

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Hidden Markov Models for Analysis of Multimodal Biomedical Images

Permalink

<https://escholarship.org/uc/item/3d03p56b>

Author

Shenoy, Renuka

Publication Date

2016

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

Hidden Markov Models for Analysis of Multimodal Biomedical Images

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy

in

Electrical and Computer Engineering

by

Renuka Vidyut Shenoy

Committee in charge:

Professor Kenneth Rose, Chair
Professor B. S. Manjunath
Professor Michael Liebling
Professor Ambuj Singh

March 2016

The dissertation of Renuka Vidyut Shenoy is approved.

Professor B. S. Manjunath

Professor Michael Liebling

Professor Ambuj Singh

Professor Kenneth Rose, Committee Chair

December 2015

Hidden Markov Models for Analysis of Multimodal Biomedical Images

Copyright © 2016

by

Renuka Vidyut Shenoy

To Amma, Annu, Ajja, Rithika and Karthik,
for your unconditional love, belief and support.

Acknowledgements

I take this opportunity to thank all those who have contributed to the completion of this dissertation.

My deepest and most sincere gratitude is extended to my advisor Prof. Kenneth Rose, for his support, guidance, and encouragement throughout my graduate studies. He has guided me to take a methodical approach to problem solving and ask pertinent questions, which have always led to more refined work. His eternal enthusiasm during our discussions (often spanning hours) ensured that I was always excited about my research. His broad knowledge and attention to detail have made him a great mentor.

I would like to extend my heartfelt gratitude to Prof. B.S. Manjunath, Prof. Michael Liebling and Prof. Ambuj Singh for accepting to be on my committee, and for always being approachable. Their advice and suggestions through the course of my doctoral studies have been invaluable, and have significantly contributed to this dissertation.

I have been fortunate enough to work with people from diverse backgrounds over the past few years. I have to thank my labmates, from both the Signal Compression Lab and the Vision Research Lab, for their help and for making my graduate school experience enjoyable. I would first like to thank Min-Chi, Vignesh, Karthik and Ludovico, with whom I actively collaborated and co-authored papers,

for helping me hone my technical skills and for giving me different perspectives on the problems I was solving. I would also like to thank Tejaswi, Rahul, Yue, Shunyao, Mehdi, Carlos, Aruna, Diana, Utkarsh, Archith, Mohaymen, Thuyen, Swapna, Niloufar, Dmitry, Lakshman, Thomas, Jiejun, Santosh, Carter, Po-Yu, Mustafa, Bohan, Sina, Weiting, Kumar, Emrah and Jingning for the discussions which helped improve this dissertation.

My sincere gratitude also goes out to Dr. Robert Marc, Dr. Bryan Jones and Dr. James Anderson from the University of Utah for providing the data used in this dissertation and for helpful inputs.

I would like to thank my summer internship mentors at GE Global Research, Dr. Alberto Santamaria-Pang, Dr. Yousef Al-Kofahi, Dr. Qing Li, Dr. Dan Meyer and Dr. Anup Sood, for the opportunity to work on exciting projects, and for their guidance throughout the internship.

Finally, I would like to acknowledge the research grants from the National Science Foundation that supported the work in this dissertation - NSF OIA 0941717 and NSF III 0808772.

Curriculum Vitæ

Renuka Vidyut Shenoy

Education

- December 2015 Doctor of Philosophy
Department of Electrical and Computer Engineering
University of California, Santa Barbara
- April 2012 Master of Science
Department of Electrical and Computer Engineering
University of California, Santa Barbara
- June 2010 Bachelor of Engineering
Department of Electronics and Communication Engineering
Visvesvaraya Technological University, India

Experience

- 2011 - 2015 Graduate Student Researcher
University of California, Santa Barbara, CA
- 2015 Graduate Research Intern
GE Global Research, Niskayuna, NY
- 2012 Teaching Assistant
University of California, Santa Barbara, CA

Publications

- Renuka Shenoy, Min-Chi Shih, Kenneth Rose, Deformable Registration of Biomedical Images using 2D Hidden Markov Models , submitted to IEEE Transactions on Image Processing (TIP) 2015
- Renuka Shenoy, Min-Chi Shih, Kenneth Rose, Segmentation of Cells in Electron Microscopy Images Through Multimodal Label Transfer, IEEE International Conference on Image Processing (ICIP) 2015
- Renuka Shenoy, Min-Chi Shih, Kenneth Rose, A Probabilistic Framework for Simultaneous Segmentation and Classification of Multiple Cells in Multi-marker Microscopy Images, IEEE International Symposium on Biomedical Imaging (ISBI) 2015

Min-Chi Shih, Renuka Shenoy, Kenneth Rose, 2D Hidden Markov Model with Spatially Adaptive State-Space for Tracing Many Cells in Image Sequence, IEEE International Symposium on Biomedical Imaging (ISBI) 2015

Renuka Shenoy, Min-Chi Shih, Kenneth Rose, Hidden Markov Model-based Multi-modal Image Fusion with Efficient Training, IEEE International Conference on Image Processing (ICIP) 2014

S. Karthikeyan, Vignesh Jagadeesh, Renuka Shenoy, Miguel Eckstein, B.S. Manjunath, From Where and How to What We See, International Conference on Computer Vision (ICCV) 2013

Ludovico Megalini, Renuka Shenoy, Kenneth Rose, James Speck, John Bowers, Shuji Nakamura, Daniel Cohen, Steven DenBaars, Estimation of roughness-induced scattering losses in III-Nitride laser diodes with a photoelectro-chemically etched current aperture, Physica Status Solidi A: Applications and Materials Science (PSSA) 2015

Ludovico Megalini, Leah Kuritzky, John Leonard, Renuka Shenoy, Kenneth Rose, Shuji Nakamura, James Speck, Daniel Cohen, Steven DenBaars, Selective and controllable lateral photoelectrochemical etching of nonpolar and semipolar In-GaN/GaN multiple quantum well active regions, Applied Physics Express (APEX) 2015

Abstract

Hidden Markov Models for Analysis of Multimodal Biomedical Images

by

Renuka Vidyut Shenoy

Modern advances in imaging technology have enabled the collection of huge amounts of multimodal imagery of complex biological systems. The extraction of information from this data and subsequent analysis are essential in understanding the architecture and dynamics of these systems. Due to the sheer volume of the data, manual annotation and analysis is usually infeasible, and robust automated techniques are the need of the hour. In this dissertation, we present three hidden Markov model (HMM)-based methods for automated analysis of multimodal biomedical images. First, we outline a novel approach to simultaneously classify and segment multiple cells of different classes in multi-biomarker images. A 2D HMM is set up on the superpixel lattice obtained from the input image. Parameters ensuring spatial consistency of labels and high confidence in local class selection are embedded in the HMM framework, and learnt with the objective of maximizing discrimination between classes. Optimal labels are inferred using the HMM, and are aggregated to obtain global multiple object segmentation. We then address the problem of automated spatial alignment of images from differ-

ent modalities. We propose a probabilistic framework, constructed using a 2D HMM, for deformable registration of multimodal images. The HMM is tailored to capture deformation via state transitions, and modality-specific representation via class-conditional emission probabilities. The latter aspect is premised on the realization that different modalities may provide very different representation for a given class of objects. Parameters of the HMM are learned from data, and hence the method is applicable to a wide array of datasets. In the final part of the dissertation, we describe a method for automated segmentation and subsequent tracking of cells in a challenging target image modality, wherein useful information from a complementary (source) modality is effectively utilized to assist segmentation. Labels are estimated in the source domain, and then transferred to generate preliminary segmentations in the target domain. A 1D HMM-based algorithm is used to refine segmentation boundaries in the target image, and subsequently track cells through a 3D image stack. This dissertation details techniques for classification, segmentation and registration, that together form a comprehensive system for automated analysis of multimodal biomedical datasets.

Contents

Curriculum Vitae	vii
Abstract	ix
List of Tables	xiii
List of Figures	xiv
1 Introduction	1
1.1 The RC1 Connectome Dataset	2
1.2 Dissertation Overview	3
2 Hidden Markov Models	6
2.1 One-dimensional Hidden Markov Models	6
2.2 Two-dimensional Hidden Markov Models	18
3 Simultaneous Segmentation and Classification of Cells in Multi- marker Images	24
3.1 Introduction	25
3.2 Proposed Method	29
3.3 Experimental Results	37
3.4 Conclusion	41
4 Deformable Registration of Multimodal Images	42
4.1 Introduction	43
4.2 Background and Related Work	44
4.3 Proposed Approach	47
4.4 Experimental Results	61
4.5 Conclusion	65

5	Segmentation and Tracking of Cells through Multimodal Label Transfer	71
5.1	Introduction	72
5.2	Proposed Method	75
5.3	Experimental Validation	83
5.4	Conclusion	87
6	Conclusions and Future Work	89
6.1	Future Work	90
	Bibliography	92

List of Tables

3.1	Comparison of results on CMP data from the RC1 connectome. Classification accuracy is not reported for mean shift segmentation as the algorithm does not directly provide classification output.	39
4.1	Performance comparison of single-channel registration methods multi-subject brain MRI data, measured by DSC of grey and white matter between subjects after warping.	63
4.2	Performance comparison of multi-channel registration methods on connectome data, measured by DSC of between ATEM images and warped CMP images.	65
5.1	F-measure statistics of segmentation results from various methods.	86
5.2	F-measure statistics comparison for tracking.	86

List of Figures

2.1	A one-dimensional hidden Markov model, with hidden states (grey squares) and observable emissions (circles).	7
2.2	A two-dimensional hidden Markov model build on a first order Markov mesh random field.	19
2.3	Turbo decoding of a 2D HMM, where alternating horizontal and vertical passes consist of separately decoded rows and columns that induce priors on each other.	22
3.1	(a) CMP data (from the RC1 connectome) consisting of 6 channels, each corresponding to a unique marker (b) RGB visualization CMP data channels taken three at a time.	27
3.2	Challenges in segmenting CMP images (a) Molecular marker artifact (b) “Hole” inside a cell (c) Cell having low contrast with background (d-e) Change in intra-cell feature variance - low variance in (d), high variance in (e)	31
3.3	Examples of segmentation results showing challenging as well as easy cases. Each column (a) - (g) shows results on a specific cell. The first row shows the ground truth segmentation, with the boundary outlined in white. The second, third and fourth rows show results obtained from mean shift segmentation, FVS and the proposed method respectively, along with the F-measure F of the resulting segmentation. (An F-measure of 0 indicates a missed detection.) To visualize each result, we choose the three channels exhibiting the highest contrast between the given cell and background. . . .	40
4.1	Mapping feature vector $\mathbf{S}_{x,y}$ at point (x, y) in the source image to a feature vector $\mathbf{T}_{x+\tau_x, y+\tau_y}$ at point $(x + \tau_x, y + \tau_y)$ in the target image using a translation τ	48

4.2	Neighboring translations are correlated.	52
4.3	Visual results on multi-subject MRI data. Results are shown as a checkerboard, where neighboring tiles come from different subjects. (a) Source-target pair before registration (b-e) After registration using (b) MIND (c) SSC (d) DROP and (e) Proposed method . . .	63
4.4	Visual results on multi-subject MRI data. Results are shown as a checkerboard, where neighboring tiles come from different subjects. (a) Source-target pair before registration (b-e) After registration using (b) MIND (c) SSC (d) DROP and (e) Proposed method . . .	64
4.5	Visual results on a relatively easy cell. (a) RGB visualization of 3 CMP channels, with ground truth of the cell of interest outlined in yellow (b) RGB visualization of the remaining 3 CMP channels, with outlined ground truth (c) ATEM image corresponding to the same region, with ground truth overlaid in red (d-h) ATEM image with results from various registration methods overlaid in reg, along with the corresponding DSC. (d) MIND (e) SSC (f) α -MI (g) CAMIR (h) Proposed method	66
4.6	Visual results on a relatively easy cell. (a) RGB visualization of 3 CMP channels, with ground truth of the cell of interest outlined in yellow (b) RGB visualization of the remaining 3 CMP channels, with outlined ground truth (c) ATEM image corresponding to the same region, with ground truth overlaid in red (d-h) ATEM image with results from various registration methods overlaid in reg, along with the corresponding DSC. (d) MIND (e) SSC (f) α -MI (g) CAMIR (h) Proposed method	67
4.7	Visual results on a challenging cell. (a) RGB visualization of 3 CMP channels, with ground truth of the cell of interest outlined in yellow (b) RGB visualization of the remaining 3 CMP channels, with outlined ground truth (c) ATEM image corresponding to the same region, with ground truth overlaid in red (d-h) ATEM image with results from various registration methods overlaid in reg, along with the corresponding DSC. (d) MIND (e) SSC (f) α -MI (g) CAMIR (h) Proposed method	68

4.8	Visual results on a challenging cell. (a) RGB visualization of 3 CMP channels, with ground truth of the cell of interest outlined in yellow (b) RGB visualization of the remaining 3 CMP channels, with outlined ground truth (c) ATEM image corresponding to the same region, with ground truth overlaid in red (d-h) ATEM image with results from various registration methods overlaid in reg, along with the corresponding DSC. (d) MIND (e) SSC (f) α -MI (g) CAMIR (h) Proposed method	69
5.1	An example of a region in an ATEM image. We observe considerable variation in appearance within and across cells, low contrast between cells and cell boundaries and background clutter.	73
5.2	Mapping a label from a SLIC superpixel in CMP to a SLIC superpixel in ATEM, using a mapping Δ	77
5.3	(a) Construction of a deformable trellis around the initial contour. The black curve is the initial contour, the dashed lines are the constructed normals, and each black dot corresponds to a state. (b) An example of an estimated contour (in red) obtained by joining a given sequence of states (red dots).	80
5.4	(Best in color) Visual results of cell segmentation in a challenging scenario. (a) Region of interest in the ATEM image (b) Ground truth of the cell overlaid in red (c-i) Image with overlaid segmentation results from (c) membrane detection method (Kaynig et al.) (d) DROP (e) α -MI (f) MIND (g) SSC (h) Graph cuts (i) HMM-based multimodal fusion (j) the proposed approach	84
5.5	Tracking results on an ATEM sequence. The first row shows ground truth, the second row shows results of Chan-Vese tracking, and the third row shows results of the HMM-based tracking algorithm. . .	87

Chapter 1

Introduction

Modern advances in imaging technology have enabled the collection of huge amounts of image data of complex biological systems. The extraction of information from this image data and subsequent analysis and interpretation on the information are the central tasks in the fast-growing field of biomedical image computing. In the recent past, there has been an increased interest in the development of “omics” fields, which are fields in biology devoted to the characterization and quantification of organisms in terms of their anatomy, physiology and dynamics. Due to the sheer size of these datasets, manual annotation and analysis is usually not feasible - the development of robust automated and semi-automated techniques is critical for analysis and diagnosis.

Connectomics [1] is an omics field which aims at comprehensively mapping an

organism's neural connections at various scales. Analyzing this map of connections is essential in understanding the the architecture and dynamics of the nervous system. To assist in the understanding and interpretation of connectomes, tissues may be imaged with two or different modalities, resulting in complementary types of information (eg., structural and functional). While multimodal image data can be of great utility in providing insights about the tissue being imaged, it is often challenging to work with and usually requires specialized methods for automated analysis.

1.1 The RC1 Connectome Dataset

The methods described in this dissertation have all been tested on the RC1 connectome [2], the first practical connectome dataset from mammalian retina. RC1 is a multimodal built by imaging a 0.25 mm diameter, 370 serial section column of rabbit retinal tissue. 341 of these slices are acquired using an automated transmission electron microscope (ATEM) at a resolution of 2.18 nm/pixel. Computational Molecular Phenotyping (CMP) is a light microscopy modality that is used to image the remaining 29 sections of the volume. In CMP, each physical slice is probed using one of six molecular markers, namely glutamate, glutamine, glycine, taurine, 4-aminobutyrate (GABA), and the excitation marker 1-amino-4-guanidobutane (AGB). The ATEM modality clearly shows the structure of cells,

but does not give any substantial functional information, while CMP images form a complementary source of information that do not show structure clearly but provide functional information. RC1 contains a total of 1132 cells belonging to 7 major cell classes. We focus our attention mainly on the first few slices, which consist of a ‘capstone’ CMP section containing six molecular marker images, followed by thirty ATEM slices, and which contains 581 cells.

Through the dataset, there are slices that were skipped due to damaged tissue. Physical defects (tissue folding/breakage) in the tissue being imaged is also present in some portions of RC1, leading to large areas that are unusable in the acquired image data. We manually mark damaged regions or slices of the data before analysis to prevent errors later in the process.

The data within RC1 totals 16.5 terabytes, and it is estimated to require over a million annotations to completely label the structures within the dataset. Large scale automated annotation can hence be of significant help in reducing time to downstream analysis.

1.2 Dissertation Overview

We begin with a short discussion on hidden Markov models in Chapter 2.

In Chapter 3, we describe a probabilistic approach to simultaneously classify and segment multiple cells of different classes in a multi-variate setting. Super-

pixels are extracted from the input vector-valued image, and a 2D hidden Markov model (HMM) is set up on the superpixel graph. HMM emission probabilities are used to ensure high confidence in local class selection based on superpixel feature vectors. Spatial consistency of labels is enforced by proper choice of transition probabilities, which are conditioned on the feature vectors of neighboring superpixels at each location. Optimal superpixel-level class labels are inferred using the HMM, and are aggregated to obtain global multiple object segmentation.

In Chapter 4, we describe a method for deformable registration with a generic theoretical formulation. Smoothness is ensured via transition probabilities of the 2D HMM and cross-modality similarity via class-conditional, modality-specific emission probabilities. The method is first derived for unimodal data and then extended to the multimodal, multi-channel setting. We also describe an edge-adaptive constraint which allows for variation in degree of smoothness across the image.

Chapter 5 describes a method to effectively utilize complementary information, if available, in ATEM segmentation. Images of both modalities are oversegmented into superpixels. A 2D HMM is set up on the superpixel graph to determine the optimal superpixel mapping between images. This mapping is used to transfer labels and generate preliminary segmentations in the ATEM domain, whose boundaries are then refined, to eliminate imprecisions due to the superpixel grid,

using a 1D HMM based contour refinement method. The refined cell boundary is now used to initialize a tracking algorithm through the ATEM stack. As an added benefit, this method intrinsically transfers cell label information from CMP to ATEM.

Finally, in Chapter 6, we discuss possible future directions and provide concluding remarks.

The complementary techniques described in this dissertation, when applied in conjunction with each other, form a complete system for classification, registration, segmentation and tracking of multimodal data.

Chapter 2

Hidden Markov Models

In this chapter, we provide a brief overview of conventional (one-dimensional) hidden Markov models and the algorithms involved in their optimization, followed by a discussion on extending to two dimensions and addressing the issue of high computational complexity associated with it.

