented reality (AR) in minimally invasive surgery (MIS) an very active research area since the early 2000s [2][3]. The aim is to grant the surgeon the ability to see "beyond" the visible surface by overlaying visual information from other available intra-operative and pre-operative data onto the endoscopic camera feed. However, registration of such data with the 3D scene remains a difficult problem since, particularly in abdominal MIS, the environment is mostly composed of soft tissue and organs that significantly deform due to the surgeon's actions as well as patient breathing and cardiovascular activity. One approach to solving this problem is to use the stereo stream from the camera to perform dense matching and provide a 3D model of the surgical scene that can then serve as a registration base for the other imaging data, e.g. as in [4]. Many methods for dense stereo matching have been proposed over the last decades [8][9], however, there are two main distinctions between the kind of data typical in MIS and the traditional reference datasets for dense stereo matching, such as the Middlebury images [10]. The first is that our binocular camera provides a video output, i.e. sequences of images with very little differences between two successive frames. Therefore, temporal smoothness gains more importance and can be enforced. Furthermore, since the MIS scenes are generally not static when captured, a significant amount of motion blur is typically introduced, which makes the stereo matching problem more difficult.

The second main difference is related to the content of the images. Datasets traditionally used in computer vision studies represent static scenes with a variety of rather simply shaped objects laid out at different depths. Moreover, the surfaces are most of time matte and the lighting is uniform, which does not induce complex lighting artifacts. On the other hand, intra-abdominal tissue is soft and presents complex reflections, due to the non-Lambertian nature of the surfaces, as well as irregular shapes, highly variable textures and various distortion . Additionally, there is a constant presence of surgical tools that severely occludes the scene with textureless plastic or highly reflective metallic parts (see figure 1). Overall, image sequences in MIS are very challenging to reconstruct and defy the robustness of current stereo matching techniques.

Few methods at dense reconstruction of stereo endoscopic images have been proposed. In [2], a method was presented for detecting and virtually removing the tools from the reconstructed scene. Later [12], Vagvolgyi et al proposed a method to overlay a kidney model onto the stereo display by registering the model to the kidney surface reconstructed from stereo data. More recently [5], Stoyanov et al presented a method to perform near real-time stereo reconstruction in MIS based on belief propagation. A similar work has been recently proposed in [6] where hybrid recursive matching was used.

All these previous methods are based on existing stereo matching algorithms that have not been designed for MIS data. For example, they all try to enforce spatial smoothness constraints, which is supposed to ensure homogeneous

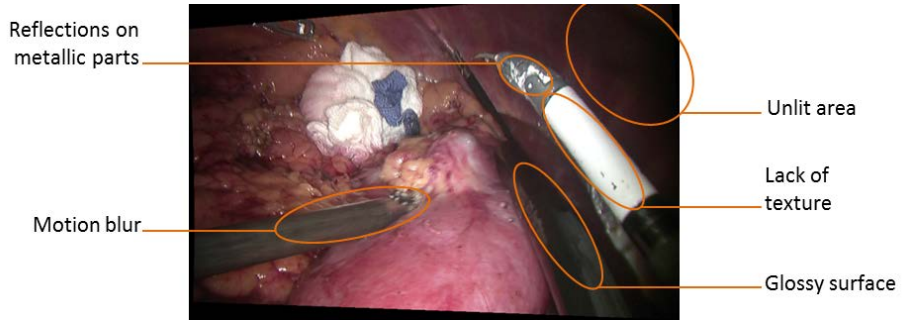


Fig. 1. Example image depicting a scene captured during a partial nephrectomy procedure. Commonly encountered artifacts are highlighted in orange.

disparities in regular areas. However, practical MIS images are more challenging, therefore mismatches are more prone to happen. If spatial smoothness is enforced, this will tend to spread errors into little clusters of homogeneous outliers, which are harder to discard than isolated ones. Also, the risk of getting rid of actual inliers is greater if they are in a group since a single pixel standing out from the rest of the depth map does not represent a realistic scene in world space. The work presented in this paper aims to address such issue by providing a 3D model of the surgical scene with emphasis on accuracy and robustness. The primary goal is to discard outliers and yet provide enough information across all frames of the video stream such that registration is still possible. To achieve this, we first detect only the most reliable matches by enforcing a series of strict criteria reflecting certainty of matching. We then enforce limited spatial smoothness to handle the few isolated outliers that still survive.

2 Methodology

2.1 Pre-processing

The output from our surgical binocular camera is an interlaced high definition video (1080i). To alleviate the problem of jagged edges in the de-interlaced images, each extracted frame is downsampled by a factor 2 down to 960×540 pixels. For each pair of frames, we then perform a sparse matching using the SIFT descriptor [13]. Occasional mismatches are discarded during the robust calculation of the fundamental matrix F using RANSAC as in [7]. The epipoles are calculated from F and used to rectify the images. This process aligns the two images into the same plane in the world space (see figure 2a-b). Then, according to the laws of epipolar geometry, every feature or pixel in one frame has its correspondence on the same row in the other frame, which greatly facilitates the matching (see figure 2b). We use polar rectification, as it is simple and guarantees minimal distortion of the images [14]. The matching of a feature yields the disparity d which is inversely proportional to the feature depth Z in

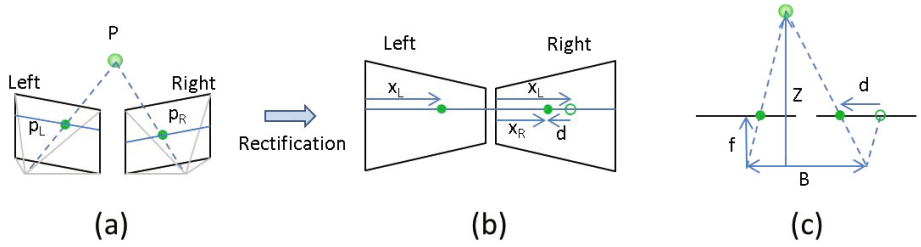


Fig. 2. Rectification and relation between depth and disparity. (a) shows the projection of a point P in world space onto the two image planes in p_L and p_R. In (b), the vertical alignment of this point in the rectified images gives the disparity d between its left and right locations, respectively x_L and x_R. The disparity d is also inversely proportional to the depth, as shown in (c) the top view of (b).

the world space (see figure 2c and equation 1, where f is the focal length and B the baseline between the two optical centers).

$$Z = -\frac{Bf}{d} \tag{1}$$

Finally, the images are converted to grayscale by averaging the three color channels with different weights as recommended in [15] where it was shown that the use of color in dense stereo matching is not beneficial.

2.2 Robust Matching via Confidence Criteria

Our dense matching is based on calculating similarity between patches from the left and right images using normalized-cross correlation (NCC) as a metric, which is efficient even in the presence of brightness change. For a given point and window in the left image, the similarity profile is calculated across the same row in the right image, bounded by a range centered at the same position. The local maxima represent the location of candidates for matching with the best candidate chosen as the one with the highest similarity value. As robustness is paramount in our application, we ensure that each point pair matching satisfies strict confidence criteria that reflect three metrics of uncertainty in the matching in our approach:

First, blurry, unlit and textureless parts of the image present very little structural information, which makes the matching difficult if not impossible. To mitigate this problem, a simple and effective gradient dispersion metric γ is considered, in order to estimate the spatial structure. Let γ_L be its value for the considered patch I_p and γ_R the one for the best candidate patch I_c. Our **first criterion** is that both of these values have to be greater than a certain threshold γ_{min} (see equation 2). If this condition is not met, the matching is declared too risky.

$$\gamma_L = stdev(\nabla^2 I_p) > \gamma_{min} \quad \text{and} \quad \gamma_R = stdev(\nabla^2 I_c) > \gamma_{min} \tag{2}$$

Second, the quality of a matching also appears in the difference of similarity score in the NCC profile between the two highest peaks, as illustrated by figure 3. The blue curve represents the matching from the left to right images, and the green one from right to left. The graph (a) of the figure reflects a patch of size of 7x7 pixels. As can be easily seen, the profile presents many peaks of approximatively the same score, which means the content of the patch is not discriminative enough and hence the window may be too small. On the graph (b) of the figure, the window size is 13x13 pixels, which allows the patch to contain more complex patterns. As a result, the correct solution stands out in the profile since the gap between the best and other peaks is significant. Our **second criterion** is thus that this difference δ has to be beyond a threshold δ_{min}. If this condition is not satisfied, the matching is declared not discriminative enough.

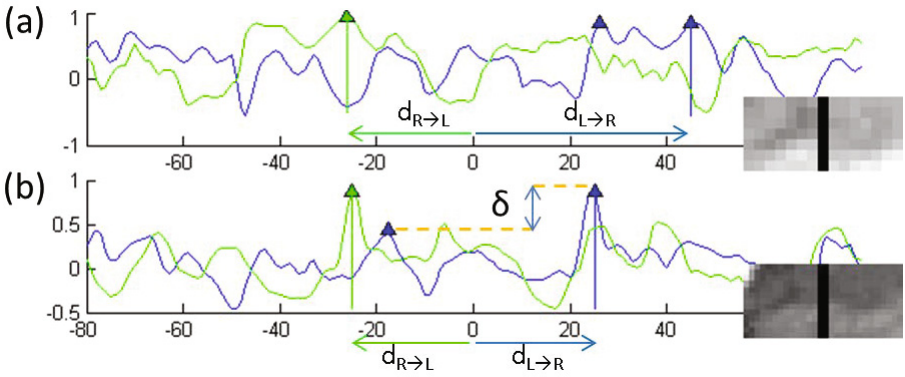


Fig. 3. Two example NCC profiles for the same point but with two different window sizes. The blue curve represents the matching from the left to right images, and the green one from right to left. On the y-axis is the similarity score and on the x-axis the disparity. The best peak is designated by a vertical line and its score difference with the second best peak is δ. The disparities of the best candidates from left to right and right to left are also displayed as $d_{L\to R}$ and $d_{R\to L}$, respectively. (b) here shows better discrimination.

Third, let us consider $d_{L\to R}$ and $d_{R\to L}$. If this correspondence is correct, then the inverse matching (from right to left) should yield the dual result: $d_{R\to L} = -d_{L\to R}$. Therefore, if both previous criteria are satisfied, then inverse matching is performed starting from the best candidate in the right image. Our **third criterion** is thus that $d_{R\to L}$ and $d_{L\to R}$ should cancel out within a threshold ϵ (see equation 3). If this is not true, the matching is then considered incorrect.

$$|d_{L\to R} + d_{R\to L}| < \epsilon \tag{3}$$

In case of failure in satisfying the above third criterion, the window is increased by a step dw in hope of providing a more discriminative matching. If the window

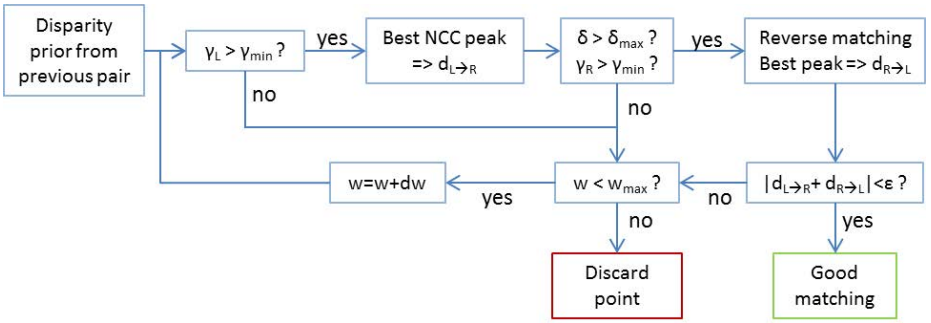


Fig. 4. Block diagram of our proposed dense matching method

size reaches a threshold w_{max} and still none of the criteria are satisfied, then this particular point in the image is discarded (see figure 4). This matching method is iterated through the image until completion.

2.3 Post-processing and Temporal Smoothness

Even though the previous criteria are very restrictive, it is still possible to have outliers slipping through. Fortunately, since spatial smoothness has not been enforced, the few surviving outliers are most of the time isolated and easy to identify. Therefore, our post-processing step consists of finding isolated pixels that are significantly different from their neighborhood. More specifically, in a window of size Δw around each point that has been matched, we consider the number N of disparities whose difference with the actual point disparity is less than a threshold Δd. If the ratio $N/\Delta w^2$ is smaller than a threshold μ, then the matching for this point is discarded. Once outliers are weeded out, one or two-pixel wide holes are filled with the median of their surrounding values. It is important to note that no other attempt at filling larger empty areas is needed, as subsequent frames will fill larger gaps (i.e. uncertain matching areas) locally across time. For computational considerations and due to the highly reliable matches, the disparities found for one image pair are used as priors for the next pair of frames by reducing the search range of a point matching around its previous disparity value, thus enforcing the temporal smoothness of the depth evaluations.

3 Results and Discussion

All experiments were carried on frames extracted from in-vivo videos recorded by our own high definition endoscopic stereo camera during four different partial nephrectomies assisted by a *da Vinci* robot. The sequences have a resolution of 1920×1080 pixels at a frame rate of 25 fps and the images color space is YCbCR.

Our experiments have shown that $\gamma_{min} = 0.5$, $\delta_{min} = 0.1$, $\epsilon = 5$, $\Delta d = 4$, $\Delta w = 5$, $\mu = 0.4$, a square window of initial width $w_{min} = 7$, maximum width $w_{max} = 30$ and step $dw = 4$, for frames of size 960×540, yield good results for a wide range of MIS scenes.

We compared our algorithm to the two latest methods in dense reconstruction from stereo in MIS – [5] and [6] – over various pair of frames from our data. Although their techniques often yielded accurate results, they would still present significant outliers in difficult areas as in figure 1. In contrast, our method has successfully discarded the vast majority of difficult areas (see figure 5), while still matching the easier parts of the image.

