Medical Image Segmentation using Weak Priors

Author(s):
Gass, Tobias

Publication Date:
2014

Permanent Link:
https://doi.org/10.3929/ethz-a-010340283

Rights / License:
In Copyright - Non-Commercial Use Permitted
Medical Image Segmentation using Weak Priors

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Tobias Gass
Dipl.-Inform., RWTH Aachen University
born on 25.07.1979
citizen of Germany

accepted on the recommendation of

Prof. Dr. Gábor Székely, examiner
Prof. Dr. Hans-Peter Meinzer, co-examiner
Prof. Dr. Orçun Göksel, co-examiner

2014
Abstract

Image segmentation is a core technique in medical image analysis. By enabling semantic interpretability of raw image data, it allows both medical experts and automatic processing software to diagnose and to plan for the treatment of pathologies with greater accuracy and efficiency. An additional use-case for segmentations is to learn models of anatomical variability, for example statistical shape models, which are an important tool in clinical workflows. Such models have been used to recognize diseases and injuries, to drive reconstructive surgery and patient-specific implant design, and in robust automatic segmentation of anatomical structures.

One critical requirement for statistical shape models or similarly powerful segmentation methods is their dependence on a large database of segmented images, which contains strong prior knowledge about the organ and/or its shape. Such database then cannot be built without human assistance, since most automatic segmentation methods require strong prior knowledge as said before. In order to break this vicious circle of mutual dependency, it is the goal of this thesis to investigate and develop reliable and accurate automatic segmentation methods which require only weak prior information such as a single annotated image, or a database of unsegmented images to learn general prior knowledge. The following two main approaches are explored in this thesis.

First, a novel technique has been developed which allows for the simultaneous registration of a single segmented image (atlas) to an unsegmented image, while also segmenting the latter based on an intensity model. This approach enables the segmentation goal function to be included into the registration criterion, while using the atlas as a weak shape prior.

The second approach utilizes the information contained in a set of unsegmented images to improve atlas-based segmentation. Motivated by semi-supervised techniques in machine learning, three different algorithms have been developed. The first proposed technique utilizes unsegmented images along with their automatic, atlas based segmentations as weak atlases, which can then be used to generate additional segmentation hypotheses for each unsegmented image. Such segmentation hypotheses are then fused analogously to multi-atlas segmentation methods. This was shown to greatly increase the accuracy and reliability of the atlas-based segmentation. Similarly, another technique has been proposed to fuse registration hypotheses, which can be obtained either by running multiple registration algorithms/parametrizations or by composing registrations along indirect
paths between images. The fusion of such registrations then optimizes a joint similarity and deformation smoothness criterion, and was shown to improve registration fidelity and atlas-based segmentation quality. Lastly, a novel algorithm has been developed which aims at rectifying pair-wise non-rigid registrations by optimizing group-wise registration consistency, which previously has been used in the literature often for estimating the fidelity of registration algorithms.

All methods have been experimentally evaluated on synthetic and clinical images from different modalities (CT and MR) and with different target structures (mandibular bone and Corpus Callosum). The fidelity of the resulting segmentations was demonstrated to be improved consistently compared to other segmentation methods that use weak priors. Crucially, worst-case performance was seen to improve substantially, which facilitates the deployment of automatic segmentation methods in scenarios without human supervision.
Zusammenfassung


Alle Methoden wurden experimentell auf synthetischen und klinischen Daten unterschiedlicher Modalität (CT und MR), und mit unterschiedlichen Zielorganen (Kieferknochen und Corpus Callosum) validiert und evaluiert. Es konnte gezeigt werden dass die Qualität der resultierenden Segmentierungen konsistent besser ist als andere Segmentierungsverfahren die schwache Vorwissen verwenden. Entscheidenderweise konnte die Qualität der schlechtesten Resultate substantiell verbessert werden, was den Einsatz von automatischen Segmentierungsmethoden in Szenarien ohne menschliche Kontrolle vereinfacht.
Acknowledgments

First and foremost, I would like to express my gratitude to Prof. Gábor Székely, who has given me the opportunity to explore the exciting field of medical image analysis as part of the Computer Vision Lab at ETH Zurich. His support and confidence in my work, along with many interesting discussions and valuable feedback, has shaped not only this thesis but also my relationship with this field of research. I would also like to thank Prof. Hans-Peter Meinzer, who has kindly agreed to review this thesis and to serve as committee member of my doctoral defense. Many thanks also to Dr. Remi Blanc, who supervised me during the beginning of my thesis and helped to set the path for this work. Furthermore, I want to express my profound gratitude to Prof. Orçun Göksel for his advice, never-ending vigor, enthusiasm and support in countless productive meetings and collaborative work.

I also owe gratitude to the CO-ME network, which has created a fantastic environment for scientific research and exchange. In addition to funding the research in this thesis as part of the Swiss NCCR CO-ME, it provided invaluable opportunities to network and collaborate.