2.1 One-dimensional Hidden Markov Models

One-dimensional hidden Markov model (1D HMMs) were introduced in the late 1960s and early 1970s, with a series of theoretical papers [3, 4, 5]. Subsequently, they have been used in applications in many fields, most notably speech processing [6, 7, 8, 9]. There are many reasons for the widespread use and success of the HMM. First, due its rich mathematical structure, the HMM can form

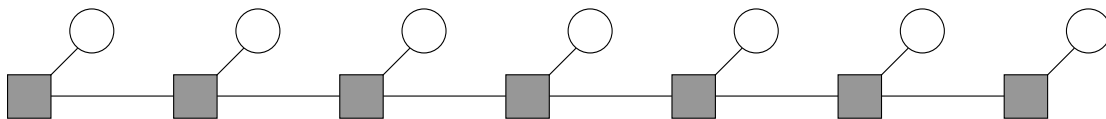


Figure 2.1: A one-dimensional hidden Markov model, with hidden states (grey squares) and observable emissions (circles).

the theoretical basis for even complex applications. Second, it has been found to work well in practice for a variety of applications when applied properly. Finally, the central problems associated with HMM - namely inferring the optimal state sequence given observation data and a model, and learning the parameters of the model - can be solved using efficient, effective algorithms. A comprehensive tutorial on the theory of HMMs and some applications in speech processing may be found in [10].

The HMM is a doubly embedded stochastic process - it consists of an underlying stochastic Markov process that is not observable (i.e., hidden), but can only be observed through a set of visible stochastic processes that produce observations (“emissions”).

An HMM is characterized by the following:

- (1) N , the number of states in the model. Although states are hidden, they may have some significance in the context of the application. The set of states is usually denoted by $S = \{S_1, S_2, \dots, S_N\}$, and the state at a given time t by q_t .

(2) M , the number of distinct observable symbols per state (in the case of discrete emissions) or a function describing the emission mechanism (in the case of continuous emissions). Individual symbols are usually denoted by $V = \{v_1, v_2, \dots, v_M\}$, and the observation at a given time t by o_t .

(3) The state transition probability distribution $A = \{a_{ij}\}$, where

$$a_{i,j} = P(q_{t+1} = S_j | q_t = S_i), \quad 1 \leq i, j \leq N.$$

(4) The observation probability distribution for each state j , given by $B = \{b_j(k)\}$, where

$$b_j(k) = P(v_k \text{ at } t | q_t = S_j), \quad 1 \leq j \leq N, 1 \leq k \leq M$$

for discrete emission HMMs and $B = \{b_j(o)\}$

$$b_j(o_t) = p(o_t | q_t = S_j), \quad 1 \leq j \leq N$$

for continuous emission HMMs.

(5) The initial state distribution $\pi = \{\pi_i\}$ where

$$\pi_i = P(q_1 = S_i) \quad 1 \leq i \leq N.$$

Given the above parameters, an HMM-based system operates as follows:

1. Set $t = 1$ and choose an initial state $q_1 = S_i$ according to π , and generate o_1 according to B .
2. Transit to a new state q_{t+1} according to A , set $t = t + 1$ and generate o_t according to B .
3. Repeat the previous step till termination, $t = T$.

This results in an observation sequence $O = o_1 o_2 \cdots o_T$, with an underlying state sequence $Q = q_1 q_2 \cdots q_T$. From the above description, it is clear that an HMM is completely specified by the probability measures. The notation

$$\lambda = (A, B, \pi) \tag{2.1}$$

is used to denote the complete parameter set of the HMM, or equivalently, the “model”.

2.1.1 The Three Basic Problems for HMMs

In order to use the HMM as described above in applications of practical interest, three fundamental problems must be solved.

Problem 1: How to efficiently compute the probability of a given observation sequence, given the model?

Problem 2: How to choose the optimal state sequence corresponding to the given observation sequence, given the model?

Problem 3: How to adjust the model parameters to maximize probability of the given observation sequence?

Effective solutions with formal mathematical foundations exist for each of these problems, and are described below.

Problem 1 - The Evaluation Problem

Under the assumption that observations are statistically independent, the probability of an observation sequence O for a given state sequence Q is

$$P(O|Q, \lambda) = b_{q_1}(o_1) \cdot b_{q_2}(o_2) \cdots b_{q_T}(o_T). \quad (2.2)$$

The probability of the state sequence is given by

$$P(Q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T}. \quad (2.3)$$

The joint probability of O and Q is the product of the above two terms.

$$P(O, Q|\lambda) = P(O|Q, \lambda)P(Q|\lambda). \quad (2.4)$$

The probability of O given the model can be obtained by summing the joint probability over all possible state sequences.

$$P(O|\lambda) = \sum_{\text{all } Q} P(O|Q, \lambda)P(Q|\lambda) \quad (2.5)$$

$$= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \cdots a_{q_{T-1} q_T} b_{q_T}(o_T) \quad (2.6)$$

Naively computing $P(O|\lambda)$ according to this equation requires $2T \cdot N^T$ computations, but with the use of the forward-backward procedure [4] described below, this computation can be solved at $O(N^2T)$ complexity.

First, the forward variable $\alpha_t(i)$ is defined as the joint probability of the partial observation sequence from time 1 to time t and the state S_i at time t , i.e.,

$$\alpha_t(i) = P(o_1 o_2 \cdots o_t, q_t = S_i | \lambda).$$

The value of $\alpha_t(i)$ can be solved for inductively, as follows:

(i) Initialization:

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N. \quad (2.7)$$

(ii) Induction:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}), \quad 1 \leq t \leq T-1, 1 \leq j \leq N. \quad (2.8)$$

(iii) Termination:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i). \quad (2.9)$$

Thus, the evaluation problem is solved using the forward variable. At this point, we also define the backward variable $\beta_t(i)$ as the probability of the partial observation sequence from time $t+1$ to time T and the state S_i at time t , i.e.,

$$\beta_t(i) = P(o_{t+1}o_{t+2}\cdots o_T | q_t = S_i, \lambda).$$

While the backward variable is not used in the solution of this problem, we use it in the following problems. The value of $\beta_t(i)$ can also be solved for inductively, as follows:

(i) Initialization:

$$\alpha_T(i) = 1 \qquad 1 \leq i \leq N. \qquad (2.10)$$

(ii) Induction:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \dots, 1, \quad 1 \leq i \leq N. \quad (2.11)$$

Problem 2 - The Decoding Problem

Here, the task is to solve for the optimal state sequence under a given optimality criterion. There are several possible reasonable criteria for decoding; we detail solutions for two commonly used criteria.

Criterion 1 Choose states which are most individually most likely, i.e.,

$$q_t = \operatorname{argmax}_{S_i} P(q_t = S_i | O, \lambda).$$

To solve this problem, we define the occupancy probability $\gamma_t(i)$ as the probability of being in state S_i and time t , given the observation sequence and the model.

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \qquad (2.12)$$

This can be expressed in terms of the forward and backward variables as

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^N \alpha_t(i)\beta_t(i)} \quad (2.13)$$

since $\alpha_t(i)$ measures probability of both the observation sequence from time point 1 through t and the state S_i at t , while $\beta_t(i)$ measures probability of the observation sequence from time point $t + 1$ through T . Due to the normalization factor, $\gamma_t(i)$ can be interpreted as a probability measure, i.e., $\sum_{i=1}^N \gamma_t(i) = 1 \quad \forall t$, and the optimal state sequence is simply the sequence of individually most likely states,

$$q_t = \operatorname{argmax}_{S_i} \gamma_t(i), \quad 1 \leq t \leq T. \quad (2.14)$$

Criterion 2 Choose states which result in the single most likely state sequence, i.e.,

$$Q^* = \operatorname{argmax}_Q P(Q, O|\lambda).$$

The Viterbi algorithm [11] is a dynamic programming method that is widely used to solve this problem, with $O(N^2T)$ complexity. We first define $\delta_t(i)$ as the highest probability along the single path at time t , which accounts for the first t

observations and ends with state S_i , i.e.,

$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P(q_1 q_2 \cdots q_{t-1}, q_t = i, o_1 o_2 \cdots o_t | \lambda). \quad (2.15)$$

By induction, we obtain $\delta_{t+1}(j)$ as

$$\delta_{t+1}(j) = [\max_i \delta_t(i) a_{ij}] \cdot b_j(o_{t+1}). \quad (2.16)$$

To retrieve the most likely state sequence, we keep track of the arguments that maximize (2.16) at each t and j . This is done via a trellis, denoted by $\psi_t(j)$, as follows:

(i) Initialization:

$$\delta_1(i) = \pi_i b_i(o_1) \quad 1 \leq i \leq N. \quad (2.17)$$

$$\psi_1(i) = 0. \quad (2.18)$$

(ii) Recursion:

$$\delta_t(j) = [\max_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}] \cdot b_j(o_{t+1}), \quad 1 \leq i \leq N, 2 \leq t \leq T \quad (2.19)$$

$$\psi_t(j) = [\operatorname{argmax}_{1 \leq i \leq N} \delta_{t-1}(i) a_{ij}], \quad 1 \leq i \leq N, 2 \leq t \leq T \quad (2.20)$$

(iii) Termination:

$$P^* = [\max_{1 \leq i \leq N} \delta_t(i)] \quad (2.21)$$

$$q_T^* = [\operatorname{argmax}_{1 \leq i \leq N} \delta_t(i)]. \quad (2.22)$$

(iv) State sequence backtracking:

$$q_t^* = \psi_{t+1}^*(q_{t+1}^*), \quad t = T - 1, T - 2, \dots, 1 \quad (2.23)$$

Problem 3 - The Training Problem

The problem of parameter estimation is the most difficult of the three problems - there are no known analytical solutions and no algorithm that guarantees optimality in estimation of the model parameters given a finite observation sequence. As a result, popular training techniques update parameters such that $P(O|\lambda)$ is maximized locally, using techniques such as gradient descent [8] or expectation-maximization(EM) [12]. Here, we discuss the Baum-Welch algorithm [5], which is an implementation of EM. (Extensive discussions on the Baum-Welch algorithm can be found in [13].) First, we define $\xi_t(i, j)$, the probability of being in states S_i and S_j at times t and $t + 1$ respectively, given the model and the observation

sequence.

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j | O, \lambda) \quad (2.24)$$

This can be rewritten in terms of the forward-backward variables as

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_t) \beta_{t+1}(j)}{P(O|\lambda)} = \frac{\alpha_t(i) a_{ij} b_j(O_t) \beta_{t+1}(j)}{\sum_{i,j=1}^N \alpha_t(i) a_{ij} b_j(O_t) \beta_{t+1}(j)}. \quad (2.25)$$

From the definition of $\gamma_t(i)$, we can relate $\gamma_t(i)$ and $\xi_t(i, j)$ as follows:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (2.26)$$

We denote the current model parameters by $\lambda = (A, B, \pi)$ and the re-estimated model parameters by $\lambda' = (A', B', \pi')$. Baum's auxiliary function is then defined as

$$\mathcal{Q}(\lambda, \lambda') = \sum_Q P(O, Q|\lambda) \log P(O, Q|\lambda'). \quad (2.27)$$

It has been proven that maximization of $\mathcal{Q}(\lambda, \lambda')$ leads to an increased likelihood, i.e.,

$$\max_{\lambda'} \mathcal{Q}(\lambda, \lambda') \implies P(O|\lambda') > P(O|\lambda), \quad (2.28)$$

resulting in the likelihood eventually converging to a critical point through repeated applications.

2.2 Two-dimensional Hidden Markov Models

A natural progression would be to leverage the ability of 1D HMMs to effectively model the behavior of one-dimensional signals, and extend to two dimensions (i.e., model spatial behavior). The Markov random field (MRF) is the 2D counterpart of the 1D Markov chain, where the notion of time-based ordering is replaced by the concept of spatial neighborhood. We consider a commonly used subclass of MRF, a first order Markov mesh random field (MMRF), where the neighborhood of a node is defined as the set of nodes which are horizontally and vertically adjacent to it. The concept of past, present and future which are used in Markov chains can now be reintroduced to MMRFs. If $Q = \{q_{x,y}, x = 1, 2, \dots, X, y = 1, 2, \dots, Y\}$ is an $X \times Y$ array of states and $Q_{x,y} = \{q_{m,n}, m < x \text{ or } n < y\}$ represents the set of states to the left or above $q_{x,y}$, the first order MMRF can be defined by the following property:

$$P(q_{x,y}|Q_{x,y}) = P(q_{x,y}|q_{x,y-1}, q_{x-1,y}). \quad (2.29)$$

With this definition, we can apply the idea of hidden states and observable emissions to the MMRF (see Fig. 2.2) as we did with Markov chains, and develop algorithms for the fundamental optimization problems for 2D HMMs. However, though the described algorithms are computationally efficient for the 1D case,

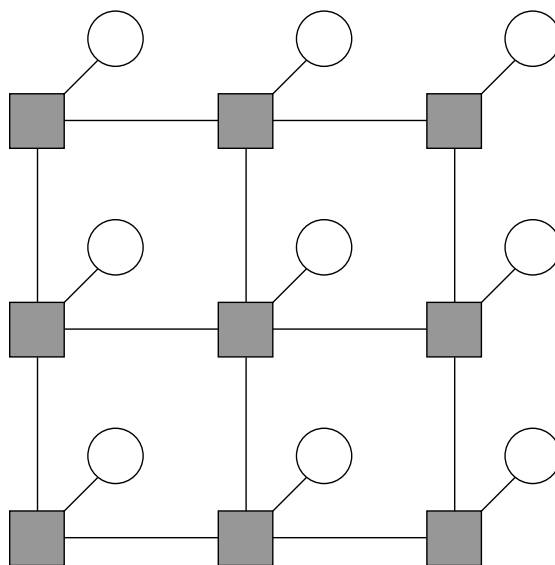


Figure 2.2: A two-dimensional hidden Markov model build on a first order Markov mesh random field.

directly extending them to 2D leads to an exponential increase in complexity and is intractable in most practical applications. There has been considerable interest in developing algorithms that approximate the performance of 2D HMMs at a lower complexity. Since the properties of 1D HMMs are well understood, most approaches approximate 2D HMMs with one [14] or more [15, 16, 17] 1D HMMs. The Path Constrained Variable State Viterbi (PCVSV) [14] limits the Viterbi search space, considering only the K state sequences having the highest observation probability. A suitable value of K is selected under the trade-off between complexity and accuracy. The Pseudo 2D HMM [15] consists of a set of “super” states, assumed to be Markovian, which subsume a set of simple Markovian states. The turbo-HMM (T-HMM) [16] assumes separability of horizontal

and vertical dependencies, and decodes the HMM separately as rows and columns that “communicate”. The Conditional Iterative Decoding (CID) [17] algorithm builds on the idea of T-HMMs, removing the requirement for horizontal-vertical separability at the cost of increased computation time. The T-HMM approximation has been shown [16] to outperform related algorithms while maintaining low run time, and is used for training and decoding 2D HMMs throughout this dissertation.

2.2.1 Turbo Hidden Markov Models

The turbo hidden Markov model was introduced in [16] and consists of horizontal and vertical 1D HMMs that are decoded separately but “communicate” in a manner similar to decoding of turbo codes. Given an $X \times Y$ two-dimensional sequence of states $Q = \{q_{x,y}, x = 1, 2, \dots, X, y = 1, 2, \dots, Y\}$, a set of observations $O = \{o_{x,y}, x = 1, 2, \dots, X, y = 1, 2, \dots, Y\}$, and model parameters λ , the joint likelihood of the 2D HMM can be written as

$$P(O, Q|\lambda) = P(O|Q, \lambda)P(Q|\lambda) \quad (2.30)$$

$$= \prod_{x,y} P(o_{x,y}|q_{x,y}, \lambda)P(q_{x,y}|q_{x,y-1}, q_{x-1,y}, \lambda) \quad (2.31)$$

The key assumption in the derivation of the T-HMM is that of separability, i.e., that the transition probability $P(q_{x,y}|q_{x,y-1}, q_{x-1,y})$ can be decomposed into the product of horizontal and vertical components. (For the sake of simplicity, λ is not explicitly included in this discussion hereafter.)

$$P(q_{x,y}|q_{x,y-1}, q_{x-1,y}) \propto P(q_{x,y}|q_{x,y-1})P(q_{x,y}|q_{x-1,y}) \quad (2.32)$$

Let the observations of row x and column y be denoted by o_x^H and o_y^V respectively, and the corresponding state sequences by q_x^H and q_y^V , where the superscripts H and V denote horizontal and vertical directions, respectively. In the T-HMM approximation, the joint likelihood is derived as

$$P(O, Q|\lambda) \approx \prod_x \left[P(o_x^H, q_x^H) \prod_y P(q_{x,y}|o_y^V) \right] \quad (2.33)$$

$$\approx \prod_y \left[P(o_y^V, q_y^V) \prod_x P(q_{x,y}|o_x^H) \right]. \quad (2.34)$$

Alternating horizontal and vertical passes are applied while decoding the 2D HMM. Each row is represented by a 1D HMM during a horizontal pass, and each column by a 1D HMM during a vertical pass (see Fig. 2.3). During the horizontal pass, a modified version of the forwards-backwards algorithm is applied to each row, and to each column in a vertical pass. Posterior probabilities obtained from

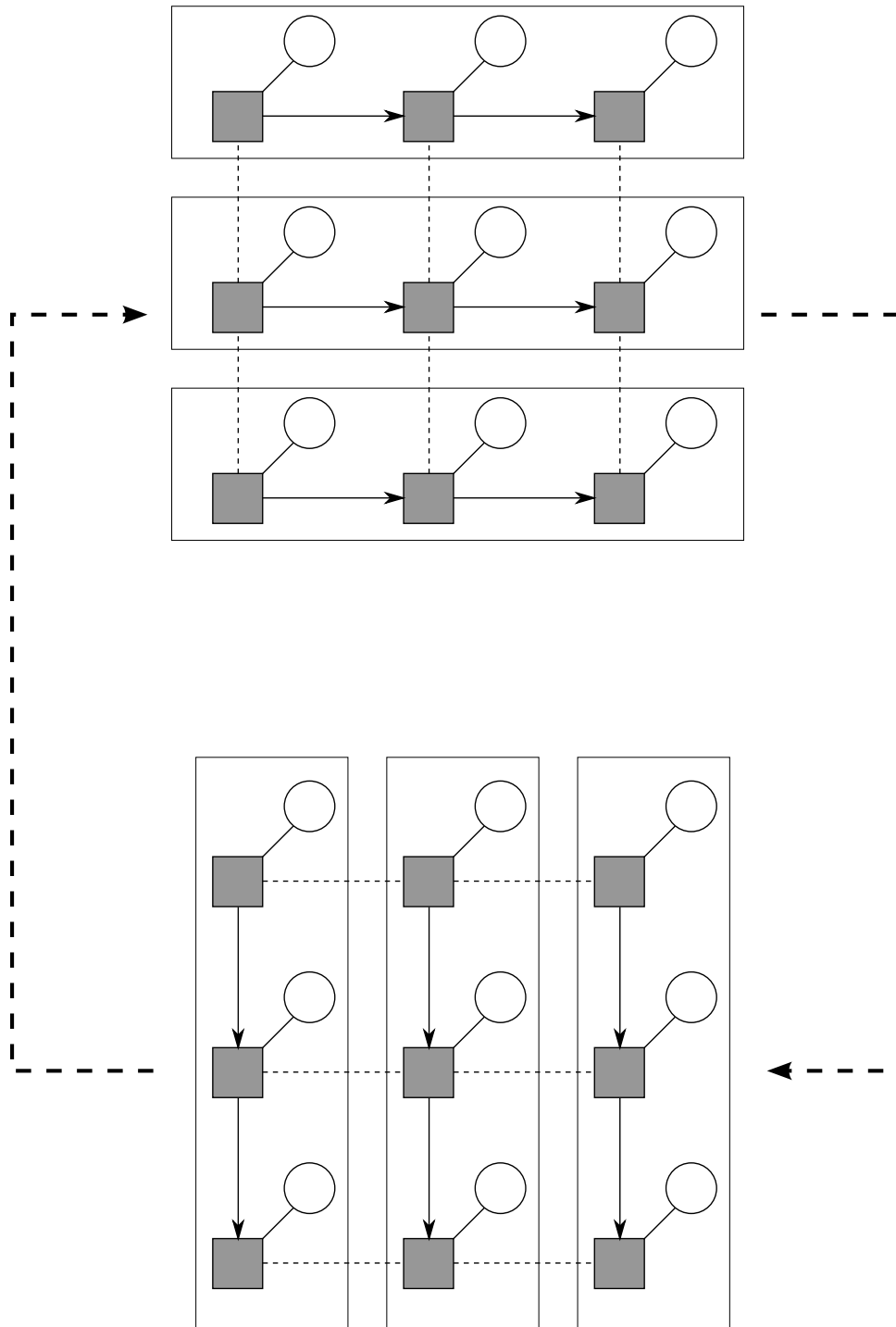


Figure 2.3: Turbo decoding of a 2D HMM, where alternating horizontal and vertical passes consist of separately decoded rows and columns that induce priors on each other.

horizontal passes are incorporated as prior probabilities on vertical passes and vice-versa. Optimal state sequences are estimated via delayed decisions on posterior probabilities at each node, rather than a greedy “winner-take-all” scheme. The iterations are repeated until vertical and horizontal passes converge to a required degree of agreement.

