Computer Aided Medical Procedures
Prof. Dr. Nassir Navab

Dissertation

Local and Global Consistency Measures in Intensity-based Medical Image Registration

Mehmet Yigitsoy

Fakultät für Informatik
Technische Universität München

TECHNISCHE UNIVERSITÄT MÜNCHEN

Computer Aided Medical Procedures & Augmented Reality / I16

Local and Global Consistency Measures in Intensity-based Medical Image Registration

Mehmet Yigitsoy

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Daniel Cremers

Prüfer der Dissertation:

1. Univ.-Prof. Dr. Nassir Navab
2. Univ.-Prof. Dr.-Ing. Joachim Hornegger
   Friedrich-Alexander-Universität Erlangen-Nürnberg

Die Dissertation wurde am 28. Oktober 2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 07. Mai 2015 angenommen.

ACKNOWLEDGMENTS

I would first like to state that this thesis would not have been possible without the support, contribution, guidance and suggestions of my family, my supervisors and colleagues.

First of all, I would like to thank Nassir Navab for his continuous guidance, encouragement and feedback thoughout my PhD studies. He provided me a great environment at the CAMP chair and I enjoyed the freedom that he always provided for exploring different ideas and realizing them. He has been not only a supervisor of my PhD studies but also a real model in all aspects of my life.

I also owe thanks to the people at Carl-Zeiss Meditech AG for making my visits there enjoyable. In particular, I would like to thank Abouzar Eslami and Konstantinos Filippatos for hosting me there and for numerous discussions we had together. I would further like to thank Sonja Kirchhoff from the Radiology Department of the LMU with whom I had the pleasure of collaborating for the first two years of my PhD.

Special thanks goes to my office mate Silvan Kraft who helped me a lot with the German translation of the abstract of this thesis.

I would like to also thank the people with whom I enjoyed the pleasure of working with at CAMP and who made my stay a pleasant one including Asad Safi, Dzhoshkun Ismail Shakir, Amin Katouzian, Kristof Ralovich, Stefanie Demirci, José Gardiazabal, Darko Zikic, Ben Glocker, Christian Wachinger, Tobias Lasser, Alexander Duliu, Aslı Okur Kuru, Ralf Stauder, Tobias Reichl, Hauke Heibel, Martin Groher, Martin Horn, Martina Hilla, Dennis Hilla, Olivier Pauly, Loic Peter, Slobodan Ilic, Pascal Fallavollita, Diana Mateus, Victor Castaneda, Richard Brosig, Stefan Hinterstoisser, Silvan Kraft, Peter Maday, Vasileios Belagiannis, Christoph Hennersperger, Markus Müller, Javad Fotouhi, Hessam Roodaki, Mohammad Alsheikhali, Shadi Albarqouni, Tingying Peng, Benjamin Becker, Lichao Wang, Sailesh Conjeti, Sebastian Pölsterl, Fausto Milletari and Ahmad Ahmadi and all others that could not mention here.

Last but not least, I would like to thank my parents, my parents-in-law and in particular my wife Mihrican for her patience, love, encourage, and for standing always with me throughout the entire period. I also owe a lot of thanks to my lovely daughter Elif Kardelen for letting me to spend some of the time for my thesis that I should have spent with her. Without them, this thesis would not have been possible.

ABSTRACT

Image registration is described as the process of bringing two or more images into spatial alignment. An important application of image registration is the construction of a larger field of view of a scene from its partially overlapping observations. Transformations to bring such partial images into an alignment can be estimated by exploiting different properties of images, ranging from extracting sparse feature sets to using full image content. *Intensity-based image registration* is a remarkable subcategory of these approaches which uses image intensity values directly. Intensity-based image registration methods address the consistency between partial images in terms of pixel intensity value correspondences in their overlapping region. Nevertheless, estimation of a plausible transformation using only local intensity correspondences can be difficult especially when the size of overlap is small or when the amount of shared information is insufficient to drive an optimization procedure. In such cases, it is often a challenge to find a solution that is consistent, both locally and globally. Hence, this thesis investigates different consistency measures to be used as regularization constraints in intensity-based registration of partial images.

We, first, present a new technique called *structure propagation* for creating structural probability maps based on the geometry of locally salient structures. This approach enables us to generate "perceptually good" artificial structures beyond the observed partial image regions. The probabilistic structures are inferred using the perceptual properties of the prominent features contained in the partial observations. We show how structural probability maps can be used to assess the consistency when creating extended field of view images from partial observations. Second, we introduce a novel regularization term into the registration framework to ensure global shape consistency. To this end, we propose to use the *Hough space parametrization* of a known target anatomy or region together with a voting scheme during registration. Learned parametrization of target shapes helps us to constrain possible transformations and avoid locally consistent but globally inconsistent transformation estimates. We demonstrate how the introduced regularization term can be effectively employed for registering partial observations of a scene. The proposed techniques have been evaluated on several medical applications including microscopy image stitching and histology volume reconstruction.

In this thesis, we aim at developing a novel approach to classical image registration problems by introducing contextual consistency which benefits from the recent advances in computer vision and machine learning. This thesis will hopefully bring new insights into computer vision tasks dealing with field of view extension.

Keywords:
Medical Image Registration, Structural Consistency, Field-of-view Extension, Mosaicing, Stitching

ZUSAMMENFASSUNG

Unter Bildregistrierung versteht man den Prozess, bei welchem zwei oder mehr Bilder räumlich aneinander ausgerichtet werden. Eine wichtige Anwendung für die Bildregistrierung ist die Erweiterung des sichtbaren Bereichs bei sich teilweise überlappenden Beobachtungsbereichen. Die Transformationen um solche Teilbereiche aneinander auszurichten können durch Ausnutzung verschiedener Eigenschaften der Einzelbilder abgeschätzt werden, angefangen beim Extrahieren von sparse-feauture-sets bis hin zur Ausnutzung des gesamten Bildinhalts. *Intensitätsbasierte Bildregistrierung* ist eine bedeutende Unterkategorie dieser Ansätze, bei welcher die Bildintensität direkt verwendet wird. Die intensitätsbasierte Bildregistrierungs-Verfahren greifen das Thema von Konsistenz zwischen Teilbilder nur im Sinne der Pixelintensitätswert-Korrespondenzen in ihrer überlappenden Region auf. Trotzdem kann die Schätzung einer plausiblen Transformation allein mit lokalen Intensitätskorrespondenzen problematisch sein, insbesondere wenn die Größe der Überlappungen klein ist oder wenn die Gesamtmenge der verfügbaren Informationen unzureichend ist, um ein Optimierungsverfahren durchzuführen. In solchen Fällen ist es oft eine Herausforderung eine Lösung zu finden, die sowohl lokal als auch global konsistent ist. Daher werden im Rahmen dieser Dissertation unterschiedliche Konsistenz-Ansätze untersucht, um sie als Regularisierungseinschränkungen bei der intensitätsbasierten Registrierung von Teilbildern einzusetzen.

Zuerst präsentieren wir eine neue Technik namens *structure propagation* für die Erstellung von Strukturwahrscheinlichkeitskarten basierend auf der Geometrie von lokal auffälligen Strukturen. Dieser Ansatz ermöglicht es uns, "gut wahrnehmbare" künstliche Strukturen zu erzeugen, die über die beobachteten Teilbildbereiche hinausgehen. Die probabilistischen Strukturen werden aus den Wahrnehmungsseigenschaften der in den Teilbeobachtungen enthaltenen auffälligen Features abgeleitet. Wir zeigen, wie die Strukturwahrscheinlichkeitskarten verwendet werden können, um während der Erzeugung eines erweiterten Sichtfeldes aus partiellen Beobachtungen die Konsistenz zu bewerten. Außerdem führen wir einen neuartigen Regularisierungs-Begriff ein, damit die globale Gesamtform des Bildes erhalten bleibt. Zu diesem Zweck schlagen wir vor, die *Houghraum Parametrisierung* einer bekannten Ziel-Anatomie oder einer Region zusammen mit einem Auswahlschema bei der Registrierung verwenden. Die gelernte Parametrisierung der Ziel-Objekte hilft uns, die mögliche Transformationen zu beschränken und die lokal konsistente aber global inkonsistente Transformationsabschätzungen zu vermeiden. Wir zeigen, wie der eingeführte Regularisierungsterm effektiv zur Registrierung von Teil-Beobachtungen einer Szene eingesetzt werden kann. Der Einsatz der vorgeschlagenen Techniken wurde für verschiedene medizinische Anwendungen evaluiert, einschließlich der Bereiche Mikroskopiebild-Stitching und Volumenrekonstruktionen in der Histologie.

Diese Dissertation zielt darauf ab einen neuartigen Ansatz für klassischen Bildregistrier-

ungs-Problemen zu entwickeln, indem eine inhaltlichen Konsistenz eingeführt wird, die von den jüngsten Fortschritten in Computervision und maschinellem Lernen profitiert. Diese Dissertation wird hoffentlich neue Erkenntnisse für Aufgabenstellungen im Bereich der Computervision erbringen, welche mit Sichtfeld-Erweiterung zu tun haben.

Schlagwörter:
Bildregistrierung, Strukturkonsistenz, Sichtfeld-Erweiterung, Mosaikierung

Contents

Introduction and Background

CHAPTER

ONE

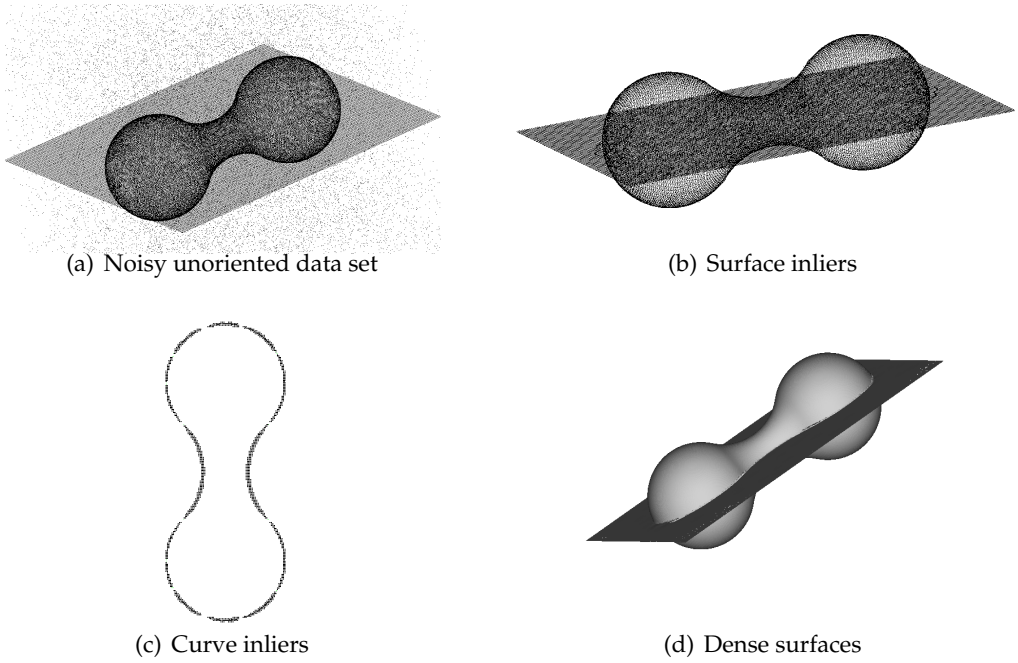
INTRODUCTION

With the advent of sophisticated image analysis and scene understanding techniques, having a large field-of-view in many applications ranging from terrestrial imaging to computer vision to medical imaging has become crucial for having more reliable measurements as well as making better generalizations from such measurements. However, the fundamental limitation of existing image acquisition devices in capturing larger regions with high resolution still remains a challenge and requires additional efforts to do the same. A typical approach for reconstructing a large field-of-view with sufficient image resolution is to combine partially overlapping high resolution images covering different parts of the scene, leading to an extended field-of-view.

There are several ways for combining partial images of a particular scene. If possible, the coordinates of the partial images can be recorded during the acquisition phase. Then, the partial images can be positioned in a global scene where post-processing steps such as blending might be necessary to account for the global illumination differences between partially overlapping images. Although such simple procedures might be applicable when the scene being imaged is static, more sophisticated techniques are necessary to account for possible deformations, which are unavoidable in most real world applications. This process gets usually more complicated if the coordinates of partial images in the global scene are inaccurately recorded or even totally unknown. Moreover, having an extended field-of-view presents further challenges if, in particular, the acquisitions are performed in an uncontrolled manner and an overlap between the partial images cannot be guaranteed.

Majority of typical approaches for having an extended field-of-view in such scenarios employ image registration techniques for estimating transformations for every partial image of the scene. Transformations are then used to bring images into a consistent configuration in the global scene. The alignment of partial images can be handled pairwise where pairs of images are aligned together in a consecutive manner or groupwise where all partial images are registered simultaneously. Simultaneous approaches have the advantage of obtaining a globally more consistent alignment at the cost of greater computational load.

In both subgroups of image registration techniques, estimation of transformations falls under two categories; geometric and iconic. Approaches employing a geometric or feature-based registration exploit the geometrical properties of the underlying scene for estimating transformation parameters. Usually, interest points derived from a particular image are



(a) Noisy unoriented data set (b) Surface inliers

(c) Curve inliers (d) Dense surfaces

FIGURE 1.1.: Inference of surfaces and surface intersections from noisy data using tensor voting. Figures and caption are from [Mord 05]. Courtesy of Philippos Mordohai.

matched to the interest points detected in another image to obtain correspondences between them. Transformations are then estimated based on the established correspondences. Iconic or intensity-based registration algorithms, on the other hand, take full advantage of image intensities either by comparing them directly or by exploiting the statistical relationships between their distributions.

Both classes of registration techniques have strengths and limitations in typical image registration setups where there is usually a sufficiently large overlap between the images to be registered together. Nevertheless, they are not directly applicable for solving registration problems of partially overlapping images with the above mentioned challenges. While it is necessary to have an overlap for feature-based, geometric registration approaches to work properly, intensity-based, iconic registration additionally suffers when the information in the limited overlap is unreliable. The unreliability of the overlap can be due to the presence of noise or other artifacts such as geometrical distortions introduced during image acquisition.

As a consequence of the limitations of classical image registration approaches mentioned above ensuring structural consistency beyond the overlapping region becomes challenging. In many applications, however, a consistent fusion of partial images is crucial for performing accurate measurements on the reconstructed scene. Any violation of properties of the underlying structures might lead to major consequences. For instance, in medical imaging, subsequent diagnoses would be directly affected by the performance of reconstruction scheme causing possible false treatment. Violations of structural consistency can be local where a sub-optimal solution obtained through a registration method such as intensity-based leads to local structural discontinuities or it can be global where, despite an optimal

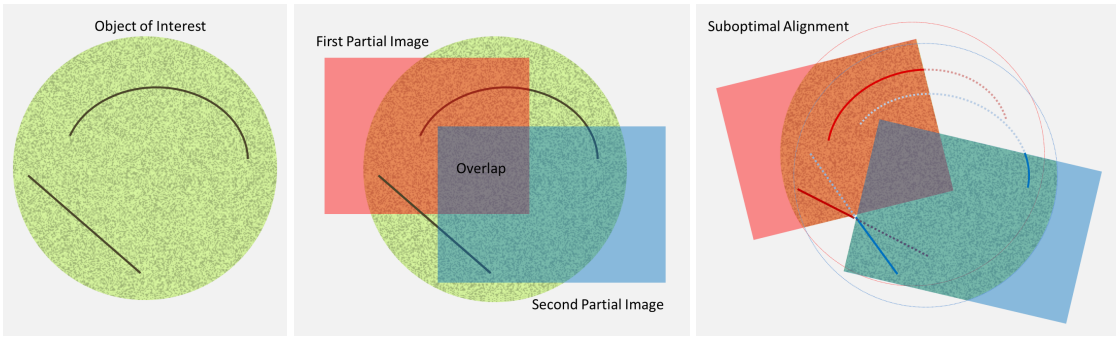
local consistency, global properties of the underlying object of interest is not respected.

In this thesis, we mainly address the problem of field-of-view extension using partially overlapping or non-overlapping images. In particular, we focus on the intensity-based registration as a means of estimating transformations for partial images. We approach the problem of structural consistency among partial images in two aspects. First, we propose a method for addressing local consistency in terms of structural continuity and smoothness. We believe that a locally consistent alignment between partially overlapping or non-overlapping images should respect the local continuity and smoothness of underlying structures. Moreover, in the absence of sufficient structural information, local neighborhoods should be taken into account for augmenting a registration algorithm. To this end, we present a structure propagation technique for inferring structures within the overlap of partial images if an overlap exists or within the extended image regions otherwise. We build upon the vast amount of established research on perceptual organization techniques focusing on modeling of human visual system. Such techniques try to infer structures from unorganized point sets by imposing laws of perceptual organization such as smoothness and continuity. Among other works on perceptual organization, *tensor voting* [Medi 00a] was introduced in the last decades and has proven to be a successful method for inferring structures from noisy, unorganized point sets (c.f. Figure 1.1). Therefore, we build upon the tensor voting framework for inferring structures in or beyond partial images through propagation. Tensor voting based structure propagation works completely model-free and respects the local structural continuity and smoothness during extension.

Secondly, we handle the global consistency between partial images during field of view extension by parametrizing an object of interest in a particular scene in a parametric space. The primary motivation here is the observation that not only the information in the overlap of partial observations should contribute to the registration but also the non-overlapping parts present important cues about the configuration of the overall scene. The vast amount of information in the non-overlapping parts is often not utilized and therefore the registration is constrained only by the information in the overlap leading to suboptimal solutions. Space transformations allow us to make local inference about the global consistency of a particular configuration of partial images independent of the amount of overlap between them. This enables the incorporation of the full amount of available information contained in partial images.

In Figure 1.2, we illustrate a possible image registration scenario where partial images of a certain object of interest are acquired with a small overlap between them. Due to the lack of salient structures in the overlapping region, the solution to the registration problem can be suboptimal and the part of the structures spanning the entire scene might be inconsistent with each other. Based on this observation, we claim that, in this case, a better strategy would be to incorporate the partially visible structures into the registration process which would provide additional regularization in terms of structural consistency.

A side benefit of our approach is the possibility of aligning partial images even without overlap, which is in general not possible otherwise. Our parametrization method is mainly inspired by the model-based object recognition techniques in typical computer vision problems. In particular, we focus on part-based model representations since they combine the appearance of object parts, represented in terms of visual vocabularies, with a geometrical relationship between them. This combination makes a detector robust to occlusions. Since partial image observations can be considered as occluded object parts, part-based object



(a) (b) (c)

FIGURE 1.2.: Illustration of the overlap dependence of intensity-based image registration and the issue of structural consistency. (a) shows an image of a very simple object with two structures visible. In (b), a possible image acquisition scenario is demonstrated. Note the limited size of the overlapping region which might be due to the limitations of the image acquisition protocol. (c) shows a possible alignment based on the information contained only in the overlapping region. The inconsistency in terms of the structures as well as the global shape of the object is illustrated by dashed lines.

models with geometrical layouts seem to be suited for detecting object in partial images.

There are different methods for defining geometrical relationships between object parts. Starting from fully inter-dependent structures as in the constellation model [Ferg 03], there can be looser dependencies as in the star shape model [Leib 04] or even no geometrical constraints as in the bag-of-words model [Csur 04]. While bag-of-words model imposes no geometrical constraints, a full connected model has the disadvantage of being computationally inefficient due to its exhaustive search mechanism for inference. Therefore, we focus on the star shape representations of object categories where parts are connected to a central reference part. In star shape representation, inference is done based on a voting mechanism inspired by the Generalized Hough Transform.

Recent research in the vision community on voting based inference mechanisms established the basic idea for this work. Such methods try to infer the existence of a known object category in a particular scene based on a model learned on object instances from the same category. The original Hough transform [Houg 62] as well as its derivatives such as [Ball 81] constitute a considerable amount of voting based detection approaches. In Hough transform, local features vote for the location or the configuration of a particular shape in the Hough space where the configuration or the location is detected by searching for the maximum in the Hough space. For a strong hypothesis, however, it is important to have a consensus between the votes in the Hough space. This constitutes the basis for our approach in this work for assessing the consistency between partial images. Similarly, it is possible to check the consistency between strongest hypotheses generated by different partial images within the same Hough space.

As the application area of above mentioned concepts, we consider only the field of medical image analysis though the proposed strategies can be employed for solving problems with similar characteristics in other fields such as general computer vision.

CHAPTER
TWO

OUTLINE

In this chapter, we first describe how this thesis is organized. Then, we briefly describe the main contributions made in this dissertation along with the problems addressed.

2.1. Thesis Organization

This thesis is organized as follows. Since the aspect of medical image analysis that we focus in this thesis is image registration, in Chapter 3, we provide a brief background on medical image registration. Then, in Chapters 4 and 5, two main contributions made in the course of this work to the intensity-based registration of partial images for field-of-view extension are presented. We present our conclusions derived from the theoretical and experimental observations from this dissertation in Chapter 6. Then we describe various ways to improve the strength of the methods presented in this thesis. Finally, we conclude with a discussion of several open issues.

2.2. Summary of Contributions

2.2.1. Structure Propagation for Locally Consistent Image Alignment

Problem: Image field-of-view extension by combining two or more partial images is a necessary practice in several applications. The main challenge in such procedures, however, is the estimation of transformation parameters for an optimal alignment between the partial images. Most of the existing image registration techniques optimize a cost function based on an overlap between images. For such methods to work properly, the overlapping region should be sufficiently large and should be salient enough in order to provide necessary information for the correspondence estimation in terms of intensities or features. While these constraints can be easily fulfilled in some applications like consumer photography, such methods are not suited for aligning partial images that have noisy or distorted overlaps or do not have an overlap at all. Such kinds of images are usually encountered in biological imaging where either the modality itself has noisy characteristic (e.g. Ultrasound), inherent distortions (e.g. MRI) or the imaging protocol introduces distortions in the images (e.g. histopathology). Moreover, while an optimal alignment is often sought for by minimizing

intensity difference or feature distances, crucial properties of underlying structures such as consistency or smoothness are often ignored. Without taking such constraints into account, an intensity-based image registration approach is doomed to result in potential suboptimal alignments with local and global structural inconsistencies.

Contribution: [Yigi 13, Muel 14] In this thesis, as our first contribution, we address the issues of missing overlap and try to ensure local structural consistency in intensity-based registration of partial images. In Chapter 4, we approach to both issues by introducing a structure propagation technique for creating salient structural probability maps. These maps are established in regions where no overlap between the images exists or in regions where an overlap is unreliable due to distortions and noise in the images. Structure propagation builds upon the powerful tensor voting framework to infer structures based on salient information in a local neighborhood. Structural probability maps are incorporated into the standard registration frameworks in the form of regularization terms. The proposed method is employed for solving various field-of-view extension problems in medical imaging. An application for mosaicing 2D microscopic images is demonstrated in Section 4.3.1. Section 4.3.2 describes how the reference-free volumetric reconstruction of histopathology sections can benefit from the proposed technique. Finally, an application for creating whole-body MRI is presented in Section 4.3.3.

2.2.2. Hough Space Parametrization for Globally Consistent Image Alignment

Problem: While our previous contribution tries to ensure local consistency through structure propagation, there is still need for methods that take into account the global properties of underlying structures and shapes. Intensity-based image registration methods are mostly based on local correspondences. Therefore, a point in one image does not provide any contribution to the overall registration process unless a corresponding point in a target image is given. Without imposing any constraints on the global configuration of partial images, a registration method might lead to locally consistent but globally inconsistent solutions. Given the fact that the parts of particular scene often provides cues about the contained objects, local information can be exploited for inferring the global consistency when aligning partial observations in the scene.

Contribution: [Yigi 14] As the second contribution of this dissertation, we propose a regularization method that ensures a globally consistent alignment of partial images through intensity-based registration. To this end, we describe a novel Hough space parametrization approach in Chapter 5 which learns an object of interest from previous data and helps to make local inference about the global consistency during image registration. We discuss several ways of how different configurations of partial images can be evaluated for assessing global consistency between them. Measures defined in a parametric space can be seamlessly plugged in as regularization terms into standard registration functionals. In Section 5.3, as an application of the proposed concept, we demonstrate how the proposed global consistency measure can be used for obtaining a wider field-of-view in Ultrasound images.

CHAPTER
THREE

BACKGROUND

In image analysis, extending the field-of-view by combining multiple images require a variety of image processing and analysis steps. The most important task in field-of-view extension is the image registration. Therefore, in this chapter, we attempt to provide a concise background on image registration where we put more emphasis on medical image registration. In what follows, we give an overview of image registration and explain its main components. Finally, while there are numerous image registration approaches proposed in the literature, we describe a particular technique based on discrete optimization which we will employ in the applications presented in this thesis.

3.1. Medical Image Registration

Image registration is a fundamental task in a wide range of image analysis applications including medical image analysis, satellite imagery and computer vision, to name a few. It is defined as the process of finding geometric transformations for two or more images of a particular scene originating from different sources, taken at different times or from varying viewpoints [Zito 03]. Depending on the application, the purpose could be, for instance, to fuse information coming from different sources, to measure certain changes over time or to observe variations in a group. The goal of an image registration method is to find a set of transformation parameters such that a cost function is optimized. In addition to the images being registered, an image registration method is composed of mainly three components: transformation model, (dis)similarity measure and optimizer.

Image registration methods are usually classified based on either the type of components such as transformation models (e.g. deformable registration) or the applications like medical image analysis (e.g. medical image registration). Sometimes, the dimensionality or the modality of images are also used as the basis for classification. Another broad classification base is the parametrization. If a transformation is defined in terms of parameters (e.g. linear transformation), then, it is called parametric registration. Second class which is called non-parametric registration usually refers to deformable registration where a dense deformation field needs to be estimated without using any parametrization. Such methods are often based on a variational framework. We do not consider non-parametric methods in this dissertation, therefore, such methods are not cov-

ered in this overview. We, however, refer the interested reader to the relevant literature [Mode 03, Fisc 04, Verc 09]. In the following, we will focus on the parametric image registration methods. Finally, although we make an attempt to give a concise overview here, for more details, we refer the reader to the excellent review papers and books on image registration [Brow 92, Main 98, Hill 01, Zito 03, Mode 03, Soti 13].

3.1.1. Medical Image Modalities

While there are numerous imaging modalities commonly used in clinical research and practice, we limit ourselves only those which are used in the applications presented in this thesis.

Ultrasound Imaging Mainly due to its speed and cost effectiveness, ultrasound has become one of the most popular imaging modalities. While it is possible to have high sampling rate in time using ultrasound, image analysis tasks for ultrasound still suffer from the low quality of images [Merc 05]. Imaging in ultrasound can be done in 2D as it is traditionally done and also in 3D thanks to the recent developments in ultrasound image acquisition techniques such as the free-hand ultrasound and phased-array transducers [Rank 93, Fens 01]. One of the main challenges of ultrasound in the clinical practice is the difficulty to penetrate through bony structures and air. Another common issue is the limited field-of-view. For this reason, often multiple partial ultrasound images are acquired and combined in order to see a wider region than a transducer normally covers. However, this task is usually challenging mainly due to the low quality of ultrasound images. We will come back to this issue when we discuss globally consistent alignment in Chapter 5.

Magnetic Resonance Imaging Magnetic Resonance Imaging (MRI) provides superb soft tissue contrast with high image resolution in contrast to other imaging modalities. Due to this characteristics of MRI, it is one of the most commonly used imaging modalities in clinical practice. MRI is also free of ionizing radiation which is another advantage over other modalities. However, long acquisition time for MRI, which prevents the acquisition of volumetric data in real time [Sieb 08], still remains as the main challenge for 3D dynamic imaging or 4D imaging. Another common issue in MRI is the geometric distortion artifacts near the boundaries of the image field-of-view which requires special treatment if multiple partial images need to be combined with the purpose of having an extended field-of-view. We will present a solution for this issue in Section 4.3.3. For more details about MRI, we refer the interested reader to [Brow 10].

Computed Tomography Computed Tomography (CT) has been considered as the gold standard imaging modality in many medical applications. Its wide availability and shorter acquisition times than MRI have rendered CT more popular than MRI. On the other hand, it does not provide the level of soft tissue contrast that MRI does. Other advantages of CT include the fact that the information contained in CT images can be directly used in radiotherapy to estimate the amount of dose to be applied to a patient [Dais 03]. However, the issue of ionizing radiation exposed to patients is the main obstacle for the usage of this modality for dynamic 3D image acquisitions.

3.1.2. Medical Image Representation

Medical images are represented as two or more dimensional finite and structured sampling grids where each node of the grid is called a *pixel* (in 2D) or *voxel* (in 3D). Depending on the imaging modality, each pixel (or voxel) contains a certain measurement value regarding a physical point in the body. In projection-based imaging modalities such as X-ray, a pixel represents the integration of attenuation along a ray cast through a patient, whereas in ultrasound a pixel or voxel value stands for echo signal of the ultrasound beam sent through the tissue. In addition to the measurements, medical image data contains also spacing information which tells us the size of the region that a pixel covers. Pixels can be isotropic if the sampling grid is equidistant in each dimension or anisotropic otherwise. Additional information such as the world coordinates of the origin of the image coordinate system as well as its orientation helps when associating a pixel with a spatial point in patient coordinate system.

3.1.3. Need for Registration

Rapid advancements in medical imaging technology have led to the introduction of diverse imaging modalities being used in clinical practice. Some of these modalities are briefly discussed previously. Furthermore, thanks to the increasing availability of digital storage capacities, huge amounts of medical images of different modalities are produced and digitally stored everyday. The analysis of this huge amount of medical imagery for several purposes such as treatment planning or studying disease progression is a crucial task in the medical field. However, such analysis tasks often require a comparison either between images in different modalities (e.g. MRI vs CT) or between the images of single modality taken at different times (e.g. pre-operative vs post-operative) or from different viewpoints.

Besides the challenges of processing this large amount of data, various factors such as tissue deformations, patient movements or variations in imaging setups make a direct comparison between medical images difficult and necessitate an image registration step. Moreover, since these factors are usually unavoidable in most cases, medical image registration has become a core part of almost every medical image analysis procedure. In most medical applications such as the prediction and monitoring of tumor growth, verification of treatments, fusion of images of different modalities, one needs to align images in order to obtain anatomical correspondences between medical images.

3.1.4. Formulation

Image registration aims at the spatial alignment of two or more images in general. However, here, for the sake of simplicity, we will consider the co-registration of two images. Let $I : \mathbf{x}_I \in \Omega_I \mapsto I(\mathbf{x}_I) \in \mathbb{R}$ and $J : \mathbf{x}_J \in \Omega_J \mapsto J(\mathbf{x}_J) \in \mathbb{R}$ be two images to be registered together where $\Omega \in \mathbb{R}^N$ defines a region in space covered by an image with dimensionality N. I is called moving or source image which is being registered and J is called fixed or target image. The goal is then to find a spatial transformation or mapping T for every point $\mathbf{x}_I \in \Omega_I$ such that

$$T : \mathbf{x}_I \mapsto \mathbf{x}_J \Leftrightarrow T(\mathbf{x}_I) = \mathbf{x}_J. \tag{3.1}$$

There are different ways for estimating the transformation T. While it is possible to find a mapping for every point in the image using a variational formulation, here, we only consider parametric approaches where the transformation is approximated with a few number of parameters. In practice, depending on how the transformation is modeled, a set of n parameters $\mathbf{t} = \{t_1, \cdots, t_n\}$ is used to parametrize the transformation T. We denote this parametrization with $T_{\mathbf{t}}$. One class of image registration approaches estimate $\mathbf{t}$ by matching features extracted from both images. Such techniques are called feature-based image registration. The most straightforward way of doing this is to let a medical expert manually annotate some point pairs in the overlapping region of the images. Then, one seeks for the optimal set of parameters using least squares. However, due to the cost and the difficulty of manual annotations, automatic methods have been developed in order to extract anatomically salient landmarks from the images. While feature-based methods are fast, their performance depends often on the availability of salient features and the reliability of feature extraction methods.

The second class of approaches use direct image intensities for estimating the best transformation parameters. For this reason, they are often referred to as intensity-based methods. Depending on the relationship between the intensities of images, a (dis)similarity measure tells the goodness of an alignment given a set of transformation parameters. Usually, an iterative optimization process is involved where, in each optimization step, the similarity measure is evaluated in the overlapping region of the images. Depending of the similarity value, the parameters are adapted (optimized) in order to achieve a better similarity value in the next step. Therefore, intensity-based image registration is often formulated as an energy minimization problem as follows

$$\hat{T_{\mathbf{t}}} = \arg\min_{T_{\mathbf{t}}} \mathcal{E}(I, T_{\mathbf{t}}, J) \tag{3.2}$$

$$= \arg\min_{T_{\mathbf{t}}} \xi(I \circ T_{\mathbf{t}}, J) + \mathcal{R}(T_{\mathbf{t}}) \tag{3.3}$$

where $T_{\mathbf{t}}$ is the parametrization of T by $\mathbf{t}$ and $I \circ T_{\mathbf{t}}$ represents the warping of I by the transformation $T_{\mathbf{t}}$. Since, in practice, I and J have different sampling grids, warping often involves a resampling step where I is resampled in the coordinate frame of J via intensity interpolation. In practice, the inverse mapping T^{-1} is estimated and I is warped backwards mainly due to the advantages of backward warping over forward warping such as the prevention of occurrence of holes and multiple intensity mappings causing blockiness artifacts during resampling [Soti 13].

The energy in Equation (3.3) usually consists of two terms; a data similarity term $\xi(I \circ T_{\mathbf{t}}, J)$ which reflects the goodness of alignment in terms of image based similarity measures and a second term $\mathcal{R}(T_{\mathbf{t}})$ to impose soft constraints on the solution. This term is often referred to as regularization. In case of deformable registration, for instance, regularization of the deformation field might be desirable in order to address the ill-posedness of the registration problem and to obtain anatomically plausible deformations. Prior knowledge about the scene under consideration can also be included in the functional as a regularization term. While regularization terms are often functions of transformations, we will show in Chapters 4 and 5 that regularization terms can be based on the images as well.

The forms of the three main components of image registration, namely transformation model, similarity measure and optimization method vary depending on the kind of the

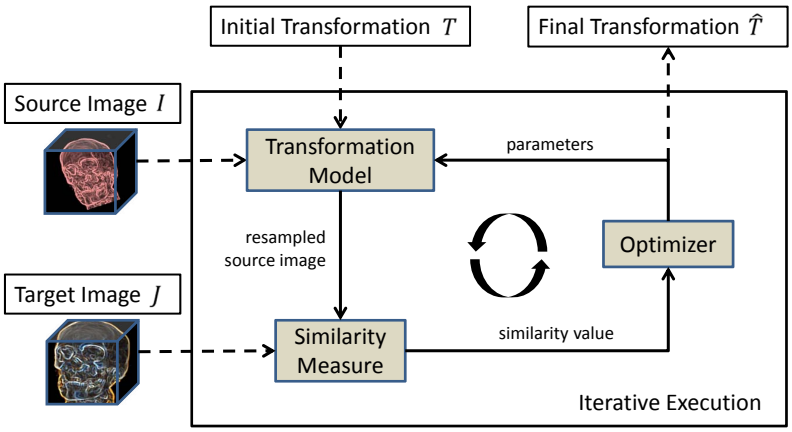


FIGURE 3.1.: Diagram showing the pipeline and the main components of an iterative intensity-based image registration algorithm.

modalities of images being registered and the amount of tissue deformations etc. In the following section, we will discuss different forms of each component briefly.

3.1.5. Registration Components

Figure 3.1 shows a diagram of a standard, iterative intensity-based image registration pipeline. Once source and target images are provided along with an initial guess for the transformation T, the registration iterates until some stopping criteria are met. Stopping criteria used in practice include the maximum number of iterations, minimum relative or absolute changes in the similarity measure as well as parameter values. At the end of registration, the final transformation parameters are used to resample the source image within the coordinate frame of the target image. In the following, we will have a deeper look in each component of the registration pipeline.

Transformation Models

Geometric deformations occurring in images can be represented using different transformation models depending on the type of application. Usually, the characteristics of the scene are considered when deciding for a transformation model. Deformations can be modeled by linear transformations if the objects of interest undergo global deformations. Such models range from simple ones such as rigid transformations where only translational and rotational motions are assumed to affine transformations if there is also scaling and shearing in the scene. Figure 3.2 illustrates the effects of different transformation types on a triangle shape.

The number of parameters to be optimized during registration is often linked to the degrees of freedom in the deformation model. While registration using linear transformation models are usually tractable even with increasing image dimensionality, non-linear registration suffers from the curse of dimensionality with a huge number of transformation parameters where a displacement for each pixel or voxel needs to estimated. In non-linear

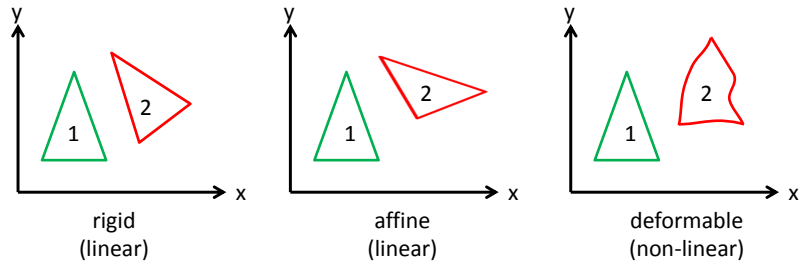


FIGURE 3.2.: Transformation types. While rigid transformation does not change the scale of the objects, affine transformations may change the scale and the aspect ratio. On the other hand, deformable transformations are local and cannot be modeled by a linear equation. Therefore, they are called non-linear transformations.

registration, the goal is to find a transformation for each point $\mathbf{p}$

$$T(\mathbf{p}) = \mathbf{p} + D(\mathbf{p}) \tag{3.4}$$

where $D(\mathbf{p})$ is a displacement field. The number of entries in D is the image size times the image dimension which results in a parameter vector of enormous length. There have been efforts in the literature to overcome the curse of dimensionality through dimensionality reduction and approximations. Such approaches often make use of interpolation theory and try to significantly reduce the number of parameters to be estimated by looking for displacements only at a subset of points in the images. Displacements for the remaining points are approximated via interpolation using different strategies including thin-plate splines [Book 89], free-form deformations (FFD) [Ruec 99] and locally affine models [Piti 06].

FFD using B-spline for interpolation is one of the most popular deformable registration techniques in medical image analysis [Soti 13]. A regular grid of control points is overlayed onto an image and only the displacements for the control points are optimized using the image content in their surroundings. Displacements for the points in between the control points are interpolated using B-splines. The support range of a control point is determined by the order of the B-splines as well as the spacing between the control points. The most commonly used type of B-splines is the cubic one, which produces smoother deformations compared to the lower order ones but comes with an increased computational complexity. Since we will be also using FFD-based transform model for deformable registration in this dissertation, let us define the equations for FFD-based transformation using B-spline interpolation.

Given a uniform grid of m control points with displacements $\mathbf{\Phi} = \{\phi_1, \cdots, \phi_m\}$, the deformation at point $\mathbf{p} = (x, y, z)$ is approximated as

$$T_{\mathbf{\Phi}}(\mathbf{p}) = \mathbf{p} + \sum_{\mathbf{j}=0}^{m} B_{\mathbf{j}}(\mathbf{p})\phi_{\mathbf{j}} \tag{3.5}$$

where $B_{\mathbf{j}}$ is the tensor product of cubic B-spline basis functions for each control point index $\mathbf{j} = \{i, j, k\}$ and $\phi_{\mathbf{j}}$ is the corresponding displacement. $\mathbf{\Phi}$ in the above equation corresponds to $\mathbf{t}$ in Equation (3.3), i.e. the parametrization of transformation $T_{\mathbf{\Phi}}$. Let us further denote the number of control points in 3D with $m_x \times m_y \times m_z$ and assume uniform spacing δ

between the control points. Then, this gives the transformation

$$T_{\mathbf{\Phi}}(\mathbf{p}) = \mathbf{p} + \sum_{a=0}^{3}\sum_{b=0}^{3}\sum_{c=0}^{3} B_a(r)B_b(s)B_c(u)\phi_{i+a,j+b,k+c} \tag{3.6}$$

where $i = \lfloor x/\delta \rfloor - 1$, $j = \lfloor y/\delta \rfloor - 1$, $k = \lfloor z/\delta \rfloor - 1$, $r = x/\delta - \lfloor x/\delta \rfloor$, $s = y/\delta - \lfloor y/\delta \rfloor$, $u = z/\delta - \lfloor z/\delta \rfloor$ and B_0 to B_3 are cubic B-spline basis functions.

The advantage of using B-spline based FFDs over other parametric deformation models include the inherent smoothness providing implicit regularization and local support of control points allowing local computation during optimization.

Similarity Measures

Another important component of intensity-based image registration is the criterion used for the measurement of similarity. Given a transformation between two images, similarity measures tell us the goodness of alignment. There are various measures for comparing images ranging from simple least squares difference to more complicated statistical measures giving the statistical similarity between the intensity distributions in the overlapping regions. The choice of a similarity measure usually depends on the type of images being registered. While difference-based measures are commonly used for the registration of intra-modality images, statistical measures are more suitable for multi-modal registration.

The most common measure of similarity is the sum of squared differences (SSD) between two images which is defined as

$$\xi_{SSD}(I, J) = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega} \left(I(\mathbf{x}) - J(\mathbf{x})\right)^2 \tag{3.7}$$

where Ω is the region of overlap. Normalization of the sum of squared differences makes the measure independent from the size of the overlap. A simpler version SSD, which is more robust to large intensity value differences, is the sum of absolute differences (SAD) defined as

$$\xi_{SAD}(I, J) = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega} |I(\mathbf{x}) - J(\mathbf{x})|. \tag{3.8}$$

SSD and SAD show poor performance if there is a constant multiplicative factor between the image intensities being registered. For such cases, a statistical similarity measure which is called the normalized cross correlation (NCC) is more appropriate where a linear relationship between image intensity values is assumed. NCC can be written as

$$\xi_{NCC}(I, J) = \frac{\sum_{\mathbf{x}\in\Omega}\left(I\left(\mathbf{x}\right) - \bar{I}\right)\left(J\left(\mathbf{x}\right) - \bar{J}\right)}{\sqrt{\sum_{\mathbf{x}\in\Omega}\left(I\left(\mathbf{x}\right) - \bar{I}\right)^2 \sum_{\mathbf{x}\in\Omega}\left(J\left(\mathbf{x}\right) - \bar{J}\right)^2}}, \tag{3.9}$$

where $\bar{I}$ and $\bar{J}$ are the mean image intensity values. NCC simply performs a cross correlation of normalized images which are obtained by subtracting their mean and dividing by their standard deviations. This normalization removes any global additive or multiplicative constants between the image intensities. Although the NCC is suited for within-modality registration scenarios, it is not always applicable to multi-modality registration tasks. For

multi-modal problems, information theoretic approaches such as mutual information (MI) [Viol 97] have been proposed in the literature. As the name already suggests, MI tries to maximize the mutual information contained in both of the images. It is defined as

$$\xi_{MI}(I, J) = h(I) + h(J) - h(I, J), \tag{3.10}$$

$h(\cdot)$ being the image entropy and $h(\cdot, \cdot)$ being the joint entropy. Modified versions of mutual information proposed to reduce its sensitivity to the size of overlap between the images include normalized mutual information [Stud 99] and entropy correlation coefficient [Maes 97].

Note that regardless of whether a similarity measure needs to minimized or maximized, we will consider image registration as a minimization problem. It should be also noted that all the similarity measures discussed here are overlap-based. That is, they are calculated only in the overlapping regions of images being registered. Therefore, they are not suitable for registering images having very small or no overlap. We are aware of only the work of August and Kanade [Augu 05] who suggested to include the intensity histograms of non-overlapping regions into a common joint distribution to avoid overlap dependence.

Optimization Procedures

Image registration is usually posed as an optimization problem where a cost function is optimized with respect to the parameters of the transformation. Depending on the type of the similarity measure, it is either minimized (e.g. SSD, SAD) or maximized (e.g. NCC, MI). The chosen similarity measure is evaluated using the current parameters of the transformation at each iteration of the optimizer. While the number of parameters to be optimized is small for linear image registration, optimization becomes difficult in the case of non-linear registration where the degrees of freedom is often very large. For such cases a reduction of dimensionality of the parameter space is helpful where the deformation can be modeled, for instance, using free form deformations.

Existing optimization methods work either gradient-based where a differentiation of the cost function (i.e. similarity measure) is needed at every iteration or gradient-free or derivative-free where no differentiation is needed but the cost function is evaluated directly. In the latter case, a modification of the parameters is done by considering the change of cost function values between the optimization iterations.

Gradient-based optimizers include simple methods such as gradient descent where only a first order derivative is needed and more complicated ones such as Gauss-Newton which has quadratic convergence. Derivative-free methods include simplex-based optimizers such as the Nelder-Mead Simplex optimizer [Neld 65] and the Powell's Method [Powe 64]. Derivative-free optimizers are good when the cost function is complicated and it is difficult to find its derivative. In Chapter 5, we will use the Nelder-Mead Simplex optimizer which is implemented in the NLopt non-linear optimization package [John].

While optimization in image registration is an inherently continuous problem, methods have been proposed to solve the registration problem in a discrete setting. The main advantages of discrete optimization approaches are the optimization speed and their success in handling non-convex cost functions. While the accuracy is more guaranteed in continuous optimization strategies given a good initialization, the same and even better accuracies can be obtained through discrete optimization with an intelligent discretization strategy. In the

following section, we will focus on a recently proposed image registration technique using discrete optimization through Markov Random Fields formulation [Gloc 08, Ziki 10]. This method constitutes the base for our intensity-based image registration approach discussed in Chapter 4.

3.1.6. Markov Random Field Registration

Markov Random Field (MRF) registration [Gloc 08] treats the estimation of transformation parameters of an image registration problem as discrete optimization. Instead of searching for continuous parameter values $\mathbf{t}$, it quantizes the parameter space and converts the optimization problem into a labeling procedure. Regardless of the deformation model, a range of possible transformation parameter values are assigned labels and the goal is set as to find the best label configuration that leads to the optimal cost function value. Once the task of parameter estimation is seen as a labeling problem, it can be expressed in terms of a discrete MRF syntax [Gloc 11].

In MRF-based formulation, parameters can be seen as the nodes (random variables) of an undirected graph $\mathcal{G} = (\mathcal{V}, E)$ where discrete labels $\mathbf{l} = \{l \mid l \in \mathbb{R}\}$ chosen from predefined parameter ranges are assigned to each node $\mathbf{p} \in \mathcal{V}$. The energy of a first order MRF is defined as

$$\mathcal{E}(\mathbf{l}) = \sum_{\mathbf{p} \in \mathcal{V}} V_{\mathbf{p}}(l_{\mathbf{p}}) + \sum_{(\mathbf{p},\mathbf{q}) \in \mathcal{N}} V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}) \tag{3.11}$$

where $\mathcal{N}$ is a neighborhood system and $l_{\mathbf{p}}$ is a labeling of the node $\mathbf{p}$. The unary term $V_{\mathbf{p}}(l_{\mathbf{p}})$ corresponds to the cost of assigning label $l_{\mathbf{p}}$ to the vertex $\mathbf{p}$ whereas the pairwise term $V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}})$ represents the cost of assigning two labels to the two neighboring nodes simultaneously reflecting the dependency between the parameters.

The MRF energy defined in Equation (3.11) can be solved using efficient discrete MRF optimization algorithms such as Fast primal-dual (FastPD) [Komo 07, Komo 08] or quadratic pseudo-boolean optimization (QPBO) [Kolm 07]. In terms of registration, the forms of unary and pairwise terms depend on the task at hand. In case of deformable registration, for instance, the similarity measure is used for the unary potential while deformation regularization replaces the pairwise term.

In the following two sections, we will give brief overviews for two recent methods for deformable [Gloc 08] and linear [Ziki 10] intensity-based registration. Both methods are based on a discrete Markov Random Field (MRF) formulation. Note that the following two overview sections heavily depend on [Gloc 08, Ziki 10, Gloc 11]. For a more detailed discussion, however, we refer the reader to these excellent papers.

Deformable Registration

Estimation of transformation parameters for deformable image registration can be solved using discrete MRF formulation. But, how is this done? As discussed previously, in order to cast the parameter estimation as a labeling problem, parameter space has to be first quantized and labels need to assigned. To explain this, let $\mathbf{d_p} = D(\mathbf{p})$ be the displacement vector for a point $\mathbf{p}$ (c.f. Section 3.1.5 for the definition of the displacement field D). While a different parameter range quantization for each parameter is possible, here, we use the same range for all parameters. Once a quantization is done, possible range of values for $\mathbf{d}$

are assigned labels l. This means that the assignment of a label $l_{\mathbf{p}}$ to the point $\mathbf{p}$ results in a displacement $\mathbf{d}_{l_{\mathbf{p}}}$ at $\mathbf{p}$. Using this notation, optimal displacement field D can be obtained by solving the dual problem of finding optimal labeling of points in the image space [Gloc 08].

In practice, instead of trying to assign labels to every image point, often, a dimensionality reduction is performed in order to reduce the computational complexity [Gloc 11]. Using the FFD-based transformation model introduced in Section 3.1.5, one aims at finding displacement vectors $\mathbf{\Phi}$ for a subset of points defined by the control point grid. Assuming $\mathbf{\Phi} = \mathbf{0}$ in the beginning, at each iteration of registration, displacement vector updates $\Delta\mathbf{\Phi} = \{\Delta\phi_1, \cdots, \Delta\phi_n\}$, which are considered as random variables, are calculated via discrete optimization. In fact, labels are not assigned to the complete displacement vectors but to their updates. The complete dense displacement field D can be calculated using B-spline interpolation as discussed in in Section 3.1.5. The neighborhood system used for the random field is similar to the FFD topology. Usually, 4-connected neighborhood in 2D and 8-connected neighborhood in 3D are used.

An interesting aspect of discrete MRF registration using FFD-based transformation model is the two-fold dimensionality reduction. In addition to the significant simplification thanks to the FFD, discrete labeling assigns integer labels to the high dimensional displacement vectors. Without labeling, one has to consider each component of a displacement vector as a single parameter, whereas labeling considers all components of a displacement vector as a single parameter.

After defining the random variables of the MRF formulation which corresponds to the control point displacement updates, a strategy is needed to define the parameter search space and the way how the discretization is performed. Glocker et al. [Gloc 08] suggests different sampling patterns for assigning labels to continuous displacement update vectors. The simplest strategy is to uniformly sample displacements along the main coordinate axes. A better and suggested way is to sample not only along the main axes directions but also along diagonal directions. Finally, a dense sampling of the search space can also be used but with the additional cost of increasing label vector size. Usually, a maximum displacement magnitude $\|\mathbf{d}_{max}\|$ is determined and s samples are taken along each direction. Maximum displacement magnitude is rescaled by a factor $0 < f < 1$ if a certain sampling does not lead to a decrease in energy. This also leads to a better accuracy which cannot be achieved if a very sparse sampling is utilized. It is also suggested that the maximum displacement is linked to spacing between the FFD control points by allowing a displacement at most 40% of the inter-grid spacing. This ensures that the final transformation is diffeomorphic [Gloc 11].

In every iteration step of the optimization, a discrete MRF labeling is solved. This requires the calculation of the MRF potential functions given in Equation (3.11) which represents Equation (3.3). While we briefly discuss the unary term here, we leave the discussion of the pairwise term to Section 3.2 where we will focus on regularization. The unary term $V_{\mathbf{p}}(l_{\mathbf{p}})$ corresponds to a similarity measure such as the ones mentioned in Section 3.1.5. However, it is computed only locally around each control point given a certain displacement label. Displacement is applied patch-based where all the points in the patch around the control point are moved using the same displacement vector. This results in an approximation of the actual deformation by ignoring the contributions from the neighboring control points. But, it also leads to an efficient computation of a look-up table to be used while searching for the optimal labeling.

Linear Registration

While discrete MRF registration was originally proposed for deformable registration in [Gloc 08], an extension to linear registration was later presented by Zikic and Glocker et al. [Ziki 10]. When seen from a parameter optimization point of view, both linear and non-linear image registration methods can be treated similarly. However, there are big differences if one wishes to apply the same discrete optimization strategy for both. While parameters of deformable registration have only local impact and thus local dependencies exist between the random variables, linear registration parameters are globally coupled by construction due to their global influence [Ziki 10]. In the following, we will briefly describe the suggested strategy to enable discrete MRF optimization also for linear image registration.

Following the same notation used in Section 3.1.4, let $\mathbf{t} = \{t_1, \cdots, t_n\}$ be a set of linear transformation parameters. In [Ziki 10], the parameter set $\mathbf{t}$ consists of translation, rotation and shearing components leading to only 6 degrees of freedom in 2D or 12 degrees of freedom in 3D. Again, parameters are treated as random variables and an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{C}_2)$ is defined where the nodes $\mathcal{V}$ represent the random variables and $\mathcal{C}_2$ represent the second order cliques of nodes.

Although a linear transformation with n parameters (i.e. degrees of freedom) requires n^{th} order MRF model, an approximation scheme was proposed in [Ziki 10], where only pairwise dependencies are considered between the parameters and complete dependency is established by taking all possible pairs simultaneously into consideration. Thanks to this approximation, the registration energy $\mathcal{E}$ in Equation (3.3) can be represented by a second order MRF model instead of an n^{th} order one if a fully connected graph was considered.

The approximated energy is written in terms of only pairwise potentials as

$$\widetilde{\mathcal{E}}(\mathbf{l}) = \frac{1}{|\mathcal{C}_2|} \sum_{\{t_i, t_j\} \in \mathcal{C}_2} \widetilde{V}_{ij}(l_{t_i}, l_{t_j}) \tag{3.12}$$

where $\widetilde{V}_{ij}(l_{t_i}, l_{t_j}) = \xi(I \circ T_{\mathbf{t}^{ij}}, J)$ represents the cost of simultaneous labeling of t_i and t_j, while keeping other variables fixed. Here, $\mathbf{t}^{ij}$ is a parameter set where only t_i and t_j are varied.

Again, the discrete nature of the MRF model requires that each parameter variable t_i takes a value from a set of discrete labels $\mathcal{L}_i = \{l \mid l \in \mathbb{R}\}$. While we assumed the same parameter space discretization for all random variables in deformable registration, here, mainly because of the different nature of linear transformation parameters, each random variable has different set of label space $\mathcal{L}_i$. The number of labels $|\mathcal{L}_i|$ for each parameter variable t_i is a free parameter, which is used to discretize a search range for that parameter by uniformly sampling a given interval $\left[t_i^{min}, t_i^{max}\right]$ about its initial value. This interval is refined in every iteration in order to achieve a better accuracy similar to the deformable case. We will discuss the specific design choices for each type of registration in the contribution chapters.

3.2. Regularization in Medical Image Registration

There are several challenges associated with medical image registration. One of the common problems we have discussed in the previous sections was the difficulty of optimization due to the dimensionality of the parameter space. Dimensionality reduction techniques using interpolation methods and parametric approaches were already mentioned in Section 3.1.5.

Another important and common issue, especially in deformable registration, is the regularization. Deformable registration seeks to find a high dimensional vector displacement for every image point by using often scalar, one dimensional, image intensity values. This leads to an ill-posed problem where the number of unknowns is larger than the number of measurements [Mode 03]. In order to overcome this issue, additional constraints have to be included into the registration. Such constraints render the registration well posed by imposing conditions like smoothness of the deformation field to penalize deformations that are physically undesired or unlikely. Other constraints include inverse consistency where the registration should still be valid when the order of the images being registered is changed and the obtained transformation is inverted. Task specific constraints such as the rigidity of underlying tissue can be also included in order to allow only anatomically plausible transformations [Star 07]. The regularization term $\mathcal{R}(T_{\mathbf{t}})$ in Equation (4.28) is often designed to achieve the previously mentioned goals. If the regularization is included in the functional via a weighting factor, then, it becomes a soft constraint. In this dissertation, we will only use soft constraints.

Typical regularization terms used in deformable image registration include diffusion regularization

$$\mathcal{R}(D) = \sum_{i=1}^{N} \int_{\Omega} \|\nabla D_i(\mathbf{p})\|_2^2 \, d\mathbf{p} \tag{3.13}$$

which is a first order regularization where ∇ is the gradient operator and curvature regularization

$$\mathcal{R}(D) = \sum_{i=1}^{N} \int_{\Omega} (\Delta D_i(\mathbf{p}))^2 d\mathbf{p} \tag{3.14}$$

which is a second order regularization [Fisc 04] where Δ is the Laplace operator. Please note that the argument to the regularization term is the displacement field rather than the transformation since regularization operates directly on the displacement field. A nice property of curvature regularization is its invariance to affine deformation components. This way, global motion is implicitly compensated without being penalized. It should be noted that these regularization strategies also apply to the discrete MRF registration where the pairwise potential take the role of a regularizer. This time, the above equations are approximated by their discrete versions in terms of labels. Special attention has to be paid if the employed regularization term makes the MRF energy non-submodular where an appropriate discrete optimization scheme has to be chosen.

The amount of regularization during registration can be controlled by a locally varying weighting function in order introduce rigidity penalty. Such spatially varying weighting functions can also be used to impose tissue specific constraints on the deformation field.

So far, regularization has always been meant for deformation field. However, prior information about the image data or other types of information extracted from local image regions can be incorporated into the registration functional as additional terms. We call such terms data-driven regularization terms. For instance, structural information in the neighborhood of a certain point in the image can be used to infer probabilistic structuredness at that point to result in higher costs if no structure is observed at a certain iteration of registration. In Chapter 4, we will show how such a data-driven term can be added to the registration functional.

So far, regularization has always been about deformable registration where the goal is to convert an ill-posed problem into a well-posed one. Another way of incorporating the observed data into the regularization of the solution space is the utilization of prior data in terms of models [Comm 05, Star 07, Mode 08]. An interesting aspect of this approach is that it can be employed also for linear image registration. Mainly due to the well-posedness of the linear image registration problem, the concept of regularization has not been explored too much for this task. However, there are often cases where images of a scene to be registered have little in common and, even for linear registration, a good initialization is required for the alignment to be successful. Prior knowledge can help in such cases by establishing a base for the registration. This can be interpreted as regularization where a non-convex parameter search space is evolved to convex one thanks to the regularization.