I would like to thank Thomas Deselaers and Henning Mueller for introducing me to (medical) computer vision and nurturing my interest in research.

Many thanks also to my colleagues at the Computer Vision Lab, who made the challenging task of stumbling towards scientific merit (and/or the PhD degree) enjoyable. In particular, I would like to thank my 'extended office' mates Fabian, Valeria, Henning, Severin, Rasmus and Marko for entertainment, advice, enough coffee and creating an overall fun atmosphere.

I want to thank my parents for being great examples in being curious, open, confident and happy, and motivating and enabling me, among a lot of things, to strive for a PhD degree. I want to thank my brother for all the games we played, sports we learned, fights we fought and adventures we shared. My grandparents for being everything I can imagine wanting them to be, and passing on their love for the mountains, which is one of the reasons I went to Switzerland. Finally, I am incredibly grateful for the chance to explore, enjoy and share life with Charlotte.
Contents

1 Introduction 1
  1.1 Semi-Automatic Segmentation 2
  1.2 Automatic Segmentation 3
    1.2.1 Priors in Automatic Segmentation 3
      1.2.1.1 Appearance Priors 4
      1.2.1.2 Shape Priors 5
      1.2.1.3 Combined Priors 5
    1.2.2 Semi-Supervised Segmentation 6
  1.3 Goals 7
  1.4 Contributions and Organization 8

2 Auxiliary Anatomical Labels for Joint Segmentation and Non-Rigid Atlas Registration 11
  2.1 Introduction 12
  2.2 Method 13
    2.2.1 Joint Registration and Segmentation 14
    2.2.2 Auxiliary Labels 15
      2.2.2.1 Coherence Potentials foe Auxiliary Labels 16
  2.3 Experimental Evaluation 16
    2.3.1 Discussion 18
  2.4 Conclusions 19

3 Simultaneous Segmentation and Multi-Resolution Nonrigid Atlas Registration 21
  3.1 Introduction 22
  3.2 Background and Notation 26
  3.3 Simultaneous Registration and Segmentation 28
  3.4 Implementation using Markov Random Fields 30
    3.4.1 Two-Layer Graph 31
    3.4.2 SRS using a Multi-Resolution Framework 32
    3.4.3 Algorithm 35
      3.4.3.1 Potential Functions 35
    3.4.4 Implementation of ARS 37
3.5 Experimental Evaluation ........................................ 38
3.6 Discussion ......................................................... 45
3.7 Conclusions ....................................................... 49

4 Semi-supervised Segmentation Using Multiple Segmentation Hypotheses from a Single Atlas 51
  4.1 Introduction ..................................................... 51
  4.2 Label Propagation ............................................. 53
  4.3 Compact Graphs .............................................. 54
  4.4 Results and Discussion ..................................... 55
  4.5 Conclusions ................................................... 59

5 Registration Fusion using Markov Random Fields 61
  5.1 Introduction .................................................... 61
  5.2 Registration Fusion using an MRF ............................ 63
     5.2.1 $\text{MRegFuse}$ with Folding Removal ($\text{MRegFuseR}$) ....... 65
     5.2.2 Pair-wise and Population-Based Registration Fusion .......... 66
  5.3 Experimental Results ...................................... 67
     5.3.1 Discussion ............................................. 70
  5.4 Conclusions .................................................. 71

6 Consistency-Based Rectification of Non-Rigid Registrations 73
  6.1 Introduction .................................................... 74
  6.2 Notation and Background .................................. 77
  6.3 Method ......................................................... 78
     6.3.1 CBRR Algorithm: Consistency .......................... 79
     6.3.2 CBRR Algorithm: Prior .................................. 80
     6.3.3 Casting into a System of Linear Equations ............... 81
     6.3.4 Iterative Refinement .................................... 83
     6.3.5 Implementation .......................................... 83
  6.4 Experimental Results ..................................... 84
     6.4.1 Synthetic Images ....................................... 84
     6.4.2 2D MR Images .......................................... 86
     6.4.3 3D CT Segmentation .................................... 89
     6.4.4 4D Liver MRI Motion Estimation ......................... 92
     6.4.5 Discussion ............................................. 92
  6.5 Conclusions .................................................. 96