The T-HMM framework allows us to utilize the efficient algorithms from 1D HMMs for both learning [18] and inference [16] problems. The computational complexity of decoding is $O(XYN^2)$, where XY is the number of all nodes and N , the number of states. In practice, the algorithm can be further accelerated using a parallel implementation.

Chapter 3

Simultaneous Segmentation and Classification of Cells in Multi-marker Images

Automated segmentation and classification of cells are basic tasks in bio-medical image processing, and form the vital initial steps in single cell analysis. We begin this chapter with a discussion on related methods in literature. Then, we describe a novel probabilistic approach to simultaneously classify and segment multiple cells of different classes in a multi-variate setting. The algorithm is designed to handle various challenging aspects of microscopy images. Finally, the performance of the proposed approach is demonstrated on a challenging microscopy dataset.

Experiments show, both quantitatively and qualitatively, that the proposed approach effectively segments and classifies cells, outperforming related techniques.

We conclude the chapter with a short closing discussion.

3.1 Introduction

Multi-variate analysis is an active area of research in bioinformatics, with applications in sub-fields as diverse as genetic studies [19], hyperspectral imaging [20] and analysis of microscopy data [21]. Multi-variate methods offer many advantages over traditional methods. Jointly analyzing data provided by several markers can provide insight into correlation between phenotypes. Further, while traditional univariate methods require specialized markers for each class, multi-variate methods can target multiple classes via different combinations of a small number of probes. It is also possible to discover new classes without explicitly probing for them. Finally, as multi-variate methods take into account the response of multiple markers, they may be more robust to variations that may be encountered over large volumes of data.

The computational molecular phenotyping [21] modality in the RC1 dataset is an example of a multi-marker microscopy modality. The six intensity images (each obtained using a different micromolecular marker) can be interpreted as a single six-channel image, in which each cell class can be characterized by its

multidimensional micromolecular “signature”. For the purpose of visualization, the intensity images may be taken three at a time as pseudocolor images (see Fig. 3.1). Segmentation and classification of CMP is a critical step in the analysis of RC1 since cell types, their molecular phenotypes and their response to stimuli form an important source of supplementary information to neuronal connectivity data.

Due to the nature of the imaging system, CMP images are often noisy and contain artefacts. Fig. 3.2 highlights some challenging aspects of the data. These include sudden spurious absence or presence (Fig. 3.2 (a)) of a marker within a cell, existence of sub-cellular bodies which are resistant to staining (Fig. 3.2 (b)), cell classes that are difficult to detect due to low background contrast (Fig. 3.2 (c)) and large changes in intra-class feature variance across classes (Fig. 3.2 (d-e)). The proposed segmentation algorithm addresses these issues in its formulation.

3.1.1 Related Work

An early approach to segmentation of cells in microscopy images, which still remains popular, is histogram-based intensity thresholding [22]. Several improvements of the basic algorithm have been suggested [23, 24], which address the issue of illumination variation common in microscopy images. The natural extension of thresholding to multi-channel data would involve processing each channel indi-

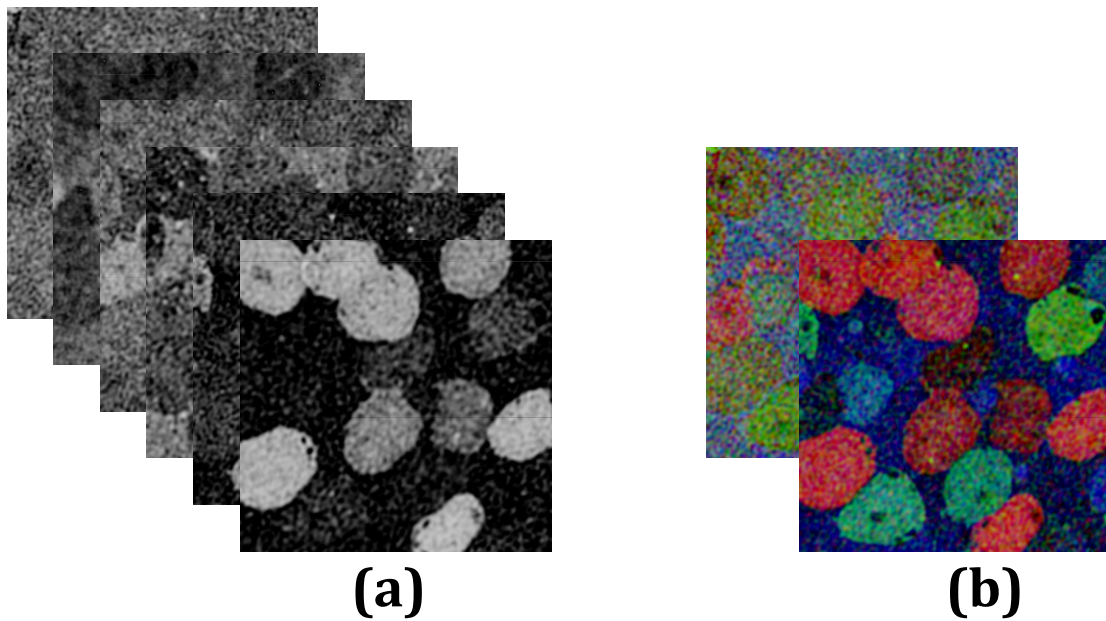


Figure 3.1: (a) CMP data (from the RC1 connectome) consisting of 6 channels, each corresponding to a unique marker (b) RGB visualization CMP data channels taken three at a time.

vidually and combining the resulting quantized data. However, such approaches ignore relations between different channels, which may provide important information for the segmentation task.

A popular method used for the segmentation of multi-variate images is the mean shift [25] algorithm. In this method, spatial coordinates are used in conjunction with feature vectors to determine local clusters, using user-defined spatial and feature bandwidths and minimum segment size parameters. However, in our experiments, we found that there is often no set of parameters that results in accurate segmentation over the large range of cell sizes and class variances present in CMP data.

There have been several papers that use random field formulations for classification and segmentation of objects. A superpixel-based technique was proposed in [26], in which a support vector machine (SVM)-based classifier is constructed on the histogram of local features and a conditional random field (CRF) is used to refine classification. In [27], a random forest classifier is used with hierarchical CRF to segment and classify images at multiple scales. The authors of [28] use a global bag of features model to combine top-down and bottom-up potentials to solve to segment multiple classes of objects. The drawback of the described methods is that they are not equipped to handle problems frequently occurring in microscopy data, such as varying contrasts and imaging artifacts.

In the method described below, we address the challenges presented by the data. In our design, superpixels are first extracted from the input vector-valued image, and a 2D hidden Markov model (HMM) is set up on the superpixel graph. HMM emission probabilities are used to ensure high confidence in local class selection based on superpixel feature vectors. Spatial consistency of labels is enforced by proper choice of transition probabilities, which are conditioned on the feature vectors of neighboring superpixels at each location. Optimal superpixel-level class labels are inferred using the HMM. Finally, contiguous regions with the same label are aggregated to obtain global multiple object segmentation.

3.2 Proposed Method

For data consisting of N cell types, we consider an M -class classification problem, where $M = N + 1$ (the additional “non-cell” class provides for background points in the image which lie between cells). We operate in the D -dimensional vector space, where D is the number of channels in the multi-marker image. Directly predicting the class label of each superpixel from its feature vector often results in incorrect labeling due to imaging artifacts or noise, or when even the label with the highest posterior probability has low confidence. Utilizing information from adjacent superpixels can help overcome this problem as there is typically a high probability of label agreement between neighboring regions. The trade-off be-

tween maintaining spatial label consistency and ensuring local selection of classes having high likelihood can be naturally modeled by embedding the system in a 2D HMM. The confidence of local classification in each superpixel is quantified by the emission probabilities of the HMM, while label coherence across neighboring superpixels is maintained by its transition probabilities. The parameters implementing are learned in a principled manner from training data. Finally, contiguous groups of superpixels bearing the same label are aggregated to obtain cell segmentation. We note that though we provide experimental results on CMP data, the method can easily be applied to a wide variety of multivariate datasets as the formulation is general and system parameters are automatically learned from data.

3.2.1 Superpixel Extraction

Rather than pixels, we use superpixels as our atomic unit. This offers two major advantages over a sliding window approach: (i) we exploit redundancies in neighboring pixels to achieve a significant reduction in the number of computations (ii) local region statistics are calculated on a meaningful neighborhood rather than a window of fixed size. The major drawback of superpixel approaches is that the final segmentation result relies on the preservation of true boundaries in the initial oversegmentation. We mitigate this issue by using SLIC (Simple Linear Iterative

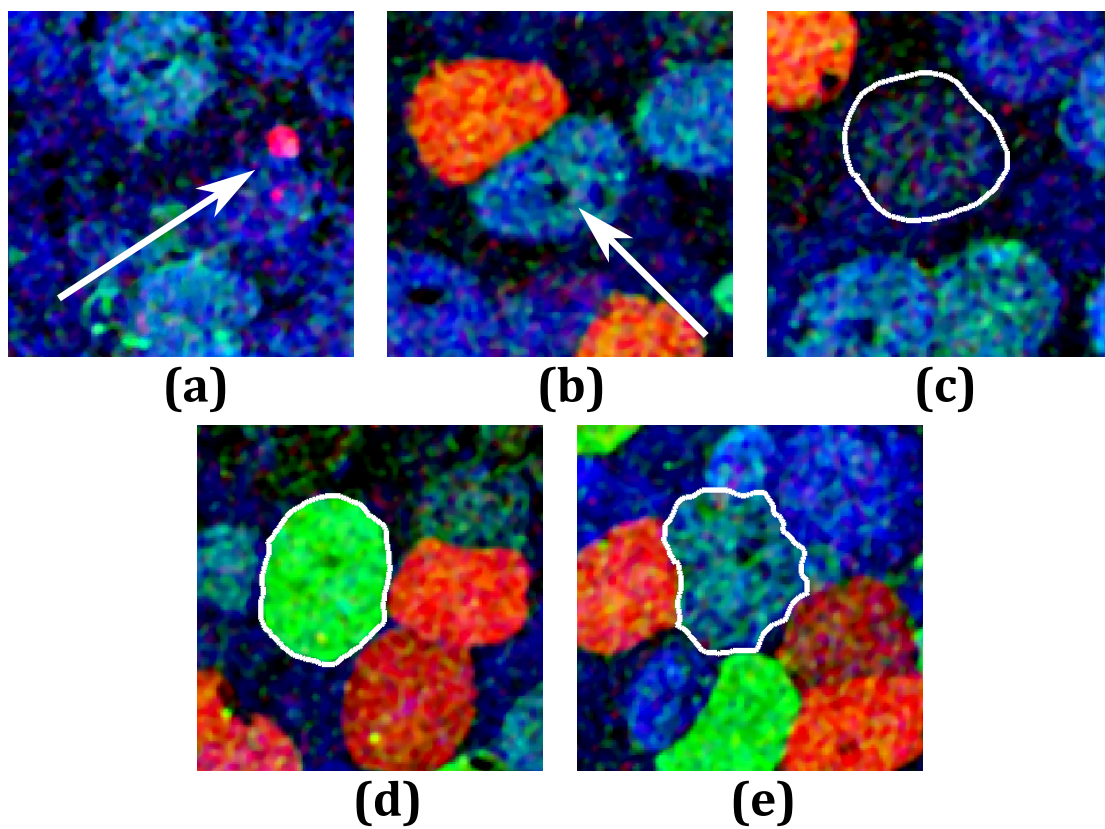


Figure 3.2: Challenges in segmenting CMP images (a) Molecular marker artifact (b) “Hole” inside a cell (c) Cell having low contrast with background (d-e) Change in intra-cell feature variance - low variance in (d), high variance in (e)

Clustering) [29] a state-of-the-art superpixel generation algorithm which has high boundary recall. We use the readily available implementation provided in [30] for convenience. SLIC superpixels are computed on the vector-valued, i.e. multi-channel, input image, and with the spatial regularization set to 1 to ensure a regular lattice structure in the extracted oversegmentation.

3.2.2 Constructing the HMM

We construct a 2D HMM over a first order Markov mesh random field of size $X \times Y$, where X and Y are, respectively, the number of superpixels per row and column of the oversegmentation. Each superpixel \mathcal{S} corresponds to a node at location (x, y) in the 2D HMM, and is denoted by $\mathcal{S}_{x,y}$. Each state q of the HMM corresponds to one of M classes $\{\omega_m, m = 1, 2, \dots, M\}$. Our aim is to find the optimal state sequence, $Q^* = \{q_{x,y}^*, x = 1, 2, \dots, X, y = 1, 2, \dots, Y\}$, and hence, segment the image. Training and testing of the HMM are performed under the turbo-HMM approximation.

3.2.3 Local Class Probabilities

Some classes are tightly packed in feature space (Fig. 3.2 (d)) while other classes have feature vectors that are more spread out (Fig. 3.2 (e)). In order to account for the difference in intra-class variance across classes while maintaining

low complexity, we employ a Bayesian quadratic discriminant analysis (QDA) classifier. The QDA model has the additional benefit of being robust to occasional outliers in training data. Class likelihood functions for each class ω_m at pixel p_k are calculated from its D-dimensional multi-channel feature vector \mathbf{f}_k :

$$p(\mathbf{f}_k|\omega_m) = \frac{\exp\{-\frac{1}{2}(\mathbf{f}_k - \mu_m)^T \Sigma_m^{-1}(\mathbf{f}_k - \mu_m)\}}{(2\pi)^{\frac{D}{2}} |\Sigma_m|^{\frac{1}{2}}} \quad (3.1)$$

The emission probability $b_{x,y}^m$ at each superpixel is obtained by combining contributions from its constituent pixels, by taking the geometric mean of all pixel-level likelihoods. In practice, $b_{x,y}^m$ is computed by taking the exponential of average log-likelihood within the superpixel, to improve numerical stability. In this model, the emission probability $b_{x,y}^m$ represents the probability of superpixel $\mathcal{S}_{x,y}$ being emitted by class ω_m , without taking neighborhood information into consideration.

$$b_{x,y}^m = \left[\prod_{p_k \in \mathcal{S}_{x,y}} p(\mathbf{f}_k|\omega_m) \right]^{\frac{1}{|\mathcal{S}_{x,y}|}} \quad (3.2)$$

3.2.4 Neighborhood Label Consistency

Each element of the transition probability matrix $a_{m',m}$, gives the probability of moving to state m' from state m . The elements of the transition probability matrix remain constant throughout the graph in the standard HMM paradigm.

However, in this work, we allow flexibility in the transition probabilities by defining a spatially varying matrix, conditioned on local feature characteristics at each node. Under this model, the individual transition probabilities can be vary based on similarity or dissimilarity of local regions, encouraging these regions to take on the same or different labels respectively.

$$a_{m',m}^H(x,y) = \begin{cases} C_m^S & m' = m \\ C_m^B & m' = M \\ \frac{1}{\beta_{m',m}} \exp\left(-\frac{R_{x,y}^H}{\beta_{m',m}}\right) & \text{otherwise} \end{cases} \quad (3.3)$$

where $R_{x,y}^H = \frac{1}{1 + K_{x,y}^H}$ and $K_{x,y}^H$ is the symmetric Kullback-Leibler (KL) divergence between the D-dimensional histogram of superpixel $\mathcal{S}_{x,y}$, given by $h_{x,y}$, and that of its horizontally neighboring superpixel, $\mathcal{S}_{x+1,y}$, given by $h_{x+1,y}$.

$$K_{x,y}^H = \sum_i h_{x,y}(i) \log \frac{h_{x,y}(i)}{h_{x+1,y}(i)} + \sum_i h_{x+1,y}(i) \log \frac{h_{x+1,y}(i)}{h_{x,y}(i)} \quad (3.4)$$

C_m^S represents the probability of self transition for each class ω_m . A high value of C^S indicates a higher probability of neighboring pixels taking on the same label. C_m^B represents the probability of transitioning from class ω_m to the background.

This model favors state transitions (i.e., cell label changes) across neighboring superpixels with abrupt changes in their histograms, as quantified by the exponential distribution with parameter $\beta_{m',m}$. We define a unique parameter $\beta_{m',m}$ for each pair of classes $\{\omega_m, \omega'_m\}$ to account for the variation in contrast between different pairs of classes.

We use a model similar to (3.3) for the transition probability matrix in the vertical direction. Since cellular microscopy images typically do not exhibit directionality along the coordinate axes, we make the simplifying (but removable) assumption that the parameters of the transition probability matrix, $\{C_m^S\}$, $\{C_m^B\}$ and $\{\beta_{m',m}\}$, are identical in the vertical and horizontal 1D HMMs.

3.2.5 Parameter Estimation

We employ a supervised learning scheme for the estimation of the the parameters relating to emission probability. The class-specific parameters of the Gaussian likelihood functions, $\{\mu_m\}$ and $\{\Sigma_m\}$, are estimated from labeled data.

The parameters of the transition probability matrix are learned using the Baum-Welch algorithm [5]. Re-estimation formulas for the parameters are derived by maximizing Baum's auxiliary function, given by:

$$\mathcal{Q}(\lambda, \lambda') = \sum_Q P(Q, I, S_{SP} | \lambda) \log P(Q, I, S_{SP} | \lambda') \quad (3.5)$$

with respect to λ' , where λ denotes the current estimate of HMM parameters, λ' the model re-estimate and Q , a sequence of states $Q = \{q_{x,y}\}$. I denotes the input image and S_{SP} , its superpixel oversegmentation.