Image mosaicing is a good example of application of image registration where the lack of a sufficiently informative overlapping region hinders the alignment and the chances of getting trapped in a local optimum is high. In such scenarios, additional information contained in non-overlapping regions of images can be utilized to regularize the solution space. Moreover, any prior knowledge about the object of interest can be incorporated into the registration process as regularizer. Such prior knowledge can be in the form of a shape model, appearance model or even geometrical heuristics about the structures contained in the images. For instance, the interest could be to have no structural discontinuity after registration. An alternative constraint is to force parametric shapes contained in the partial images to be consistent with each other after an alignment. In Chapter 5, we will show how prior information can be incorporated in the registration functional to act as a regularizer for ensuring the consistency of an alignment with respect to the prior.

Contributions

CHAPTER

FOUR

STRUCTURE PROPAGATION FOR LOCAL STRUCTURAL CONSISTENCY

Field-of-view extension is a commonly used technique in many medical imaging applications where partial observations of a certain scene are stitched or stacked together in order to obtain a larger field-of-view. The step of stitching or stacking usually requires a registration between partial images either in their overlapping region or at their interface. This registration task becomes often challenging when the information shared by partial images in their overlapping region is absent or small. While it is not possible to perform an alignment without overlap using existing registration techniques, imaging artifacts such as noise or geometrical distortions towards image boundaries present further complications during registration by decreasing the reliability of available information. Even in the presence of an overlap, without taking such issues into consideration, a registration approach might result in a suboptimal solution and therefore create inconsistencies such as the violation of the continuity and the smoothness of structures across partial observations.

To alleviate these issues, in this chapter, we propose a novel registration approach for the registration of partial images in such challenging scenarios. By employing a perceptual grouping approach, we extend partially observed images beyond their acquisition boundaries. This is done by propagating available structures in order to create structural probability maps in the extended regions representing the possible extensions of observed structures. These structural maps are then used either to establish correspondences between partial images when the shared information is absent, small or unreliable or they can be employed in a data-based regularization term in addition to the existing correspondences to incorporate structural constraints on the solution. Our approach ensures that the structures across sub-image boundaries maintain their inherent smoothness and continuity properties. Furthermore, since only structures are used for propagation, the proposed method can also be used for the stitching or stacking of multi-modal partial images.

In the following, we first provide a brief introduction to the problems addressed here along with the shortcomings of existing approaches. Then, a detailed description of the *structure propagation* method is given where structural probability maps are formulated using the tensor voting framework. Finally, the effectiveness of our method is demonstrated in linear and non-linear registration settings through experiments on synthetic and real medical images.

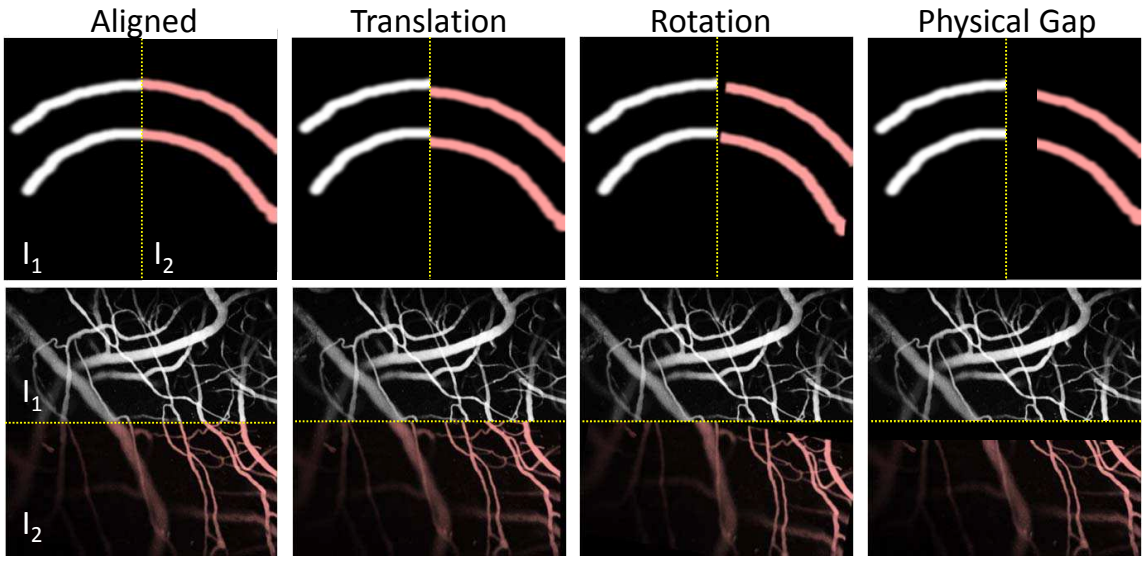


FIGURE 4.1.: Illustration of the possible challenging scenarios in partial image alignment. Initially aligned subimage pairs on the left can be misaligned by translational or rotational transformations applied to one of the subimages. In this case, it is very difficult to restore the applied transformations without having an overlapping region to be used as a base for the alignment. It is much more challenging when there is an occluded region between the partial observations.

4.1. Introduction and Related Work

Image field-of-view extension is the process of combining two or more images in order to obtain an extended field-of-view (FOV). The simplest form of extension is in lateral dimension where 2D images are combined together to get a larger 2D image. The extension can also be in the form of 3D reconstruction where several 2D images are stacked along their axial dimension in order to extend the FOV in that direction yielding a 3D image. The most complex form is the combination of several 3D images with the purpose of extending the visible area in all possible directions.

Image FOV extension has many application areas ranging from computational photography to computer vision to medical imaging. Especially in the field of medical imaging, there has been an increasing interest in having larger FOV images. 3D histology reconstruction [Lee 08, Cape 09, Bagc 10, Feue 11, Saal 12], extended FOV microscopy [Verc 06, Chow 06, Prei 09, Tsai 11, Loew 11], laparoscopy [Moun 09], endoscopy [Atas 08, Totz 12, Warr 12], 3D ultrasound mosaicing [Wach 07, Brat 11], and whole-body Magnetic Resonance Imaging (MRI) [Wach 08, Dzyu 13] are some of the applications where an extended FOV is created by stitching subimages or subvolumes or stacking lower dimensional images. Having a large FOV is especially important in microscopy images since it, for example, enables studying several questions that arise in neuroscience [Tsai 11] and therefore having a high resolution images with a wide FOV is crucial. Moreover, mosaicing is often a necessity since most of the existing microscopes have a limited view either in lateral (e.g. confocal, multi-photon) or axial (e.g. histology) dimensions.

The main step of FOV extension is the combination of subimages by estimating transfor-

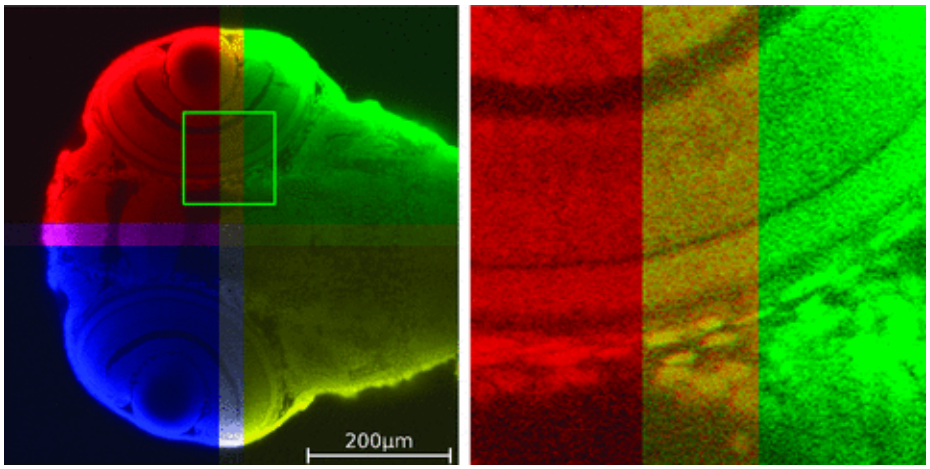


FIGURE 4.2.: Zebrafish embryo stained with TOTO3 to visualize all cell nuclei. Middle section of a confocal stack, gamma corrected ($\gamma = 0.3$). For quality control each tile is colored differently. The homogeneous mix color in the overlapping regions indicates a good placement of the tiles. (Figure and caption reprinted from [Emme 09], ©2009 The Authors Journal compilation ©2009 The Royal Microscopical Society)

mations that will bring them into a spatial alignment in a common coordinate space (c.f. Figure 4.1). This is often done by using an image registration algorithm. However, classical image registration techniques usually assume minor differences in the viewpoints of images to be registered. On the other hand, this is not the case in microscopic imaging such as confocal microscopy where a large FOV is obtained by stitching subimages each covering a limited portion of the sample [Tsai 11]. In such settings, the amount of overlap between neighboring sub-images is often small in size and contains distortions. Figure 4.2 shows an example of such stitching scenario where the overlap regions are relatively smaller than the non-overlapping image regions.

On the other hand, in 3D digital pathology, in order to have a better analysis in 3D, several histology slices are stacked along the axial direction in order to reconstruct a 3D histology volume [Feue 11, Bagc 10, Saal 12]. The challenges for the reconstruction, however, are the artifacts caused by the sectioning process. Besides having no physical overlap between adjacent slices due to the characteristics of slice preparation procedure, artifacts such as tissue tears or folds reduce the amount of reliable correspondence between neighboring slices. Finally, whole-body MRI for extended FOV has similar challenges where partially overlapping MR images are deformably stitched in their overlapping regions. The main challenge, however, is geometrical distortions in overlapping regions making an anatomically correct stitching difficult. Therefore, special registration techniques are needed for extending the field-of-view in either lateral or axial directions. Such techniques should be able to handle 1) very small or distorted overlap between subimages or subvolumes (as in the confocal microscopy or whole-body MRI cases), 2) adjacent images having no overlap at all (maybe just touching each other as in 3D histology case).

In the conventional image stitching, the amount of overlap between subimages to be stitched is important for finding correspondences for the estimation of necessary transformation parameters through image registration. Usually, a predefined value for the

amount of overlap is used to decide whether a stitching between two subimages is feasible [Cape 09, Kayn 10, Tsai 11, Khur 11]. Besides not being able to extract sufficient features (if a feature-based registration is used) or intensity correspondences, the *continuity and smoothness of structures* across subimages cannot be guaranteed unless a sufficient amount of overlap is present. Moreover, in particular in the case of deformable FOV extension, even if sufficient overlap is present, the registration is done only in the overlapping region, guaranteeing alignment in the overlap but leaving the gross natural alignment of distant structures. Having a consistent alignment in such cases also between distant structures is essential especially in microscopic imaging since the subsequent measurements on the structures will also be influenced by inconsistencies introduced during the alignment procedure.

To motivate the problem further, let us assume that one of the initially aligned subimages in Figure 4.1, which are simply two pieces obtained by cutting a whole image into two, undergoes some misalignments such as translational or rotational transformations. It is very challenging to bring these two pieces into a spatial alignment again even in the case of a simple translational misalignment without having an overlapping region. This obvious issue becomes even more challenging and complicated to solve when there is also a physical gap between the pieces, in other words, when there is an occluded region between the pieces as illustrated in the last column of Figure 4.1. Although these images are in 2D, one could imagine the same scenario in 3D where partial images are either 3D subimages or 2D slices to be stacked together.

Since an overlap between the subimages cannot always be guaranteed, there have been efforts in the literature to address the overlap dependence. One scheme is, assuming that there is a shared border between the subimages, which holds for 3D digital pathology, to perform contour matching after extracting contours at image boundaries [Kong 01, Tsam 10]. Although this approach works on puzzle solving tasks, where partial images usually have clean shared borders, it fails when applied on medical images where boundaries usually have low signal to noise ratios or geometric distortions. Another approach, which is often used for 3D digital pathology, is the use of slices at subvolume boundaries. Subvolumes are created by stacking together a small set of 2D slices by assuming minor structural differences. Using boundary slices of such subvolumes directly as done in [Cape 09, Berl 11], however, the final reconstruction might be sensitive to potential distortions in boundary slices [Lee 08]. Based on this observation, Bagci et al. [Bagc 10] proposed to select *best reference slices* from each subvolume, which are then registered together assuming that they are anatomically similar. Although this seems to be a better strategy than using only boundary slices, the difficulty associated with the selection of a "good" representative slice as well as the potential anatomical variations due to the distance or physical gap between the selected slices are the main challenges associated with this approach.

All of the techniques mentioned so far rely on the information contained either at the boundary or in a specific slice of a subvolume. *However, structures in medical data often have certain morphological smoothness and continuity properties which can be exploited to ensure that a final stitching is consistent with respect to the continuity of structures spanning the stitched subimages.* Following this idea, Lee & Bajcsy [Lee 08] proposed a feature-based approach for the volumetric reconstruction from the confocal laser scanning microscopy subvolumes. Feature trajectories extracted from vessel centerlines in subvolumes are used for fitting

polynomial curves that are fused with a corresponding trajectory in a neighboring subvolume. However, they search for an affine transformation only in the lateral plane while ignoring transformations in the stitching direction. Moreover, their method heavily depends on the extraction of centerlines through vessel segmentation. It is usually hard to find feature correspondences and segmentation is often an error prone task.

Reconstruction of histology volumes are even more difficult when the sections or the slices are registered separately. The main difficulty comes from the fact that all slices undergo a cutting process which introduces deformations, foldings or tears, therefore, there is no "target slice" to which a slices can be registered. Therefore, when the slice are simply registered to each other (e.g. consecutively), there is no guarantee that the original shapes of the structures will be preserved in the final volume. This is known as the classical "banana problem" [Stre 97]. A possible solution for this issue is to use external reference images such as block-face images [Bard 02, Feue 11]. Block-face images are simply the photographs of sections taken before cutting. In this setting, each slice is registered to both its block-face image and its neighboring slice. However, the main limitation of such methods is the difficulty related to the acquisition setup and requirement for multi-modal registration.

To alleviate the above-mentioned issue, there have been also reference-free histology reconstruction attempts in the literature where no external references are required. Gaffling et al. [Gaff 11] introduced a reference-free method that uses the regression of manually extracted landmarks to restrict the deformation. By using a polynomial regression over corresponding landmark positions, they obtain a smooth and consistent reconstruction of histology slices. Although the results of this method seem promising, given the large size of histology slices, manual extraction of landmarks is not feasible in practice. Cifor et al. [Cifo 11] proposed a smoothness-guided approach where a min-max curvature flow only in lateral directions is calculated. Once the in-place flows are computed, transformations of different complexity can be extracted from the flow. This method, however, requires the extraction of surfaces in order to perform a curvature flow. Saalfeld et al. [Saal 12] proposed a hybrid reconstruction method where an initial feature-based alignment is followed by block-matching. Deformation of a slice is coupled with the deformations of neighboring slices using a spring model. Gaffling et al. [Gaff 14] proposed a very interesting approach where they separate anatomical information which has low frequency from the high frequency errors due to deformations.

Based on these observations, we believe that there is still no promising solution to the problem of registering/stitching images/volumes in the absence of sufficient or reliable overlap. *Ensuring morphological continuity and local consistency of structures has remained a challenging task in various applications in medical imaging including digital pathology and wide-field microscopy.* Thus, in this chapter, we aim at proposing novel registration techniques addressing the local structural consistency in medical image registration.

The main difference between the presented scheme and the other existing solutions for extending image field-of-view is that a sufficient overlap region is not assumed to be *readily available* from the outset. It is enough for our method if the subimages are "sufficiently close" to each other. Furthermore, we assume that their relative positions are roughly known. That is to say, we know the adjacency relationship between the subimages in advance. Such challenges associated with medical image mosaicing like the initial positioning of partial images by identifying adjacency relationships or avoiding error accumulations

during mosaicing have been extensively studied and addressed in several recent works [Brow 07, Emme 09, Prei 09, Tsai 11, Loew 11]. Moreover, most acquisition protocols record the physical coordinates and extents of images being acquired that can be used for rough initialization. *Therefore, in this work, we will focus on strategies for the registration of subimages having limited or no overlap while ensuring the continuity and the smoothness of structures crossing their boundaries.* It should be noted that, for a successful registration, it is important that image structures have inherent continuity, which is a property that can be observed and exploited in most medical images.

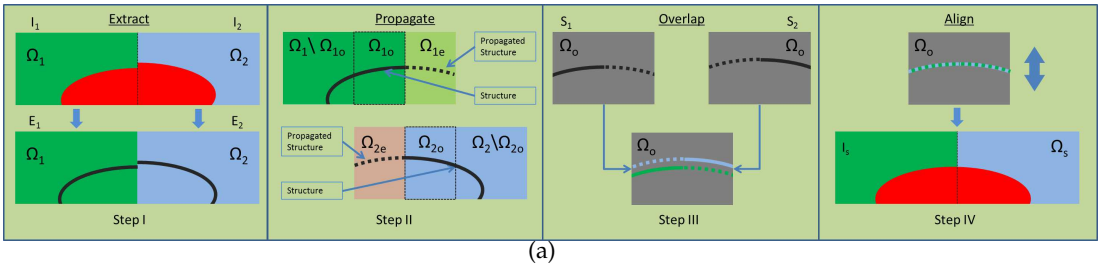
In this work, we are inspired mostly by the ability of the human visual system to perceive the "good continuity" of structures by integrating information from pieces. It has been reported in the literature that the human visual system can construct a whole from its pieces by integrating the continuity information of apparent contours within each piece, i.e. by perceptually grouping them based on the Gestalt law of good continuation [Wert 38]. The principle of good continuity states that, when perceiving a scene, the human visual system tends to prefer a smooth continuity of pieces rather than allowing sharp changes leading to high curvatures. Similarly, we believe that a computer aided solution to the problem of image FOV extension should mimic this property of the human visual system by reconstructing a scene from its partial observations in such a way that, at the end, structures are perceptually smooth and continuous.

Recently, there has been extensive research in the field of computer vision to understand and model the perceptual grouping mechanism of the human visual system [Pare 89, Fiel 93, Augu 03, Sing 07]. Several methods for modeling the perceptual grouping have been developed for curve and surface inference in computer vision [Shas 88, Sark 93, Thor 95, Medi 00a, Mord 06]. Among these methods, tensor voting [Medi 00a] has been employed in numerous applications [Tong 04, Fran 06, King 08, Jia 08, Loss 11]. Besides the problem of curve and surface inference, it has been also used for solving many other computer vision problems such as image repairing [Jia 03], color correction [Jia 05] and terrain extraction [King 08]. More interestingly, several relatively recent works in the field of medical imaging have successfully employed tensor voting in applications such as gap filling for vascular structures [Riss 08], catheter detection in fluoroscopic images [Fran 06] and detection of curvilinear structures in microscopic images [Loss 11]. These works also exploit the continuity property of structures in medical images to reconstruct or detect objects.

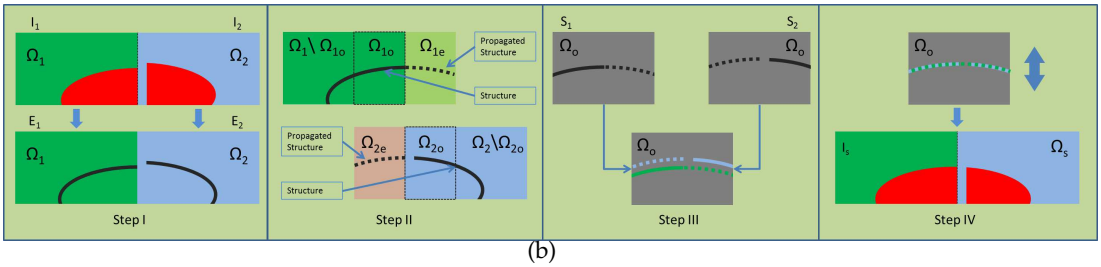
In this work, we propose a novel technique that uses tensor voting for the inference of structures beyond observed image regions. The goal here is to establish a region shared by the subimages to be stitched as demonstrated in Figure 4.3(a). To this end, we create structural probability maps by propagating salient structures from observed image regions into other regions. Then, these structural probability maps are used for the subsequent alignment of subimages. Finally, resulting transformations are transferred to the original subimages for the optimal alignment with respect to the smoothness and the continuity of structures across subimage boundaries. So far, we are not aware of the use of a conceptual grouping technique for extending the field-of-views of medical images.

The same technique can be used also for inferring existing structures based on the available information in the surroundings. Such an approach is quite useful for scenarios where the available measurement is corrupted and there is no ground truth for the underlying shape, which is often the case in histology images. When aligning digital histology slices

with the purpose of having a digital histology volume reconstruction, artifacts due to the sectioning process renders information within each slice unreliable to be used as basis for registration. However, we will demonstrate that, using the proposed structure propagation technique, it is possible to infer the reliability of existing structures by letting their neighborhoods vote for their existence. This voting process results in a map of structural probability that can be used as a self-reference during reconstruction.



(a)



(b)

FIGURE 4.3.: (a) Overview of the proposed four-step stitching method using structure propagation. Here, we use a simple example where two pieces of a partial ellipse are to be stitched together. In Step I, the edge images, E_i are obtained by extracting structures , i.e. the edges, from the input images, I_i. This is followed by the propagation of structures into a predefined region Ω_{ie} for E_i in Step II. We create structure images denoted by S_i in Ω_o, which is constructed by combining the extended, Ω_{ie}, and the overlapped, Ω_{io}, regions of the edge image E_i in Step III. Finally, these newly created images are registered to each other followed by the application of resulting transformations to the original images, I_i in Step IV. (b) The same technique can be directly applied to the stitching of images having a physical gap (see missing part in I_2). The proposed technique is capable of aligning image even in this case thanks to the fact that the structures are extrapolated by making sure that the global continuity of structures is not violated.

In summary, there are several benefits of using structure propagation;

- First of all, in the absence of an overlap region, structural probability maps can serve as a basis for the estimation of necessary transformation that will bring subimages into a spatial alignment.

- Secondly, if an existing overlap region is small or has severe geometrical distortions or deformations arising from slide preparations etc., then, this region can be supported by structural probability maps created by extrapolating the structures that are salient or that present with no or less distortion into this region.

- Finally, an image stitched together from subimages using structural probability maps will have smooth connections between the corresponding structures of subimages.

Once structural probability maps are created, then, either they can be registered to each other directly or one can employ them in a regularization term while registering partial observations with independent measurements. The transformations needed to combine partial observations can be linear or non-linear depending on the needs of applications. For instance, while a linear transformation is sufficient to register partial images for microscopic image mosaicing of a fixed sample, in 3D digital pathology, one has to correct for local deformations which requires a non-linear transformation model. Here, we will demonstrate how both types of transformations can be estimated using structural probability maps obtained through structure propagation. For both types of registration, we use a registration framework based on Markov Random Fields (MRF) [Gloc 08]. For linear registration, we employ the method proposed in [Ziki 10] whereas for non-linear (deformable) registration we base our work on the registration method proposed in [Gloc 08]. A brief review of these methods was presented in Section 3.1.6.

The organization of this chapter is as follows. In Section 4.2, we explain our method where we give the theory of tensor voting in Section 4.2.2 and describe how it is used for structure propagation in Section 4.2.3. Alignment of partial observations with the support of structural probability maps in several applications is discussed in Section 4.3.1. Experimental results demonstrating the performance and the effectiveness of the proposed method on synthetic and real medical images are presented in Section 4.4. Sections 4.5 and 4.6 conclude the chapter with discussions.

4.2. Method

The main idea behind our approach is the extension or propagation of information available in partial observations or subimages in order to allow registration when it is not possible otherwise. A secondary motivation is to improve registration performance by augmenting the existing registration frameworks with data-based regularization terms. The extension is in terms of image structures such as strong edges in the subimages that are extrapolated beyond subimage boundaries. By posing this as an inference problem, we employ the tensor voting method to infer structures based on the orientation and the saliency properties of existing structures in the neighborhood. Inferred structures can be then used in estimating transformations at different scales for the optimal alignment between subimages.

In this section, we first give an overview of our approach followed by theoretical background on tensor voting and then we explain how structures are propagated using tensor voting. Finally, the complete registration strategy is discussed at the end of this section.

4.2.1. Overview

Let $I_1 : \Omega_1 \subset \mathbb{R}^N \rightarrow \mathbb{R}$ and $I_2 : \Omega_2 \subset \mathbb{R}^N \rightarrow \mathbb{R}$ be two images with dimensionality N to be stitched together in order to extend the field-of-view. In this work, we use strong edges as structural information to be propagated. Therefore, we further define $E_1 = \zeta(I_2) : \Omega_1 \subset \mathbb{R}^N \rightarrow \mathbb{R}$ and $E_2 = \zeta(I_2) : \Omega_2 \subset \mathbb{R}^N \rightarrow \mathbb{R}$ as the structural representations where $\zeta(\cdot)$ is an edge operator. Strong edges in images can be detected by using even the simplest methods

like calculating the gradient magnitude and applying a threshold on it. We have chosen to use an efficient implementation of edge detection with recursive filtering [Mong 91][1]. It should be noted that no post processing such as thinning is necessary for our method to work.

In order to be able to register I_1 and I_2, one has to have a rough initial idea about their positions relative to each other. In classical registration, based on the assumption of minor viewpoint differences, the images are usually placed such that a large overlap is guaranteed. The capability of a registration algorithm to refine the relative positions of images often depends on the employed correspondence estimation techniques. The prior positional information can be obtained in several ways such as using a step motor in microscopic imaging or using other existing methods as mentioned in Section 4.1. Moreover, in most image acquisition protocols, image coordinates are saved in headers. Therefore, in this work, we will assume that such prior information is available. In other words, we will assume that we know the rough adjacency relationship between the subimages. Note that the goal here is to find the necessary transformation parameters that will bring these two subimages into spatial alignment such that the structures in the subimages are smooth and continuous across the boundary between them.

Let Ω_{1e} and Ω_{2e} be the regions corresponding to the extensions of image regions Ω_1 and Ω_2, respectively. Using structure propagation, a scalar structural probability map in the extended region Ω_{ie} is created from the structures available in the edge image E_i. Ω_{ie} is now assumed to be overlapping with a small, if not entire, portion of region Ω_j ($j \neq i$ and $i, j \in \{1, 2\}$) of the other subimage. This region of overlap will be denoted by Ω_{jo}, i.e. the overlapping portion of Ω_j with the extension from Ω_i. Further we define

$$\Omega_o = \Omega_{io} \cup \Omega_{ie}, \quad i \in \{1, 2\} \tag{4.1}$$

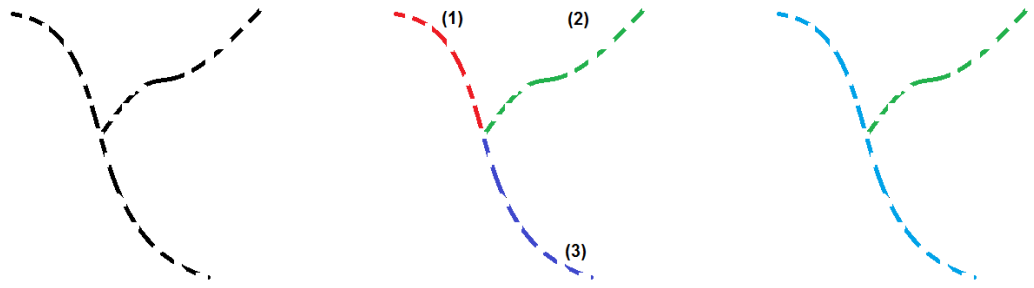
which is simply the union of the overlapped and extended regions for the edge image E_i. We denote the corresponding images with $S_i : \Omega_o \subset \mathbb{R}^N \rightarrow \mathbb{R}$, i.e. the structure images that are constructed by combining the extended, Ω_{ie}, and the overlapped, Ω_{io}, regions of the edge image E_i. This is illustrated graphically in Figure 4.3(a). In this way, we create an overlap region between subimages by means of a two-sided structure propagation. For the optimization of transformation parameters to align the original images I_1 and I_2, the information present in the established overlap region Ω_o can be used. Finally, the stitched image will be denoted by $I_s : \Omega_s \subset \mathbb{R}^N \rightarrow \mathbb{R}$ where Ω_s is the union of the domains of two subimages as shown in the fourth step of Figure 4.3(a).

Note that if an overlap is already available but not reliable enough, then, the overlapping region can be defined as the region into which structures should be propagated. For the registration, however, one could use only the propagated structures or couple it with the unreliable information already contained in the overlapping region. A marginal example for this case is the 3D deformable reconstruction from 2D slices where the overlapping region can be chosen as the entire stack itself. Therefore, one could also infer structures already inside the observed partial images.

[1]Implementation available from `ftp://ftp-sop.inria.fr/epidaure/Softs/Malandain/`

4.2.2. Tensor Voting

The Tensor Voting is a perceptual organization method designed for the inference of perceptually salient structures from a set of incoherent input points [Medi 00b, Medi 04]. In [Medi 04], "perceptual saliency" is defined as *the quality of features to be important, stand out conspicuously, be prominent and attract our attention.* It provides a computational framework based on the Gestalt principles for perceptual organization to extract information regarding the underlying structures. The Gestalt principles describe the properties that the human visual system exploits when organizing input signals into structures [Wert 38]. These principles include proximity, good continuity, closure, simplicity and similarity. We can get some intuition about some of these principles if we consider the fragments in Figure 4.4(a). Using the closure principle alone would lead to three groups while using the good continuity principle alone will result in only two curves. According to the latter principle, for the human visual system, it is more likely to have a curve out of the fragments (1) and (3), rather than (1) and (2) as shown in Figure 4.4(c). This is an evidence that our visual system prefers to perceive structures with minimal amount of abrupt changes.



(a) (b) (c)

FIGURE 4.4.: Example for perceptual grouping based on the Gestalt principles. Colors indicate different groups. (a) A set of fragments as input. (b) Perceptual organization based on the closeness principle. (c) Perceptual organization into curves based on the good continuity principle. Note the grouping with respect to minimal total curvature.

In tensor voting, structure inference is done locally based on the contextual support from neighborhood. It is based on a communication scheme where every point, called *voter*, casts its local structural information encoded as a second order real symmetric positive semi-definite tensor vote, $\mathbf{T}$, to other points in its neighborhood, called *votee*, over a hypothesized smooth curve with a low total curvature. Tensor representation enables us, at a certain point, to encode all possible structure types and their saliences in a single term. Such structure types include point and curve in 2D/3D and surface in 3D. The employed tensor here is similar to the structure tensor used in computer vision for encoding local geometry. The shape of a ellipsoid corresponding to a tensor intuitively represents the local shape of the underlying structure. The positive semi-definiteness of the tensor ensures that the saliences are non-negative. In addition to this, a second order tensor is needed to encode

the orientation information in 2D and higher dimensional spaces.

The strength of the vote cast depends on the voter's perceptual saliency, the voter-to-votee distance and the curvature of the assumed curve connecting them. In the following two subsections, we provide more details about structure representation and tensor communication.

Structure Representation

In tensor voting, every structure type in N-dimensional space is characterized by the dimensionality d of its normal space $\mathcal{N}_d$ and its associated saliency s_d [King 08]. For instance, in 2D an isolated point has a ball-shaped tensor ($d = 2$, i.e. the normal space has the dimensionality of 2), whereas a point on a curve has a stick-shaped tensor ($d = 1$, i.e. the normal space has the dimensionality of 1) where the dominant component of the corresponding high dimensional ellipsoid represents the orientation of the normal (c.f. Figure 4.5). The strength of a d-structure is related to the magnitude of its saliency s_d. In the following, we will show how all possible structure types are represented by a single tensor.

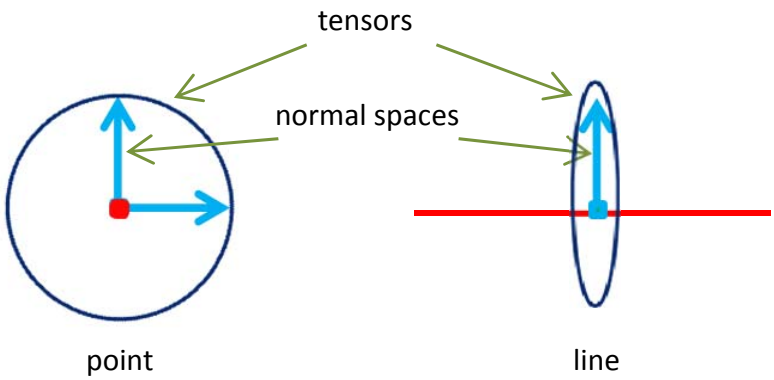


FIGURE 4.5.: Structure types in 2D, their normal spaces and corresponding tensor representations. Left: point structure (red). Right: linear structure (red). Note stick tensor is shown as a normal ellipse for the ease of representation. In reality, its minor axis radius goes to zero, i.e. its rank is 1.

A second order symmetric tensor $\mathbf{T}$ can be decomposed as [Itsk 09]

$$\mathbf{T} = \sum_{d=1}^{N} \lambda_d \hat{\mathbf{e}}_d \hat{\mathbf{e}}_d^T \tag{4.2}$$

where $\lambda_1 \geq \ldots \geq \lambda_N \geq 0$ and $\hat{\mathbf{e}}_1 \ldots \hat{\mathbf{e}}_N$ are eigenvalues and eigenvectors of $\mathbf{T}$, respectively and N is the dimensionality of the input space. Due to [Medi 00a], we can reformulate it as

$$\mathbf{T} = \sum_{d=1}^{N-1} (\lambda_d - \lambda_{d+1}) \sum_{k=1}^{d} \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^T + \lambda_N \sum_{k=1}^{N} \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^T \tag{4.3}$$

which can be further written as

$$\mathbf{T} = \sum_{d=1}^{N} s_d \mathcal{N}_d \quad \text{with} \quad s_d = \begin{cases} \lambda_d - \lambda_{d+1} & , d < N \\ \lambda_N & , d = N \end{cases} \tag{4.4}$$

where s_d is the saliency and $\mathcal{N}_d = \sum_{k=1}^{d} \hat{\mathbf{e}}_k \hat{\mathbf{e}}_k^T$ is the d-dimensional normal space associated with the d-structure type [King 08]. This representation can be interpreted as the decoding of the encoded information in the tensor into different structure types. Every possible structure type has a normal space and an associated saliency. The rank of the normal space gives hints about the type of the structure.

As discussed earlier, such a representation is suitable for combining information related to different structure types in a single term. By using a joint representation with saliences, it is not necessary to classify underlying structures into discrete class labels. This could also be related to Gaussian mixture models in the sense that at each point, a mixture of N models is fitted where structural saliences can be seen as the model weights. Another advantage of this representation is the ease with the aggregation of encoded information coming from different sources through tensor addition. This property will be especially important during tensor-based communication which we discuss in the following.

Tensor Communication

Tensor communication consists of two stages; encoding and decoding. Encoding happens at the *voter* site while decoding is done at the *votee*. A votee at point $\mathbf{q} \in \mathbb{R}^N$ accumulates incoming votes from all voters $\mathbf{p} \in \mathbf{P} \subset \mathbb{R}^N$, using tensor addition

$$\mathbf{T}(\mathbf{q}) = \sum_{\mathbf{p} \in \mathbf{P}} \mathbf{A}^{\mathbf{p}}(\mathbf{q}) \tag{4.5}$$

where $\mathbf{P}$ denotes the set of voters in the local neighborhood of the votee. Decoding of this tensor into structures can be done by using the decomposition described in Equations (4.2) to (4.4). Here, $\mathbf{A}^{\mathbf{p}}(\mathbf{q})$ represents the incoming second order symmetric tensor vote from voter at $\mathbf{p}$ encoding the local structural information.

Using the same decomposition scheme, $\mathbf{A}^{\mathbf{p}}(\mathbf{q})$ can be written as the sum of vote components for every d-structure as

$$\mathbf{A}^{\mathbf{p}}(\mathbf{q}) = \sum_{d=1}^{N} s_d^{\mathbf{p}} \mathbf{A}_d^{\mathbf{p}}(\mathbf{q}) \tag{4.6}$$

where $s_d^{\mathbf{p}}$ is the saliency of the d-structure type and $\mathbf{A}_d^{\mathbf{p}}(\mathbf{q})$ represents the vote component for the d-structure at $\mathbf{p}$.

A vote component for a d-structure is defined as

$$\mathbf{A}_d^{\mathbf{p}}(\mathbf{q}) = \sum_{j=1}^{d} \mathbf{S}_{d,j}^{\mathbf{p}}(\mathbf{q}) \quad \text{with} \quad \mathbf{S}_{d,j}^{\mathbf{p}}(\mathbf{q}) = w(\mathbf{p}, \mathbf{q}, \mathcal{N}_d^{\mathbf{p}}) \hat{\mathbf{v}}_{c,j} \hat{\mathbf{v}}_{c,j}^T \tag{4.7}$$

where $\mathbf{S}_{d,j}^{\mathbf{p}}(\mathbf{q})$ is a *stick voting field* for the j^{th} basis vector of $\mathcal{N}_d^{\mathbf{p}}$, i.e. the normal space of the d-structure. Before we continue with the explanation of stick voting field, let us give an intuitive definition of vote components. In plain words, a vote component for a certain d-structure is composed of contributions from the basis vectors of its normal space. A contribution of a certain basis vector is also in tensor form obtained by the covariance matrix of the inferred version of this vector.

To give more intuition, let us assume that, in 2D, the underlying structure at a point $\mathbf{p}$ is a curve, then, according to Figure 4.5 and Equation (4.6), the saliency of the 2-structure

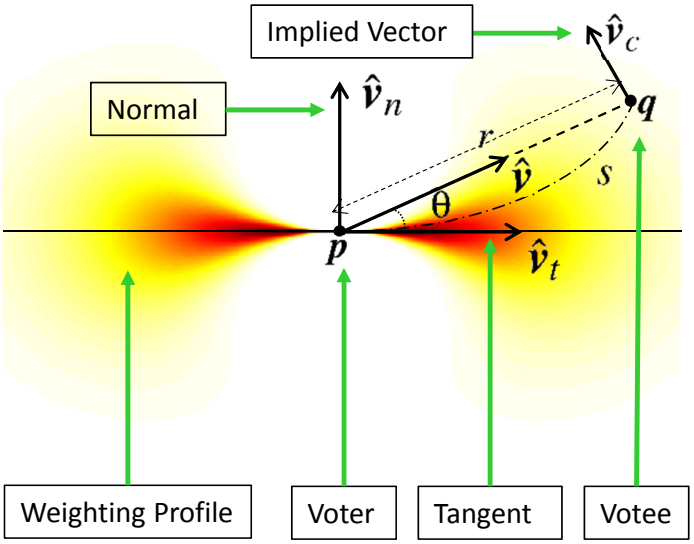


FIGURE 4.6.: Illustration of tensor voting. A voting scenario in 2D from a voter at $\mathbf{p}$ to a votee at $\mathbf{q}$ along an hypothesized circular arc s. A weighting function (color) determines the strength of a vote based on the distance between the voter and the votee. Here only a vote for a 1-structure, i.e. curve, which has information regarding only its normal space, $\mathcal{N}_1$, is sent.

becomes zero and we are left with a 1-structure having $\mathcal{N}_1$, i.e. only a single normal vector. Therefore, the tensor vote cast by the voter reduces to

$$\mathbf{A}^{\mathbf{p}}(\mathbf{q}) = s_1^{\mathbf{p}} \mathbf{A}_1^{\mathbf{p}}(\mathbf{q}) \quad \text{with} \quad \mathbf{A}_1^{\mathbf{p}}(\mathbf{q}) = \mathbf{S}_{1,1}^{\mathbf{p}}(\mathbf{q}) \tag{4.8}$$

where s_1 is a constant depending on the eigenvalues of the local tensor. This means only a single stick voting field for a single basis vector, i.e. the normal vector, is used. If there would be only a 2-structure type (isolated point), then, the only vote component for this structure will be the sum of stick voting fields for each basis vector of the corresponding normal space $\mathcal{N}_2$.

The stick voting field, $\mathbf{S}_{d,j}^{\mathbf{p}}(\mathbf{q})$ for the j^{th} basis vector, $\hat{\mathbf{v}}_{n,j}$, of $\mathcal{N}_d$ is the product of a weighting term $w(\mathbf{p}, \mathbf{q}, \mathcal{N}_d^{\mathbf{p}})$ with the covariance matrix of the implied vector $\hat{\mathbf{v}}_{c,j}$. The implication process for $\hat{\mathbf{v}}_{c,j}$ is depicted in Figure 4.6. It is based on the observation that the most appealing curve connecting the voter to the votee is a circular arc [Guy 97, Medi 04, King 08]. The implied normal can be seen as the vector that we would obtain if the normal vector, $\hat{\mathbf{v}}_{n,j}$, is slid on an assumed circular arc as shown in Figure 4.6 while preserving its orthogonality to the tangent of the circle. This circle is unique once $\mathbf{p}, \mathbf{q}$ and the normal at $\mathbf{p}$ is given.

Following this assumption and by projecting $\hat{\mathbf{v}}_{c,j}$ onto the unit basis vectors of the normal space $\mathcal{N}_d$ at $\mathbf{p}$, we get

$$\hat{\mathbf{v}}_{c,j} = \hat{\mathbf{v}}_{n,j} \cos(2\theta) - \hat{\mathbf{v}}_t \sin(2\theta) \tag{4.9}$$

where $\hat{\mathbf{v}}_t$ is the unit vector in the direction of the projection of $\mathbf{v} = \mathbf{q} - \mathbf{p}$ in the associated tangent space and θ is the vote angle defined as the complementary of the angle between $\hat{\mathbf{v}}_{n,j}$ and $\hat{\mathbf{v}}$. This simply gives the inferred normal at $\mathbf{q}$ if there would be a circular arc

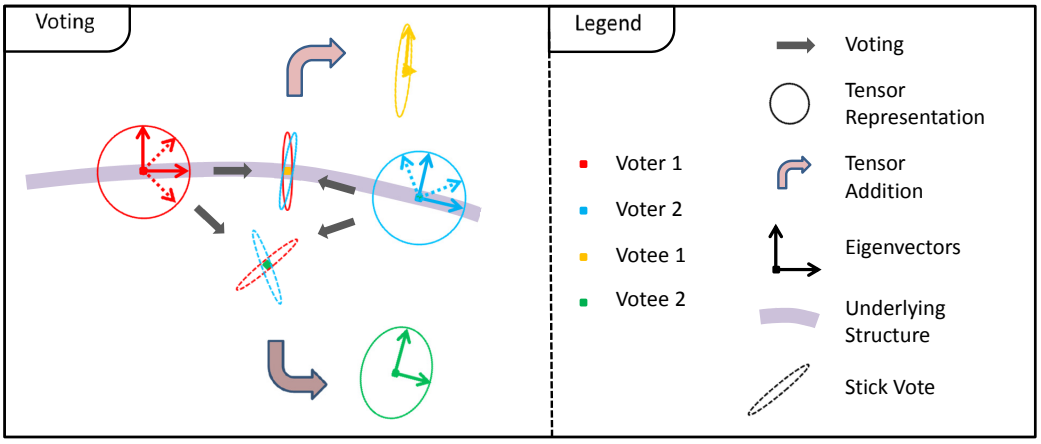


FIGURE 4.7.: Communication between ball tensors. Note how the normal space of a voter (dashed vectors) is oriented for each votee and how different accumulated tensor votes depend on the location of votees. Votee 1 received more stick votes in a certain orientation than votee 2 since it is more aligned with the voters than votee 2.

connecting $\mathbf{p}$ and $\mathbf{q}$. Mordohai [Mord 05] and later King [King 08] observed that if the first basis vector in $\mathcal{N}_d$ is chosen to be the unit vector in the direction of the projection of $\mathbf{v}$ onto this normal space, the implication operator in Equation (4.9) reduces to identity for the remaining orthogonal basis vectors spanning $\mathcal{N}_d$. This is due to the orthogonality of the remaining basis vectors to $\hat{\mathbf{v}}$ and it leads to a simplification in the computation of component votes.

Since a low total curvature of the assumed curve connecting the points is desired [Medi 00b], the magnitude of votes cast over arcs having high curvatures should be punished in order to reduce the likelihood of high curvature arcs. Furthermore, the requirement that the strength of the vote should decrease as the distance between the points justifies the usage of a weighting term based on the curvature of the assumed curve and the distance between the points.

The weighting term $w(\mathbf{p}, \mathbf{q}, \mathcal{N}_d^{\mathbf{p}})$ in Equation (4.7) essentially controls the magnitude of a tensor vote depending on the angle, θ, and the distance between the points, $r = \|\mathbf{v}\|$. It is defined as

$$w(\mathbf{p}, \mathbf{q}, \mathcal{N}_d^{\mathbf{p}}) = e^{-(\frac{s^2 + c\kappa^2}{\sigma^2})} \tag{4.10}$$

$$s = r\theta / \sin(\theta) \tag{4.11}$$

$$\kappa = 2\sin(\theta)/r \tag{4.12}$$

as proposed in [Medi 00a] where s is the arc-length, κ is the curvature of the assumed curve, σ is a scale parameter controlling the voting distance and c is a parameter that can be tuned to change the compactness of the voting field. Although Medioni et al. [Medi 00a] fixed this value as $c = \frac{-16\log(0.1)(\sigma-1)}{\pi^2}$, it can be adapted depending on the characteristics of applications.

While increasing r penalizes the votes cast to the points beyond some scale, larger values of θ reduces the effects of votes cast through a hypothesized curve. $w(\mathbf{p}, \mathbf{q}, \mathcal{N}_d^{\mathbf{p}})$, as shown in color in Figure 4.6, is simply a scaling function for the vote components. The strength of

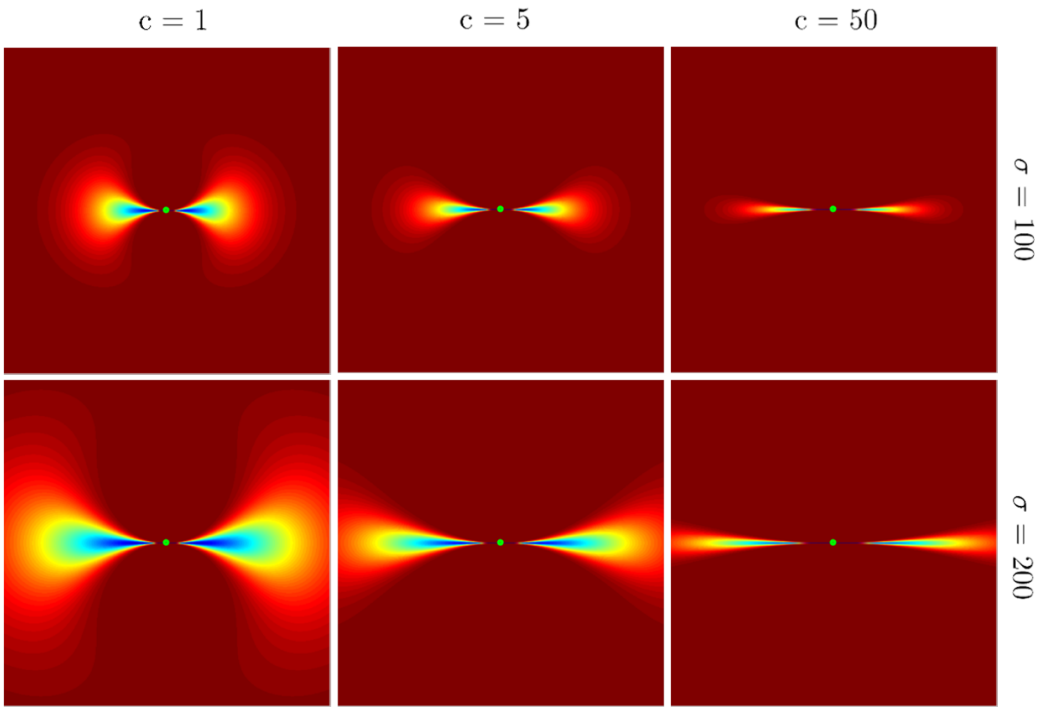


FIGURE 4.8.: Tensor weight profiles with changing free parameter and scale. Free parameter c can be used to control the sharpness of the weighting profile. The image size for each parameter pair is 500x500 pixels.

the stick vote is larger if the curvature is zero and if the voter-to-votee distance is minimal. Figure 4.8 shows different weighting profiles for changing values of σ and c.

It should be noticed that the weighting profiles shown in Figure 4.8 are based purely on the normal at the voter. The local curvature of the underlying structure is not considered for voting. We believe that including the local structural curvature would lead to a better inference by respecting more to the properties of the local structure as well as a longer range inference. In Section 4.2.4 we give the description of curvature based tensor voting along with some sample results on synthetic images.

The final accumulated vote cast by a set of voters, $\mathbf{P}$, to a certain votee at $\mathbf{q}$ becomes

$$\mathbf{T}(\mathbf{q}) = \sum_{\mathbf{p}\in\mathbf{P}} \sum_{d=1}^{N} s_d^{\mathbf{p}} \sum_{j=1}^{d} e^{-(\frac{s^2+c\kappa^2}{\sigma^2})} \hat{\mathbf{v}}_{c,j} \hat{\mathbf{v}}_{c,j}^T. \tag{4.13}$$

A subsequent decomposition applied to $\mathbf{T}(\mathbf{q})$ as described in Equations (4.2) to (4.4) will allow us to extract the saliences of different structure types. In summary, given a set of points with tensors assigned to them, one can let all points cast votes to a certain point where it becomes possible to infer a structure. In the following section, we will describe how the points are initialized and how they are used to propagate existing structures.

4.2.3. Structure Propagation

Tensor Initialization

Structure propagation is based on the local structural composition of voters. This means that before casting structural information through voting, every point should be aware of the composition of its underlying structures. This can be accomplished in several ways. The most straightforward way is to first calculate the structure tensor at every point in the image. Structure tensor is used quite often in computer vision to estimate the local orientation and coherence which parallels our definition of local structural composition [Schu 09]. If structure tensor is used as the initial tensor, then, every point in the image can be considered as voter. Although this would introduce robustness to the inference, the voting process becomes extremely expensive in term of computational cost.

We, instead, limit the set of voters to the most prominent structures in the image to simplify the voting complexity. This can be achieved by performing an edge detection and keeping only the edge pixels as the set of voters. This way, the complexity reduces by several factors since we do not include the points in regions with no structural significance.

For initializing the edge pixels we can either calculate and use the local structure tensor or assume local structural isotropy. What is interesting in the latter case is that we can let edge pixels, i.e. tokens, cast votes between each other in order to find mutually agreed representations for local structural compositions. Moreover, using the latter approach leads to a more robust normal estimation with respect to possible outliers during edge detection. Therefore, we have chosen to estimate the orientation directly from the unoriented tokens in order to provide a scheme that is applicable in cases where local gradient computation from image data is difficult.

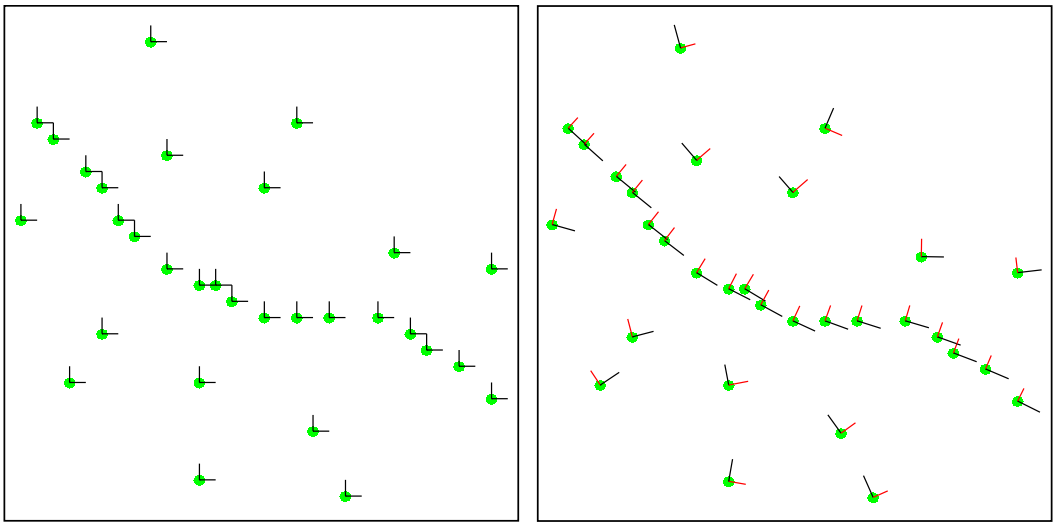
The process of estimating normals from unordered point data is called token refinement or sparse tensor voting which we explain in the following.

Tensor Refinement

We define a set of tokens, i.e. edge pixels, $\mathbf{P}_i \in \Omega_i$, for each edge image, E_i. Assuming that these tokens do not have any preferred orientation in the beginning, each of them is assigned a unit ball tensor with an arbitrary orthogonal basis as depicted in Figure 4.9(a). Making this assumption helps eliminating the dependence on the orientations of the edges as detected by the edge detector. Next, every token casts vote to every other token in its neighborhood using the voting mechanism described before. Since their local structural composition is only a unit ball tensor, the only non-zero saliency is $s_N = 1$ which follows from the eigen-decomposition of the identity matrix. Also their normal spaces are of dimension N.

At this stage, a vote component is created for each token that consists of contributions from each basis vector of its normal space. If the orthogonal basis of the normal space is chosen such that the projection of voter-to-votee vector is along the first unit basis vector, then, the total sum of the vote components reduces to a vote that will reinforce linearity. That is, unorganized tokens lying on a locally linear structure such as curve or plane will agree on a preferred normal space, which is the normal space of the underlying structure.

For instance, in 2D, without any preferred orientations, a voter will always vote in the favor of a locally linear curve, i.e. stick votes, passing through a certain votee and itself.



(a) Before token refinement. (b) After token refinement.

FIGURE 4.9.: Token refinement during tensor voting using a set of unorganized input tokens. Green dots represent the token positions while sticks stand for the eigenvectors of associated tensors. (a) Tokens are assigned ball tensors initially. They have no preferred orientations. (b) After a refinement step, they agreed on a certain orientation by exchanging information with each other in a local neighborhood. Note how tokens lying on a (possible) structure have similar orientation preferences.

This practically means that the stick voting field in Figure 4.6 is always aligned with the voter-to-votee vector $\hat{\mathbf{v}}$. This is possible due to the fact that the basis vectors of the normal space of a unit ball tensor can always be chosen such that one of the vectors is along $\hat{\mathbf{v}}$ and others are orthogonal to it. The contribution from the first basis vector will be negligible due to the large angle, while the contributions from the other remaining vectors will be strong due to the vote angle being zero (c.f. Figure 4.7).

At the end of this procedure, each token accumulates stick votes coming from other tokens in its neighborhood into a tensor which is then decomposed according to Equation (4.4). Tokens that are close to each other or lie on a structure mutually agree on a preferred orientation and have higher saliences for the underlying structure type as depicted in Figures 4.7 and 4.9(b). In our 2D example, if a votee receives more line votes from voters along a certain orientation, then, after decomposition, this will be its preferred orientation, which is an indication of the presence of a linear structure passing through this token with an associated saliency. At the end of this step, edge pixels will be assigned tensors encoding possible structure types and their saliences.

After the refinement step, it is possible to eliminate voters which are non-salient. This can be achieved by analyzing the structural composition of tensors. A votee with orientation uncertainty has larger ball saliency (i.e. s_N) compared to remaining structure types. At this point, a more suitable measure for local anisotropy is the *fractional anisotropy* (FA), which is a measure commonly used in diffusion tensor analysis [Bass 96]. It gives a normalized value in $[0, 1]$ where 0 means full isotropy and 1 means full anisotropy. For $N = 3$, FA is

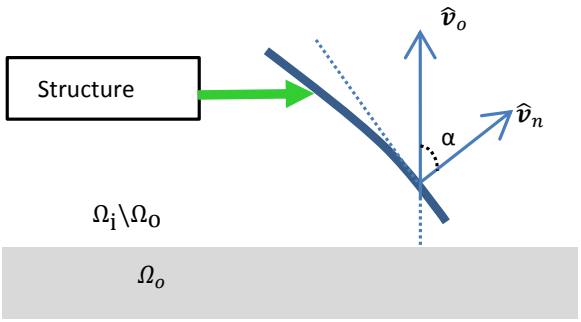


FIGURE 4.10.: Token refinement based on the angle of incidence. If the angle α for a token is smaller than a threshold it can be discarded from structure propagation.

defined as

$$FA = \sqrt{\frac{3}{2}} \frac{\sqrt{(\lambda_1 - \hat{\lambda})^2 + (\lambda_2 - \hat{\lambda})^2 + (\lambda_3 - \hat{\lambda})^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{4.14}$$

where $\hat{\lambda} = (\lambda_1 + \lambda_2 + \lambda_3)/3$. For $N = 2$, one can employ the *normalized coherence measure* (NCM) to assess local anisotropy [Jahn 93]. Normalized coherence is defined as

$$NCM = \frac{(\lambda_1 - \lambda_2)^2}{(\lambda_1 + \lambda_2)^2}. \tag{4.15}$$

In both Equations (4.14) and (4.15), the case where all eigenvalues are zero can be detected and eliminated. This is actually unlikely in the sparse voting case since all voters are assigned ball tensors which is fully isotropic. Thus, FA and NCM can be used as normalized anisotropy measures for 3D and 2D, respectively. By using these measures along with a threshold in $[0, 1]$, it is possible after the initial sparse voting step to choose a subset of voters that are structurally anisotropic. Such a threshold ensures that only tokens lying on structures are used in the subsequent propagation stage.

Structure Propagation

After having assigned tensors to edge pixels, i.e. tokens, encoding structure types and their associated saliences, every token can now propagate its information again using tensor voting. In other words, based on their self information about the underlying structures, tokens vote for possible structures in other regions. We use this for the inference of structures in the extended regions Ω_{ie} in Figure 4.3(a). Practically, each token in Ω_i sends a vote to every pixel in Ω_{ie}.