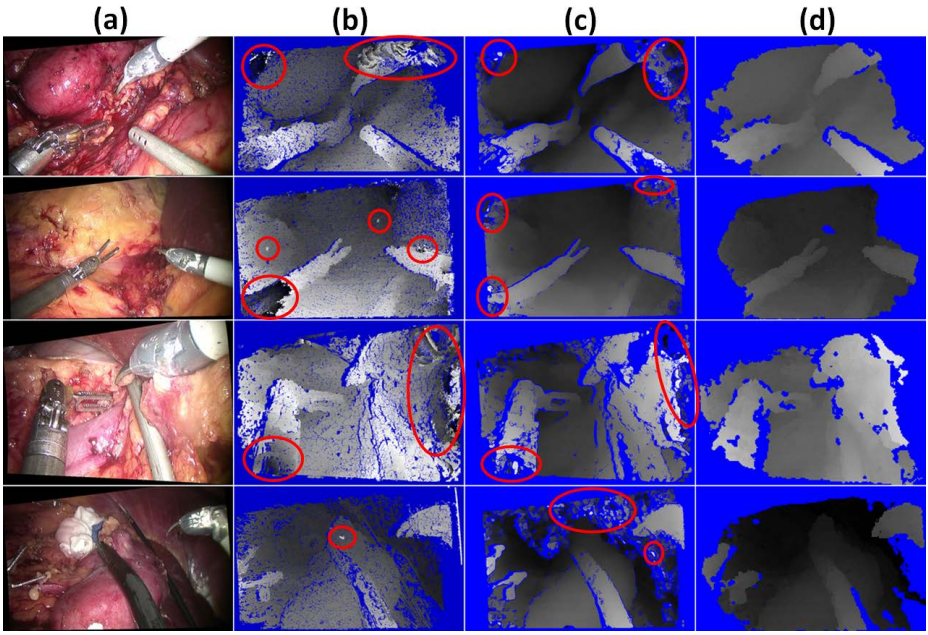


Fig. 5. Comparison with other methods. (a) Original image; (b) Depth map from [5]; (c) Depth map from [6] and (d) our method. Our method successfully dismisses error-prone areas while the two other methods present outliers (highlighted in red). In the depth maps, whiter is shallower.

Given that certain difficult areas may be discarded in our robust matching process which may result in occasional localized loss of reconstruction information, the temporal nature of our matching ensures that successful reconstruction is attained within a few frames. For example, in the central parts of the frame, most pixel reconstructions are updated within 10 frames which represent less than 0.5 seconds, as illustrated in figure 6. Therefore, the region of interest can always be successfully reconstructed locally in time.

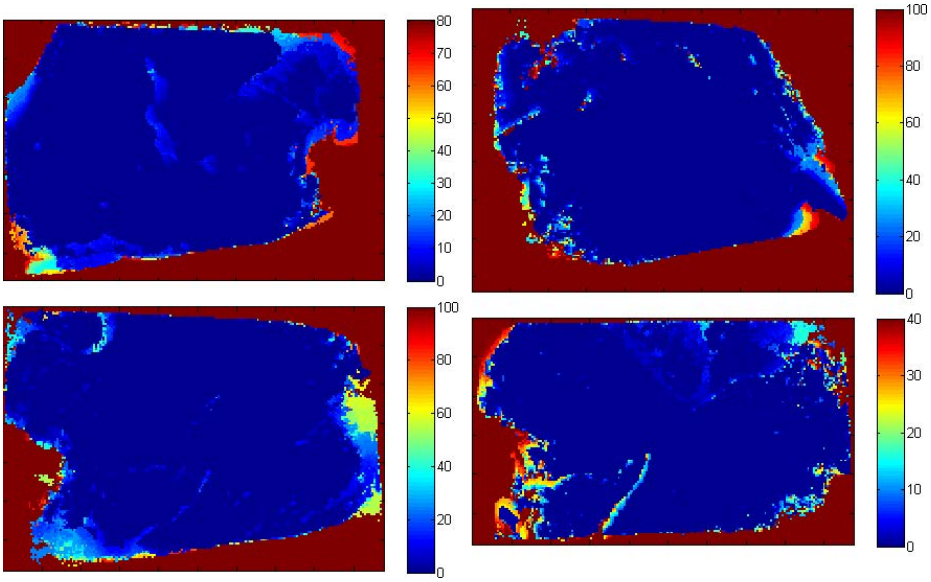


Fig. 6. Depth update with respect to image region. This figure presents four different sequences. For each graph, the pixel value represents the largest number of successive frames for the corresponding pixel reconstruction to be updated again in the sequence. The colormap scales from 0 to the total number of frames. Pixels with the maximum value (red) are those which have never been matched, either for being out of the rectified image or because of difficult regions. However, in all four example sequences shown, the central parts of the image is always blue i.e. these pixels are matched very regularly.

4 Conclusions

The purpose of this work was to provide an accurate and robust stereo dense matching method that is suited to surgical scene reconstruction. By enforcing strict matching confidence criteria and relaxing spatial smoothness, we have shown that our method is capable of discarding most outliers in frame pairs from real in-vivo clinical data where current state-of-the-art techniques fail.
Since all matchings are independent of each other, our method is highly parallelizable and will reach its full potential once implemented on GPU, which is our plan for the near future.

Acknowledgements. The author would like to thank Dr. Danail Stoyanov for providing the code of his method, Sebastian Roehl for submitting results from his program on our images, the Hamad Medical Corporation hospital for providing the surgical stereo sequences and Mitchell Vu for his help on compiling results.

References

1. Babbar, P., Hemal, A.K.: Robot-assisted partial nephrectomy: current status, techniques, and future directions. International Urology and Nephrology 44(1), 99–109 (2012)
2. Mourgues, F., Devernay, F., Coste-Maniere, E.: 3D Reconstruction of the Operating Field for Image Overlay in 3D-Endoscopic Surgery. In: International Symposium on Augmented Reality, pp. 191–192 (2001)
3. Sielhorst, T., Feuerstein, M., Navab, N.: Advanced Medical Displays: A Literature Review of Augmented Reality. Journal of Display Technology 4(4), 451–467 (2008)
4. Su, L.-M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D.: Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3D-CT to stereoscopic video registration. Journal of Urology 73(4), 896–900 (2009)
5. Stoyanov, D., Visentini-Scarzanella, M., Pratt, P., Yang, G.-Z.: Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 13(pt 1), pp. 275–282 (2010)
6. Roehl, S., et al.: Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. The International Journal of Medical Physics Research and Practice 39(3), 1632–1645 (2012)
7. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (2004) ISBN: 0521540518
8. Scharstein, D., Szeliski, R., Zabih, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: IEEE Workshop on Stereo and Multi-Baseline Vision, vol. (1), pp. 131–140 (2002)
9. Brown, M.Z., Burschka, D., Hager, G.D.: Advances in computational stereo. EEE Transactions on Pattern Analysis and Machine Intelligence 25(8), 993–1008 (2003)
10. `http://vision.middlebury.edu/stereo/data/`
11. Stoyanov, D., Elhelw, M., Lo, B.P., Chung, A., Bello, F., Yang, G.-Z.: Current Issues of Photorealistic Rendering for Virtual and Augmented Reality in Minimally Invasive Surgery. In: International Conference on Information Visualization, pp. 350–358 (2003)
12. Vagvolgyi, B.P., Su, L.-M., Taylor, R.H., Hager, G.D.: Video to CT Registration for Image Overlay on Solid Organs. In: Augmented Reality in Medical Imaging and Augmented Reality in Computer-Aided Surgery (AMIARCS), pp. 78–86 (2008)
13. Lowe, D.G.: Object recognition from local scale-invariant features. In: International Conference on Computer Vision, pp. 1150–1157 (1999)
14. Pollefeys, M., Koch, R., Van Gool, L.: A simple and efficient rectification method for general motion. In: International Conference on Computer Vision, pp. 496–501 (1999)
15. Bleyer, M., Chambon, S.: Does Color Really Help in Dense Stereo Matching? In: International Symposium on 3D Data Processing, pp. 1–8 (2010)

Model-Based Human Teeth Shape Recovery from a Single Optical Image with Unknown Illumination

Aly Farag[1], Shireen Elhabian[1], Aly Abdelrehim[1],
Wael Aboelmaaty[2], Allan Farman[2], and David Tasman[2]

[1] Computer Vision and Image Processing Laboratory,
University of Louisville, Louisville, KY 40292
[2] School of Dentistry, University of Louisville, Louisville, KY, 40202, USA

Abstract. Several existing 3D systems for dental applications rely on obtaining an intermediate solid model of the jaw (cast or teeth imprints) from which the 3D information can be captured. In this paper, we propose a model-based shape-from-shading (SFS) approach which allows for the construction of plausible human jaw models *in vivo*, without ionizing radiation, using fewer sample points in order to reduce the cost and intrusiveness of acquiring models of patients teeth/jaws over time. The inherent relation between the photometric information and the underlying 3D shape is formulated as a statistical model where the effect of illumination is modeled using Spherical Harmonics (SH) and the partial least square (PLS) approach is deployed to carry out the estimation of dense 3D shapes. Moreover, shape and texture alignment is accomplished using a proposed definition of anatomical jaw landmarks which can be automatically detected. Vis-à-vis dental applications, the results demonstrate a significant increase in accuracy in favor of the proposed approach. In particular, our approach is able to recover geometrical details of tooth occlusal surface as well as mouth floor and ceiling as compared to SFS-based approaches.

1 Introduction

Object modeling from a single image, augmented with prior information, facilitates various studies and applications in art, design, reverse engineering, rapid prototyping and basic analysis of deformations and uncertainties. Without the use of ionizing radiation (e.g. X-ray and Computer Tomography - CT), object modeling involves constructing a 3D representation for the information conveyed in the given 2D images. This problem has been studied in the past four decades resulting in many solutions bundled under the name *shape-from-X*. In particular, techniques, such as shape-from-shading provide promise of image-based 3D reconstruction when the imaging environment is somewhat precise.

To motivate the contribution of this work, we consider a dental application; 3D reconstruction of the visible part of the human jaw. Dentistry usually require accurate 3D representation of the teeth and jaw for diagnostic and treatment purposes. For instance, orthodontic treatment involves the application, over time, of force systems to teeth for malocclusion correction. Several existing 3D systems for dental applications found in literature rely on obtaining an intermediate solid model of the jaw (cast or teeth imprints) and then capturing the 3D information from that model, e.g. [1]. There may

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 263–272, 2013.

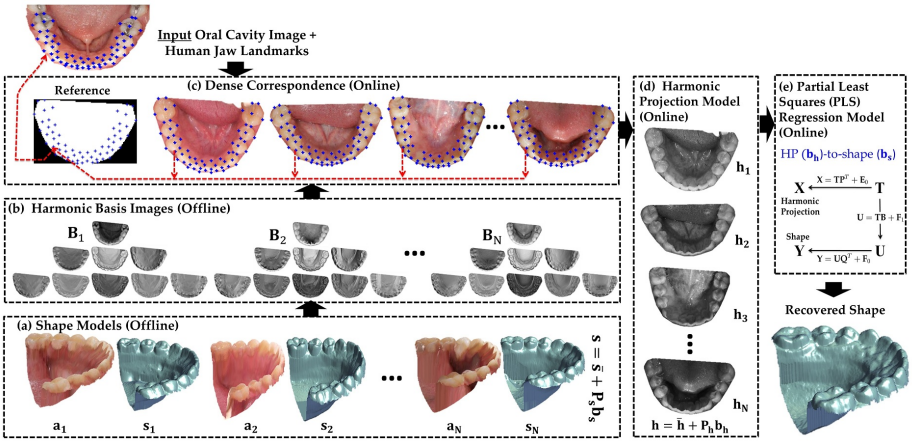


Fig. 1. Block diagram of the proposed model-based human jaw shape recovery: (a) An aligned ensemble of the shapes and textures (oral cavity images) of human jaws is used to build the 3D shape model. (b) Given the texture and surface normals (defining the shape) of a certain jaw in the ensemble, harmonic basis images are constructed. Given an input oral cavity image under general unknown illumination and a set of human jaw anatomical landmark points: (c) Dense correspondence is established between the input image and each jaw sample in the ensemble, where each pixel position within the convex hull of a reference jaw shape corresponds to a certain point on a sample jaw (shape and texture) and in the same time to a certain point on the input image. (d) The input image, in the reference frame, is projected onto the subspace spanned by the harmonic basis of each sample in the ensemble which are scaled (using the projection coefficients) and summed-up to construct the harmonic projection (HP) images which encodes the illumination conditions of the input image. Such images are then used to construct an HP model of the input image. (e) The inherit relation between the HP images and the corresponding shape is cast as a regression framework where partial least squares is used to solve for shape coefficients to recover the shape of the input image.

therefore be a demand for intraoral measurement that could be fulfilled by photogrammetry, which has been applied to the measurement of many small objects, even the measurement of dental replicas. Thus photogrammetry seems to offer a reduced cost technique while avoiding the need for castings.

Our argument of image-based approach for 3D reconstruction as an alternative to CT-scanning is based on the following. During the exposure to diagnostic imaging using x-ray (ionizing/ electromagnetic radiation), the patient body is penetrated by millions of x-ray photons whose ionization can damage the body's molecules especially DNA in chromosomes. Most DNA damage is repaired immediately, but rarely a portion of a chromosome may be permanently altered (a mutation) leading ultimately to the formation of a tumor [2]. While doses and risks for dental radiology are small, a number of epidemiological studies have provided evidence of an increased risk of brain [3], salivary gland [4] and thyroid tumors [5] for dental radiography. Also, pregnant mothers undergoing diagnostic or therapeutic procedures involving ionizing radiation, or who may be exposed to environmental radiation, there is a great potential for damage to the early embryo [6]. These effects are believed to have no threshold radiation dose below

which they will not occur [7]. On the other hand, CT-scanning is considered expensive and not paid by insurance companies unless disease oriented. Meanwhile, dental offices in rural areas do not have such a luxury. Thus our intent is to develop a purely image-based reconstruction mechanism as a cost-effective information tool for the dentist.