7 Summary and Perspectives .................................... 97

Bibliography ....................................................... 101
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Data requirements and examples for strong and weak priors</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Vicious circle for robust automatic segmentation</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>Auxiliary-label joint segmentation and registration</td>
<td>13</td>
</tr>
<tr>
<td>2.2</td>
<td>Pipeline for obtaining the auxiliary-labelled atlas segmentation</td>
<td>16</td>
</tr>
<tr>
<td>2.3</td>
<td>Progression of coherence potential</td>
<td>17</td>
</tr>
<tr>
<td>2.4</td>
<td>Example segmentations of the corpus callosum</td>
<td>18</td>
</tr>
<tr>
<td>2.5</td>
<td>Effect of coherence potential weight on single- and aux-label ARS.</td>
<td>19</td>
</tr>
<tr>
<td>3.1</td>
<td>A synthetic example comparing SRS to traditional methods</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Product-label SRS vs. our sum-label SRS</td>
<td>23</td>
</tr>
<tr>
<td>3.3</td>
<td>Bayesian network and factor graph for the SRS MAP problem</td>
<td>29</td>
</tr>
<tr>
<td>3.4</td>
<td>Schematic of a local pixel neighborhood in the SRS MRF defined on the pixel level</td>
<td>31</td>
</tr>
<tr>
<td>3.5</td>
<td>Illustration of the SRS-MRF using different resolutions for registration and segmentation subgraphs</td>
<td>33</td>
</tr>
<tr>
<td>3.6</td>
<td>Progression of the SRS algorithm on a 2D MRI corpus callosum segmentation task</td>
<td>34</td>
</tr>
<tr>
<td>3.7</td>
<td>Minimizing one instance of the SRS energy using ( \alpha )-expansion graph-cuts and TRW-S.</td>
<td>38</td>
</tr>
<tr>
<td>3.8</td>
<td>Deformed segmentation ( S_A(T) ) and segmentation estimate ( S_X ) for single-label SRS and multi-label SRS.</td>
<td>41</td>
</tr>
<tr>
<td>3.9</td>
<td>Sample segmentations using the different approaches</td>
<td>42</td>
</tr>
<tr>
<td>3.10</td>
<td>Distribution of the results for mandibular bone and CC IIo experiments</td>
<td>44</td>
</tr>
<tr>
<td>3.11</td>
<td>Results of all the methods on five different examples</td>
<td>45</td>
</tr>
<tr>
<td>3.12</td>
<td>Facial feature tracking example in a video sequence</td>
<td>46</td>
</tr>
<tr>
<td>4.1</td>
<td>Illustration of standard atlas based segmentation and our proposed method</td>
<td>53</td>
</tr>
<tr>
<td>4.2</td>
<td>Sample results for both datasets</td>
<td>56</td>
</tr>
<tr>
<td>4.3</td>
<td>Segmentation accuracy (Dice) of traditional single atlas based segmentation and our proposed method with LAW fusion for each atlas and target image combination</td>
<td>57</td>
</tr>
<tr>
<td>4.4</td>
<td>Dice metric for support sample selection for a single atlas using random selection, GSW-based, individual ARE, and ARE-G ranking criteria</td>
<td>58</td>
</tr>
</tbody>
</table>
5.1 Schematic of the proposed MRF fusion ............................................. 63
5.2 Unary potentials for MRegFuse ..................................................... 65
5.3 Population-based registration hypotheses generation ....................... 66
5.4 Example fusion result for independently computed registrations ....... 68
5.5 Two registration examples showing the Jacobian determinant values for
MRegFuse and MRegFuseR ............................................................. 69
5.6 Regularization and folding in MRegFuse ....................................... 70

6.1 (a) Group-wise registration to the mean image (center), implicitly imply-
ing complete consistency via the inferred inter-image registrations (dashed).
CBRR utilises the information from the complete set of pair-wise registra-
tions (b). In this set, it aims at reducing the registration inconsistency by
modifying pair-wise transformations such that the differences $\vec{r}$ between
direct and indirect transformations are minimised (c). This is done simulta-
aneously for all pair-wise transformations using all possible registration
triplets. This allows us to exploit the redundant information from multiple
loops for each transformation (d). .................................................. 76

6.2 Inconsistency and target registration error (TRE) for registration loops in
4D liver MRI sequences. Each point indicates the inconsistency of one
loop of three registrations and the average TRE of those three registra-
tions, and the blue line is a least-square fit with $r=0.77$ denoting the Pear-
son correlation. For more details on the data and registration algorithms
please see Sec. 6.4.4. ................................................................. 79

6.3 Local normalized cross correlation (LNCC) weight used for the CBRR
regularizer. Shown are target image, deformed source, and LNCC weight
computed with $\sigma = 4$ mm and $\gamma = 10$ for sample pairs from the MRI
(top) and CT (bottom) datasets. ...................................................... 80

6.4 Relative improvement using CBRR on synthetic data using random ground-
truth deformations of varying complexity, additive random errors with
variable magnitude and frequency, and varying CBRR grid resolution . . . . 86

6.5 Data generation for the 2D MR experiment. First, image $X_1$ is registered
to the remaining images $X_i$, resulting in transformations $T_{1,i}$ (top left).
Then, $X_1$ is deformed using the estimated transformation and 50% SNR
Gaussian noise ($\zeta$) is added, generating new images $X_i'$ (top right). Pair-
wise transformations $T_{i,j}$ between all $X_i'$ are known analytically through
the registrations from $X_1$. The image set $\{X_1, X_2', \ldots, X_N'\}$ is then used
as the experimental dataset and the known dense correspondences $T_{i,j}$ as
the ground-truth in evaluation. ..................................................... 87

6.6 Relative ATE improvement using CBRR with an increasing number of 2D
MR images. Each experiment is repeated 10 times with randomly selected
images. The standard deviation of the results is shown by the error bars. . 88

6.7 Right: Improvements through CBRR over multiple iterations .......... 89
6.8 CBRR of MRF (top) and demons (bottom) registration on the 2D MRI data. Shown are from left to right: target image, deformed source, true registration error, error of rectified registration and the source image deformed by the rectified registration. The hue of the color in the error visualizations indicates its orientation, while the saturation corresponds to the error magnitude.