At this point, it is possible to put a restriction on the voters such that only the ones lying on a structure which makes a sufficient angle of incidence with the boundary of Ω_o are included in the propagation. The reason for this could be, for instance, to filter out those points lying on structure with small probability of passing through the image boundary. Therefore, $\mathbf{P}_i$'s can be further refined depending on the angle $\alpha = \arccos(\hat{\mathbf{v}_n}^T \hat{\mathbf{v}}_o)$ where $\hat{\mathbf{v}}_o$ is the unit normal vector of the boundary of Ω_o as shown in Figure 4.10. This restriction may help to avoid ineffective votes due to the voter-to-vote angle leading to computational

overhead. If curvature information is available, a similar strategy can be employed to eliminate points with high curvatures. A dense voting is performed using the remaining salient voters by

$$\mathbf{T}(\mathbf{q}_i) = \sum_{\mathbf{p}_i \in \mathbf{P}_i} \mathbf{A}^{\mathbf{p}_i}(\mathbf{q}_i), \quad i \in \{1,2\}, \mathbf{p}_i \in \Omega_i, \mathbf{q}_i \in \Omega_{ie} \tag{4.16}$$

where every pixel position $\mathbf{q}_i$ in the extended region Ω_{ie} is considered as a votee.

It should be noted that, at this stage, any point in space can be considered as a votee. Employing such a strategy, for instance, might be helpful to obtain intrinsic structural representation contained in a noisy or distorted image. Once dense voting is finished, structural saliency maps, which we consider as the structural probability maps, can be extracted from the accumulated tensors as follows

$$S_i(\mathbf{q}_i) = \lambda_1(\mathbf{q}_i) - \lambda_2(\mathbf{q}_i) \;\; i \in \{1,2\}, \mathbf{q}_i \in \Omega_{ie} \tag{4.17}$$

where $\lambda_1(\mathbf{q}_i)$ and $\lambda_2(\mathbf{q}_i)$ are calculated using the eigendecomposition of $\mathbf{T}(\mathbf{q}_i)$ according to Equations (4.2) to (4.4). This structural saliency corresponds to curve saliency and surface saliency in 2D and 3D, respectively. Although we make use only one type of structural saliency for creating the scalar structural probability maps, it is also possible to create a vector valued map where each component corresponds to a different structure type.

Finally, we create the structure images S_i by setting intensity values from the overlapping region $\Omega_o = \Omega_{io} \cup \Omega_{ie}$. This is done simply by cropping the portion of the extended edge image E_i falling in Ω_o as shown in Figure 4.3(a). This is a combination of probabilistic and deterministic structural maps. Now, we have a pair of completely overlapping structure images, S_1 and S_2. The observed structures in S_1 corresponds to the estimated structures in S_2, and vice versa. Alternatively, one could also obtain a structural probability map for the observed structures if the observed structures are too noisy or distorted. Here, for the sake of simplicity, we will assume that our observed structures are free of noise. However, in the experiments, we will show that the proposed method is already robust against noise. This property will be important when using structural probability maps for digital pathology images where structures are often distorted. In order to make sure that structures at different scales are treated appropriately, one could also employ a multi-scale approach for structure propagation, which we will mention in the context of registration.

4.2.4. Curvature-based Structure Propagation

Curvature is one of the important descriptors for the local surface geometry. Although it is straightforward to estimate if the surface is represented analytically, special techniques are necessary in case of discrete representations such as triangular meshes. Therefore, estimating curvature has been studied extensively for the analysis of mesh data. Two major approaches include local analytical surface fitting and discrete approximation. While fitting methods benefit from locally having an analytical representation for the surface that can be used to analytically derive curvature, discrete methods enjoy the lower computational cost compared to the former methods [Gatz 06].

One of the discrete approaches for estimating curvature is the curvature tensor. Taubin [Taub 95] provides a very elegant formulation for discretely estimating the curvature tensor

using local neighborhoods. Curvature was introduced into the tensor voting framework by Tang and Medioni [Tang 02]. They included an additional step after the initial sparse voting for the estimation of local surface shape with the help of the sign of curvature. They classify each point into labels $\{planar, elliptic, parabolic, \{hyperbolic, outlier, discontinuity\}\}$ that are considered during the last dense voting step along with the estimated sign of curvature. [Tong 05] later extended this work by introducing the curvature tensor into the voting framework. Using this approach it is possible to estimate the principle curvatures and associated directions at each input token using tensor voting. However, this is only limited for surfaces where 2D curvature tensors are estimated in the tangent plane of each token.

Curvature Tensor

In 3D, the local surface geometry can be represented by the Darboux frames [Peti 02] which is defined as

$$\Delta\mathbf{q} = (\mathbf{q}, \hat{\mathbf{v}}_n, \hat{\mathbf{v}}_\kappa^1, \hat{\mathbf{v}}_\kappa^2, \kappa_1, \kappa_2). \tag{4.18}$$

Here, $\mathbf{q}$ is a surface point, $\hat{\mathbf{v}}_n$ is the normal at $\mathbf{q}$ and $\hat{\mathbf{v}}_\kappa^i$, κ_i are respectively the maximum and minimum curvature directions and associated curvatures. The curvature tensor is defined as a mapping form every point $\mathbf{q}$ to the directional curvature $\kappa(\hat{\mathbf{v}}_t)$ along a unit length tangent vector $\hat{\mathbf{v}}_t$ [Taub 95]. If $\hat{\mathbf{v}}_t$ is written in terms of an orthonormal basis $\hat{\mathbf{v}}_t^1, \hat{\mathbf{v}}_t^2$ in the tangent space of the surface at $\mathbf{q}$ as $\hat{\mathbf{v}}_t = t_1\hat{\mathbf{v}}_t^1 + t_2\hat{\mathbf{v}}_t^2$, then, $\kappa(\hat{\mathbf{v}}_t)$ can be expressed in the quadratic form

$$\kappa(\hat{\mathbf{v}}_t) = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}^T \begin{pmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \tag{4.19}$$

κ_{11} and κ_{22} are the curvatures along the basis vectors $\hat{\mathbf{v}}_t^1, \hat{\mathbf{v}}_t^2$, respectively, and $\kappa_{12} = \kappa_{21}$. If $\kappa_{12} = \kappa_{21} = 0$, then, $\hat{\mathbf{v}}_t^1, \hat{\mathbf{v}}_t^2$ become the principle directions ($\mathbf{v}_\kappa^1, \mathbf{v}_\kappa^2$ in Equation (4.18)) and κ_{11}, κ_{22} become the associated principle curvatures (κ_1, κ_2 in Equation (4.18)), respectively. Hereafter, we will use the principle vectors as the orthonormal basis of the tangent space.

Let $\hat{\mathbf{v}}_t^\theta$ be a arbitrary vector in the tangent plane of the surface at $\mathbf{q}$ where θ is the angle between the first principle vector $\hat{\mathbf{v}}_t^1$ and $\hat{\mathbf{v}}_t^\theta$. Then, one writes the curvature along $\hat{\mathbf{v}}_t^\theta$ as

$$\kappa(\hat{\mathbf{v}}_t^\theta) = \kappa_1 \cos^2(\theta) + \kappa_2 \sin^2(\theta). \tag{4.20}$$

A second order symmetric matrix can be obtained by integrating this for all possible unit tangent vectors

$$M = \frac{1}{2\pi} \int_{-\pi}^{\pi} \kappa(\hat{\mathbf{v}}_t^\theta) \hat{\mathbf{v}}_t^\theta (\hat{\mathbf{v}}_t^\theta)^T d\theta \tag{4.21}$$

which represents the curvature tensor at $\mathbf{q}$. Taubin [Taub 95] shows that the eigenvectors of M correspond to the principle directions $\{\hat{\mathbf{v}}_t^1, \hat{\mathbf{v}}_t^2\}$ in addition to the normal vector and the corresponding eigenvalues lead to the principle curvatures using the following linear map

$$\kappa_1 = 3a_1 - a_2 \tag{4.22}$$

$$\kappa_2 = 3a_2 - a_1 \tag{4.23}$$

with a_1 and a_2 being the eigenvalues of M corresponding to $\{\hat{\mathbf{v}}_t^1, \hat{\mathbf{v}}_t^2\}$. Equation (4.21) can be approximated by

$$\hat{M} = \sum_{\mathbf{p} \in V_\mathbf{q}} \omega_\mathbf{p} \kappa_{\hat{\mathbf{v}}_t^\mathbf{p}} \hat{\mathbf{v}}_t^\mathbf{p} (\hat{\mathbf{v}}_t^\mathbf{p})^T \tag{4.24}$$

where $V_{\mathbf{q}}$ is the neighborhood of $\mathbf{q}$, $\hat{\mathbf{v}}_t^{\mathbf{p}}$ is the projection of $\mathbf{r} = \mathbf{p} - \mathbf{q}$ onto the tangent space of the surface at $\mathbf{q}$. $\omega_{\mathbf{p}}$ is a normalized weighting term for neighboring element. This can be computed from the local triangular mesh if the surface is estimated on a mesh [Taub 95] or from the votee and voter normals in terms of vote consistency in tensor voting [Tang 02]. The directional curvature $\kappa_{\hat{\mathbf{v}}_t^{\mathbf{p}}}$ in the direction $\hat{\mathbf{v}}_t^{\mathbf{p}}$ can be estimated using differential geometry as

$$\kappa_{\hat{\mathbf{v}}_t^{\mathbf{p}}} = \frac{2\mathbf{n}^T\mathbf{r}}{\|\mathbf{r}\|^2}. \tag{4.25}$$

In fact, this is equivalent to our definition in Equation (4.12) for the curvature of the osculating circle that passes through $\mathbf{p}$ and $\mathbf{q}$.

In order to factor out the principle directions and corresponding curvatures from Equation (4.24), we use an approach similar to [Taub 95]. The matrix $\hat{M}$ in Equation (4.24) has the normal vector $\hat{\mathbf{v}}_n$ as one of its eigenvectors with an eigenvalue equal to zero. This is due to the projection of the vector $\mathbf{r}$ onto the tangent plane leading to a rank-deficient matrix $\hat{M}$. Taubin [Taub 95] shows that using a Householder transformation for restricting the matrix to the tangent plane and a subsequent Givens rotation for diagonalizing the remaining minor matrix, it is possible to find the principle directions and curvatures.

Curvature Tensor Estimation using Tensor Voting

Using the formulation above, curvature information can be obtained for curves in 2D and surfaces 3D. Although this formulation is developed for estimating curvature for vertices on a mesh with known normal directions, it can be used also for estimating the curvature information for points in an unorganized point set. However, it is necessary to first extract normals for curves in 2D and surfaces in 3D. This can be accomplished through sparse tensor voting as described in Section 4.2.2. This results in a set of points with encoded local structural saliences that can be used to filter out non-salient points in terms of curve or surface.

Similar to the sparse voting step for normal estimation, curvatures can be estimated by using the decoded normals at each point. Each point receives weighted "curvature votes" from its neighbors as shown in Equation (4.24). Weights can be computed by using the same weighting term as the one used for tensor voting Equation (4.10). Scale parameter σ should be chosen such that a sufficient local neighborhood is used for estimating the curvature of the underlying structure. This often depends on the curvature of the underlying structure.

During curvature voting, the sign of curvature is efficiently encoded by using the approach of [Tong 05] where the votee-to-voter vector $\mathbf{r}$ is projected onto the normal vector and the sign of the projection is used as the sign of curvature. In fact, this is the sign of the cosine between the normal at the votee and the votee-to-voter.

At the end of the second tensor voting pass, all the input points contain their surface saliency as well as curvature information. This means that at this point, the underlying structure, i.e. curve in 2D or surface in 3D, can be fully described by the available information obtained via the two sparse voting steps.

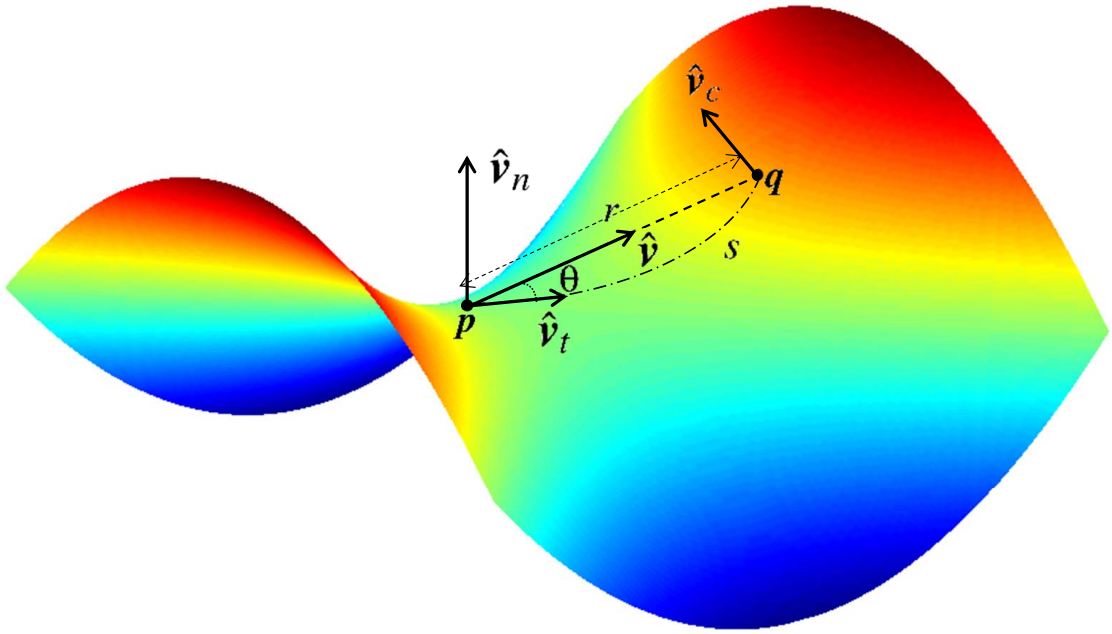


FIGURE 4.11.: Illustration of curvature-based tensor voting. Vote cast by the voter will be weighted depending on the local surface shape.

Curvature-based Structure Propagation

Local structural information can be propagated in a manner similar to the previously discussed locally linear structure propagation. Despite its success demonstrated in this work, the main limitation of the locally linear structure propagation approach was the limited amount of information (first order, i.e. orientation) that was being propagated. Thanks to the scheme described here, it is now possible to encode and propagate second order structural information using curvature tensors.

Curvature-based structure propagation works in a very similar way to the linear structure propagation discussed previously. The major difference is the modification of the decay function (the weighting term in Equation (4.10)). We propose to add the underlying curvature to the curvature of the assumed circular arc connecting a voter to a votee. The new weighting term for a vote cast by $\mathbf{p}$ at $\mathbf{q}$ becomes

$$w(\mathbf{p}, \mathbf{q}, \mathcal{N}_d^{\mathbf{p}}) = e^{-(\frac{s^2 + c\hat{\kappa}^2}{\sigma^2})} \tag{4.26}$$

$$\hat{\kappa} = \kappa_0 \pm \kappa_{\hat{\mathbf{v}}_t^{\mathbf{q}}} \tag{4.27}$$

where s is given by Equation (4.11), κ_0 is given by Equation (4.12) and $\kappa_{\hat{\mathbf{v}}_t^{\mathbf{q}}}$, the directional curvature at $\mathbf{p}$ in the direction of $\hat{\mathbf{v}}_t^{\mathbf{q}}$, can be obtained using Equation (4.20) given principle directions and corresponding curvatures. The directional curvature is added to or subtracted from the assumed curvature depending on whether the votee $\mathbf{q}$ is in the lower or the upper half space of the tangent plane at $\mathbf{p}$. This can be easily determined by $sign(\mathbf{r} \cdot \hat{\mathbf{v}}_n)$.

Figure 4.11 shows a scenario where a voter at $\mathbf{p}$ casts a stick vote to a votee at $\mathbf{q}$. Using curvature-based tensor voting, the vote will not be weighted down due to the distance of

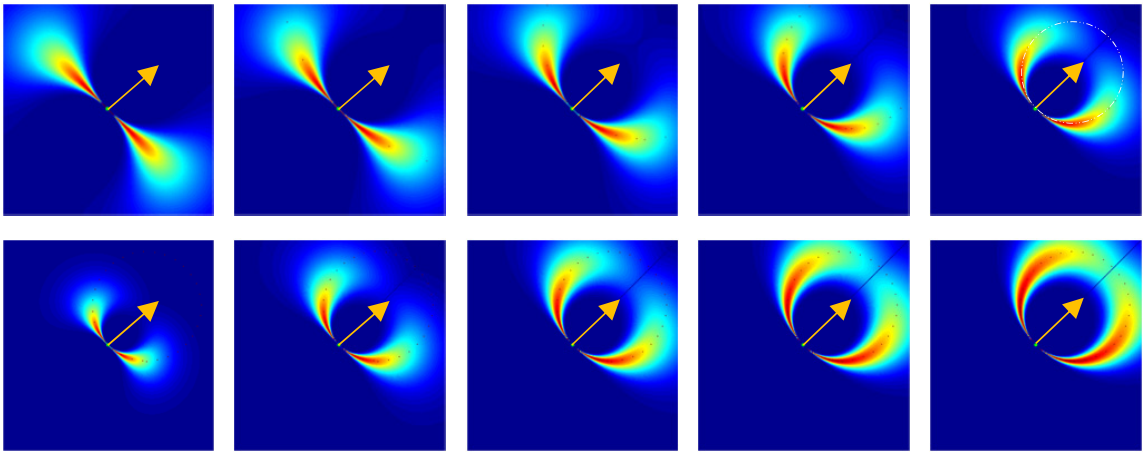


FIGURE 4.12.: Top row: Curvature-based vote weighting function with increasing local curvature from left to right. A circle with the corresponding radius of curvature is shown as reference on the rightmost figure. Bottom row: Weighting function with increasing scale parameter from left to right. Note the capture range is limited by the radius of curvature.

the votee from the tangent plane of the voter. The directional curvature will be taken into account so that the vote cast to the votees that are far from the circular arc contained in the normal plane defined by the normal of the votee and the voter-to-votee vector are weighted down.

Figure 4.12 shows in the top row different weighting profiles for a voter with changing local curvatures. It is clear that the votees in the support of osculating circles get higher weights than the others. In the bottom row are the 2D profiles for a fixed curvature but changing scale parameter. With a curvature, the strength of a vote is limited with the osculating circle. The weighting profiles get more complicated in 3-dimensional space. Depending on the sign of the Gaussian curvature, defined as the product of principle curvatures, one differentiates between elliptical, hyperbolic and parabolic points. Weighting profiles for these cases are shown in Figure 4.13.

We use the curvature-based tensor voting approach to extend one of the structure images shown in Figure 4.15. We applied both linear and non-linear structure propagation techniques and varied the scale parameter to observe the differences as shown in Figure 4.14. It can be observed from the figure that non-linear approach has much more consistent propagation compared to the linear one especially in the region where the structure is non-linear. However, what is also noteworthy is the decreasing propagation quality of the non-linear approach with increasing scale parameter. This is especially the case where the structure is highly curved. The main reason for this is the usage of a much larger neighborhood for the estimation of curvature than necessary with increasing σ.

The above observation suggests that the neighborhood size for curvature tensor estimation should be adapted to the non-linearity of underlying structures. In this sense, curvature estimation bears some similarities with the non-linear dimensionality reduction techniques where the goal is to recover the underlying manifold surface using local information. Therefore, we believe that a local neighborhood analysis such as estimating

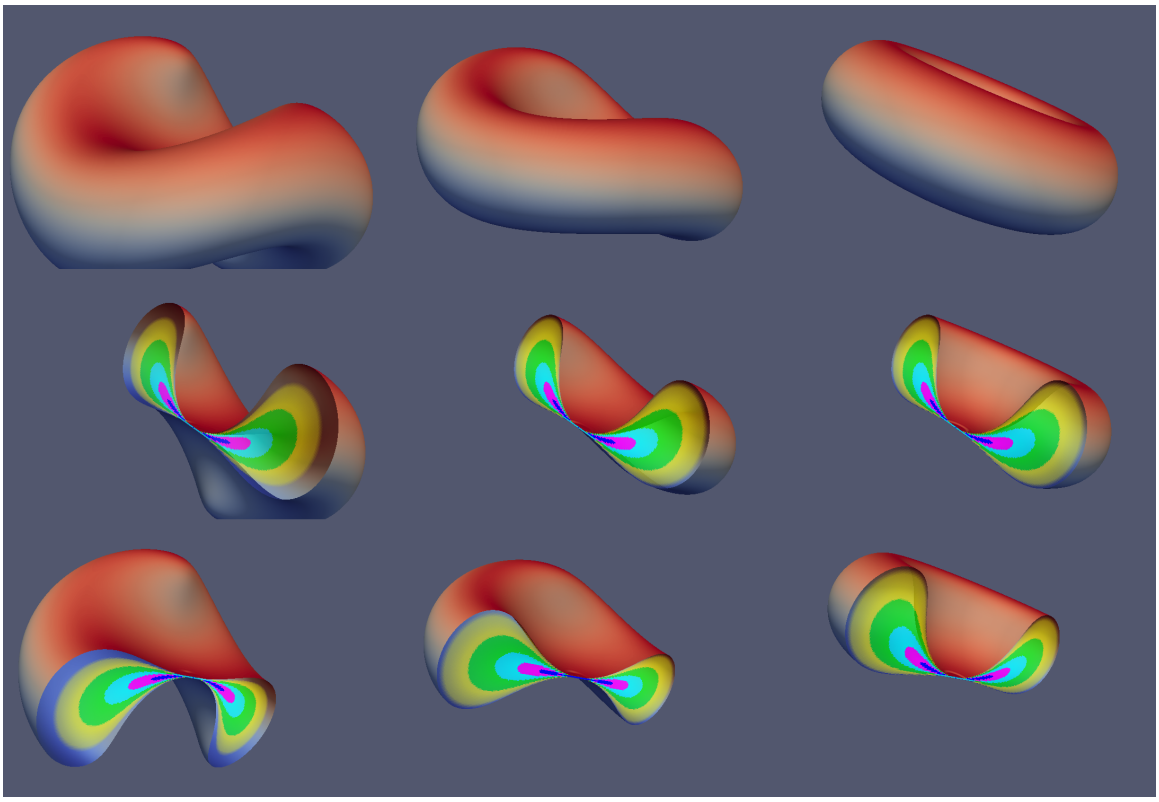


FIGURE 4.13.: 3D weight profiles for curvature-based tensor voting. Columns show weighting profiles for hyperbolic, parabolic and elliptic surface points, respectively. Rows show surface rendering for each case and cuts of the surfaces along x and y planes, respectively.

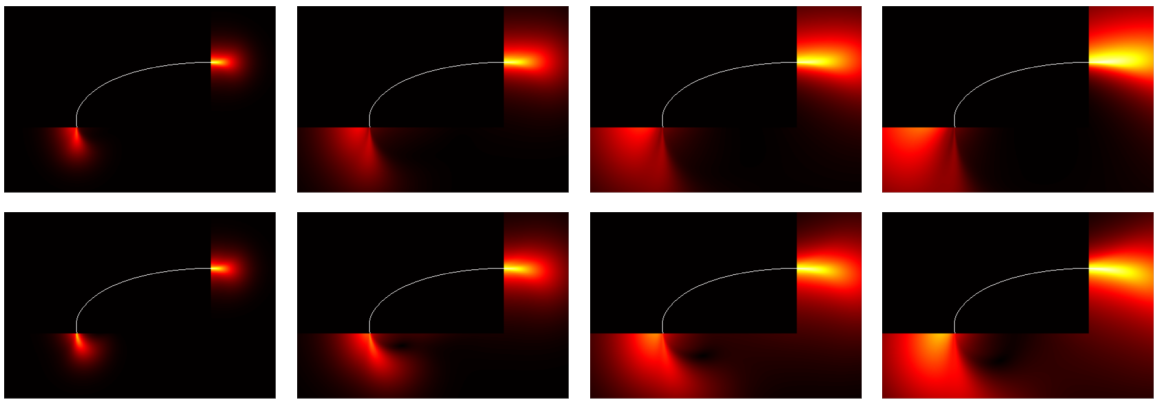


FIGURE 4.14.: Comparison of linear vs non-linear structure propagation. Top row shows linear propagation (i.e. using only normals) whereas bottom row shows non-linear (i.e. normal and curvature-based) propagation. Columns show structure propagation with varying scale parameter σ.

geodesic distances would help improve curvature estimation, thus structure propagation.

This section has shown how local curvature information can be estimated and exploited for structure propagation. Results on synthetic images indicate potential improvement in structure propagation when applied on real images. The benefits of using non-linear structure propagation would be the improved consistency of propagated structures especially in cases of occlusion or gaps. Finally, it should be noted that this section provides only a theoretical extension to the previous linear structure propagation and it has not been experimentally validated.

4.2.5. Implementation Details

The implementation of structure propagation was done purely in `C++`. We use an adapted version of the tensor voting library presented in [King 08] with its implementation as a part of VXL [2]. The implementation is both in 2D and 3D. While tensor voting related parts are implemented based on the VXL syntax, pre- and post image processing tasks were implemented using an in-house developed image processing library called Colibri.

4.3. Applications

In this section, we will describe some of the potential applications of structure propagation. First, we will show how stitching of 2D images can be formulated within the structure propagation framework. Due to the ease of demonstration in 2D, we used this application in previous sections for explaining our method. This application was published in [Yigi 13].

Although it is possible in some applications to secure a shared region to be used later for alignment purposes, there are cases where it is not possible to have an overlap. Creating 3D digital volumes from 2D slices in the context of digital pathology is an example for this. Therefore, we will demonstrate how existing histology reconstruction techniques can benefit from the structure propagation framework for an improved local structural consistency. This application was published in [Muel 14].

Finally, as the last application of our contribution, we will show how structure propagation can be employed for distortion correction with the purpose of creating whole-body Magnetic Resonance Images (MRI). This application was published in [Yigi 15].

4.3.1. 2D Image Stitching

When creating a wide field-of-view of a certain scene, partial observations have to be stitched together in a consistent way. This consistency is usually ensured by allowing an overlap region between subimages. This overlap is later used for estimating the transformation parameters that will bring the subimages into a consistent alignment. In this work, we claim that for a successful alignment of subimages such an overlap does not need to be secured during acquisition. In previous sections, we have demonstrated how such an overlap can be established in the absence of an existing one. In order to achieve a consistent alignment, a registration method is needed to recover the transformation parameters using the information in the established overlap region. Stitching of non-overlapping images can

[2] VXL library is available for download at `http://vxl.sourceforge.net/`.

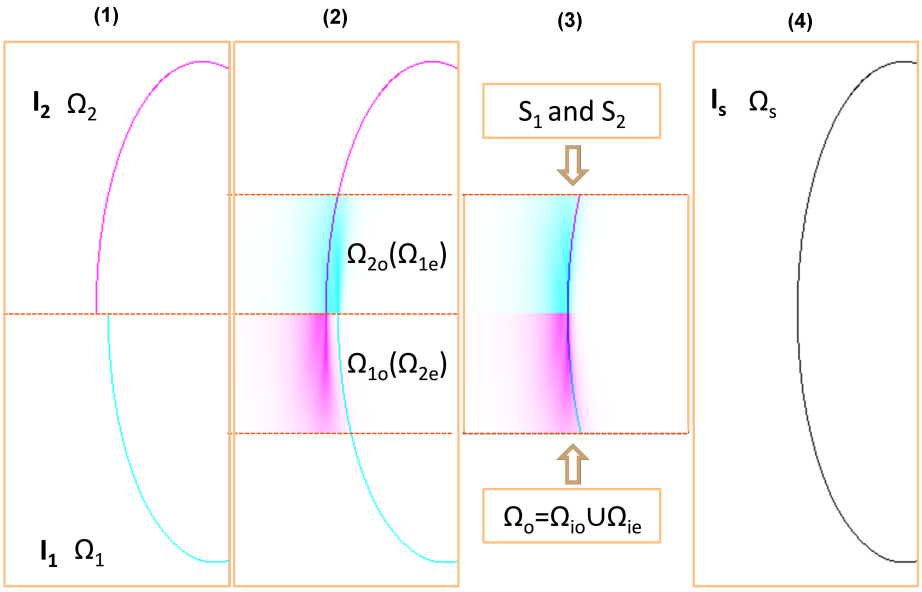


FIGURE 4.15.: 2D image stitching by structure propagation using a synthetic image pair. (1) Shown are only the structure images. (2) Structures in one of the subimages are propagated towards the other one. (3) Then, the subimages in the established overlapping region are used for estimating the necessary transformation parameters. (4) Finally, resulting parameters are applied to the original images yielding a smooth and continuous stitching of structures at the stitching boundary.

be cast as a problem of registering structural images S_i which contain propagated structures in the overlapping region. Transformations obtained from the successful registration of structural images can be transferred to the original images I_i. Figure 4.15 shows an example of registering non-overlapping partial observations in 2D.

Considering again two partial observations introduced earlier in Section 4.2.1, registration of S_1 and S_2, which we can put as the source and the target images respectively, is posed as an optimization problem such that a transformation T optimizes an energy $\mathcal{E}$. Transformation T can be parametrized in terms of a set of n parameters $\mathbf{t} = \{t_1, \cdots, t_n\}$. Optimal transformation parameters $\hat{\mathbf{t}}$ can be estimated via the following equation

$$\hat{\mathbf{t}} = \arg\min_{\mathbf{t}} \mathcal{E}(\mathbf{t}) \tag{4.28}$$

where $\mathcal{E}$ is defined in terms of a similarity measure as

$$\mathcal{E}(\mathbf{t}) = \xi(S_1 \circ T_{\mathbf{t}}, S_2) \tag{4.29}$$

with $T_{\mathbf{t}}$ being the parametrization of T by $\mathbf{t}$.

Misalignment between subimages can be global, which can be corrected by using a linear registration technique, but it can also be local where a non-rigid or deformable registration method is needed. In this section, we consider only global, linear misalignments. We employ a recently proposed intensity-based affine registration method [Ziki 10] which is based on a discrete Markov Random Field (MRF) formulation. For the details about the used affine registration and discrete optimization methods, we refer the reader to Section 3.1.6.

For the optimization of Equation (3.12), Fast-PD [Komo 07, Komo 08], an efficient MRF optimization algorithm is used. Registration is performed in a multi-resolution setting to avoid local minima, where the optimal transformation is communicated between levels. Note that both structure propagation and stitching take place in a multi-resolution setting. Such an approach during structure propagation ensures that structures present at larger scales are considered first for the optimization whereas structures at a smaller scale are used for refining the transformation parameters. Although not implemented here, alternatively, one could also combine different scales of structure propagation and registration to make use of a larger scale propagation also in a finer scale registration for regularization purposes.

In this work, after experimenting with several standard similarity measures for intensity based image registration, we ended up using the normalized cross-correlation (NCC) measure (Equation (3.9)) as the similarity measure $\xi(\cdot,\cdot)$ in Equation (4.29). We integrated the structure propagation library into the discrete MRF based linear image registration framework [Gloc 08, Ziki 10].

For the demonstration of locally consistent stitching of 2D subimages by respecting the local smoothness and the continuity of structures across subimage boundaries, we have used synthetic as well as real image pairs. In Section 4.4.1, we show and interpret the results of the experiments on these images.

4.3.2. Deformable Histology Reconstruction

In digital pathology, the reconstruction of a 3D volume from a stack of 2D digital histology slices remains as a challenge especially if no external references are available. Without a reference, standard registration approaches tend to align structures that should not be perfectly aligned.

In this section, we show how structural probability maps can be effectively used as a regularization term for the deformable, reference-free reconstruction of histology volumes. Structural probability map gives an estimate of the original 3D structure of the sample from the misaligned and possibly corrupted 2D slices. Here, we describe two reconstruction approaches, consecutive and simultaneous, that incorporate this estimate in a deformable registration framework. In Section 4.4.2, we demonstrate experiments on synthetic and mouse brain data sets showing that using structural probability maps yields similar results when compared to techniques that operate reference-based on synthetic data sets. Moreover, it improves the smoothness of the reconstruction compared to standard registration techniques on real data with no references.

Let $\mathcal{I} = (I_1, ..., I_g)$ be a stack of 2D images that we consider as a volumetric image. We further define structural probability map images $\mathcal{S} = (S_1, ..., S_g)$ where $S_i(\mathbf{y})$ is the structural saliency at $\mathbf{y} \in \mathbb{R}^2$. As discussed in Section 4.2.3, we consider strong edges in images I_i detected by a standard 2D edge detector as the set of voters $E = (E_1, ..., E_g)$ and every point in S as votees. After the initial sparse voting performed in 3D, inference is also done in 3D by performing a dense voting for each votee and then extracting surface saliences from the accumulated tensors $\mathbf{T}(\mathbf{y})$ using Equations (4.2) to (4.4). Finally, surface saliences are set as scalar values for $S(\mathbf{y})$.

Consecutive Registration We assume that the histology stack was already roughly prealigned by a standard rigid registration, thus our method aims to improve the smoothness

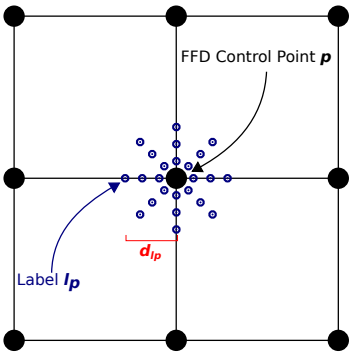


FIGURE 4.16.: Example for MRF labeling around a control point.

by performing a modified deformable registration based on 2D Free-Form Deformations (FFD). We refer to Section 3.1.5 for more details about FFD-based image registration. We pose the deformable registration as a discrete optimization problem using Markov Random Fields (MRFs) [Gloc 08]. A 2D FFD grid G^i is assigned to every slice I_i, thus each control point $\mathbf{p}$ represents a node in the MRF. In order to model the actual displacement of control points we designate a labeling l of discrete values to all nodes. Each label $l_\mathbf{p}$ therefore describes the displacement $\mathbf{d}_{l_\mathbf{p}}$ of the control point $\mathbf{p}$ (see Figure 4.16). The labeling problem can then be solved with a quadratic pseudo-boolean optimization (QPBO) algorithm [Kolm 07].

In our first approach the labeling is solved consecutively for each slice I_i in the stack by minimizing the following cost function

$$\mathcal{E}_i(l) = \sum_{\mathbf{p}\in G^i} \Big(\mathcal{E}_{data}(I_i, I_{i+1}, l_\mathbf{p}) + \gamma \mathcal{E}_S(E_i, S_i, l_\mathbf{p}) + \rho \mathcal{R}(l_\mathbf{p}) \Big) \tag{4.30}$$

where $\mathcal{S} = (S_1, ..., S_g)$ depict the structural probability maps obtained through structure propagation and $E = (E_1, ..., E_g)$ the edge maps for each slice in the stack, γ and ρ weight the contributions of the respective terms. $\mathcal{E}_{data}$ is defined as

$$\mathcal{E}_{data}(I_i, I_{i+1}, l_\mathbf{p}) = NCC(I_i^{'}, I_{i+1}) \tag{4.31}$$

and it compares the deformed slice (denoted by prime) with the undeformed neighbor using Normalized Cross Correlation (NCC) which yields robust results with lower computational cost. It is also possible to replace this data term with more sophisticated measures such as mutual information.

Notice that we do not register to the deformed neighboring slice because this would accumulate deformations and therefore introduce a strong drift. Therefore this term alone can only align the slices roughly since the deformations of two neighboring slices are independent from each other.

The novel data-based regularization term $\mathcal{E}_S$, on the other hand, introduces further control regarding the consistency of structures in the neighborhood. This term measures the similarity between the existing structures in a slice, represented by E_i, with the estimated structural probability map S_i through structure propagation from its local 3D neighborhood. Since the structural probability map provides a perceptually salient approximation

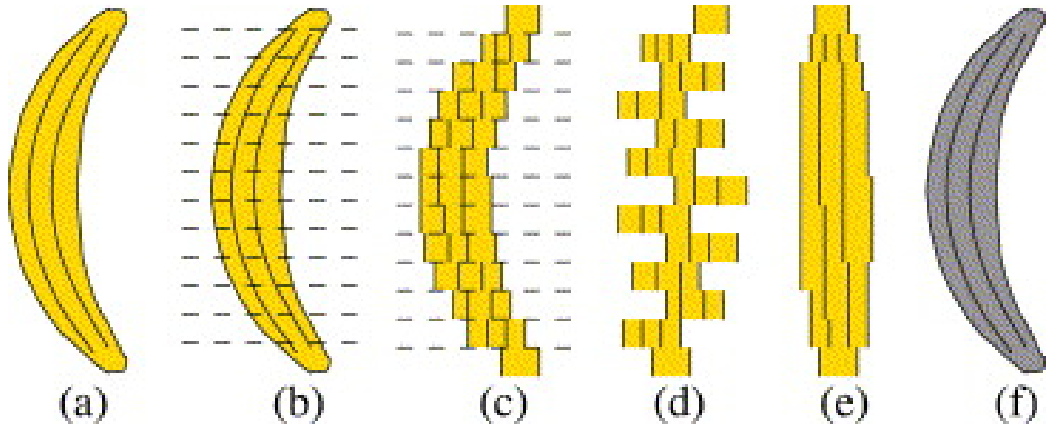


FIGURE 4.17.: The banana problem: the 3-D reconstruction of a 3-D curved object is not easy. (a) Take a 3-D curved object (e.g., a banana); (b) cut it into slices; (c) digitize the slices; (d) mix the digitized slices; (e) the 3-D reconstruction results in a cylindrical banana. (f) Using a shape prior (e.g., MRI) may help to reconstruct the curved banana. (Figure and caption reprinted from [Mala 04], Copyright(2014), with permission from Elsevier).

of the local 3D structure of the stack, it also constrains the 2D FFDs in order to respect the local structural consistency and avoid the clustering of slices which is enforced by the data term. This data-based regularization term enables us to cope with the "banana problem" (c.f. Figure 4.17) [Stre 97, Mala 04, Daug 07, Gaff 11] by imposing local intrinsic structural consistency without having to resort to an external reference, such as block-face images (also called as photographic volume [Daug 07]) or in vivo MRI of the target region before cutting.

The data-based regularization term can also be written in terms of NCC as follows

$$\mathcal{E}_S(E_i, S_i, l_{\mathbf{p}}) = NCC(E_i', S_i) \tag{4.32}$$

where it is again possible to replace NCC with a more specific similarity term that takes the probabilistic nature of the structural representation into account. Finally, the deformation regularization term $\mathcal{R}$ in Equation (4.30) penalizes implausible or unnatural deformations and, in our case, depends only on the distance of all in-plane neighbors $\mathcal{N}(\mathbf{p})$ of each control point $\mathbf{p}$

$$\mathcal{R}(l_{\mathbf{p}}) = \sum_{\mathbf{q}\in\mathcal{N}(\mathbf{p})} ||\mathbf{d}_{l_{\mathbf{p}}} - \mathbf{d}_{l_{\mathbf{q}}}||^2. \tag{4.33}$$

The consecutive registration method has the advantage of being fast and performs well when the stack does not involve too complex structures or deformations (e.g. the synthetic data used in Section 4.4.2). However, because the deformations are not directly linked between slices, its performance deteriorates when the tissue deformations are complicated which is often the case in real histology data.

Simultaneous Registration In an attempt to improve the registration, we extended our algorithm to register the whole stack simultaneously, which is computational more expensive but also yields better results. For this we employ a method similar to the one proposed

in [Feue 11], which splits the MRF energy into one pair-wise term and two unary ones as follows

$$\mathcal{E}(l) = \sum_{i=1}^{g-1} \sum_{\substack{\mathbf{p}\in G^i \\ \mathbf{q}\in G^{i+1}}} \mathcal{E}_{data}(I_i, I_{i+1}, l_\mathbf{p}, l_\mathbf{q}) + \sum_{i=1}^{g} \sum_{\mathbf{p}\in G^i} \Big(\gamma \mathcal{E}_S(E_i, S_i, l_\mathbf{p}) + \rho \mathcal{R}(l_\mathbf{p})\Big). \tag{4.34}$$

In this formulation, the unary terms $\mathcal{E}_S$ and $\mathcal{R}$ remain the same as in Equations (4.32) and (4.33) respectively and act as regularizers in terms of both structural consistency and deformation field smoothness. However, to make the model even better, we recapitulate the data term $\mathcal{E}_{data}$ in a pair-wise manner by coupling the deformations of neighboring slices.

$$\mathcal{E}_{data}(I_i, I_{i+1}, l_\mathbf{p}, l_\mathbf{q}) = NCC(I'_i, I'_{i+1}) \tag{4.35}$$

Contrary to Equation (4.31), by comparing deformed version of slices, the term now links the control points (or more accurately the corresponding graph nodes) of neighboring slices, thus significantly increasing their possible alignment simultaneously. This approach is similar to the spring-based model proposed in [Saal 12] in that the deformations of neighboring slices are coupled.

This also means that the weighting γ should be treated differently in the simultaneous case (Equation (4.34)) than in the consecutive one (Equation (4.30)), because $\mathcal{E}_S$ is now only responsible for the regularization in terms of structural consistency and no more for the smoothness itself. For the consecutive one, we empirically found out that $\gamma = 1.0$ is a good overall value but can be increased if the structure map is of good quality, which depends on the amount of deformations in the slices. Whereas for the simultaneous registration, a lower value around 0.5 usually produces regularized but still structurally consistent results. ρ can be again chosen empirically by observing the amount of deformation smoothness. Large values lead to more global transformation while very small values result in unregularized deformation fields.

4.3.3. Whole Body MRI

As another application of the structure propagation method introduced in this work, we have chosen the creation of whole-body MRI. This type of imaging is becoming popular thanks to the improvements in MRI acquisition and the diagnostic quality of MRI. In clinic, whole-body MRI is useful for several studies ranging from oncology applications to forensics [Wach 08]. However, the main limitation of whole-body MRI is the long acquisition time and limited field-of-view making it difficult to have a single, consistent high resolution scan. An established way of overcoming this issue is performing a multi-station acquisition using, for instance, step-wise moving table. Although there are alternative acquisition techniques such as continuous table movement, the step-wise acquisition is technically less complicated compared to the former [Diet 08]. However, the main disadvantage, other than the relative longer acquisition time compared to continuous table movement, is the geometric distortion artifacts near the boundaries of the field-of-view.

In multi-station scanning, multiple scans covering overlapping regions of the body are successively acquired. The subimages are than fused together in their overlapping regions. The geometric distortions in the regions far from the iso-center often leads to discontinuities

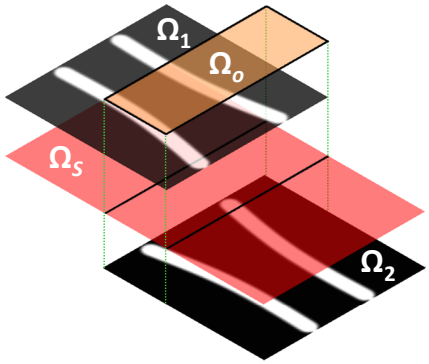


FIGURE 4.18.: Illustration of stitching, Ω_s, overlap, Ω_o, and image, $\Omega_{\{1,2\}}$, domains. Note the distortion in Ω_o.

in the composite volume. Wachinger et al. [Wach 08] proposed a deformable registration approach to handle such discontinuities. They suggested to iteratively register overlapping areas to their linearly weighted average. Although, the resulting composite images look visually pleasing, it remains a question whether linear average converges to the "natural" shape of the structures in the overlap.

We instead propose to utilize the reliable information in the undistorted regions to support the fusion of the distorted regions. To do this, we propose to propagate the reliable structures in the overlapping region and use the structural maps for regularization purposes similar to our approach for histology reconstruction.

Given two subimages I_1, I_2 with a certain overlap between them as illustrated in Figure 4.18, using the same philosophy of 2D image stitching, we pose the fusion in the overlapping region as an optimization problem. In a way similar to [Wach 08], we define

$$\hat{T}_{1,2} = \arg\min_{T_{1,2}} \int_{\Omega_o} \mathcal{E}(E_1, S_1, T_1, E_2, S_2, T_2, \mathbf{x}) d\mathbf{x} \tag{4.36}$$

where $\mathbf{x} \in \mathbb{R}^N$ is pixel position, $T_{1,2}$ is the set of transformation parameters for T_1 and T_2 bringing the two images into a spatial alignment in Ω_o, $\mathcal{E}(\cdot)$ is a cost function based on the edge images E_i and the structural representations S_i. We look for one transformation T_i per image E_i since, in our case, both images will have possibly unreliable information, i.e. geometric distortions, within their overlapping regions. Different from [Wach 08], we warp each edge image E_i towards the weighted average, $\mathcal{W}$, of the structural probability maps, S_i, in Ω_o. $\mathcal{W}$ is defined as

$$\mathcal{W}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{x} \in \Omega_o \\ S_1(T_1(\mathbf{x})), & \text{if } \mathbf{x} \in \Omega_1 \setminus \Omega_2 \\ S_2(T_2(\mathbf{x})), & \text{if } \mathbf{x} \in \Omega_2 \setminus \Omega_1 \end{cases} \quad \text{with} \quad f(\mathbf{x}) = \sum_{i=1}^{2} \mu_i(\mathbf{x}) S_i(T_i(\mathbf{x})) \tag{4.37}$$

where $f(\mathbf{x})$ is a weighting function with $\mu_i(\mathbf{x})$ being linear in $[0, 1]$, which is 0 at $\partial\Omega_i$ and 1 at $\partial(\Omega_i \setminus \Omega_o)$ on the stitching side. We use the Normalized Cross Correlation (NCC) as the

similarity measure between $\mathcal{W}$ and $E_i(T_i)$. Now Equation (4.36) becomes

$$\hat{T}_{1,2} = \arg\min_{T_{1,2}} \sum_{i=1}^{2} \int_{\Omega_o} \xi(\mathcal{W}(\mathbf{x}), E_i(T_i(\mathbf{x})))d\mathbf{x}. \tag{4.38}$$

We alleviate the optimization by defining free-form deformation (FFD) grids [Ruec 99], G_i, in Ω_s in order to parametrize T_i in terms of cubic B-splines. Displacement of a grid point $\mathbf{p}$ by $\mathbf{d_p}$ induces a deformation of the underlying structure. FFD-based image registration is discussed in more detail in Section 3.1.5. An MRF is obtained by assigning nodes to grid points and defining a discrete labeling such that an assignment of a label to a grid point induces a displacement of that point. Definition of the first-order MRF's is given in Equation (3.11) and we repeat here for completeness.

$$E(\mathbf{l}) = \sum_{\mathbf{p}\in G} V_{\mathbf{p}}(l_{\mathbf{p}}) + \sum_{(\mathbf{p},\mathbf{q})\in\mathcal{N}} V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}). \tag{4.39}$$

Our data term in Equation (4.38) corresponds to the unary term in a first order MRF, while an explicit grid regularization can be enforced by defining a regularization term as the pairwise potential of the MRF. The unary term is defined as

$$V_{\mathbf{p}}(l_{\mathbf{p}}) = \int_{\Omega_o} \hat{\eta}(|\mathbf{x}-\mathbf{p}|) \cdot \xi(\mathcal{W}(\mathbf{x}), S_i(T_i(\mathbf{x}+\mathbf{d}_{l_{\mathbf{p}}})))d\mathbf{x} \quad \forall \mathbf{p} \in G_i, \quad i \in \{1,2\} \tag{4.40}$$

where $\hat{\eta}(|\mathbf{x}-\mathbf{p}|)$ is a B-spline weighting term controlling the effect of $\mathbf{x}$ on $\mathbf{p}$ [Gloc 08]. The pairwise terms, $V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}})$, are used for explicit grid regularization and are defined as

$$V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}) = \rho \left\| (\mathbf{d}_{l_{\mathbf{p}}}) - (\mathbf{d}_{l_{\mathbf{q}}}) \right\|^2 \qquad \forall \mathbf{p}, \mathbf{q} \in G_i \wedge (\mathbf{p}, \mathbf{q}) \in \mathcal{N}. \tag{4.41}$$

Since the data term is defined over Ω_o, the resulting deformations might lead to discontinuities at the boundaries between the overlapping and the non-overlapping regions of each image. Finally, deformations are smoothly propagated beyond Ω_o due to the incorporated explicit regularization term as well as the implicit smoothness properties of the FFD model, which is defined over Ω_s.

For finding the optimal labeling, Fast-PD [Komo 07, Komo 08], an efficient MRF optimization is used. We put everything in a multi-resolution setting to avoid local minima where the optimal labeling is communicated between levels. Note that both structure propagation and stitching take place in a multi-resolution setting. During structure propagation, this ensures that structures present in coarser scales are considered first for the optimization whereas structures at finer scales are used for refining the local deformations.

4.4. Experiments and Results

For the evaluation of the proposed structure propagation framework, several experiments have been conducted using synthetic and real medical images for each application discussed in the previous section. Synthetic images were used to demonstrate the capacities of the proposed approach in a controlled environment where different challenging scenarios can be synthetically produced. Real medical image datesets are used to assess the performance of the proposed technique on clinical data sets.

4.4.1. 2D Image Stitching

Field-of-view extension by stitching partial observations together by means of registration can be quite challenging. In particular, this is the case where not only there is no overlap but also there is a physical gap between the subimages to be stitched together. Similar challenges exist when the images to be stitched are too noisy or corrupted. It can be further difficult if the information in the overlapping region can not be reliably used for the stitching due to geometric distortions or severe deformations. We have created several synthetic image pairs in order to evaluate the performance of the stitching of subimages in such scenarios without violating the smoothness and the continuity of structures across subimage borders. In the following, we show and interpret the results of synthetic experiments.

The proposed method has also been applied to the stitching of real medical images. To do this, we have extracted 2D slices from a two-photon microscopic image data set acquired from a rat brain. Again, as in the synthetic experiment case, slices were cut into two. Furthermore, four microscopic image pairs with overlapping subimages to be stitched are used. Both for the experiments with the synthetic as well as the real microscopic images, manual linear transformations were applied to one of subimages. The goal of these experiments is to demonstrate that, using structure propagation, it is possible to recover transformations in various challenging scenarios. These experiments are important in that they showed the feasibility of the proposed method for being used for the stitching of real medical images.

For the evaluation of results, in addition to the visual assessment, we have compared the stitched images to the ground truth data using a correlation technique. Moreover, we evaluated the performance of the method in recovering the applied individual linear transformation parameters. Transformations are designed by varying one of the parameters at a time in a predefined range while keeping the others fixed to their initial values. In this way, we assess the sensitivity of our method to the variations in transformation parameters. In an ideal recovery of the parameters, the sum of the applied and the recovered translational or rotational parameters should add up to zero. For the scaling parameters, the multiplication of the applied and the recovered parameter values should be equal to one. Moreover, since a variation in a single parameter can also be compensated by variations in other parameters during optimization, we have also evaluated the recovery error for transformation matrices as a whole where all parameters (including shearing) have non-trivial values.

As an additional quantitative evaluation of the proposed method, we have developed a continuity index called *structural continuity index* (SCI), which is designed to evaluate the smoothness of structures across stitched subimages. We compute SCI again using tensor voting where structures extracted from one side of the stitching boundary *after* the alignment vote for the possible locations of structures on the other side. The real structures are then correlated with the estimated ones via tensor voting using the same notation in Section 4.2. SCI can be formulated as

$$SCI = \sum_{\substack{i,j\in\{1,2\}\\ i\neq j}} \sum_{\mathbf{p}\in\mathbf{P}_{io}} \frac{S_i(\mathbf{p})S_j(\mathbf{p})}{|\mathbf{P}_{io}|} \tag{4.42}$$

where $\mathbf{P}_{io} \subset \mathbf{P}_i$ is the set of structure points in Ω_{io}. Note that this measure is in fact a cross

correlation where only the detected structure points are used in the computation. This measure can also be used as a cost function during registration. Here, we use it only for evaluation in order to have an objective testing.

The robustness of the proposed approach against noise was also tested where varying degrees of white noise was added to a pair of synthetic images before applying any initial transformation. The magnitude of the noise added ranges between 0%-50% of the maximum image intensities.

There are several parameters to be set both for the tensor voting and for the employed registration method. First of all, the size of the overlapping region Ω_{oi} to be established for each subimage in the stitching direction was set to be 20mm. This was considered to be sufficient for performing a registration on the used datasets that will be described later. This value can be adapted depending on the application and the image acquisition settings. A prior knowledge about the acquisition setup can be exploited for this purpose. Moreover, in cases where subimages are already touching each other, if the images to be stitched are well structured, then, smaller sizes for the overlapping regions can be preferred for increased computational performance. A value of 40 was used for the scale parameter σ in Equation (4.10). For this experiment, this value guarantees that a structure point 20mm inside a subimage can also vote for an image point 20mm inside of the other image although the influence decreases by the increasing distance. As a rule of thumb this value can be chosen as the double of the estimated overlap size. The edges are detected using the recursive edge detection method proposed by Deriche [Deri 87].

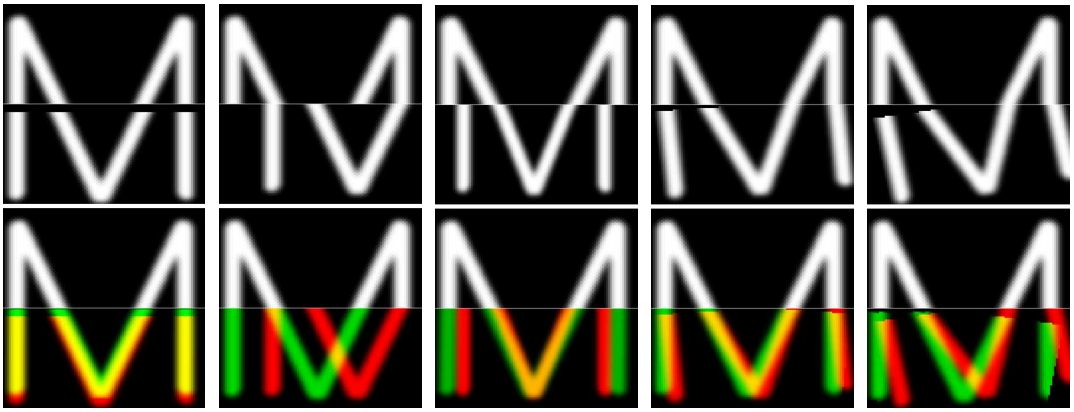
For defining the search interval for the affine transformation parameters in Equation (3.12), we set the search range to ± 40mm for the translational and rotational components and to 1 ± 0.20 for the scaling components. Two pyramid levels are used for all synthetic experiments except for the experiments with noisy images where only a single level is used. The number of pyramid levels for the experiments with microscopic images is set to 2. The size of the discrete label space $|\mathcal{L}_i|$ for each parameter is set to be 7 which is refined by a factor of 0.6 as suggested in [Ziki 10] in every pyramid level. The maximum number of iterations on each pyramid level is set to 512 while, in practice, the registration is terminated if the change of energy is very small in 5 consecutive iterations.

For all experiments, a PC equipped with Intel® Core™ i7 2820QM, 230 GHz and 12GB RAM was used. It takes about 2s to stitch an image of size 512x512 pixels with 1.6x1.6mm pixel size.

Synthetic Images

Synthetic experiments were conducted to demonstrate the capabilities of the proposed method for the stitching of subimages without overlap in different scenarios. In order to make sure that the method successfully copes with the lack of an overlapping region, we simply cut a full image into two pieces and created simple scenarios by applying affine transformations to one of the pieces. Given a full 2D image with smooth structures, it was cut into two subimages I_1 and I_2, which are stitched again to obtain the original image I. Note that by simply cutting into two, we do not allow any overlap between the subimages. This is important because this is one of the main challenges that we are trying to address in this work.

In Figure 4.19, we demonstrate how stitching can be done in different scenarios. A



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

FIGURE 4.19.: Image stitching without overlap. A synthetic image of "M" letter is cut into two and the lower piece is translated, scaled or rotated which is then restored using the proposed approach. In the first row, the initial stitching before the alignment is shown. The gray horizontal line delineates the boundary between the pieces. In the second row, the final stitching is overlaid onto the initial one where the red and green colors show the lower piece of the letter before and after the alignment, respectively. In (a)-(b), translations in horizontal and vertical directions, respectively, are applied to the lower piece which are then successfully restored using the proposed stitching method. Note the physical gap in (a) created by applying a vertical translation. (c) demonstrates the ability to restore a change in scaling. In (d)-(e), a small and a relatively large rotation is recovered. Again, there is a gap as a result of the applied rotation which makes the stitching much more challenging than the case where there is no gap between the pieces to be stitched together.

synthetic image of the letter "M" of size 102x124 pixels with 1mmx1mm uniform pixel size was divided into two and the lower piece was translated, scaled or rotated as shown in the figure. These transformations were then recovered using the proposed approach. In the first row, the initial stitching before alignment is shown. The gray horizontal line stands for the boundary between the subimages. In the second row, the final stitching is overlayed onto the initial one, where the red and green colors show the lower piece of the letter before and after the alignment, respectively. In Figures 4.19(a) and 4.19(b) translations in horizontal and vertical directions, respectively, are applied to the lower piece which is successfully restored using the proposed stitching method. Note the physical gap in Figure 4.19(a) which is caused by the applied vertical translation. Figure 4.19(c) shows the ability to restore a change in scaling. In Figures 4.19(d) and 4.19(e), a small and a relatively large rotation, respectively, are applied and then recovered. Again, there is a gap as a result of rotation which makes stitching much more challenging than the case where there is no gap between the pieces to be stitched together. In all cases, it is visually obvious that a stitching is possible in different scenarios, where the common issue is the lack of image overlap.