In this paper we aim at making it easy and feasible for doctors, dentists, and researchers to obtain models of a person's jaw *in vivo*, without ionizing radiation, using fewer sample points in order to reduce the cost and intrusiveness of acquiring models of patients teeth/jaws over time. This is a challenging problem due to the "unfriendly" environment of taking measurements inside a person's mouth [8]. Further assumptions of the presence of distinct features or texture regions on the object in stereo images and the photo consistency in space carving are rarely valid in practice.

Due to the lack of surface texture, shape-from-shading (SFS) algorithms have been used to reconstruct the 3D shape of human teeth/jaw due to the significant shading cue presented in an intra-oral image, e.g. [9]. Nonetheless, in principle, SFS is an ill-posed problem, Prados and Faugeras [10] showed that constraining the SFS problem to a specific class of objects can improve the accuracy of the recovered shape. Thus the main objective of the presented work is to develop and validate a holistic approach for image-based 3D reconstruction of the human jaw based on statistical shape-from-shading (SSFS), covering regions which the classical SFS approach does not handle, using scanned molds and images of the oral cavity to estimate the shape of a human jaw in order to create a more accurate jaw 3D model. In specific, the structure of human jaw reveals what can be acquired in terms of prior information to enhance the SFS process where the upper and lower jaws are symmetric and lined up according to specific anatomical features and landmarks. We believe that this approach has the potential to greatly improve plausibility of the resulting shape from shading models.

2 Related Work

There has been a substantial amount of work regarding statistical shape recovery for human face modeling and biomedical structures with distinct shapes - e.g., modeling the corpus callosum, the kidney and spinal cord; it is an active research area under shape and appearance modeling (e.g., [11, 12]). Atick et al. [13] proposed the first statistical SFS method where principal component analysis (PCA) was used to parameterize the set of all possible facial surfaces. Scene parameters such as pose and illumination were estimated in the process of a morphable model fitting using a stochastic gradient descent-based optimization. By considering the statistical constraint of [13] and the geometric constraint of symmetry in [14], Dovgard and Basri [15] introduced a statistical symmetric SFS method. Smith and Hancock [11] modeled surface normals within the framework of statistical SFS. Based on active appearance models (AAM) concept of Cootes et al. [16], Castelan et al. [17] developed a coupled statistical model to recover the 3D shape from intensity images with frontal light source, where the 2D shape model in [16] is replaced with a 3D shape model composed of height maps. The main advantage of the Castelan approach over the 3D morphable model framework [18] is the straightforward recovery of the 3D face shape, without undergoing a costly optimization process.

One of the main challenges that confront SFS algorithms is dealing with arbitrary illumination. Basri and Jacobs [19] proved that images of convex Lambertian object taken under arbitrary distant illumination conditions can be approximated accurately using low-dimensional linear subspace based on spherical harmonics. This has also been validated for near illumination conditions [20]. Since then, SH was incorporated in SFS framework to tackle the problem of illumination [21–23, 12].

3 Contributions

In this paper, we propose to investigate the SSFS approach on the human jaw where face and jaw modeling carry similarities and differences. Facial images can be easily obtained and databases of various imaging conditions are already in place, along with a significant body of algorithmic development. Human faces are easy to annotate and automate the process of face cropping and feature extraction. On the other hand, the human jaw is not a friendly environment to image, as indicated before, while no databases exist to carry out a SSFS methodology.

Fig. 1 illustrates the SSFS problem for reconstruction of the human jaw using a series of textures and shapes (obtained from CT scans of molds) for a group of subjects. The process starts with annotating the jaw at the known anatomical landmarks, in order to co-register the shapes and textures needed to construct the corresponding models. We use spherical harmonics to provide the optimal basis for illumination representation, and the partial least square (PLS) approach to carry out the estimation of dense 3D shapes. Key requirements for successful SSFS are the availability of a comprehensive database that describe the teeth/jaw variability per age, gender and ethnic factors. Our work aims to undertake such a task and make the databases available for researchers worldwide.

Vis-à-vis dental applications, the results demonstrate a significant increase in accuracy in favor of the proposed approach. In particular, our approach is able to recover geometrical details of tooth occlusal surface as well as mouth floor and ceiling as compared to shape-from-shading based approaches.

4 Proposed Definition of Anatomical Jaw Landmarks

4.1 Landmarks Definition

In this work, we mainly focus on the reconstruction of the *clinical crowns* which are defined to be the portion of the teeth that is visible in the mouth. As such, we limit the jaw's anatomical landmarks to such a space as follows according to their location, i.e. on the tooth surface or on the interface between the tooth and the gum. Typically a landmark represents a distinguishable point which is present in most of the images under consideration, for example, the location of central grooves of each tooth. Fig. 2 illustrates the location of 72 landmark points for a fourteen-teeth jaw.

In case of posterior teeth (i.e. cuspids, premolars and molars) which are responsible for chewing food, we are interested in the coalescence of the crown lobes. In particular, a *central pit or groove* can be considered as a landmark which is the deepest portion of

a tooth fossa. While anterior teeth (i.e. incisors) whose job is to rip food apart is identifiable by a convex elevation of the crown surface which forms the biting edge. Hence we consider the midpoint of the *incisal edge or ridge* as a landmark for an anterior tooth.

The fibrous tissue covering the alveolar bone and surrounds the necks of the teeth, i.e. the gum, forms what is denoted as *gingival line*. This line marks the level of termination of the non-attached soft tissue surrounding the tooth. It separates the clinical crown and the root. We define the *gingival line midpoint* to be the minimum or maximum point on the gingival line formed by a single tooth. On the other hand, *gingival embrasure* is the respective point in the open space between the proximal surfaces of two adjacent teeth in the same dental arch.

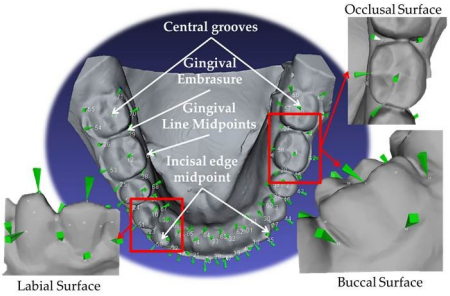


Fig. 2. Illustration of the proposed human jaw anatomical landmarks

4.2 Landmark Localization in Optical Images

In the online stage of our approach, a single image of the visible crowns is given from which the defined landmarks should be identified. This guides the alignment of the input image to the prior model, e.g. [24]. Hence, it is essential to automate the detection of such landmarks. In the training set, we manually annotate an ensemble of human jaws surfaces (based on CT-scanning of molds) in order to construct a sparse version of the jaw shape. These landmarks serve as a correspondence operator between individual training samples where we use the generalized Procrustes analysis [25] to filter out translation, scale and rotation. We deployed the Active Shape Model (ASM) by Cootes [26] to search for the landmarks in the given image. The ASM repeats the following two steps until convergence: (i) suggest a tentative shape by adjusting the locations of shape points by template matching of the image texture around each point (ii) conform the tentative shape to a global shape model. The individual template matches are unreliable and the shape model pools the results of the weak template matchers to form a stronger overall classifier. The entire search is repeated at each level in an image pyramid, from coarse to fine resolution. The initialization of the mean shape onto the given image is accomplished by segmenting the teeth region based on fitting a Gaussian mixture to the image intensity with two dominant classes; jaw and background.

5 Illumination-Invariant Statistical Shape from Shading

When the light source and the viewer are far from the object, the image intensity I at a pixel $\mathbf{x}$ can be obtained from the image irradiance of the corresponding surface point, which is defined as the surface radiance being modulated by the surface texture $a(\mathbf{x})$, i.e. $I(\mathbf{x}) = a(\mathbf{x})\mathcal{R}(\boldsymbol{n}(\mathbf{x}))$. The classical brightness constraint in SFS measures the total brightness of the reconstructed image compared to the input image, it can be defined as;

$$\epsilon = \int\int \left(I(\mathbf{x}) - a(\mathbf{x})\mathcal{R}(\boldsymbol{n}(\mathbf{x}))\right)^2 d\mathbf{x} \tag{1}$$

where $a(.)$ is the surface texture at point $\mathbf{x}$ while $\mathcal{R}(.)$ is the radiance of the surface patch with unit normal $\boldsymbol{n}(\mathbf{x})$, also known as surface reflectance function.

The brightness constraint in (1) can be rewritten in the discrete domain as a linear combination of harmonic basis images resulted from the 2nd order SH approximation to the reflectance function [19]. Thus the image intensity I can be expressed as; $I(\mathbf{x}) = \sum_{i=0}^{n-1} \alpha_i b_i(\mathbf{x})$ where $b_i(\mathbf{x}) = f_i\left(a(\mathbf{x}), \mathbf{n}(\mathbf{x})\right)$ are the harmonic basis images which are functions of surface texture $a(\mathbf{x})$ and surface normals $\mathbf{n}(\mathbf{x})$ at pixel $\mathbf{x}$ (refer to [19] for their definition). The coefficient α_i denotes the ith coefficient in the illumination spectrum being modulated by the Lambertian kernel spectrum.

In matrix notation, let $I \in \mathbb{R}^{d\times 1}$ be an image vector with d pixels, $\mathbf{B} = [b_0(\mathbf{x}), ..., b_{n-1}(\mathbf{x})] \in \mathbb{R}^{d\times n}$ be the matrix of harmonic basis images as its columns, where n is the number of basis images, typically $n = 9$, and $\alpha \in \mathbb{R}^{n\times 1}$ vector of SH coefficients[1]. Hence the discrete version of the brightness constraint becomes,

$$\epsilon = \sum_{\mathbf{x}} \left(I(\mathbf{x}) - \mathbf{B}(\mathbf{x})\alpha\right)^2 = \|I - \mathbf{B}\alpha\| \tag{2}$$

Representing the surface reflectance function in terms of SH allow us to infer the illumination of a given image; given an input image I, the harmonic basis images $\mathbf{B}$ of a 3D object (a human jaw in particular), defined by its shape $\mathbf{s} = [\mathbf{n}(\mathbf{x}_0), ..., \mathbf{n}(\mathbf{x}_{d-1})]^T$ and texture $\mathbf{a} = [a(\mathbf{x}_0), ..., a(\mathbf{x}_{d-1})]^T$), are obtained to deduce the coefficients $\widehat{\alpha}$ that best matches the input image. This results in an over-determined linear system of equations $I = \mathbf{B}\alpha$ which can be solved for $\widehat{\alpha}$ using singular value decomposition (SVD).

If the input image and the basis images used to compute the coefficients $\widehat{\alpha}$ belong to the same object, we can reconstruct the input image from these coefficients, i.e. $h = \mathbf{B}\widehat{\ell} = I$, where h denotes what we call *harmonics projection* (HP) image. However in the general case, the basis images $\mathbf{B}$ would belong to an object which is different from the one in the input image I, nonetheless they belong to the same object class e.g. different realizations of a human jaw. Thus the reconstructed image h provide a mean of encoding the illumination of the input image while maintaining the identity of the object whose basis images are used in the reconstruction process.

While (1) can be solved in an iterative manner to infer the underlying shape as in [22], the inherit relation between the HP images $\mathbf{h}$ and the corresponding shape $\mathbf{s}$ can be cast

[1] Since the information of this harmonic expansion mainly lies in the analytic form of the SH basis, we denote its coefficients as SH coefficients.

into a regression framework resulting into the HP-to-shape model. In this case, the shape is solved for using a series of matrix operations guaranteeing faster shape recovery when compared to its iterative counterpart. This was proven to yield comparable results in terms of reconstruction accuracy [12].

Dimensionality reduction is performed using PCA to construct 3D shape model (offline step) and HP model (online step) where the coefficients are used to build the regression model rather than the original shape and HP instances. In particular, the 3D shape model can be constructed by performing PCA on a set of aligned samples of 3D shapes, the resulting shape model is $\mathbf{s} = \bar{\mathbf{s}} + \mathbf{P_s}\mathbf{b_s}$ where $\bar{\mathbf{s}}$ is the mean shape, $\mathbf{P_s}$ are the shape eigenvectors and $\mathbf{b_s}$ is the set of shape coefficients. On the other hand, the HP model is trained online which incorporate the illumination conditions of the input image; given an image I and the basis images $\mathbf{B}_k$ of object instance k, the HP image h_k is obtained, where $h_k = \mathbf{B}_k\widehat{\alpha}_k$ with $\widehat{\alpha}_k$ obtained by solving the linear system of equations $I = \mathbf{B}_k\alpha_k$. After reconstructing the projection images of all the instances in the jaw database, we can model the HP images using PCA as $\mathbf{h} = \bar{\mathbf{h}} + \mathbf{P_h}\mathbf{b_h}$ where $\bar{\mathbf{h}}$ is the mean HP image, $\mathbf{P_h}$ are the HP images eigenvectors and $\mathbf{b_h}$ is the set of HP coefficients. Thus, instead of using the high dimensional vectors $\mathbf{s}_k$ and $\mathbf{h}_k$ into the regression, they are replaced by their respective coefficients $\mathbf{b_s}_k$ and $\mathbf{b_h}_k$, where the HP coefficients are considered the independent variable while the shape coefficients are the dependent variables. We use partial least squares regression (PLS) instead of the classical least squares to avoid random noise which might exist in the dependent and independent variables. Fig. 1 shows a block diagram of the offline/online processes for the proposed shape recovery approach.