6.9 Surface distance of ground-truth jawbone segmentation to MRF-registered source (left) and subsequent CBRR (right). Best viewed in color.

6.10 Segmentation overlap percentiles on the 3D CT dataset. Each datapoint shows the maximum Dice coefficient for a given percentage of the total pair-wise registrations.

6.11 Inconsistency and TRE of one loop of sequence 1 before and after using CBRR for the PTV registration method. Mid-sagittal slices are shown for MRIs, inconsistency, and TRE extrapolated using thin-plate splines. CBRR was computed on a longer sequence using 11 images and 110 registrations in total.

6.12 TRE before and after CBRR on the 4D Liver MRI sequences. Each point shows the error for a single landmark before (x-axis) and after (y-axis) CBRR. Note that points blow the y = x line indicate a reduction in error using CBRR.

6.13 Mid-coronal slices of the mean images of the 3D CT dataset for (a) group-wise and (b) CBRR registration. For CBRR, one image was chosen as reference to which all images were aligned for averaging. We deformed the group-wise mean to the same image to allow a better comparison.

List of Tables

2.1 Segmentation results for joint segmentation and registration

3.1 Evaluation datasets

3.2 Parameters for the different experiments

3.3 Average segmentation quality metrics in the 3D CT mandibular bone segmentation task

3.4 Average segmentation quality metrics in the 2D MR corpus callosum segmentation task

3.5 Landmark registration error (mm) on the CC
4.1 Mean segmentation accuracy from leave-one-out evaluation on the MR (2D) and CT (3D) datasets.

5.1 Results using independently computed registrations and their fusion

5.2 Fusing registrations obtained by automatic hypothesis generation in sets of mutually registered images

6.1 Left: Average transformation error (ATE), inconsistency $C$ and pearson correlation $\rho$ of estimated and true registration error magnitude for the 2D MRI experiments using simulated transformations before and after using CBRR. We also compare CBRR to locally averaged weighting (LWA), an earlier version of this work (CLERC) presented in [Gass et al., 2014c], and AQUIRC [Datteri and Dawant, 2012], which computes a dimensionless estimate of registration error which can be correlated with the true error.

6.2 Quantitative evaluation of CBRR and comparison to group-wise registration on the 3D CT dataset. Mean TRE, mean surface distance (MSD), Hausdorff distance (HD) and inconsistency $C$ are given in mm.

6.3 Quantitative evaluation of CBRR on the 4D liver MRI sequences. The average run time is the average time required to process one of the sequences with 11 breathing phases of 3D MR.
Introduction

Imaging has become an invaluable tool in medical workflow by enabling non-invasive mapping of subject anatomies. In order to maximize the utility of image data, it is often imperative to semantically interpret the images, for example by delineating organs. Typical imaging technologies only allow for the extraction of low-level semantic information, e.g. the local radiation absorption in computer tomography (CT). While some imaging responses can be clearly identified (e.g. air in CT), the appearance of organs is often only unique with respect to the tissue which leads to ambiguities when discriminating between organs of a similar tissue type, e.g. bones or muscles. An additional challenge is that most imaging technologies require a trade-off between image quality versus imaging time (e.g. in MR) and radiation exposure (e.g. in CT). Consequently, low resolution and low signal-to-noise-ratio further impair the interpretability of medical images.

In order to facilitate semantic interpretation of medical images, one of the most important techniques is the segmentation of such images. Here, a segmentation is a partition of an image into (typically disjunct) regions. Such partitions can then be labeled according to their anatomical or functional characteristics, thereby enabling a semantic interpretation of the data. This then opens up a multitude of possibilities of further processing and analysis in clinical scenarios. For diagnosis, a segmentation can for example be used to measure brain atrophy which is an indicator for Alzheimer’s disease [Seahill et al., 2002]. Segmentations also enable many treatment procedures, where one of the most widely deployed techniques is organ delineation for radiotherapy (RT) planning [Pekar et al., 2004]. While this is done mostly manually by an expert, a visualization of the surrounding organs and the tumor can help in guiding the planning to avoid tissue damage [Bonduau et al., 2005]. A rich database of segmented organs can also be used for further automatic analysis of appearance and morphology. For instance, automatically trained appearance models can then be used to detect tumors or aid automatic segmentation methods. Furthermore, statistical models of organ shapes can be inferred, leading to a wide range of possible applications: Given a pathological case (for example due to trauma or malformation), the most similar normal organ can be computed and used for reconstructive surgery or implant design [Zachow et al., 2005]. It also allows for the estimation of 3D shapes from 2D images [Kurazume et al., 2009], which can reduce the radiation exposure of
patients. Furthermore, statistical shape models are widely used in automatic segmentation methods [Heimann and Meinzer, 2009] due to their adherence to anatomical prior knowledge.