For the quantitative evaluation of the proposed method, several transformations, in which only one of the parameters t in Equation (3.12) was varied, were applied to the lower piece of the letter image. In each case, only one of the parameters changed its value in a

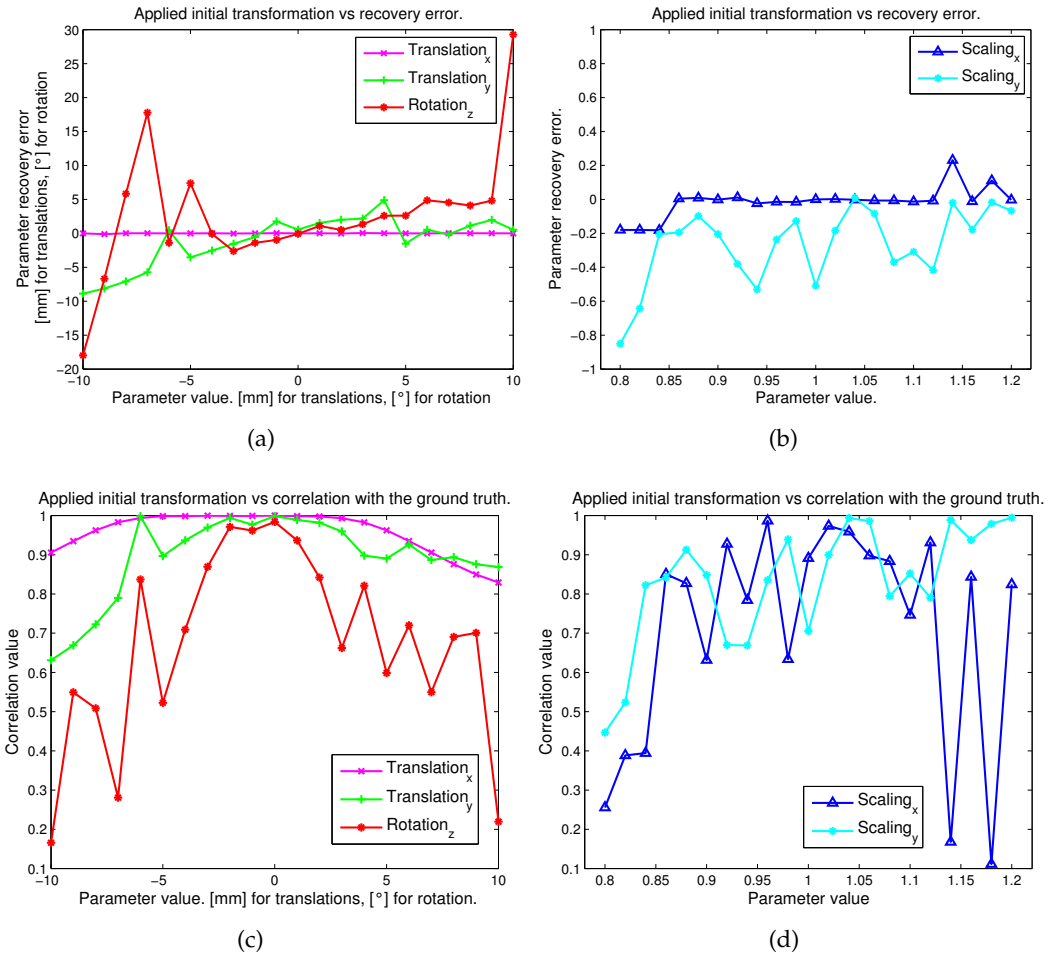


FIGURE 4.20.: Parameter recovery and ground truth correlation experiments. Several transformations, which differ from the identity only by the variation of one of the parameters in p in Equation (3.12), have been applied to the lower piece of the letter image shown in Figure 4.19. In each case, only one of the parameters changed its value in specific interval which was chosen to be ± 10 for translations (mm) and rotations (degrees) and 1 ± 0.20 for scalings. 21 equally spaced values for each parameter were sampled from the given interval. In each case, the error between the applied and the recovered transformation parameter was computed for (a) translation and rotation and (b) scaling. (c)-(d) Correlations of the aligned subimage with the ground truth subimage for (c) translation and rotation and (d) scaling.

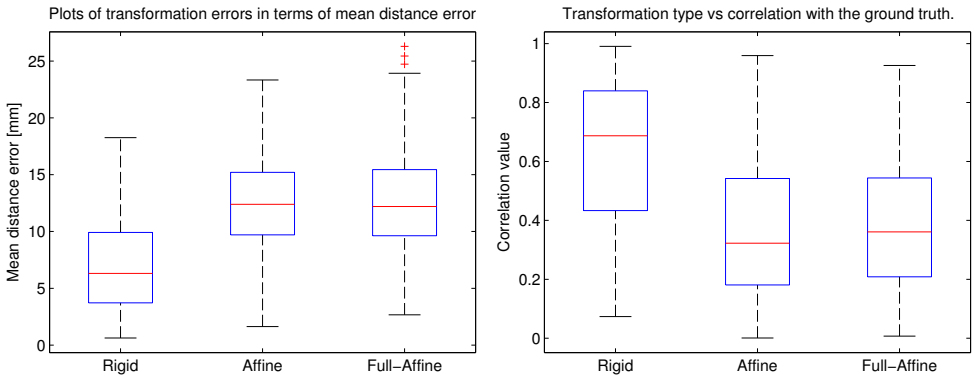
specific interval which was chosen to be ± 10 for translations (mm) and rotations (degrees) and 1 ± 0.20 for scalings. Shearing parameter was not tested here as it makes no difference without a non-identity scaling. Instead, it was tested together with the other parameters while evaluating the transformation recovery performance which is explained later in this section. 21 equally spaced values were sampled from the given interval. Afterward, for each case, the error between the applied and the recovered parameter was computed. For translation and rotation parameters, the absolute value of the summation of the applied

and the recovered parameter values is computed whereas for scaling the deviation of their ratio from identity is used as a measure of error. The error plots for each parameter in the specified interval are shown in Figures 4.20(a) and 4.20(b).

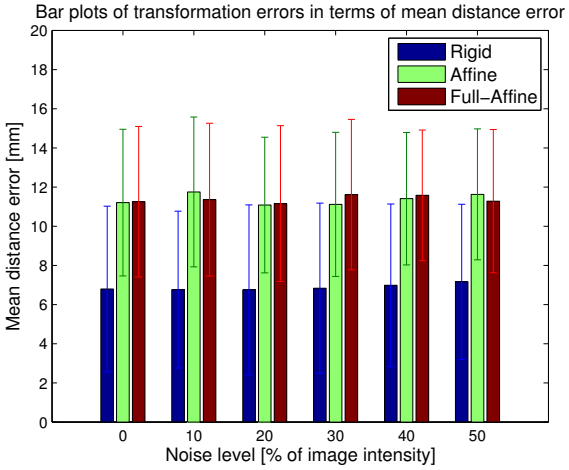
Several observations can be made based on the results. First of all, it is obvious that the most problematic parameter is the rotation. Although the parameter recovery error is reasonable in the vicinity of 0 degrees, i.e. small rotations, it gets quite large towards the marginal values. Another observation is that the recovery error for the translation parameter in the direction orthogonal to the stitching boundary (y) is quite larger than its counterpart in the parallel direction (x). Same behaviour can be observed for the scaling, too. The reason for this is obvious; the physical gap created in both cases makes the recovery more difficult. The recovery error for the other parameters are almost zero, meaning that all the parameters are successfully restored even for the marginal parameter values.

We also compute the correlation of the aligned subimages with the ground truth before cutting to see whether the alignment was correct in terms of the image intensities. Figures 4.20(c) and 4.20(d) show the correlation results for each parameter for the specified value range. What is noteworthy is that although the correlation results for the rotation and the translation parameters are consistent with the parameter recovery error shown in Figure 4.20(a), this is not the case for the other parameters especially for the scaling parameter in the direction parallel to the stitching boundary, which can be observed in Figure 4.20(b). This indicates that the misalignment caused by the variation of this parameter has been mostly compensated by the variations in other parameters. This is quite possible as all of the parameters were optimized simultaneously for the best alignment. Furthermore, we take the middle of the stitching border as the origin of the image coordinate space. For this reason, although a scaling in the direction orthogonal to the stitching boundary (y) does not lead to a padding, a scaling in the parallel direction (x) definitely results in padding during registration. Therefore, a direct correlation of image intensities might be misleading in this case. Yet another observation that can be made from Figure 4.20 is that, in almost all cases, the alignment performance decreases towards the marginal values of the varied parameters. This is an expected behavior since the performance of every registration algorithm highly depends on the initial position.

In addition to the evaluation in terms of individual parameter errors, we also compute the recovery error for transformation matrices as a whole. The error is computed as the mean distance of a set of points before applying an initial transformation and after restoring it. The point set consists of 200 points drawn randomly from a square with a side length of 200mm centered at the origin. Ideally, a successful recovery should lead to zero mean distance error. The applied transformations are composed by drawing parameter values randomly from the following ranges; translations from ± 10mm, rotation from $\pm 10^{\circ}$, scalings from 1 ± 0.20, and finally the shearing from $\pm 10^{\circ}$. Three cases were tested; only rigid parameters (rigid), rigid plus anisotropic scaling (affine) and finally including all six parameters (full-affine). In each case, initial transformations were composed of only relevant parameters and only those parameters were optimized for. 200 random registration experiments were performed in each cases. The statistics of transformation recovery errors are shown in Figure 4.21(a). From the figure, it is obvious, as expected, that the performance decreases as the number of degrees of freedom increases. The median error for each case are 6.3mm, 12.4mm and 12.2mm, respectively. It is interesting that including shearing in transformations did not introduce a significant change in the performance. Non-trivial



(a) (b)



(c)

FIGURE 4.21.: Stitching experiments testing the sensitivity of the proposed technique to non-trivial transformations. The applied transformations are composed by randomly drawing transformation parameters from the following ranges; translations from ± 10mm, rotation from $\pm 10^{\circ}$, scaling range from 1 ± 0.20, and finally the shearing from $\pm 10^{\circ}$. Three cases were tested; only rigid parameters (rigid), including anisotropic scaling (affine) and finally including all six parameters (full-affine). In each case, 200 random initial transformations were composed of only relevant subset of parameters and only those parameters were optimized for. (a) Box plots show the transformation recover error statistics in each case. It is clear that the error increases as the number of degrees of freedom increases. Furthermore, adding shearing parameter did not make a significant difference in the overall performance. (b) Mean correlations of the aligned subimage with the ground truth subimage for different transformations. (c) Mean distance error in the presence of white noise. The added noise ranges from 0% to 50% of image pixel intensities. The results present parallelism with (a) and (b) with a good performance in rigid case and a decreased performance for affine and full-affine transformations. Moreover, it is obvious that the noise did not affect the performance of the proposed method.

random transformations were also used for ground-truth correlation of image intensities. The results shown in Figure 4.21(b) bear parallelism with the results of transformation recovery error as shown in Figure 4.21(a). The median correlation is about 70% for rigid case dropping to 35% as the complexity increases.

The proposed method was further assessed in terms of its robustness against noise. To this end, varying degrees of white noise was added to the initial pair of synthetic images before applying any transformation. The noise added ranges between 0%-50% of image intensities with 10% bins. For each noise level and for each transformation type (i.e. rigid, affine, full-affine), 100 random transformations were applied to one of the synthetic images which was then recovered by the proposed approach. Since a multi-resolution setting would eliminate the added noise during image pyramid creation, we performed registration only on the finest level to make sure that the noise is present during structure propagation. The results shown in Figure 4.21(c) demonstrate the robustness of the proposed approach to the amount of noise contained in subimages, which also confirms the the robustness of structure propagation. Performance with respect to the order of applied transformations has similar patterns with the previous experiments.

The last method that we use for the assessment of the stitching performance is the structure continuity index, SCI, as defined in Section 4.4.1. Here, for the same experiments described above, we computed the SCI before and after a stitching was performed. Statistics on the SCI values computed from 21 experiments for each parameter are presented in Figure 4.22. Figure 4.22(a) shows the box plots for the SCI values before doing a stitching. Parameters tx, sx, i.e. translation and scaling in the direction parallel to the stitching boundary, respectively, as well as the rotation parameter, rz, have a low median SCI value, which is mainly due to the fact that the initially applied transformations resulted in misalignments between the structures of subimages. Such misalignments occurred less in the cases of ty and sy, i.e. translation and scaling parameters in the direction orthogonal to the stitching boundary. This is mainly because the structures are either scaled or shifted in the stitching direction which did not influence the structure alignments that much. SCI values were also computed after the stitching as presented in Figure 4.22(b) and the changes in SCI values is shown in Figure 4.22(c).

We can make several observations looking at Figure 4.22(c). First of all, one can say that there has been an improvement in the estimation of the translation and the scaling parameters in the direction parallel to the stitching boundary. On the other hand, there has been little or no improvement for their counterparts in the orthogonal direction. Rotation parameters also present an increased SCI value. One can conclude that stitching performance for correcting the misalignments caused by the variations in in translation and scaling parameters in the direction orthogonal to the stitching boundary is not as good as the performance for correcting other types of misalignments. However, one should note that the structure continuity index, SCI, is based on the correlation between the votes cast by structure points in one of the subimages and the actual structure points in the other subimage. A transformation in the direction of stitching usually leads to a loss of structure points, which also influences the value of the SCI. Therefore, SCI should not be taken as a sole performance indicator when evaluating the proposed method.

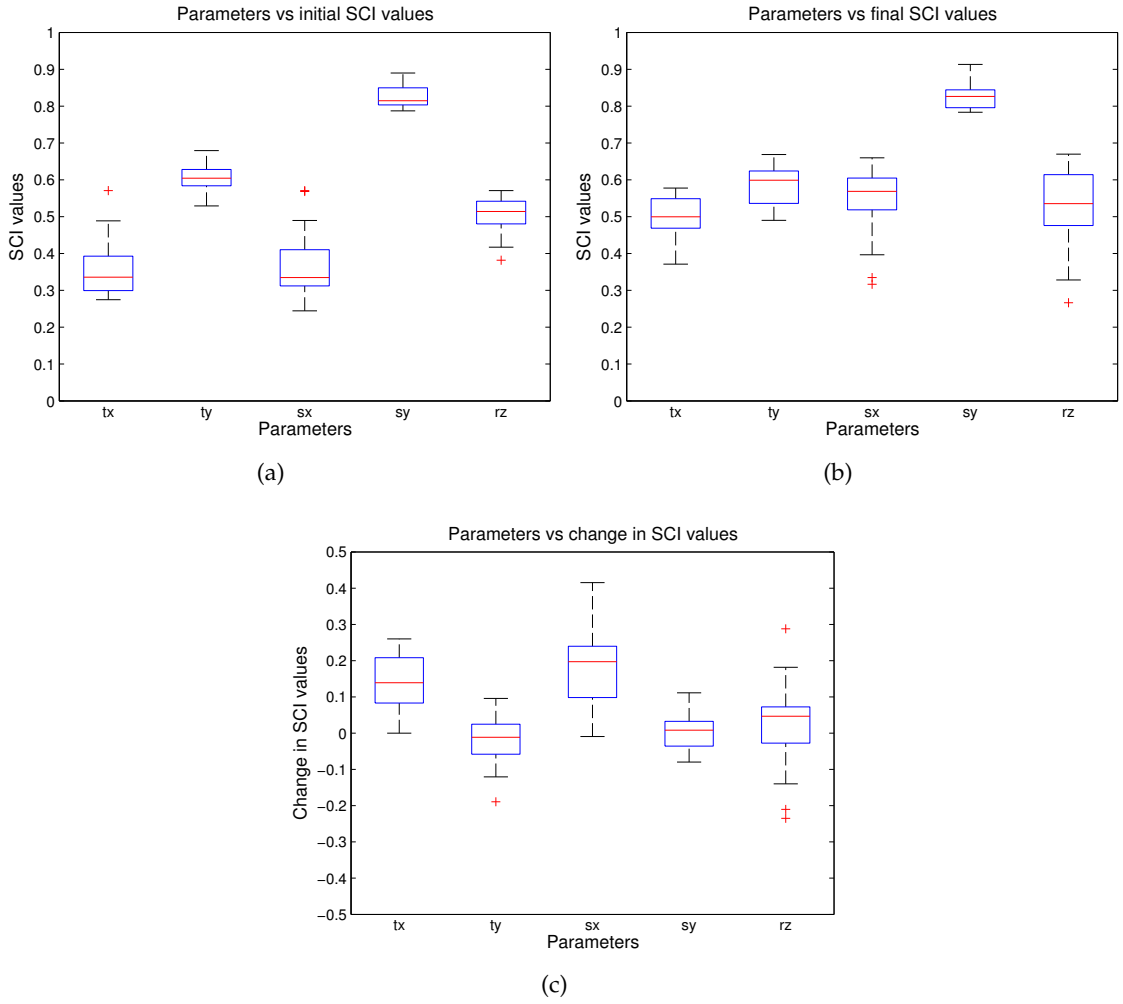
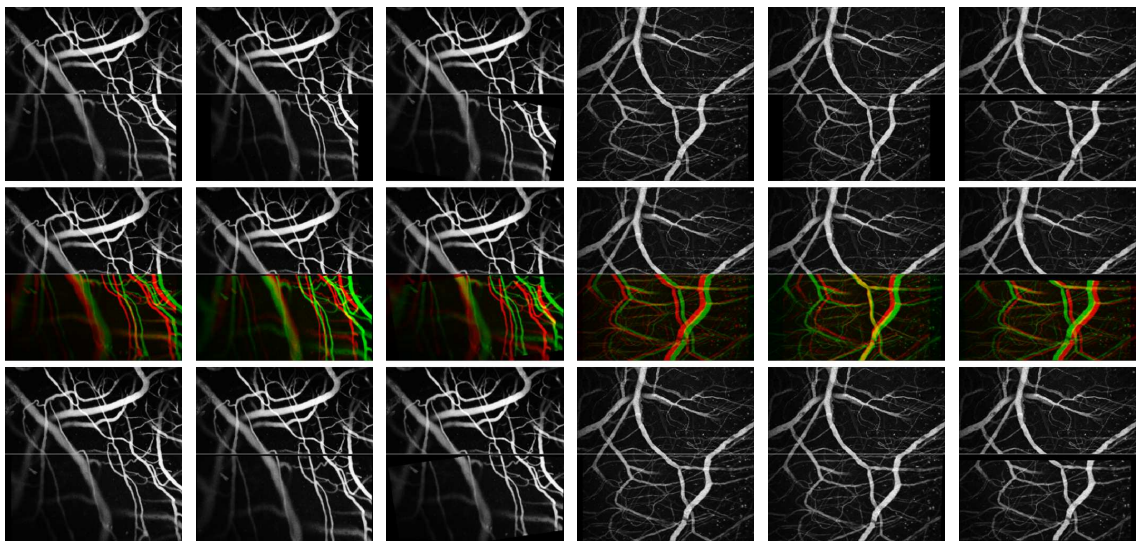


FIGURE 4.22.: Stitching performance in terms of the structural continuity index (SCI). SCI was computed for each stitching task as shown in Figure 4.20. For each parameter there are 21 experiments where the parameter takes a value from an interval as discussed in Section 4.4.1. Here, we show the box plots of SCI values for each parameter. (a) Initial SCI values before performing a stitching. (b) Final SCI values after stitching. (c) Change in SCI after stitching. Looking at (c), one can say that there has been a positive change for the translation and the scaling parameters in the direction parallel to the stitching boundary. On the other hand, there has been little or no improvement for their counterparts in orthogonal direction. Rotation parameter also presents with an increased SCI value.

Microscopic Images

Experiments have also been conducted on medical images to show the effectiveness of the proposed technique on real images. To this end, we have extracted 2D slices from a two-photon microscopy data set, which was taken from a mouse brain. The images were taken to examine the vasculature in the brain. There are two slices of size 512x512 pixels with

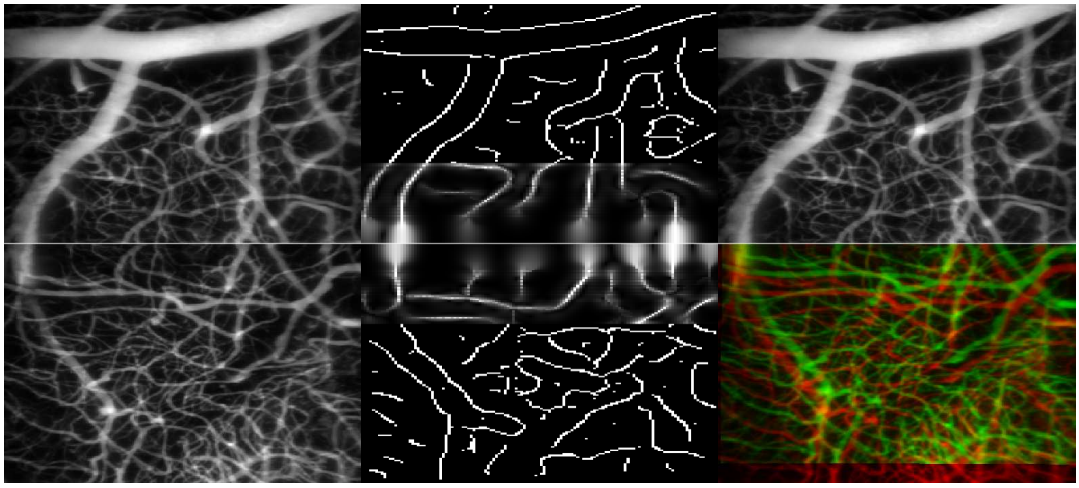


(a) Translation (b) Scaling (c) Rotation (d) Translation (e) Scaling (f) Trans.+Gap

FIGURE 4.23.: Stitching experiments on 2D microscopy images. There are two microscopy image pairs P1 [(a)-(c)] and P2 [(d)-(f)] used for the experiments. The first row shows the initial mosaics obtained by applying affine transformations. In the second row, the final mosaics after the alignment is overlayed onto the initial one where red and green colors are used for initial and final versions of the aligned subimage. The last row shows the final mosaics without overlay. In (a)-(b), similarly in (d)-(a), misalignments caused by the variations in the translation and scaling parameters, respectively, are restored. In (c), the correction of a misalignment due to rotation is demonstrated. Finally, a misalignment caused by a translational transform in the presence of a physical gap between the subimages is recovered in (f). See the text for the interpretation of this figure.

1.16μm uniform pixel size. Again, as it was done in the previous synthetic experiments, the slices were cut into two pieces where one of the pieces underwent a transformation in order to induce a misalignment of structures across the cutting border. Figure 4.23 shows a series stitching experiments on the created pairs demonstrating the performance of the proposed method on microscopic images with synthetic misalignments.

In order to test the proposed registration technique using structure propagation on a real data without synthetic transformations and also to compare it to a competitive registration method, four microscopic image pairs, with subimages of 256x256 pixels resolution and 2.32μm uniform pixel size, were used for the experiments. An example pair along with corresponding structure images are shown in Figures 4.24(a) and 4.24(b). The overlap region was further removed from one of the subimages for each pair in order to assess the proposed approach in no overlap case. Ground truth rigid transformations were obtained by using corresponding landmarks manually annotated in both subimages which resulted in overlap sizes of 53.47μm, 62.76μm, 55.79μm and 62.76μm, respectively. For the experiments with overlap, three different scenarios were used for registration with and without structure propagation. First, a good initialization of transformation close to the ground truth was provided in the beginning. Secondly, white noise in the range of 50% of the im-



(a) Before alignment (b) Structure images (c) After alignment

FIGURE 4.24.: Pair of microscopy images used for the evaluation of the proposed method in comparison to the classical approach. (a) Before an alignment was performed. (b) Corresponding structure images with propagated structures. (c) After the subimages were aligned using the proposed method. Aligned subimage is overlayed onto the original one.

age intensity were added to the images. Finally, subimages were placed only side-by-side where stitching boundaries were touching each other. The last scenario was further used for aligning without overlap. Again, mean distance error between the landmarks was used for the quantitative evaluation of the registration performance. For structure propagation, the value of scale parameter was fixed to 50μm for each pair.

The errors shown in Figure 4.25 indicate the comparable performance of the proposed approach in the presence of an overlap when a good initial estimate of the transformation is provided. Our approach already outperforms the classical approach when the overlap region is degraded with white noise. However, when the alignment is initialized by placing the subimages only side-by-side (i.e. touching), the classical approach fails to find a transformation close to the ground truth whereas the current approach can still successfully align the subimages. Moreover, it was not possible to use the classical approach in no-overlap case where, again, a good performance was obtained by using structure propagation. The results reveal that registration with structure propagation can be successfully used also for correcting misalignments in real images.

Based on the previous experiments, it is clear that the physical gap between subimages has a negative influence on the stitching performance. Since one of the goal of the proposed techniques it to overcome the limitations caused by the physical gap between the subimages, we have conducted further experiments to test the sensitivity of the technique to the size of physical gap between the subimages. To this end, we have manually introduced a physical gap of varying size to the lower subimage of the microscopic image pair P2 as illustrated in Figure 4.23(f). There are 21 different gap sizes sampled from the interval [0μm,20μm]. For each gap size, misalignments of varying magnitude were introduced by applying 21 different translational transformations from the interval $\pm$20μm in the direc-

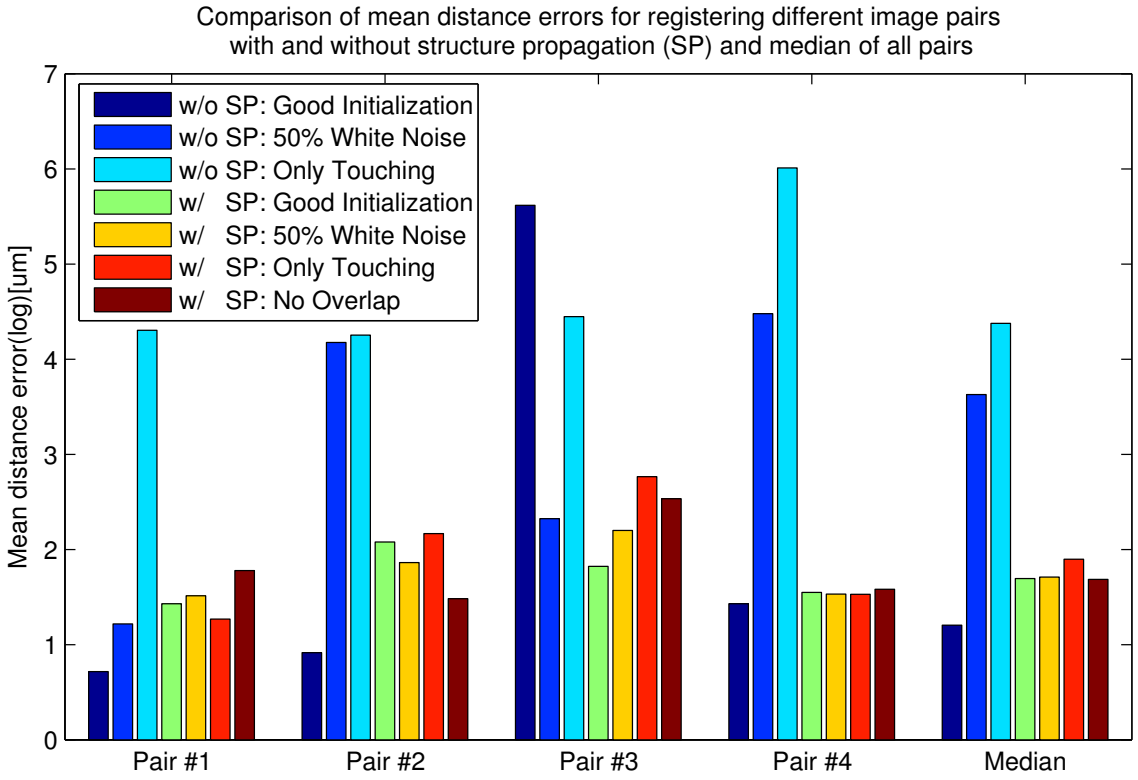
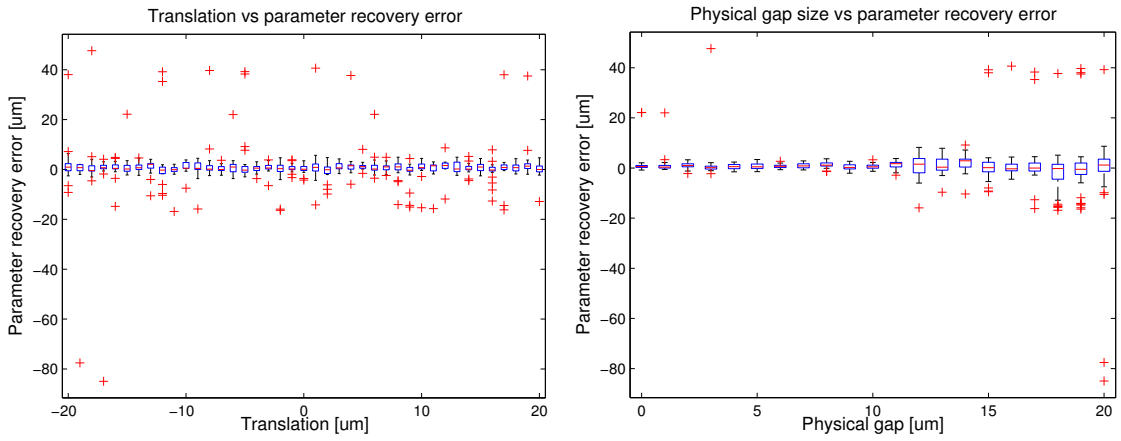


FIGURE 4.25.: Comparison of the proposed technique using structure propagation to registration without structure propagation. Four different microscopic image pairs are used. For the experiments with overlap, three different scenarios are used. First, a good initialization close to the solution is provided in the beginning. Secondly, white noise in the range of 50% of the image intensity is added to the images. Finally, subimages are placed only side-by-side where stitching boundaries are touching each other. The last scenario is further used for aligning without overlap. Mean distance error of a randomly drawn point set is used for the evaluation. Note the errors are shown in log scale for better visualization. The results indicate the good performance of the proposed approach in various scenarios with and even without overlap whereas it is only possible to do an alignment in the presence of an overlap and when a good initialization is provided.

tion parallel to the stitching boundary. We have calculated the parameter recovery error in order to evaluate the influence of the amount of translation and the size of physical gap, respectively. Figure 4.26(a) shows that the stitching performance is not influenced too much by the amount of misalignment in the presence of a physical gap. However, when the parameter recovery error is plotted against the size of the introduced physical gap as shown in Figure 4.26(b), it becomes obvious that the stitching performance decreases with the increasing gap size. This is again an expected behavior since initially it was claimed that a non-overlap stitching is possible if the subimages are sufficiently close to each other. This is a much looser condition than the one that classical stitching techniques often require. It is still noteworthy that with the proposed technique it is still possible to stitch two subimages even if there is a physical gap on the order of 20μm between them. However, looking at



(a) (b)

FIGURE 4.26.: Stitching experiments testing the sensitivity of the proposed technique to the physical gap size. 21 different gap sizes from the interval [0μm,20μm] are introduced to one of the subimages in the direction of stitching as illustrated in Figure 4.23(f). For each gap size, 21 translational transforms from the interval $\pm$20μm are applied in the direction parallel to the stitching border. In each case, parameter recovery error is computed. In (a), the parameter recovery error is plotted against the amount of translation. It is clear that there is not much correlation between the amount of translation and the associated recovery error in the presence of a physical gap. However, when plotted against the size of the introduced physical gap as shown in (b), it becomes clear that the parameter recovery error increases with the increasing gap size, which is an expected behavior. It should be noted that it is still possible to perform a stitching even if the size of the gap is on the order of 20μm, which is the novel side of the proposed method.

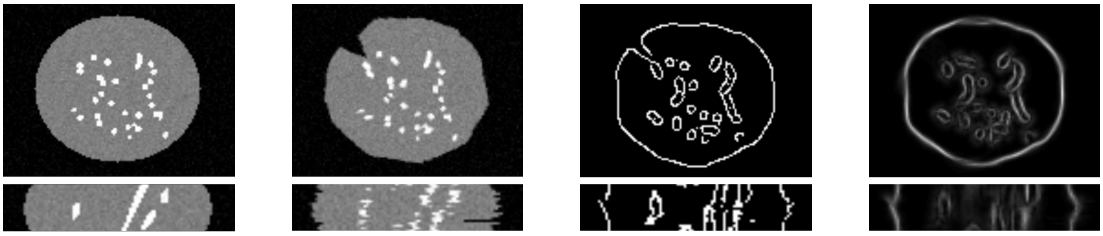
Figure 4.26(b), in order to guarantee a successful stitching, one can allow a physical gap of around 12μm as a threshold after which the deviation of the error from the median starts to increase.

4.4.2. Deformable Histology Reconstruction

In this section, we demonstrate the two histology reconstruction approaches based on structural probability maps, i.e. consecutive and simultaneous, that we have described in Section 4.3.2. To this end, in Section 4.4.2, we describe our experiments on synthetic and mouse brain data sets showing that using structural probability maps yields similar results when compared to techniques that operate reference-based on synthetic data sets.

Synthetic Dataset

The synthetic data consists of a stack of 20 slices with a resolution of 128x96 and a pixel spacing of 1mm. An example of an original slice is shown in Figure 4.27(a). The slices contain a circular tissue which grows towards the middle and several skewed vascular structures. Each slice is deformed by a random FFD with a maximum displacement of



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

FIGURE 4.27.: Deformable reconstruction using synthetic dataset. Top and bottom rows show respectively the axial and coronal views of the volume in each case. (a) undeformed original stack, (b) stack corrupted with non-linear deformations, (c) volume reconstructed using the edge maps of corrupted slices, (d) volume reconstructed using the structural probability maps of corrupted slices.

Method	EE		AE	
	Error (mm)	STD (mm)	Error (°)	STD (°)
Consecutive Registration w/o SPM	2.91	1.78	60.31	31.41
Consecutive Registration w/ SPM	1.69	1.25	45.39	26.63
Simultaneous Registration w/o SPM	2.37	1.27	57.31	33.17
Simultaneous Registration w/ SPM	1.68	1.13	45.95	26.89

TABLE 4.1.: End point error (EE) and angular error (AE) of the presented methods using structural probability map (SPM).

5mm for each node and the corresponding deformation field is saved as ground truth for evaluation. Additional tears are introduced in randomly selected slices to simulate the real histology cutting process. Distorted slices are shown in Figure 4.27(b). This is the same synthetic dataset used in [Feue 11].

In order to quantify the results of our approach, we calculated the absolute end point error (EE) and the relative angular error (AE) between the resulting deformation fields and the ground truth fields [Bake 11]. Table 4.1 shows the errors after the application of different reconstruction approaches discussed in Section 4.3.2.

For the registration we used a grid spacing of 15mm and 2 grid levels. In the consecutive case both ρ and γ in Equation (4.30) were set to 1.0 in order to put more emphasis on the structural probability map (SPM). For the simultaneous method (Equation (4.34)), we used 0.5 for both instead. The results in Figure 4.28 show the same coronal slice as Figure 4.27 after reconstructing by using different methods.

The consecutive registration without the SPM regularization performs the worst in terms of the error but also in its visual appearance (c.f. Figure 4.28(a)). All curvilinear structures get straightened in stack direction and especially two vessel structures on the right side cluster into four distinctive structures. This is a case for the banana problem that we discussed in Figure 4.17. Extending this with the use of SPM for structural regularization maintains the outer round shape but also preserves the curvilinearity of the vessels inside (Figure 4.28(b)). This is similar for the simultaneous method: while the unregularized

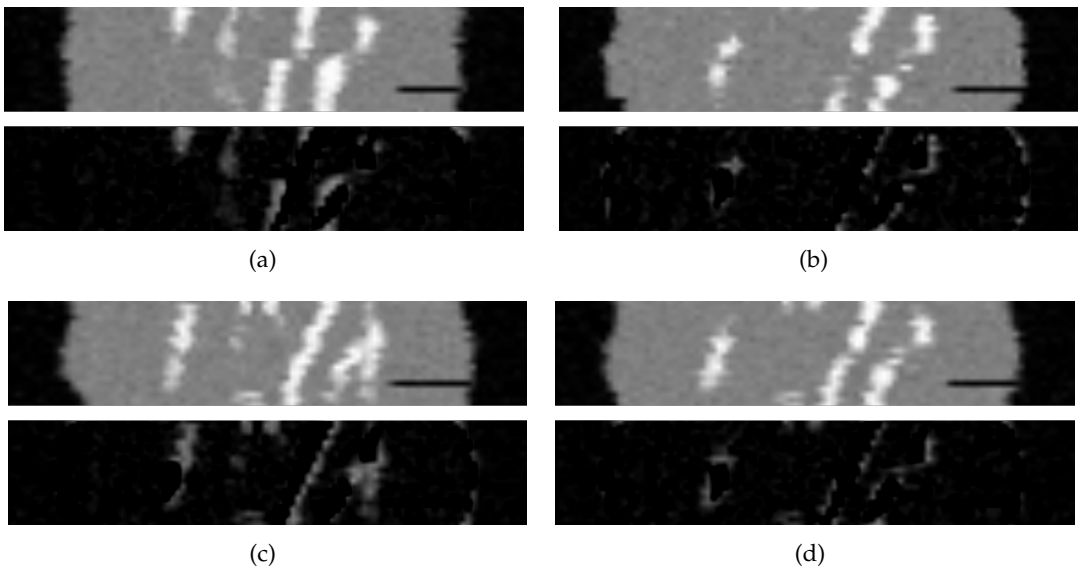
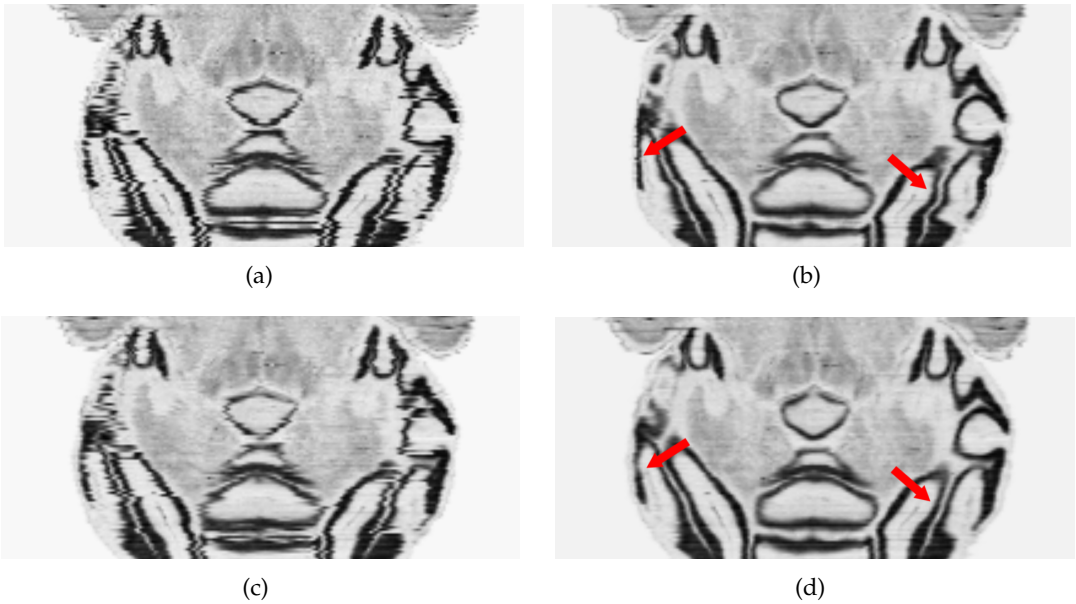


FIGURE 4.28.: Deformable reconstruction results using different methods. First row shows the volumes while in the second row absolute differences with Figure 4.27(a) are shown. Reconstruction by using (a) consecutive registration w/o SPM, (b) consecutive registration w/ SPM, (c) simultaneous registration w/o SPM, (d) simultaneous registration w/ SPM.

registration (Figure 4.28(c)) does perform significantly better than the unregularized consecutive method from a visual perspective, it still produces a high error which can be again compensated with the use of the SPM. Since the simultaneous methods put more emphasis on aligning the actual image data, the reconstructed stacks have a slightly smoother appearance (Figure 4.28(c)). However, without the regularization through a structural map, there is a drift error. Also the results of both of our methods are visually closer to the original stack in Figure 4.27(a) than the unregularized methods.

Real Dataset

We also performed experiments on a mouse brain dataset of 100 slices with 213x168 pixels that was provided online by [Ju 06]. Since the spacing infomation was missing, we assumed it to be 1mm. The FFD grid size was therefore set to 20mm and subdivided on 3 grid levels. γ and ρ in Equation (4.34) were set to 0.5 again. The slices were aligned rigidly beforehand (see Figure 4.29(a)). Since there is no ground truth available, only visual results are provided. As indicated before, the consecutive method (Figure 4.29(c)) improves the structural consistency over the source stack but the result is less smoother than the unregularized simultaneous method (Figure 4.29(b)). Our simultaneous method (Fig 4.29(d)), however, improves the results significantly over the other two approaches. It especially corrects drift errors that are present on the highlighted structures in Figure 4.29(b). The overall simultaneous reconstruction including the tensor voting took around 20 minutes while the consecutive one only needs around 7 minutes which is a potential advantage on big datasets. For all experiments, a PC equipped with Intel® Core™ i7, 2.80 GHz, 8 logical



(a) (b)

(c) (d)

FIGURE 4.29.: Deformable reconstruction of mouse brain from histology sections. Shown are the coronal views of the reconstructed mouse brain dataset. (a) rigidly aligned source stack, (b) simultaneous registration w/o SPM, (c) consecutive registration w/ SPM, (d) simultaneous registration w/ SPM.

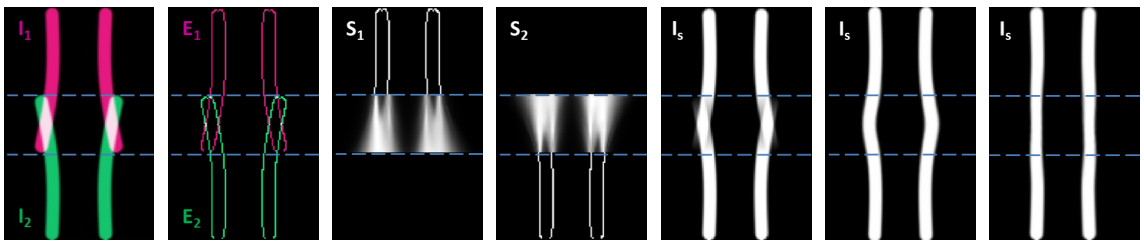
cores and 12GB RAM was used.

4.4.3. Whole Body MRI

For the multi-station scanning based whole-body MRI imaging, a post-processing stage is necessary for the fusion of partial observation in their overlapping regions [Diet 08]. However, approaches for fusing the images in the overlapping region do not consider the natural continuity of the geometrically distorted structures beyond this overlapping region. Although, it is not possible, without a shape prior, to infer the distorted structures using the existing salient far structures, it is still possible to exploit the saliency of relatively close structures to make an inference about the distorted structures.

As described in Section 4.3.3, we propose to use the propagated structures for regularization in terms of consistency and smoothness in the overlapping region. To demonstrate the effectiveness of structure propagation for this application, we have designed experiments on synthetic as well as real images. In the synthetic case, we use two partial images of size 100x100 pixels with 1mm uniform pixel spacing, containing two sticks, which exhibit distortions in the overlapping region (40mm) as shown in Figure 4.30(a). The corresponding edge and the structure images are shown in Figures 4.30(b) to 4.30(d).

For the registration, two levels are used for both the grid and the image pyramids while the number of discrete optimization iterations on each level is set to 20. The inter-grid spacing is set to 30mm and the maximum allowed displacement of each control point is set to 5mm with 5 steps for the sparsely sampled label sets. For the structure propagation, the scale parameter (σ) is set to 30mm and the compactness parameter (c) in Equation (4.11) is



(a) (b) (c) (d) (e) (f) (g)

FIGURE 4.30.: Stitching on synthetic images with distortions in the overlap. Blue-dashed lines show the overlap boundary. (a) Overlayed images to be stitched (originals are shown in Figure 4.18). (b) Corresponding edge images. (c)-(d) Structure images each spanning Ω_s. (e) Initial stitch without deformable registration. (f) Stitched image by using only intensity information. (g) Stitching result when structural information is utilized. The effect of using structure propagation can be clearly observed in the final stitched images.

set to 30. The edges are detected using the Gaussian approximation of Deriche [Deri 93] with a scale parameter of 3.

By registering the edge images to the average structure image in the overlapping area and applying the resulting transformation to the original images, we stitch the two images while simultaneously correcting the distortions. Initial stitching result and the result after registering with and without structure propagation are shown in Figures 4.30(e) to 4.30(g). This example visually demonstrates the effectiveness of using structure information. Note that without using structures, images are iteratively registered to the weighted average of the original images. Although this will lead to a locally optimum correspondence in terms of intensities, a alignment that respects the natural continuation of structures cannot be guaranteed.

For the quantitative evaluation, we apply varying degrees of synthetic distortions with known displacements to non-distorted stick images. Then, we recover displacements fields by performing stitching using our method with structure propagation and the method of [Wach 08] without structure propagation. Two displacement error measures, mean angular error (AE) and mean end-point-error (EE), which are commonly used in optical flow [Bake 11], are used for evaluation. For the synthetic distortion, we use a distortion model $F(l) = c_1 e^{-l^2/c_2}$ where l is the distance from the stitching border in the stitching direction, c_1 and c_2 are variables controlling the amount and the sharpness of the distortion, respectively. Both 2D and 3D experiments are conducted where a 3D image is obtained by stacking 20 2D stick images. Several pairs were created and stitched together by using both methods where c_1 and c_2 were varied between [10-20] and [20-30], respectively. These ranges were empirically chosen in order to generate meaningful synthetic data. Note that these values are not part of the algorithm. Two registration levels with 20 pixels control point spacing were used. For each run, AE and EE values as well as their standard deviations were computed for both methods. We first tested the sensitivity of our method alone to the given ranges as shown in Figure 4.31, which demonstrates that the method is more sensitive to the amount of distortion (i.e. c_1) than its sharpness (i.e. c_2). Overall statistics of errors for both methods are shown in Figure 4.32. It is clear that the use of structural

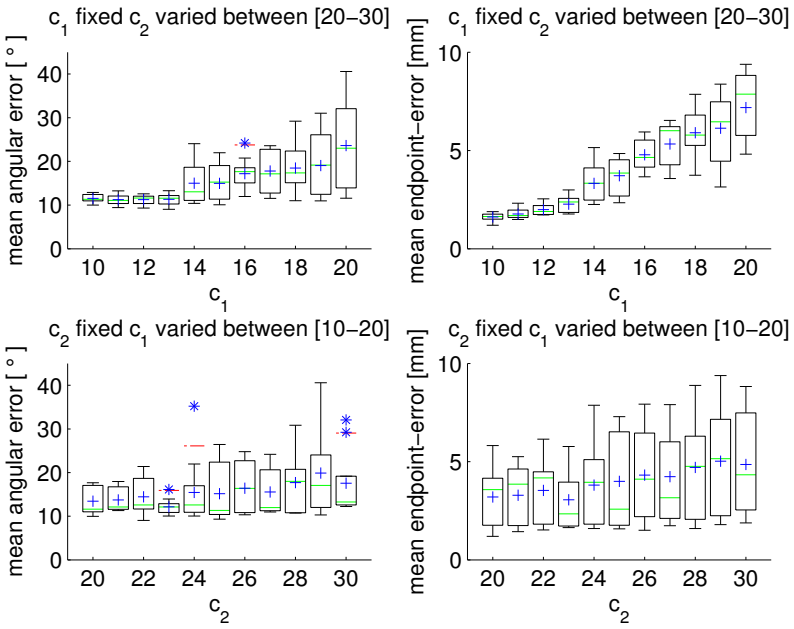


FIGURE 4.31.: Performance analysis of structure propagation using synthetic distortion model. Performance analysis of our method by varying the parameters of the synthetic distortion model.

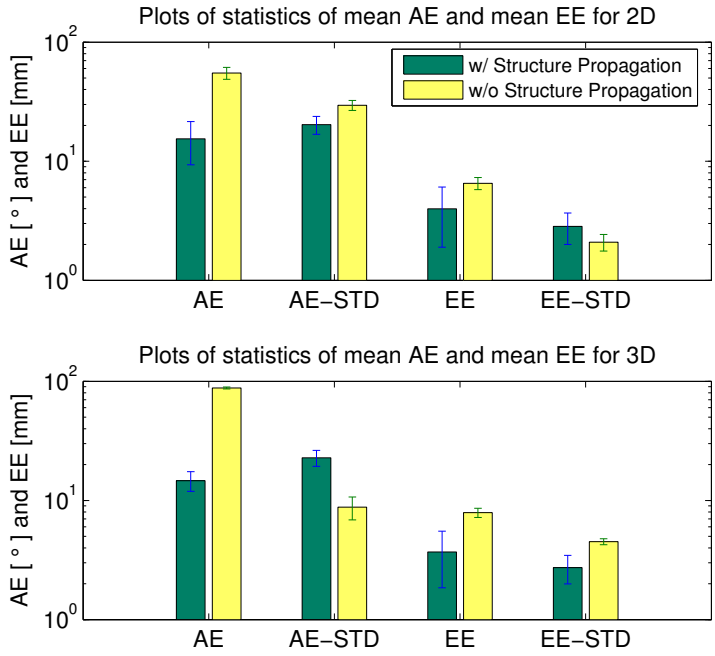


FIGURE 4.32.: Performance analysis of structure propagation using synthetic distortion model. Evaluation of the method on synthetic images in 2D and 3D. Please note that the y-axis has logarithmic scale. It is obvious that the use of structure propagation leads to smaller mean angular and endpoint errors as well as smaller standard deviations in both cases.

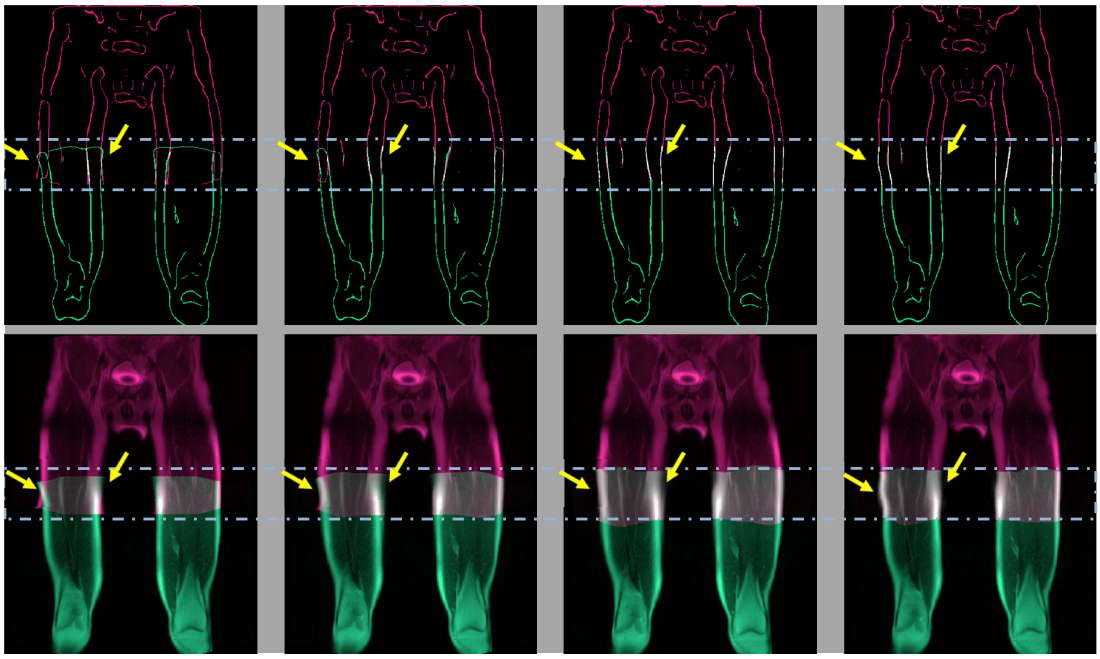


FIGURE 4.33.: Whole-body MRI stitching experiments. Edge and intensity representations are shown on the first and second rows, respectively. Columns from left to right: (1) Initially overlayed input images. (2) Stitching with structure propagation. (3) Structure stitching combined with a subsequent intensity stitching. (4) Result with only the intensity stitching without taking the structures into account. Arrows are added to emphasize changes in regions with severe distortions.

maps rather than the intensities alone leads to much smaller mean errors in both 2D and 3D cases.

For experiments on real data sets, two 3D MR images from the lower body are used for creating a larger field-of-view. The resolution of both images is $448 \times 318 \times 30$ pixels with $1 \times 1 \times 7.8$ mm spacing and a known overlapping region of 70 mm along the head-to-foot direction. For the registration, two levels are used for both the grid and the image pyramids while the number of discrete optimization iterations on each level is set to 20. The inter-grid spacing is set to 30mm and the maximum allowed displacement of each control point is set to 5mm with 5 steps for the sparsely sampled label sets. For the structure propagation, the scale parameter (σ) is set to 50mm and the compactness parameter (c) in Equation (4.11) is set to 30 while the size of the region for propagation is set to 70mm in accordance with the prior information. The edges are detected using the Gaussian approximation of Deriche [Deri 93] with a scale parameter of 5.

Only qualitative experiments with visual inspection have been performed as there was no ground truth stitching to quantify the stitching error. We compare to the case where no structural information is utilized. After stitching by using structural probability maps, we perform another intensity based stitching deploying the method presented in [Wach 08]. This was only to correct for very local fine intensity misalignments while the dominant structures are already aligned by registering to the propagated structures. Results with comparison to the case where only intensity based stitching is used are shown in Figure 4.33.

Overlaid edge and intensity images are shown in the first and second rows, respectively. Note that edge and intensity images are shown simultaneously to highlight the influence of stitching in both representations. Furthermore, no blending is performed in the overlapping region for a better assessment of matching quality. Edge images are registered if a structure-based stitching is performed and intensity images registered otherwise. The resulting deformation fields are applied to both representations simultaneously.

The first column in Figure 4.33 shows the original images where the distortion is visible in the overlapping region (highlighted with dashed lines) especially in the lateral direction. In the second column, non-rigidly aligned structure images are shown in the upper image with the corresponding intensity images in the lower one. Note the improvement of alignment in the highlighted regions with respect to the first column. Although structures are aligned to a great extent in the overlapping area, a subsequent intensity based registration is still needed in order to align the intensities. This is due to the fact that only the edges are considered during registration. In the third column, we show the results after this additional intensity-based alignment. Note the improvement of alignment of intensities compared to the previous column. Finally, in the last column, we show the stitching result where only intensity information is utilized without considering the structures at all.

It is clear that the use of structural information alone already leads to smoother edges as well as better continuation of structures across the overlapping area. This is mainly due to the gross alignment of structures. Subsequent intensity-based alignment helps further refinement of local intensity correspondences. On the other hand, if only intensity information is utilized without taking the structures into account, as depicted in the last column, distortions do not vanish leading to bulges in the middle of the overlapping region. Finally, based on these observations, we can argue that a simultaneous use of structure and intensity information in a joint optimization fashion could result in even better outcomes.

4.5. Discussion

The proposed stitching method presented here is designed to overcome the limitations of classical techniques by means of structure propagation by enabling a "perceptually good" alignment of images under difficult conditions such as subimages having small or no overlap, or distortions in the overlapping region. Although the method is addressing the state of art, there are still several limitations of the framework arising either from the employed method for the structure propagation or from the used image registration technique.

The tensor voting technique is capable of propagating the information regarding the underlying structure. However, in the presented technique only the local anisotropy is taken into account for the propagation of the information. Therefore, currently, structures can be propagated only as locally linear structures. Although, this is sufficient for structural propagation in short ranges, it cannot properly extend curvilinear structures in its current design. There have been attempts to include the local curvature of the underlying structure into the voting procedure [Tong 05]. In Section 4.2.4 we also showed how local curvature can be included into the structure propagation framework enabling curvature-based structure propagation. We demonstrated the feasibility of estimating the tensors of curvature and exploiting them for an improved structure propagation respecting to the geometry of

underlying structures.

The scale parameter σ of the weighting term Equation (4.10) for tensor voting is another factor that has an influence on the quality of propagation. Having a too large value for σ causes blurry propagation of structures while having it too small yields to a limited propagation of local structural information. The choice of this parameter is strongly related to the size of the region of extension, Ω_{ie} in Figure 4.3(a). In our synthetic 2D stitching experiments, for instance, we used a value of 20mm for the overlap region which proved to be enough. Accordingly, the scale parameter was chosen to be 40, which guaranteed that a structure point located 20mm into Ω_i can send a vote with sufficient saliency to the farthest point within the region of extension, Ω_{ie}. The choice of this parameter depends also on the size of physical gap. Obviously, for a large gap size, a larger value of scale parameter is needed. An interesting direction of research here would be adapting the scale parameter to the non-linearity of the underlying structure. Such an approach should allow larger extensions for relatively linear structures.

Although promising results were observed during experiments, the obtained registration errors, especially the ones shown in Figure 4.21 for the 2D image stitching task, indicate that the proposed registration is still far from being compared to the errors obtained using classical registration techniques. However, it should be noted that these results are despite the absence of an overlap whereas no errors have been reported so far in such challenging cases. On the other hand, the experimental results on the deformable reconstruction of histology sections show that there is big potential in the analysis of local structural consistency for a reference-free reconstruction. However, the quality of reconstruction is still not comparable to reference-based methods, which is an expected outcome. Nevertheless, the reconstruction quality was visually comparable to the results obtained by min-max curvature flow based reference-free reconstruction proposed by Cifor et. al. [Cifo 11] and another novel reference-free reconstruction based on frequency separation proposed in the recent work of Gaffling et. al. [Gaff 14]. For the validation of our approach to histology reconstruction on real images, we have used only visual inspection. This was mainly due to the absence of ground truth data for histology images. However, a more elaborate quantification of performance can be performed by using simulated histology images as done in [Gaff 14], which we leave as a subject of future work.

As another application, whole-body MRI is used to show how structure propagation can be exploited for a locally consistent fusion of multi-station MRI volumes in their overlap. Experiments indicate that the structural information alone is not sufficient to find a consistent matching in terms of full intensity correspondences. This is due to the sparsity of the structural correspondences when only the edges are used. In order to smoothly propagate the deformation obtained from the registered structural representation, it is necessary to employ structural consistency as a regularization term in the registration functional.

As a result of structure propagation for local structural consistency, we obtain structural probability maps. It should be however mentioned that these probability maps do not obey the rules of a probability distribution such as the sum of the entries should sum up to 1. We still call it probability maps as they reflect the likelihood of the presence of a certain structure in a particular region in the image. For the assessment of similarity between these maps as well as when comparing to deterministic structural images (i.e. egde maps), we have so far used normalized cross correlation. Since these images have probabilistic nature, the study of more appropriate similarity measures designed for comparing probabilistic

distributions such as the Kullback-Leibler divergence [Chun 02], or the Jensen-Renyi divergence [He 03] would be another interesting direction for future work.