During the E-step, the modified forward-backward algorithm of the T-HMM is used to estimate the occupancy probabilities (3.6) and ancillary training variables (3.7) in the horizontal and vertical directions.

$$\gamma_{x,y}^{H,\tau} = P(q_{x,y}^H = \tau | I, S_{SP}, \lambda) \quad (3.6)$$

$$\gamma_{x,y}^{V,\tau} = P(q_{x,y}^V = \tau | I, S_{SP}, \lambda)$$

$$\xi_{x,y}^H(m', m) = P(q_{x+1,y} = m', q_{x,y} = m | \lambda) \quad (3.7)$$

$$\xi_{x,y}^V(m', m) = P(q_{x,y+1} = m', q_{x,y} = m | \lambda)$$

During the M-step, $Q(\lambda, \lambda')$ is maximized with respect to each parameter to obtain the following re-estimation formulas:

$$\hat{C}_m^S = \frac{\sum_{x,y} [\xi_{x,y}^H(m, m) + \xi_{x,y}^V(m, m)]}{\sum_{x,y} [\gamma_{x,y}^H(m, m) + \gamma_{x,y}^V(m, m)]} \quad (3.8)$$

$$\hat{C}_m^B = \frac{\sum_{x,y} [\xi_{x,y}^H(M, m) + \xi_{x,y}^V(M, m)]}{\sum_{x,y} [\gamma_{x,y}^H(M, m) + \gamma_{x,y}^V(M, m)]} \quad (3.9)$$

$$\hat{\beta}_{m',m} = \frac{\sum_{x,y} [\xi_{x,y}^H(m', m) R_{x,y}^H + \xi_{x,y}^V(m', m) R_{x,y}^V]}{\sum_{x,y} [\xi_{x,y}^H(m', m) + \xi_{x,y}^V(m', m)]} \quad (3.10)$$

Transition probability matrices are uniformly initialized, and parameters are re-estimated using iterative passes of the Baum-Welch algorithm until the likelihood of the training set converges. The optimal label sequence is inferred using the decoding procedure with modified forward-backward iterations [16].

3.2.6 Obtaining Segmentation

The final segmentation results are obtained by aggregating contiguous super-pixels having the same label. This may lead to cells of the same class being clumped together. In this case, cells are separated using the method described in [31].

3.3 Experimental Results

We compare our method with two well known algorithms - mean shift segmentation [25] and the algorithm proposed by Fulkerson, Vedaldi and Soatto (FVS)

in [26] in terms of the results obtained on CMP data. The optimal parameters for mean shift were found to be $(h_s, h_r, M) = (20, 8, 5000)$. We use 4-fold cross validation to train and test FVS as well as the proposed method. While running FVS, we set $K = 25$ as increasing beyond this value resulted in overfitting the data. SLIC was used to produce superpixels with an average size of 250 pixels for the proposed method. Segmentation results were obtained by aggregating contiguous superpixels having the same label. For fair comparison, we apply the clump separation algorithm to the two competing methods as well.

The accuracy of segmentation is measured by comparing with ground truth. Using magnitude of pixel overlap, each ground truth cell is associated with at most one cell in the segmentation output. F-measure (F) is used as a measure of similarity between each ground truth cell (S_{gt}) and its corresponding segmented cell (S_{seg}).

$$F = \frac{2|S_{seg} \cap S_{gt}|}{|S_{seg}| + |S_{gt}|} \quad (3.11)$$

where $|\cdot|$ denotes number of pixels. The area (in pixels) of each cell in the ground truth is used to weight the corresponding F-measure in calculation of F-measure statistics. The weighted mean and standard deviation of the F-measure across all cells is used to compare the accuracy of segmentation results obtained from the three methods.

Classification accuracy is evaluated at the pixel level. The label of each su-

Method	Mean \pm Std of F-measure	Classification Accuracy	Running Time (s)
Mean shift [25]	0.7424 \pm 0.1793	-	30.51
FVS [26]	0.8093 \pm 0.1515	77.04%	21.30
Proposed	0.8372 \pm 0.1305	85.97%	23.04

Table 3.1: Comparison of results on CMP data from the RC1 connectome. Classification accuracy is not reported for mean shift segmentation as the algorithm does not directly provide classification output.

perpixel is assigned to all the pixels within it. The resulting pixel-level labeling is compared with ground truth labeling. Accuracy is calculated as the percentage of pixels that are correctly classified.

Numerical results comparing the performance of the proposed approach to related methods are provided in Table 3.1, along with average run time for a 1024 \times 1024 pixel region for each method. We observe significant improvement over the competing methods in both segmentation and classification accuracy.

Visual results on some example cells, along with the corresponding F-measure of each segmentation, are presented in Fig. 3.3. The proposed approach demonstrates the ability to handle challenging scenarios such as the presence of “holes” within cells (see Fig. 3.3 (a)), having large intra-cell variance (see Fig. 3.3 (b)) and cells having low contrast with background and surrounding cells (see Fig. 3.3 (e)), whereas competing methods are unable to accurately capture cell boundaries in these cases.

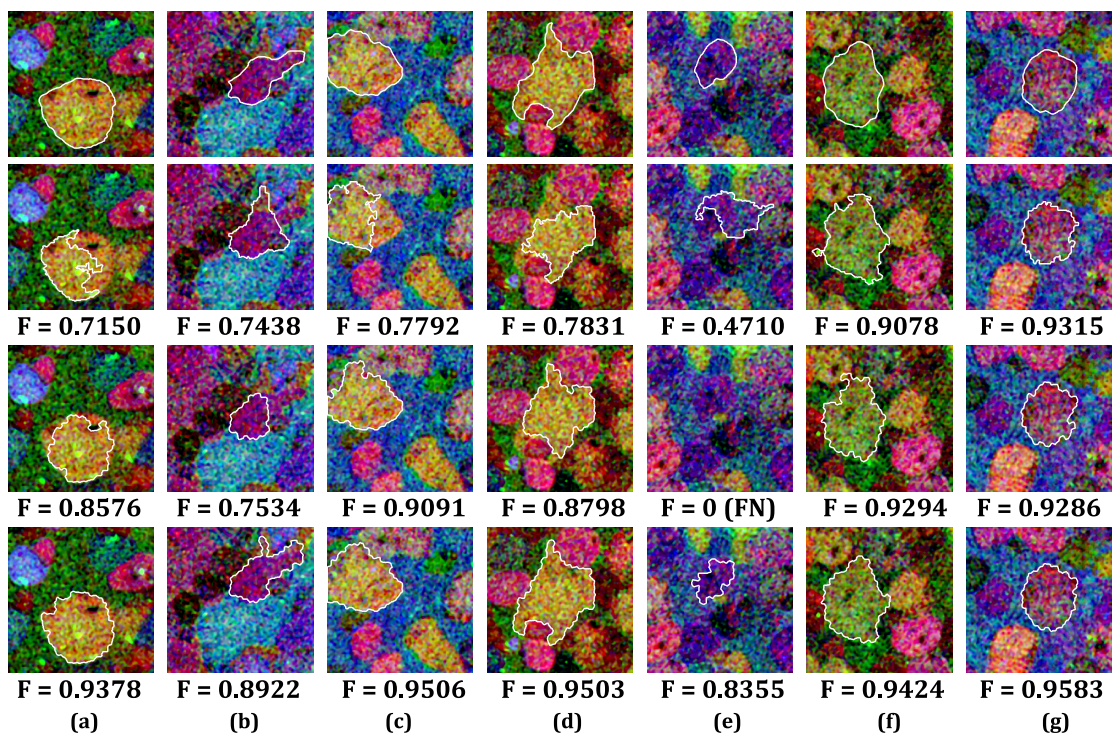


Figure 3.3: Examples of segmentation results showing challenging as well as easy cases. Each column (a) - (g) shows results on a specific cell. The first row shows the ground truth segmentation, with the boundary outlined in white. The second, third and fourth rows show results obtained from mean shift segmentation, FVS and the proposed method respectively, along with the F-measure F of the resulting segmentation. (An F-measure of 0 indicates a missed detection.) To visualize each result, we choose the three channels exhibiting the highest contrast between the given cell and background.

3.4 Conclusion

In this chapter, we describe a novel probabilistic algorithm for simultaneous segmentation and classification of cells in multi-marker images. Costs associated with neighborhood label coherence and local class membership probabilities are embedded in a 2D HMM framework. The T-HMM approximation is used to learn parameters of the HMM and to infer the optimal solution. We provide experimental validation on cellular microscopy data. The proposed method overcomes some of the main pitfalls of segmentation of such challenging data. As a result, we observe significant gains over competing methods in terms of both segmentation performance and classification accuracy.

Chapter 4

Deformable Registration of Multimodal Images

Robust registration of unimodal and multimodal images is a key task in biomedical image analysis, and is often utilized as an initial step on which subsequent analysis techniques critically depend. In this chapter, we first present an overview of related work in biomedical image registration, ranging from early work to recent techniques. The next section describes a method for deformable registration with a generic theoretical formulation, where the underlying framework is applicable to a broad spectrum of problems in the domain. The method is first derived for unimodal data and then extended to the multimodal, multi-channel setting. We also describe an edge-adaptive constraint which allows for variation in degree of

smoothness across the image. We provide experimental evaluation of the proposed algorithm, for unimodal registration of MRI data as well as multimodal registration of connectome data, in comparison with state-of-the-art deformable registration techniques in literature. Finally, we provide concluding remarks.

4.1 Introduction

Image registration, which aims at accurately aligning structures or regions across related images, is an important problem in biomedical image computing, and is an active area of current research. Image registration is used in a variety of applications in biomedical image analysis. Assessing the efficacy of treatments often requires registration in order to accurately compare pre-treatment and post-treatment scans. Atlas-based methods, which analyze subject data in comparison to one or more standard models, critically depend on accurate registration. Often, different types of information, eg., structural and functional, can be extracted from different image modalities, and registering these multimodal images is an important step in combining the complementary sources of information. Alignment of structures across temporal or depth-based volumes are important in time-lapse and 3D network reconstruction. Hence, accurate automated deformable registration is the need of the hour.

4.2 Background and Related Work

Registration is a classic problem in biomedical image analysis and has been widely studied over the past three decades. Early research in unimodal registration focused on rigid alignment [32, 33]. Due to the nature of the unimodal registration problem, the sum of squared differences (SSD) [34, 35, 36] and the mean squared difference (MSD) [37, 38] of image intensities have been widely and successfully used to measure data similarity.

Early work on multimodal registration used intensity levels [39] and joint entropy [40] as the measure of information across modalities. Mutual information (MI), which measures the statistical dependence of two random variables, was proposed in [41, 42] and [43] for rigid alignment problems and quickly gained popularity as a measure of similarity for both rigid and deformable registration, and in both unimodal and multimodal settings. (A survey of various MI-based methodologies for medical image registration is available in [44].) In [45], a deformable registration technique was introduced which uses both a global affine transformation and local transformations, with MI as a measure of information. The local transformations are modeled with Free Form Deformations (FFD) on a non-rigid lattice of control points using cubic B-splines for interpolation. Several subsequent papers leveraged FFDs, extending their use to multimodal data [46], [47] and using variants of MI (such as Conditional MI (CMI) in [48] and Normalized

MI (NMI) in [49]). In [50] and subsequently [51], dense deformable registration is performed by modeling it as a minimal cost graph problem on a Markov random field (MRF) built upon the FFD framework. Smoothness constraints are imposed through connectivity of nodes and labels correspond to deformations. For the multimodal case, MI (or a variant of MI) is used as the matching criterion.

A significant drawback of these techniques lies in the estimation of the joint histogram between modalities where one or both of the modalities proffers multiple channels of information. A natural extension of MI-based approaches to include data with multiple channels (e.g., RGB data) would be to use multivariate MI. However, the complexity of populating higher dimensional joint histograms grows exponentially with the number of channels, and these methods quickly become impractical for multichannel modalities. Further, inadequate population of such high dimensional histograms due to sparse availability of data could lead to inaccuracies in the inferred deformation.

Some notable prior research has been devoted to the problems associated with estimation of multivariate MI. A simplifying approximation of the general multivariate MI was proposed as a similarity measure in [52]. In [53], an entropic graph-based implementation was used to estimate α -MI of multiple channels.

The concept of self-similarity was introduced as a means for multimodal registration in two recent papers, [54] and [55], which introduce the modality inde-

pendent neighborhood descriptor (MIND) and the self-similarity context (SSC) descriptor respectively. These descriptors exploit local structural similarities between multimodal image pairs and are calculated on a defined spatial search region in each modality. SSD is used as a measure of distance between descriptors, and final deformation is estimated using Gauss-Newton optimization. These methods, however, rely on significant anatomical similarity between images of the two modalities, which may often be absent, as in the case of connectome [2] data. In addition, they require modification to be used with multi-channel inputs.

In this chapter, we propose a novel probabilistic framework, based on the 2D hidden Markov model (2D HMM), to capture the deformation between pairs of images. The HMM is tailored to capture spatial transformations across images via state transitions, and modality-specific data costs via emission probabilities. The method is derived for the unimodal setting (where simpler matching metrics may be used) as well as the multimodal setting, where different modalities may provide very different representation for a given class of objects, necessitating the use of advanced similarity measures. We also introduce a local edge-adaptive constraint to allow for varying degrees of smoothness between object boundaries and homogeneous regions. The parameters of the described method are estimated in a principled manner from training data via maximum likelihood learning, and the deformation is subsequently estimated using an efficient dynamic programming

algorithm.

4.3 Proposed Approach

Our aim is to find the deformation that best explains the relation between one image (the “source”) and a second image, from the same or different modality (the “target”). We propose a probabilistic method that estimates the global deformation with a set of local deformations. There is a clear trade-off between flexibility in local deformations so that high accuracy is achieved in discerning the true structural relationship between the source and the target, and the need to impose global coherence and avoid highly “non-smooth” deformations. Moreover, every local transformation results in a mapping between the two modalities and there must be a way to measure and maximize the goodness of this match within the constraints of the deformation framework.

Translation consistency in neighborhoods and cross-modality matching costs are embedded into a 2D HMM built on a first order Markov mesh random field (MMRF). Local translations are “hidden” and their impact is felt through the corresponding data matching costs. The parameters of these data matching costs are data dependent, and are learned from source-target pairs of training images. In addition, the degree of smoothness of the optimal deformation can differ considerably across different types of data, and hence the training images are also

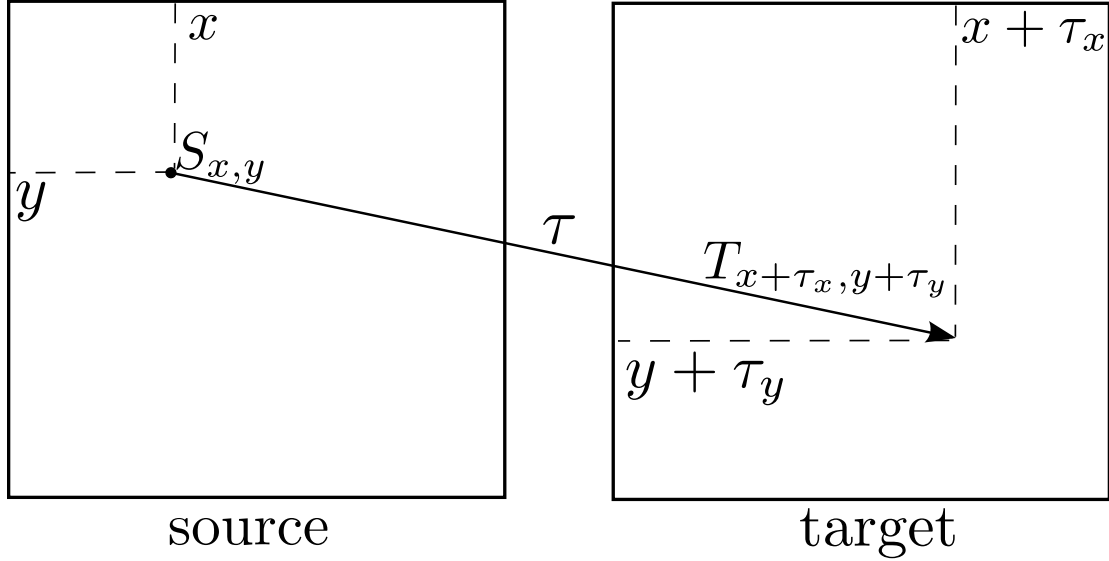


Figure 4.1: Mapping feature vector $\mathbf{S}_{x,y}$ at point (x, y) in the source image to a feature vector $\mathbf{T}_{x+\tau_x, y+\tau_y}$ at point $(x + \tau_x, y + \tau_y)$ in the target image using a translation τ .

utilized in learning smoothness parameters.

4.3.1 Data Similarity Measure

The probability of matching a feature vector in the source image to a feature vector in the target image is captured by the emission probabilities of the HMM.

In our design, each state, q , of the HMM corresponds to a specific, unique translation τ relating the source and target, and whose components are τ_x and τ_y in the x- and y-directions, respectively. A state with translation τ maps a point (x, y) in the source to $(x + \tau_x, y + \tau_y)$ in the target. Therefore, the emission probability $b_{x,y}^\tau$ represents the probability of matching the source feature vector

at (x, y) , given by $\mathbf{S}_{x,y}$, to the target feature vector at the given translation, given by $\mathbf{T}_{x',y'}$. (see Fig. 4.1.)

$$b_{x,y}^\tau = P(\mathbf{T}_{x',y'} | \mathbf{S}_{x,y}) \quad (4.1)$$

This probability can be modeled in various ways in single modality problems, for example, using a simple similarity metric such as sum of absolute differences in intensity or correlation on patches from source and target images. In our experiments, we use an term based on SSD of intensities to model emission probability.

$$b_{x,y}^\tau = \frac{1}{\sqrt{2\pi}\sigma_{SSD}} \exp \left\{ -\frac{\epsilon_{x,y}^\tau}{2\sigma_{SSD}^2} \right\} \quad (4.2)$$

where $\epsilon_{x,y}^\tau$ is the sum of squared differences in sliding windows of size $(2W + 1) \times (2W + 1)$ centered at the points of interest in the source and target images.

$$\epsilon_{x,y}^\tau = \sum_{x_w, y_w = -W}^W (\mathbf{s}_{x+x_w, y+y_w} - \mathbf{T}_{x'+x_w, y'+y_w})^2 \quad (4.3)$$

4.3.2 Extending to Multimodal Data

Since different modalities may vary greatly in the way they represent objects from the same class, we cannot directly apply intensity-based similarity measures for registration of multimodal data. Instead, we assume that the feature vectors

of the source and target images at specific locations are not directly related, but rather, related only through the object type at the corresponding locations in the underlying “true” arrangement. In other words, the source feature vector, the underlying object type at the corresponding location in the source ($\omega_{x,y}^S$), the underlying object type at the location after translation in the target ($\omega_{x',y'}^T$) and the target feature vector form a Markov chain.

$$\mathbf{S}_{x,y} \longleftrightarrow \omega_{x,y}^S \longleftrightarrow \omega_{x',y'}^T \longleftrightarrow \mathbf{T}_{x',y'} \quad (4.4)$$

In our model, the underlying object types and their spatial relations are hidden, information from them can only be extracted from the observable features in the images from each modality.