1.1 Semi-Automatic Segmentation

Due to the critical nature of medical procedures for the well-being of patients, all steps in such workflows should be as accurate as possible. For the segmentation of medical images, this has traditionally led to a high dependency on human experts, who are trained well and have the necessary prior knowledge on both the anatomical and the imaging side. A critical aspect of such manual segmentation is the availability and cost of time spent by an experienced clinician to annotate images. In order to alleviate this burden, several semi-automatic techniques have been developed to assist an expert in performing this task. Aside from directly drawing a segmentation on the image or directly manipulating a template surface [McInerney and Terzopoulos, 1999], one of the most common methods in this context is region growing [Hojjatoleslami and Kittler, 1998], where a user sets an adaptive threshold that is used to label connected regions. A refinement then typically facilitates connecting multiple such regions allowing for labeling anatomies with inhomogeneous intensities. An advanced method is grab-cut [Rother et al., 2004], where a user labels distinctive foreground and background pixels. These labels are then used to train generative appearance models for foreground and background, and a Markov random field (MRF) is used to find a spatially coherent segmentation. This segmentation is then used as new initial labels and the process is repeated until convergence.

The cost associated with human-expert based segmentation methods limits the availability of advanced medical procedures to a broader range of patients [Olabarriaga and Smeulders, 2001]. Additionally, such segmentations are hardly ever objective, and significant variability in segmentation results has been found in both intra- and inter-rater comparison studies [Pizer et al., 1990, Fontana et al., 1993, Udupa, 1997, Yamamoto et al., 1999]. Furthermore, clinical segmentations are often performed with a specific goal or technology in mind, limiting their generalization capabilities. For instance, a segmentation outline meant to be used for radiotherapy (RT) targeting can be adjusted based on the prior knowledge of the clinician with regards to characteristic of the RT beam and the known/assumed locations of cancer, e.g. such that margins are wide around such critical segmentation borders. This means that such label might be originally intended to delineate a specific organ, but is altered in practice with the intended goal of RT in mind. This limits its generalizability, prohibiting the usage in different scenarios as for example learning of organ appearance. A further complication in human-assisted segmentation is that current computer systems mostly visualize 3D anatomies as 2D renderings. This leads to a tedious and error prone process when relying on human feedback, where it is for example still very common that segmentations are performed on a slice-by-slice basis.
1.2 Automatic Segmentation

The aforementioned challenges of manual or assisted segmentation workflows have led to a strong interest in fully automated segmentation methods. However, the main challenge is that for most medical applications, a segmentation algorithm needs to be highly dependable and accurate to avoid negative interference with clinical procedures. To this end, many algorithms attempt to mimic the strength of human experts by utilizing prior knowledge about the anatomy that is to be segmented.

1.2.1 Priors in Automatic Segmentation

There are several different ways to categorize prior knowledge. A typical medical expert has (or should have) seen many instances of a certain organ over the course of his training.
in at least one imaging modality, and potentially also observed it directly during surgery or as anatomical or histological specimen. This enables such expert to form an expectation of the typical shape and appearance of the organ, along with an expectation of the variability of these parameters. Analogously, algorithms can learn prior knowledge from annotated data automatically. For the purpose of this thesis, prior knowledge which is similar to the knowledge of an expert will be denominated as a strong prior, as it is inferred from a population of organ-specific samples. In contrast, a weak prior is then either organ-specific or population-based. This means that it can for instance be inferred from a single instance of the organ (organ-specific) or from a population of images without annotation. Fig. 1.1 visualizes the above-mentioned prior concept and gives examples of segmentation methods which use such priors.

It is further possible to differentiate priors by their type, where appearance and shape are distinctive characteristics, which will be explained in the following sections.

1.2.1.1 Appearance Priors

The term “appearance” is most commonly used in computer vision to describe features of the intensity distribution of an image. Herein, it will be used to denote the features that support the intended segmentation of an organ or anatomical structure. A typical weak appearance prior is for example an intensity value used for thresholding [Fu and Mui, 1981]. A notable application for this is the segmentation of cortical bone in CT, where such bone typically has the highest X-ray absorption rate and therefore appears as the brightest structure in images. Note that such prior is then unspecific with regard to bone (as prosthesis and teeth may appear even brighter) or regard to a particular bone, and thus can be considered weak in the proposed categorization.