Finally, although the demonstrations for the 2D image stitching application include only two images, the proposed method could also be extended for the alignment of multiple images. If prior information in terms of adjacency relationships is available, a simultaneous optimization strategy similar to bundle adjustment can be employed to obtain the optimal transformation parameters for neighboring images. Otherwise, a method similar to the one employed in [Brow 07] can be used to first identify adjacency relationships between subimages. Here, the proposed structure propagation technique can be employed to enable the comparison between subimages when no overlap is available.

4.6. Conclusion

In many medical imaging applications ranging from microscopy to ultrasound to digital pathology, a wide field-of-view is usually desired to enable a better analysis at different scales. However, having a wide field-of-view is often limited by the capabilities of the imaging devices in guaranteeing a large field-of-view at the highest resolution. Therefore, it is a common practice to acquire smaller tiles which are then stitched together in order to get a larger field-of-view. There have been many solutions proposed in the past for the stitching of subimages in various applications. The common requirement for almost all of these techniques is that the subimages to be stitched have to have at least a sufficient amount of overlap in order to be able to optimize some transformation parameters using the information shared in this overlapping region. The amount of overlap has remained as a challenging issue which is often circumvented by the undesired solution of enforcing a sufficient overlap during image acquisition. Although this is a solution in some cases, it not only brings further complications to the acquisition protocol such as using accurate motorized stages as in the acquisition of large field-of-view microscopic images but also introduces further computational complexity during the final stitching.

Registration of neighboring histology sections with the purpose of a 3D volumetric reconstruction can also be cast as an overlap issue. Due to the characteristics of the imaging modality, having overlap along the stack direction is not possible. Independent artifacts introduced to each slice have a direct effect on the amount of structural overlap between neighboring slices. Although reference-based reconstruction techniques are able to handle such inconsistencies, complicated acquisition protocols due to the additional reference volume remain as a major limitation of such approaches. Ideally, without using any external references, it should be possible to estimate the local geometry of anatomical structures from the distorted partial observations from digital histology slices.

Although recent advancements in the imaging technologies have enabled better ways for creating whole-body MRI, such technologies are not commonly available yet. A whole image of a human body using MRI is usually possible only by using multi-station image techniques. In this acquisition setting, multiple images covering different overlapping regions of the body are acquired and fused together to reconstruct a whole-body MRI. However, fusing multiple images in their overlapping area needs special consideration due to the geometrical distortions inherent to MRI. A straightforward approach would be to acquire more images than needed to cover to the entire body and stitch together after

cropping distorted regions. However, this may result not only in a much longer overall acquisition time but also motion artifacts due to the extended acquisition time which would lead to inconsistencies. Therefore, the optimal protocol should be the one that acquires the minimum possible number of partial images in the shortest possible acquisition time. In this case, geometric distortions should be handled in such a way that local continuity and consistency of structures should be respected.

Here, we have proposed a novel structure propagation technique to address the local structural consistency in medical computer vision problems. Our approach was mainly inspired by the perceptual grouping principles from the field of the Gestalt psychology. Principles such as closeness and good continuity for inferring curves and surfaces from noisy data has been the starting point for our approach. Using a similar approach on feature sets such as edge points extracted from image data, we posed the structure propagation as an inference problem. Then, we solved the inference problem using the tensor voting framework where feature points such as edge points first learn their self structural saliences and, then, propagate their local structural information using tensor voting by respecting the perceptual grouping principles. Propagated structures result in a structural probability map indicating the saliences of types of structures at the inferred locations.

We have applied the proposed structure propagation method to solve the above mentioned problems in medical image analysis. First, 2D image stitching using structure propagation was demonstrated, which addresses specifically the classical overlap issue in image stitching. The main motivation for incorporating a perceptual grouping mechanism in the stitching task is our belief that such an approach will ensure the continuity and the smoothness of structures across subimages to be stitched. This belief is based on the fact that perceptual grouping techniques are usually inspired by the perceptual power of the human visual system that can visually reconstruct an occluded region by imposing the smoothness and the continuity principles on the observed signals. Smoothness and continuity of structures are essential properties for medical images. Although ensuring these properties is possible with the existing methods when a sufficient amount of overlap is present, it gets challenging as the amount of overlap decreases. It becomes even impossible to perform a stitching with the existing stitching methods if there is no overlap between the pieces. Therefore, the proposed method is designed to overcome this limitation by allowing a "perceptually good" alignment in such scenarios by means of structure propagation which helps to create an overlap region where the information is extrapolated from the non-overlapping regions of subimages. Experiments on synthetic as well as real medical images have shown the potential use of the proposed technique for creating wide-field-of-view microscopy images. The only requirement for our method that the structures should have inherent smoothness indicate that the proposed technique can be employed for solving many similar scenarios in the field of medical image analysis.

As the second application, we employed the structure propagation for creating structural probability maps to be used for regularization during deformable reconstruction of histology sections. In the absence of a ground-truth reference image, it is hard to know whether a local optimum is indeed leading to a structurally smooth and consistent volumetric reconstruction. To alleviate this issue, we infer the most locally consistent locations of distorted structures through local voting. While registering neighboring slices to each other based on intensity features, we use the inferred structures as regularization to ensure the loyalty of the optimum to the local structural consistency. Results of experiments on

synthetic and real histology images revealed that using structural probability maps yielded a better deformable reconstruction of histology sections.

Finally, we applied structure propagation to the problem of fusing multi-station partial MR images to create a whole-body MRI. The distorted image region was augmented by the structural probability map created by propagating structures that are relatively salient. Aligning structures for initial alignment and then refining the registration by using intensity information, we are able to fuse partial images in a natural manner. We compared our approach to a baseline method where the partial images are registered to their linearly weighted average. Experimental results showed that using structural probability maps as a regularization term leads to much more smoother fusion of structures in the region where the images overlap.

In this chapter, we have demonstrated by several applications that structural probability maps obtained through structure propagation can be easily incorporated into any registration functional as a regularization term to guarantee the local structural consistency. However, the application of the proposed regularization technique is not limited to the ones presented here. It can help solving similar problems in the field of medical image analysis where distorted or missing structures is an issue.

CHAPTER

FIVE

HOUGH SPACE PARAMETRIZATION FOR GLOBAL STRUCTURAL CONSISTENCY

Field-of-view extension via mosaicing of partially overlapping images has been used in various applications ranging from remote sensing to computer photography to medical imaging. The main challenge in most applications is having an extended field-of-view without sacrificing spatial resolution. A variety of methods have been developed over the decades for aligning partially overlapping images. These include matching features extracted from partial images and intensity-based alignment where correspondences are built between image intensities in overlapping regions.

While feature-based approaches may fail in the absence of salient, distinctive features, intensity-based approaches suffer when the amount of information in overlapping regions is insufficient for estimating transformation parameters. In particular, in the absence of dominant structures such as strong edges in such regions, obtaining a solution that ensures consistency between the structures spanning partial images becomes difficult.

Towards addressing this issue, in Chapter 4, we have presented a method for ensuring local structural consistency by propagating salient structures from non-overlapping parts. While this approach yields locally consistent fusion of partial images, it is, however, hard to assure a globally consistent alignment of all partial observations by employing only a local approach. Thus, in this chapter, we propose a global regularization term for consistency in intensity-based image registration. This term exploits the vast amount of available information beyond the overlapping area, irrespective of the distance to the overlap. This approach has advantages over the previous local one particularly in terms of being unbounded to the support range of structures as well as to the size of overlap between images.

In this work, the global regularization term is formulated in terms of the Generalized Hough Transform (GHT) [Ball 81]. Using GHT-based object representation, we learn a flexible model of the object of interest using prior data by parametrizing the object shape in the Hough space. This parametrization is then used for global regularization when registering partially overlapping images of a particular scene including the learned object without using any prior data.

This chapter is organized as follows. In Section 5.1 we start by motivating the problem of registering partially overlapping images. Then, we present the relevant prior art on this subject. Especially, we look into the object recognition approaches using part-based

object representations. Then, we relate our shape-based representation with similar techniques for object recognition in computer vision. Section 5.2 provides the details of the proposed global consistency measure where we also give an overview of the Generalized Hough Transform. Our empirical evaluation through experiments on synthetic as well as on sample real medical images is presented in Section 5.3. Finally, we close the chapter with concluding remarks after discussing the results of the experiments.

5.1. Introduction and Related Work

Construction of high resolution images with a wide field-of-view has been an active area of research for many decades. Applications where field-of-view extension is used include consumer photography for panorama imaging, remote sensing, satellite imagery and medical imaging. While most of the existing solutions are based on extracting and matching features from images covering different partially overlapping regions of a scene, such approaches are likely to fail when it is difficult to find salient features in the overlapping regions. Intensity-based alignment is an alternative approach where full intensity content is used. Intensity-based registration offers better accuracy with the cost of increased computational cost compared to the relatively more sensitive feature-based methods.

Traditional approaches to field-of-view extension rely solely on the overlap between images covering different parts of a scene. Therefore, existing solutions cannot be employed in cases where images do not overlap. Moreover, the quality and the amount of information in the overlap is a crucial determining factor in the success of any alignment method. Degradations due to imaging setup or lack of structural information present challenges both for feature-based and for intensity-based registration. In addition, and more importantly, to the best of our knowledge, no image mosaicing technique using overlap considers the global properties of underlying structures during alignment. In fact, local primitives including structures as high level features and intensities as the lowest level features provide valuable cues about the composition of a particular scene. Therefore, such local primitives can vote for the global properties of the scene. In this work, we present an approach to extending image field-of-view by incorporating local primitives in the estimation of optimal transformations respecting the global properties of underlying objects or structures such as position or scale.

The primary observation that led to this work is the contextual relationship of the information available in partial images of a certain scene. It can be easily observed that objects or structures in natural images have certain geometrical or physical constraints. For instance, knowing the positions of the legs of a cat gives cues about the possible location of its head. Once such constraints are learned and partial observations of the scene are available, it should be possible to make predictions about the spatial configuration of these observations within the scene. Constraints on the objects may be represented through models of objects of interest in different forms. In medical image analysis, for instance, models are quite popular through which prior knowledge can be exploited while performing certain analysis tasks on images. The most successful application of models is the segmentation of objects of interest in medical images. There is a vast amount of literature on different methods using prior shape and appearance information in terms of statistics for segmenting objects of interest [Heim 09].

In some applications, however, the goal is not an accurate segmentation of an object in the first place but rather its detection or localization in a particular scene. Such approaches also rely on prior knowledge in terms of models or training images. Our approach is also in line with such methods in regards to using prior data. However, the main difference between this work and previous works on object detection is that we are not directly interested in detecting objects. In this work, we perform soft detections on partial images and independently generate object configuration hypotheses. We, then, use the consistency between hypotheses to assess the consistency of alignment between partially overlapping images. Here, we use the term *soft* to indicate that our final goal is not the detection or localization of the object within the scene. Due to this relationship with object recognition methods, we will have a deeper look into the literature about detection approaches based on training data before describing our measure of consistency using partial soft detections.

5.1.1. Related Work on Object Recognition

Object recognition task can be classified as specific vs generic [Grau 11]. In the first category, the goal is to detect the presence and the location of a specific object, such as a certain book or a particular car. The usual steps in specific object recognition include extraction of features on the object in a scene and matching of these features. Finally, a verification step is needed to ensure a consistent geometric configuration of the feature correspondences. In the second class of object recognition tasks, one tries to identify and locate objects belonging to a generic object category [Grau 11]. For instance, detecting humans in video sequences is a good example for this class of tasks. An algorithm for detecting humans does not make any distinction between different humans. Since we are interested in the registration of partial images of generic objects such as human organs, in the following, we will focus on generic object recognition methods.

In generic object recognition, we differentiate between window-based and part-based methods. Window-based approaches model the object appearance within a region of interest in terms of image gradients or texture. Feature descriptors extracted from the entire region or within its sub-windows are used to summarize the appearance. Descriptors for window-based representations include SIFT [Lowe 04], SURF [Bay 06], HoG [Dala 05] and Haar-like box filters [Viol 04] among others. On the other hand, part-based representations handle object as a combination of its parts and try to establish geometrical relationships between parts for introducing flexibility. Although window-based recognition approaches have been successfully used for many applications, they have certain limitations when the object of interest is textureless or goes under global appearance changes or occlusions. On the other hand, part-based models combine the local appearance in the form of a visual vocabulary with spatial relationships making it robust to global changes in illumination or appearance as well as partial occlusions [Grau 11].

The goal in this work is to incorporate local information contained in partial images in assessing the global consistency of alignment between them. Similar to part-based object recognition approaches, local parts in partial images give cues about the underlying object. Due to this similarity, in what follows, we give more details about part-based models for object category recognition.

Part-based Object Recognition

In part-based models, geometrical relationships between object parts can be formulated in different ways. The simplest form is the bag of visual words model [Csur 04] where there is no geometrical relationship between the parts. The most complicated form, on the other hand, is a fully connected graph where all parts are assumed to be connected to each other [Ferg 03]. Leibe et al. [Leib 04] introduced a star-shaped spatial model for layout representation which stands in between. In this type of representation, local parts are connected to only a central reference part, thus, they are conditionally independent from each other. Such a representation has computational advantage compared to a fully connected model. Many subsequent works for visual object recognition in the last decade use similar approaches in part-based modeling due to its efficiency. Therefore, we mainly focus on the star shape representations in the following.

Since each part of an object is treated independently given a reference part, the instantiation of an object is also relative to the central part where local parts make predictions about the location of the reference. Usually, local parts are chosen as representations for appearance in terms of codebook entries. Codebooks or visual vocabularies can be constructed by grouping together similar parts of different objects in the same category. Therefore, each codebook entry represents varying appearances of a part in different training images. During recognition, a local patch around an interest point can be compared against codebook entries and then corresponding hypotheses stored in the codebook are used for the prediction of the reference part. A powerful mechanism called *Implicit Shape Model* (ISM) for implementing this idea is proposed in [Leib 04]. This approach models the implicit geometric layout between the parts of an object along with part appearances.

The hypothesis generation mechanism employed in ISM is inspired by the evidence gathering technique used in the Generalized Hough Transform (GHT). We will present more details about the GHT in the methods section. In contrast to the GHT where the orientations of object contour points are used for indexing a table of offsets storing the vectors for the reference point, in the ISM, patch appearances are matched against a visual codebook casting probabilistic votes for the reference part. Votes cast are accumulated in the Hough space and then peaks are detected as agreed hypotheses for the object reference point.

Despite its popularity, robustness to partial occlusions and proven success in detecting objects with large number of parts, voting-based methods based on visual codebooks come with prohibiting computational cost with increasing number of object parts [Gall 13]. To address this issue Gall et al. [Gall 13] formalized the Hough voting-based detection process within a discriminative random forest framework, which is called *Hough Forests*. In Hough Forests, mappings between image patches and probabilistic votes for the reference point of an object are learned from annotated training data along with the class probabilities. The main difference between the two approaches is that while codebook-based approaches have a unsupervised way of learning an explicit appearance codebook from the training data, Hough Forests does the same in a supervised manner yielding an optimized implicit codebook for voting [Gall 11].

Random Forest, which is the basis for Hough Forest, consists of decision trees that are trained on random subsets of a training data set. It can be applied both for classification or regression leading to either discrete or continuous result on a test image patch. The

power of random forests comes mainly from its capacity of processing large training data sets as well as from its speed and generalization capabilities [Fane 13]. During training, each randomized tree gets a random subset of the training data and at each non-leaf node of the tree, patches from the training images are sent either to the left or the right child node based on a binary test that optimizes the information gain after split. This procedure continues until some criteria such as the maximum tree depth or the minimum number of points in a child node are met. The binary test applied on each node is chosen to be the one from a randomly chosen subset of binary tests that yields the best information gain, thus minimum uncertainty in the set going to each child node.

Hough Forest [Gall 11] extends the Random Forest training procedure by changing two components. First, the input to the decision tree includes not only the labeled pixel data, but also, depending on whether it is from the figure or ground, it contains the corresponding offsets for the object reference point. Secondly, the function optimized at each node is modified by adding another term that minimizes the uncertainty among the offset vectors belonging to a child node. During training, one of these two functions is chosen also randomly at each node. Hough Forest exceeds the performance in the state of the art. It is suitably designed to handle multi-class and multi-aspect views, and adapts to detect deformed and occluded instances of the model in cluttered environments. We will birefly discuss the potential use of the Hough Forest for the purpose of globally consistent alignment in the discussion section.

Contour-Based Object Representation

In addition to the Hough Forest, several other extensions of the ISM framework [Leib 04] have been proposed for voting-based inference in part-based object category recognition. Some of them have focused on representing the shape of an object instead or in addition to the appearance [Shot 05, Opel 06, Shot 08]. The principle idea behind using shape is that shape elements such as boundaries provide a natural representation of the object. This kind of representation is particularly useful for less textured objects [Opel 06]. In addition, it provides a better means of representation for objects having unreliable textures due to degradation or noise. It should be noted that boundary-based representations can be coupled with texture-based representations for a better overall performance. We will discuss this among the possible extensions of our framework in Section 5.4. Other advantages of contour- or boundary-based representation include robustness to the changes in appearance due to illumination, color or texture [Shot 08].

Opelt at al. [Opel 06] proposed a *Boundary Fragment Model* (BFM) where they train weak detectors using small subsets of boundary fragments and obtain a strong detector via boosting. They use star-based representation as the spatial prior where votes for the object center are stored along with the contour fragments. Shotton et al. [Shot 05] independently proposed a similar approach. Both methods use Chamfer distance for matching contours from a test image to the codebook of learned boundary fragments. Similar to Leibe et al. [Leib 04], based on the matched codebook entries, votes for the object center are cast and peaks in the Hough space are detected as the instance of the object class. Another related approach is proposed by Riemenschneider et al. [Riem 10] where contour descriptors based on angular representations are matched partially to the model object contour. Then, the votes generated by the matched contour fragments are accumulated in the Hough space for subsequent

hypothesis detection. Angular descriptors are used to improve the discriminative power of the matching process.

Inspired by part-based object recognition approaches for detecting objects through parts, we introduce a global consistency measure to be used when putting parts of an object or a scene together for widening the available field-of-view. Our consistency measure depends on the detection hypotheses generated via partial images of an object or a scene of interest. A consensus between the strongest detection hypotheses for partially overlapping images in a parametric space indicates a consistent configuration of images with respect to the underlying object of interest. Thus, global consistency when registering partial images can be seen as the consistency between the detected object hypotheses in partial images.

5.1.2. Related Work on Field-of-view Extension

In the previous section, we draw a parallelism between object detection approaches using part-based modeling and the idea of assessing global consistency via local information. It this section, we will discuss the prior art on field-of-view extension in different fields, particularly in medical imaging.

Computer Vision

The problem of field-of-view extension has been studied extensively in the literature. While generic methods exists, most approaches are tailored towards specific applications. Of these, consumer photography is the most popular application area of field-of-view extension. A excellent review on image stitching is presented in [Szel 06]. In his review, Szeliski states that film photography was the first area where a panoramic imaging was used for getting a wide field-of-view. Later, it was also used for creating panoramas from consumer cameras [Mann 94, Szel 96]. Subsequent works focused on getting a globally consistent alignment (still based on overlap) [Shum 00].

Researchers have also focused on automatic determination of adjacency relationships between the overlapping partial images of a particular scene [Brow 07, Shum 00, McLa 02]. These methods use the idea of bundle adjustment [Trig 00] for obtaining the globally most consistent alignment of partial images. Despite the proven success of these methods in many data sets and applications leading to various commercial products, their performance is bounded mainly by the amount and the quality of overlap between partial images. For instance, it could be difficult to solve the parallax error (created by seeing an object from different viewpoints) when the overlap size is limited [Szel 06].

Medical Imaging

Extending the field-of-view of biomedical samples for visualization under microscopes without sacrificing spatial resolution is another related active area of research. In biological imaging, for instance, it is necessary to see an entire section of a sample at high resolution which is usually higher than the resolution offered by acquisition devices. Therefore, in order to acquire a high resolution image, often multiple tiles are captured with or without overlapping regions and combined or stitched together [Rank 05, Verc 06, Chow 06, Prei 09, Emme 09, Saal 10, Tsai 11, Loew 11]. Usually, a step motor is used during image acquisition

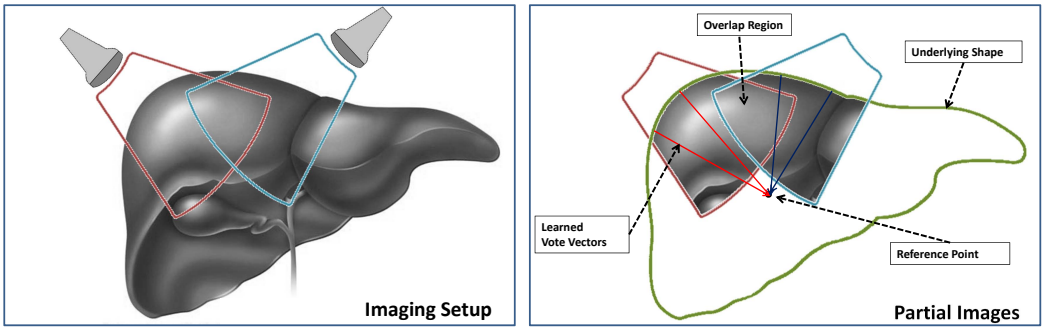
to record the positions of tiles which are used for initializing a mosaicing or stitching process. If the initial configuration of tiles is unknown, feature-based matching can be used given sufficient overlaps between the tiles are available [Saal 10]. Ensuring global consistency in microscopic imaging is remaining as a challenge since most micron level structures do not have predefined global shapes that can be learned and used as constraints. Instead local consistency is ensured through structural continuity and smoothness [Yigi 13]. See Chapter 4 for details.

Other applications of field-of-view extension include in-vivo micro-image mosaicing [Loew 11], in-vivo navigation for minimally invasive surgery [Moun 09], endoscopy [Reef 06, Sesh 06, Atas 08], ophthalmic image analysis [Catt 06] and cystoscopy [Hern 10]. Ultrasound (US) imaging is another area of research where field-of-view widening is often required. The primary reason for this need is that the acoustic window of the transducer is limited; therefore, to capture a large field-of-view, several acquisitions are necessary where partial images of a target region have to be stitched together [Wach 07, Brat 11, Oye 12]. However, primarily due to the typical imaging artifacts inherent to US, information contained in the overlapping region is often not salient enough for a purely intensity-based field-of-view extension.

Wachinger et al. [Wach 07] proposed an algorithm for aligning multiple US volumes of a baby phantom acquired from different viewpoints. Even though promising results are reported and this work pioneered the groupwise alignment of US volumes, their approach is purely intensity-based and does not utilize any structural or geometrical properties of region of interest. In [Ni 08], a feature-based US mosaicing method is presented. 3D SIFT [Lowe 04] features are detected and transformations are obtained via matching. Schneider et al. [Schn 12] suggested a similar approach where small changes between partial images are assumed and simpler features are employed yielding a real-time performance. This method however requires sufficient amount of overlap between images in addition to the salient feature points in those regions. Similarly, Brattain and Howe [Brat 11] presented a real-time method where electrocardiograph (ECG) gating and electromagnetic (EM) tracking are used for compounding. Thanks to the ECG gating, similar partial volumes were used for compounding, eliminating the need for intensity-based registration. However, again, a complicated setup involving gating and tracking is required.

3D transcranial ultrasound (TC-US) [Ahma 11] imaging serves as an example of such challenging setups where partial images or volumes are acquired through the temporal window from both sides and then reconstructed to create a larger field-of-view of the brain. However, stitching partial US images using only intensity information, especially in this setting, is very difficult due to the the typical imaging artifacts inherent to US such as low signal-to-noise ratio, large speckle patterns and low image quality [Ahma 13]. Therefore, the reconstruction is often done via a freehand ultrasound system [Wein 06]. In this system, US probe is tracked using optical tracking which keeps track of the coordinates of acquisition planes. A subsequent registration is also needed to correct for tracking errors. This however requires an additional tracking setup that is not often available in clinical settings.

Ahmadi [Ahma 13] proposed a joint segmentation, registration, reconstruction (JSR2) technique to address the challenges in multi-view 3D TC-US imaging. The reconstruction is performed in an iterative manner where the target region (the mid-brain in that case) is first segmented in every partial image as well as in the initially reconstructed image.



(a) Imaging setup (b) Partial images

FIGURE 5.1.: (a) Illustrates a possible imaging setup where partial images may have limited overlap. (b) Given that the underlying shape is known, other structures from non-overlapping regions can be utilized for regularization. **Copyright**: Liver illustrations reproduced with the permission of [Sing 14].

Then, the individual segmentations are deformably registered to the segmentation of the reconstructed image (joint segmentation). Through iteration, this leads to the optimal deformations of partial images minimizing the distances between their segmentations and the joint segmentation. Although promising results are demonstrated, this approach has strong dependency on the segmentation method involved. An incorrect segmentation can easily lead to an incorrect deformation, thus an incorrect reconstruction. To address this issue, reconstruction step is regularized through an anatomical shape prior in the form of statistical shape model (SSM). Furthermore, the available structural information is not utilized directly but rather indirectly through segmentation.

5.1.3. Globally Consistent Alignment

Having discussed the prior art on field-of-view extension, we return back to the fundamental issue of global consistency and overlap dependence when expanding field-of-view by putting many partial images together. Figures 5.1(a) and 5.1(b) show a case where two US images of the liver with little overlap need to be registered to have an extended field-of-view. Obviously, most purely intensity-based registration methods would fail here due to the limited structural information contained in the overlap to drive the registration process. On the other hand, there is a rich amount of information beyond the overlap that can support the registration. An existing solution for this problem is to register each partial image to an atlas of the target organ thus indirectly registering partial images to each other. Alternatively, one can propagate structures contained in the surroundings of the overlapping region. See Chapter 4 for details.

Several works have been done to alleviate the above mentioned issues regarding overlap. For instance, acquisitions can be performed such that there are large overlaps between partial images as done in [Oye 12]. Another typical approach is to augment the registration using previous data such as computed tomography (CT) scans [Haci 13] where each partial image is registered to the prior as well as to other partial images simultaneously [Kutt 09, Wein 09]. For such approaches to work, however, a previous scan of the same patient has

to be available, which may not always be the case in particular when no intervention is planned [Brat 11].

When no previous scans are available, other kinds of prior data such as anatomical atlases are used for registration. Although it is possible to have good registration results using atlases, an explicit use of the intensity distribution of an atlas biases registration towards the chosen reference, therefore, making it inflexible and necessitating additional regularization. Our previously described approach seeks for a consistent alignment of structures beyond overlapping region [Yigi 13]. This method, however, provides only local smoothness and continuity; thus, it cannot be generalized as a global regularization constraint.

Therefore, all the previously mentioned existing approaches for field-field-view extension will have difficulties when the overlap size is small, non-salient or no prior data is given. As a result, they cannot ensure a globally consistent alignment of partial images that respects the shape of the underlying structures. In this work, we propose not to use prior data explicitly, but we learn a global parametrization $\mathcal{P}$ of the anatomy in question. Although parametrization or learning can be represented in a number of ways as mentioned in Section 5.1.1, we use the Generalized Hough Transform (GHT) [Ball 81] to learn $\mathcal{P}_H$ of the target anatomy in the Hough space. GHT has the favorable property of being robust to partial occlusions and noise. Here, we exploit such properties of GHT to enable global regularization for intensity-based image registration. When coupled with a local similarity measure, $\mathcal{P}_H$ will serve as a regularization that ensures global consistency while registering partially overlapping images to each other.

In the following sections, we first describe the GHT. Then, we explain how known shapes can be learned in the Hough space. Afterward, the details of the entire intensity-based image registration framework are given. We also describe how non-overlapping partial images can be consistently aligned thanks to our novel intensity-based registration framework.

5.2. Hough Space Parametrization

In this work, we use the parametrization of the shape of an object of interest in the Hough space to enable a global regularization through local voting. The parametrization is done by learning the shape from training data. Learning is done via Generalized Hough Transform (GHT), which is a voting-based object recognition method. Voting-based schemes have the advantage of being robust to outliers, missing data, occlusions and small deformations.

5.2.1. Hough Transform

The origins of the Hough Transform (HT) dates back to 1950s. It was named after its inventor who originally proposed it for tracking particles in bubble chamber photographs [Houg 62]. In his patent, Hough described the basic idea of representing lines composed of co-linear points as the intersection of geometric shapes in a parametric space [Hart 09]. However, in his invention disclosure, Hough did not give any algebraic definition of the transform. It was Rosenfeld [Rose 69] who discovered the invention and in his seminal book, he gave an algebraic definition of the HT [Hart 09]. He described how zero dimensional points can be mapped to one dimensional lines in a parametric space. Later, Duda

and Hart [Duda 72] proposed to use polar representation of lines in order to resolve the issue of detecting vertical lines in images whose points map to lines in parametric space intersecting at infinity. They also extended the standard formulation to detect higher order analytical shapes such as circles [Hart 09]. Kimme et. al. [Kimm 75] included gradient information into the Hough Transform leading to more efficient detection of circular patterns. Finally, Merlin and Faber [Merl 75] demonstrated how arbitrary shapes at a certain orientation and scale can be detected using the HT. This work was later extended by Ballard [Ball 81] as the Generalized Hough Transform (GHT) for detecting arbitrary shapes at any orientation and any scale.

Linear Hough Transform

The idea behind the HT is as simple as transforming a shape of interest to its parameter space.[1] Simple case of two dimensional lines in Cartesian space is introduced as transforming points' spatial coordinates to a parametric space of their slope and intercept using point-to-line transformation. Every point in image space corresponds to a line in Hough space representing lines passing through this point with all possible orientations. Each point adds votes along a line in an accumulator corresponding to the discrete Hough space, and eventually, the areas with maximum votes are most probably the parameters of the line passing through the voting points in the image space.

Given an image $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ every point $\mathbf{p} = (x, y) \in \mathbb{R}^2$ in the image space can hypothetically lie on a line with the Cartesian equation

$$f((m, n), (x, y)) = y - mx - n = 0 \tag{5.1}$$

where (m, n) is a pair of parameters for slope and intercept, respectively. For every fixed point (x_i, y_i), the HT plots the trace of all possible (m, n) pairs that satisfy the following equation

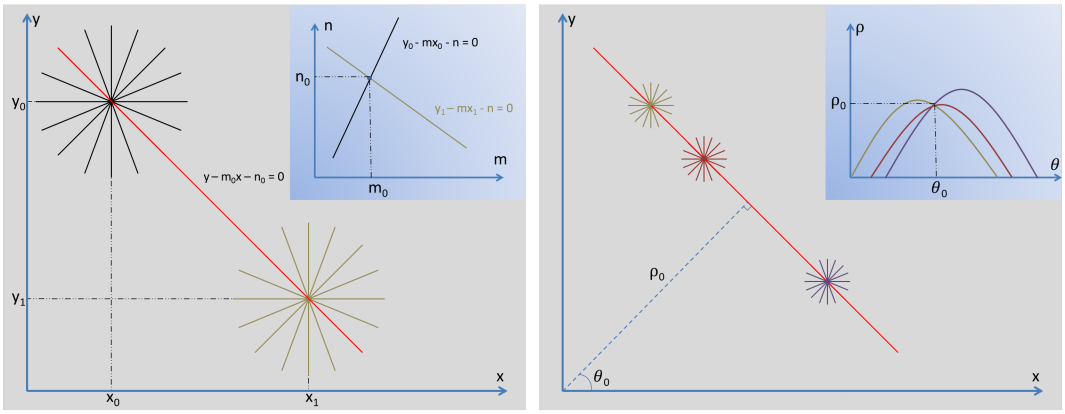
$$(x, y) = \{(x_i, y_i) \in \zeta(I)\}_{i=1}^{g} \tag{5.2}$$

$$f((m, n), (x, y)) = m + \frac{1}{x}n - \frac{y}{x} = 0 \tag{5.3}$$

letting $\zeta(\cdot)$ an operator over the image space to detect features (e.g. edge detector). The most common features used are the edge points. Its popularity comes from the fact that edges in images play an essential role in the perception of objects in human visual system. Human eye primarily decodes the appearance of shapes based on their boundaries. The linear model defines a line in the parameter space. Strictly speaking, all potential lines which a point might lie on are gathered in the parameter space, and this trace will appear as a single line in the parameter space defined by slope and intercept. If one continues repeating this for all g edge points in the image space, it will produce g lines in the parameter space.

The pipeline of a HT-based line detector starts with detecting edge points in the image space. An empty accumulator array with axis m (slope) and n (intercept) is allocated. For every line trace relating to an edge point, the corresponding accumulator cells on the trace are incremented by 1. By the end of the process, the accumulator cell containing the largest

[1] Parts of the background study on Hough Transform have been conducted by Javad Fotouhi as part of his master thesis.



(a) Representation in Cartesian coordinates (b) Representation in polar coordinates

FIGURE 5.2.: Different representations of line in the Hough space

peak is most probably defining the parameters (m_0, n_0) of the common line passing through the majority of points [Rose 69]. This is illustrated in Figure 5.2(a).

This type of line representation might lead to issues since vertical lines in image space have a gradient of infinity. Therefore, in a discrete, bounded parameter space, it would not be possible to identify the intersection of lines which most probably happens to be at infinity. Alternatively, polar representation of a line is used [Duda 72]. In Figure 5.2(b) a line is described by its perpendicular distance ρ to the origin of the image, and the angle θ of the perpendicular vector $\vec{\rho}$ with the positive x-axis. Then, the polar representation of the line follows as

$$\rho = x \cos\theta + y \sin\theta. \tag{5.4}$$

While the absolute value of ρ is bounded with the diagonal size of the image, θ value varies from 0 to 180. Unlike the Cartesian representation, the points in the image space are no more mapped to straight lines in the parameter space. The parameter space with ρ and θ axes will map image points as bounded sinusoidals.

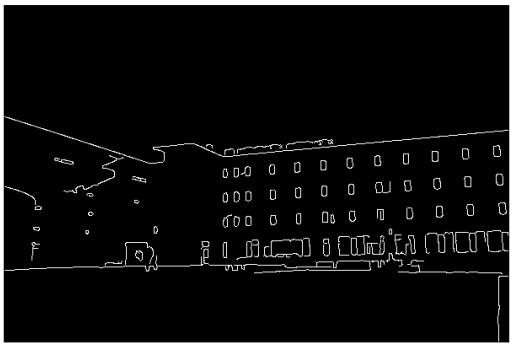
Figure 5.3 illustrates the common steps of a polar linear HT process. Linear HT is used when there are several line segments after a poor edge detection and results in locating potential lines. Methods integrating further knowledge can constrict the number of edge pixels by sampling approaches, and restrict the scanning area employing gradient information extracted from the image [Kimm 75].

Generalized Hough Transform

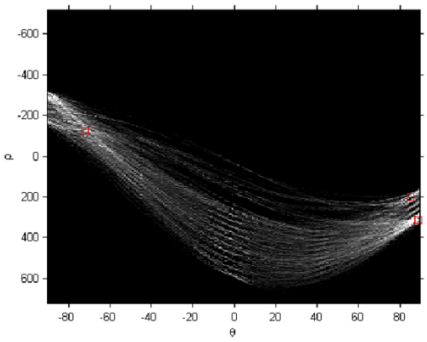
While the standard HT is able to detect analytical shapes, most objects in the world cannot be described in an analytical way. On the other hand, in most object recognition applications, the shape of an object is usually known a priori. Generalized Hough Transform (GHT) is designed to learn priors and facilitate the detection of object instances. Likewise, it enables locating complex shapes using a set of elementary structures.

GHT was introduced as an extension to the standard HT where any arbitrary shape can be used as the prior for object detection [Ball 81]. GHT parametrizes a shape with the

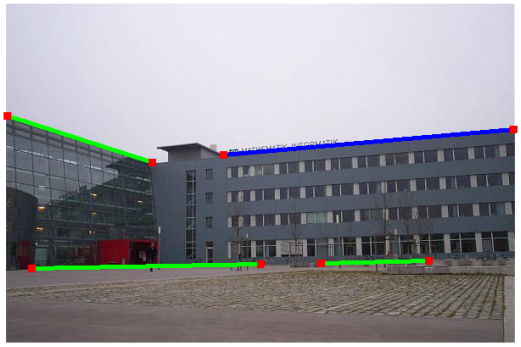
(a) Input image



(b) Edge map



(c) Parameter space accumulator



(d) Back-projected Hough lines

FIGURE 5.3.: Linear Hough Transform using polar representation. (a) Target image where dominant lines need to be located. (b) Edge map using Canny edge detector. (c) Polar accumulator computed over the edge map with respect to distance ρ and angle θ. Red squares are positioned on the peaks of the array. (d) Back-projected lines detected through the linear Hough Transform are overlaid on the input image. The line with the maximum number of votes is indicated by blue. Figures are courtesy of Javad Fotouhi.

offsets of each shape element from a reference point. This parametrization is stored in a look-up table. This table is later used for detection where elements in a target image vote for the hypotheses that might have generated them. The peaks in the Hough space created by accumulating such votes correspond to the parameters (such as the chosen reference point) of possible target object.

GHT parametrizes or learns an object shape by mapping shape features to an arbitrary reference point. The mapping is defined in terms of offset vectors which are stored in a lookup table R, so-called the *R-table* [Ball 81]. The R-table is in fact a codebook which groups offset vectors according to the discrete orientations of corresponding gradients. During testing or detection, edge points as well as their gradients in the target image are extracted. Then, for each edge point, the R-table is indexed based on the gradient orientation. The offset vectors stored in the corresponding orientation bin are added to the coordinate of the edge point. Resulting coordinates define the bins in the accumulator that

Gradient Direction Bin $\Delta\phi_j$		**Voting Vectors $\mathbf{V}_j$**
0	$\Delta\phi$	$\mathbf{V}_0$
1	$2\Delta\phi$	$\mathbf{V}_1$
2	$3\Delta\phi$	$\mathbf{V}_2$
$\vdots$	$\vdots$	$\vdots$

TABLE 5.1.: R-table $R(\phi)$ used in the Generalized Hough Transform

need to be incremented.

A common choice for the reference point of the shape is the center of mass, which is stable and introduces less accumulation errors. Let $\Delta\phi$ be the orientation step, for every edge pixel in the set $\mathbf{P} = \{\mathbf{p} = (x, y) \in \zeta(I)\}$ where $\zeta(\cdot)$ is an edge operator, the orientations $\phi(\mathbf{p})$ and offset vectors to the reference point $\mathbf{o}$ is computed and stored in the R-table $R(\phi)$ as illustrated in Table 5.1.

Offset vectors corresponding to a single bin are calculated as

$$\mathbf{V}_j = \{\mathbf{r} | \mathbf{o} - \mathbf{r} = \mathbf{p}, \quad \mathbf{p} \in \mathbf{P}, \quad j\Delta\phi \leq \phi(\mathbf{p}) < (j+1)\Delta\phi\} \tag{5.5}$$

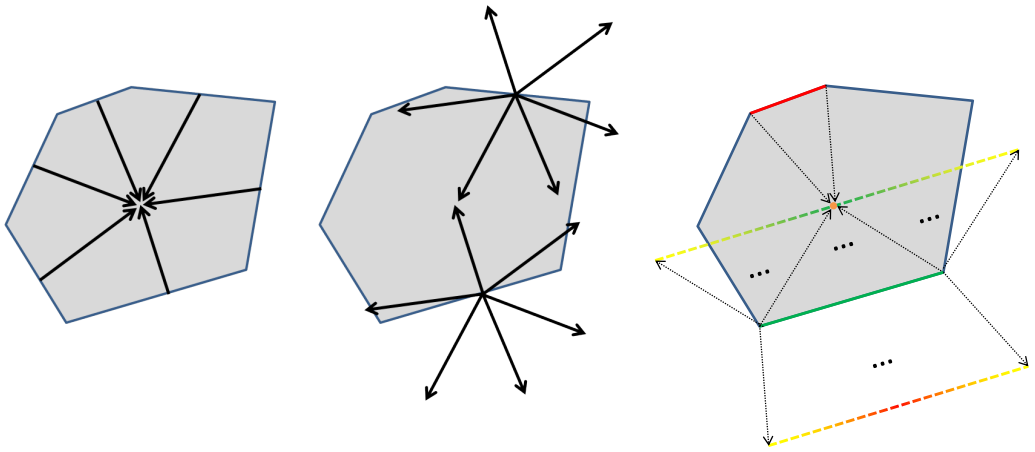
Figure 5.4 depicts the typical voting process with and without using edge orientation information in the R-table. Figure 5.4(a) shows the offset vector for a sample subset of edge points of an arbitrary shape with a given reference point. In Figure 5.4(b), a voting scenario on the same shape without using orientation information is shown. This is in fact equivalent to having only a single row in the R-table during training. In this case, for each candidate edge point, all of the stored offset vectors are retrieved and applied. Obviously, this leads to a very dispersed accumulator in the Hough space as well as a higher computational complexity.

Figure 5.4(c) shows how orientation information can be used when indexing an R-table. For clarity only the points on the line segment highlighted with green color are used for voting. Since there is another line segment with similar edge orientation (highlighted with red color), offset vectors coming from that segment during learning are also used for generating votes for the green segment. For clarity again, only the offsets corresponding to the endpoints of the segments are shown. The dashed lines depict the distribution of votes in the image space. This clearly will lead to a more structured accumulator array in the end and significantly reduces the number of false votes.

Once the R-table is built, starting with an empty accumulator A with dimensions equal to the size of a test image, every edge pixel $\mathbf{p} \in \mathbf{P}$ in the test image with gradient direction $\phi(\mathbf{p})$, we let all vectors in the corresponding row of the R-table vote by incrementing $A(\mathbf{a})$, where $\mathbf{a}$ is defined as

$$\mathbf{a} = \mathbf{p} + \mathbf{V}_j, \quad \forall \mathbf{p} \in \mathbf{P} \quad j\Delta\phi \leq \phi(\mathbf{p}) < (j+1)\Delta\phi. \tag{5.6}$$

The estimate of reference point in the test image is detected as the maximum in the accumulator. The resolution of the accumulator as well as the orientation step size can be used to



(a) Offset vectors stored in R-table (b) Voting without orientation (c) Voting with orientation

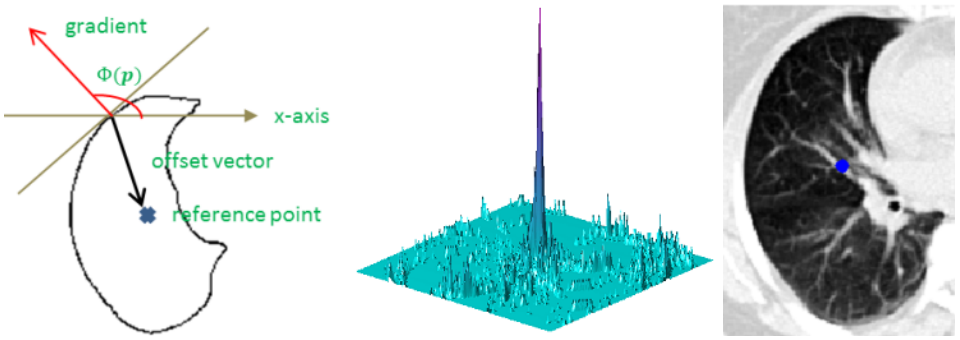
FIGURE 5.4.: Illustration of Generalized Hough Transform voting process. (a) Stored offset vectors corresponding to a sample set of edge points. (b) Voting based on the same shape with a subset of edge points without using orientation of edges. Obviously, there are too many votes in arbitrary directions coming from all edge points in the template making a inference difficult. (c) Voting from a selected boundary segment (green) using gradient orientation information. Green and red dashed lines show the distribution of votes coming from the segments with the same color.

control the speed and the accuracy of detection. While a coarser sampling in the parameter space increases the robustness of detection to the discretization errors and having sharper peaks in the vicinity of the parameters for the correct reference point, a coarser quantization of the R-table rows (gradient orientations) reduces the sensitivity of detection to errors in gradient estimation.

Figure 5.5 demonstrates the detection of lung reference point in a Computed Tomography (CT) image. Figure 5.5(a) shows how the orientation bin for an offset vector is determined. In Figure 5.5(b) 2D accumulator in the Hough space is shown. Notice the peak corresponding to the parameters (coordinates) of the lung reference point. The detected reference point overlayed on the input lung CT image is shown in Figure 5.5(c).

Depending on the image quality, level of noise, distortion, and occlusion the incrementation strategy might go beyond a single parameter. Gaussian weighted voting is an accepted strategy to increase confidence by casting weighted votes also in the neighborhood of a parameter. Voting along with gradient modulus, local curvature, or local consistency information is widely used as well. A set of connected boundary points are more consistent and reliable than unconnected segments of edges. However, incorporating neighboring information into the incrementation strategy requires more iterations since the points voting for a certain parameter have to be identified. Dynamic programming-based Hough transform was introduced to answer this problem [Ball 76].

Instances of the object can be rotated along the image axis by τ degrees or scaled by a factor of s. So far, the translation t is addressed in GHT. Letting v the model and u the expected shape in the target image, T_s and T_τ the respective scaling and rotation transformations,



(a) GHT template shape (b) GHT accumulator space (c) Detected reference point overlayed

FIGURE 5.5.: GHT in lung CT image (a) Shape template of lung used for R-table construction. Gradient binning includes 100 entries, leading to 100 rows in the R-table. (b) 2D accumulator array constructed from the lung CT image. A canny edge operator is applied on the input image. CT lung image introduces many false edges as there are complicated structures of bronchi channels, which are absent in the template model; therefore, the accumulator carries relatively large number of false votes. (c) The blue circle in the image reflects the position of the reference in the input image. Having the shape template and the input images of the same size has resulted identical reference points in the model and the input test. Figures are courtesy of Javad Fotouhi.

the relation between the model and the expected shape becomes

$$u(t, s, \tau) = t + T_s \left[T_\tau \left[v \right] \right]. \tag{5.7}$$

Assuming an isotropic scaling s along the entire image, the rigid body transformation and scaling of the template results in 4 degrees of freedom

$$\mathbf{a} = (x_o, y_o, s, \tau). \tag{5.8}$$

Computing all 4 degrees of freedom requires running an exhaustive search over a 4D accumulator. Ballard [Ball 81] studied the application of scaling and rotation through R-table transformation. Scaling is addressed as following

$$T_s[R(\phi)] = sR(\phi). \tag{5.9}$$

This implies that scaling can be directly encoded as scaling of individual vectors in the R-table. Rotating a structure will both change the gradient direction and the vector's direction

$$T_\tau[R(\phi)] = Rot[R(\phi - \tau), \tau] \tag{5.10}$$

Vectors stored as pairs of Cartesian coordinates $R(\phi) = \begin{bmatrix} \mathbf{r}_x(\phi) \\ \mathbf{r}_y(\phi) \end{bmatrix}$ are rotated and scaled as

$$T_s \left[T_\tau \begin{bmatrix} \mathbf{r}_x(\phi) \\ \mathbf{r}_y(\phi) \end{bmatrix} \right] = s \begin{bmatrix} \mathbf{r}_x(\phi) \cos\tau - \mathbf{r}_y(\phi) \sin\tau \\ \mathbf{r}_x(\phi) \sin\tau + \mathbf{r}_y(\phi) \cos\tau \end{bmatrix}. \tag{5.11}$$

The GHT transforms a global object recognition task into a much simpler peak detection in parameter space (i.e. Hough space) [Illi 88]. Locating a maximum or a cluster in parameter space can be achieved in a number of ways. These range from simple methods such as using a global threshold for the minimum number of votes to more complicated searching strategies like using hierarchical searching by dividing parameter space into multiple subspaces. More complicated tools such as mode-seeking methods including Mean-Shift [Chen 95] and its derivatives Medoid-Shift [Shei 07] and Quick-Shift [Veda 08] have also found their applications in Hough-voting based object recognition [Leib 04, Leib 08, Wood 14].

There is an ever increasing number of applications of GHT and its voting scheme in computer vision including object classification [Leib 04], detection [Gall 13] and tracking [Gode 13] to name a few. However, it has fewer applications in intensity-based image registration where there is a large potential of usage. In [Varn 13], GHT was used to estimate the initial pose of intraoperative images where possible poses of target object are learned a priori. Shams et al. [Sham 07] addressed the initialization of intensity based rigid registration by using standard Hough Transform on gradient fields. To the best of our knowledge, GHT has not been used for global regularization of intensity-based image registration.

5.2.2. Global Consistency in Intensity-based Image Registration

As discussed before, GHT has the favorable property of being robust to the presence of partial occlusions, noise and clutter in target images. Therefore, this property allows us to detect a previously learned object even if it is only partially visible in a novel image.

This observation lets us to exploit this property of the GHT to enable global regularization for intensity-based image registration of partially overlapping images *independent of the size of overlap between them*. To do this, given an object of interest, we learn a Hough space parametrization $\mathcal{P}_H$ from its prior data such as previous scans, statistical shape models or atlases. Afterward, following the traditional Hough-like voting frameworks discussed in previous sections, and using $\mathcal{P}_H$, the voting elements[2] in partial images vote for the hypotheses about the instance as well as the parameters of the object in a parameter space. Finally, a global consistency measure (GCM) for the alignment of partial images in image space is inferred from the resulting independent vote distributions.

Distribution where strong hypotheses are gathered together indicates a better global consistency in contrast to a scattered distribution of strong hypotheses. This in fact means that local voting elements contained in partial observations agree on the parameters of the global shape of the underlying object. Although this procedure seems at first as an object detection scheme, the ultimate goal is not the localization of the object in a novel image, but rather *the assessment of consistency in the configuration of the parts of the object observed in various partial images*. This is the key difference between the existing object recognition works and the presented work.

In the voting-based inference, various types of voting elements can be used ranging from contour elements [Opel 06, Shot 08, Ferr 10] to more complex features such as key-

[2]The elements do not have to be necessarily in the overlapping region which is one of the key idea behind this work.

points [Leib 04] or image patches [Gall 13]. Here, only edge pixels are considered as voting elements which can be extracted via standard edge detection methods such as Canny.

Global Consistency Measure

Let

$$\mathbf{I} = \left\{ I_i : \Omega_i \rightarrow \mathbb{R}, \Omega_i \subset \mathbb{N}^N \right\}_{i=1}^{g} \tag{5.12}$$

be a set of g partial images of an object to be registered together in order to widen the field-of-view, where N is the image dimension. Furthermore, let Ω_c be the common spatial domain of partial images and

$$\mathbf{T} = \left\{ T_i : \mathbf{x}_c \mapsto \mathbf{x}_i, | \mathbf{x}_i, \mathbf{x}_c \in \mathbb{R}^N \right\}_{i=1}^{g} \tag{5.13}$$

be the corresponding set of transformations parametrized by $\mathbf{t} = \{t_i\}_{i=1}^{g}$, which, when optimal, will bring partial images into spatial alignment in the common coordinate frame.

We further define Hough images

$$\mathbf{H} = \left\{ H_i(\mathbf{x}|I_i, t_i, \mathcal{P}_H) : \Omega_{\mathbf{a}} \rightarrow \mathbb{R}, \Omega_{\mathbf{a}} \subset \mathbb{N}^M \right\}_{i=1}^{g} \tag{5.14}$$

where M is the dimensionality of Hough space and $\mathbf{a}$ is a coordinate in the Hough parameter space. The goal in this work is to define a measure using the Hough images (i.e. vote distributions in parameter space) which should reflect the consistency in a particular configuration of partial images that have generated them. We will investigate different ways of defining such a measure and discuss their strengths and limitations in what follows.

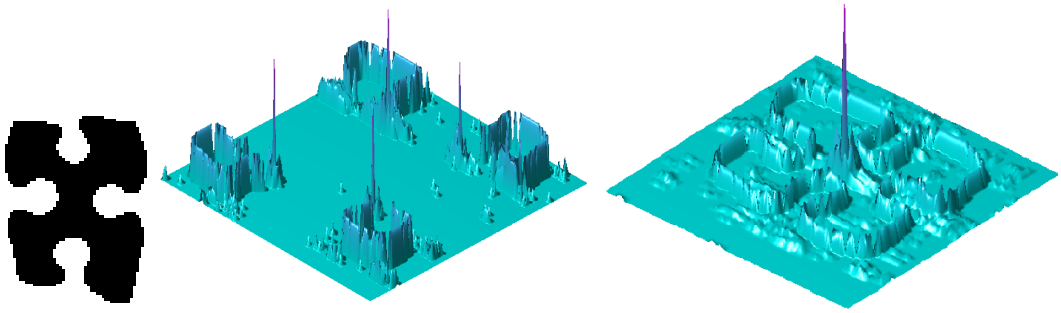
We interpret the compactness of votes in the joint Hough space as the global consistency criterion among the sub-images. In this context, compactness is defined as the measure of clusteredness in the joint parameter space, or simply as the peak alignment between the sub-accumulators. The motivation behind the compactness is demonstrated in Figure 5.6 where changes in the vote distributions are observed based on partial image configuration. As a result of proper alignment, vote distributions coming from partial images lead to the maximum compactness in the joint parameter space.

In the following we assess metrics which represent the amount of compactness in a distribution function. The goal is to cluster a dispersed distribution in Figure 5.7(a) and achieve a compact distribution as in Figure 5.7(b) [3].

Entropy of Joint Vote Distributions The accumulator in GHT demonstrates the number of counts of votes for a parameter for being the predefined reference point. Therefore, if we normalized the vote distribution, the GHT voting space stands as a probability distribution function (PDF) for the parameter of the reference point. Thus, one can treat it like a normalized histogram.

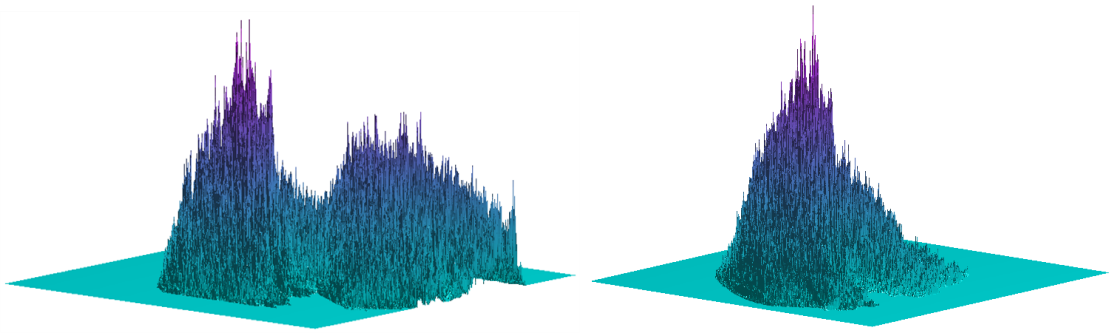
Entropy of a distribution is a powerful metric of uncertainty and is widely used to assess the structuredness of a probability distribution. This is in fact a measure of the amount of information content in a distribution. The Shannon entropy [Shan 48] h of a random

[3] Hough-space based similarity measures discussed in this section have been implemented by Javad Fotouhi as part of his master thesis.



(a) Template (b) Sub-images set apart (c) Sub-images aligned properly

FIGURE 5.6.: Superimposed votes in Hough space (a) Template model learned in the R-table. (b) The template in (a) is split into four partial images and are set apart from each other. In the next step, sub-images are set to cast votes for position of the parameters of the learned model. In this figures, votes are overlaid with the edge points. Partial images construct four small peaks in their local coordinate frames. (c) Sub-images are registered properly with respect to the global model. The four small local peaks are now superimposed, thus constructing a single global reference. The largest possible peak happens at this setup, therefore resulting in the most probable parameters for the global reference. Figures are courtesy of Javad Fotouhi.



(a) Dispersed accumulator (b) Compact accumulator

FIGURE 5.7.: Compactness in Hough space (a) Initial joint vote distibution in the joint parameter space constructed from two sub-accumulators. (b) Compact joint vote distributions resulting in the proper alignment of corresponding sub-images. Figures are courtesy of Javad Fotouhi.

variable X with possible values in $[x_1, \cdots, x_n]$ and a probability distribution function (PDF) $p(X)$ is defined as

$$h(X) = \sum_{x_i \in X} p(x_i) \log_2(\frac{1}{p(x_i)}) = -\sum_{x_i \in X} p(x_i) \log_2(p(x_i)). \tag{5.15}$$

Entropy can be interpreted as a measure of uncertainty about the random variable. Therefore, a uniform distribution has the maximum entropy since it has the most uncertain form of distribution. In contrast, a peaky distribution has always smaller entropy values due to

the fact that the probability of the random variable taking certain values is higher, leading to having more information about the behavior of the random variable.