We learn the distribution of the source feature vectors, and rather than making a hard decision on object type at each node, we obtain its posterior probability. Thus, for an M class problem, we learn $P(\omega_m | \mathbf{S}_{x,y})$ for each class $m \in \{1, 2, \dots, M\}$ at every node. Applying the law of total probability to (4.1) under the Markov assumption (4.4) results in:

$$b_{x,y}^\tau = \sum_{m=1}^M P(\omega_m | \mathbf{S}_{x,y}) P(\mathbf{T}_{x',y'} | \omega_m) \quad (4.5)$$

Since linear combinations of Gaussians can approximate arbitrarily shaped densities, we use a mixture of K Gaussians to model $P(\mathbf{T}_{x',y'}|\omega_m)$ for each object class ω_m .

$$P(\mathbf{T}_{x',y'}|\omega_m) = \sum_{k=1}^K w_m^k P(\mathbf{T}_{x',y'}|\omega_m^k) \quad (4.6)$$

where mixture component weights must satisfy the constraint:

$$\sum_{k=1}^K w_m^k = 1 \quad \forall m \in \{1, 2, \dots, M\} \quad (4.7)$$

Each individual component density is a Gaussian having dimensionality D , equal to that of the target feature space.

$$P(\mathbf{T}_{x',y'}|\omega_m^k) = \frac{\exp\{-\frac{1}{2}(\mathbf{T}_{x',y'} - \mu_m^k)^T \boldsymbol{\Sigma}_m^k^{-1} (\mathbf{T}_{x',y'} - \mu_m^k)\}}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}_m^k|^{\frac{1}{2}}} \quad (4.8)$$

where μ_m^k and $\boldsymbol{\Sigma}_m^k$ are the mean and covariance of the Gaussian respectively.

4.3.3 Deformation Smoothness Model

Each state of the HMM corresponds to a translation τ relating the source and target. The translations of neighboring nodes are correlated as quantified by the transition probabilities of the HMM (see Fig. 4.2). Equivalently, in an

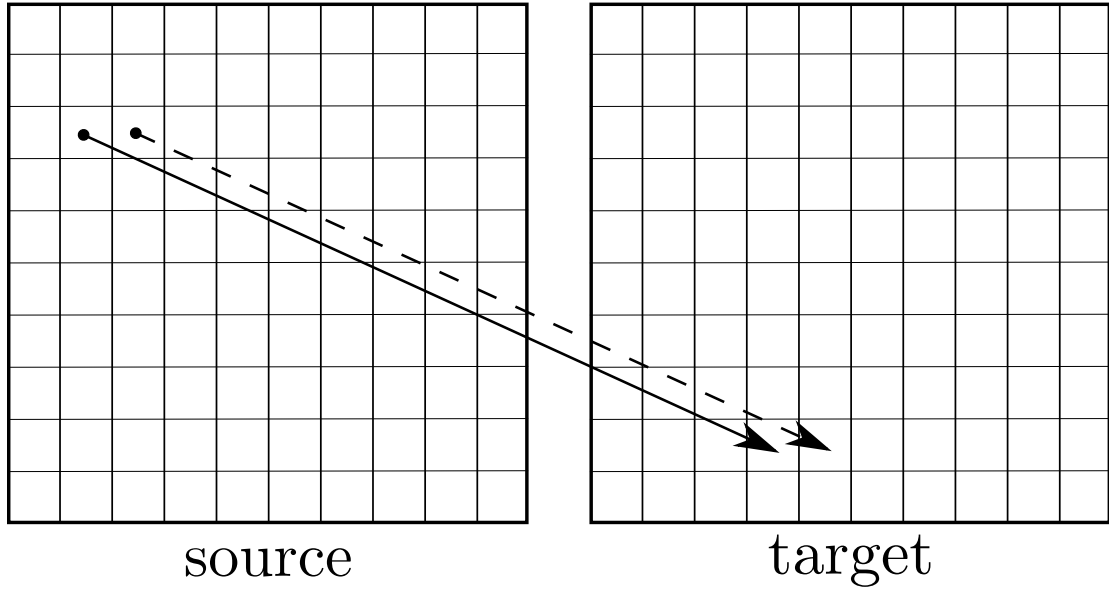


Figure 4.2: Neighboring translations are correlated.

HMM constructed over a first order Markov mesh random field, the state of a node depends on the state of its adjacent neighbors in the horizontal and vertical directions. While an arbitrary transition probability matrix may be used, we introduce some assumptions in order to reduce the number of free parameters in the system. These are outlined below.

Assuming a stationary HMM, the transition probabilities of the horizontal and vertical 1D HMMs are given by

$$a^H(\tau, \tau') = P(q_{x,y} = \tau' | q_{x,y-1} = \tau)$$

and

$$a^V(\tau, \tau') = P(q_{x,y} = \tau' | q_{x-1,y} = \tau)$$

respectively. In the case of cellular microscopy images, we make the simplifying assumption that parameters of horizontal and vertical HMMs are identical as these images typically do not exhibit directionality along coordinate axes (as is often the case in faces [18], man-made scenes, and certain natural images):

$$a^V(\tau, \tau') = a^H(\tau, \tau') = a(\tau, \tau') \quad (4.9)$$

We also impose shift invariance so that probability of moving from one state to another only depends on the difference in the corresponding translations:

$$a(\tau, \tau') = a(\tau' - \tau) = a(\delta\tau) \quad (4.10)$$

where $\delta\tau = [\delta\tau_x \ \delta\tau_y]^T$ and $\delta\tau_x$ and $\delta\tau_y$ are, respectively, the horizontal and vertical components of the difference in translations.

Further, we restrict ourselves to parametric transition probabilities to increase robustness. For a Gaussian model with a covariance matrix Σ , the general expression is

$$a(\delta\tau) \propto \exp \left\{ -\frac{1}{2}(\delta\tau)^T \Sigma^{-1}(\delta\tau) \right\} \quad (4.11)$$

Noting again the lack of consistent directionality in microscopy data, we simplify to an isotropic model with a single variance parameter σ^2 .

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}$$

Incorporating the isotropic model into (4.11) results in a simplified expression for transition probability, given below.

$$a(\delta\tau) \propto \exp \left\{ -\frac{1}{2} \left(\frac{\delta\tau_x^2 + \delta\tau_y^2}{\sigma^2} \right) \right\} \quad (4.12)$$

4.3.4 Edge-adaptive Smoothness Constraint

For data consisting of multiple objects, large translations occur more frequently near object boundaries than well inside object. Similar behavior is noticed while computing optical flow in images with multiple objects, and has been addressed by including an structure-adaptive regularization constraint in the cost function [56, 57]. We model this variation in smoothness by introducing a spatially varying transition matrix, parametrized by two Gaussians. The transition probability matrix at each point is calculated using local “edgeness” at that point.

To calculate the edgeness at a given point (x, y) , we consider the set of all estimated object labels in a window centered at that point. By normalizing the

histogram of all unique labels l in the window, we obtain $[P(l)]$, the local label probability vector. The label entropy around the point (x, y) is given by

$$\mathcal{H}_{x,y} = - \sum_l P(l) \log_2 P(l) \quad (4.13)$$

The label entropy in each window measures label uncertainty in the window and is used to estimate edgeness. We thus quantify edgeness \mathcal{E} at (x, y) as

$$\mathcal{E}_{x,y} = 1 - e^{-\mathcal{H}_{x,y}} \quad (4.14)$$

Note that this measure of edgeness approaches zero at low entropy, and approaches one at high entropy. The transition matrix at each point is modeled as a linear combination of two Gaussian-parametrized matrices.

$$a_{x,y}(\delta\tau) = \mathcal{E}_{x,y} a^E(\delta\tau) + (1 - \mathcal{E}_{x,y}) a^I(\delta\tau) \quad (4.15)$$

where $a^E(\delta\tau)$ denotes the transition matrix for points on cell edges and $a^I(\delta\tau)$, that of interior points.

$$a^E(\delta\tau) \propto \exp \left\{ -\frac{1}{2} \left(\frac{\delta\tau_x^2 + \delta\tau_y^2}{\sigma_E^2} \right) \right\} \quad (4.16)$$

$$a^I(\delta\tau) \propto \exp \left\{ -\frac{1}{2} \left(\frac{\delta\tau_x^2 + \delta\tau_y^2}{\sigma_I^2} \right) \right\} \quad (4.17)$$

Hence, we allow for flexibility in the transition probability matrix according to the edgeness of the point in question.

4.3.5 Estimation of Deformation Field

We infer the optimal state sequence using the Viterbi algorithm with the modified forward-backward iterations described in [16].

4.3.6 Parameter Estimation

Baum-Welch Training

The parameters of the HMM are estimated in an unsupervised manner from source-target image pairs. In the Baum-Welch algorithm ([5, 13]). Re-estimation formulas for these parameters are derived by maximizing Baum's auxiliary function, given by

$$\mathcal{Q}(\lambda, \lambda') = \sum_Q P(Q, S, T | \lambda) \log P(Q, S, T | \lambda') \quad (4.18)$$

with respect to λ' , where λ denotes the current estimate of HMM parameters, λ' the model re-estimate and Q , a sequence of states $Q = \{q_{x,y}, x = 1, 2, \dots, X, y =$

$1, 2, \dots, Y\}$. S and T denote corresponding source and target images.

During the expectation step, we estimate the occupancy probabilities of the horizontal and vertical 1D HMMs,

$$\gamma_{x,y}^{H,\tau} = P(q_{x,y}^H = \tau | S, T, \lambda)$$

$$\gamma_{x,y}^{V,\tau} = P(q_{x,y}^V = \tau | S, T, \lambda)$$

and the overall occupancy probability

$$\gamma_{x,y}^{\tau} = \frac{\gamma_{x,y}^{H,\tau} + \gamma_{x,y}^{V,\tau}}{2}$$

We also estimate the ancillary training variables,

$$\xi_{x,y}^H(\tau, \tau + \delta\tau) = P(q_{x,y+1} = \tau + \delta\tau, q_{x,y} = \tau | S, T, \lambda)$$

$$\xi_{x,y}^V(\tau, \tau + \delta\tau) = P(q_{x+1,y} = \tau + \delta\tau, q_{x,y} = \tau | S, T, \lambda)$$

During the maximization-step, we maximize Baum's auxiliary function with respect to each parameter to derive re-estimation formulas.

For the SSD-based metric, there is only one emission parameter, σ_{SSD} , whose re-estimation equation is given by

$$\hat{\sigma}_{SSD}^2 = \frac{\sum_{x,y,\tau} \gamma_{x,y}^{\tau} \epsilon_{x,y}^{\tau}}{\sum_{x,y,\tau} \gamma_{x,y}^{\tau}} \quad (4.19)$$

To update emission parameters for the multimodal case, we must first use the updated variables from the expectation step to calculate the per-component posterior probability at each node, given by

$$\phi_{x,y,m}^{\tau,k} = \frac{P(\omega_m | \mathbf{S}_{x,y}) w_m^k P(\mathbf{T}_{x',y'} | \omega_m^k)}{\sum_{m=1}^M P(\omega_m | \mathbf{S}_{x,y}) \sum_{k=1}^K w_m^k P(\mathbf{T}_{x',y'} | \omega_m^k)} \quad (4.20)$$

where $x' = x + \tau_x$ and $y' = y + \tau_y$.

Emission parameters are re-estimated using the following update equations:

$$\hat{w}_m^k = \frac{\sum_{x,y,\tau} \gamma_{x,y}^{\tau} \phi_{x,y,m}^{\tau,k}}{\sum_{x,y,\tau,k} \gamma_{x,y}^{\tau} \phi_{x,y,m}^{\tau,k}} \quad (4.21)$$

$$\hat{\mu}_m^k = \frac{\sum_{x,y,\tau} \gamma_{x,y}^{\tau} \phi_{x,y,m}^{\tau,k} \mathbf{T}_{x',y'}}{\sum_{x,y,\tau} \gamma_{x,y}^{\tau} \phi_{x,y,m}^{\tau,k}} \quad (4.22)$$

$$\hat{\Sigma}_m^k = \frac{\sum_{x,y,\tau} \gamma_{x,y}^{\tau,k} \phi_{x,y,m}^{\tau,k} (\mathbf{T}_{x',y'} - \hat{\mu}_m^k)(\mathbf{T}_{x',y'} - \hat{\mu}_m^k)^T}{\sum_{x,y,\tau} \gamma_{x,y}^{\tau,k} \phi_{x,y,m}^{\tau,k}} \quad (4.23)$$

For transition probability matrices parameterized by a single Gaussian, the update equation is given by:

$$\hat{\sigma}^2 = \frac{\sum_{x,y,\tau,\delta\tau} [\xi_{x,y}^H(\tau, \tau') + \xi_{x,y}^V(\tau, \tau')] [\delta\tau]^2}{\sum_{x,y,\tau,\delta\tau} [\xi_{x,y}^H(\tau, \tau') + \xi_{x,y}^V(\tau, \tau')]} \quad (4.24)$$

For edge-adaptive transition probabilities, the update equations are given by:

$$\hat{\sigma}_E^2 = \frac{\sum_{x,y,\tau,\delta\tau} [\xi_{x,y}^H(\tau, \tau') + \xi_{x,y}^V(\tau, \tau')] \left(\frac{\mathcal{E}_{x,y}}{a_{x,y}(\delta\tau)} \right) [\delta\tau]^2}{\sum_{x,y,\tau,\delta\tau} [\xi_{x,y}^H(\tau, \tau') + \xi_{x,y}^V(\tau, \tau')] \left(\frac{\mathcal{E}_{x,y}}{a_{x,y}(\delta\tau)} \right)} \quad (4.25)$$

$$\hat{\sigma}_I^2 = \frac{\sum_{x,y,\tau,\delta\tau} [\xi_{x,y}^H(\tau, \tau') + \xi_{x,y}^V(\tau, \tau')] \left(\frac{1 - \mathcal{E}_{x,y}}{a_{x,y}(\delta\tau)} \right) [\delta\tau]^2}{\sum_{x,y,\tau,\delta\tau} [\xi_{x,y}^H(\tau, \tau') + \xi_{x,y}^V(\tau, \tau')] \left(\frac{1 - \mathcal{E}_{x,y}}{a_{x,y}(\delta\tau)} \right)} \quad (4.26)$$

Transition matrices are initialized uniformly. Parameters relating to emission probabilities may be initialized either by learning the data from the target feature vectors independently of the source image, or by performing rigid matching between the source and the target to estimate the parameters. While training multiple Gaussians per cell class, we perform standard Expectation-Maximization for Gaussian mixture models (EM-GMM) on each class to learn parameters of each desired component, and use these for initialization. $P(\omega_m | \mathbf{S}_{x,y})$ is learned by applying EM-GMM on the source modality.

4.3.7 Complexity

The complexity of training and decoding under the T-HMM approximation is $O(N^2XY)$, where N is the number of states of the HMM and XY is the number of nodes in the HMM. N , in turn, is proportional to Δ^2 , where Δ is the maximum translation allowed per direction. The complexity of algorithm is hence $O(\Delta^4XY)$.

In order to reduce complexity, we employ a multi-resolution coarse-to-fine scheme, approximating a group of 4 nodes at each resolution with a single node at the nearest coarser resolution. For The complexity at each level is $O(n^2X_LY_L)$, where L is the index of each level, taking values from 1 (the finest level) to L_{max} (the coarsest level), n is the number of states at each level of resolution (a constant), and X_LY_L is the number of nodes at the given level. Therefore,

$$X_LY_L = \frac{XY}{4^{(L-1)}}.$$

The total complexity considering all levels is $O(n^2XY([\sum_{L=1}^{L_{max}} 4^{-(L-1)}]))$. Since $[\sum_{L=1}^{L_{max}} 4^{-(L-1)}] \approx 1.33$, the complexity of the hierarchical approach is $O(n^2XY)$, independent of Δ .

4.4 Experimental Results

We present the performance of the proposed approach on two biomedical image datasets. To evaluate registration quality, we generated automated segmentations by warping the source image segmentations using the transformation obtained from each method. The resulting automated segmentation (S_{trans}) was compared to the manual target segmentation (T) using the Dice similarity measure (DSC) [58] as a measure of overlap.

$$\text{DSC} = \frac{2|S_{trans} \cap T|}{|S_{trans}| + |T|} \quad (4.27)$$

where $|\bullet|$ denotes cardinality in terms of number of pixels.

To check the statistical significance of improvement in results, we obtained p -values by performing two-sided Wilcoxon tests [59] on DSC values obtained using the proposed approach paired with DSC values from each competing method. A value of $p < 0.05$ was considered to indicate statistical significance.

4.4.1 Multi-subject Brain Data

The MR brain data sets were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at the Neuroimaging Informatics Tools and Resources Clearinghouse (<https://www.nitrc.org/projects/ibsr/>).

The dataset consists of MR brain images of resolution $256 \times 256 \times 128$, along with manual expert segmentations of white and grey matter, for 16 subjects. For each subject, the T1-weighted volumetric images have been positionally normalized into the Talairach orientation (rotation only). We randomly select one subject as the target and register images from each of the other 15 subjects to the selected target image.

We compare the performance of the proposed method with that of two self-similarity based registration approaches - MIND [54] and SSC [55] - as well as dense (iconic) registration based on discrete optimization, DROP [51]. For DROP, SSD was used as a measure of similarity since it resulted in the highest DSC values, and the weighting factor λ was empirically found to be 0.01. The optimal regularization term α for MIND and SSC was found to be 0.1. We used three levels of resolution for all methods to compare performance.

The quantitative results on the IBSR dataset are presented in Table 4.1. In addition, visual results are shown in Fig. 4.3 and Fig. 4.4. We observe that the proposed approach shows statistically significant improvement over related approaches, for both grey matter and white matter.

Method	Mean	Median	Std Dev	p -value
Grey Matter				
MIND [54]	0.7469	0.7358	0.0276	$7 \cdot 10^{-3}$
SSC [55]	0.7513	0.7493	0.0263	$3 \cdot 10^{-2}$
DROP [51]	0.7192	0.7350	0.0808	$1 \cdot 10^{-2}$
Proposed Method	0.7756	0.7788	0.0244	—
White Matter				
MIND [54]	0.7213	0.7233	0.0285	$3 \cdot 10^{-4}$
SSC [55]	0.7203	0.7240	0.0265	$1 \cdot 10^{-4}$
DROP [51]	0.6589	0.6727	0.0773	$8 \cdot 10^{-5}$
Proposed Method	0.7612	0.7645	0.0194	—

Table 4.1: Performance comparison of single-channel registration methods multi-subject brain MRI data, measured by DSC of grey and white matter between subjects after warping.