More sophisticated appearance models for example based on support vector machines [Cortes et al., 1995, Ricci and Perfetti, 2007] or random forests [Breiman, 2001, Lempitsky et al., 2009] are typically learned from a collection of pre-annotated images and can therefore be categorized as strong priors. Due to the amount and variety of information captured, e.g. texture or local shape, such classifiers can then be targeted more specifically at individual organs. One of the crucial drawbacks of such methods is that they typically treat each pixel in the image independent of all others. This local approach is therefore highly susceptible to noise or inhomogeneous appearance. These challenges can be alleviated by incorporating spatial context in the segmentation method. Here, one general assumption is that appearance should be homogeneous within image regions, and inhomogeneous between them [Fu and Mui, 1981]. Such spatial context can again be general, anatomy-unspecific and therefore weak, as for example, employed in MRF-segmentation [Boykov and Jolly, 2000] in which a first-order Markov assumption encourages solutions that are pair-wise constant. Such pair-wise terms can also be adapted to favor boundaries between the target object and background, where it is then common to learn such terms from populations of images in conditional random field (CRF) approaches [Lafferty et al., 2001, Lee
1.2. Automatic Segmentation

et al., 2005], which is thus considered here to use a strong prior. Alternative methods learn appearance with spatial context using random forests [Criminisi et al., 2011].

1.2.1.2 Shape Priors

Shape prior knowledge can also be included in the form of example shapes or templates. One of the most popular techniques in this framework is atlas-based segmentation [Miller et al., 1993, Collins et al., 1995, Rohlfing et al., 2004]. Here, an image with a gold-standard segmentation (the atlas) is registered to an image which is to be segmented. The registration applied to the gold-standard segmentation of the atlas then yields a segmentation of the target image. Such registration typically does not use any information about the appearance of the object to be segmented, but merely considers an overall similarity metric between the deformed atlas and target image as a goal function. General assumptions about valid anatomical deformations are used to make the problem well-posed [Fischer and Modersitzki, 2008], which is again typically not organ-specific and can therefore be considered as a weak prior. Such registration-based segmentation is still prone to poor local minima in the optimization due to poor initialization or atlas-target mismatch. It has therefore become popular in recent years to register multiple atlases to a target image and probabilistically combine the resulting manifold of segmentation hypotheses [Heckemann et al., 2006, Warfield et al., 2004]. This method then requires multiple (manually) annotated atlas images, and is therefore considered to use strong priors in the proposed categorization. Similarly, multiple segmented images can be used to generate a statistical atlas which is typically the mean of the registered images and should lie in the center of the distribution [Guimond et al., 2000].

A different form of a weak shape prior can be that of a deformable model. Such models can be parametric [Kass and Witkin, 1988] or non-parametric [Cremers et al., 2007], and are typically characterized by general topological assumptions, e.g. desired maximal curvature, and a general intensity based matching criterion which favors the adherence of the shape surface to the edges in the image. An extension of such weak priors can then be the knowledge learned about the statistical variation of the shape. The resulting statistical shape models (SSM) [Cootes et al., 1992] are then considered as a strong prior for segmentation and can help achieving robust organ-specific segmentations [Heimann and Meinzer, 2009], for example using active contours [Cootes and Taylor, 1992].

1.2.1.3 Combined Priors

A natural way to improve the segmentation performance of such methods as presented previously is to combine multiple priors. A typical approach combines appearance priors with shape information, where the earliest approaches used fixed shape templates, e.g. elliptical or star-shaped priors [Slabaugh and Unal, 2005, Veksler, 2008] to drive a
MRF-based segmentation. In order to enable improved organ-specific segmentations, such general templates were replaced by shape models derived from aligned training shapes [Leventon et al., 2000, Rousson and Paragios, 2002, Ali et al., 2007]. In these approaches, the shape model is aligned to an image to be segmented which allows for penalizing unlikely shapes. To overcome the often limited generalizability of SSMs alone, they are often used as regularizers, for example, by spatially constraining a graph-cut based segmentation [Majeed et al., 2012]. In that work, a shape model can also be re-fitted using information from graph-cut segmentation and the process can be iterated until convergence. Alternatively, it is also possible to learn the statistical spatial variation of the appearance in combination with the variability of the shape. Such active appearance models (AAM, [Cootes et al., 2001]) combine a strong appearance model learned from the training population with a strong shape prior in the form of a statistical shape model.

1.2.2 Semi-Supervised Segmentation

Besides such priors learned from annotated images as detailed in the previous sections, semi-supervised learning is a thriving research domain, which originated in the machine learning community [Chapelle et al., 2006]. Its underlying idea is to utilize the information that is contained in unlabeled samples for augmenting any other given prior knowledge.