6 Experimental Results

In this section, we show experiments to evaluate the performance of the proposed framework in recovery 3D models for human jaws. Upper and lower jaw models are constructed from eight young-aged subjects using their oral cavity images and the CT-scan of their respective molds. There are two samples per subject, one pre-repair jaw and another post-repair jaw. The original 3D scans are converted into a Monge patch format which represents the surface as $(x, y, f(x, y))$. We use a landmark-based approach to establish the dense correspondence between database samples, where a set of sparse anatomical landmark points are manually annotated (refer to Fig. 1 for their illustration). Generalized Procrustes Analysis (GPA) is first performed to align the set of shapes to a common reference frame. The average of the aligned shapes define the reference shape which is crucial in establishing dense correspondence between the jaw samples, see Fig. 1. Each pixel within the convex hull of the reference shape corresponds to a certain point on each jaw sample scan through a physically motivated thin-plate splines warping function.

To evaluate the proposed approach, out-of-training jaw samples are reconstructed and compared against the ground truth CT-scan. Four types of samples are considered: (a) pre-repair and (b) post-repair lower jaw, (c) pre-repair and (d) post-repair upper jaw.

Fig. 3(c) shows a sample reconstruction of a human jaw based on the proposed approach. Notice that it is close to its ground truth shape, as illustrated in Fig. 3(b).

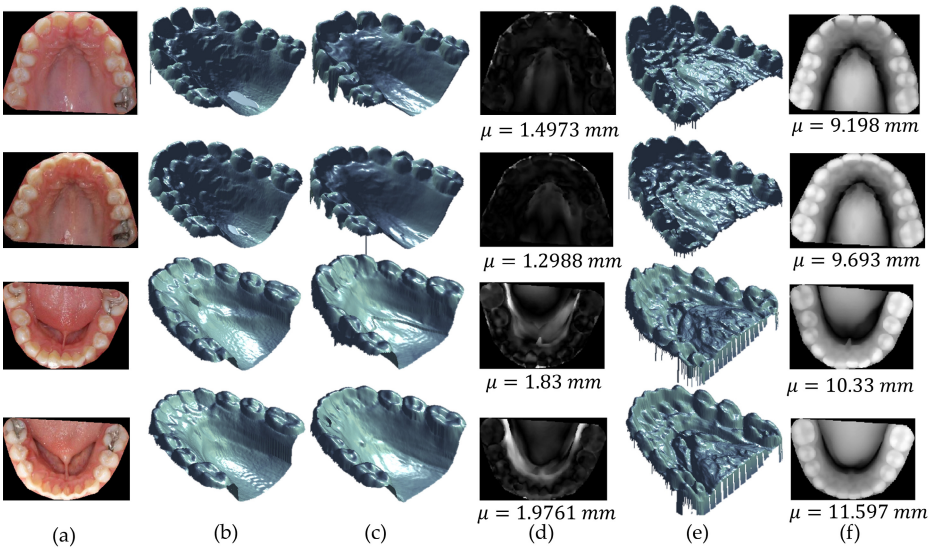


Fig. 3. Sample reconstruction result of a single subject: (first row) upper, pre-repair jaw, (second row) upper, post-repair jaw, (third row) lower, pre-repair jaw, and (fourth row) lower, post-repair jaw. (a) Input image being masked using the convex hull of the jaw landmarks. (b) Ground truth shape from the CT-scanner. (c) Reconstructed shape based on our approach. (d) root-mean-squared error map with average error shown in mm. (e) Reconstructed shape based on SFS of [9]. (f) root-mean-squared error map of (e) with average error shown in mm.

While Fig. 3(e) shows inaccurate reconstructions based on SFS of [9] which was recently proposed for jaw shape recovery. This emphasizes the role of incorporating prior-information for shape recovery as well as illumination modeling.

Table 1 reports the root-mean-square error in mm between the 3D points from the CT scan and the corresponding reconstructed surface points. Notice that the error values are minimal when compared to SFS-based reconstruction. Post-repair error values are also smaller than pre-repair values in most of the samples, indicating that it is more difficult to reconstruct human jaws with irregular tooth shapes and locations. One can observe higher errors in case of SFS for the lower jaw when compared to the upper one where there is no occlusion due to the tongue.

A natural question to be asked is how to make use of SFS results and SSFS? Of course, SFS is based on the visible surface of the jaw; at best the crown would be

Table 1. Average surface reconstruction accuracy (RMS) in mm

Jaw type	SSFS	SFS
Upper, pre-repair	2.08999	8.80572
Upper, post-repair	2.02334	8.96832
Lower, pre-repair	3.11911	10.02804
Lower, post-repair	2.57112	11.42853

possibly constructed, while SSFS constructs the entire jaw. On the other hand, SFS provides the object-specific constructions. A logical thing would be to enhance the SSFS with SFS, by morphing the upper part of the model with the crown portion generated from SFS. With a good database of objects, credible SSFS models would be possible, which when morphed to the crown reconstructions would produce a more realistic jaw.

7 Conclusion and Future Work

In this paper, we presented an affordable, flexible, automatic dental tool for the reconstruction of the clinically visible part of the human jaw. It was based on a single captured optical image and a statistical shape recovery approach which makes use of a small number of measured points to construct a plausible 3D model through a learned correspondence based on a measured human jaw dataset. We expressed the surface reflectance function in terms of spherical harmonics to provide the optimal basis for illumination representation. The brightness constraint was then cast as a Partial Least Squares (PLS) regression problem, which allows for the rapid computation of the solution. The PLS algorithm is composed of a sequence of matrix operations; the approach in this work can recover 3D shapes much faster than its iterative counterpart, without compromising the integrity of the results. The results demonstrated the effect of adding statistical prior as well as illumination modeling on the accuracy of the recovered shape. The next step is to investigate the fusion of SFS and SSFS where SFS provides the object-specific constructions while SSFS is perform shape recovery based on partial information. This will lend benefits to tasks such as teeth restoration in dental applications.

References

1. Goshtasby, A.A., Nambala, S., de Rijk, W.G., Campbell, S.D.: A system for digital reconstruction of gypsum dental casts. IEEE Transactions on Medical Imaging 16, 664–674 (1997)
2. European Commission, European guidelines on radiation protection in dental radiology. Radiation Protection Issue number 136 (2004)
3. Maillie, H.D., Gilda, J.E.: Radiation-induced cancer risk in radiographic cephalometry. Oral Surg. Oral Med. Oral Pathol. 75, 631–637 (1993)
4. Horn-Ross, P.L., Ljung, B.M., Morrow, M.: Environmental factors and the risk of salivary gland cancer. Epidemiology 8, 414–419 (1997)
5. Hallquis, A., Hardell, L., Degerman, A., Wingren, G., Boquist, L.: Medical diagnostic and therapeutic ionizing radiation and the risk for thyroid cancer: a case-control study. Eur. J. Cancer Prevention 3, 259–267 (1994)
6. Wilson, K., Sun, N., Huang, M., Zhang, W., Lee, A., Li, A., Wang, S., Wu, J.: Effects of ionizing radiation on self renewal and pluripotency of human embryonic stem cells. Cancer Res. 70(13), 5539–5548 (2010)
7. European Commission, Low dose ionizing radiation and cancer risk. Radiation Protection Issue number 125 (2001)
8. Yamany, S.M., Farag, A.A., Tasman, D., Farman, A.G.: A 3-d reconstruction system for the human jaw using a sequence of optical images. IEEE Transactions on Medical Imaging 19, 538–547 (2000)

9. Abdelrahim, A.S., Abdelrahman, M.A., Abdelmunim, H., Farag, A., Miller, M.: Novel image-based 3d reconstruction of the human jaw using shape from shading and feature descriptors. In: Proceedings of the British Machine Vision Conference, pp. 41.1–41.11. BMVA Press (2011)
10. Prados, E., Faugeras, O., Camilli, F.: Shape from shading: a well-posed problem? RR 5297, INRIA Research (August 2004)
11. Smith, W.A.P., Hancock, E.R.: Recovering facial shape using a statistical model of surface normal direction. IEEE Trans. on Pattern Analysis and Machine Intelligence 28
12. Rara, H., Elhabian, S., Starr, T., Farag, A.: 3d face recovery from intensities of general and unknown lighting using partial least squares. In: 17th IEEE International Conference on Image Processing (ICIP 2010) (September 2010)
13. Atick, J.J., Griffin, P.A., Norman Redlich, A.: Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images. Neural Comput. 8(6), 1321–1340 (1996)
14. Zhao, W.Y., Chellappa, R.: Symmetric shape-from-shading using self-ratio image. Int. J. Comput. Vision 45(1), 55–75 (2001)
15. Dovgard, R., Basri, R.: Statistical Symmetric Shape from Shading for 3D Structure Recovery of Faces. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3022, pp. 99–113. Springer, Heidelberg (2004)
16. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23, 681–685 (2001)
17. Castelan, M., Smith, W., Hancock, E.: A coupled statistical model for face shape recovery from brightness images. IEEE Transaction on Image Processing 16, 1139–1151 (2007)
18. Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. IEEE Trans. Pattern Anal. Mach. Intell. 25(9), 1063–1074 (2003)
19. Basri, R., Jacobs, D.W.: Lambertian reflectance and linear subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(2), 218–233 (2003)
20. Frolova, D., Simakov, D., Basri, R.: Accuracy of Spherical Harmonic Approximations for Images of Lambertian Objects under Far and Near Lighting. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 574–587. Springer, Heidelberg (2004)
21. Ahmed, A., Farag, A.A.: A New Statistical Model Combining Shape and Spherical Harmonics Illumination for Face Reconstruction. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Paragios, N., Tanveer, S.-M., Ju, T., Liu, Z., Coquillart, S., Cruz-Neira, C., Müller, T., Malzbender, T. (eds.) ISVC 2007, Part I. LNCS, vol. 4841, pp. 531–541. Springer, Heidelberg (2007)
22. Rara, H., Elhabian, S., Starr, T., Farag, A.: Model-based shape recovery from single images of general and unknown lighting. In: 16th IEEE International Conference on Image Processing (ICIP 2009), November 7-10, pp. 517–520 (2009)
23. Castelan, M., Van Horebeek, J.: Relating intensities with three-dimensional facial shape using partial least squares. Computer Vision, IET 3, 60–73 (2009)
24. Blanz, V., Mehl, A., Vetter, T., Seidel, H.P.: A statistical method for robust 3d surface reconstruction from sparse data. In: Int. Symp. on 3D Data Processing, Visualization and Transmission, pp. 293–300 (2004)
25. Gower, J.C.: Generalized procrustes analysis. Psychometrika 40(1), 33–51 (1975)
26. Taylor, C.J., Cootes, T.F.: Technical report: Statistical models of appearance for computer vision, The University of Manchester School of Medicine (2004), `www.isbe.man.ac.uk/bim/refs.html`

Brain Tumor Cell Density Estimation from Multi-modal MR Images Based on a Synthetic Tumor Growth Model

Ezequiel Geremia[1], Bjoern H. Menze[1,2,3], Marcel Prastawa[4], M.-A. Weber[5], Antonio Criminisi[6], and Nicholas Ayache[1]

[1] Asclepios Research Project, INRIA Sophia-Antipolis, France
[2] Computer Science and Artificial Intelligence Laboratory, MIT, USA
[3] Computer Vision Laboratory, ETH Zurich, Switzerland
[4] Scientific Computing and Imaging Institute, University of Utah, USA
[5] Diagnostic and Interventional Radiology, Heidelberg University Hospital, Germany
[6] Machine Learning and Perception Group, Microsoft Research Cambridge, UK

Abstract. This paper proposes to employ a detailed tumor growth model to synthesize labelled images which can then be used to train an efficient data-driven machine learning tumor predictor. Our MR image synthesis step generates images with both healthy tissues as well as various tumoral tissue types. Subsequently, a discriminative algorithm based on random regression forests is trained on the simulated ground truth to predict the continuous latent tumor cell density, and the discrete tissue class associated with each voxel. The presented method makes use of a large synthetic dataset of 740 simulated cases for training and evaluation. A quantitative evaluation on 14 real clinical cases diagnosed with low-grade gliomas demonstrates tissue class accuracy comparable with state of the art, with added benefit in terms of computational efficiency and the ability to estimate tumor cell density as a latent variable underlying the multimodal image observations. The idea of synthesizing training data to train data-driven learning algorithms can be extended to other applications where expert annotation is lacking or expensive.

1 Introduction

Brain tumors are complex patho-physiological processes representing a series of pathological changes to brain tissue [1]. Increasing effort is invested in modelling the underlying biological processes involved in brain tumor growth [2, 3]. As brain tumors show a large variety of different appearances in multi-modal clinical images, the accurate diagnosis and analysis of these images remains a significant challenge. We show in the example of gliomas, the most frequent brain tumor [4], how a *generative* patho-physiological model of tumor growth can be used in conjunction with a *discriminative* tumor recognition algorithm, based on random regression forests. Applied to real data the random forest is capable of predicting the precise location of the tumor and its substructures.

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 273–282, 2013.

In addition, our model can also infer the spatial distribution of (unobservable) latent physiological features such as tumor cell densities, thus avoiding the need for expensive patho-physiological model inversion [5].

Generative probabilistic segmentation models of spatial tissue distribution and appearance proved to generalize well to previously unseen images [6–9]. In [6], tumors are modeled as outliers relative to the expected appearance of healthy tissue following a related approach for MS lesion detection [10]. Other methods [7, 8] provide explicit models for the tumor class. For instance, [8] builds a tumor appearance model for channel specific segmentation of the tumor, combining a tissue appearance model with a latent tumor class prior from [9]. Tumor growth models (*e.g.* reaction-diffusion models) have been used repeatedly to improve image registration [11] and, hence, atlas-based tissue segmentation [12]. Similarly, [13] relies on a bio-mechanical tumor growth model to estimate brain tissue loss and displacement. Generative approaches require a detailed formal description of the image generation process and may need considerable modifications when applied to slightly different tasks. These approaches also tend to be computationally inefficient.