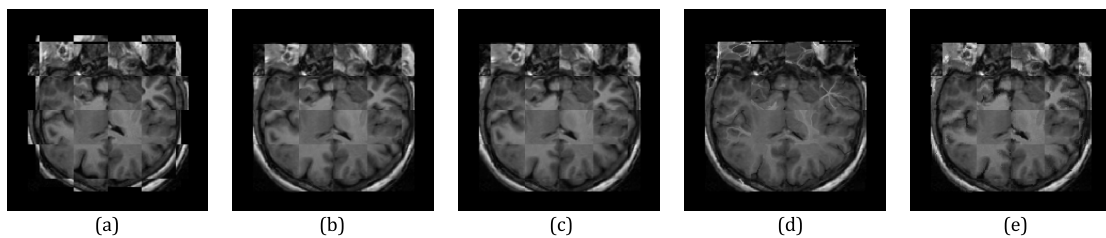


Figure 4.3: Visual results on multi-subject MRI data. Results are shown as a checkerboard, where neighboring tiles come from different subjects. (a) Source-target pair before registration (b-e) After registration using (b) MIND (c) SSC (d) DROP and (e) Proposed method

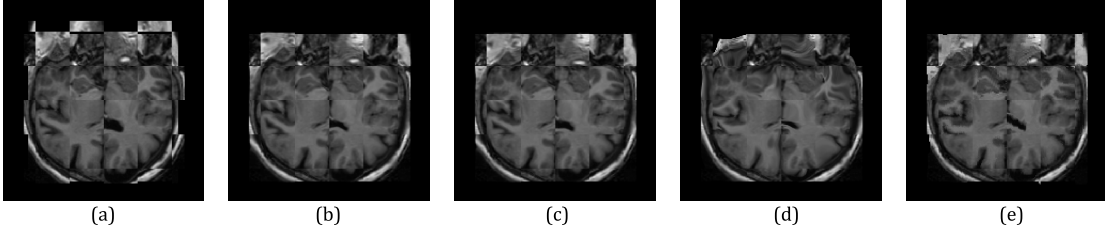


Figure 4.4: Visual results on multi-subject MRI data. Results are shown as a checkerboard, where neighboring tiles come from different subjects. (a) Source–target pair before registration (b–e) After registration using (b) MIND (c) SSC (d) DROP and (e) Proposed method

4.4.2 Rabbit Retinal Connectome Data

We focus our attention on the capstone region of the RC1 dataset. We compare our method with CAMIR [52] embedded in the FFD [45] framework, the α -MI-based approach outlined in [53], MIND [54] and SSC [55]. The optimal value of parameters for the α -MI approach were empirically found to be $\alpha = 0.99$ and $k = 7$ respectively. The feature vector for the target image in the proposed approach is the average pixel intensity in a 5×5 neighborhood. The hyper-parameters used for this dataset are $M = 7$ and $K = 2$. Three levels of resolution were used for all approaches.

Quantitative results on the RC1 connectome dataset are presented in Table 4.4.2. Visual results are shown in Figs. 4.5 to 4.8. We see the proposed method performs well in both easy as well as challenging scenarios. The most improvement over competing methods is observed when there is a large deformation between modalities. One explanation for this is that our approach can account

Method	Mean	Std Dev	Median	p -value
MIND [54]	0.7661	0.1547	0.7896	$3.9 \cdot 10^{-3}$
SSC [55]	0.7682	0.1536	0.7927	$6.4 \cdot 10^{-3}$
α -MI [53]	0.7708	0.1436	0.7863	$1.1 \cdot 10^{-3}$
CAMIR [52]	0.7810	0.1495	0.7924	$5.2 \cdot 10^{-3}$
Proposed Method	0.8185	0.1338	0.8345	—

Table 4.2: Performance comparison of multi-channel registration methods on connectome data, measured by DSC of between ATEM images and warped CMP images.

for large deformations during the training phase, by iteratively optimizing emission and transition parameters. Relatively easy scenarios are shown in Figs. 4.5 and 4.6. Examples of cells with large changes across modalities can be seen in Figs. 4.7 and 4.8. We observe that the proposed approach shows both quantitative and qualitative improvement of registration accuracy in comparison to competing approaches.

4.5 Conclusion

We have presented a novel approach for registration of unimodal as well as multimodal image data, with the deformation system embedded in the probabilistic framework of a 2D HMM and solved using the T-HMM approximation. The formulation is general and different types of transformation models may be used. For registration of images with multiple objects, we allow flexibility in the smoothness

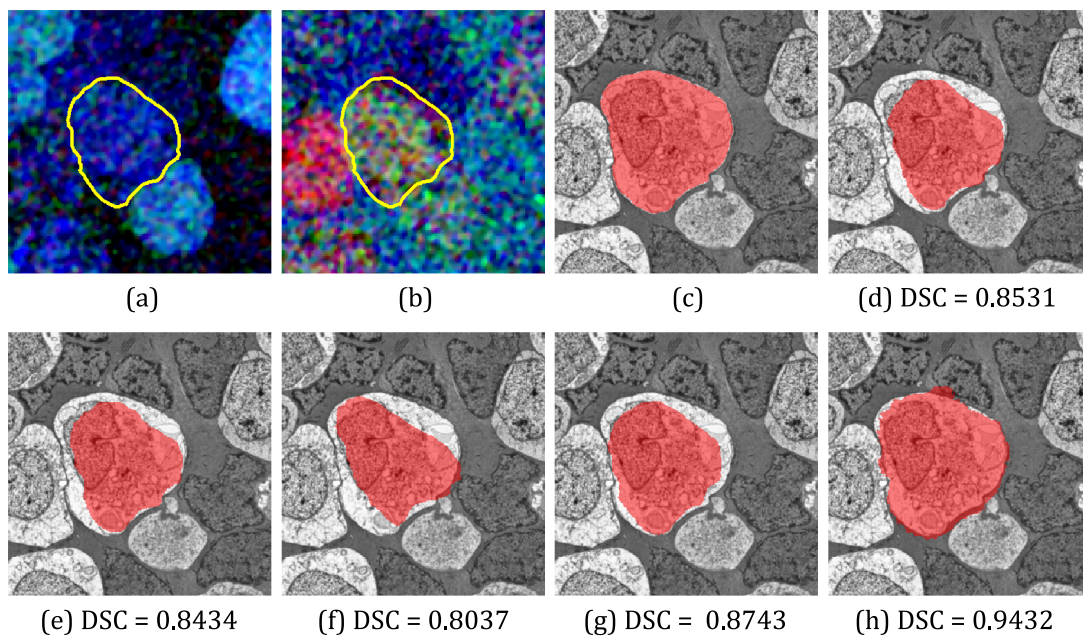


Figure 4.5: Visual results on a relatively easy cell. (a) RGB visualization of 3 CMP channels, with ground truth of the cell of interest outlined in yellow (b) RGB visualization of the remaining 3 CMP channels, with outlined ground truth (c) ATEM image corresponding to the same region, with ground truth overlaid in red (d-h) ATEM image with results from various registration methods overlaid in red, along with the corresponding DSC. (d) MIND (e) SSC (f) α -MI (g) CAMIR (h) Proposed method

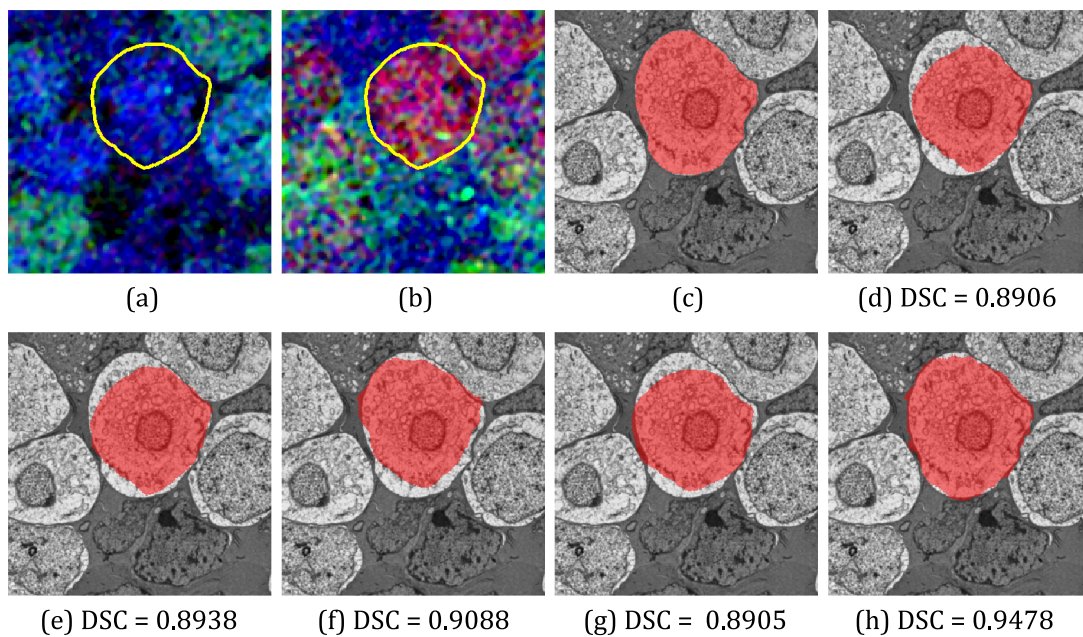


Figure 4.6: Visual results on a relatively easy cell. (a) RGB visualization of 3 CMP channels, with ground truth of the cell of interest outlined in yellow (b) RGB visualization of the remaining 3 CMP channels, with outlined ground truth (c) ATEM image corresponding to the same region, with ground truth overlaid in red (d-h) ATEM image with results from various registration methods overlaid in red, along with the corresponding DSC. (d) MIND (e) SSC (f) α -MI (g) CAMIR (h) Proposed method

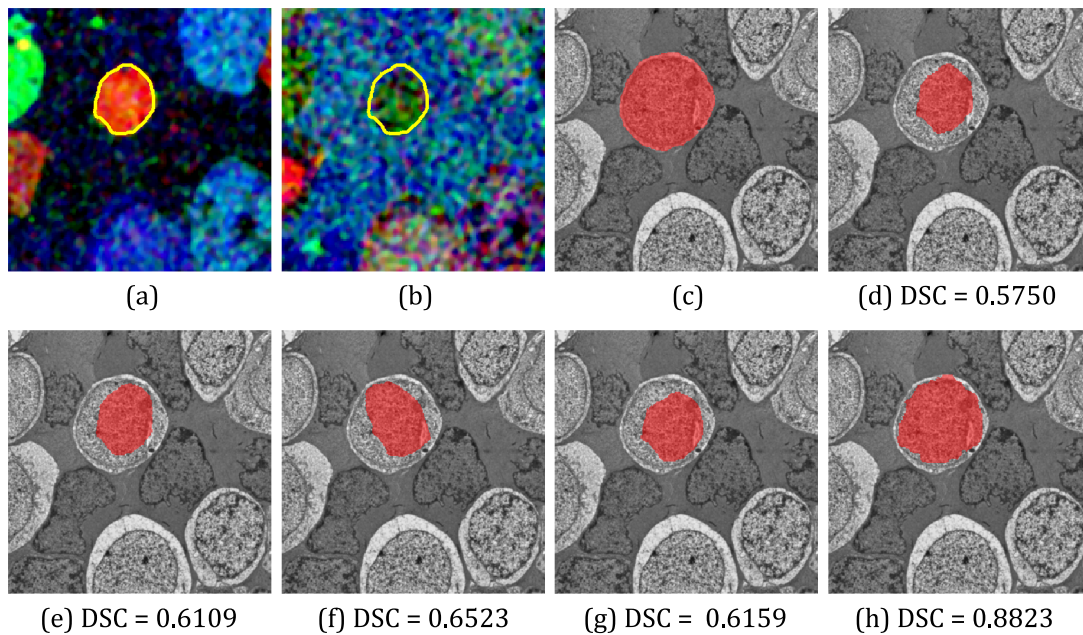


Figure 4.7: Visual results on a challenging cell. (a) RGB visualization of 3 CMP channels, with ground truth of the cell of interest outlined in yellow (b) RGB visualization of the remaining 3 CMP channels, with outlined ground truth (c) ATEM image corresponding to the same region, with ground truth overlaid in red (d-h) ATEM image with results from various registration methods overlaid in red, along with the corresponding DSC. (d) MIND (e) SSC (f) α -MI (g) CAMIR (h) Proposed method

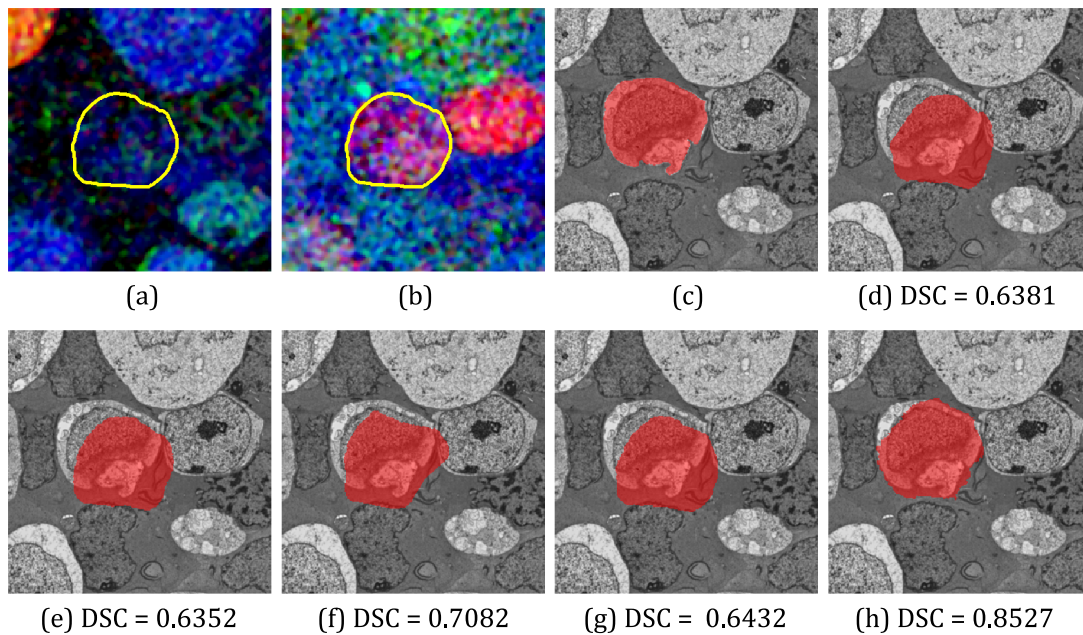


Figure 4.8: Visual results on a challenging cell. (a) RGB visualization of 3 CMP channels, with ground truth of the cell of interest outlined in yellow (b) RGB visualization of the remaining 3 CMP channels, with outlined ground truth (c) ATEM image corresponding to the same region, with ground truth overlaid in red (d-h) ATEM image with results from various registration methods overlaid in red, along with the corresponding DSC. (d) MIND (e) SSC (f) α -MI (g) CAMIR (h) Proposed method

of the transformation by allowing local adaptation in the transition probability matrix. Multi-channel input data, if available, is utilized in an efficient manner by incorporating it into the emission probabilities of the HMM. Further, we use an efficient approximation to train and decode the T-HMM at reduced complexity. The results of our method show substantial gains over state-of-the-art deformable registration techniques on both intra-modal and inter-modal problems.

Chapter 5

Segmentation and Tracking of Cells through Multimodal Label Transfer

Automated segmentation of electron microscope (EM) images is a challenging problem, but the presence of related images of a different modality can be a valuable resource. The chapter begins with a discussion on related methods in literature. Then, we describe a method to effectively utilize the complementary information in ATEM segmentation, using a multi-step approach. Each segmented cell boundary is then used to initialize a tracking algorithm through the ATEM stack. Combined with the CMP segmentation, this method forms a completely automated

system for segmentation, classification and tracking of the connectome data. We then provide experimental results for both segmentation and tracking of ATEM images, followed by concluding remarks.

5.1 Introduction

A central task in connectome analysis is building the underlying network of connections from EM image data, which relies on accurate segmentation of electron microscopy images. Segmentation of cells in the ATEM images in RC1 is particularly difficult, due to considerable variation in appearance of cells, cluttered background, and, in some cases, low contrast between cell interiors and boundaries. On the other hand, cells in the CMP image can be segmented and classified by performing multivariate analysis on 6 layers taken together, as described in Chapter 3. This observation led us to explore methods that utilize the information, available from the CMP modality, to achieve reliable EM segmentation.

5.1.1 Related work

In the recent past, there has been considerable interest in developing reliable segmentation algorithms for electron microscope (EM) images [60, 61, 62, 63, 64,

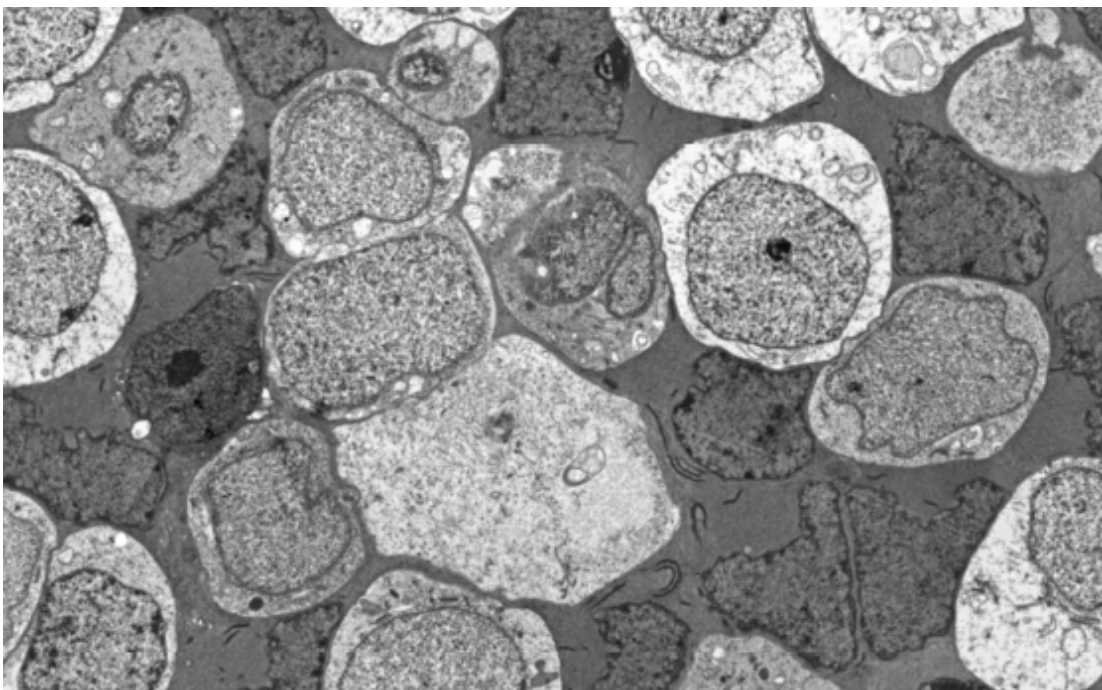


Figure 5.1: An example of a region in an ATEM image. We observe considerable variation in appearance within and across cells, low contrast between cells and cell boundaries and background clutter.

65]. In [60], membrane detection is performed using a random forest classifier followed by gap completion. A neural network approach to predict membranes is described in [62]. In [61] and [63], the authors use support vector machines (SVMs) to learn shape-based and context-based features respectively, which are used to segment mitochondria in EM images. Recently introduced methods have used hierarchical clustering with active learning [64] and merge trees [65] to perform EM segmentation. However, most existing EM segmentation algorithms are unable to reliably segment ATEM images from RC1 due to the variation in cell appearance through the image and lack of a clearly defined membrane in some cells. Further, due to background clutter and variation of cell characteristics through the image, traditional region-based segmentation approaches such as graph cuts [66, 67, 68] require careful initialization on each cell to perform well on such data.

Since the CMP data is available as an additional source of information, an alternative approach is to apply multimodal registration to deform CMP segments to the corresponding region in the ATEM modality. There are many recent methods that aim at solving the multimodal alignment problem, as discussed in Chapter 4. The drawback of using such methods is that they attempt to optimize a criterion that measures the quality of the mapping itself, which may be mismatched with our ultimate objective of optimal ATEM segmentation.

In this chapter, we propose a multi-stage technique for segmentation of ATEM

images. Information from automatically labeled and segmented CMP images is transferred to corresponding ATEM images by finding the optimal mapping between the two images. The initial contours obtained in the ATEM image as a result of this mapping are then further processed to get improved segmentation. The described system has the added advantage of transferring cell type information (which is difficult to determine using ATEM images alone) to each segment. Further, obtained cell segmentations are used to initialize tracking through the ATEM stack, resulting in an automated system for network reconstruction within the connectome.