Therefore, the Shannon entropy can be utilized to evaluate the compactness of the vote distribution in the joint parameter space. This should also reflect the extent of agreement between the predictions of partial images about the global properties of the underlying object. The compactness measure in the joint parameter space is defined as

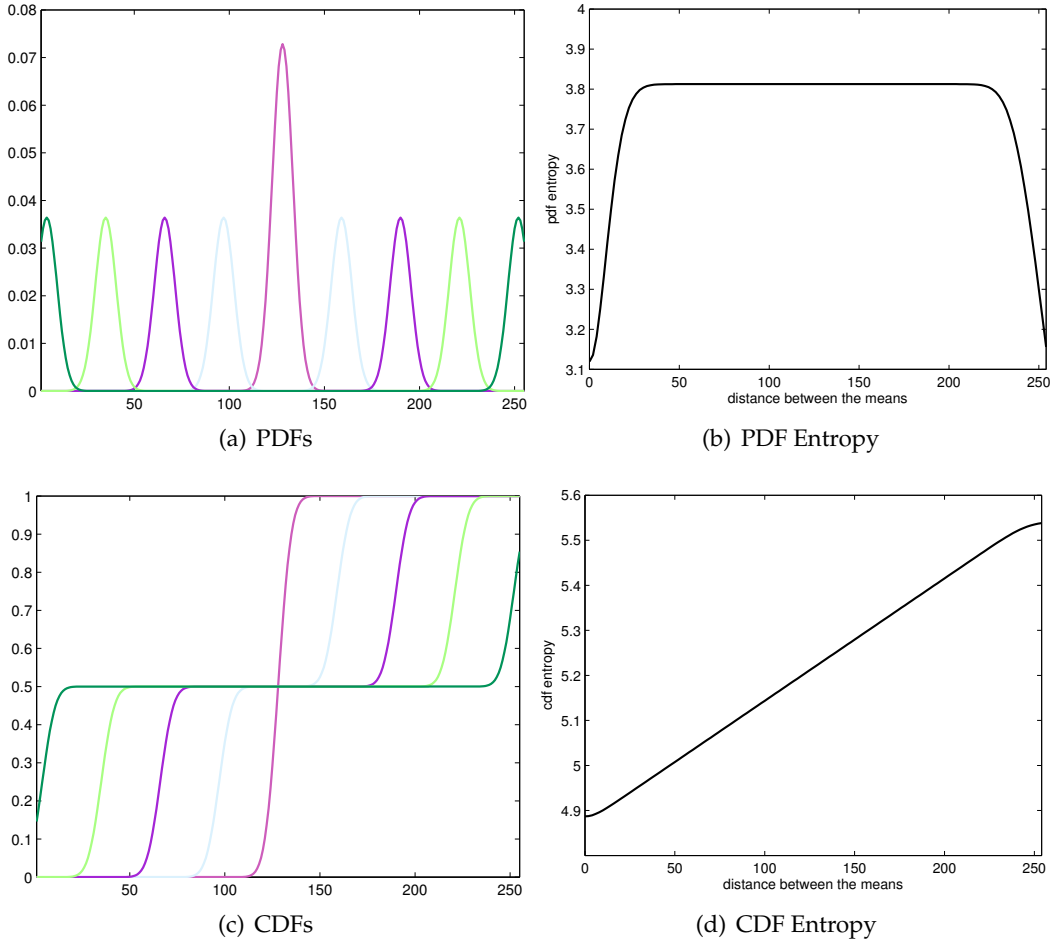
$$CM(\mathbf{I}, \mathbf{T}, \mathcal{P}_H) = h(\mathcal{H}) \quad \text{with} \quad \mathcal{H} = \bigcup \mathbf{H} \tag{5.16}$$

where $h(\cdot)$ is the Shannon entropy, $\mathbf{H}$ is the set of individual vote distributions and $\mathcal{H}$ is the joint vote distribution. The union is obtained simply by adding up the individual vote distributions. Ideally, a smaller entropy values are expected when the joint distribution has a single peak. Unfortunately, however, it is possible to have smaller entropy values when there are multiple peaks in the distribution. This behavior is in fact exploited in the mutual information similarity measure [Viol 97, Maes 97] for multi-modal image registration where a more clustered joint intensity histogram is desired regardless of the number of peaks.

In order to study the applicability of vote distribution entropy for consistency assessment, we have conducted an experiment. To this end, we created distributions consisting of multiple Gaussian components. First, we used only two components with equal variances. As shown in Figure 5.8(a), starting from a configuration where the components coincide, we increased the distance between their means. The idea was to observe the behavior of entropy of PDF with varying configuration of peaks in the distribution. Figure 5.8(b) shows that in most cases entropy cannot capture how the peaks are close or far from each other. Actually, this conclusion can be also drawn from the formulation of the Shannon entropy. A random shuffling of values in a distribution does not change the value of the entropy. Therefore, minimization of entropy will only result in a more clustered joint vote distribution which may not contain a single peak. However, if we change the definition of consistency between partial images from having the same predictions for reference point to having consistent predictions for multiple properties of the object, entropy of vote distributions can be successfully used as a consistency measure. We will go into more details about this in the discussion section.

To make the measure sensitive to spatial configuration of clusters in a distribution, in other words, to have a measure that captures the difference in terms of closeness of peaks, we considered Cumulative Distribution Function (CDF). CDF is obtained by integrating a PDF. Therefore, different CDFs can be obtained if peaks are configured differently in a distribution. To this end, we calculated CDFs as shown in Figure 5.8(c) corresponding to the PDFs shown in Figure 5.8(a) and evaluated the entropies of them. Figure 5.8(d) shows the resulting entropies. It is clear that entropy of CDFs tell us more about how the peaks are located within the distributions. This gives an initial impression that the entropy of CDFs of joint vote distributions can be used as measure of consistency in terms of predictions of reference points in the Hough space.

In order to further evaluate the entropy of CDF, we introduced another random Gaussian component to the distributions and repeated the same experiment in Figure 5.8. Results shown in Figure 5.9 indicate that the entropy of CDF is still able to differentiate between far peaks and close peaks despite another randomly introduced peak in the distribution.



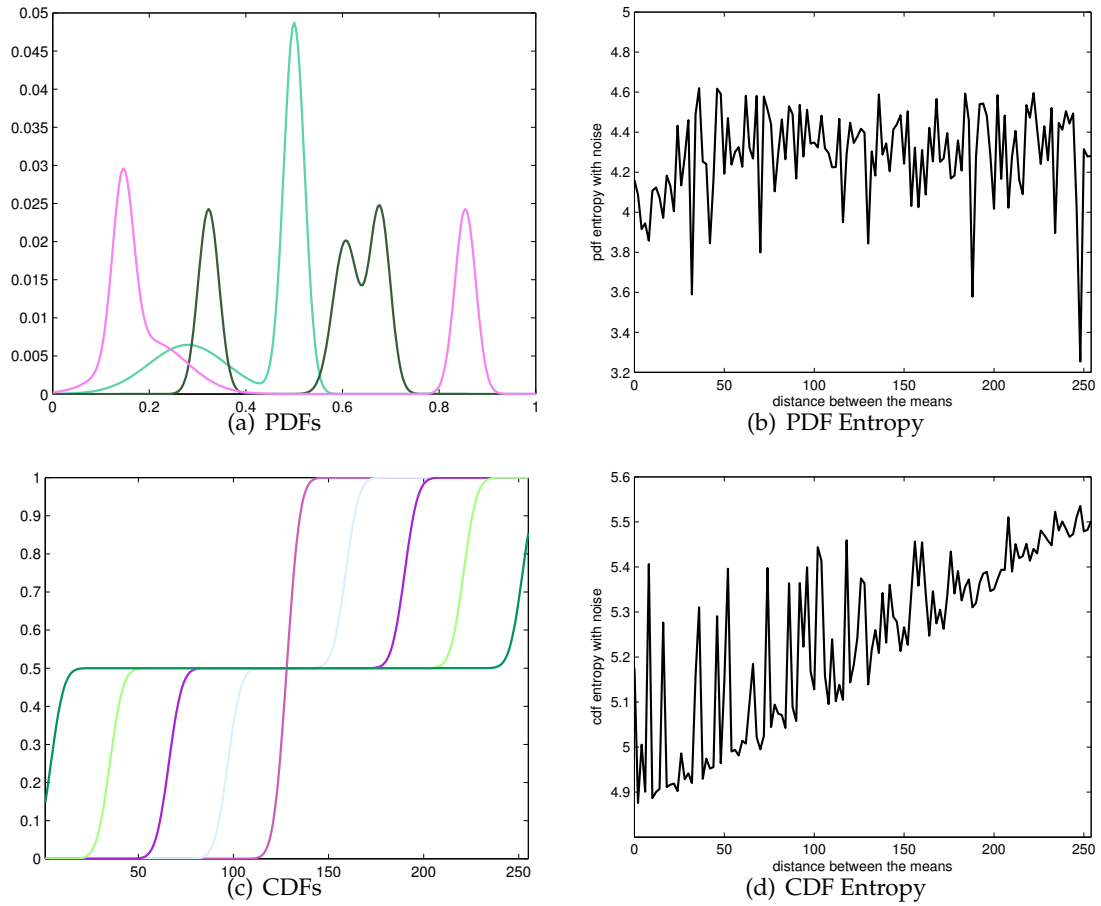
(a) PDFs (b) PDF Entropy

(c) CDFs (d) CDF Entropy

FIGURE 5.8.: Analysis of sensitivity of entropy to peak locations in distributions. (a) shows PDFs with varying distances between Gaussian component means. For clarity only a subset is shown with different colors. (b) Entropy of PDFs with varying distances. (b) shows clearly that there is a large range where entropy cannot capture how the peaks are close or far from each other. (c) shows CDFs corresponding to the PDFs shown in (a). (d) Entropy of CDFs with varying distances. It is clear from (b) that CDF entropy is sensitive to how peaks are located in the distributions.

However, when the distribution is very peaky, that is, the two components are very close to each other, the third random component changes the expected behavior of this approach.

CDFs in higher dimensions can be obtained via integral images. Although this approach seems promising in toy examples, it does not produce sufficient results when applied on real joint vote distributions in higher dimensions. The reason for this could be the fact that CDFs are not probability distributions. Therefore, they may contain large flat regions in higher dimensions. Nevertheless, we still believe that further investigation of the usage of entropy is necessary for a better understanding of reasons of failure in real cases.



(a) PDFs (b) PDF Entropy (c) CDFs (d) CDF Entropy

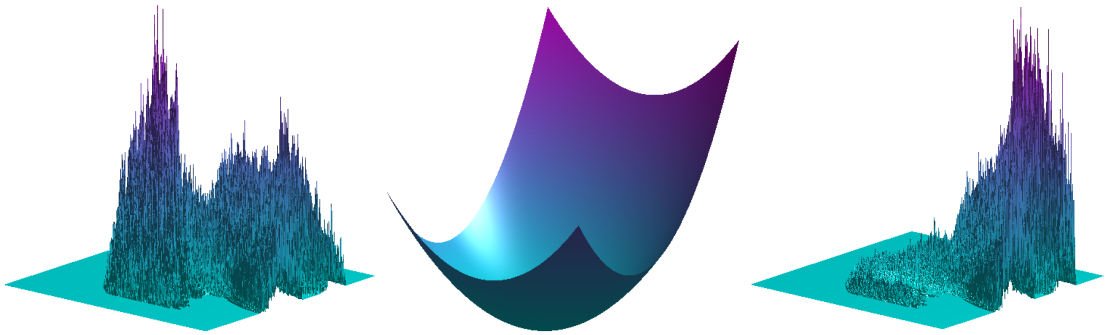
FIGURE 5.9.: Analysis of sensitivity of entropy to peak locations in distributions with noise. (a) shows PDFs with varying distances between two Gaussian component means along with another Gaussian component with random mean. For clarity only three of them are shown with different colors. (b) Entropy of PDFs with varying distances. (c) shows CDFs corresponding to the PDFs shown in (a). (d) Entropy of CDFs with varying distances. Entropy of CDF still gives reliable information regarding the spatial closeness of largest peaks in the distribution.

Weighted Distribution Functions The original goal here was to achieve a joint vote distribution where there is only a single peak which is the superimposition of the strongest parameter hypotheses generated by partial images. Entropy of the distribution is insensitive to locations of the peaks in the distribution and it tries to cluster the votes instead of forcing a single peak.

In order to have a single sharp peak in the distribution, we studied weighting the bins in the joint vote distribution with respect the strongest peak. The perception of weighted distribution functions is originated from k-means clustering. The compactness term in this case is defined as

$$CM(\mathbf{I}, \mathbf{T}, \mathcal{P}_H) = \sum d(\mathbf{a})\mathcal{H}(\mathbf{a}). \tag{5.17}$$

The distance map $d(\cdot)$ associates larger weights to distant coordinates $\mathbf{a}$ in the parameter



(a) Joint vote distributions (b) Distance function $d(\cdot)$ (c) Weighted joint vote distribution

FIGURE 5.10.: Weighted joint vote distribution. (a) shows a joint distribution with two distinct peaks. (b) Distance function has its minimum at the position of the largest peak. Bins farther from the peak are weighted by their distances. (c) The largest peak in the distribution is suppressed to zero. Minimization of Equation (5.17) will force the distant elements to get closer to the largest peak. Figures are courtesy of Javad Fotouhi.

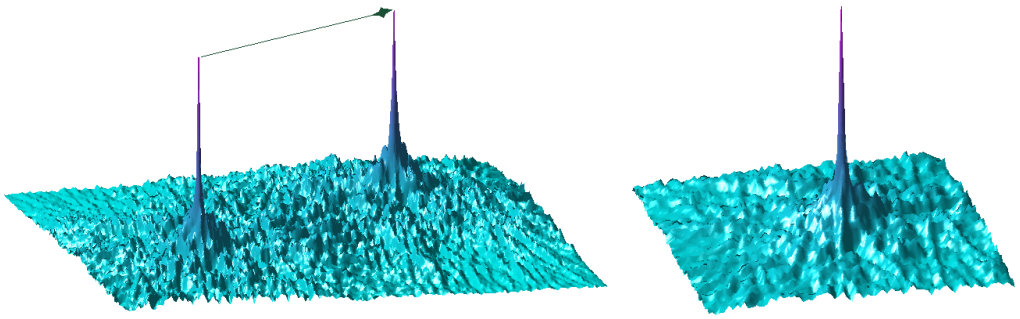
space from the peak. It is described as

$$d(\mathbf{a}) = \|\mathbf{a} - \mathbf{a}_{max}\| \quad \text{with} \quad \mathbf{a}_{max} = \arg\max_{\mathbf{a}} \mathcal{H}(\mathbf{a}). \tag{5.18}$$

In this way, the votes in the vicinity of $\mathbf{a}_{max}$ are suppressed, therefore, when minimized, this term forces all strong peaks far from $\mathbf{a}_{max}$ to get closer by assigning higher costs to them. Other peaks far from the largest one have usually large vote counts. Thus, they are affected more than any other bin with less number of votes. Figure 5.10 illustrates the effect of weighting by the distance map $d(\cdot)$. Although this approach seems to be more promising in terms of being sensitive to the spatial locations of the peaks, it is not robust enough to spurious peaks with large number of votes which do not correspond to the reference point of the object of interest in a partial image. Due to boosting with the distance, such spurious peaks can be easily falsely aligned with the largest peak.

Peak Distance Minimization Peaks are the only meaningful properties of the Hough space. Superimposing the peaks generated by different partial images is equivalent to solving the problem of configuration, which is the maximum compactness in this context. If the hypotheses are in an agreement, partial images are positioned correctly relative to the object parametrization. However, having all the peaks clustered does not answer the problem of rotation. Rotation of a partial image around the reference point may result in a similar distribution which is the reason for not being sensitive to rotation. Including more constraints in terms of structural consistency can solve the issue of rotation. We will come back to this issue in the following paragraphs.

Now, assuming that rotation is not an issue, the most straightforward way of aligning peaks in partial vote distributions can be achieved by first detecting them and then minimizing the distance between them. Following this idea, we can define a global consistency



(a) Joint vote distribution with distant peaks

(b) Vote distribution peaks superimposed

FIGURE 5.11.: An illustration of peak distance minimization. (a) Joint vote distribution with respect to the initial guess. (b) Peaks are superimposed in the proper alignment in terms of translation. Figures are courtesy of Javad Fotouhi.

measure (GCM) as the following

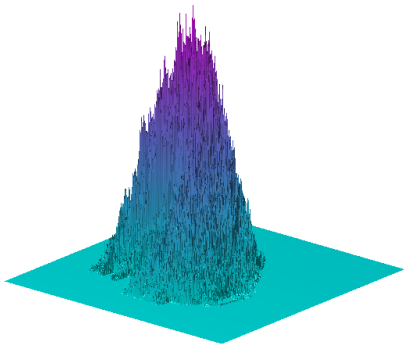
$$GCM(\mathbf{I}, \mathbf{T}, \mathcal{P}_H) = \frac{2}{g(g-1)} \sum_{i=1}^{g} \sum_{j=i+1}^{g} \left\| \arg\max_{\mathbf{a}_i} H_i(\mathbf{a}_i) - \arg\max_{\mathbf{a}_j} H_j(\mathbf{a}_j) \right\|. \tag{5.19}$$

Equation (5.19) minimizes the overall distance between all pairs of largest peaks in individual Hough images. Figure 5.11 illustrates superimposed partial distribution peaks. Note that as long as the peaks in partial distributions are accurately identified, this measure forces the partial images to come to an alignment in terms of reference point location. Peaks are detected using a very simple $\arg\max$ operation where the index with the largest number of votes is selected. In order to cope with quantization artifacts, Gaussian smoothing is performed before the peak detection. Later, we will discuss the possible issues with peak detection and other kinds of deformations.

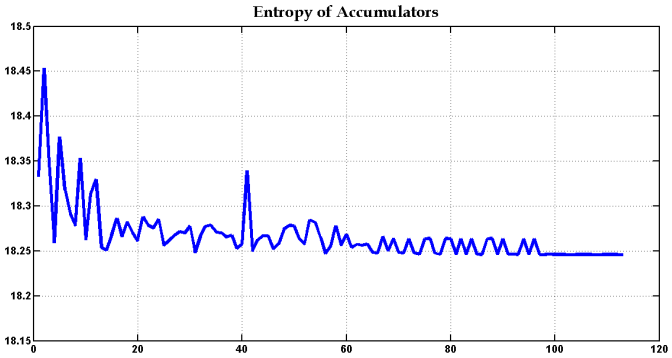
Evaluation of the Compactness Measures Figure 5.12 compares two main compactness measures of entropy and GCM that we have discussed so far. The peak size for the same alignment problem increases from entropy to GCM indicating that GCM has the highest potential of superimposing peaks. Consequently, the vote distribution in Figure 5.12(c) has the sharpest distribution compared to distributions in Figure 5.12(a).

Table 5.2 compares the statistics of the obtained joint accumulator. As expected from the formulation, the peak distance is smaller in GCM compared to entropy. However, the number of non-zero votes in the joint distribution is larger in the case of GCM. This is indeed not a suprise since entropy tries to cluster the votes but not to reduce the number of peaks in the distribution. GCM, on the other hand, only overlays the peaks, while other votes are more spread.

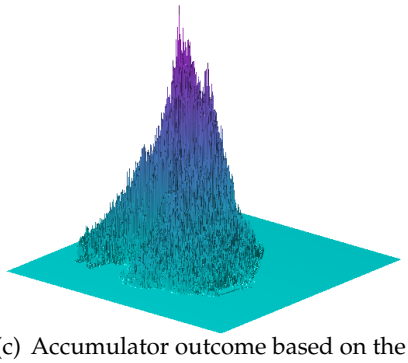
In this work, the sum of pairwise distances between the maxima in Hough images is considered as the global consistency measure Equation (5.19). Using the same Hough space size for each partial image, the goal is to bring the strongest hypotheses into a cluster. Equation (5.19) is a function of parametrization, partial images and corresponding



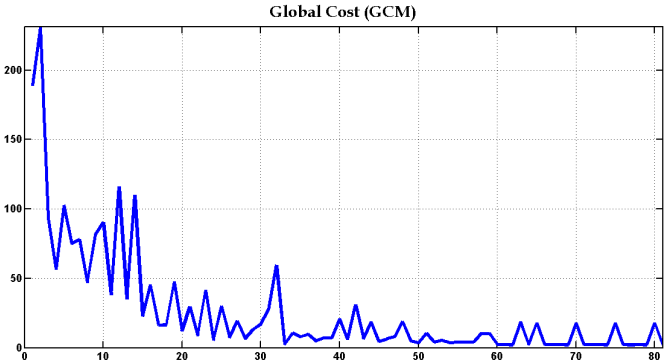
(a) Accumulator outcome based on the entropy of PDF



(b) Entropy of PDF



(c) Accumulator outcome based on the GCM



(d) GCM

FIGURE 5.12.: Comparison among the compactness measures. Each distribution is constructed from two sub-distributions by superimposing them and are obtained by minimizing the corresponding compactness cost. (a,c) Final joint distributions based on the entropy and GCM minimization, respectively. (b,d) Compactness measures of entropy and GCM over iterations of an alignment procedure. Figures are courtesy of Javad Fotouhi.

transformations. The goal here is then to minimize the GCM in order to increase the agreement between the prediction of partial images about the location of reference point in the parameter space.

So far, transformations other than translation have been ignored for simplicity. Since partial observations of the objects may not be in the same pose as the learned template, it is necessary to include other parameters such as scaling and rotation depending on the expected amount of deformation in the partial images. For instance, in 2D, isotropic scaling and rotation around image center can be included in the Hough space, leading to a 4D parameter space. Higher order transformations can be handled in a similar manner, but leading to a very high dimensional parameter space. We will discuss such issues related to the complexity due to higher order transformation is the discussion section.

Using the proposed setting, it is possible to find the parameters of transformations that generated partial observations from a template. Afterward, one can simple apply the inverse of each transformation to map partial images to the coordinate frame of the template.

Method	Final peak distance	Final non-zero bins
Entropy of PDF	36	412817
GCM	0	553324

TABLE 5.2.: Comparison of entropy-based and peak distance minimization based compactness measures.

However, although this can solve the registration problem, it becomes merely a pose estimation procedure for partial images. Our goal in this work is not pose estimation but rather to employ the agreement between the hypotheses in parameter space as a global regularization in registration where both global and local consistency are simultaneously ensured. Therefore, we prefer to iteratively compute the GCM during registration in order to assess the global consistency as the optimization evolves. This measure will be then coupled with a local intensity-based similarity term which will account for the quality of alignment in terms of intensity in the overlapping region.

Intensity-based Image Registration

We define the local similarity measure (LSM) in terms of intensities as

$$LSM(\mathbf{I},\mathbf{t}) = \frac{2}{g(g-1)} \sum_{k=1}^{|\Omega_c|} \sum_{i=1}^{g} \sum_{j=i+1}^{g} \xi(I_i(T_{t_i}^{-1}(\mathbf{x}_k)), I_j(T_{t_j}^{-1}(\mathbf{x}_k))) \tag{5.20}$$

with T_{t_i} being the parametrization of T_i by t_i. Finally, we pose the alignment of all partial observations as an optimization problem such that optimal transformations $\mathbf{T}$ optimizes an energy $\mathcal{E}$. Optimal parameters $\mathbf{t}$ then can be estimated via the following equation

$$\hat{\mathbf{t}} = \arg\min_{\mathbf{t}} \mathcal{E}(\mathbf{t}|\mathbf{I}, \mathcal{P}_H) \tag{5.21}$$

where $\mathcal{E}$ is defined as

$$\mathcal{E}(\mathbf{t}|\mathbf{I}, \mathcal{P}_H) = LSM(\mathbf{I},\mathbf{t}) + e^{GCM(\mathbf{I},\mathbf{t},\mathcal{P}_H)-\rho}. \tag{5.22}$$

In this functional, LSM serves as a data fidelity term evaluated in the overlapping regions, whereas the exponential part is a global regularization term evaluated in the Hough space. We use the exponent of the GCM in order to exponentially punish globally inconsistent misalignment though it should still work without an exponent if a weighting factor is carefully chosen. ρ in the exponent is a constant controlling the amount of regularization based on the uncertainty in GCM. The choice of ρ depends on several factors such as the quantization of the parameter space as well as on the image content. For large parameter spaces, there is more uncertainty regarding the location of maxima in the Hough space, therefore, inconsistent local alignments should be penalized less. Similarly, object deformations as well as clutter in the scene cause dispersions in Hough space, thus, leading again to an increased uncertainty. When there is more uncertainty, ρ should be set to a smaller value to give more flexibility to the local term. In the experiments we will perform a sensitive analysis for this parameter.

The data term $\xi(\cdot)$ can be chosen according to the modalities being registered. We have defined several similarity measures for intensity-based image registration in Section 3.1.5. In this work, we have chosen to use Normalized Cross Correlation (NCC). Although different kinds of transformations given in Section 3.1.5 can be recovered using the proposed technique, in this work, we will demonstrate only similarity transformations in 2D by assuming that the optimal transformations $\mathbf{T}$ include translation, rotation and isotropic scaling only, leading to 4D Hough images. For the optimization of Equation (5.21), various optimization techniques as listed in Section 3.1.5 can be used. While some of the procedures require the derivative of the registration functional, others operate without derivatives. For practical purposes, we have chosen Nelder-Mead Simplex algorithm as part of the NLopt package [John] as the optimizer in our registration framework.

5.2.3. Global Consistency Measure as a Regularization Term

Before moving on to the experimental validation of globally consistent intensity-based image registration, in this section, we will theoretically demonstrate that the global consistency measure (GCM) that we defined in Section 5.2.2 increases the capture range of an intensity-based image registration even when there is a sufficient overlap between images to be registered together.

Initial positioning is often an issue in general intensity-based image registration which is usually addressed by allowing large step sizes in the beginning of the transformation optimization. Another common technique is to use a multi-resolution approach where global motion is captured in the higher pyramid levels.

Here, we will show that a registration method becomes insensitive to initial positioning of images thanks to the global consistency measure which is independent of the overlap. This sounds similar to using a hybrid image registration method where global motion is compensated by the feature-based term. However, although feature-based image registration approaches need an overlap to match detected landmarks, the method proposed here works even when there is no overlap between the images to be aligned together.

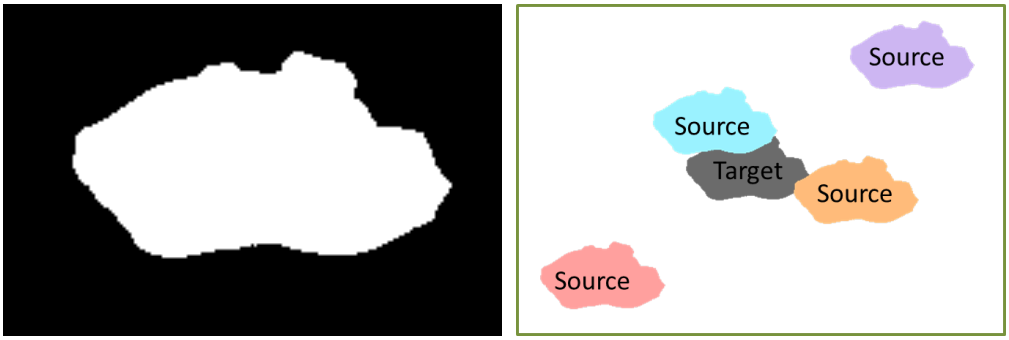
To this end, here, we study an alternative version of the functional given in Equation (5.22). We propose the following formulation

$$\mathcal{E}(\mathbf{t}|\mathbf{I}, \mathcal{P}_H) = LSM(\mathbf{I}, \mathbf{t}) + GCM(\mathbf{I}, \mathbf{t}, \mathcal{P}_H) e^{(1-\rho/GCM(\mathbf{I},\mathbf{t},\mathcal{P}_H))} \tag{5.23}$$

where GCM is not only in the exponent any more. In this formulation, GCM dominates the overall cost until its value is on the order of ρ. As its value gets smaller, the contribution of GCM to the overall cost will be negligible leading to an implicit handshake between the local and global terms. In the following, we will study different behaviors of Equation (5.23) as we vary the parameter ρ. For this experiment, we use two identical images of size 180x120 as shown in Figure 5.13(a).

To see the effect of the GCM, we varied ρ in $[0, 400]$ with increments of 100 where a value of 0 implies full regularization whereas a value of 400 means negligible amount of regularization in this setting. To plot the cost function values, we have chosen to change only the position of the source images within a global image canvas. The ranges of translations are $[-150, 150]$ and $[-100, 100]$ for x and y coordinates respectively with increments of 10.

Figure 5.14 shows the cost function behaviors in the vicinity of optimum. Each row shows the results obtained by using the respective values of ρ. From left to right, we show



(a) Test image (b) Different initial positionings

FIGURE 5.13.: (a) Test image used in the experiments. (b) Examples of initial positioning of source image with respect to the target image.

the function plots for the logarithm total cost, only intensity cost and only regularization term, respectively. The logarithm of the total cost was plotted to be able to visualize the local minima in the cost function. It can be clearly observed from the figure that while a full regularization allows no room to the intensity term, the total cost has a lot of local optima in the vicinity of the optimum when no regularization is used. The parameter ρ in fact shapes the regularization function as seen in the last column of Figure 5.14. Large values lead to a flat function around the largest peak in the Hough space and renders the regularization effective only in regions far from the peak. On the other hand, small values result in a more effective regularization also around the optimum.

While it is often an issue to fine tune a parameter in similar settings, we have observed that choosing the regularization parameter value close to the size of the object of interest gives sufficient results in practice. Moreover, the shape of the overall cost function does not change much unless a very large value is used for this parameter.

5.3. Experimental Validation and Results

We have conducted several experiments on synthetic as well as on real Ultrasound images to demonstrate the performance of the proposed registration framework. Since our claim is to have a globally consistent alignment between partial images regardless of the presence and the size of the overlap between them, through the synthetic experiments, we have analyzed the robustness of our approach to the size of overlap and to the varying degrees of imaging noise. For this purpose, a binary image with 400x190 pixels containing a certain shape was used as a template to learn the object. Two partially overlapping images are extracted from the same shape as the observed images as shown in Figure 5.15(a).

This pair of synthetic images is deliberately chosen to demonstrate the need for a global regularization in cases when the information contained in the overlapping region is not sufficient for an optimal alignment. Therefore, it should be noted that if the shape information is not utilized for the alignment of partial images in Figure 5.15(a), a registration method will be insensitive to horizontal translations in the overlapping region to a certain extent. The ultimate goal in this experiment is to reconstruct the same geometry starting from a

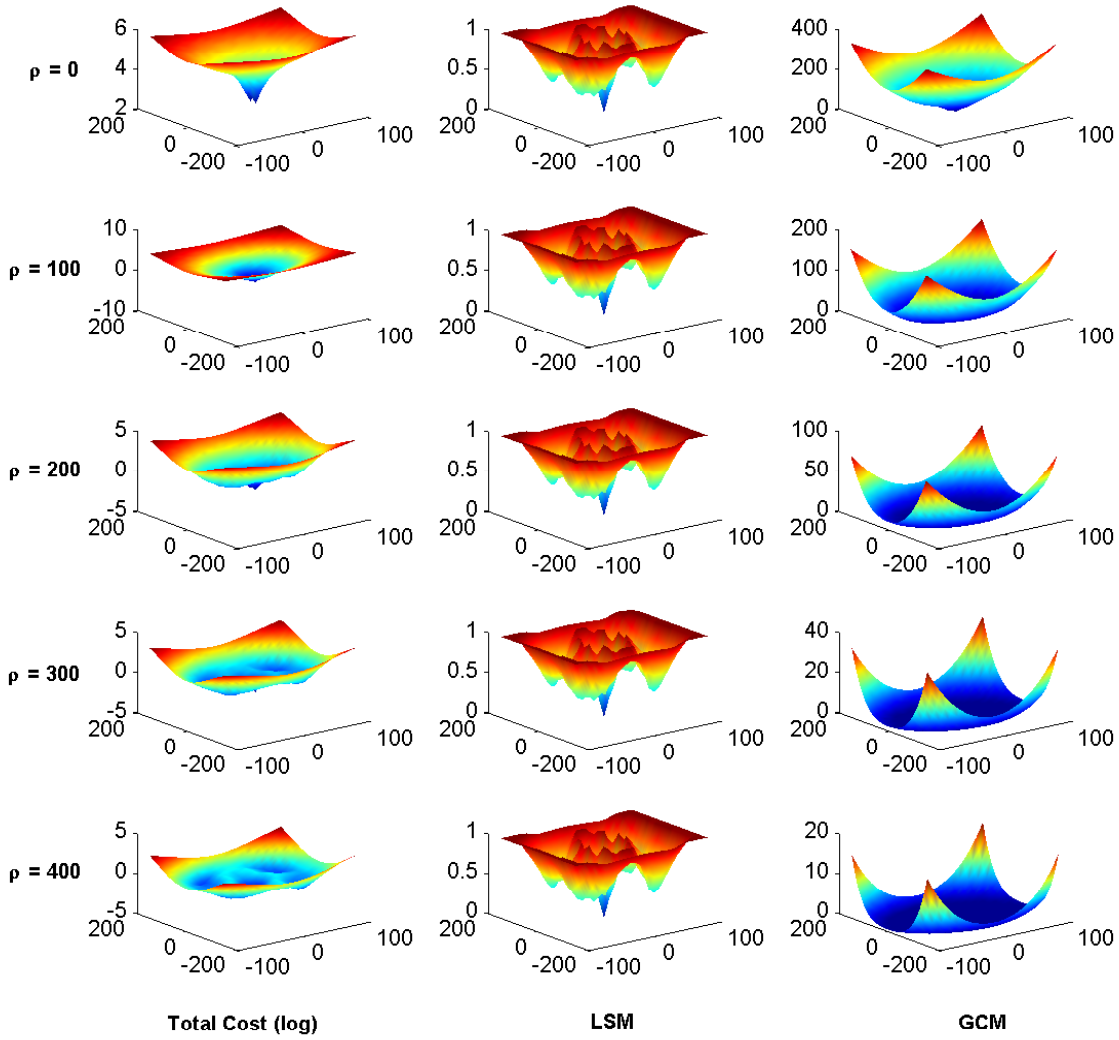
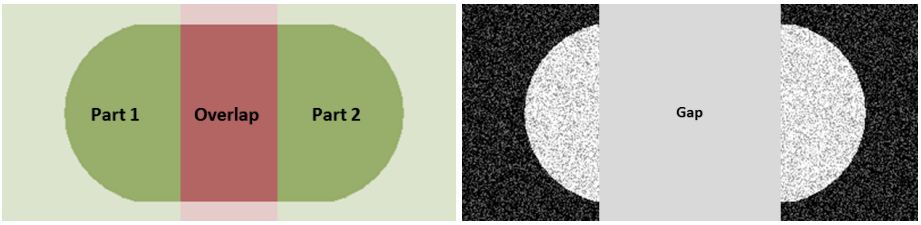


FIGURE 5.14.: Cost functions depending on the choice of the regularization parameter ρ in Equation (5.23). Please refer to text for the details.

randomly chosen initial relative positioning of partial images within a specified range.

To study the effect of the size of overlap, starting from a full overlap of objects in the partial images, we varied the overlap size by increments of 10% of the image width till we get a -60% overlap, which is a gap of size 60% of the image width. In order to simultaneously evaluate the robustness to the amount of image noise, for each increment of overlap size, we added uniform noise by varying its maximum relative to the the image dynamic range. Then, for each overlap size and noise level, we applied 20 combinations of initial rigid transformations with translations and rotations chosen from $[-200, 200]$ pixels and $[-50, 50]$ degrees respectively.

For synthetic experiments, the value of the peak uncertainty parameter ρ is set to 10. For



(a) Synthetic image with overlap (b) Synthetic image with gap and noise

FIGURE 5.15.: (a) An arbitrary shape and sample selection of partial images for experiments. (b) Same partial images with a physical gap of 40% of image width and uniform noise with a maximum of 100% of maximum intensity added.

the Nelder-Mead simplex optimizer as part of the NLopt package, $[100, 100, 20]$ are used as the initial steps for translations and rotation, respectively. Moreover, the parameter bounds are set to $[\pm 200, \pm 200, \pm 50°]$. The size of the canvas used as the common coordinate is set to three times the size of the template shape. Finally, 100 cost function evaluations are allowed for the optimizer.

For the evaluation, we warped the original partial images without noise using the optimal transformations obtained and compared to the corresponding part of the original full image using Dice similarity score. Dice similarity coefficient between two segmentations A and B is defined defined as $DSC = 2\,|A \cap B| \,/\, |A| + |B|)$ and its value is in $[0, 1]$ where 0 means no overlap and 1 means full overlap. Finally, in order to clearly show the effect of the global regularization term, we have conducted the same experiments without using the GCM.

Registration results, in terms of the mean Dice score of 20 runs and their standard deviation (STD), obtained using the proposed global consistency measure are shown in the top row in Figure 5.16. When compared to results on the bottom without using GCM, the results have two indications; first, the proposed method is able to align partial images even when there is gap between the partial images; second, it is robust to the amount of uniform noise in the images. It is also clear that the usage of GCM improves the capture range of local cost function and avoids undesired local optima. Please refer to Section 5.2.3 for detailed explanation regarding the cost function capture range.

In order to demonstrate the performance of the proposed intensity-based image registration technique on real images, we conducted an experiment where we took a pair of slices, each having 512x384 pixels with a pixel spacing 0.45mm, from a co-registered Ultrasound-Computed Tomography (US-CT) pair corresponding to the liver area. After segmenting the liver region in the CT image, we learned the parametrization $\mathcal{P}_H$ of the shape of the liver in the Hough space. Then, we cut the US image into two partially overlapping sub-images as shown in Figure 5.17(a). The goal of this experiment is then to reconstruct the original US image using both the local information contained in the US partial images and the global information through $\mathcal{P}_H$. The size of the overlap was about 15% of the original US image size.

With the purpose of quantifying the performance of registration, we applied random initial rigid transformations in a range around the optimum to achieve a statistically relevant result. 50 initial rigid transformations composed of translations in x and y and rotations

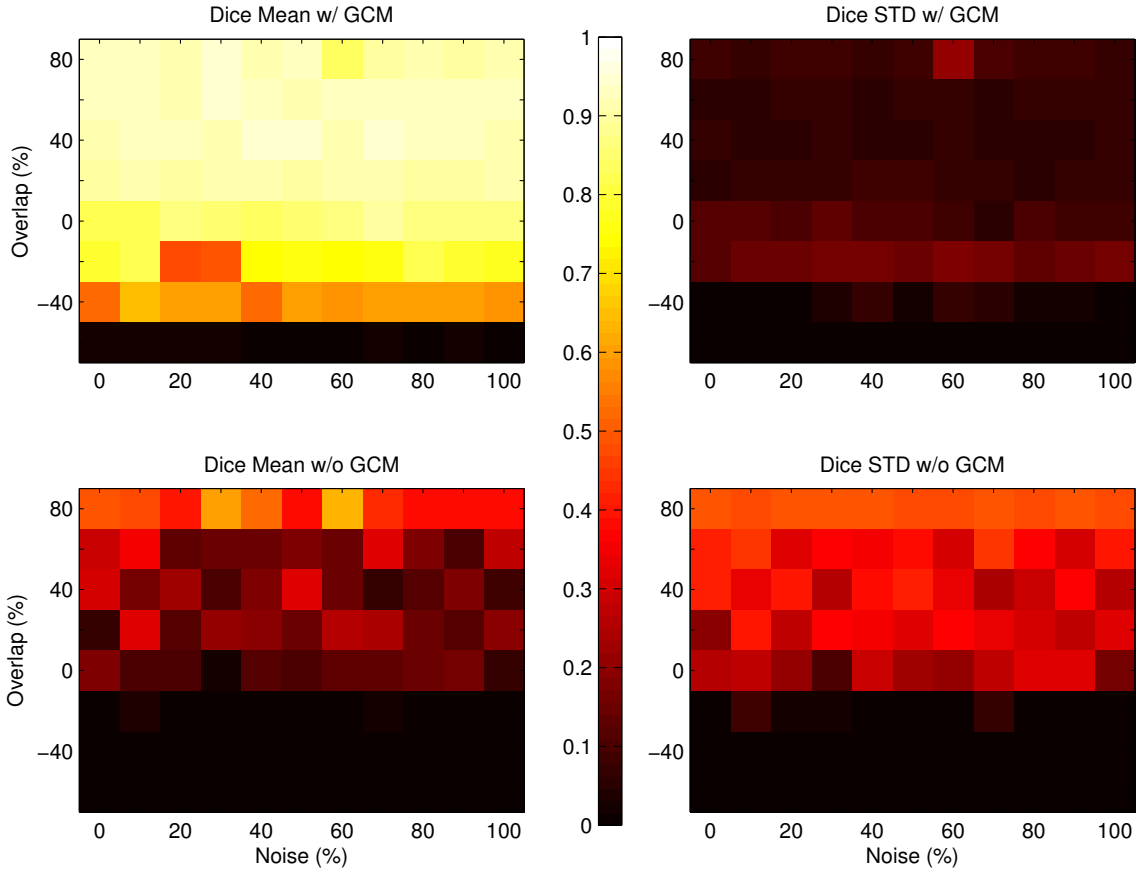
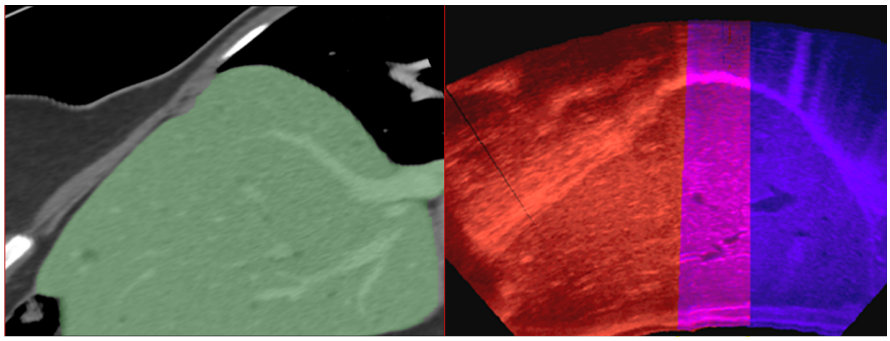


FIGURE 5.16.: Performance evaluation against noise and the amount of overlap using synthetic images. Top row: Dice score and its standard deviation (STD) using GCM. Bottom row: Dice score and STD without using GCM. A negative overlap value indicates a gap between images. Noise added to each pixel is uniformly drawn between $[0, p * max(I)]$ where p is in $[0, 100]$ as x-axis labels and $max(I)$ is the maximum intensity value in the image. It is clear that the method is robust to noise and can tolerate even gaps between the images if the GCM is used. Whereas, without GCM, Dice score is very low with a high STD even when the overlap is sufficient and it is not possible to register with a gap.

around the z-axis were randomly sampled from $[(-100, 300), (100, 500), \pm 30]$ relative to the optimum, respectively.

Parameter of the optimizer for the US experiment are set as follows. The value of the peak uncertainty parameter ρ is set to 50. $[100, 100, 20]$ are used as the initial steps of the optimizer for translations and rotation, respectively. Moreover, the parameter bounds are set to $[(-100, 200), (100, 600), \pm 50]$ for translation in x and y and rotations, respectively. The size of the canvas used as the common coordinate is set to three times the size of the template shape. 100 cost function evaluations are allowed for the optimizer. Finally, minimum relative changes in cost function values and parameter values are set to e^{-16} and e^{-8}, respectively.

For the evaluation of each registration run, we warped the corresponding segmentation



(a) US-CT image pair

FIGURE 5.17.: US-CT pair with CT segmentation overlayed on CT (left) and registered partially overlapping US images (right) where the overlapping region is highlighted.

masks of partial US images and compared with the segmentation of the original ground truth US image using Dice score. The overall mean, median and STD values with and without using GCM were recorded as 0.95, 0.99, 0.07 and 0.48, 0.45, 0.34, respectively. Results shown also in Figure 5.19(a) support our previous observations in terms of robustness. Moreover, we calculated the scores with respect to changing uncertainty parameter ρ in Equation (5.22) by varying its value in $[0, 100]$. Figure 5.19(b) shows that the large values of uncertainty lead to a degradation of performance which is expected due to the reduced amount of regularization giving more emphasis to the local similarity term which is more sensitive to initial positions. This is also valid for small values resulting in a very strict regularization, thus, making it sensitive to the possible errors in detecting peaks in the Hough space.

For further evaluation in terms of Target Registration Error (TRE), we used 16 pairs of landmarks manually extracted from the overlapping region. The same experiments described above were repeated by using US and CT segmentation masks as well as their slightly deformed versions respectively for learning $\mathcal{P}_H$. In this case US mask is also used for learning while in the previous experiments CT mask was employed as the template shape. Finally, the goal of introducing deformations to the template shapes is to demonstrate the performance of the method to variations in the object shape.

As seen in Figure 5.19(c), the best median TRE (8.46 pixels) was obtained by using CT for learning followed by using US (13.38 pixels). Slight performance reductions were observed in each case when their deformed versions were used for learning, indicating the tolerance of the method to small deformations. A median error of 127.88 pixels was obtained when GCM was not employed. Obviously, most of the registration runs without using GCM failed due to the small size of the overlap and the sensitivity to the initial parameters indicating a very limited capture range of the cost function. The slightly worse performance when using the US mask as the shape template compared to CT is due to the speckle and shadows in US images leading to false edges, thus, more uncertainty.

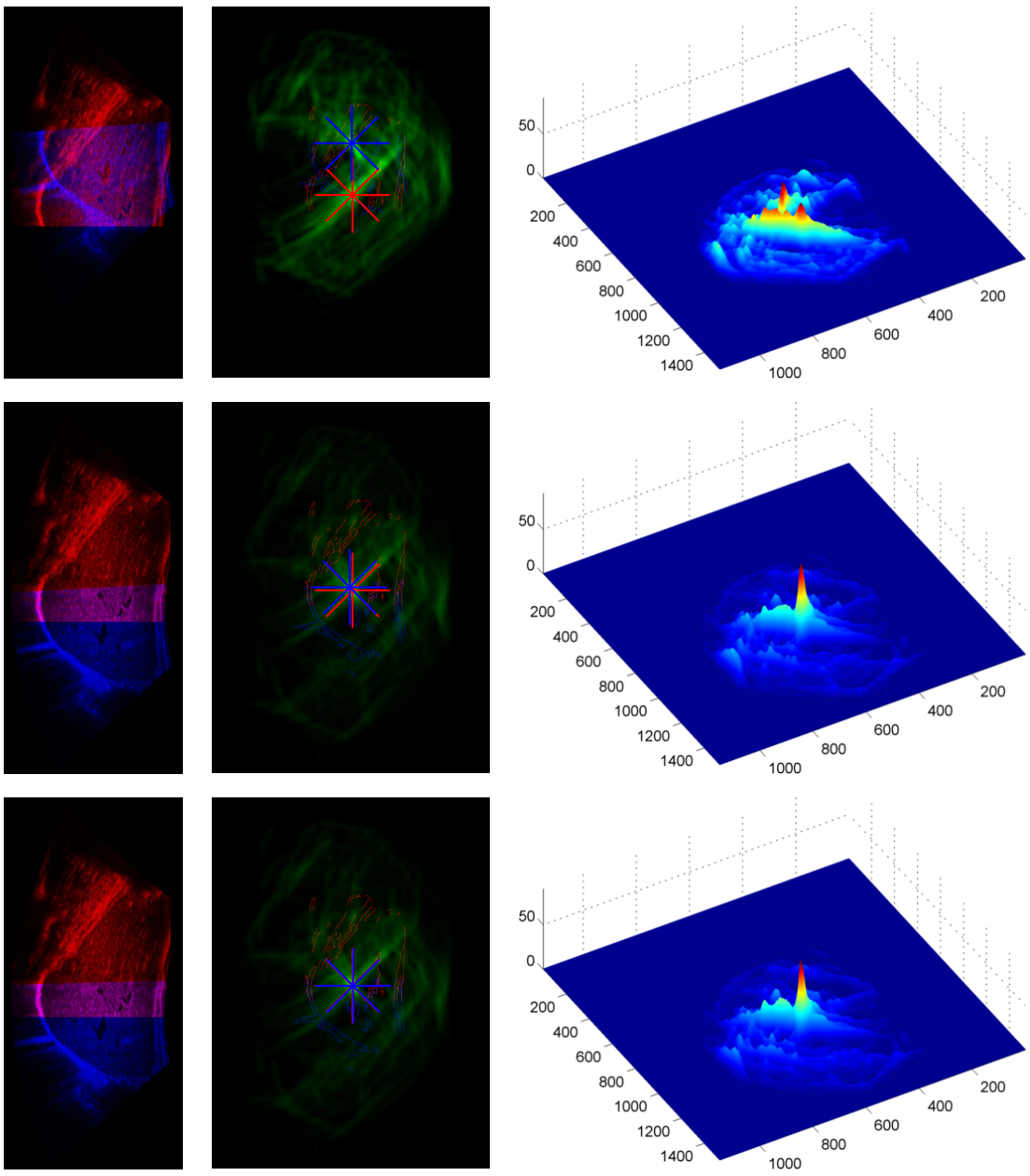
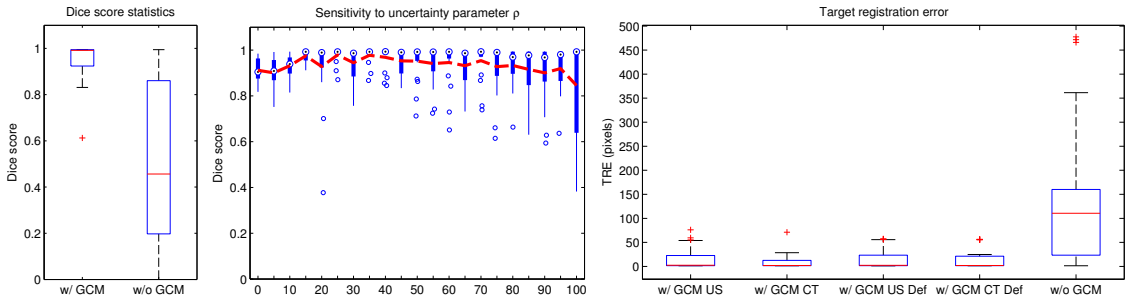


FIGURE 5.18.: Intermediate results from the US field-of-view extension experiment. From top to bottom: Results for the first, middle and last iterations of registration, respectively. From left to right: overlay of partial images within the image canvas, overlay of edge maps as well as votes, surface plot of joint vote distributions.

5.4. Discussion

The concept of registration that we have demonstrated in this chapter differs from model-based segmentation and registration methods in that we do not make any explicit use of prior data. We only learn a parametrization in order to employ it later for registering partial



(a) (b) (c)

FIGURE 5.19.: GCM evaluation on partial US images. (a) Dice scores with and without using GCM. (b) Sensitivity analysis with respect to the changing uncertainty parameter ρ (x-axis) in Equation (5.22). Dashed line represents the mean value. (c) TREs when US(w/ GCM US), CT(w/ GCM CT), deformed US (w/ GCM US Def) and deformed CT (w/ GCM CT Def) are used for learning. The last one is when GCM is not used at all.

images. This differentiates our work also from the model-to-image alignment methods [Toew 13] where the goal is to match a model to an observed image. Therefore, in our case, once the parametrization is learned, the full intensity distribution of the prior image is not required for registering partial instances of an object.

GCM term does not depend on the modality of images as long as features required for the GHT can be extracted from the images. Given that features as simple as edges can be efficiently used, this property makes the proposed framework suitable also for multi-modal applications where, for instance, partial observations of an object in different modalities need to be registered to have an widened field-of-view. Theoretically, as prior data, it is possible to use 1) a different modality, 2) an image of the slightly deformed target anatomy, 3) or a statistical shape model for learning the parametrization. This is one of the key features of the proposed concept allowing flexibility in model-based reconstruction. It should be further noted that we are not making any comparisons to the atlas-based registration techniques. Here, we propose an alternative concept with its own advantages.

Here, it should be emphasized that, during GCM computation, we did not employ any sophisticated techniques for finding modes in the Hough space, emphasizing the simplicity of the proposed concept. This, in turn, slightly effected the performance of regularization due to false detections by simply finding the parameter with maximum number of votes. Nevertheless, it is possible to augment the proposed Hough space parametrization approach by using advanced voting and mode finding techniques such as Mean-Shift [Chen 95] and its derivatives Medoid-Shift [Shei 07] and Quick-Shift [Veda 08] which are quite popular in computer vision applications. In this work, a theoretical concept along with a minimal implementation and its key features are presented.

We see several directions in which this line of approach can be further developed. Including more local information such as the contextual features could improve the inference power of Hough space parametrization based partial image alignment. Furthermore, due to the deformable nature and huge shape variability of most objects of interest, an optimal alignment between the partial images of an object may not be in terms of just a reference

point. Incorporating more constraints such as the smoothness and continuity of local structures, which can be again inferred via a Hough space parametrization, will further increase the robustness of the presented approach.

Although we have chosen to learn the shape of an object of interest in this work, an alternative formulation of global consistency would be an implicit consistency measure that takes the available structures into account. For instance, the continuity of straight lines and curvilinear structures could be integrated into this framework to enable the measurement of structural consistency directly in terms of structure and without knowing the underlying object shape. Such a concept, however, should employ multiple Hough models to handle different models of structures. One can think of a linear Hough model that tries to align linear structures combined with a circular Hough model to ensure the global consistency of such structures beyond the overlap between the images. In the case of a linear model, the objective should be having a more clustered parameter space in the case of a good alignment compared to a more dispersed one when images are not aligned in a consistent manner in terms of linear structures. Similarly, having a structured vote distribution in the circular case would mean that circular structures spanning multiples partial images agree to form a consistent shape.

For the assessment of how structured a vote distribution in the parameter space is, the approach of calculating the entropy of vote distributions in the Hough space would be a suitable choice. We have discussed the limitations and strengths of the entropy-based measure in Section 5.2.2. Moreover, in addition to the parametric shapes such as lines and circles, any prior information about the shape of an object can be seamlessly integrated again using the Generalized Hough Transform as demonstrated here.

The classical GHT is invariant to rotations and isotropic scaling of the object and tolerates small deformations which was confirmed by our experiments using slightly deformed segmentation masks for learning. However, it still cannot handle large deformations. One of the reasons is the approximation of complex transformations with a few degrees of freedom. The GHT framework is not suitable for detecting object with large shape variations. This limitation, however, can be relaxed by employing large training datasets in the learning phase in a decision trees framework [Gode 13]. This stands as a potential future extension of the proposed method by posing it as a regression problem and by solving it in a random forest framework. When implemented in such a framework, this would allow employing higher order features such as interest points or image patches as voting elements [Gall 13]. The edge contour-based approach studied here can be extended by including deformable shape models as done in [Ferr 10]. Finally, using higher order transformations such as affine or deformable as well as mosaicing in 3D are the potential immediate extensions of the proposed concept.

5.5. Conclusion

In this chapter, we have demonstrated an intensity-based image registration approach where a novel Generalized Hough Transform based global regularization term called Global Consistency Measure (GCM) is used to enable globally consistent alignments of partially overlapping as well as non-overlapping images. The key idea is the parametrization of the shape of an object of interest in the Hough space and its usage for generating parameter

hypothesis in a joint parameter space about the underlying objects via local features in partial images.

The main strength of the proposed approach comes from its ability to infer global properties using local information. This allows us to use information contained in non-overlapping image regions to jointly estimate the global consistency of an alignment. Thanks to this property of being independent to the size of the overlap between images being registered, different from the conventional image registration techniques, it can be employed for aligning even non-overlapping parts of an object. If an overlap is available, this term can be further combined with a local similarity metric based on feature matching or intensity correspondences. Such a combination profits from a globally consistent alignment thanks to the GCM and locally consistent alignment due to the local similarity metric. The GCM can be efficiently used for the alignment of partially overlapping images when the local information in the overlapping regions is insufficient or corrupted.

In order to demonstrate the above mentioned features of the proposed global regularization term, we have conducted experiments on synthetically generated data as well as on real data. Through experiments, it was demonstrated that the proposed method is able to align partial images with large physical gaps between them. It was also shown that such an approach can handle corrupted as well as noisy overlaps where only using local similarity measure might lead to an incorrect alignment.

Thanks to the continuous development in Hough-based object recognition techniques in computer vision, the presented approach has a huge potential of further improvements and developments. Also, given the novelty and demonstrated strengths of the proposed method, there is a variety of potential applications both in general computer vision as well as in medical image processing waiting to be discovered. Last but not the least, while a basic implementation of our approach is shown in this work, the proposed concept is massively parallelizable and it is suited for reconstructing sparsely sampled scenes.

CHAPTER
SIX

CONCLUSIONS AND OUTLOOK

In our final chapter, we present our conclusions derived from the theoretical and experimental observations from this dissertation. Then we describe various ways to improve the strength of the methods presented in this thesis. Finally, we conclude this chapter with a discussion of several open issues.

In this dissertation, we have looked into different structural consistency measures to be employed in various intensity-based image registration applications. We have devised regularization terms within the existing image registration functionals to include additional constraints with the purpose of having structurally consistent registration solutions. We have presented two ways of establishing measures of local and global consistency.

6.1. Local Consistency

First, for the local structural consistency, we proposed to create structural probability maps by propagating salient structures into regions where the existing information is either missing or corrupted. Our main motivation was the availability of salient contextual information in the neighborhood of any image region that can be exploited while doing a registration. To this end, we have extended the popular tensor voting method to infer structures based on the neighborhoods. Such propagated structures serve to impose additional constraints when the existing information is not enough to find a good estimation of the correct transformation parameters. We have mainly considered structural consistency where our goal was to make sure that the structures spanning a region wider than the one used for registration are smooth and continuous after the registration.

We have demonstrated a few of several applications where such a consistency constraint is needed. Applications demonstrated here can be listed under the category of image field-of-view extension, in which the goal is to have a broader picture of the scene by putting together two or more partially observed images. We have used the term field-of-view extension to mean the widening of image field-of-view in any direction. Therefore, we have put the 2D image mosaicing where images are stitching together in the plane and the 3D volume reconstruction where 2D images are stacked along the z-axis in the same category.

First, we have attacked the problem of 2D image mosaicing. In traditional approaches,

only the overlap between sub-images is used for the estimation of transformation parameters. Here, we showed how the information contained in the non-overlapping image regions can be incorporated into the registration process for a more consistent combination of sub-images. Our experimental results showed that, besides improving the registration quality if an overlap exists, such a strategy even enables the registration of non-overlapping images. The latter result has opened a new perspective in the field of image mosaicing where the alignment of non-overlapping images had been so far avoided.

Secondly, we have demonstrated how the reconstruction of 3D digital histopathology images could benefit from the proposed structural probability maps. The main challenge in this application is the independent distortions in the 2D images as a result of the cutting process. Due to this reason, reconstruction becomes an ill-posed problem where additional constraints such a regularization or anatomical priors in terms of reference images are required. Our method has followed the recent trend of reference-free reconstruction approaches where only the implicit information is used to obtain an anatomically sound 3D reconstruction. Our experiments on synthetic and real data showed that the use of structural probability maps for local structural consistency leads to a reconstruction quality which is comparable to the approaches that use external references for anatomical consistency between neighboring histology slices.

Our final application on whole-body MRI reconstruction was chosen to demonstrate the capacity of our structure propagation method in handling geometric distortions that might be present in the image data. Without having any reference data with the ground truth image information, reconstruction of whole-body MRI solely based on the overlap between partial images often leads to an alignment that might be far from having consistency between the anatomical structures passing through the overlap. Experiments have indicated that extending structures from the less distorted image regions and incorporating them in the registration in the overlapping region leads to a structurally more consistent reconstruction.