In medical imaging, such information about shape and appearance can be obtained by analyzing groups of images and introduce high-level assumptions about the relations of the images in this group. These assumptions are often general, and not anatomy-specific and can therefore be categorized as weak priors. For example, it can be assumed that an anatomical structure to be segmented should have similar appearance across multiple images. This assumption can then be used to co-segment two or more images by training foreground/background appearance models on both images jointly which is expected to increase robustness [Rother et al., 2006, Han et al., 2011]. Since intensities alone are rarely a sufficient information cue for medical image segmentation, it is often required that the images are brought into correspondence such that corresponding voxels then should have the same segmentation label across all images. Two main approaches have been developed to achieve this goal: Joint segmentation and registration, pioneered by [Wyatt and Noble, 2003], aims at improving both registration and segmentation by modeling them as a joint process. This work has inspired a large body of literature, which extends the originally rigid, MRF-based technique to different non-rigid registration methods [Mahapatra and Sun, 2012], variational segmentation methods [Ghosh et al., 2010], or additional bias field correction [Pohl et al., 2005]. While the original approach of [Wyatt and Noble, 2003] was formulated for any number of images, most of the joint segmentation and registration approaches focus on pairs of images.
1.3 Goals

As can be seen from the literature overview of the previous sections, many of the current state-of-the-art segmentation methods use strong priors of shape and appearance to achieve the goal of anatomically correct and accurate segmentations. However, the segmentations that are required to generate such strong priors can then evidently not be obtained automatically, as most current automatic segmentation algorithms require strong priors in the first place, creating a vicious circle as shown in Fig. 1.2. Furthermore, it
was frequently observed that strong priors do not generalize well over different populations [Blezek and Miller, 2007, Blanc et al., 2009], thus furthering the need for large, diverse bodies of segmented data from which to generate population-specific priors. It is the goal of this thesis to develop efficient automatic methods for segmenting groups of images, such that the resulting data can then be used to break such vicious circle, i.e. as a set of annotations to build strong priors from. The developed segmentation methods will thus not use any form of strong priors, since such information for most anatomies do not exist or require large amounts of time to obtain by means of manual annotation.

1.4 Contributions and Organization

To achieve the aforementioned goals, the approaches investigated in this thesis focus on atlas-based segmentation, where such atlas is to be registered non-rigidly to an unsegmented image with the purpose of transferring the atlas segmentation onto the target. This choice is motivated by the fact that such atlas-based segmentation can achieve reasonable segmentation quality even with a limited, weak prior in the form of only a single atlas image with segmentation annotation. It is, however, well-known that image registration is tremendously difficult. Even if images are noise-free, and correspondences between images exist, the widely used intensity based registration goal functions are only surrogate measures of registration quality, which leads to an ill-posed problem. The inclusion of spatial regularization then results in an NP-complete elastic matching problem [Keysers and Unger, 2003], which makes it impossible to guarantee globally optimal solutions using typical polynomial time registration algorithms. This thesis therefore aims at developing algorithms to alleviate these challenges in non-rigid registration. Two main directions are then investigated:

1. As far as segmentation is concerned, the non-rigid registration is only a means to an end, with the goal function only indirectly related to the actual segmentation goal. It is thus investigated to which extent the inclusion of a segmentation criterion in the pair-wise registration goal function can benefit both the registration and the segmentation targets.

2. This thesis also investigates means to explicitly utilize the information contained in the group of un-segmented images. In the spirit of semi-supervised learning, methods that can utilize such information to improve the segmentation of all images in a set using only weak priors are developed.

Specifically, the contributions presented in the thesis, and subsequently the structure of this manuscript, are as follows:
In the next two chapters, a novel method for simultaneous segmentation and non-rigid atlas registration using discrete optimization is developed. Chapter 2 focuses on the problem of weak appearance priors for segmentation, which can lead to ill-posedness of joint registration and segmentation (JRS). An algorithm is proposed that iterates between registering the atlas and segmenting the target image, while penalizing the difference between target segmentation and deformed atlas segmentation using a novel distance-based penalty in a discrete optimization framework. In order to deal with the ambiguity of the weak intensity prior, which is learned from the atlas image and segmentation, the JRS model is augmented by an auxiliary segmentation label which corresponds to tissue with the same appearance as the target anatomy, e.g. other bone when a specific bone should be segmented. Chapter 3 then utilizes these techniques, but further develops a truly simultaneous optimization of segmentation and registration in contrast to the alternating solution of the previous chapter. A fundamental challenge of such simultaneous optimization is the different levels of detail in which segmentation and registration are commonly solved, with segmentation being estimated at pixel-level and registration being commonly estimated based on a coarse grid of control points. To overcome this, a novel two-layer Markov Random-Field (MRF) algorithm is proposed, where the structure is justified by a rigorous probabilistic analysis of the underlying Bayesian optimization problem.

Chapter 4 then investigates a semi-supervised segmentation strategy where each unsegmented image is first segmented by regular atlas-based segmentation, and then utilized as a weak atlas which can be used to generate additional segmentation hypotheses for the remaining unsegmented images. This results in several segmentation estimates for each image, which are fused analogously to multi-atlas segmentation. In the straight-forward implementation, a quadratic number of registrations is required, registering each image to all remaining images. To overcome this computational burden, a support-sample selection strategy is devised which enables the automatic selection of a limited number of unsegmented images as weak atlases in linear time, thus reducing the number of required additional registrations.