5.2 Proposed Method

The proposed method aims at segmenting cells in ATEM images by leveraging information transferred from CMP to ATEM. We use a multi-stage technique to perform segmentation. The first step involves segmenting and labeling cells in CMP using a superpixel grid. Supersegmentation is then applied to the ATEM image, and a 2D HMM is used to find the optimal mapping between the two images. The cell segmentations are then transferred from CMP to ATEM, and an HMM-based contour refinement method is used to further improve the quality of segmentation. Each individual cell is then tracked through the ATEM stack using the refined contour as initialization.

5.2.1 CMP Segmentation

The multi-marker classification and segmentation procedure outlined in Chapter 3 is used to initialize the ATEM segmentation algorithm.

5.2.2 Multimodal Segmentation Transfer

Cell segmentations from the CMP image are transferred to the ATEM image using a 2D HMM framework built on a superpixel lattice. We oversegment the ATEM image using SLIC superpixels, in a similar setting to that used for CMP oversegmentation.

We construct a 2D HMM over a first order Markov mesh random field of size $X \times Y$, where X and Y are, respectively, the number of superpixels per row and column in the CMP image. Each superpixel in the CMP image corresponds to a node at a location (x, y) in the 2D HMM, and is denoted by $\mathbf{C}_{x,y}$. The label of $\mathbf{C}_{x,y}$ is denoted by $\omega_{x,y}$. Similarly, $\mathbf{A}_{x,y}$ denotes a superpixel at a location (x, y) in the ATEM superpixel lattice.

Each state q of the 2D HMM corresponds to a unique mapping Δ from a superpixel $\mathbf{C}_{x,y}$ in CMP to $\mathbf{A}_{x+\Delta_x, y+\Delta_y}$ in ATEM. Our aim is to find the optimal state sequence, $Q^* = \{q_{x,y}^*, x = 1, 2, \dots, X, y = 1, 2, \dots, Y\}$, which describes a deformation of the superpixel lattice. This deformation is used to map labels from superpixels in CMP to superpixels in ATEM.

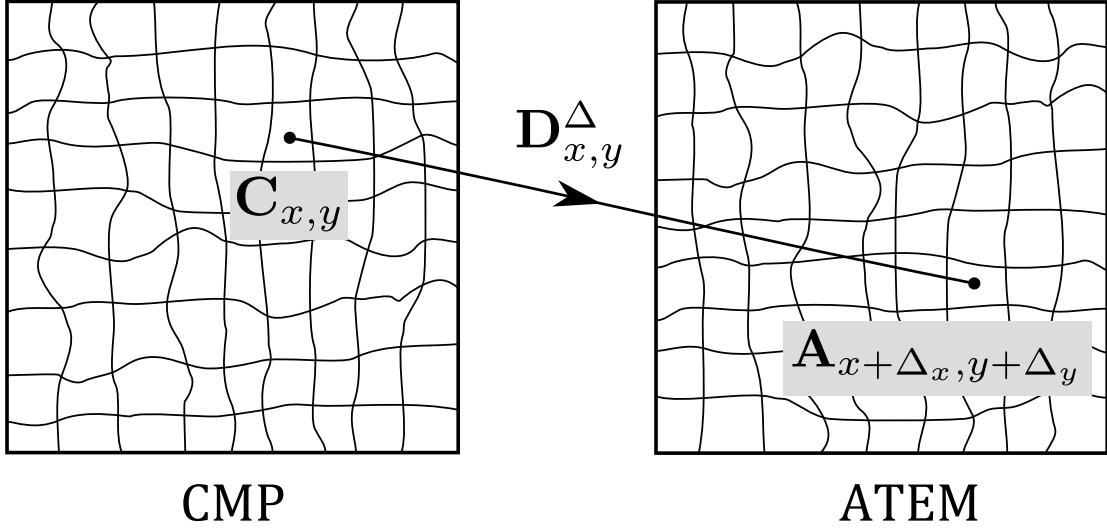


Figure 5.2: Mapping a label from a SLIC superpixel in CMP to a SLIC superpixel in ATEM, using a mapping Δ .

Data Matching Costs

The cost of matching a superpixel from the CMP image to a superpixel from the ATEM is captured by the emission probabilities of the 2D HMM. The emission probability $b_{x,y}^{\Delta}$ represents the probability of matching $C_{x,y}$ to $A_{x+\Delta_x, y+\Delta_y}$ (see Fig5.2).

$$b_{x,y}^{\Delta} = P(\mathbf{A}_{x+\Delta_x, y+\Delta_y} | \mathbf{C}_{x,y}) \quad (5.1)$$

We extract two features from every superpixel $\mathbf{A}_{x,y}$ in the ATEM image, the average and median intensity in the superpixel, and combine them into a single feature vector $\mathbf{f}_{x,y}$. Given the label $\omega_{x,y}$ from $C_{x,y}$, emission probability can be

rewritten as

$$b_{x,y}^\Delta = P(\mathbf{f}_{x+\Delta_x,y+\Delta_y} | \omega_{x,y}) \quad (5.2)$$

We model $P(\mathbf{f}_{x+\Delta_x,y+\Delta_y} | \omega_m)$ for the class ω_m with a mixture of K Gaussians.

$$P(\mathbf{f}_{x+\Delta_x,y+\Delta_y} | \omega_{x,y}) = \sum_{k=1}^K w_m^k P(\mathbf{f}_{x+\Delta_x,y+\Delta_y} | \omega_{x,y}^k) \quad (5.3)$$

where the set of mixture component weights for each class must satisfy $\sum_{k=1}^K w_m^k = 1 \forall m \in \{1, 2, \dots, M\}$. Each individual component density is a Gaussian of dimension 2.

$$P(\mathbf{f}_{x+\Delta_x,y+\Delta_y} | \omega_m^k) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{f}_{x+\Delta_x,y+\Delta_y} - \boldsymbol{\mu}_m^k)^T \boldsymbol{\Sigma}_m^k{}^{-1} (\mathbf{f}_{x+\Delta_x,y+\Delta_y} - \boldsymbol{\mu}_m^k)\right\}}{(2\pi)^{|\boldsymbol{\Sigma}_m^k|} \frac{1}{2}} \quad (5.4)$$

where $\boldsymbol{\mu}_m^k$ and $\boldsymbol{\Sigma}_m^k$ are the mean and covariance, respectively.

Neighborhood Consistency

Consistency between translations of neighboring superpixels is ensured by the transition probabilities of the 2D HMM. The neighborhood consistency measure from 4 is adapted for use in this application. In contrast to previous chapter (where the 2D HMM was set up on the regular source image grid), the grid in this approach is set up on the superpixel lattice. In order to account for the

irregularity of the grid, we define a spatially varying transition probability matrix which depends on the centroids of superpixels in both images.

A state corresponding to a mapping Δ maps the superpixel $\mathbf{C}_{x,y}$ in CMP to $\mathbf{A}_{x+\Delta_x,y+\Delta_y}$ in ATEM. Let $\mathbf{D}_{x,y}^\Delta$ be a 2D vector representing the distance between the centroids of superpixels $\mathbf{C}_{x,y}$ and $\mathbf{A}_{x+\Delta_x,y+\Delta_y}$. Each element of the horizontal transition probability matrix, $a_{\Delta',\Delta}^H(x,y)$, represents the probability of moving from state Δ at a location $(x-1,y)$ to the state Δ' at (x,y) in the 2D HMM, and is modeled by a Gaussian given by

$$a_{\Delta',\Delta}^H(x,y) \propto \exp \left\{ -\frac{1}{2} \left(\frac{\|\mathbf{D}_{x,y}^{\Delta'} - \mathbf{D}_{x-1,y}^\Delta\|^2}{\sigma^2} \right) \right\} \quad (5.5)$$

where $\|\cdot\|$ denotes Euclidean distance. This model ensures smoothness in the resulting deformation by encouraging neighboring superpixels to take similar translations. Since cellular microscopy images do not typically exhibit different behavior in different directions, we assume a similar model for the vertical transition probability matrix.

Parameter Estimation and Inference

The parameters of the 2D HMM are learned using Baum-Welch training [5], and the optimal state sequence is inferred using the Viterbi decoding algorithm with modified forward-backward iterations, as described in [18].

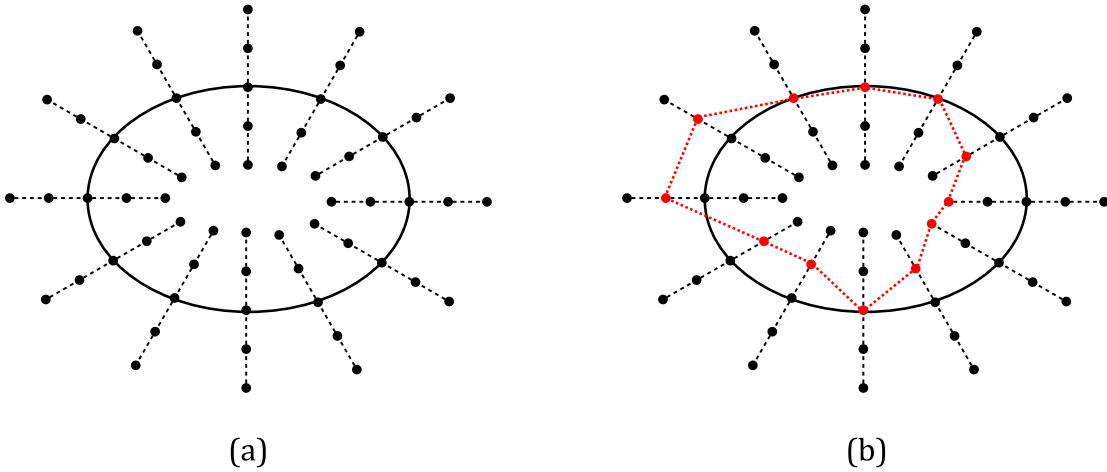


Figure 5.3: (a) Construction of a deformable trellis around the initial contour. The black curve is the initial contour, the dashed lines are the constructed normals, and each black dot corresponds to a state. (b) An example of an estimated contour (in red) obtained by joining a given sequence of states (red dots).

5.2.3 Contour Refinement

We apply a basic version of the tracking algorithm from [69] to refine the obtained segmentation. Each segmented cell is taken individually, and an initial contour is set up on perimeter of the cell. Nodes for the HMM are initialized by sampling points at constant spacing along this contour, resulting in N_ϕ total nodes. A normal line is constructed at each node, and N_ψ equally spaced points (each corresponding to a state) are placed symmetrically along the normal (see Fig.5.3(a)), resulting in a deformable trellis. A given sequence of states, $Q = \{q_\phi, \phi = 1, 2, \dots, N_\phi\}$, corresponds to a path through the trellis, which forms a contour (see Fig.5.3(b)). Our aim is to estimate the optimal cell boundary by

deforming the initial contour.

Emission Probability

The emission probabilities of the HMM represent the cost of the contour passing through a given point on the trellis. We extract three features to measure the emission probability b_ϕ^ψ for a state ψ at a location ϕ on the contour - one region-based feature which captures local object characteristics around the point of interest and two edge features which involve the first and second order average gradients along the normal line. The three features are concatenated into a single feature vector \mathbf{f}_ϕ^ψ , and the emission probability is modeled with a 3-dimensional Gaussian with mean $\boldsymbol{\mu}_c$ and variance $\boldsymbol{\Sigma}_c$.

$$b_\phi^\psi = \frac{\exp\{-\frac{1}{2}(\mathbf{f}_\phi^\psi - \boldsymbol{\mu}_c)^T \boldsymbol{\Sigma}_c^{-1} (\mathbf{f}_\phi^\psi - \boldsymbol{\mu}_c)\}}{(2\pi)^{\frac{3}{2}} |\boldsymbol{\Sigma}_c|^{\frac{1}{2}}} \quad (5.6)$$

Transition Probability

The transition probabilities of the HMM ensure smoothness in the estimated contour, and are modeled with a tilted Gibbs distribution. Each element of the transition probability matrix, $a_{\psi',\psi}$, is the probability of moving to state ψ' from

state ψ , and is given by

$$a_{\psi',\psi} \propto \exp \left\{ -\frac{(\psi - \psi')^2}{2\theta_1^2} \right\} \cdot \exp \left\{ -\frac{(\psi' - \psi_0)^2}{2\theta_2^2} \right\} \quad (5.7)$$

where ψ_0 is the middle state, which corresponds to the initial contour. The term containing θ_1 ensures consistency between consecutive points on the contour, whereas the term containing θ_2 penalizes large deviations from the initial contour.

Parameter Estimation and Inference

The emission parameters for the HMM used for contour refinement are learned using a support vector machine (SVM) [70] on labeled training data. Transition parameters are trained using maximum likelihood (Baum-Welch) on the training dataset. The optimal sequence is inferred using the Viterbi algorithm, and the resulting contour forms the refined segmentation boundary of the cell.

5.2.4 Cell Tracking

We note that the method described in Section 5.2.3 relies on SVM-based parameter estimation. Therefore, for successful tracking using this algorithm, the data must satisfy one of two conditions: (i) the intensity profile of cells must remain fairly similar while traversing the stack or (ii) there must be extensive

ground truth available for all frames. Since manual annotation is tedious for large volumes of data and appearance of ATEM images varies considerably from frame to frame, we cannot use the described algorithm directly, and instead use the variant described in [71] for tracking. The contour from the first frame of ATEM is used to initialize the tracking, and the final contour from each frame is used as initial contour in the following frame, with HMM parameters updated after each iteration of deformation. As this is a topology aware algorithm, it has the ability to handle splits and merges of cells, as well as large displacement between layers. This is particularly useful in case of skipped or damaged slices, which are common in the RC1 data.

5.3 Experimental Validation

5.3.1 ATEM Segmentation

We test the performance of several methods on images obtained from the RC1 connectome [2]. The test data consists of 85 total cells in the first frame. To evaluate the accuracy of segmentation of each method, we compare the result with manually guided expert annotated ground truth. Segmented cells are optimally matched with ground truth cells using magnitude of overlap. The F-measure F is used as a measure of similarity between each ground truth cell (S_{GT}) and its

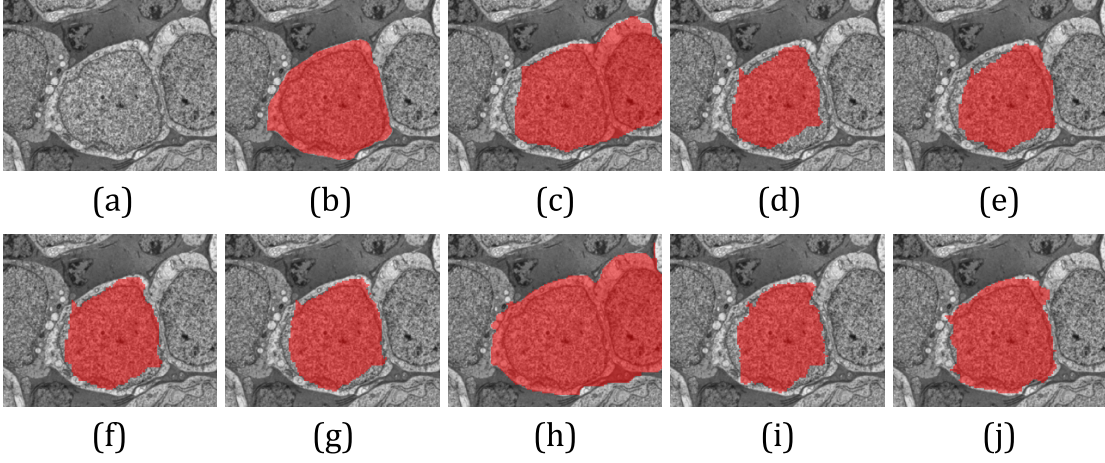


Figure 5.4: (Best in color) Visual results of cell segmentation in a challenging scenario. (a) Region of interest in the ATEM image (b) Ground truth of the cell overlaid in red (c-i) Image with overlaid segmentation results from (c) membrane detection method (Kaynig et al.) (d) DROP (e) α -MI (f) MIND (g) SSC (h) Graph cuts (i) HMM-based multimodal fusion (j) the proposed approach

corresponding segmentation (S_{SEG}).

$$F = \frac{2 \cdot PR \cdot RC}{PR + RC} \quad (5.8)$$

where the precision PR is given by $PR = \frac{|S_{GT} \cap S_{SEG}|}{|S_{SEG}|}$ and recall RC is given by $RC = \frac{|S_{GT} \cap S_{SEG}|}{|S_{GT}|}$. The operator $|\cdot|$ denotes number of pixels. The area (in pixels) of each ground truth cell is used as weight in the calculation of F-measure statistics.

We compare the performance of the proposed approach with related approaches that fall in two categories, direct segmentation methods and segmentation trans-

fer via multimodal registration. The methods are listed below along with their respective optimal setting (each empirically found).

Direct ATEM Segmentation Methods: (i) Membrane detection using random forest classification followed by gap completion [60] with manually selected cells (since only contours are detected) (ii) Graph cut segmentation [66] in a multiple object setting, with parameters learned via manual seeding of a small subset of cells.

Label transfer-based Methods: (i) Graph cut segmentation [66], seeded with the results of CMP segmentation (ii) DROP [51], with $\lambda = 0.1$ (iii) α -MI-based registration [53], with $\alpha = 0.99$ and $k = 7$ (iv) MIND [54], with $\alpha = 0.1$ (v) SSC [55], with $\alpha = 0.1$ (vi) Our recent approach for multimodal fusion [72] (vii) the proposed method.

A comparison of quantitative results is provided in Table 5.1. We see that the proposed approach shows considerable improvement over competing methods in terms of segmentation accuracy. Visual results for a challenging scenario (touching cells without a well-defined membrane separating them) are shown in Fig. 5.4. We observe that direct EM segmentation methods are unable to separate the cell of interest from the visually similar adjoining cell, while registration-based methods utilize the label information from CMP to mitigate this problem. The additional contour refinement step of the proposed method results in further improved cell

Method	Mean	Std Dev
Membrane detection [60]	0.7016	0.2732
Graph Cuts [66] (automatic)	0.6765	0.2889
DROP [51]	0.7412	0.1674
α -MI [53]	0.7288	0.1620
MIND [54]	0.7537	0.1688
SSC [55]	0.7523	0.1696
Graph Cuts [66] (seeded from CMP)	0.7808	0.2014
Multimodal fusion [72]	0.7978	0.1480
Proposed method	0.8651	0.1287

Table 5.1: F-measure statistics of segmentation results from various methods.

Method	Mean	Std Dev
Chan-Vese Algorithm [73]	0.8802	0.1439
HMM-based Tracking	0.9268	0.0604

Table 5.2: F-measure statistics comparison for tracking.

segmentation.

5.3.2 ATEM Cell Tracking

We compare the results of our algorithm with a level set tracking algorithm [73], initialized with the refined segmentation obtained in our segmentation refinement step. F-measure statistics of both methods on 10 cells over 10 frames against manually annotated ground truth are reported in Table 5.2. The proposed algorithm is observed to outperform the level set-based method in tracking. We note that the tracking algorithm [71] was developed for use with manual initialization.