The three applications of structure propagation presented here have shown that established registration methods benefit from including more contextual information from the surrounding of regions that are not salient enough for registration. While we have attempted to demonstrate the potential uses and variations of structure propagation, there is a variety of ways of how our approach can be further improved or extended. Firstly, the structure propagated here are only edge features extracted from images. We believe that inclusion of higher order features as well as direct intensity values would lead to a better inference quality. Moreover, structural information inferred so far was the salience and the orientation of structures. We have shortly discussed how higher order structural information such as curvature can be employed in the inference to enable a non-linear extension of structures. Finally, in this work, only one type of structural saliency was used as a scalar value. For instance, in 2D, we used curve saliency whereas in 3D surface saliency was used. It is, however, possible to extract all kinds of saliences such as point, curve or surface at a votee position. Therefore, if needed, one could use other types of structural saliences, too.

Since the focus in this dissertation is image registration, our examples were also chosen from this field. Other potential applications include the creation of wide field-of-view Optical Tomography Images (OCT), stitching of two-photon microscopy tiles among others. Furthermore, it should be noted that the application of the structure propagation method is not limited to only image registration. This idea can be used equally well for segmentation,

for instance. Especially, the segmentation method that are based on the boundary information could benefit from the structure propagation. Weak and missing boundaries can be augmented by the structural probability maps which can be plugged in again as regularization terms into the segmentation functionals. Furthermore, detection and segmentation of elongated structures in low quality and noisy medical images would be among the possible applications of curvature-based structure propagation.

6.2. Global Consistency

As a second measure of consistency, we have proposed a parametrization method for the alignment of partial observation of a previously seen object or scene. The goal was to utilize the parametrization of the object in a different space to ensure a globally consistent alignment. The consistency measure devised here acts as a regularization term in a standard image registration method where the usual intensity-based similarity term accounts for the local consistency.

The primary motivation in this part was similar to the one in the case of local consistency where we made the observation that not only the information in the overlap of partial observations should contribute to the registration but also the non-overlapping parts present important cues about the configuration of the overall object. To this end, we proposed to learn or parametrize the shape of an object in the Hough space using the Generalized Hough Transform. Then, we let each partial image make an inference about the global properties of the object of interest in the current scene. For the inference, we have used simple edge features that cast votes in the parametric space. Once every partial image has made an inference, we proposed to use the agreement between the inferences coming from partial observations as a measure of consistency in terms of the global configuration of the object parts.

Experiments on synthetic as well as on partial ultrasound images have shown that such a consistency measure renders the search space of a registration problem more convex, thus, leads to a better convergence. Experiments have also indicated that, as a regularization term, the global consistency measure devised here relaxes the dependency on the initial configuration of the parts of an object during intensity-based image registration. This is a very important property since most of the existing image registration methods tend to end up in a local optima unless a good initialization in terms of parameters is provided.

While we have assumed only linear transformation in our applications, with some modifications and methodological improvements, the proposed concept is suitable also for the registration scenarios where higher order transformations are sought for. To make this happen, first, the method of parametrization needs to be improved where the Generalize Hough Transform should be abandoned. Instead, a more sophisticated parametrization that is based on training on large data sets including deformations should be utilized. Recent advancements in learning based object detection and tracking methods in computer vision should be utilized while doing this. Decision forest based learning mechanisms, in particular Hough Forest, would be a good direction to look into.

Another theoretical extension we see is the exploration of different ways of handshaking mechanisms between the global consistency measure and the local similarity term. We have proposed to use an uncertainty measure to enable this handshake where the influence of

the global consistency measure becomes negligible when the value of (in)consistency falls in the range of the uncertainty parameter. From this point on, the local similarity measure becomes more effective in refining the solution. While we have made use of heuristics for determining this uncertainty parameter in our experiments, it would be beneficial to have an implicit strategy for handing over the responsibility of alignment.

The term consistency in this context was used for the agreement of hypotheses in the parameter space. The hypotheses generated were for the reference point of the shape of interest. An interesting theoretical extension, which we have also discussed in Chapter 5, would be the further investigation of measures in the parametric space that would reflect the internal consistency of existing structures in the partial images. This could be done either by carefully studying the statistics of natural shape of structures in the target object or by using heuristics about the possible configuration of structures. For instance, it might be desirable to have an orientational consensus between the linear structures spanning two partial images. This could impose additional constraint if the information in the overlap of partial images could not lead to a unique solution. Similarly, partial circular structures could be forced to compose a unique large circular structure during the alignment which would again bring a regularization to the registration. For this purpose, going into the directions of multi-modal Hough transforms as well as studying natural image statistics would be beneficial.

In terms of applications, we have demonstrated the use of the proposed concept for the mosaicing of partial ultrasound images of the human liver. However, it should be noted that the application of our method is not limited to only this kind of cases. Similar scenarios where partial observations of a previously seen object of interest need to be registered together could be potential applications of our method. For instance, industrial applications where a reconstruction of a scene from partial views is required could also benefit from the proposed approach.

Final Words: This dissertation has investigated various ways of including additional information into the intensity-based image registration with the purpose of having more regularized solutions. While we have presented two separate method for ensuring local and global structural consistency, respectively, we do not see any reason for not combining them in a unified framework for ensuring both types of consistencies during image registration. Finally, we hope that this work will give inspirations for new approaches for structural and contextual consistency in intensity-based image registration.

Appendix

APPENDIX

A

OTHER MAJOR CONTRIBUTIONS

A.1. Out-of-Plane Motion Compensation in Cine-MRI

Mehmet Yigitsoy [1], Maximilian Reiser [2], Nassir Navab[1] and Sonja Kirchhoff[2]
[3] [4]

Significant socio-economic burden of colonic motility disorders necessitates in-depth analysis of this pathology. Current analysis techniques are based on diameter measurements of colonic lumen on cine-MR images. Interleaved multi-plane acquisition makes it difficult to perform simultaneous measurements on the line of plane intersections due to the out-of-plane motion (OPM) caused by respiration affecting the underlying anatomy. Low temporal acquisition rate and dark-banding artifact are the challenging factors for OPM compensation. In this paper, we propose the use of manifold learning in combination with in-plane motion tracking for estimating OPM. We evaluate the effectiveness of our approach on 8 MR patient data sets. Experimental results show the good performance of our approach. The proposed method is independent of the acquisition rate and is not limited to this specific application.

A.1.1. Introduction

Functional gastrointestinal disorders, such as diarrhea and chronic constipation, are conditions presenting with a significant socioeconomic burden. Chronic constipation is one of the most common of these conditions being one of the leading diagnoses for gastrointestinal (GI) disorders in the United States [Tall 08]. One important type of constipation is the slow-transit constipation associated with slow colonic transit time which is usually attributed to colonic motility disorders [Bass 05]. Therefore, it is necessary to study colonic motility in order to understand its effects on colon pathologies mentioned above leading to an improved and more adequate therapy in the end.

Existing well-established examination techniques such as manometry or scintigraphy are either invasive and inconvenient or expose patients to ionizing radiation creating the

[1]Computer Aided Medical Procedures, Technische Universität München, Germany
[2]Institute of Clinical Radiology, Klinikum der Universität München, Germany
[3]This work was funded by DFG (German Research Foundation)
[4]The original publication [Yigi 12a] is available at www.springerlink.com

demand for a fast and non-invasive monitoring technique for the evaluation and quantification of colonic motility. On the other hand, functional cine magnetic resonance imaging (cine-MRI) allows for non-invasive, fast dynamic imaging with a superb soft tissue contrast [Lien 00].

The first use of cine-MRI for the analysis of colon motility was reported in [Buhm 05] and the authors considered luminal diameter changes as motility parameter and manual diameter measurements were performed in the ascending, transverse and descending parts of the colon. Sets of 2D dynamic image sequences from the same imaging plane over time were acquired. Due to the respiratory gating, the sampling in time was irregular making the continuous tracking of lumen diameters infeasible. In [Gloc 07], this approach was extended by addressing the irregular sampling in time and manual measurements on each 2D frame. A semi-automatic tool was developed to perform automatic diameter measurements on the data acquired during free breathing. However, measurements were still performed in 2D and the analysis of complex colon motion in 3D was not feasible by using this approach. Kutter et al. [Kutt 08] proposed a multi-plane image acquisition of the colon by concentrating only on the descending part. In this setting, the images of the descending colon were acquired in multiple quasi-orthogonal planes in an interleaved way in order to perform simultaneous measurements on sagittal and coronal planes at various points on the line of intersection (LOI) between the planes and, then, to combine these values to have an approximation of the colon motion in 3D. However, since the acquisition in two planes is not simultaneous but sequential, the colon undergoes an out-of-plane movement caused by respiratory motion. This necessitates an OPM compensation in order to perform simultaneous diameter measurements at anatomically corresponding locations on two intersecting planes.

One approach to deal with this issue, is to approximate the breathing curve by fitting a sinusoidal to in-plane displacements via a Fourier analysis, as proposed in [Kutt 08]. However, this idea is based on the assumption that the sampling rate per orientation plane is high enough according to the Nyquist-Shannon sampling theorem [Shan 49] for the recovery of the breathing curve with a frequency approximately 0.20 Hz. Apparently this is not feasible in our case where sampling rates per orientation are approximately 0.25 Hz. Another approach is the usage of the shared information, i.e. the intensity profiles (IP) along the lines of intersection (LOI), to have a rigid alignment as in [Kim 10]. However, due to dark banding artifact in the images caused by fast MRI acquisition [Dera 10], the IPs from two images present large variations resulting in uncorrelated intensity patterns along the LOI.

Therefore, in this work, we propose the use of manifold learning, a method that has recently been successfully applied to the image-based recovery of breathing motion [Geor 08], for the recovery and the compensation of the OPM in sagittal and coronal cine-MRI sequences. To this end, we perform dimensionality reductions on both sequences independently and use the average of the parameterizations of these embeddings to approximate the breathing curve in order to establish a basis for relating the sequences. In combination with in-plane motion tracking, significant compensation of the OPM can be achieved. The performance of the method is not affected by the data sampling rate as long as enough samples per orientation are collected.

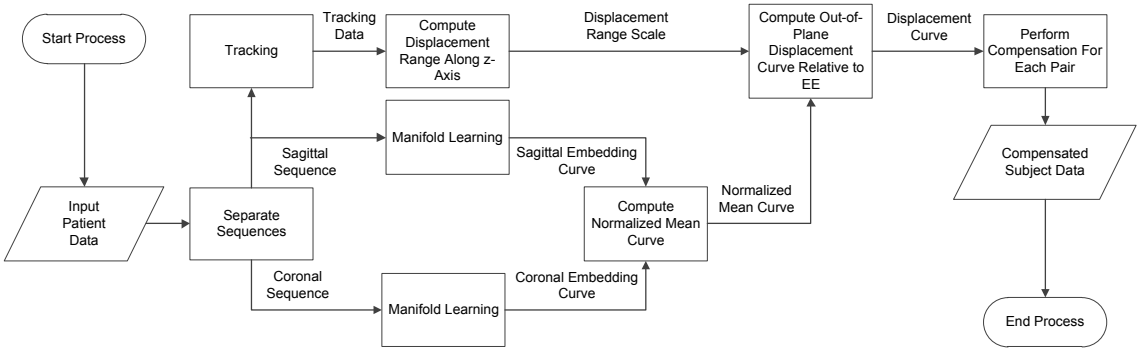


FIGURE A.1.: Flow diagram of the OPM compensation algorithm.

A.1.2. Materials and Methods

In this study 8 subjects were examined with functional cine-MRI [Kirc 11]. The acquisition was performed using a half Fourier acquisition single shot turbo spin echo sequence (HASTE; TR: 476 ms, TE: 54 ms, SL: 5 mm, FOV: 320 x 400 mm, i-Pat-factor: 3.0, voxel size: 1.25 x 1.25 x 5.0 mm, duration: 8 min). Three orientations oblique to each other were set, examining 1 slice per orientation (transversal, sagittal, coronal) with a time interval of 4 s between image-sets to be able to physiologically image and track bowel motility. An alternating acquisition scheme, $\{\ldots, \mathbf{u}_i^t, \mathbf{u}_{i+1}^s, \mathbf{u}_{i+2}^c, \mathbf{u}_{i+3}^t, \mathbf{u}_{i+4}^s, \mathbf{u}_{i+5}^c, \ldots\}$, was used in order to equally sample data from each orientation. The transversal sequence is not used in this work as it is not used for colon motility analysis. For more details regarding the image acquisition protocol, please refer to [Kirc 11].

Manifold Learning

The general idea of manifold learning is to project a manifold in high dimensional space $\mathbb{R}^N$ to a low dimensional space $\mathbb{R}^n$, while preserving the local neighborhood. In our case, we consider one dimension of the ambient space for each image pixel, so N is corresponding to the resolution of the MRI images. For the low dimensional space, we set $n = 3$, in order to keep as much relevant information as possible while performing a significant dimensionality reduction. Considering k MR images $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ that are acquired over several breathing cycles in one orientation, the manifold learning $\mathcal{M}$ assigns each image to a coordinate in the low dimensional space ϕ_i

$$\mathcal{M} : \mathbb{R}^N \rightarrow \mathbb{R}^n \tag{A.1}$$

$$\mathbf{u}_i \mapsto \phi_i, \tag{A.2}$$

with $1 \leq i \leq k$. The suggestion that images lie on a low dimensional manifold in the ambient space seems to be justified because variations between neighboring slices are smooth, and furthermore, slices from the same breathing phase but different acquisition times share similar information. Moreover, since manifold learning techniques try to optimally preserve local information [Belk 03], meaning that similar images are mapped to similar positions in the low dimensional space, it is reasonable to use ϕ_i as an estimate for the respiratory phase.

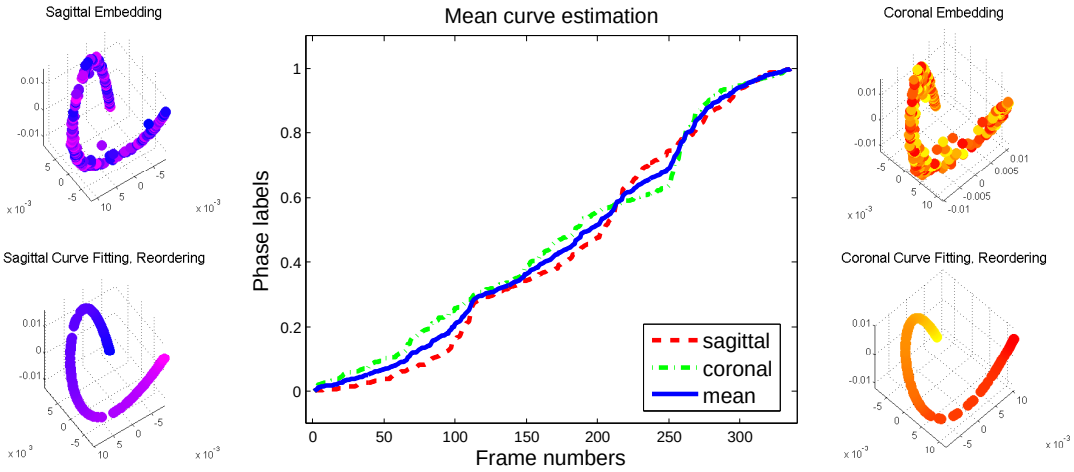


FIGURE A.2.: Estimation of the mean motion curve via manifold learning. Curve fitting and parametrization are applied to each of the embeddings. In the figures on left and right, the color encodes the index before (upper) and after (lower) reordering.

We propose the application of Laplacian eigenmaps [Belk 03] for the respiratory phase estimation because the technique is well founded on mathematical concepts (Laplace Beltrami operator) and computationally efficient. Laplacian eigenmaps build upon the construction of a graph, which represents the neighborhood information of the data set. Subsequently, the graph Laplacian is applied to calculate a low-dimensional representation of the data that preserves the local neighborhood information in an optimal way.

We construct a graph with a node for each point $\mathbf{u}_i$ and with edges connecting neighboring nodes. In order to deal with contrast differences between frames, we use cross correlation (NCC) as our similarity measure which is essential for neighborhood selection and weighting. We select for each image $\mathbf{u}_i$ the l nearest neighbors, by evaluating the term $NCC(\mathbf{u}_i, \mathbf{u}_j)$. Further, heat kernel-based weights are assigned to the edges with

$$w_{ij} = e^{-(1-|NCC(\mathbf{u}_i,\mathbf{u}_j)|)^2/(2\cdot\sigma^2)} \tag{A.3}$$

and σ^2 the variance [Belk 03]. Once the neighborhood graph is constructed, the eigenvectors of the graph Laplacian provide the embedding map. After performing independent dimensionality reductions on coronal and sagittal sequences, these two embeddings need to be related in order to do further processing. The fact that both sequences are affected by the same breathing motion enables us to estimate the mean motion curve by using low dimensional embeddings of two sequences. Since the embeddings approximate curves in 3D, see Figure A.2, we first fit a 3rd order polynomial curve onto which we project original embedding points. Then, each curve is parametrized by point distances from one side of the curves. The direction of parametrization is chosen so that both of them are parametrized from end-exhale (EE) to end-inhale (EI). After normalizing the parametrized curves, the mean breathing curve to be used as a basis for compensation is computed by taking the average of the normalized curves.

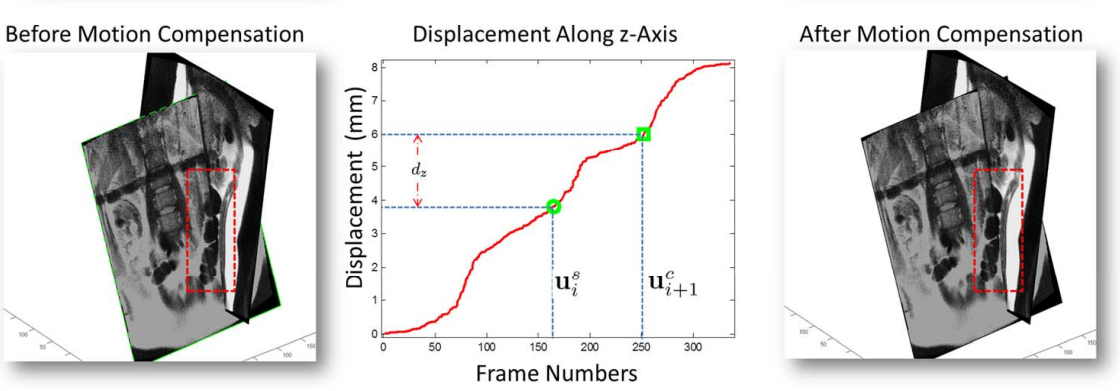


FIGURE A.3.: Illustration of the motion compensation process. The region of interest around the LOI is highlighted where the difference after the compensation can be clearly observed by looking at the lumen borders. Contrast enhanced for better visibility.

Out-of-plane Motion Compensation

Since it is not possible to directly infer metric displacement values from the mean breathing curve, we need to compute the range of displacements in mm's, d_z, along the direction of dominant breathing motion, i.e. z-axis [Rohl 04] in the reference coordinate frame. Then, we use this factor to scale the mean breathing curve in order to get the metric displacement values along the z-axis relative to the EE phase that is assumed to have no displacement.

In order to find d_z, we first estimate the range of in-plane displacements, d_p, in the sagittal orientation plane and then back project it onto the z-axis in order to approximate d_z. To this end, we perform tracking using block matching on the sagittal sequence to find a sequence of in-plane displacements, $\mathcal{D} = \{d_{p1}, \ldots, d_{pk}\}$, relative to the first frame. A region that is highly influenced by the respiratory motion and close to the diaphragm is chosen for tracking. This is the visible part of the liver in our case. In-plane displacement range, d_p, is computed as $d_p = max(\mathcal{D}) - min(\mathcal{D})$. This value is projected on z-axis as

$$d_z = \begin{cases} \dfrac{d_p}{sin(\alpha)} & \text{if } \alpha > 0 \quad \text{(A.4a)} \\ 0 & \text{otherwise} \quad \text{(A.4b)} \end{cases}$$

where α is the angle between the sagittal plane normal and the z-axis calculated from the DICOM orientation data. The sagittal plane is chosen for estimating d_z since it is more aligned with the z-axis than the coronal plane. Once d_z is obtained, the mean breathing curve is updated by scaling it with d_z. Then, for each pair $\{\mathbf{u}_i^s, \mathbf{u}_{i+1}^c\}$, we compute the difference between their z-displacements relative to the EE phase, see Figure A.3. We shift the coronal plane by applying a translational transform along the z-axis to bring it to the breathing phase of the sagittal plane leading to real anatomical correspondence along the LOIs.

A.1.3. Experiments and Results

We have conducted experiments on 8 different patient data to evaluate the performance of the proposed method. For each data set, we first separated coronal and sagittal 2D se-

quences each consisting of 336 frames with a size of 256x320 pixels. Then the proposed method was applied to compensate for the OPM. Sagittal sequences of each subject underwent a tracking in order to estimate the in-plane-displacement range. In each case, Laplacian eigenmaps was used as the manifold learning technique. We set neighborhood size to 70 which empirically seemed to be sufficient for our application.

As discussed in Section A.1.1, direct use of IPs for alignment is not possible due to large variations along the LOIs. Due to the same reason, we can not use IPs directly for validation purposes. Instead, we used a landmark that is affected by the breathing motion and that can easily be located in both of the IPs before and after the application of the motion compensation technique. A good landmark candidate is the edge of the liver which lies on the LOI and has a higher contrast that can be distinguished from the rest. In the case of a simultaneous acquisition, the positions of this landmark would match on both planes. Therefore, we used the difference between the positions of this landmark to asses the performance of compensation. A reduction of the difference indicates a good performance of the compensation technique. The location of the landmark in IPs was detected by searching for the maximum intensity around a predefined point. This point is associated with a high intensity standard deviation (STD) in time, see Figure A.4(a). There are also other points having high STD values, but, this one can be simply identified by picking the first one from the top.

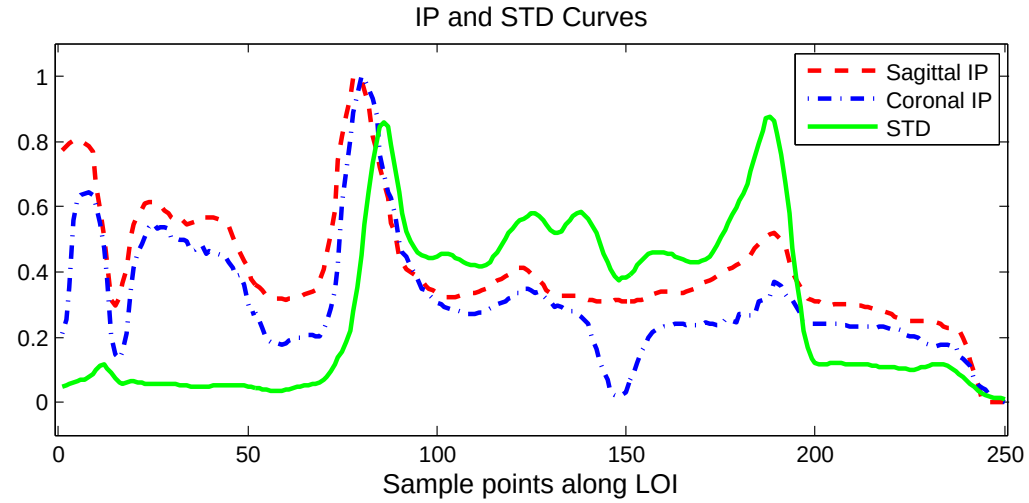
For each subject, we performed difference measurements on 336 different pairs. We plot the statistics of the landmark error for each patient in Figure A.4(b). In each case, compensation resulted in a reduction of landmark error showing the effectiveness of the proposed approach.

A.1.4. Discussion and Conclusion

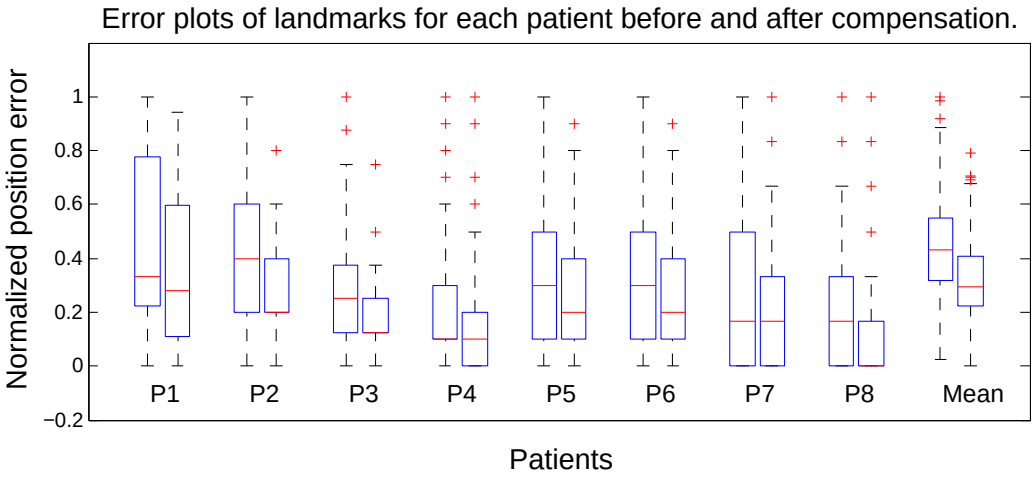
For an accurate analysis of the colon motility, motion due to respiration must be suppressed both in- and out-of-plane. It is of great importance to compensate for the OPM since the subsequent diameter measurements are directly based on LOIs. Therefore, in this work, we proposed a novel OPM compensation technique for dynamic, multi-plane, cine-MRI sequences of the colon. Experimental results show the good performance of our method and high potential for being used in similar scenarios with multi-plane acquisitions and OPM artifacts.

The novelty of our method is based on the use of manifold learning for the estimation of out-of-plane breathing motion on multiple planes. This gives a ground for relating sagittal and coronal planes in terms of breathing motion. Since low dimensional embedding gives a relative distribution of high dimensional data in the low dimensional space, it is not possible to deduce metric displacements using manifold learning. Therefore, in-plane motion tracking on sagittal sequences is included to assign metric values to the low dimensional embeddings relative to EE phase. Correlations between tracking and manifold learning for sagittal sequences were above 0.95 in average. This also reveals the effectiveness of manifold learning in detecting motion patterns.

Due to the nature of the problem, a simple validation approach based on IPs was not feasible. Dark banding artifact along the LOIs did not allow us to use IPs directly for the evaluation. Instead, we detected landmarks along LOIs that are highly influenced by the breathing motion. By measuring the difference between positions of landmarks on two



(a)



(b)

FIGURE A.4.: (a) Mean IPs for each orientation plane and the intensity STD curve along the LOI. The landmark is searched for in the vicinity of the first peak from the right side in the STD curve. (b) Plot of normalized landmark errors for each patient. Each patient data has 336 pairs of coronal and sagittal to be aligned. Therefore, in this figure, each box pair represents error statistics for 336 image pairs before (left) and after (right) compensation.

planes before and after motion compensation, we were able to evaluate the performance of the proposed method. Experimental results revealed the possibility of recovering the out-of-plane breathing motion using a combination of in-plane motion tracking and manifold learning on image sequences.

A.2. Dynamic Graph Cuts for Colon Segmentation in Functional Cine-MRI

Mehmet Yigitsoy[1], Maximilian Reiser[2], Sonja Kirchhoff[2] and Nassir Navab[1]
[5] [6]

A major limitation of graph cuts for the segmentation of large 2D image sequences is its interactive nature. The user has to provide seeds for almost every image frame to get an accurate segmentation. Straightforward approaches like direct copying of seeds provided in the first frame to other frames of the sequence fail in cases of great contrast or topological changes that occur when there is a large temporal distance between frames. In this work, we propose a dynamic seed propagation technique which can automatically adjust to any contrast or topological changes in a sequence. To this end, distance transform and skeleton extraction methods are employed to initialize the segmentation of the current frame using the result of the previous one. The proposed methods were used for the segmentation of functional cine-MRI sequences of colon which is especially challenging due to the large temporal distance between its frames. Both quantitative and qualitative results show good performance of the proposed method.

A.2.1. Introduction

Functional gastrointestinal disorders, to name the most common diarrhea, chronic constipation, irritable bowel syndrome and pseudo-obstruction, are conditions presenting with a significant socioeconomic burden. Chronic constipation is one of the most common of these conditions which is one of the leading diagnoses for gastrointestinal (GI) disorders in the United States [Tall 08]. One important type of constipation is the slow-transit constipation associated with slow colonic transit time which is usually attributed to colonic motility disorders [Bass 05]. Therefore, it is necessary to study colonic motility in order to understand its effects on such conditions leading to an improved and more adequate therapy in the end.

Existing well-established examination techniques such as manometry or scintigraphy are either invasive and inconvenient or expose patients to ionizing radiation creating the demand for a fast and non-invasive monitoring technique for the evaluation and quantification of colonic motility. On the other hand, functional cine magnetic resonance imaging (cine-MRI) allows for non-invasive, fast dynamic imaging with a superb soft tissue contrast [Lien 00].

The first use of cine-MRI for the analysis of colon motility was reported in [Buhm 05] where luminal diameter changes were considered as a motility parameter and manual diameter measurements were performed in the ascending, transverse and descending parts of the colon. Sets of 2D dynamic image sequences from the same imaging plane over time were acquired using a T2-weighted single shot HASTE sequence. Due to the respiratory gating the sampling in time was irregular, therefore, a continuous tracking of lumen diameters was not feasible. Glocker et al. [Gloc 07] extended this approach by addressing the irregular sampling in time and manual measurements on each 2D frame. A semi-automatic

[5] This work was funded by DFG (German Research Foundation)

tool was developed to perform automatic diameter measurements on the data acquired during free breathing. Prior to diameter measurements, the colon was segmented in 2D dynamic sequences using the graph cuts [Boyk 01] approach where the user provides seeds for object and background in the first frame. Due to the inconvenience of placing seeds in each frame separately, the seeds on the first frame were directly propagated to the other frames of the sequence and a subsequent segmentation was performed by considering a 2D sequence as a volumetric image.

Although graph cuts provides globally optimal solutions, one needs to fine tune the region and boundary parameters for every segmentation. Despite the global motion compensation, the colonic contractions between successive frames might be large enough to leave the seeds from the first frame invalid leading to an inconsistent segmentation in the end. A possible solution to this is to update the seed points for every frame to be segmented by using the segmentation of the previous frame reducing the human interaction, thus bias. Such a dynamic seed placement procedure will provide more flexibility for the segmentation method to allow for large morphological differences of the colon over time.

Therefore, in this work, we propose the use of a dynamic seed placement procedure for the segmentation of 2D dynamic MR-image sequences where topological changes do not allow for direct use of the same seed brushes for subsequent frames.

A.2.2. Materials and Methods

Functional Cine-MRI Acquisition

Functional cine-MRI was acquired using a half Fourier acquisition single shot turbo spin echo sequence (HASTE; TR: 476 ms, TE: 54 ms, SL: 5 mm, FOV: 320 x 400 mm, i-Pat-factor: 3.0, voxel size: 1.25 x 1.25 x 5.0 mm, duration: 8 min). Three orientations oblique to each other were set, examining 1 slice per orientation (transversal, sagittal, coronal) with a time interval of 4 s between image-sets to be able to physiologically image bowel motility. For more details about the image acquisition protocol, we refer the reader to [Kirc 11].

Dynamic Graph Cuts

Graph Cuts The interactive graph cuts approach was proposed by Boykov and Jolly [Boyk 01] and has since been used extensively in a multitude of applications. Let P be the set of pixels p in the image $I : \Omega \rightarrow \mathbb{R}, \Omega \subset \mathbb{Z}^2$, $\mathcal{N}$ set of neighboring pixel pairs (p, q) in P, $L = (L_1, \ldots, L_p, \ldots, L_{|P|})$ a label configuration for a binary segmentation where $L_p \in \{obj, bkg\}$, $G = (V, E)$ a graph with nodes, V, and edges, E, $\mathcal{S}_{obj}$ and $\mathcal{S}_{bkg}$ sets of pixels defined by the user for object and background, respectively. The label configuration, i.e. segmentation, can be formulated as an energy minimization problem

$$E(L) = R(L) + \lambda \cdot B(L) \tag{A.5}$$

which when minimized assigns image pixels into two subsets $\mathcal{O}$ (object) and $\mathcal{B}$ (background) where

$$L_{p \in \Omega} = \begin{cases} \text{“}obj\text{”} & \text{if } p \in \mathcal{O} \\ \text{“}bkg\text{”} & \text{if } p \in \mathcal{B}. \end{cases} \tag{A.6}$$

Here, the regional term R penalizes the labeling of the pixels based on the intensity models of object and background. The function R is defined as

$$R(L) = \sum_{p \in P} R_p(L_p) \tag{A.7}$$

where

$$R_p(L_p) = -\log \Pr(I_p \mid L_p) \tag{A.8}$$

is the negative log-likelihood of object or background intensity distributions obtained from the user seeds. For the pixels the label of which are provided by the user, the regional term is defined as

$$R_p(L_p) = \begin{cases} \infty & \text{if } L_p = \text{“}obj'' \wedge p \in \mathcal{S}_{bkg} \\ \infty & \text{if } L_p = \text{“}bkg'' \wedge p \in \mathcal{S}_{obj} \\ 0 & \text{otherwise.} \end{cases} \tag{A.9}$$

The second part B of the segmentation energy is called the boundary term representing the energy for pairs of neighboring pixels $p, q \in \mathcal{N}$ to have to the same segmentation label. It is defined as

$$B(L) = \sum_{p,q \in \mathcal{N}} B_{p,q} \cdot \delta(L_p, L_q) \tag{A.10}$$

where

$$\delta(L_p, L_q) = \begin{cases} 1 & \text{if } L_p \neq L_q \\ 0 & \text{otherwise.} \end{cases} \tag{A.11}$$

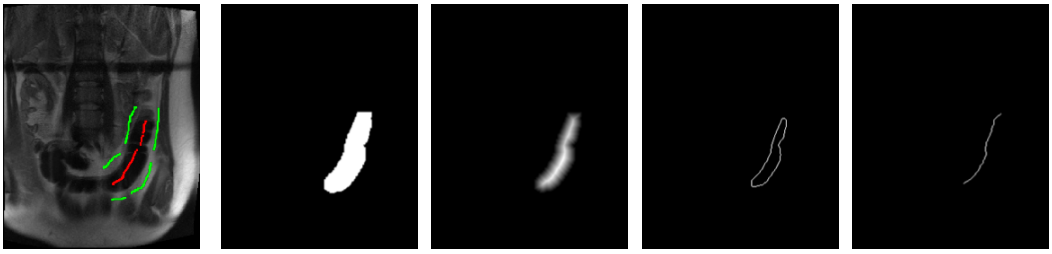
$B_{p,q}$ can be defined as a penalty function

$$B_{p,q} \propto exp(-\frac{(I_p - I_q)^2}{2\sigma^2}) \cdot \frac{1}{dist(p,q)} \tag{A.12}$$

which penalizes the dissimilar neighbors to have the same labels. The minimum of this functional can be computed by finding the minimum cut on the corresponding graph using efficient min-cut/max-flow algorithms [Boyk 04].

Dynamic Seed Placement The traditional graph cut segmentation algorithm proposed by Boykov and Jolly [Boyk 01] is applicable to N-D data, including 3D medical image data and temporal volumes constructed from video frames. However, when the size of the data to be segmented gets larger, the segmentation becomes a difficult problem to solve and the performance turns out to be an issue. Furthermore, the large size of the data leads to an inconvenient interaction for the user. For the segmentation of volumetric images, one could initialize the segmentation on one of the slices and use the region properties, such as the intensity distribution, extracted from that slice for the entire volume since they are similar for the object in every slice of the volume. For live video sequences, the flow recycling approach [Kohl 07] provides good initialization on a frame using the resulting flow from the previous cut. However, it is based on the assumption that changes between consecutive frames are *small*.

The above mentioned methods cannot be directly applied to our data sets where there are significant contrast changes between consecutive frames. Several factors, such as the



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

FIGURE A.5.: (a) Seed brushes in one frame of the dynamic image sequence. (b) Resulting segmentation of the descending colon. (c) Distance map of the segmentation. (d) Extracted level curve. (e) Extracted centerline.

dynamic nature of the abdominal region causing out-of-plane motion, fast imaging requirements for capturing colon motility which leads to low resolution and the low quality of images, require special treatment for cine-MRI sequences.

Due to the changes in regional properties for the object and the background between successive frames and topological changes throughout an image sequence, the hard constraints provided by the user have to be updated for every frame. It is inconvenient for the user to place seeds in every frame in a large sequence. Instead we propose an automatic initialization procedure by using centerline extraction and distance transform. For every frame to be segmented, we initialize the object seeds using the samples taken from the centerline and/or the level curves of the distance map. Given a binary image corresponding to the segmentation of the previous frame, we obtain the distance map using a distance transform $D(p) : \Omega \rightarrow \mathbb{R}$ defined as

$$D(p) := \begin{cases} min\{d(p,q)|q \in \mathcal{B}\} & \text{if } p \in \mathcal{O} \\ 0 & \text{otherwise} \end{cases} \tag{A.13}$$

where

$$d(p,q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \tag{A.14}$$

is the Euclidean distance. A level curve is extracted from the distance map by thresholding at a user defined level c. The centerline of the binary image is extracted using a thinning algorithm[Pala 01]. The level curve and the centerline are sampled equidistantly and these samples are used as object seeds for the next frame. We place the background seeds around the object making sure that they are placed in the surrounding tissue and that they never cross the object boundary throughout the whole sequence.

A.2.3. Experiments and Results

Experiments have been conducted to demonstrate the effectiveness of the proposed method. A dynamic 2D coronal Cine-MRI sequence was used for the study. The sequence consists of 336 frames of size 256x320 pixels and presents significant contrast and topology changes. Four different initialization techniques were used for the segmentation of this sequence; a standard copy/paste (CP) technique where the seeds for the first frame are directly copied

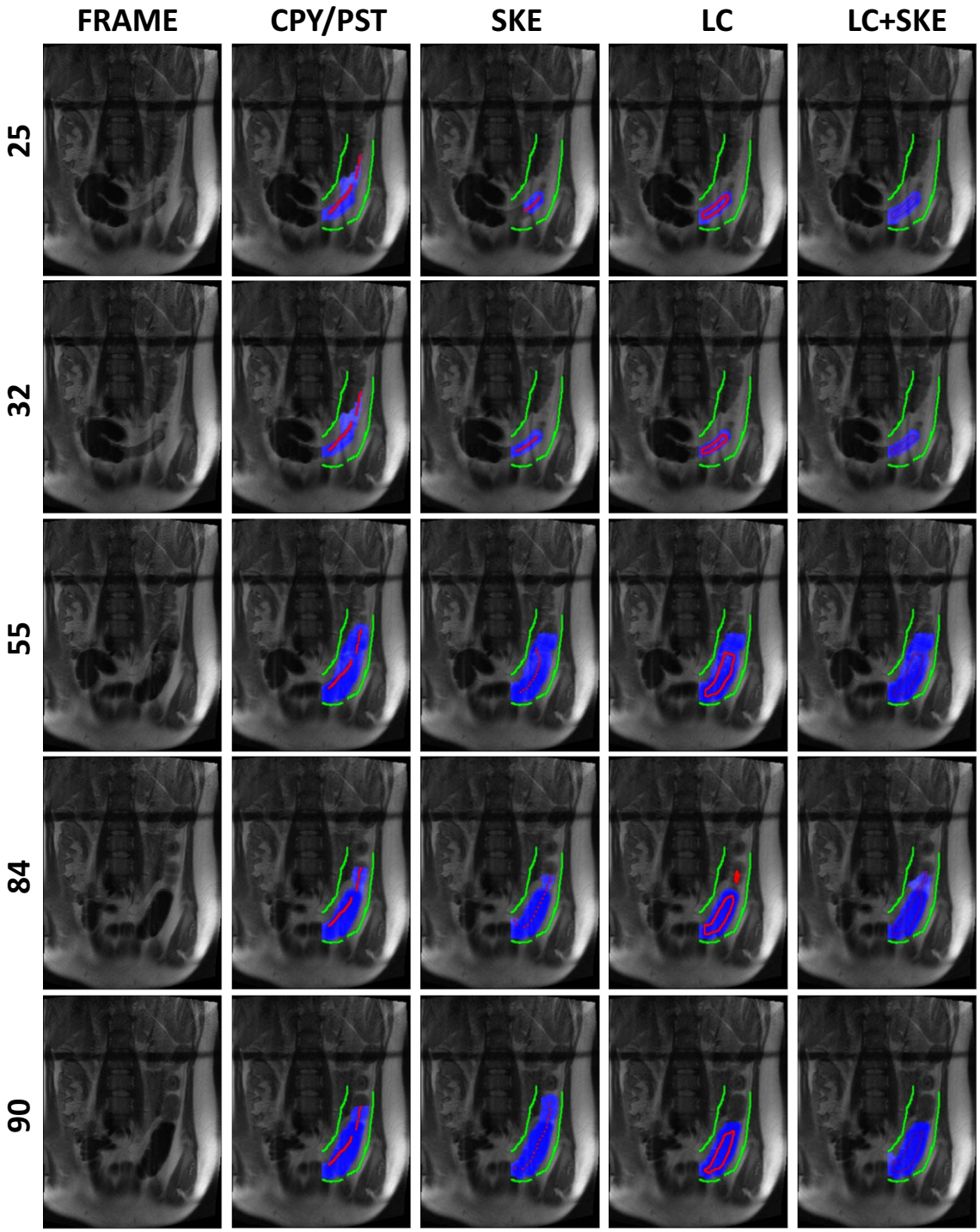


FIGURE A.6.: Sample frames from the dynamic image sequence together with their segmentation results obtained by using different seed placement approaches. The flexible nature of the dynamic approaches is quite obvious especially for frame 25 where the colon shape is very different from the initial frame shown in Fig. A.5(a). Note: Please refer to the electronic version of this paper to make full use of colors in this figure.

Statistics	CP	SKE	LC	SKE+LC
Dice Mean	0.946	0.966	0.966	0.964
Dice STD	0.088	0.027	0.029	0.019

TABLE A.1.: Mean and standard deviation of Dice scores for different seed placement approaches.

to every other frame, proposed dynamic seed placement techniques using skeleton (SKE) and level curve (LC), finally, a combination of skeleton and level curve extraction (SKE+LC). We set $c = max(D)/2$ for the level curve extraction.

User seeds are provided for the first frame as shown in Fig. A.5(a). The background seeds are copied always from the first frame. Object seeds are updated depending on the chosen seed placement technique. Sample frames from the sequence are shown in Fig. A.6 where the unsegmented frames are shown in the first column and the segmentation results overlayed on the original frames are shown in columns 2 to 5. Note that the region of interest (ROI) is defined by the user seeds. Therefore, only the descending part of the colon is segmented by placing the seeds only around this part of the colon.

The results in Fig. A.6 qualitatively demonstrate the poor performance of the copy/paste procedure based on the assumption that minor changes in terms of contrast and topology occur in the sequence. On the other hand, the results also reveal the good performance of dynamic seed placement techniques based on the previous segmentation and their flexibility during colon contractions leading to large topological changes. Dice coefficient was used for the quantitative evaluation of the results. Segmentation of each frame was compared to a ground truth where object seeds were placed manually on every frame. The Dice scores of each scenario for every frame in the sequence are plotted in Fig. A.7. Dynamic seed placement based segmentations represent a higher consistency over time. The mean and the standard deviations (STD) of the Dice scores for each case are shown in Table A.1. Dynamic techniques present not only with higher mean Dice scores but also with lower standard deviations throughout the sequence.

A.2.4. Discussion and Conclusion

In this paper, a dynamic seed placement technique for the semi-automatic segmentation of dynamic 2D Cine-MRI image sequences using graph cuts has been proposed. The qualitative and quantitative evaluation of the experimental results suggest that using dynamic seeds for the segmentation of abdominal Cine-MRI increases the robustness of the segmentation against intrinsic topological and contrast changes throughout the sequence.

Currently, only the object seeds are being updated. The background seeds have to placed on the first frame so that they never fall in the object region throughout the sequence. This requires a careful selection of neighboring tissues that remain stable over time. Although it is easy to find such regions in the ROI, a semi-automatic selection of background seeds would make intuitively more sense within the context of this paper. A straightforward approach would be the use of outer level curve from the signed distance function. However, this is not an easy task as in the case of inner level curve where the seeds always fall in the object region. During the distension of the colon, the extracted seeds might fall into the

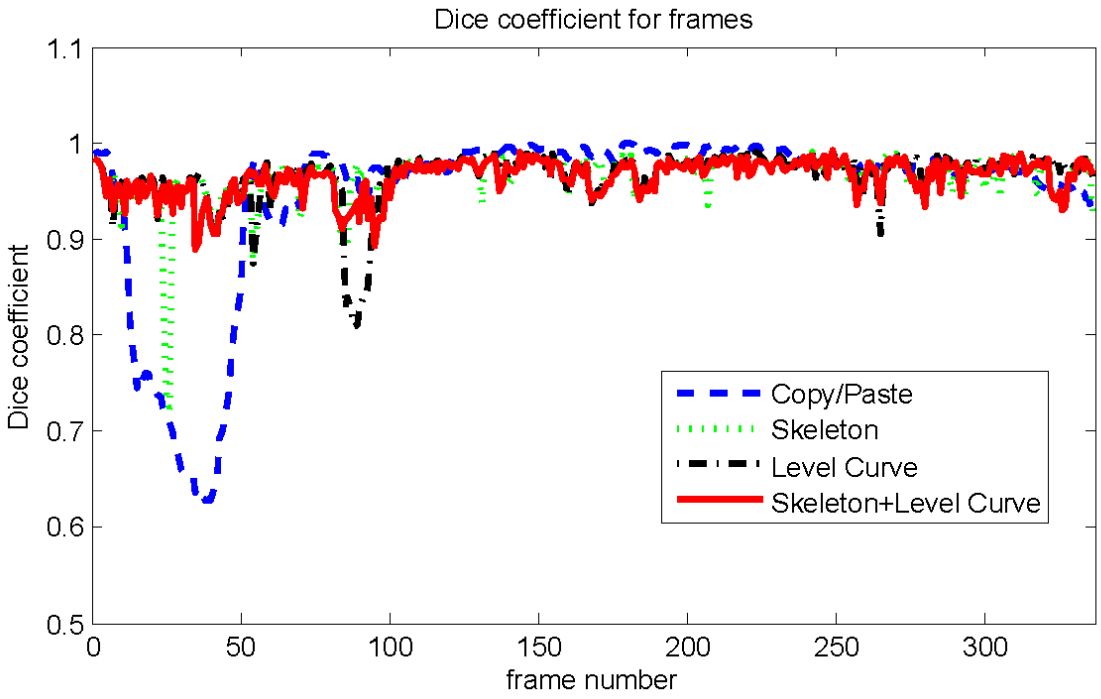


FIGURE A.7.: Dice scores for every frame plotted against frame number using different seed placement approaches.

colon region in the current frame resulting in an erroneous segmentation. Extension of the dynamic placement approach to the background seeds is the subject of future research.

A.3. Random Ferns for Multiple Target Tracking in Microscopic Retina Image Sequences

M. Yigitsoy[1], V. Belagiannis[1], A. Djurka[7], A. Katouzian[1], S. Ilic[1], F. Pernuš[7], A. Eslami[1], N. Navab[1]

[8]

Accurate and robust tracking of retina in operating microscope images is critical for an augmented reality assistance system for retinal surgery. Most retinal surgeries such as the peeling are performed using hand-held intraocular light and hence the tool and its shadow have two different motions, independent from the motion of the retina. In this paper, we propose multi-object motion estimation in high definition operating microscopic images by using a parallel network of random ferns, followed by RANSAC in order to achieve a simultaneous and robust tracking of the retina, the tool, and the tool shadow. Thanks to the separate tracking of each object, the number of outliers is dramatically reduced and the extracted motions are more accurate and reliable even in complex scenes which are considerably occluded by the tool and its shadow. The proposed method is evaluated on several challenging sequences in comparison with SIFT tracking, direct visual tracking, and single random ferns tracking of the retina. The experimental results show that the proposed method has a significantly higher success rate in comparison to the other three approaches with the accuracy of 4 pixels in tractable frames which is comparable with the intra- and inter-observer error of manual tracking (3.4 and 8.5 pixels, respectively).

A.3.1. Introduction

Retinal surgeries are among the most difficult operations due to the limited field of view, the reduced depth perception and uncertainty in localizing the tool by the surgeons. Several methods have been proposed in the past for assisting the surgeon either by means of robots or by fusion or overlay of pre- and intra-operative images [Pitc 12]. In general, fast and accurate tracking of the retina and the tool is the key component of a successful solution for computer assisted eye surgery. However, the inherent characteristics of the interventional imaging of retinal surgeries such as uneven illumination of the retinal surface, occlusion by tool or its shadow, variable zooming scale and variable appearances pose challenges for the intraoperative tracking in retina. For these reasons the conventional computer vision techniques are not completely successful and a specially designed method is required for tracking the retina in operating microscopy images.

Different approaches have been proposed for interventional retina and tool tracking including feature- or intensity-based or hybrid methods. A gradient-based tracker is employed in combination with a classifier-based detector for tool tracking in [Szni 12]. In their recent work, Richa et al. [Rich 12] proposed a hybrid approach based on direct visual and feature-based tracking using SURF features for retina tracking. The feature-based tracking in [Rich 12] is employed as backup to recover the transformation when the direct visual tracking may not reliably perform and its confidence goes below a threshold. Although, direct visual tracking performs reasonably good in case of lower resolution videos such, it

[7]Fac. of Electr. Eng., Univ. of Ljubljana, Ljubljana, Slovenia

[8]This work has been submitted to ISBI 2015 and it is currently under review.

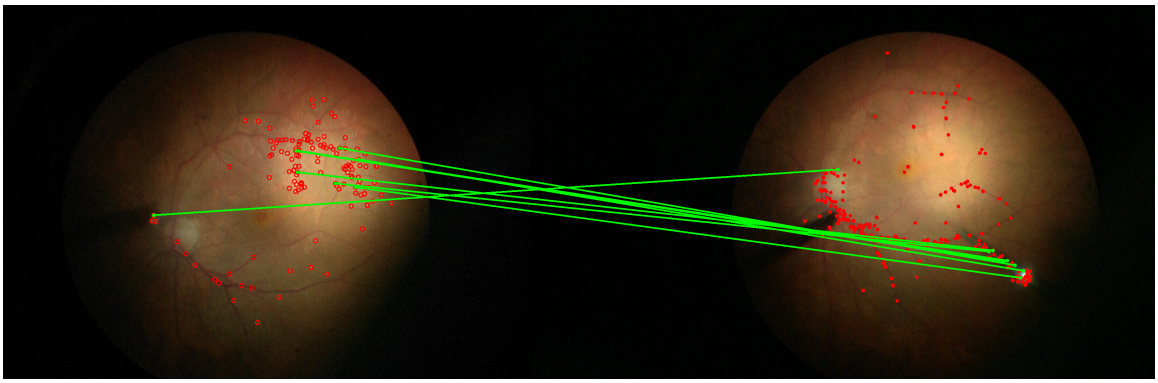


FIGURE A.8.: An example of matches between the reference frame on the left, and live frame on the right. The false matches with the tool in the live frame may cause an error in the estimated homography.

is not computationally efficient for the new high resolution microscopes.

On the other hand, feature based methods rely on the performance of their feature descriptors as well as their detectors. New feature descriptors have been proposed which were specially customized for retinal images [Stew 03, Broe 11]. Nevertheless, they are computationally expensive which makes them more suitable for registration but not as much for interventional retina tracking.

In this paper, we propose a purely feature-based multi-object tracking approach using multiple random ferns classifiers [Ozuy 10] along with RANSAC [Fisc 81] in order to efficiently and accurately track the retina without excessive computation. We introduce the novel concept of using multiple random ferns classifiers which are trained on a reference frame using robust keypoints and used simultaneously on live frames to classify detected keypoints into retina, surgical tool and its shadow even in the occluded scenes. During training, we perform a keypoint sparsification step that helps prevent sample selection bias which can happen when the keypoints (i.e. the training data) are not uniformly sampled. During testing, we perform a pre-grouping of detected live keypoints based on the posteriors returned by classifiers. Thanks to the exclusive parallel tracking, the proposed solution is more robust to the occlusions created by the tool and its shadow in the scene. As a result, the estimated homography for retina is more accurate and reliable even with less retinal features. Three different random ferns classifiers are employed for tracking by detecting the corresponding features exclusively on the tool, retina and the tool shadow.

Fig.A.8 illustrates more the rationale behind the proposed multi-target feature-based tracking. The green lines show the inliers detected by RANSAC after extracting and matching live and reference keypoints using a single random ferns classifier. The remaining keypoints are those which correspond either to matches which were detected as outliers after RANSAC, or were not matched at all. It should be noted that the RANSAC could not discriminate between different objects but, by randomly selecting a subset of putative pairs, tries to find a transformation leading to the most inliers. Hence, the performance crucially depends on the scene. If the scene is occluded by the tool and its shadow, then it is most likely that many of the inliers be the tool or shadow features, as observed in Fig.A.8, and

the final homography be wrong. Besides that, large number of false matches requires more RANSAC iterations leading to an increased runtime. Therefore, as we demonstrate experimentally in this paper, separate handling of each object by employing multiple classifiers simultaneously will increase the performance of homography estimation by decreasing the number of false matches for each object in the scene. Moreover, the pre-grouping of detected keypoints will enhance the tracking further by limiting the number of correspondences passed to the estimator.

A.3.2. Method

We pose retina tracking as a classification problem and employ random ferns classifier [Ozuy 10] to estimate the motion of the retina from the reference frame to each subsequent frame in a sequence. After training on the reference frame, detected keypoints from subsequent frames are matched via random ferns and a motion per frame is robustly estimated through RANSAC. However, occlusions due to multiple objects in retina images leads to outliers which has a negative influence on the estimation performance. Therefore, we track each object independently via multiple ferns classifiers, each separately trained on the reference frame for each object along with individual robust estimations. In the following, we provide a brief overview of the random ferns and discuss the tracking of multiple targets via multiple random ferns classifiers simultaneously.

Random Ferns

Ferns classification approach has successfully treated the keypoint matching as a classification problem [Ozuy 10]. Based on random trees [Lepe 06], ferns has introduced a non-hierarchical structure for classification. The combination of this structure with binary features have led to a fast algorithm.

For the ferns classifier, the object of interest is expressed as a set of keypoints. Each keypoint defines a patch centered on it and the number of keypoints is equal to the number of classes. The goal is then to learn these classes during training and detect them using a Semi-Naive Bayesian formulation afterwards [Ozuy 10]. Consider $c_i, i = 1, ...H$ classes to be learned. In addition, a set of binary features, $f_j, j = 1, ...N$, are calculated over the image patches. Binary features are selected randomly and divided into M ferns (i.e. groups) of size $S = \frac{N}{M}$. In order to estimate the joint probability of the binary features f_j conditioned on the learned classes c_i, independence between the ferns is assumed. Then, the conditional probability of an image patch given the classes becomes

$$P(f_1, f_2, ..., f_N \mid C = c_i) = \prod_{k=1}^{M} P(F_k \mid C = c_i) \tag{A.15}$$

where $F_k = \left\{f_{\sigma(k,1)}, f_{\sigma(k,2)}, ..., f_{\sigma(k,S)}\right\}, k = 1, ...M$ and $\sigma(k, j)$ a random permutation function for grouping the binary features randomly.

During training, there is a reference image from which robust keypoints are being detected. This is done by transforming the reference frame, detecting and matching keypoints iteratively. The keypoints with the most votes are kept for learning. In the last step of training, a normalization is done for each class separately and on each fern independently. In our problem, the process of training is repeated for each object separately.

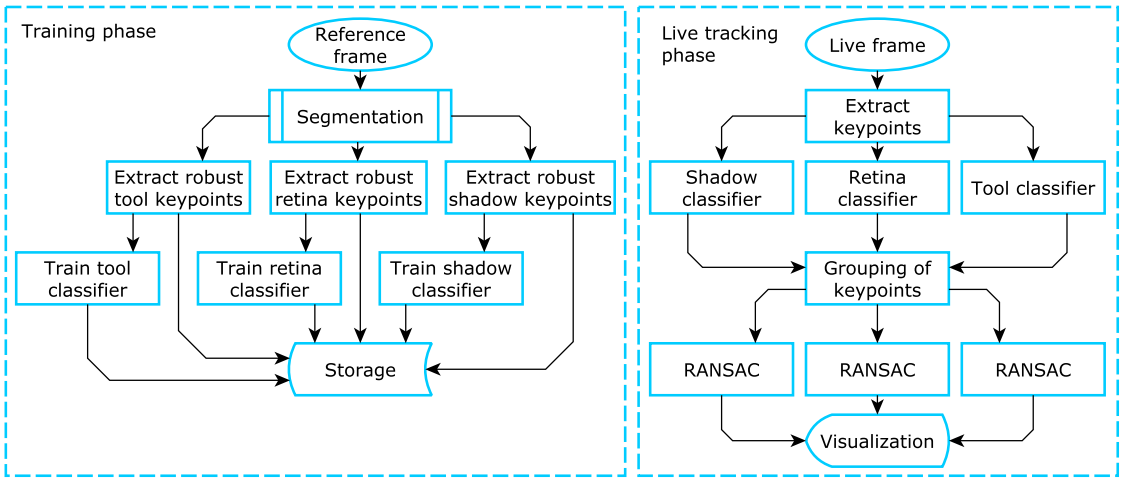


FIGURE A.9.: Flow chart of the proposed multi-target tracking algorithm. The entire process consists of two phases, namely, the training phase(left) where we train our classifiers individually and the testing phase (right) where we test our classifiers on detected live keypoints simultaneously.

Multi Target Tracking

In our image data sets, in addition to retina as the main object of interest, there is also a surgical tool and its shadow. The presence of these additional objects leads to occlusions and/or illumination differences creating large number of outliers in the detected correspondences by random ferns reducing the tracking performance. To overcome this issue, we propose to track each object separately by employing a random ferns classifier for each object.