In contrast, discriminative techniques focus on modeling the difference between *e.g.* a lesion and healthy tissues, directly [14–16]. A number of recent techniques based on decision tree ensembles have shown strong generalization capabilities and computational efficiency, even when applied to large data sets [17–19]. In [20], for example, a *classification* forest is used for segmenting multiple sclerosis lesions using long-range spatial features. In [15], the authors derived a constrained minimization problem suitable for min-cut optimization that incorporates an observation model provided by a discriminative Probabilistic Boosting Trees classifier into the process of segmentation. For multi-modal brain lesion segmentation, [16] propose a hierarchical segmentation framework by weighted aggregation with generic local image features. Unfortunately, fully supervised discriminative approaches may require large expert-annotated training sets. Obtaining such data is often prohibitive in many clinical applications.

This paper proposes a new way of combining the best of the generative and discriminative world. We use a generative model of glioma [21] to synthesize a large set of heterogeneous MR images complete with their ground truth annotations. Such images are then used to train a multi-variate *regression* forest tumor predictor [20, 22]. Thus given a previously unseen image the forest can perform an efficient, per-voxel estimation of both tumor infiltration density *and* tissue type. The general idea of training a discriminative predictor (a classifier or a regressor) on a large collection of synthetic training data is inspired by the recent success of the Microsoft Kinect for XBox 360 system [23]. This approach has great potential in different domains and especially for medical applications where obtaining extensive expert-labelled is nearly impossible.

2 Learning to Estimate Tissue Cell Density from Synthetic Training Data

This section describes the two basic steps of our algorithm: i) synthesizing heterogeneous MR images showing tumors, and ii) training a tumor detector which works on *real* patient images.

2.1 Generative Tumor Simulation Model

The automatic generation of our synthetic training dataset relies on the publicly available brain tumor simulator presented in [21]. It builds on an anisotropic glioma growth model [24] with extensions to model the induced mass-effect and the accumulation of the contrast agents in both blood vessels and active tumor regions. Then, multi-sequence MR images are synthesized using characteristic image textures for healthy and pathological tissue classes.

We generate synthetic pathological cases with varying tumor location, tumor count, levels of tumor expansion and extent of edema. The resulting synthetic cases successfully reproduce mass-effect, contrast enhancement and infiltration patterns similar to what observed in the real cases. The synthetic dataset contains 740 synthetic cases. It includes a large variability of brain tumors ranging from very diffusive tumors, showing a large edema-infiltration pattern without necrotic core, to bulky tumors with a large necrotic core surrounded by an enhanced vascularization pattern. For each case, the simulation provides four MR sequences (cf. Fig. 1) which offer different views of the same underlying tumor density distribution.

This synthetic ground truth provides a diverse view of the pathological process including mass-effect and infiltration, but also very detailed annotations for the healthy structures of the brain. The ground truth consists of voxel-wise annotations on the data that are: white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), edema, necrotic tumor core, active tumor rim and blood vessels. Unlike binary annotations which provide a mask for each tissue class, the ground truth consists of a continuous scalar map for each tissue class. Each scalar map provides, for every voxel in the volume, the density of every tissue class.

2.2 Regression Forests for Estimating Tissue Cell Density

Problem setting. We adapt a regression forests similar to the one of [17] to train an estimator of tissue cell densities from visual cues in the multi-channel MR images. For each voxel $\mathbf{v}$, the ground truth provides the density $R_c(\mathbf{v}) \in [0, 1]$ of each tissue class $c \in \mathcal{C}$. The density distribution R is normalized so that it satisfies $\sum_{c \in \mathcal{C}} R_c(\mathbf{v}) = 1$ in every voxel $\mathbf{v}$.

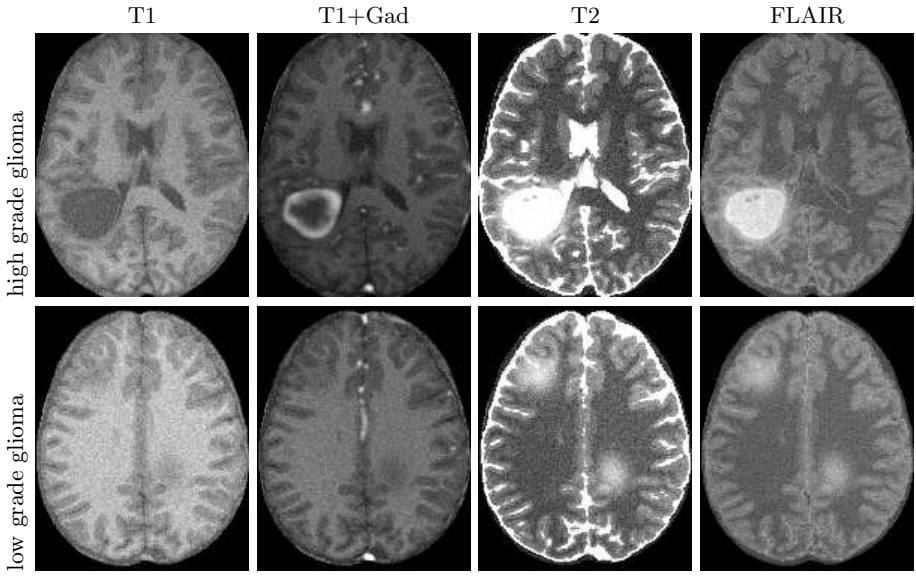


Fig. 1. Synthetic MR images. From left to right: T1, T1+Gad, T2, and FLAIR MR images. Top row: bulky tumor characterized by a large necrotic and a surrounding vascularization pattern. Bottom row: very infiltrating tumor characterized by the extended of the edema.

Feature representation. To calculate the local image features – both during training and for predictions – we sub-sample or interpolate all images to $1 \times 1 \times 2$ mm^3 resolution. We perform a skull-stripping and an intensity normalization [25] so that real MR images match the intensity distribution of synthetic MR sequences. Then image features are calculated for each voxel $\mathbf{v}$. Features include local multi-channel intensity (T1, T1+Gad, T2, Flair) as well as long-range displaced box features such as in [20]. In addition we also incorporate symmetry features, calculated after estimating the mid-sagittal plane [26]. In total, every voxel is associated with a 213−long vector of feature responses.

Regression forest training. The forest consists of T trees indexed by t. During training observations of all voxels $\mathbf{v}$ are pushed through each of the trees. Each internal node p applies a binary test $t_p = \tau_{low} \leq \theta(\mathbf{v}) < \tau_{up}$ implementing a double thresholding (τ_{low}, τ_{up}) of the visual feature $\theta(\mathbf{v})$ evaluated at voxel $\mathbf{v}$. The voxel $\mathbf{v}$ is then sent to one of the two child nodes based on the outcome of this test. As a result, each node p receives a partition of the input training data $\mathcal{T}_p = \{\mathbf{v}, R(\mathbf{v})\}_p$, composed of a voxel $\mathbf{v}$ and a vector $R(\mathbf{v}) \in [0,1]^{|\mathcal{C}|}$ storing the cell density value for each tissue class. We model the resulting distribution via a multi-variate Gaussian $\mathcal{N}_p(\mu_p, \Gamma_p)$ where μ_p and Γ_p are the mean and covariance matrix of all $R(\mathbf{v}) \in \mathcal{T}_p$, respectively. During training, the parameters (τ_{low}, τ_{up}) of the node test function and the employed visual feature θ are

optimized to maximize the information gain. We define the information gain $IG(t_p)$ to measure the quality of the test function t_p which splits $\mathcal{T}_p$ into $\mathcal{T}_p^{left}$ and $\mathcal{T}_p^{right}$. The information gain is defined as $IG(t_p) = -\sum_{k\in\{left,right\}} \omega_k log\rho_k$ with $\omega = |\mathcal{T}_p^k|/|\mathcal{T}_p|$ and $\rho_k = max|eig(\Gamma_k)|$ where eig denotes all matrix eigenvalues. In contrast to the more conventional information gain used in [17], our formulation gives a robust estimate of the dispersion. Indeed, the information gain presented in [17] models the dispersions as $|\Gamma_k|$ which evaluates to 0 in the case a tissue class is missing from the input partition $\mathcal{T}_p$. Our definition of the information gain focuses on the direction showing maximum dispersion, i.e. ρ_k, and ignores the missing information on tissue classes.

At each node p, the optimal test $t_p^* = \arg\max_\Lambda IG(t_p)$ is found by exhaustive search over a random subset of the feature space $\Lambda = \{\tau_{low}, \tau_{up}, \theta\}$. Maximizing the information gain encourages minimizing ρ_p, thus decreasing the prediction error when approximating $\mathcal{T}_p$ with $\mathcal{N}_p$. The trees are grown to a maximum depth D, as long as $|\mathcal{T}_p| > 100$.

After training, the random forest embeds a hierarchical piece-wise Gaussian model which captures the multi-modality of the training data. In fact, each leaf node l_t of every tree t stores the Gaussian distribution $\mathcal{N}_{l_t}$ associated with the partition of the training data arrived at that leaf $\mathcal{T}_{l_t}$.

The employed random regression forest approximates the multi-variate distribution R by a piece-wise Gaussian distribution $\hat{R}$.

Regression forest prediction. When applied to a previously unseen test volume $\mathcal{T}_{test} = \{\mathbf{v}\}$, each voxel $\mathbf{v}$ is propagated through all the trained trees by successive application of the relevant binary tests. When reaching the leaf node l_t in all trees $t \in [1..T]$, estimated cell densities $r_t(\mathbf{v}) = \mu_{l_t}$ are averaged together to compute the forest tissue cell density estimation $r(\mathbf{v}) = (\sum_{t\in[1..T]} r_t(\mathbf{v}))/T$. Note that in each leaf l_t we maintain an estimate of the confidence Γ_{l_t} associated to the cell density estimation μ_{l_t}.

3 Experiments

We conducted two main experiments. First, as a proof of concept, we tested how well the learned forest reproduces the tissue cell densities in the synthetic model. In a second experiment we applied our method to real, previously unseen, clinical images and measured accuracy by comparing the detected and ground truth tumor outlines.

We evaluate the predictions for every test case using two complementary metrics: a segmentation metric and a robust regression metric. The segmentation metric compares binary versions of the physiological maps, independently normalized for each tissue class. The binary masks are obtained by thresholding the prediction and the ground truth at the same value. Then, we evaluate the true postive rate $TPR = TP/(TP + FN)$, the false positive rate $FPR = TP/(TP+FP)$ and the positive predictive value $PPV = TP/(TP+FP)$, where

TP, FP, and FN are the number of true positives, false positives, and false negatives, respectively. Finally, we compute the area under the ROC and the one precision-recall curves to measure how well the prediction fits the ground truth.

The robust regression metric evaluates the estimation error between the predicted continuous map and the ground truth. For every tissue class c, we compute the mean over the voxels v of the estimation error, defined as $E_c(v) = |R_c(v) - r_c(v)|$. In order to avoid artificial decrease of the mean error, we make this metric robust by only considering regions of the physiological map showing at least 10% signal in either the prediction or the ground truth.

In both experiments, we used the same forest containing $T = 160$ trees of depth $D = 20$ trained on 500 synthetic cases. The values of these meta-parameters were tuned by training and testing on a different synthetic set.

3.1 Experiment 1: Estimating Cell Density in Synthetic Cases

We tested the random forest on a previously unseen synthetic dataset with 240 cases. Results (Fig. 2) show a good qualitative match between predicted and ground truth physiological maps. As a segmentation metric we calculate the true and false positive rates as well as the positive predictive value for each possible

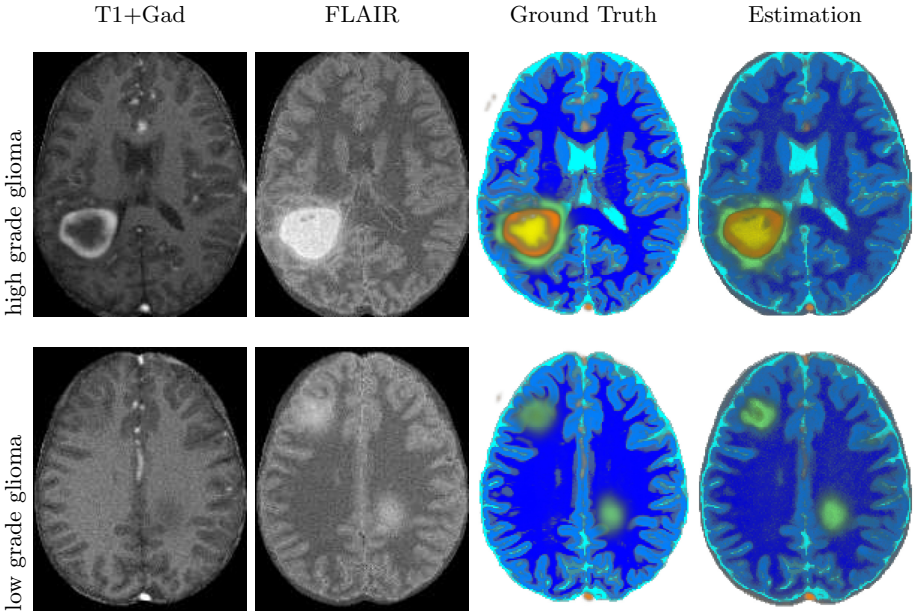


Fig. 2. Estimation of tissue cell densities. From left to right: T1+Gadolinium, FLAIR image, the ground truth provided by the simulator, the estimation of our random regression forest. Each voxel of the ground truth maps displays the mixed density between predefined tissue classes: WM (dark blue), GM (light blue), CSF (cyan), edema (green), blood vessels (orange), and necrotic core (yellow).