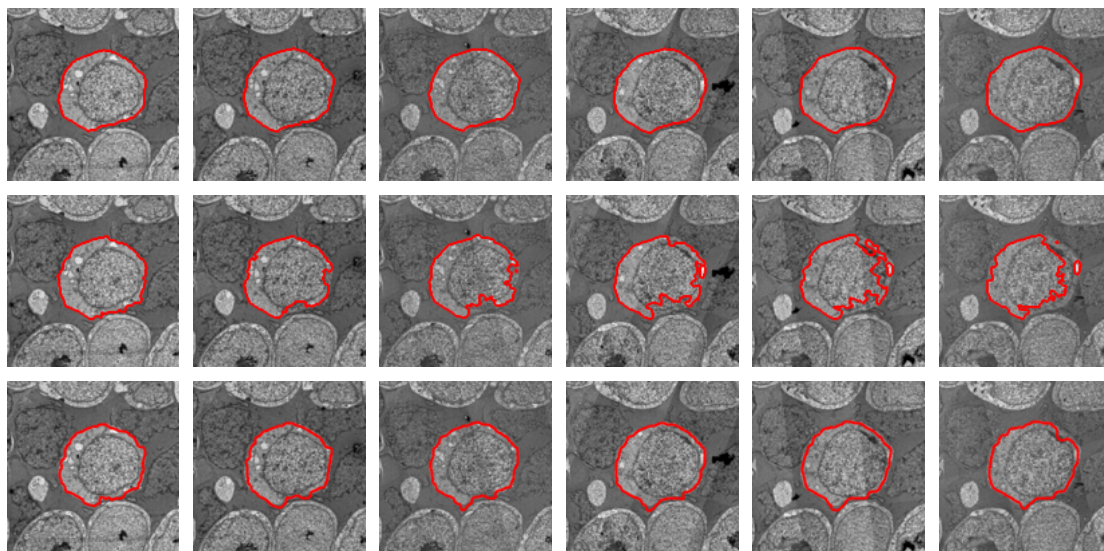


Figure 5.5: Tracking results on an ATEM sequence. The first row shows ground truth, the second row shows results of Chan-Vese tracking, and the third row shows results of the HMM-based tracking algorithm.

However, we have demonstrated how it may be adapted for use with automated initialization.

5.4 Conclusion

This chapter presents a novel approach to segment and track objects in a multimodal setting, where information from a supplementary source is used to facilitate the segmentation of a challenging dataset. The labels from segmented cells in light microscopy images are transferred to electron micrographs using a 2D HMM-based mechanism built over a superpixel lattice. The obtained cell segmentations are refined using a HMM-based contour refinement technique. The

refined cell boundaries are then used to initialize a closed-contour cell tracking mechanism. Experimental results show the capability of the proposed approach to effectively segment and track cells in ATEM images.

Chapter 6

Conclusions and Future Work

In this dissertation, we focus on developing techniques for the automated analysis of biomedical data, with an emphasis on multimodal microscopy data.

In Chapter 3 we describe an algorithm to simultaneously segment and classify cells of multiple classes in a multi-marker image, using a superpixel grid and HMM-based optimization. In Chapter 4 a multi-modal, multi-channel deformable registration framework is introduced, wherein modality specific costs and smoothness constraints are quantified by the parameters of a 2D HMM. With the knowledge that standard EM segmentation algorithms are unable to perform well on the described ATEM data, Chapter 5 introduces a multi-stage label transfer based segmentation system for ATEM. Following segmentation in the first frame, tracking of the cell through the stack is also performed.

To the best of our knowledge, this is the first attempt at developing algorithms for the comprehensive automated reconstruction of a multi-modal biological dataset.

6.1 Future Work

We suggest a few possible future directions, both from a broad point of view as well as specific to the methods described in previous chapters.

The most widely used paradigm in microscopic data analysis separates computer vision tasks such as image segmentation and registration from the downstream statistical analysis. It is conceivable that integrating statistical tools at the image analysis stage could lead to improved performance. (For example, explicitly considering the subclass attributes discovered [2] in CMP data could lead to improved cell segmentation).

Multimodal image acquisition is very common in biomedical image analysis, and often, a specific task is much easier to perform on one modality than the other(s). It will be interesting to explore how domain adaptation could be applied in such cases to leverage the knowledge for challenging datasets.

With respect to the methods described in this dissertation, we propose a few extensions that could improve upon obtained results. A multi-channel extension of the contour refinement algorithm can be designed to correct superpixel-induced

errors in CMP segmentation. Cells can be tracked until the next CMP layer is reached, and the data in the CMP image can be used to check and correct for the accuracy of the tracked sequence. If necessary, another cell track could be run in the reverse direction to correct errors.

Another possible area of work is how to deal with cells or regions of low confidence in algorithm output. This would entail computing a local uncertainty score along with the result. We may choose to exclude regions of low confidence from further statistical analysis, or use an active learning framework with expert input.

Finally, while the algorithms in this dissertation focus on 2D-HMMs, it would be interesting to see how the ideas presented here could be extended to higher dimensional data.

Bibliography

- [1] T. E. J. Behrens and O. Sporns, *Human connectomics*, *Current Opinion in Neurobiology* **22** (2012), no. 1 144–153. [1](#)
- [2] J. R. Anderson, B. W. Jones, C. B. Watt, M. V. Shaw, J.-H. Yang, D. DeMill, J. S. Lauritzen, Y. Lin, K. D. Rapp, D. Mastronarde, *et. al.*, *Exploring the retinal connectome*, *Molecular Vision* **17** (2011) 355. [2](#), [46](#), [83](#), [90](#)
- [3] L. Baum and T. Petrie, *Statistical inference for probabilistic functions of finite state Markov chains*, *The Annals of Mathematical Statistics* (1966) 1554–1563. [6](#)
- [4] L. E. Baum and J. A. Eagon, *An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology*, *Bulletin of the American Mathematical Society* **73** (1967), no. 3 360–363. [6](#), [11](#)
- [5] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*, *The Annals of Mathematical Statistics* **41** (1970), no. 1 164–171. [6](#), [16](#), [35](#), [56](#), [79](#)
- [6] J. K. Baker, *The DRAGON system—an overview*, *IEEE Transactions on Acoustics, Speech and Signal Processing* **23** (1975), no. 1 24–29. [6](#)
- [7] F. Jelinek, *Fast sequential decoding algorithm using a stack*, *IBM Journal of Research and Development* **13** (1969), no. 6 675–685. [6](#)
- [8] S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, *An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition*, *The Bell System Technical Journal* **62** (1983), no. 4 1035–1074. [6](#), [16](#)

- [9] B.-H. Juang, *On the hidden Markov model and dynamic time warping for speech recognition—a unified view*, *AT&T Bell Laboratories Technical Journal* **63** (1984), no. 7 1213–1243. [6](#)
- [10] L. R. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition*, *Proceedings of the IEEE* **77** (1989), no. 2 257–286. [7](#)
- [11] A. J. Viterbi, *Error bounds for convolutional codes and an asymptotically optimum decoding algorithm*, *IEEE Transactions on Information Theory* **13** (1967), no. 2 260–269. [14](#)
- [12] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, *Journal of the Royal Statistical Society. Series B (Methodological)* (1977) 1–38. [16](#)
- [13] L. R. Welch, *Hidden Markov models and the Baum-Welch algorithm*, *IEEE Information Theory Society Newsletter* **53** (2003), no. 4 1–14. [16](#), [56](#)
- [14] J. Li, A. Najmi, and R. M. Gray, *Image classification by a two-dimensional hidden Markov model*, *IEEE Transactions on Signal Processing* **48** (2000), no. 2 517–533. [19](#)
- [15] S.-S. Kuo and O. E. Agazzi, *Keyword spotting in poorly printed documents using pseudo 2-D hidden Markov models*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16** (1994), no. 8 842–848. [19](#)
- [16] F. Perronnin, J.-L. Dugelay, and K. Rose, *Iterative decoding of two-dimensional hidden Markov models*, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. III–329, 2003. [19](#), [20](#), [23](#), [37](#), [56](#)
- [17] M. E. Sargin, A. Altinok, K. Rose, and B. S. Manjunath, *Conditional iterative decoding of two dimensional hidden Markov models*, in *IEEE International Conference on Image Processing*, pp. 2552–2555, 2008. [19](#), [20](#)
- [18] F. Perronnin, J.-L. Dugelay, and K. Rose, *A probabilistic model of face mapping with local transformations and its application to person recognition*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27** (2005), no. 7 1157–1171. [23](#), [53](#), [79](#)
- [19] A. Maity, P. F. Sullivan, and J.-Y. Tzeng, *Multivariate phenotype association analysis by marker-set kernel machine regression*, *Genetic epidemiology* **36** (2012), no. 7 686–695. [25](#)

- [20] G. Lu and B. Fei, *Medical hyperspectral imaging: a review*, *Journal of biomedical optics* **19** (2014), no. 1 010901–010901. [25](#)
- [21] R. E. Marc, R. F. Murry, and S. F. Basinger, *Pattern recognition of amino acid signatures in retinal neurons*, *The Journal of neuroscience* **15** (1995), no. 7 5106–5129. [25](#)
- [22] N. Otsu, *A threshold selection method from gray-level histograms*, *Automatica* **11** (1975), no. 285-296 23–27. [26](#)
- [23] F. H. Y. Chan, F. K. Lam, and H. Zhu, *Adaptive thresholding by variational method*, *IEEE transactions on Image Processing* **7** (1998), no. 3. [26](#)
- [24] B. Gatos, I. Pratikakis, and S. J. Perantonis, *Adaptive degraded document image binarization*, *Pattern recognition* **39** (2006), no. 3 317–327. [26](#)
- [25] D. Comaniciu and P. Meer, *Mean shift: A robust approach toward feature space analysis*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** (2002), no. 5 603–619. [28](#), [37](#), [39](#)
- [26] B. Fulkerson, A. Vedaldi, and S. Soatto, *Class segmentation and object localization with superpixel neighborhoods*, in *IEEE 12th International Conference on Computer Vision, 2009*, pp. 670–677, 2009. [28](#), [38](#), [39](#)
- [27] M. Y. Yang and W. Forstner, *A hierarchical conditional random field model for labeling and classifying images of man-made scenes*, in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 196–203, 2011. [28](#)
- [28] D. Singaraju and R. Vidal, *Using global bag of features models in random fields for joint categorization and segmentation of objects*, in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2313–2319, 2011. [28](#)
- [29] R. Achanta *et. al.*, *SLIC superpixels compared to state-of-the-art superpixel methods*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34** (2012), no. 11 2274–2282. [32](#)
- [30] A. Vedaldi and B. Fulkerson, *VLFeat: An open and portable library of computer vision algorithms*, . [32](#)
- [31] J. Cheng and J. C. Rajapakse, *Segmentation of clustered nuclei with shape markers and marking function*, *IEEE Transactions on Biomedical Engineering* **56** (2009), no. 3 741–748. [37](#)

- [32] R. P. Woods, S. R. Cherry, and J. C. Mazziotta, *Rapid automated algorithm for aligning and reslicing PET images*, *Journal of computer assisted tomography* **16** (1992), no. 4 620–633. 44
- [33] P. Thevenaz, U. E. Ruttimann, and M. Unser, *A pyramid approach to subpixel registration based on intensity*, *IEEE Transactions on Image Processing* **7** (1998), no. 1 27–41. 44
- [34] K. Ø. Noe, K. Tanderup, J. C. Lindegaard, C. Grau, and T. S. Sørensen, *GPU accelerated viscous-fluid deformable registration for radiotherapy*, *DAIMI Report Series* **36** (2007), no. 583. 44
- [35] M. Foskey, B. Davis, L. Goyal, S. Chang, E. Chaney, N. Strehl, S. Tomei, J. Rosenman, and S. Joshi, *Large deformation three-dimensional image registration in image-guided radiation therapy*, *Physics in Medicine and Biology* **50** (2005), no. 24 5869. 44
- [36] J. McClelland, J. Blackall, S. Tarte, A. Chandler, S. Hughes, S. Ahmad, D. B. Landau, and D. J. Hawkes, *A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy*, *Medical Physics* **33** (2006), no. 9 3348–3358. 44
- [37] R. Shekhar, P. Lei, C. R. Castro-Pareja, W. L. Plishker, and W. D. DSouza, *Automatic segmentation of phase-correlated CT scans through nonrigid image registration using geometrically regularized free-form deformation*, *Medical Physics* **34** (2007), no. 7 3054–3066. 44
- [38] D. Yang, W. Lu, D. A. Low, J. O. Deasy, A. J. Hope, and I. El Naqa, *4D-CT motion estimation using deformable image registration and 5d respiratory motion modeling*, *Medical Physics* **35** (2008), no. 10 4577–4590. 44
- [39] R. P. Woods, J. C. Mazziotta, S. R. Cherry, *et. al.*, *MRI-PET registration with automated algorithm*, *Journal of computer assisted tomography* **17** (1993), no. 4 536–546. 44
- [40] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal, *3D multi-modality medical image registration using feature space clustering*, in *Computer Vision, Virtual Reality and Robotics in Medicine*, pp. 193–204, Springer, 1995. 44

- [41] C. Studholme, D. L. G. Hill, and D. J. Hawkes, *Automated 3-D registration of MR and CT images of the head*, *Medical image analysis* **1** (1996), no. 2 163–175. 44
- [42] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, *Multimodality image registration by maximization of mutual information*, *Medical Imaging, IEEE Transactions on* **16** (1997), no. 2 187–198. 44
- [43] P. Viola and W. M. Wells III, *Alignment by maximization of mutual information*, *International journal of computer vision* **24** (1997), no. 2 137–154. 44
- [44] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever, *Mutual-information-based registration of medical images: a survey*, *IEEE Transactions on Medical Imaging* **22** (2003), no. 8 986–1004. 44
- [45] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, *Nonrigid registration using free-form deformations: application to breast MR images*, *IEEE Transactions on Medical Imaging* **18** (1999), no. 8 712–721. 44, 64
- [46] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, et. al., *A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations*, in *Medical Image Computing and Computer-Assisted Intervention(MICCAI)*, pp. 573–581, Springer, 2001. 44
- [47] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, *PET-CT image registration in the chest using free-form deformations*, *IEEE Transactions on Medical Imaging* **22** (2003), no. 1 120–128. 44
- [48] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, *Nonrigid image registration using conditional mutual information*, *IEEE Transactions on Medical Imaging* **29** (2010), no. 1 19–29. 44
- [49] M. Modat, G. R. Ridgway, Z. A. Taylor, M. Lehmann, J. Barnes, D. J. Hawkes, N. C. Fox, and S. Ourselin, *Fast free-form deformation using graphics processing units*, *Computer methods and programs in biomedicine* **98** (2010), no. 3 278–284. 45

- [50] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, *Dense image registration through MRFs and efficient linear programming*, *Medical image analysis* **12** (2008), no. 6 731–741. 45
- [51] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, *Deformable medical image registration: Setting the state of the art with discrete methods*, *Annual review of biomedical engineering* **13** (2011) 219–244. 45, 62, 63, 85, 86
- [52] G. K. Rohde, S. Pajevic, C. Pierpaoli, and P. J. Basser, *A comprehensive approach for multi-channel image registration*, in *Biomedical Image Registration*, pp. 214–223. Springer, 2003. 45, 64, 65
- [53] M. Staring, U. A. van der Heide, S. Klein, M. A. Viergever, and J. Pluim, *Registration of cervical MRI using multifeature mutual information*, *IEEE Transactions on Medical Imaging* **28** (2009), no. 9 1412–1421. 45, 64, 65, 85, 86
- [54] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Martin, F. V. Gleeson, S. M. Brady, and J. A. Schnabel, *MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration*, *Medical Image Analysis* **16** (2012), no. 7 1423–1435. 45, 62, 63, 64, 65, 85, 86
- [55] M. P. Heinrich, M. Jenkinson, B. W. Papież, M. Brady, and J. A. Schnabel, *Towards realtime multimodal fusion for image-guided interventions using self-similarities*, in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, pp. 187–194. Springer, 2013. 45, 62, 63, 64, 65, 85, 86
- [56] J. Weickert and C. Schnörr, *A theoretical framework for convex regularizers in PDE-based computation of image motion*, *International Journal of Computer Vision* **45** (2001), no. 3 245–264. 54
- [57] A. Wedel, D. Cremers, T. Pock, and H. Bischof, *Structure-and motion-adaptive regularization for high accuracy optic flow*, in *IEEE International Conference on Computer Vision*, pp. 1663–1668, 2009. 54
- [58] L. R. Dice, *Measures of the amount of ecologic association between species*, *Ecology* **26** (1945), no. 3 297–302. 61
- [59] F. Wilcoxon, *Individual comparisons by ranking methods*, *Biometrics* **1** (1945), no. 6 80–83. 61

- [60] V. Kaynig, T. Fuchs, and J. M. Buhmann, *Neuron geometry extraction by perceptual grouping in ssTEM images*, in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2902–2909, 2010. 74, 85, 86
- [61] A. Lucchi, K. Smith, R. Achanta, G. Knott, and P. Fua, *Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features*, *IEEE Transactions on Medical Imaging* **31** (2012), no. 2 474–486. 74
- [62] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, *Deep neural networks segment neuronal membranes in electron microscopy images*, in *Advances in Neural Information Processing Systems*, pp. 2843–2851, 2012. 74
- [63] A. Lucchi, C. J. Becker, P. Marquez Neila, and P. Fua, *Exploiting enclosing membranes and contextual cues for mitochondria segmentation*, in *Medical Image Computing and Computer Assisted Intervention*, pp. 65–72, 2014. 74
- [64] J. Nunez-Iglesias, R. Kennedy, T. Parag, J. Shi, and D. B. Chklovskii, *Machine learning of hierarchical clustering to segment 2D and 3D images*, *PLoS One* **8** (2013), no. 8 e71715. 74
- [65] T. Liu, C. Jones, M. Seyedhosseini, and T. Tasdizen, *A modular hierarchical approach to 3D electron microscopy image segmentation*, *Journal of Neuroscience Methods* **226** (2014) 88–102. 74
- [66] Y. Boykov and G. Funka-Lea, *Graph cuts and efficient ND image segmentation*, *International Journal of Computer Vision* **70** (2006), no. 2 109–131. 74, 85, 86
- [67] Y. Boykov and V. Kolmogorov, *An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26** (2004), no. 9 1124–1137. 74
- [68] V. Kolmogorov and R. Zabini, *What energy functions can be minimized via graph cuts?*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26** (2004), no. 2 147–159. 74
- [69] M.-C. Shih, *Probabilistic graphical models for contour tracking and segmentation in electron microscopy images*, 2014. 80

- [70] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector machines*, *ACM Transactions on Intelligent Systems and Technology* **2** (2011) 27:1–27:27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. 82
- [71] M.-C. Shih and K. Rose, *Hidden Markov models for tracking neuronal structure contours in electron micrograph stacks*, in *IEEE International Symposium on Biomedical Imaging*, pp. 1377–1380, 2012. 83, 86
- [72] R. Shenoy, M.-C. Shih, and K. Rose, *Hidden Markov model-based multi-modal image fusion with efficient training*, in *IEEE International Conference on Image Processing*, 2014. 85, 86
- [73] T. F. Chan, L. Vese, *et. al.*, *Active contours without edges*, *IEEE transactions on Image Processing* **10** (2001), no. 2 266–277. 86