Let $\mathcal{R} = R_1, \ldots, R_n$ be a set of classifiers for n objects $\mathcal{O} = O_1, \ldots, O_n,\ i = 1, \ldots, n$. We start by training each classifier R_i on the reference frame where O_i is visible. In order to do that, we perform a manual segmentation of each object region for initialization. Then, each classifier, R_i, is trained in one of this subregions as described in Section A.3.2 where the number of robust keypoints in each subregion determines the number of classes, H_{O_i}, for that classifier. After the extraction of robust reference keypoints, we apply a non-maximum suppresion criteria to sparsify the occurrences of the robust keypoints in densely sampled regions to have a uniform distribution of keypoints for each object in the scene. This novel sparsification step helps prevent sample selection bias for the estimated homography towards densely sampled regions. Once each classifier is trained, the following steps are illustrated in Fig. A.9.

We use a very simple keypoint detector [Lepe 06] along with binary features proposed in [Ozuy 10]. Another novel aspect of our approach is the pre-grouping of detected keypoints based on their matching scores, which improves the estimation performance significantly. Class posteriors are used as match scores where a live keypoint is removed from the correspondence list of classifier R_i if the posterior assigned by R_i is smaller than the posteriors assigned by the others classifiers. This pre-grouping step decreases the probability that

	Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	Error	S.Rate	Error	S.Rate	Error	S.Rate	Error	S.Rate
NCC	82±74	32%	570±540	0%	210±130	2%	340±280	0%
SIFT	4±1	100%	620±1100	22%	310±200	0%	78±67	8%
Single Classifier	5±2	96%	84±140	52%	270±340	24%	14±18	72%
Multiple Classifiers	4±2	100%	31±50	68%	28±45	60%	16±12	54%

TABLE A.2.: Comparison of tracking errors (pixels) and success rates of each method on 4 different datasets. It is obvious that the proposed approach performs much better that the others especially on datasets 2 and 3 where the scene had occlusions.

candidate keypoints for an object O_i actually belong to a different object O_j and become outliers for O_i.

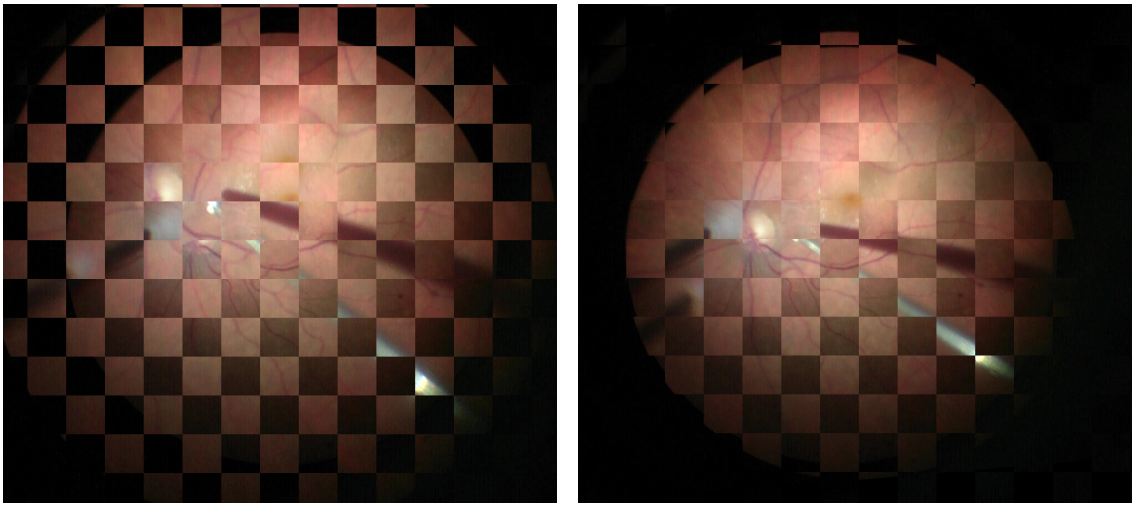
A.3.3. Experiments and Results

For evaluation, the proposed method was applied to 4 in vivo human image sequences from peeling operation. The images were acquired by Carl-Zeiss Lumera 700 operating microscope with 1080p resolution at 25 fps. Ground truth for each data set was generated by choosing several (at least 6) landmarks on the reference images as well as in subsequent frames (c.f. Fig. A.11(a)). The process was repeated to determine the intra- and inter-observation variations. All experiments were performed using an unoptimized C++ implementation of ferns, running on an Intel Core2Duo 3GHz machine. The runtime limited us to process up to 5 frames per second. However, an efficient implementation by exploiting GPU would lead to considerably speedup.

For each dataset, from the reference frame, we extracted 150, 100 and 100 robust keypoints from retina, shadow and tool regions, respectively, which we used to train 3 random ferns classifiers. The robust keypoints were selected by creating 3000 random warps of the reference image, extracting keypoints from each warped image, and then back-projecting the locations of the extracted keypoints. Then, we applied the sparsification method as described in Section A.3.2. Each octave was treated separately, and the ferns were trained using patches on the corresponding octave level, with patches of sizes 32x32. After the classifiers were trained, each live frame was processed by extracting 300 keypoints. All keypoints were matched using each of the 3 classifiers. We only kept 1-to-1 matches, i.e. each reference keypoint of each object had maximum 1 corresponding keypoint with the highest probability of match on the live frame. We selected the match with the highest probability from the classifier.

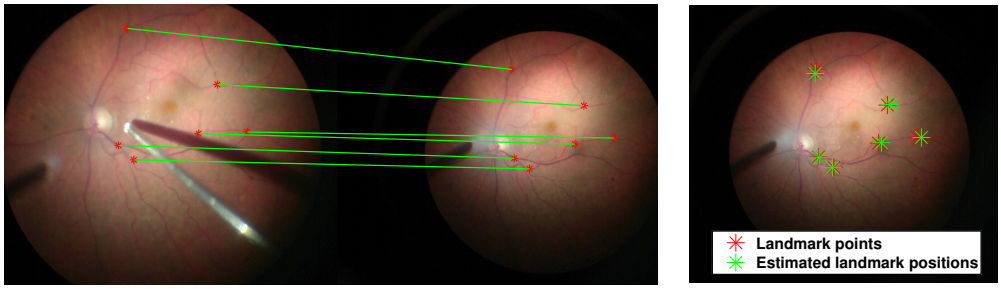
The effectiveness of the proposed method in terms of motion estimation accuracy can be first visually inferred from Fig. A.10 where a pair of reference and live frames before and after motion compensation are overlayed using checkerboard pattern. Despite the occlusion due to the tool and its shadow in the live frame and their absence in the reference frame, the estimated motion is still accurate and the vascular structures are continuous in Fig. A.10(b). The performance can be further inferred from Fig. A.11(b) where the consistency between the ground truth landmark in the live frame (red) and warped landmarks of the reference frame (green) can be clearly observed.

For quantitative evaluation, the proposed multi-target tracking approach was compared with three different approaches; (1) SIFT-based tracking [Lowe 99], (2) the NCC-based di-



(a) Overlay before tracking (b) Overlay after tracking

FIGURE A.10.: Demonstration of the effectiveness of the proposed tracking approach on a pair of reference and live images. The images are overlayed (a) before and (b) after the motion was compensated. Note the consistency of vascular structures after compensation despite the presence of occlusions.



(a) (b)

FIGURE A.11.: Illustration of the matching performance in terms of ground truth landmark matching. (a) Manual selection of corresponding landmarks on the reference (left) and live (right) frames. (b) Annotated landmarks in the live frame (red) and mapped landmarks from the reference frame by using the proposed homography estimation for retina. Note the agreement between the mapped and the manually extracted landmarks.

rect visual tracking, (3) tracking only a single object using a single random ferns classifier (without pre-grouping of keypoints) and RANSAC. The average distance error for retina landmarks is used as the evaluation criterion. Fig. A.12 and Table A.2 compares the overall performance of our approach to others. A success rate of each method on each dataset is calculated by deeming a tracking failed if the error is more than 15px and then taking the ratio of the number of successful and the total number frames. Clearly, the proposed method outperforms the single-target tracking and other approaches especially on datasets 2 and 3 where the scenes were considerably occluded by the tool whereas a similar performance

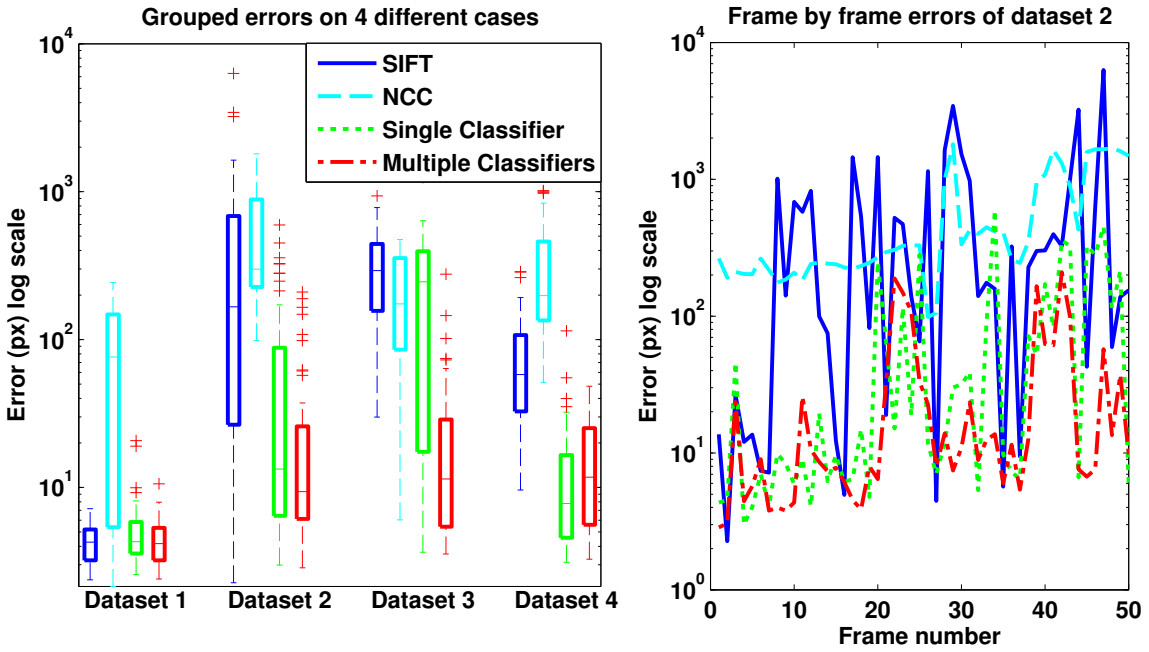


FIGURE A.12.: Comparison of the performance of our approach with others. (Left) Tracking errors for each method on 4 datasets. The proposed multiple classifiers led to significantly lower tracking errors compared to the others, especially for the case of the single classifier. (Right) Tracking errors per frame for dataset 2 where the scene was often occluded.

of single and multi-object tracking is observed in dataset 1 where there are no occlusions. Dataset 4 also contained no tool but low contrast frames with extreme non-uniform illumination leading to a decrease in the performance. Overall, the average success rate was 70.5% whereas it was 64% on average in cases with occlusions which is comparable to the error reported in [Rich 12] considering the differences in frame resolutions. The intra- and inter-observer variations of manual extraction of landmarks from our 1080p image data sets were 3.4 and 8.5 pixels, respectively.

A.3.4. Discussion and Conclusion

In this paper, we have presented a novel multi-object tracking approach using multiple random ferns classifiers in order to efficiently and robustly track the retina. Other contributions include the sparse sampling of reference keypoints during the training of classifiers which helps prevent sample selection bias and the pre-grouping of matched live keypoints into object categories reducing the number of false matches and resulting in an increased performance of subsequent robust homography estimation. The proposed multi-target tracking method proved to be robust to the occlusions which was experimentally demonstrated on retina image sequences. We compared our approach to tracking with single classifier and other state of the art methods such as SIFT-based tracking. Note that although we estimate

the homographies also for the tool and its shadow, we did not quantify this since our main object of interest is the retina itself. However, we use the classification response for the tool to boost the matching performance for the retina.

A nice extension of our approach would be the detection and/or avoidance of occlusions during surgery for an augmented reality assistance system for retinal surgery. Besides that, several other improvements can be made. For instance, although the keypoints are uniformly sampled from the reference frame, extraction of live keypoints are still biases towards highly illuminated regions. Therefore, employment of the same sparsification approach could further improve the matching performance. Moreover, a quadratic motion model can be incorporated which can help reduce the tracking error.

APPENDIX

B

AUTHORED AND CO-AUTHORED PUBLICATIONS

[Esla 11] A. Eslami, M. Yigitsoy, and N. Navab. "Manifold learning for shape guided segmentation of Cardiac boundaries: Application to 3D+t Cardiac MRI". In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pp. 2658–2662, IEEE, 2011.

[Esla 12] A. Eslami, A. Teimori, M. Yigitsoy, and N. Navab. "A new framework for morphological and morphometric study of fish species based on groupwise registration of otolith images". In: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pp. 1679–1682, IEEE, 2012.

[Mill 14] F. Milletari, M. Yigitsoy, N. Navab, and S.-A. Ahmadi. "Left Ventricle Segmentation in Cardiac Ultrasound Using Hough-Forests With Implicit Shape and Appearance Priors". In: *MICCAI Challenge on Endocardial Three-dimensional Ultrasound Segmentation (CETUS)*, Boston, MA, September 2014.

[Muel 14] M. Mueller, M. Yigitsoy, H. Heibel, and N. Navab. "Deformable Reconstruction of Histology Sections using Structural Probability Maps". In: *International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Springer, Boston, USA, September 2014.

[Peng 14] T. Peng, M. Yigitsoy, A. Eslami, C. Bayer, and N. Navab. "Deformable Registration of Multi-modal Microscopic Images Using a Pyramidal Interactive Registration-Learning Methodology". In: *Biomedical Image Registration*, pp. 144–153, Springer, 2014.

[Wach 10] C. Wachinger, M. Yigitsoy, and N. Navab. "Manifold learning for image-based breathing gating with application to 4D ultrasound". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pp. 26–33, Springer, 2010.

[Wach 12] C. Wachinger, M. Yigitsoy, E.-J. Rijkhorst, and N. Navab. "Manifold learning for image-based breathing gating in ultrasound and MRI". *Medical image analysis*, Vol. 16, No. 4, pp. 806–818, 2012.

[Yigi 11a] M. Yigitsoy, C. Wachinger, and N. Navab. "Manifold learning for image-based breathing gating in mri". In: *SPIE Medical Imaging*, pp. 796210–796210, International Society for Optics and Photonics, 2011.

[Yigi 11b] M. Yigitsoy, C. Wachinger, and N. Navab. "Temporal groupwise registration for motion modeling". In: *Information Processing in Medical Imaging*, pp. 648–659, Springer, 2011.

[Yigi 12a] M. Yigitsoy, S. Kirchhoff, M. F. Reiser, and N. Navab. "Out-of-plane motion compensation in cine-MRI". In: *Abdominal Imaging. Computational and Clinical Applications*, pp. 107–115, Springer, 2012.

[Yigi 12b] M. Yigitsoy, M. F. Reiser, N. Navab, and S. Kirchhoff. "Dynamic graph cuts for colon segmentation in functional cine-MRI". In: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pp. 1268–1271, IEEE, 2012.

[Yigi 13] M. Yigitsoy and N. Navab. "Structure Propagation for Image Registration". *Medical Imaging, IEEE Transactions on*, Vol. 32, No. 9, pp. 1657–1670, Sept 2013.

[Yigi 14] M. Yigitsoy, J. Fotouhi, and N. Navab. "Hough Space Parametrization: Ensuring Global Consistency in Intensity-based Registration". In: *International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Springer, Boston, USA, September 2014.

APPENDIX
C

LIST OF FIGURES

APPENDIX

D

LIST OF TABLES

APPENDIX
E

BIBLIOGRAPHY

[Ahma 11] S.-A. Ahmadi, M. Baust, A. Karamalis, A. Plate, K. Boetzel, T. Klein, and N. Navab. "Midbrain segmentation in transcranial 3d ultrasound for parkinson diagnosis". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 362–369, Springer, 2011. 87

[Ahma 13] S.-A. Ahmadi. *Image Analysis Methods in Multi-View Transcranial 3D Ultrasound for Diagnosis of Neurological Movement-Disorders*. PhD thesis, Technische Universität München, München, 2013. 87

[Atas 08] S. Atasoy, D. P. Noonan, S. Benhimane, N. Navab, and G.-Z. Yang. "A global approach for automatic fibroscopic video mosaicing in minimally invasive diagnosis". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pp. 850–857, Springer, 2008. 26, 87

[Augu 03] J. August and S. Zucker. "Sketches with curvature: The curve indicator random field and markov processes". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 25, No. 4, pp. 387–400, 2003. 30

[Augu 05] J. August and T. Kanade. "The role of non-overlap in image registration". In: *Information Processing in Medical Imaging*, pp. 713–724, Springer, 2005. 16

[Bagc 10] U. Bagci and L. Bai. "Automatic best reference slice selection for smooth volume reconstruction of a mouse brain from histological images.". *IEEE Trans Med Imaging*, Vol. 29, No. 9, pp. 1688–1696, Sep 2010. 26, 27, 28

[Bake 11] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. "A database and evaluation methodology for optical flow". *IJCV*, Vol. 92, No. 1, pp. 1–31, 2011. 69, 72

[Ball 76] D. Ballard and J. Sklansky. "A Ladder-Structured Decision Tree for Recognizing Tumors in Chest Radiographs". *Computers, IEEE Transactions on*, Vol. C-25, No. 5, pp. 503–513, May 1976. 94

[Ball 81] D. H. Ballard. "Generalizing the Hough transform to detect arbitrary shapes". *Pattern recognition*, Vol. 13, No. 2, pp. 111–122, 1981. 6, 81, 89, 90, 91, 92, 95

[Bard 02] E. Bardinet, S. Ourselin, D. Dormont, G. Malandain, D. Tandé, K. Parain, N. Ayache, and J. Yelnik. "Co-registration of histological, optical and MR data of the human brain". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2002*, pp. 548–555, Springer, 2002. 29

[Bass 05] G. Bassotti, G. de Roberto, D. Castellani, L. Sediari, and A. Morelli. "Normal aspects of colorectal motility and abnormalities in slow transit constipation". *World J Gastroenterol*, Vol. 11, No. 18, pp. 2691–2696, 2005. 123, 130

[Bass 96] P. J. Basser and C. Pierpaoli. "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI". *Journal of Magnetic Resonance, Series B*, Vol. 111, No. 3, pp. 209–219, 1996. 41

[Bay 06] H. Bay, T. Tuytelaars, and L. Van Gool. "Surf: Speeded up robust features". In: *Computer Vision–ECCV 2006*, pp. 404–417, Springer, 2006. 83

[Belk 03] M. Belkin and P. Niyogi. "Laplacian eigenmaps for dimensionality reduction and data representation". *Neural computation*, Vol. 15, No. 6, pp. 1373–1396, 2003. 125, 126

[Berl 11] M. Berlanga, S. Phan, E. Bushong, S. Wu, O. Kwon, B. Phung, S. Lamont, M. Terada, T. Tasdizen, M. Martone, *et al.* "Three-dimensional reconstruction of serial mouse brain sections: solution for flattening high-resolution large-scale mosaics.". *Frontiers in neuroanatomy*, Vol. 5, p. 17, 2011. 28

[Book 89] F. L. Bookstein. "Principal warps: Thin-plate splines and the decomposition of deformations". *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 11, No. 6, pp. 567–585, 1989. 14

[Boyk 01] Y. Y. Boykov and M.-P. Jolly. "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images". In: *Proc. Eighth IEEE Int. Conf. Computer Vision ICCV 2001*, pp. 105–112, 2001. 131, 132

[Boyk 04] Y. Boykov and V. Kolmogorov. "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 26, No. 9, pp. 1124–1137, 2004. 132

[Brat 11] L. Brattain and R. Howe. "Real-time 4d ultrasound mosaicing and visualization". *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 105–112, 2011. 26, 87, 89

[Broe 11] A. M. Broehan, T. Rudolph, C. A. Amstutz, and J. H. Kowal. "Real-time multimodal retinal image registration for a computer-assisted laser photocoagulation system". *Biomedical Engineering, IEEE Transactions on*, Vol. 58, No. 10, pp. 2816–2824, 2011. 138

[Brow 07] M. Brown and D. Lowe. "Automatic panoramic image stitching using invariant features". *International Journal of Computer Vision*, Vol. 74, No. 1, pp. 59–73, 2007. 30, 77, 86

[Brow 10] M. Brown and R. Semelka. *MRI: basic principles and applications*. Wiley-Blackwell, 2010. 10

[Brow 92] L. G. Brown. "A survey of image registration techniques". *ACM computing surveys (CSUR)*, Vol. 24, No. 4, pp. 325–376, 1992. 10

[Buhm 05] S. Buhmann, C. Kirchhoff, C. Wielage, T. Mussack, M. F. Reiser, and A. Lienemann. "Assessment of large bowel motility by cine magnetic resonance imaging using two different prokinetic agents: a feasibility study.". *Invest Radiol*, Vol. 40, No. 11, pp. 689–694, Nov 2005. 124, 130

[Cape 09] M. Čapek, P. Brůža, J. Janáček, P. Karen, L. Kubínová, and R. Vagnerová. "Volume reconstruction of large tissue specimens from serial physical sections using confocal microscopy and correction of cutting deformations by elastic registration". *Microscopy research and technique*, Vol. 72, No. 2, pp. 110–119, 2009. 26, 28

[Catt 06] P. C. Cattin, H. Bay, L. Van Gool, and G. Székely. "Retina mosaicing using local features". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 185–192, Springer, 2006. 87

[Chen 95] Y. Cheng. "Mean shift, mode seeking, and clustering". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 17, No. 8, pp. 790–799, 1995. 96, 113

[Chow 06] S. Chow, H. Hakozaki, D. Price, N. MacLean, T. Deerinck, J. Bouwer, M. Martone, S. Peltier, and M. Ellisman. "Automated microscopy system for mosaic acquisition and processing". *Journal of microscopy*, Vol. 222, No. 2, pp. 76–84, 2006. 26, 86

[Chun 02] A. C. Chung, W. M. Wells III, A. Norbash, and W. E. L. Grimson. "Multi-modal image registration by minimising kullback-leibler distance". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2002*, pp. 525–532, Springer, 2002. 77

[Cifo 11] A. Cifor, L. Bai, and A. Pitiot. "Smoothness-guided 3-D reconstruction of 2-D histological images". *Neuroimage*, Vol. 56, No. 1, pp. 197–211, 2011. 29, 76

[Comm 05] O. Commowick, R. Stefanescu, P. Fillard, V. Arsigny, N. Ayache, X. Pennec, and G. Malandain. "Incorporating statistical measures of anatomical variability in atlas-to-subject registration for conformal brain radiotherapy". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pp. 927–934, Springer, 2005. 21

[Csur 04] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. "Visual categorization with bags of keypoints". In: *Workshop on statistical learning in computer vision, ECCV*, pp. 1–2, 2004. 6, 84

[Dais 03] J. Daisne, M. Sibomana, A. Bol, G. Cosnard, M. Lonneux, and V. Grégoire. "Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency". *Radiotherapy and Oncology*, Vol. 69, No. 3, pp. 237–245, 2003. 10

[Dala 05] N. Dalal and B. Triggs. "Histograms of oriented gradients for human detection". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pp. 886–893, IEEE, 2005. 83

[Daug 07] J. Dauguet, D. Bock, R. C. Reid, and S. K. Warfield. "Alignment of large image series using cubic B-splines tessellation: application to transmission electron microscopy data.". *Med Image Comput Comput Assist Interv*, Vol. 10, No. Pt 2, pp. 710–717, 2007. 53

[Dera 10] J. J. Derakhshan, M. A. Griswold, S. G. Nour, J. L. Sunshine, and J. L. Duerk. "Characterization and reduction of saturation banding in multiplanar coherent and incoherent steady-state imaging.". *Magn Reson Med*, Vol. 63, No. 5, pp. 1415–1421, May 2010. 124

[Deri 87] R. Deriche. "Using Canny's criteria to derive a recursively implemented optimal edge detector". *International journal of computer vision*, Vol. 1, No. 2, pp. 167–187, 1987. 58

[Deri 93] R. Deriche. "Recursively implementating the Gaussian and its derivatives". 1993. 72, 74

[Diet 08] O. Dietrich and S. O. Schoenberg. "Technical Prerequisites". In: M. F. Reiser, G. van Kaick, C. Fink, and S. O. Schoenberg, Eds., *Screening and Preventive Diagnosis with Radiological Imaging*, pp. 77–126, Springer Berlin Heidelberg, 2008. 54, 71

[Duda 72] R. O. Duda and P. E. Hart. "Use of the Hough transformation to detect lines and curves in pictures". *Communications of the ACM*, Vol. 15, No. 1, pp. 11–15, 1972. 90, 91

[Dzyu 13] O. Dzyubachyk, B. P. F. Lelieveldt, J. Blaas, M. Reijnierse, A. Webb, and R. J. van der Geest. "Automated algorithm for reconstruction of the complete spine from multistation 7T MR data". *Magnetic Resonance in Medicine*, Vol. 69, No. 6, pp. 1777–1786, 2013. 26

[Emme 09] M. Emmenlauer, O. Ronneberger, A. Ponti, P. Schwarb, A. Griffa, A. Filippi, R. Nitschke, W. Driever, and H. Burkhardt. "XuvTools: free, fast and reliable stitching of large 3D datasets.". *J Microsc*, Vol. 233, No. 1, pp. 42–60, 2009. 27, 30, 86

[Fane 13] G. Fanelli, M. Dantone, J. Gall, A. Fossati, and L. Van Gool. "Random forests for real time 3d face analysis". *International Journal of Computer Vision*, Vol. 101, No. 3, pp. 437–458, 2013. 85

[Fens 01] A. Fenster, D. Downey, and H. Cardinal. "Three-dimensional ultrasound imaging". *Physics in medicine and biology*, Vol. 46, p. R67, 2001. 10

[Ferg 03] R. Fergus, P. Perona, and A. Zisserman. "Object class recognition by unsupervised scale-invariant learning". In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, pp. II–264, IEEE, 2003. 6, 84

[Ferr 10] V. Ferrari, F. Jurie, and C. Schmid. "From images to shape models for object detection". *International Journal of Computer Vision*, Vol. 87, No. 3, pp. 284–303, 2010. 96, 114

[Feue 11] M. Feuerstein, H. Heibel, J. Gardiazabal, N. Navab, and M. Groher. "Reconstruction of 3-D histology images by simultaneous deformable registration". *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 582–589, 2011. 26, 27, 29, 54, 69

[Fiel 93] D. J. Field, A. Hayes, and R. F. Hess. "Contour integration by the human visual system: evidence for a local "association field".". *Vision Res*, Vol. 33, No. 2, pp. 173–193, Jan 1993. 30

[Fisc 04] B. Fischer and J. Modersitzki. "A unified approach to fast image registration and a new curvature based registration technique". *Linear Algebra and its applications*, Vol. 380, pp. 107–124, 2004. 10, 20

[Fisc 81] M. A. Fischler and R. C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". *Communications of the ACM*, Vol. 24, No. 6, pp. 381–395, 1981. 138

[Fran 06] E. Franken, M. van Almsick, P. Rongen, L. Florack, and B. ter Haar Romeny. "An efficient method for tensor voting using steerable filters". *ECCV 2006*, pp. 228–240, 2006. 30

[Gaff 11] S. Gaffling, V. Daum, and J. Hornegger. "Landmark-constrained 3-D Histological Imaging: A Morphology-preserving Approach.". In: *VMV*, pp. 309–316, The Eurographics Association, 2011. 29, 53

[Gaff 14] S. Gaffling, V. Daum, S. Steidl, A. Maier, H. Kostler, and J. Hornegger. "A Gauss-Seidel Iteration Scheme for Reference-Free 3-D Histological Image Reconstruction". *Medical Imaging, IEEE Transactions on*, Vol. PP, No. 99, pp. 1–1, 2014. 29, 76

[Gall 11] J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky. "Hough forests for object detection, tracking, and action recognition.". *IEEE transactions on pattern analysis and machine intelligence*, Vol. 33, No. 11, p. 2188, 2011. 84, 85

[Gall 13] J. Gall and V. Lempitsky. "Class-specific hough forests for object detection". In: *Decision Forests for Computer Vision and Medical Image Analysis*, pp. 143–157, Springer, 2013. 84, 96, 97, 114

[Gatz 06] T. Gatzke and C. M. Grimm. "Estimating curvature on triangular meshes". *International journal of shape modeling*, Vol. 12, No. 1, pp. 1–28, 2006. 43

[Geor 08] M. Georg, R. Souvenir, A. Hope, and R. Pless. "Manifold learning for 4D CT reconstruction of the lung". In: *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pp. 1–8, IEEE, 2008. 124

[Gloc 07] B. Glocker, S. Buhmann, C. Kirchhoff, T. Mussack, M. Reiser, and N. Navab. "Towards a Computer Aided Diagnosis System for Colon Motility Dysfunctions". In: *SPIE Medical Imaging*, International Society for Optics and Photonics, San Diego, California, USA, February 2007. 124, 130

[Gloc 08] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. "Dense Image Registration through MRFs and Efficient Linear Programming". *Med Image Anal*, Vol. 12, No. 6, pp. 731–741, 2008. 17, 18, 19, 32, 51, 52, 56

[Gloc 11] 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*, Vol. 13, pp. 219–244, 2011. 17, 18

[Gode 13] M. Godec, P. M. Roth, and H. Bischof. "Hough-based tracking of non-rigid objects". *Computer Vision and Image Understanding*, Vol. 117, No. 10, pp. 1245–1256, 2013. 96, 114

[Grau 11] K. Grauman and B. Leibe. *Visual object recognition*. Morgan & Claypool Publishers, 2011. 83

[Guy 97] G. Guy and G. Medioni. "Inference of surfaces, 3D curves, and junctions from sparse, noisy, 3D data". *IEEE Trans. PAMI*, Vol. 19, No. 11, pp. 1265–1277, 1997. 37

[Haci 13] I. Hacihaliloglu, D. R. Wilson, M. Gilbart, M. A. Hunt, and P. Abolmaesumi. "Non-iterative partial view 3D ultrasound to CT registration in ultrasound-guided computer-assisted orthopedic surgery". *International journal of computer assisted radiology and surgery*, Vol. 8, No. 2, pp. 157–168, 2013. 88

[Hart 09] P. E. Hart. "How the Hough transform was invented [DSP History]". *Signal Processing Magazine, IEEE*, Vol. 26, No. 6, pp. 18–22, 2009. 89, 90

[He 03] Y. He, A. B. Hamza, and H. Krim. "A generalized divergence measure for robust image registration". *Signal Processing, IEEE Transactions on*, Vol. 51, No. 5, pp. 1211–1220, 2003. 77

[Heim 09] T. Heimann and H.-P. Meinzer. "Statistical shape models for 3D medical image segmentation: A review". *Medical image analysis*, Vol. 13, No. 4, pp. 543–563, 2009. 82

[Hern 10] Y. Hernandez-Mier, W. Blondel, C. Daul, D. Wolf, and F. Guillemin. "Fast construction of panoramic images for cystoscopic exploration". *Computerized Medical Imaging and Graphics*, Vol. 34, No. 7, pp. 579–592, 2010. 87

[Hill 01] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. "Medical image registration". *Physics in medicine and biology*, Vol. 46, No. 3, p. R1, 2001. 10

[Houg 62] P. V. Hough. "Method and means for recognizing complex patterns". Dec. 18 1962. US Patent 3,069,654. 6, 89

[Illi 88] J. Illingworth and J. Kittler. "A survey of the Hough transform". *Computer vision, graphics, and image processing*, Vol. 44, No. 1, pp. 87–116, 1988. 96

[Itsk 09] M. Itskov. *Tensor Algebra and Tensor Analysis for Engineers: With Applications to Continuum Mechanics*. Springer Publishing Company, Incorporated, 2nd Ed., 2009. 35

[Jahn 93] B. Jähne. *Spatio-temporal image processing: theory and scientific applications*. Vol. 751, Springer, 1993. 42

[Jia 03] J. Jia and C.-K. Tang. "Image repairing: Robust image synthesis by adaptive nd tensor voting". In: *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, pp. I–643, IEEE, 2003. 30

[Jia 05] J. Jia and C. Tang. "Tensor voting for image correction by global and local intensity alignment". *IEEE Trans. PAMI*, Vol. 27, No. 1, pp. 36–50, 2005. 30

[Jia 08] J. Jia and C. Tang. "Image stitching using structure deformation". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 30, No. 4, pp. 617–631, 2008. 30

[John] S. G. Johnson. "The NLopt nonlinear-optimization package". `http://ab-initio.mit.edu/wiki/index.php/NLopt`. Accessed: 2014-21-02. 16, 106

[Ju 06] T. Ju, J. Warren, J. Carson, M. Bello, I. Kakadiaris, W. Chiu, C. Thaller, and G. Eichele. "3D volume reconstruction of a mouse brain from histological sections using warp filtering". *Journal of Neuroscience Methods*, Vol. 156, No. 1, pp. 84–100, 2006. 70

[Kayn 10] V. Kaynig, B. Fischer, E. Mueller, and J. Buhmann. "Fully automatic stitching and distortion correction of transmission electron microscope images.". *Journal of structural biology*, Vol. 171, No. 2, pp. 163–173, 2010. 28

[Khur 11] P. Khurd, R. Oketokoun, T. Gajera, J. Frangioni, L. Grady, H. Sundar, S. Gibbs-Strauss, and A. Kamen. "Global error minimization in image mosaicing using graph connectivity and its applications in microscopy". *Journal of Pathology Informatics*, Vol. 2, 2011. 28

[Kim 10] K. Kim, P. A. Habas, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme. "Intersection based motion correction of multislice MRI for 3-D in utero fetal brain image formation.". *IEEE Trans Med Imaging*, Vol. 29, No. 1, pp. 146–158, Jan 2010. 124

[Kimm 75] C. Kimme, D. Ballard, and J. Sklansky. "Finding circles by an array of accumulators". *Communications of the ACM*, Vol. 18, No. 2, pp. 120–122, 1975. 90, 91

[King 08] B. King and R. P. Institute. *Range Data Analysis by Free-space Modeling and Tensor Voting*. Rensselaer Polytechnic Institute, 2008. 30, 35, 36, 37, 38, 49

[Kirc 11] S. Kirchhoff, M. Nicolaus, J. Schirra, M. Reiser, B. Göke, and A. Lienemann. "Assessment of colon motility using simultaneous manometric and functional cine-MRI analysis: preliminary results". *Abdominal imaging*, Vol. 36, No. 1, pp. 24–30, 2011. 125, 131

[Kohl 07] P. Kohli and P. Torr. "Dynamic graph cuts for efficient inference in markov random fields". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2079–2088, 2007. 132

[Kolm 07] V. Kolmogorov and C. Rother. "Minimizing nonsubmodular functions with graph cuts-a review". *TPAMI*, Vol. 29, No. 7, pp. 1274–1279, 2007. 17, 52

[Komo 07] N. Komodakis and G. Tziritas. "Approximate labeling via graph cuts based on linear programming". *IEEE Trans. PAMI*, Vol. 29, No. 8, pp. 1436–1453, 2007. 17, 51, 56

[Komo 08] N. Komodakis, G. Tziritas, and N. Paragios. "Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies". *Computer Vision and Image Understanding*, Vol. 112, No. 1, pp. 14–29, 2008. 17, 51, 56

[Kong 01] W. Kong and B. B. Kimia. "On solving 2D and 3D puzzles using curve matching". In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, pp. II–583, IEEE, 2001. 28

[Kutt 08] O. Kutter, S. Kirchhoff, M. Berkovic, M. Reiser, and N. Navab. "Spatio-temporal registration in multiplane MRI acquisitions for 3D Colon Motility analysis". In: *SPIE Medical Imaging*, International Society for Optics and Photonics, San Diego, California, USA, February 2008. 124

[Kutt 09] O. Kutter, W. Wein, and N. Navab. "Multi-modal registration based ultrasound mosaicing". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pp. 763–770, Springer, 2009. 88

[Lee 08] S.-C. Lee and P. Bajcsy. "Trajectory Fusion for Three-dimensional Volume Reconstruction.". *Comput Vis Image Underst*, Vol. 110, No. 1, pp. 19–31, Apr 2008. 26, 28

[Leib 04] B. Leibe, A. Leonardis, and B. Schiele. "Combined Object Categorization and Segmentation With an Implicit Shape Model". In: *Workshop on Statistical Learning in Computer Vision (ECCV)*, Springer, May 2004. 6, 84, 85, 96, 97

[Leib 08] B. Leibe, A. Leonardis, and B. Schiele. "Robust object detection with interleaved categorization and segmentation". *International journal of computer vision*, Vol. 77, No. 1-3, pp. 259–289, 2008. 96

[Lepe 06] V. Lepetit and P. Fua. "Keypoint recognition using randomized trees". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 28, No. 9, pp. 1465–1479, 2006. 139, 140

[Lien 00] A. Lienemann, D. Sprenger, H. O. Steitz, M. Korell, and M. Reiser. "Detection and mapping of intraabdominal adhesions by using functional cine MR imaging: preliminary results.". *Radiology*, Vol. 217, No. 2, pp. 421–425, Nov 2000. 124, 130

[Loew 11] K. Loewke, D. Camarillo, W. Piyawattanametha, M. Mandella, C. Contag, S. Thrun, and J. Salisbury. "In vivo micro-image mosaicing". *IEEE Trans. BME*, Vol. 58, No. 1, pp. 159–171, 2011. 26, 30, 86, 87

[Loss 11] L. A. Loss, G. Bebis, and B. Parvin. "Iterative tensor voting for perceptual grouping of ill-defined curvilinear structures.". *IEEE Trans Med Imaging*, Vol. 30, No. 8, pp. 1503–1513, Aug 2011. 30

[Lowe 04] D. G. Lowe. "Distinctive image features from scale-invariant keypoints". *International journal of computer vision*, Vol. 60, No. 2, pp. 91–110, 2004. 83, 87

[Lowe 99] D. G. Lowe. "Object recognition from local scale-invariant features". In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, pp. 1150–1157, IEEE, 1999. 141

[Maes 97] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. "Multimodality image registration by maximization of mutual information". *Medical Imaging, IEEE Transactions on*, Vol. 16, No. 2, pp. 187–198, 1997. 16, 99

[Main 98] J. Maintz and M. A. Viergever. "A survey of medical image registration". *Medical image analysis*, Vol. 2, No. 1, pp. 1–36, 1998. 10

[Mala 04] G. Malandain, Éric Bardinet, K. Nelissen, and W. Vanduffel. "Fusion of autoradiographs with an MR volume using 2-D and 3-D linear transformations". *NeuroImage*, Vol. 23, No. 1, pp. 111 – 127, 2004. 53

[Mann 94] S. Mann and R. W. Picard. "Virtual bellows: Constructing high quality stills from video". In: *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, pp. 363–367, IEEE, 1994. 86

[McLa 02] P. F. McLauchlan and A. Jaenicke. "Image mosaicing using sequential bundle adjustment". *Image and Vision Computing*, Vol. 20, No. 9, pp. 751–759, 2002. 86

[Medi 00a] G. Medioni, M. Lee, and C. Tang. *A computational framework for segmentation and grouping*. Vol. 1, Elsevier Science, 2000. 5, 30, 35, 38

[Medi 00b] G. Medioni, C. Tang, and M. Lee. "Tensor voting: Theory and applications". *Proceedings of RFIA, Paris, France*, 2000. 34, 38

[Medi 04] G. Medioni and S. Kang. *Emerging topics in computer vision*. Prentice Hall PTR, 2004. 34, 37

[Merc 05] L. Mercier, T. Langø, F. Lindseth, and L. Collins. "A review of calibration techniques for freehand 3-D ultrasound systems". *Ultrasound in medicine & biology*, Vol. 31, No. 2, pp. 143–165, 2005. 10

[Merl 75] P. M. Merlin and D. J. Farber. "A parallel mechanism for detecting curves in pictures". *IEEE Transactions on Computers*, Vol. 24, No. 1, pp. 96–98, 1975. 90

[Mode 03] J. Modersitzki. *Numerical methods for image registration*. OUP Oxford, 2003. 10, 20

[Mode 08] J. Modersitzki. "FLIRT with rigidity—image registration with a local non-rigidity penalty". *International Journal of Computer Vision*, Vol. 76, No. 2, pp. 153–163, 2008. 21

[Mong 91] O. Monga, R. Deriche, G. Malandain, and J.-P. Cocquerez. "Recursive filtering and edge tracking: two primary tools for 3-D edge detection". *Image and Vision Computing*, Vol. 9, No. 4, pp. 203–214, August 1991. 33

[Mord 05] P. Mordohai. *A Perceptual Organization Approach for Figure Completion, Binocular and Multiple-View Stereo and Machine Learning using Tensor Voting*. PhD thesis, University of Southern California, Los Angeles, California, 2005. 4, 38

[Mord 06] P. Mordohai and G. Medioni. "Tensor voting: a perceptual organization approach to computer vision and machine learning". *Synthesis Lectures on Image, Video, and Multimedia Processing*, Vol. 2, No. 1, pp. 1–136, 2006. 30

[Moun 09] P. Mountney and G.-Z. Yang. "Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping". In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 1184–1187, IEEE, 2009. 26, 87

[Muel 14] M. Mueller, M. Yigitsoy, H. Heibel, and N. Navab. "Deformable Reconstruction of Histology Sections using Structural Probability Maps". In: *International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Springer, Boston, USA, September 2014. 8, 49

[Neld 65] J. A. Nelder and R. Mead. "A simplex method for function minimization". *The computer journal*, Vol. 7, No. 4, pp. 308–313, 1965. 16

[Ni 08] D. Ni, Y. Qu, X. Yang, Y. P. Chui, T.-T. Wong, S. S. Ho, and P. A. Heng. "Volumetric ultrasound panorama based on 3D SIFT". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pp. 52–60, Springer, 2008. 87

[Opel 06] A. Opelt, A. Pinz, and A. Zisserman. "A boundary-fragment-model for object detection". In: *Computer Vision–ECCV 2006*, pp. 575–588, Springer, 2006. 85, 96

[Oye 12] O. K. Øye, W. Wein, D. M. Ulvang, K. Matre, and I. Viola. "Real time image-based tracking of 4d ultrasound data". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 447–454, Springer, 2012. 87, 88

[Ozuy 10] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua. "Fast keypoint recognition using random ferns". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 32, No. 3, pp. 448–461, 2010. 138, 139, 140

[Pala 01] K. Palágyi, E. Balogh, A. Kuba, C. Halmai, B. Erdőhelyi, E. Sorantin, and K. Hausegger. "A sequential 3D thinning algorithm and its medical applications". In: *Information Processing in Medical Imaging*, pp. 409–415, Springer, 2001. 133

[Pare 89] P. Parent and S. W. Zucker. "Trace inference, curvature consistency, and curve detection". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 11, No. 8, pp. 823–839, 1989. 30

[Peti 02] S. Petitjean. "A survey of methods for recovering quadrics in triangle meshes". *ACM Computing Surveys (CSUR)*, Vol. 34, No. 2, pp. 211–262, 2002. 44

[Pitc 12] J. Pitcher, J. Wilson, S. Schwartz, and J. Hubschman. "Robotic Eye Surgery: Past, Present, and Future". *J Comput Sci Syst Biol*, Vol. 3, No. 1, 2012. 137

[Piti 06] A. Pitiot, E. Bardinet, P. M. Thompson, and G. Malandain. "Piecewise affine registration of biological images for volume reconstruction". *Medical image analysis*, Vol. 10, No. 3, pp. 465–483, 2006. 14

[Powe 64] M. J. Powell. "An efficient method for finding the minimum of a function of several variables without calculating derivatives". *The computer journal*, Vol. 7, No. 2, pp. 155–162, 1964. 16

[Prei 09] S. Preibisch, S. Saalfeld, and P. Tomancak. "Globally optimal stitching of tiled 3D microscopic image acquisitions.". *Bioinformatics*, Vol. 25, No. 11, pp. 1463–1465, 2009. 26, 30, 86

[Rank 05] V. Rankov, R. J. Locke, R. J. Edens, P. R. Barber, and B. Vojnovic. "An algorithm for image stitching and blending". In: *Biomedical Optics 2005*, pp. 190–199, International Society for Optics and Photonics, 2005. 86

[Rank 93] R. Rankin, A. Fenster, D. Downey, P. Munk, M. Levin, and A. Vellet. "Three-dimensional sonographic reconstruction: techniques and diagnostic applications". *American Journal of Roentgenology*, Vol. 161, No. 4, p. 695, 1993. 10

[Reef 06] M. Reeff, F. Gerhard, P. C. Cattin, and G. Székely. "Mosaicing of endoscopic placenta images". *GI Jahrestagung (1)*, Vol. 2006, pp. 467–474, 2006. 87

[Rich 12] R. Richa, B. Vágvölgyi, M. Balicki, G. Hager, and R. H. Taylor. "Hybrid tracking and mosaicking for information augmentation in retinal surgery". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 397–404, Springer, 2012. 137, 143

[Riem 10] H. Riemenschneider, M. Donoser, and H. Bischof. "Using partial edge contour matches for efficient object category localization". In: *Computer Vision–ECCV 2010*, pp. 29–42, Springer, 2010. 85

[Riss 08] L. Risser, F. Plouraboué, and X. Descombes. "Gap filling of 3-D microvascular networks by tensor voting". *IEEE Transactions on Medical Imaging*, Vol. 27, No. 5, pp. 674–687, 2008. 30

[Rohl 04] T. Rohlfing, C. Maurer Jr, W. O'Dell, and J. Zhong. "Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images". *Medical Physics*, Vol. 31, p. 427, 2004. 127

[Rose 69] A. Rosenfeld. "Picture processing by computer". *ACM Computing Surveys (CSUR)*, Vol. 1, No. 3, pp. 147–176, 1969. 89, 91

[Ruec 99] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. "Nonrigid registration using free-form deformations: application to breast MR images". *Medical Imaging, IEEE Transactions on*, Vol. 18, No. 8, pp. 712–721, 1999. 14, 56

[Saal 10] S. Saalfeld, A. Cardona, V. Hartenstein, and P. Tomančák. "As-rigid-as-possible mosaicking and serial section registration of large ssTEM datasets". *Bioinformatics*, Vol. 26, No. 12, pp. i57–i63, 2010. 86, 87

[Saal 12] S. Saalfeld, R. Fetter, A. Cardona, and P. Tomancak. "Elastic volume reconstruction from series of ultra-thin microscopy sections". *Nature Methods*, Vol. 9, No. 7, pp. 717–720, 2012. 26, 27, 29, 54

[Sark 93] S. Sarkar and K. L. Boyer. "Integration, inference, and management of spatial information using Bayesian networks: perceptual organization". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 3, pp. 256–274, 1993. 30

[Schn 12] R. J. Schneider, D. P. Perrin, N. V. Vasilyev, G. R. Marx, P. J. del Nido, and R. D. Howe. "Real-time image-based rigid registration of three-dimensional ultrasound". *Medical image analysis*, Vol. 16, No. 2, pp. 402–414, 2012. 87

[Schu 09] T. Schultz, J. Weickert, and H.-P. Seidel. "A Higher-Order Structure Tensor". *Visualization and Processing of Tensor Fields: Advances and Perspectives*, p. 263, 2009. 40

[Sesh 06] S. Seshamani, W. Lau, and G. Hager. "Real-time endoscopic mosaicking". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 355–363, Springer, 2006. 87

[Sham 07] R. Shams, N. Barnes, and R. Hartley. "Image registration in Hough space using gradient of images". In: *Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on*, pp. 226–232, IEEE, 2007. 96

[Shan 48] C. E. Shannon. "A Mathematical Theory of Communication". *Bell System Technical Journal*, Vol. 27, No. 3, pp. 379–423, 1948. 97

[Shan 49] C. Shannon. "Communication in the presence of noise". *Proceedings of the IRE*, Vol. 37, No. 1, pp. 10–21, 1949. 124

[Shas 88] A. Shashua and S. Ullman. "Structural Saliency: The Detection Of Globally Salient Structures using A Locally Connected Network". In: *Proc. Second Int Computer Vision. Conf*, pp. 321–327, Massachusetts Institute of Technology, 1988. 30

[Shei 07] Y. A. Sheikh, E. A. Khan, and T. Kanade. "Mode-seeking by medoidshifts". In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pp. 1–8, IEEE, 2007. 96, 113

[Shot 05] J. Shotton, A. Blake, and R. Cipolla. "Contour-based learning for object detection". In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, pp. 503–510, IEEE, 2005. 85

[Shot 08] J. Shotton, A. Blake, and R. Cipolla. "Multiscale categorical object recognition using contour fragments". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 30, No. 7, pp. 1270–1281, 2008. 85, 96

[Shum 00] H.-Y. Shum and R. Szeliski. "Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment". *International Journal of Computer Vision*, Vol. 36, No. 2, pp. 101–130, 2000. 86

[Sieb 08] M. von Siebenthal. *Analysis and Modelling of Respiratory Liver Motion using 4DMRI*. PhD thesis, Eidgenössische Technische Hochschule ETH Zürich, 2008. 10

[Sing 07] M. Singh and J. M. Fulvio. "Bayesian contour extrapolation: geometric determinants of good continuation.". *Vision Res*, Vol. 47, No. 6, pp. 783–798, Mar 2007. 30

[Sing 14] H. Singh. "ZD Solutions". `http://www.zdsolutions.it`, 2014. info[at]zdsolutions.it, [Online; accessed 1-October-2014]. 88

[Soti 13] A. Sotiras, C. Davatzikos, and N. Paragios. "Deformable medical image registration: A survey". *Medical Imaging, IEEE Transactions on*, Vol. 32, No. 7, pp. 1153–1190, 2013. 10, 12, 14

[Star 07] M. Staring, S. Klein, and J. P. Pluim. "A rigidity penalty term for nonrigid registration". *Medical physics*, Vol. 34, No. 11, pp. 4098–4108, 2007. 20, 21

[Stew 03] C. V. Stewart, C.-L. Tsai, and B. Roysam. "The dual-bootstrap iterative closest point algorithm with application to retinal image registration". *Medical Imaging, IEEE Transactions on*, Vol. 22, No. 11, pp. 1379–1394, 2003. 138

[Stre 97] J. Streicher, W. J. Weninger, and G. B. Mueller. "External marker-based automatic congruencing: a new method of 3D reconstruction from serial sections". *The Anatomical Record*, Vol. 248, No. 4, pp. 583–602, 1997. 29, 53

[Stud 99] C. Studholme, D. L. Hill, and D. J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment". *Pattern recognition*, Vol. 32, No. 1, pp. 71–86, 1999. 16

[Szel 06] R. Szeliski. "Image alignment and stitching: A tutorial". *Foundations and Trends® in Computer Graphics and Vision*, Vol. 2, No. 1, pp. 1–104, 2006. 86

[Szel 96] R. Szeliski. "Video mosaics for virtual environments". *Computer Graphics and Applications, IEEE*, Vol. 16, No. 2, pp. 22–30, 1996. 86

[Szni 12] R. Sznitman, K. Ali, R. Richa, R. H. Taylor, G. D. Hager, and P. Fua. "Data-Driven visual tracking in retinal microsurgery". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 568–575, Springer, 2012. 137

[Tall 08] N. J. Talley. "Functional gastrointestinal disorders as a public health problem.". *Neurogastroenterology & Motility*, Vol. 20 Suppl 1, pp. 121–129, May 2008. 123, 130

[Tang 02] C.-K. Tang and G. Medioni. "Curvature-augmented tensor voting for shape inference from noisy 3d data". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 24, No. 6, pp. 858–864, 2002. 44, 45

[Taub 95] G. Taubin. "Estimating the tensor of curvature of a surface from a polyhedral approximation". In: *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pp. 902–907, IEEE, 1995. 43, 44, 45

[Thor 95] K. K. Thornber and L. R. Williams. "Analytic solution of stochastic completion fields". In: *Proc. Symp. Int Computer Vision*, pp. 617–622, Springer, 1995. 30

[Toew 13] M. Toews and W. M. Wells III. "Efficient and robust model-to-image alignment using 3D scale-invariant features". *Medical image analysis*, Vol. 17, No. 3, pp. 271–282, 2013. 113

[Tong 04] W. Tong, C. Tang, P. Mordohai, and G. Medioni. "First order augmentation to tensor voting for boundary inference and multiscale analysis in 3d". *IEEE Trans. PAMI*, Vol. 26, No. 5, pp. 594–611, 2004. 30

[Tong 05] W. Tong and C. Tang. "Robust estimation of adaptive tensors of curvature by tensor voting". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 27, No. 3, pp. 434–449, 2005. 44, 45, 75

[Totz 12] J. Totz, K. Fujii, P. Mountney, and G.-Z. Yang. "Enhanced visualisation for minimally invasive surgery.". *Int J Comput Assist Radiol Surg*, Vol. 7, No. 3, pp. 423–432, May 2012. 26

[Trig 00] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. "Bundle adjustment—a modern synthesis". In: *Vision algorithms: theory and practice*, pp. 298–372, Springer, 2000. 86

[Tsai 11] C.-L. Tsai, J. Lister, C. Bjornsson, K. Smith, W. Shain, C. Barnes, and B. Roysam. "Robust, globally consistent and fully automatic multi-image registration and montage synthesis for 3-D multi-channel images.". *J Microsc*, Vol. 243, No. 2, pp. 154–171, 2011. 26, 27, 28, 30, 86

[Tsam 10] E. Tsamoura and I. Pitas. "Automatic Color Based Reassembly of Fragmented Images and Paintings". *IEEE Transactions on Image Processing*, Vol. 19, No. 3, pp. 680–690, 2010. 28

[Varn 13] A. Varnavas, T. Carrell, and G. Penney. "Fully automated initialisation of 2D-3D image registration". In: *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, pp. 568–571, IEEE, 2013. 96

[Veda 08] A. Vedaldi and S. Soatto. "Quick shift and kernel methods for mode seeking". In: *Computer Vision–ECCV 2008*, pp. 705–718, Springer, 2008. 96, 113

[Verc 06] T. Vercauteren, A. Perchant, G. Malandain, X. Pennec, and N. Ayache. "Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy". *Medical Image Analysis*, Vol. 10, No. 5, pp. 673–692, 2006. 26, 86

[Verc 09] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. "Diffeomorphic demons: Efficient non-parametric image registration". *NeuroImage*, Vol. 45, No. 1, pp. S61–S72, 2009. 10

[Viol 04] P. Viola and M. J. Jones. "Robust real-time face detection". *International journal of computer vision*, Vol. 57, No. 2, pp. 137–154, 2004. 83

[Viol 97] P. Viola and W. M. Wells III. "Alignment by maximization of mutual information". *International journal of computer vision*, Vol. 24, No. 2, pp. 137–154, 1997. 16, 99

[Wach 07] C. Wachinger, W. Wein, and N. Navab. "Three-dimensional ultrasound mosaicing". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pp. 327–335, Springer, 2007. 26, 87

[Wach 08] C. Wachinger, B. Glocker, J. Zeltner, N. Paragios, N. Komodakis, M. S. Hansen, and N. Navab. "Deformable mosaicing for whole-body MRI". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pp. 113–121, Springer, 2008. 26, 54, 55, 72, 74

[Warr 12] A. Warren, P. Mountney, D. Noonan, and G.-Z. Yang. "Horizon stabilized–dynamic view expansion for robotic assisted surgery (HS-DVE).". *Int J Comput Assist Radiol Surg*, Vol. 7, No. 2, pp. 281–288, Mar 2012. 26

[Wein 06] W. Wein, F. Pache, B. Röper, and N. Navab. "Backward-warping ultrasound reconstruction for improving diagnostic value and registration". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 750–757, Springer, 2006. 87

[Wein 09] W. Wein, E. Camus, M. John, M. Diallo, C. Duong, A. Al-Ahmad, R. Fahrig, A. Khamene, and C. Xu. "Towards guidance of electrophysiological procedures with real-time 3D intracardiac echocardiography fusion to C-arm CT". In: *Medical Image Computing and Computer-assisted Intervention–MICCAI 2009*, pp. 9–16, Springer, 2009. 88

[Wert 38] M. Wertheimer. "Laws of organization in perceptual forms". *A source book of Gestalt psychology*, pp. 71–88, 1938. 30, 34

[Wood 14] O. J. Woodford, M.-T. Pham, A. Maki, F. Perbet, and B. Stenger. "Demisting the Hough transform for 3D shape recognition and registration". *International Journal of Computer Vision*, Vol. 106, No. 3, pp. 332–341, 2014. 96

[Yigi 12a] M. Yigitsoy, S. Kirchhoff, M. F. Reiser, and N. Navab. "Out-of-plane motion compensation in cine-MRI". In: *Abdominal Imaging. Computational and Clinical Applications*, pp. 107–115, Springer, 2012. 123

[Yigi 12b] M. Yigitsoy, M. F. Reiser, N. Navab, and S. Kirchhoff. "Dynamic graph cuts for colon segmentation in functional cine-MRI". In: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pp. 1268–1271, IEEE, 2012. 130

[Yigi 13] M. Yigitsoy and N. Navab. "Structure Propagation for Image Registration". *Medical Imaging, IEEE Transactions on*, Vol. 32, No. 9, pp. 1657–1670, Sept 2013. 8, 49, 87, 89

[Yigi 14] M. Yigitsoy, J. Fotouhi, and N. Navab. "Hough Space Parametrization: Ensuring Global Consistency in Intensity-based Registration". In: *International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Springer, Boston, USA, September 2014. 8

[Yigi 15] M. Yigitsoy, A. Katouzian, and N. Navab. "Structure propagation for deformable image stitching". In: *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pp. 1081–1084, IEEE, 2015. 49

[Ziki 10] D. Zikic, B. Glocker, O. Kutter, M. Groher, N. Komodakis, A. Kamen, N. Paragios, and N. Navab. "Linear intensity-based image registration by Markov random fields and discrete optimization.". *Med Image Anal*, Vol. 14, No. 4, pp. 550–562, Aug 2010. 17, 19, 32, 50, 51, 58

[Zito 03] B. Zitová and J. Flusser. "Image registration methods: a survey". *Image and Vision Computing*, Vol. 21, pp. 977–1000, 2003. 9, 10