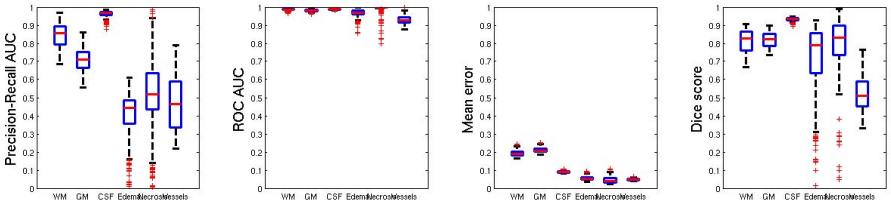


Fig. 3. Evaluation of the predictions on the synthetic dataset for each cell density map. Each label in the x-axis represents a tissue class: WM, GM, CSF, edema, necrotic core, blood vessels, respectively. We show from left to right: the area under the precision-recall curve, the area under the ROC curve, the estimation of the mean prediction error, and the dice score Each point of the ROC and precision-recall curves is built by thresholding the prediction and the ground truth at the same value. The ground truth and the prediction density maps were thresholded at the same value, i.e. 0.3.

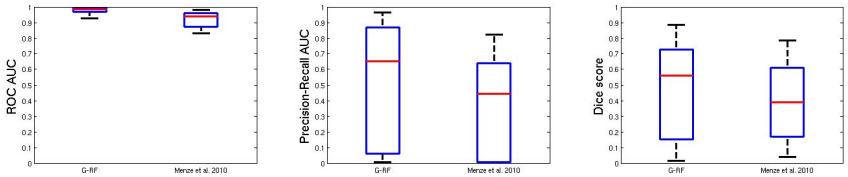


Fig. 4. Evaluation of the predictions on the clinical dataset. Box plots of the area under the ROC curve (left), under the precision-recall curve (right), and the dice score. Comparison of the proposed method (G-RF) with the method presented in [8].

threshold jointly on r and R and summarize it through ROC and precision-recall curves. For every tissue class c, we also compute the mean approximation error, defined as $E_c(v) = |R_c(v) - r_c(v)|$ (integrating over voxels with $> .001$ tumor cell density for tumor classes). Results in Fig. 3 show excellent results for WM, GM, CSF. The predicted tumor cell density is in good agreement with ground truth. A systematic bias leads to a slightly larger variance in the error metric due to the small size of the tumor classes compared to the healthy tissue classes.

3.2 Experiment 2: Segmenting Tumors in Clinical Images

We tested the same random forest on 14 clinical cases showing low and high grade glioma (Fig. 5) with T1, T1+Gad, T2 and FLAIR images. None of the clinical cases was used during training. Training was done exclusively on synthetic images. The manually-obtained ground truth consists of a binary tumor mask delineating the tumor+edema region. We calculated the same tumor outline from the predicted continuous physiological masks as done for the synthetic model [21]. Segmentation results (Fig. 4) are in excellent agreement with a

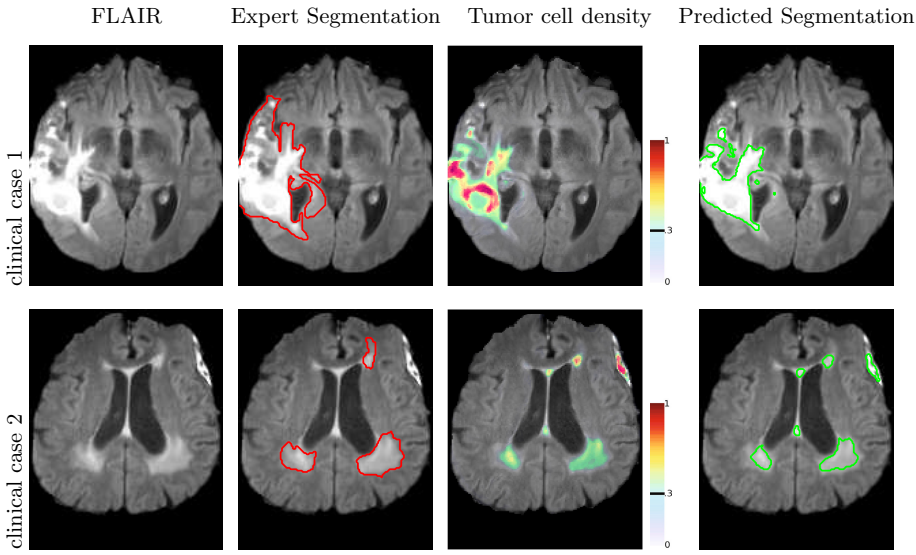


Fig. 5. Segmentation and tumor cell distribution. From left to right: preprocessed Flair MR image, FLAIR MR image overlayed with the segmentation of an expert, the normalized tumor cell density, and the predicted tumor segmentation (threshold at 0.3).

state-of-the-art unsupervised multimodal brain tumor segmentation method that also outperformed standard EM segmentation in an earlier study [8]. Note that the method presented in [8] significantly outperformed [6]. Interestingly, in a qualitative evaluation (cf. Fig. 5), the tumor cell density map shows smooth transition between the active rim of the tumor (red) and the edema (green).

4 Conclusions

This paper presented a new generative-discriminative algorithm for the automatic detection of glioma tumors in multi-modal MR brain images. A regression forest model was trained on multiple synthetically-generated labelled images. Then the system demonstrated to work accurately on previously unseen synthetic cases. It showed promising results on real patient images which led to state of the art tumor segmentation results. Our algorithm can estimate continuous tissue cell densities both for healthy tissues (WM, GM, CSF) as well as tumoral ones.

Acknowledgments. This work was partially funded by the ERC MedYMA grant. We would like to thank Marc-André Weber from the Diagnostic and Interventional Radiology Group in Heidelberg University Hospital, Germany for providing us with the clinical data.

References

1. Angelini, E.D., Delon, J., Bah, A.B., Capelle, L., Mandonnet, E.: Differential MRI analysis for quantification of low grade glioma growth. Medical Image Analysis 16, 114–126 (2012)
2. Cristini, V., Lowengrub, J.: Multiscale Modeling of Cancer: An Integrated Experimental and Mathematical Modeling Approach, pp. 185–205. Cambridge University Press (2010)
3. Deisboeck, T.S., Stamatakos, G.S.: Multiscale Cancer Modeling, pp. 359–406. CRC Press (2010)
4. Angelini, E.D., Clatz, O., Mandonnet, E., Konukoglu, E., Capelle, L., Duffau, H.: Glioma dynamics and computational models: A review of segmentation, registration, and in silico growth algorithms and their clinical applications. Cur. Med. Imag. Rev. 3, 262–276 (2007)
5. Menze, B.H., Van Leemput, K., Honkela, A., Konukoglu, E., Weber, M.-A., Ayache, N., Golland, P.: A Generative Approach for Image-Based Modeling of Tumor Growth. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 735–747. Springer, Heidelberg (2011)
6. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. Medical Image Analysis 8, 275–283 (2004)
7. Zou, K.H., Wells III, W.M., Kaus, M.R., Kikinis, R., Jolesz, F.A., Warfield, S.K.: Statistical Validation of Automated Probabilistic Segmentation against Composite Latent Expert Ground Truth in MR Imaging of Brain Tumors. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002, Part I. LNCS, vol. 2488, pp. 315–322. Springer, Heidelberg (2002)
8. Menze, B.H., van Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., Golland, P.: A Generative Model for Brain Tumor Segmentation in Multi-Modal Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 151–159. Springer, Heidelberg (2010)
9. Riklin-Raviv, T., Leemput, K.V., Menze, B.H., Wells, W.M., Golland, P.: Segmentation of image ensembles via latent atlases. Medical Image Analysis 14, 654–665 (2010)
10. Leemput, K.V., Maes, F., Vandermeulen, D., Colchester, A.C.F., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE Trans. Med. Imaging 20(8), 677–688 (2001)
11. Cabezas, M., Oliver, A., Lladó, X., Freixenet, J., Cuadra, M.B.: A review of atlas-based segmentation for magnetic resonance brain images. Comp. Meth. Prog. Biomed. 104, 158–164 (2011)
12. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint Segmentation and Deformable Registration of Brain Scans Guided by a Tumor Growth Model. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 532–540. Springer, Heidelberg (2011)
13. Zacharaki, E.I., Hogea, C.S., Shen, D., Biros, G., Davatzikos, C.: Non-diffeomorphic registration of brain tumor images by simulating tissue loss and tumor growth. NeuroImage 46, 762–774 (2009)
14. Lee, C.-H., Wang, S., Murtha, A., Brown, M.R.G., Greiner, R.: Segmenting Brain Tumors Using Pseudo–Conditional Random Fields. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 359–366. Springer, Heidelberg (2008)

15. 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 3-D MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 67–75. Springer, Heidelberg (2008)
16. Corso, J.J., Sharon, E., Dube, S., El-Saden, S., Sinha, U., Yuille, A.L.: Efficient multilevel brain tumor segmentation with integrated bayesian model classification. IEEE Transactions on Medical Imaging 27, 629–640 (2008)
17. Criminisi, A., Shotton, J., Robertson, D., Konukoglu, E.: Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In: Menze, B., Langs, G., Tu, Z., Criminisi, A. (eds.) MICCAI 2010. LNCS, vol. 6533, pp. 106–117. Springer, Heidelberg (2011)
18. Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled Decision Forests and Their Application for Semantic Segmentation of CT Images. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 184–196. Springer, Heidelberg (2011)
19. Gray, K.R., Aljabar, P., Heckemann, R.A., Hammers, A., Rueckert, D.: Random Forest-Based Manifold Learning for Classification of Imaging Data in Dementia. In: Suzuki, K., Wang, F., Shen, D., Yan, P. (eds.) MLMI 2011. LNCS, vol. 7009, pp. 159–166. Springer, Heidelberg (2011)
20. Geremia, E., Clatz, O., Menze, B.H., Konukoglu, E., Criminisi, A., Ayache, N.: Spatial decision forests for ms lesion segmentation in multi-channel magnetic resonance images. NeuroImage 57, 378–390 (2011)
21. Prastawa, M., Bullitt, E., Gerig, G.: Simulation of brain tumors in MR images for evaluation of segmentation efficacy. Medical Image Analysis 13, 297–311 (2009)
22. Criminisi, A., Shotton, J., Konukoglu, E.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical report, Microsoft (2011)
23. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: Proc CVPR, pp. 1297–1304 (2011)
24. Clatz, O., Sermesant, M., Bondiau, P.Y., Delingette, H., Warfield, S.K., Malandain, G., Ayache, N.: Realistic simulation of the 3d growth of brain tumors in mr images coupling diffusion with mass effect. IEEE Transactions on Medical Imaging 24, 1334–1346 (2005)
25. Coltuc, D., Bolon, P., Chassery, J.M.: Exact histogram specification. IEEE TIP 15, 1143–1152 (2006)
26. Prima, S., Ourselin, S., Ayache, N.: Computation of the mid-sagittal plane in 3d brain images. IEEE Transactions on Medical Imaging 21, 122–138 (2002)

Current-Based 4D Shape Analysis for the Mechanical Personalization of Heart Models

Loïc Le Folgoc[1], Hervé Delingette[1], Antonio Criminisi[2], and Nicholas Ayache[1]

[1] Asclepios Research Project, INRIA Sophia Antipolis, France
[2] Machine Learning and Perception Group, Microsoft Research Cambridge, UK

Abstract. Patient-specific models of the heart may lead to better understanding of cardiovascular diseases and better planning of therapy. A machine-learning approach to the personalization of an electro-mechanical model of the heart, from the kinematics of the endo- and epicardium, is presented in this paper. We use 4D mathematical currents to encapsulate information about the shape and deformation of the heart. The method is largely insensitive to initialization and does not require on-line simulation of the cardiac function. In this work, we demonstrate the performance of our approach for the joint estimation of three parameters on one heart geometry. We manage to retrieve parameters such that the model matches the 4D observations with an accuracy below the voxel size, in less than three minutes of computation.

Keywords: patient-specific heart model, mechanical personalization, currents, machine-learning.

1 Introduction

Patient-specific models may help better understand the role of biomechanical and electrophysiological factors in cardiovascular pathologies. They may also prove to be useful in predicting the outcome of potential therapeutic interventions for individual patients. In this paper we focus on the mechanical personalization of the Bestel-Clement-Sorine (BCS) model, as described in [2][4].

Model personalization aims at optimizing model parameters so that the behaviour of the personalized model matches the acquired patient-specific data (e.g. cine-MR images). Several approaches to the problem of cardiac model personalization have been suggested in the recent years, often formulating the inverse problem via the framework of variational data assimilation[6] or that of optimal filtering theory[14][13][3]. The output of these methods is dependent on the set of parameters used to initialize the algorithm; for this reason calibration procedures are introduced as a preprocessing stage, such as the one developed in [16]. Furthermore these approaches rely on on-line simulations, as an accurate estimation of the effect of parameter changes along several directions in the parameter space is required to drive the parameter estimation. Due to the complexity of the direct simulation these approaches are costly in time and computations.

In this paper, we explore a novel machine-learning approach, in which the need for initialization and on-line simulation is removed, by moving the analysis of the parameter effects on the kinematics of the model (and thus the bulk of the computations) to an

B.H. Menze et al. (Eds.): MCV 2012, LNCS 7766, pp. 283–292, 2013.

off-line learning phase. In this work we assume the tracking of the heart motion from images to be given (e.g. via [15]) and focus on the mechanical personalization of the cardiac function from meshes. Our work makes use of currents, a mathematical tool which was originally introduced to the medical imaging community in the context of shape registration[18][8] and offers a unified, correspondence-free statistical representation of geometrical objects. Our main contributions include the construction of 4D currents to represent, and perform statistics on $3D + t$ beating hearts and the proposal of a machine-learning framework to personalize electromechanical cardiac models.

The remaining of this article is organized as follows. In the first part we introduce the background on currents necessary to present the rest of our work. We develop our method in the following section, then present and discuss experimental results in the final sections.

2 Currents for Shape Representation

2.1 A Statistical Shape Representation Framework

Currents provide a unified representation of geometrical objects of any dimension, embedded in the Euclidean space $\mathbb{R}^n$, that is fit for statistical analysis. The framework of currents makes use of geometrically rich and well-behaved data spaces allowing for the proper definition of classical statistical concepts. Typically the existence of an inner product structure provides a straightforward way to define the mean and principal modes of a data set for instance, as in the Principal Component Analysis (PCA). These comments motivate an approach of currents from the perspective of kernel theory in this section, although currents are formally introduced in a more general way *via* the field of differential topology. The connection to differential topology is particularly relevant to outline the desirable properties of currents when dealing with discrete approximations of continuous shapes, in terms of convergence and consistence of the representation [7].

A well-known theorem due to Moore and Aronszajn[1] states that for any symmetric, positive definite (p.d.) kernel on a set $\mathcal{X}$, there exists a unique Hilbert space $\mathcal{H}_K \subset \mathbb{R}^{\mathcal{X}}$ for which K is a reproducing kernel. This result suggests a straightforward way of doing statistics on $\mathcal{X}$ as long as a p.d. kernel K can be engineered on this set, by mapping any point $\mathrm{x} \in \mathcal{X}$ to a function $K(\mathrm{x}, \cdot) \in \mathcal{H}_K$ and exploiting the Hilbert space structure in $\mathcal{H}_K$. Furthermore, practical computations can be efficiently tracted thanks to the *reproducing kernel* property - namely, for any $\mathrm{x}, \mathrm{y} \in \mathcal{X}$, we have

$$(K(\mathrm{x}, \cdot) \mid K(\mathrm{y}, \cdot))_{\mathcal{H}_K} = K(\mathrm{x}, \mathrm{y}) \,, \tag{1}$$

and more generally yet, for any $f \in \mathcal{H}_K$, $(K(\mathrm{x}, \cdot)|f)_{\mathcal{H}_K} = f(\mathrm{x})$. Expanding on this, one can compute statistics on pairs of points and m-vectors $(\mathrm{x}, \eta) \in \mathbb{R}^n \times \Lambda_m \mathbb{R}^n$ by mapping them to functions $K(\mathrm{x}, \cdot)\eta$ and making use of the reproducing property

$$(K(\mathrm{x}, \cdot)\eta | K(\mathrm{y}, \cdot)\nu) = \eta^\top \nu K(\mathrm{x}, \mathrm{y}) \,. \tag{2}$$

Eq. 2 simply extends Eq. 1 to vector-valued functions, making use of the fact that the tensor product of two kernels is again a kernel over the product space. Expanding the

framework even further, we can regard a discrete shape as a finite set $\{(\mathrm{x}_i, \eta_i)\}_{1\leq i\leq p}$, where η_i describes the tangent space at x_i, and associate to it a signature function $\sum_{1\leq i\leq p} K(\mathrm{x}_i, \cdot)\eta_i$. The correlation between two discrete shapes $\{(\mathrm{x}_i, \eta_i)\}_{1\leq i\leq p}$ and $\{(\mathrm{y}_j, \nu_j)\}_{1\leq j\leq q}$ can then be measured by the inner product

$$(\sum_i K(\mathrm{x}_i, \cdot)\eta_i \mid \sum_j K(\mathrm{y}_j, \cdot)\nu_j) = \sum_{i,j} \eta_i^\top \nu_j K(\mathrm{x}_i, \mathrm{y}_j) \,. \tag{3}$$

This construction may in fact be acknowledged as a special case of the convolution kernel on discrete structures described in [11] and [10]. The above defines a correspondence-free way to measure proximity between shapes, trading hard correspondences for an aggregation of the measures of proximity between each simplex of one shape with every simplex of the other shape in the sense of a kernel $K(\cdot, \cdot)$. We have yet to specify a choice of kernel K. In the following, we will consider the multivariate Gaussian kernel with variance Σ:

$$K_\Sigma(\mathrm{x}, \mathrm{y}) = \frac{1}{\{(2\pi)^n |\Sigma|\}^{1/2}} \exp -\frac{1}{2}(\mathrm{x}-\mathrm{y})^\top \Sigma^{-1}(\mathrm{x}-\mathrm{y}) \,.$$

The choice of kernel width Σ can be interpreted as a choice of scale at which the shape of interest is observed: shape variations occurring at a lower scale are likely to be smoothed by the convolution and go unnoticed. This mechanism naturally introduces some level of noise insensitivity in the analysis. This parameter should thus be decided with regard to the mesh resolution and the level of noise in the data.

Finally, the linear pointwise-evaluation functional $\delta_{\mathrm{x}}^\eta \colon \omega \mapsto \omega(\mathrm{x})(\eta)$ is continuous and dual to $K(\mathrm{x}, \cdot)\eta$ by the reproducing kernel property. In the following we will refer to δ_{x}^η as a *delta-current* or a *moment*. To summarize, the discretized m-manifold $\{(\mathrm{x}_i, \eta_i)\}_{1\leq i\leq p}$ admits equivalent representations as the current $\sum_i \delta_{\mathrm{x}_i}^{\eta_i}$, its dual differential m-form $\sum_{1\leq i\leq p} K(\mathrm{x}_i, \cdot)\eta_i^\top$ or its dual vector field $\sum_{1\leq i\leq p} K(\mathrm{x}_i, \cdot)\eta_i$.

2.2 Computational Efficiency and Compact Approximate Representations

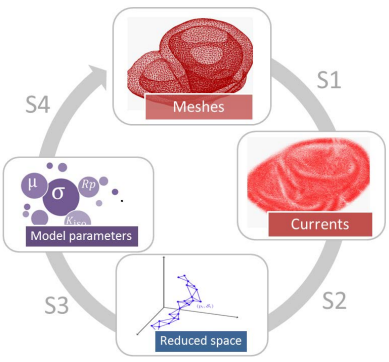
This framework lends itself to an efficient implementation. Firstly, the inner product between two discrete shapes can be computed in linear time with respect to the number of momenta through the use of a translation invariant kernel. Indeed $\gamma(\cdot) = \sum_i K(\mathrm{x}_i, \cdot)\eta_i$ may then be precomputed at any desired accuracy on a discrete grid by convolution, and rewriting $\sum_{i,j} \eta_i^\top \nu_j K(\mathrm{x}_i, \mathrm{y}_j)$ as $\sum_j \gamma(\mathrm{y}_j)^\top \nu_j$ demonstrates the claim.

Secondly, if the mesh diameter is small with respect to the scale Σ, the initial delta-current representation will be highly redundant. Durrleman et al.[9] introduced an iterative method to obtain compact approximations of currents at a chosen scale and with any desired accuracy. We rely on this procedure at training time to fasten computations and reduce the memory load. This algorithm is inspired from the Matching Pursuit method[5]. A compact current is built from the current S to approximate (of dual field γ) by iteratively adding a single delta current $\delta_{\mathrm{x}_n}^{\eta_n}$ to the previous approximation S_{n-1}, in such a way that the difference $\|S - S_n\|_{\mathcal{W}'_\Sigma}$ steadily decreases. This is achieved by greedily placing the moment at the maximum (in $\|\cdot\|_2$ norm) x_n of the residual field $\gamma(\cdot) - \gamma_{n-1}(\cdot)$, then choosing the optimal η, i.e. the one that minimizes

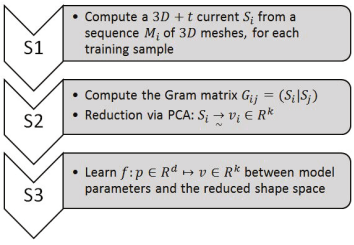
$\|\gamma - \{\gamma_{n-1} + K(\mathrm{x}_n, \cdot)\eta\}\|^2_{\mathcal{H}_\Sigma}$. It is shown in [9] that this algorithm is greedy in $\|\cdot\|_{\mathcal{H}_\Sigma}$ norm, and converges both in $\|\cdot\|_{\mathcal{H}_\Sigma}$ norm and $\|\cdot\|_\infty$ norm. The stopping criterion is on the residual norm $\|\gamma(\cdot) - \gamma_n(\cdot)\|^2_{\mathcal{H}_\Sigma}$. Our implementation uses a discrete kernel approximation of the Gaussian kernel, rather than an FFT based scheme, for fast local updates of the residual field.

3 Method

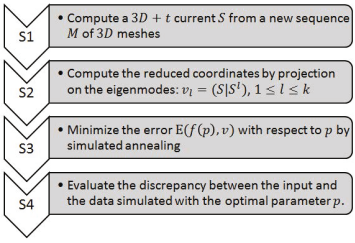
The workflow for the proposed machine-learning based parameter estimation method couples three successive processing steps: the first one aims at generating a current from an input sequence of meshes, so as to obtain a statistically relevant representation; the second one consists in a dimensionality reduction step, so as to derive a reduced shape representation in $\mathbb{R}^k$, which leads to computationally efficient statistical learning; the third step tackles the matter of finding a relationship between the reduced shape space and the (biophysical) model parameters. The three modules are mostly independent and can easily be adjusted in their own respect. As a machine learning based method, our work involves an off-line learning stage and an on-line testing stage: all three modules of the pipeline are involved during each stage. Fig. 1 gives a visual overview of our approach. The rest of this section describes the three afore-mentioned processing steps and their use during learning and testing stages.



(a) Step-by-step process



(b) Learning phase



(c) Testing phase

Fig. 1. Overview of the learning and testing phases

3.1 Current Generation from Mesh Sequences

Let us briefly describe the way we build a current from a time sequence of 3D meshes. We first extract surface meshes from the volumetric meshes. This choice derives from the assumption that the displacement of surface points can be recovered more easily than the displacement of all points within the myocardium, given a sequence of images; thus learning from surface meshes may be more relevant for real applications. In this work we assume the trajectory of surface points to be entirely known, as opposed to the displacement in the direction normal to the contour only (aperture problem). Several variants to derive currents for 4D object representation can be discussed (e.g. [7]), but their relevance largely depends on the application and complete processing work flow from the original data.

In this work, we rely on the remark that the concatenation of smoothly deformed surface meshes can be visualized as a (3D) hyper-surface in 4D (Fig. 2). The ith simplex of this hyper-surface generates a current $\delta_{\mathrm{x}_i}^{\eta_i}$, where x_i is its barycenter and η_i is the vector of $\mathbb{R}^4$ normal to its support and of length the volume of the simplex. The current associated to the series of meshes is the aggregation of such delta currents, $\sum_i \delta_{\mathrm{x}_i}^{\eta_i}$. This construction captures both the geometry of the heart and its motion.

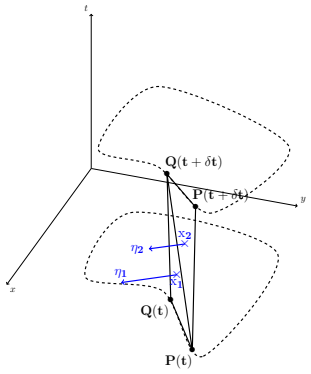


Fig. 2. Current generation from a mesh element, illustrated on an element of contour in 2D deformed in time. The simplex PQ is followed over two consecutive timesteps, which gives a quad embedded in 3D. The quad is divided into two triangles, from which we get two current deltas, applied at each triangle barycenter, orthogonal to the support of their corresponding triangles and of norm the area of the triangle. For a surface in 3D deformed over time, each element of the triangulation followed over two consecutive timesteps generates a hyper-prism embedded in 4D, which is in turn decomposed in three tetrahedra from which we obtain three momenta.

3.2 Shape Space Reduction

Since learning a direct mapping between the space of model parameters and the space of 3D+t currents is a cumbersome task, we introduce an intermediate step of dimensionality reduction via PCA. During the learning stage, we compute the mean current

and principal modes of variation from the learning database of N currents $\{S_i\}_{1\leq i\leq N}$ generated from the N training mesh sequences $\{M_i\}_{1\leq i\leq N}$ as described in §3.1. This is achieved efficiently by computing the Gram matrix of the data $\mathbf{G}_{ij} = (S_i|S_j)$ column by column and using the "kernel trick"[17]. Each column of $\mathbf{G}$ is computed in $\mathcal{O}(N \cdot P)$, where P is the maximum number of momenta among all currents S_j (cf. §2.1). Finally, we compute an approximate compact representation at the scale Σ of the mean current $\bar{T}$ and of the K first modes of variation $\{T_k\}_{1\leq k\leq K}$ to accelerate computations of inner products involving these currents[9].

At testing time and given a new current S, we derive its coordinates $v = \left(v_1, \cdots, v_K\right)$ in the reduced shape space by projection on the principal modes of variation, $v_k = (S - \bar{T}|T_k)$.

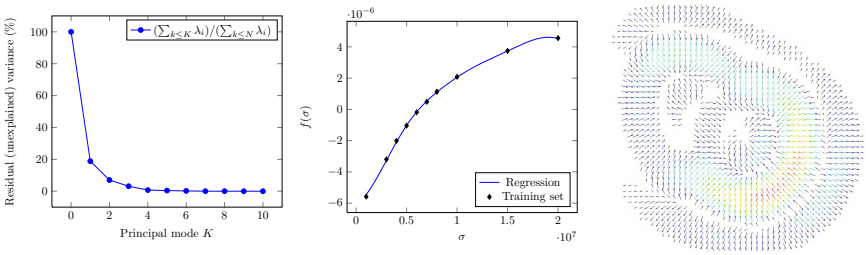
3.3 Regression Problem for Model Parameter Learning

It remains to link the physiological (model) parameters to the reduced shape space. Although we are ultimately interested in finding an optimal set of parameters $\mathrm{p} \in \mathbb{R}^d$ from an observation $v \in \mathbb{R}^K$ we will actually learn a mapping in the other direction, $f\colon \mathrm{p} \in \mathbb{R}^d \mapsto v \in \mathbb{R}^K$. We motivate this choice by three arguments. Firstly, the observation v is a deterministic output of the cardiac model given a parameter set p and thus the mapping f is well-defined; however there may be several parameter sets resulting in the same observable shape and deformation, as parameter identifiability is not *a priori* ensured. Secondly, the parameter space is expected to be of smaller dimensionality than the reduced shape space and therefore easier to sample for combinatorial reasons. Finally, we can also expect that the set of biologically admissible model parameters be relatively well-behaved; on the other hand few points in the shape space may actually relate to anatomically reasonable hearts: thus mapping every $v \in \mathbb{R}^k$ to a parameter set could be impractical.

The regression function f is learned by kernel ridge regression using a Gaussian kernel[12], and admits a straightforward close-form expression. During the testing phase, given a new observation v, we solve the optimization problem $\arg\min_{\mathrm{p}} \|f(\mathrm{p}) - v\|^2$ by Simulated Annealing[19]. This optimization problem involves an analytical mapping between low-dimensional spaces, as opposed to optimizing directly over the 4D meshes or currents. Thus it will not constitute a computational bottleneck regardless of the chosen optimization scheme. Naturally, if a prior on the likelihood of a given parameter set $\mathrm{p} \in \mathbb{R}^d$ were known (e.g. via a biophysical argument), it could be integrated in the cost function in the form of a prior energy term $\lambda \cdot R(\mathrm{p})$.

4 Experimental Results

In our first experiment we focus on the prediction of the maximum contractility parameter σ_0 of the BCS model, defined globally for the whole cardiac muscle. Building on the sensitivity analysis from [16], we consider that σ_0 covers the range of values from 10^6 to $2\ 10^7$ in an anatomically plausible way. We form a training base of ten cases $\{\mathrm{p}_i, \mathcal{M}_i\}$ by sampling this range deterministically and launching simulations with the corresponding parameter sets, for a single heart geometry from the STACOM'2011



(a) Residual variance **(b)** Regression function **(c)** First mode of PCA

Fig. 3. PCA results for the first experiment. The projection of the first mode of variation on a plane orthogonal to the z-axis at a fixed time step is shown in (c), and can be interpreted as capturing the variability in the extent of the contraction of the muscle.

dataset. Following the PCA, the first principal mode of variation is found to explain 81% of the variance, thus we set the reduced shape space to be of dimension 1 ($K = 1$); the regression function ($\sigma = 0.3$, $\lambda = 10^{-5}$) bijectively maps the model parameter space and the reduced shape space. In all experiments, the model parameters are affinely mapped to $[-1, 1]$ for convenience, for the regression and optimization stages. We use an isotropic Gaussian kernel of width 1cm in space and 50ms in time.

In the spirit of cross-validation procedures, we evaluate the performance of our approach on an independent test set $\{\mathrm{p}_j, \mathcal{M}_j\}_{0 \leq j < N'}$ by randomly choosing parameter sets in the admissible range of parameters and launching the corresponding simulations. We thereafter refer to p_j as the *real* parameter (value) and to the output of our approach p_j^* as the *optimal* parameter (value). Our test set is of size $N' = 100$ samples. The whole personalization pipeline, from the current generation to the parameter optimization phase, takes roughly 2 minutes per sample on a regular laptop. We define the relative error on the parameter value for a given test sample j as $\epsilon_r \mathrm{p}_j = |\mathrm{p}_j^* - \mathrm{p}_j|/\mathrm{p}_j$. In addition to the relative error, we consider the absolute error over the range of admissible parameters, $\epsilon_a \mathrm{p}_j = |\mathrm{p}_j^* - \mathrm{p}_j|/|\mathrm{p}_{\max} - \mathrm{p}_{\min}|$. We refer to $\epsilon_a \mathrm{p}$ as an absolute error but express it for convenience as a percentage of the admissible parameter variation. Over the test set, we found a mean relative (resp. absolute) error of 9.2% (resp. 4.5%) and a median relative (resp. absolute) error of 6.8% (resp. 2.3%).

We are also interested in a preliminary evaluation of the robustness of our approach with respect to geometry changes. Ten samples are generated following the same procedure as before, but using another heart geometry of the STACOM dataset. The 10 mesh sequences are manually registered (*via* a similarity transform) to the training geometry based on the end-diastole mesh before applying the normal pipeline, as described in Section 3. The mean relative (resp. absolute) error on the contractility parameter over our sample is 25% (9.3%), with 15% (resp. 7.5%) median relative (absolute) error.

The second experiment proceeds similarly to the first one, but we simultaneously estimate the contractility σ_0, the relaxation rate k_{rs} and the viscosity μ. For the training phase, the parameter space is sampled on a $7 \times 7 \times 7$ grid with σ_0 in the range $[10^6, 2\ 10^7]$, k_{rs} in $[5, 50]$ and μ in $[10^5, 8\ 10^5]$. The explained variance with

1 eigenmode of the PCA (resp. 2 to 5) out of the $N = 343$ modes equals 63.2% of the total variance (resp. 80.3%, 89.5%, 94.1%, 96.7%). We set the dimension of the reduced shape space to $K = 3$. The performance is tested on $N' = 100$ random samples. Because we can no longer assume the parameter set to be identifiable *a priori*, we introduce another measure of the goodness of fit of our personalization by directly evaluating the error on the observations. Given two surface mesh sequences $\mathcal{M} = \{\mathcal{M}_i\}_{1\leq i\leq T}$ and $\mathcal{M}' = \{\mathcal{M}'_i\}_{1\leq i\leq T}$, we define the pseudo-distance $d_{\text{sur}}(\mathcal{M}, \mathcal{M}') = \max_i d_s(\mathcal{M}_i, \mathcal{M}'_i)$ where $d_s(\mathcal{M}_i, \mathcal{M}'_i)^2$ is the mean square distance of the points of the surface $\mathcal{M}_i$ to the surface $\mathcal{M}'_i$. Additionally given one-to-one correspondences between $\mathcal{M}$ and $\mathcal{M}'$, we can define the distance $d_{\text{nod}}(\mathcal{M}, \mathcal{M}') = \max_i d_p(\mathcal{M}_i, \mathcal{M}'_i)$, where $d_p(\mathcal{M}_i, \mathcal{M}'_i)$ is the mean distance between corresponding nodes of $\mathcal{M}_i$ and $\mathcal{M}'_i$. While d_{sur} intuitively relates to an upper bound for the matching between surface meshes at any time step, d_{nod} conveys more information about the quality of the matching of point trajectories. The results for this experiment are reported in Table 1. As a comparison, two mesh sequences corresponding to extreme values in the parameter set will yield a value for $d_{\text{sur}}(\mathcal{M}, \mathcal{M}')$ (resp. $d_{\text{nod}}(\mathcal{M}, \mathcal{M}')$) of the order of 6mm (resp. 8mm).

Table 1. Experiment 2 - results

	$\epsilon_r\sigma_0$ ($\epsilon_a\sigma_0$)	$\epsilon_r k_{rs}$ ($\epsilon_a k_{rs}$)	$\epsilon_r\mu$ ($\epsilon_a\mu$)	d_{sur} (mm)	d_{nod} (mm)
Mean	15.2% (8.0%)	48.8% (26.4%)	40.5% (20.0%)	0.92mm	1.42mm
Median	13.2% (6.3%)	44.7% (19.6%)	32.1% (17.5%)	0.80mm	1.32mm

In addition we compute the optimal parameters and performance indicators for a different choice of the reduced space dimension K, obtaining quasi-identical statistics for $K = 4$. Finally, we test here again the robustness with respect to changes of the heart geometry. Using the same procedure as before on 10 test samples on a different geometry, we find a mean error of 1.4mm and a median value at 1.3mm for d_{sur} (respectively, 1.8mm and 1.6mm for d_{nod}).

5 Discussion

Despite working around the bias and error introduced by the model and image processing in real applications, our synthetic experiments show promising performance for our framework in terms of accuracy, tolerance to non-linear effects of parameters, robustness and computational efficiency. The accuracy of our approach was found to be below the typical voxel dimension (1mm), while a priori optimizing among a very wide range of parameter values at test time, and using a reasonable number of training samples at learning time. Although a single geometry is used for the training phase, the accuracy was of the same order on similar (non-pathological) heart geometries. Naturally, further work should handle geometry variability in a proper way, taking it into account at the training stage, and adding "shape factors" to the model parameter space capturing 3D shape variability. Moreover the addition in the pipeline of a pre-clustering stage

with respect to the heart geometry, so as to distinguish very different geometries and treat them separately, should reduce the number of samples required to cover the whole parameter space while achieving better model personalization.

The proposed framework also brings an interesting perspective on the issue of parameter identifiability. It should be noticed that we achieve good results in terms of spatial distance between the matched model and observations while significant differences in the parameter space may still be observed. Parameter identifiability encompasses two distinct aspects. Firstly, small variations of the parameter values may result in changes that are not noticeable at the scale of reference. This sensitivity to parameters partially explains the error on the retrieved set of parameters. In our approach, the kernel width for currents impacts the ability of the algorithm to discern shape differences. In the future we will experiment with smaller kernel widths and improve algorithms to handle increased computational cost. Secondly in joint parameter estimation, a whole subset in the parameter space may result in identical observations, which also affects parameter identifiability. Such considerations can be analyzed in depth at the regression or optimization steps: several parameter sets with similar costs along with a measure of local sensitivity around these values may be additionally output by the Simulated Annealing algorithm. Biophysical priors may also be introduced at the optimization step by penalizing unlikely parameter sets without adding significant computational cost.

Finally more efficient machine learning algorithms should be tested in lieu of PCA, so as to capture non-linear 4D shapes variation, and to obtain and exploit precise information about the manifold structure of 4D heart shapes. Not only will this be of help with parameter identifiability and to derive efficient representations in the reduced shape space, but it could also provide valuable feedback for "smart" sampling of the parameter space.

6 Conclusion

A machine-learning current-based method has been proposed in this paper for the personalization of electromechanical models of the heart from patient-specific kinematics. A framework to encapsulate information regarding shape and motion in a way that allows the efficient computation of statistics via 4D currents has been described. This approach has been evaluated on synthetic data using the BCS model, with the joint estimation of the maximum contraction, relaxation rate and viscosity. It is found that the proposed method is accurate, computationally efficient and robust.

Acknowledgments. This work was partly funded by Microsoft Research through its PhD Scholarship Programme and by the ERC Advanced Grant MedYMA.

References

1. Aronszajn, N.: Theory of reproducing kernels. Harvard University (1951)
2. Bestel, J., Clément, F., Sorine, M.: A Biomechanical Model of Muscle Contraction. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 1159–1161. Springer, Heidelberg (2001)

3. Chabiniok, R., Moireau, P., Lesault, P.F., Rahmouni, A., Deux, J.F., Chapelle, D.: Estimation of tissue contractility from cardiac cine-MRI using a biomechanical heart model. Biomechanics and Modeling in Mechanobiology 11(5), 609–630 (2012)
4. Chapelle, D., Le Tallec, P., Moireau, P., Sorine, M.: An energy-preserving muscle tissue model: formulation and compatible discretizations. IJMCE 10(2), 189–211 (2012)
5. Davis, G., Mallat, S., Avellaneda, M.: Adaptive greedy approximations. Constructive Approximation 13(1), 57–98 (1997)
6. Delingette, H., Billet, F., Wong, K., Sermesant, M., Rhode, K., Ginks, M., Rinaldi, C., Razavi, R., Ayache, N., et al.: Personalization of Cardiac Motion and Contractility from Images using Variational Data Assimilation. IEEE Trans. Biomed. Eng. 59(1), 20 (2012)
7. Durrleman, S.: Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution. Ph.D. Thesis, INRIA (March 2010)
8. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Measuring Brain Variability Via Sulcal Lines Registration: A Diffeomorphic Approach. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 675–682. Springer, Heidelberg (2007)
9. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Sparse Approximation of Currents for Statistics on Curves and Surfaces. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 390–398. Springer, Heidelberg (2008)
10. Gärtner, T., Flach, P., Kowalczyk, A., Smola, A.: Multi-instance kernels. In: Proceedings of the 19th International Conference on Machine Learning, pp. 179–186 (2002)
11. Haussler, D.: Convolution kernels on discrete structures. Tech. rep., Technical report, UC Santa Cruz (1999)
12. Hoerl, A., Kennard, R.: Ridge regression: Biased estimation for nonorthogonal problems. Technometrics pp. 55–67 (1970)
13. Imperiale, A., Chabiniok, R., Moireau, P., Chapelle, D.: Constitutive Parameter Estimation Methodology Using Tagged-MRI Data. In: Metaxas, D.N., Axel, L. (eds.) FIMH 2011. LNCS, vol. 6666, pp. 409–417. Springer, Heidelberg (2011)
14. Liu, H., Shi, P.: Maximum a posteriori strategy for the simultaneous motion and material property estimation of the heart. IEEE Trans. Biomed. Eng. 56(2), 378–389 (2009)
15. Mansi, T., Pennec, X., Sermesant, M., Delingette, H., Ayache, N.: ilogdemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. International Journal of Computer Vision 92(1), 92–111 (2011)
16. Marchesseau, S., Delingette, H., Sermesant, M., Rhode, K., Duckett, S.G., Rinaldi, C.A., Razavi, R., Ayache, N.: Cardiac Mechanical Parameter Calibration Based on the Unscented Transform. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 41–48. Springer, Heidelberg (2012)
17. Schölkopf, B., Smola, A.: Learning with kernels: Support vector machines, regularization, optimization, and beyond. The MIT Press (2002)
18. Vaillant, M., Glaunès, J.: Surface Matching via Currents. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 381–392. Springer, Heidelberg (2005)
19. Xiang, Y., Gubian, S., Suomela, B., Hoeng, J.: Generalized simulated annealing for efficient global optimization: the GenSA package for R. The R Journal (2012) (forthcoming)

Author Index