Chapter 5 extends this concept to registrations, in contrast to segmentations as in the previous chapter. Here, compositions of registrations along indirect paths between two images are used to generate additional registration hypotheses. Since no sound method to fuse such multiple registrations existed yet, a novel registration fusion method is developed which allows for the joint optimization of an image similarity criterion and deformation smoothness in a discrete optimization framework. This general method can also be used to fuse registrations obtained in other settings as the group-wise scenario, for example, running different registration methods or parametrizations. This can alleviate the selection of the most appropriate registration algorithm or parametrization for new data or application scenarios.

While the registration fusion presented in the previous chapter focuses on registration smoothness and post-registration image similarity, a novel registration post-processing method based on registration consistency is proposed in Chapter 6. Such consistency re-
flects the assumption that correspondences should generally exist not only between image pairs, but also between all images in a set of images. This criterion has been frequently used to quantify the quality of a registration algorithm, but had not yet been used as a goal function to be optimized within groups of images. In order to tackle the challenging optimization task of finding consistent correspondences between all images in a set of images, the problem is formulated as a registration post-processing step and it is cast as a linear least-squares optimization problem, for which efficient algorithms exist. The resulting consistent registrations are shown to improve not only the dense registration accuracy, but also the pair-wise post-registration segmentation overlap between the segmentation estimate and the ground-truth.

The thesis is concluded by a summary in Chapter 7, which briefly compares the proposed techniques and gives an outlook on possible further research.

Chapters 2-5 have been published in peer-reviewed journals and conference proceedings as [Gass et al., 2014a] [Gass et al., 2014e] [Gass et al., 2012] [Gass et al., 2014d]. Chapter 6 is currently under review as [Gass et al., 2014b]. A preliminary version of the work in Chapter 6 has been presented in [Gass et al., 2014c].
This paper studies improving joint segmentation and registration by introducing auxiliary labels for anatomy that has similar appearance to the target anatomy while not being part of that target. Such auxiliary labels help avoid false positive labelling of non-target anatomy by resolving ambiguity. A known registration of a segmented atlas can help identify where a target segmentation should lie. Conversely, segmentations of anatomy in two images can help them be better registered. Joint segmentation and registration is then a method that can leverage information from both registration and segmentation to help one another. It has received increasing attention recently in the literature. Often, merely a single organ of interest is labelled in the atlas. In the presence of other anatomical structures with similar appearance, this leads to ambiguity in intensity based segmentation; for example, when segmenting individual bones in CT images where other bones share the same intensity profile. To alleviate this problem, we introduce automatic generation of additional labels in atlas segmentations, by marking similar-appearance non-target anatomy with an auxiliary label. Information from the auxiliary-labeled atlas segmentation is then incorporated by using a novel coherence potential, which penalizes differences between the deformed atlas segmentation and the target segmentation estimate. We validated this on a joint segmentation-registration approach that iteratively alternates between registering an atlas and segmenting the target image to find a final anatomical segmentation. The results show that automatic auxiliary labelling outperforms the same approach using a single label atlases, for both mandibular bone segmentation in 3D-CT and corpus callosum segmentation in 2D-MRI.
2.1 Introduction

In computer vision, in general, and medical imaging, in particular, the segmentation of an image is a common problem that has been studied extensively in the literature. Possible approaches can be categorized by the amount of prior knowledge they use, such as intensity-based segmentation on the one end of the spectrum and model-based segmentation on the other. Intensity-based segmentation typically involves a local classifier to determine the type of tissue that a pixel most likely belongs to. It can additionally regularize such segmentation, for example using Markov random fields (MRF) [Boykov and Funke-Lea, 2006]. This method is typically fast, but lacks anatomical correctness as neither absolute nor relative spatial locations (e.g., shapes) are taken into account. Due to its weak dependency on manual annotations, many recent studies have focused on registration-based segmentation [Glocker et al., 2008] where a single reference (atlas) image is registered to a target image. The resulting transformation is then applied to the labelled atlas, which yields a segmentation of the target image. While being widely used in medical imaging, image registration alone cannot solve the segmentation task by itself as it is known to be an ill-posed problem [Fischer and Modersitzki, 2008]. This is because anatomical correspondences, which are not guaranteed to exist, are computed using surrogate criteria such as intensity similarity. In addition, the problem of non-rigid registration is known to be NP-complete [Keysers and Unger, 2003], thus making approximative algorithms prone to local optima, for example due to poor initialization. Increasing the amount of prior knowledge is a common remedy for such problems, for example by including multiple or statistical atlases [Rohlfing et al., 2005, Heckemann et al., 2006, Glocker et al., 2007].

In order to alleviate such problems, joint optimization of segmentation and registration was first proposed by [Wyatt and Noble, 2003]. Such methods allow segmentation and registration processes to mutually benefit from one another. On the one hand, knowing the segmentation of both target and atlas images one can greatly improve their registration. On the other hand, knowing the perfect correspondences between such two images will improve, and even solve completely, the (atlas-based) segmentation problem. Accordingly, several approaches for such joint registration and segmentation have been proposed in the literature, focusing on alternating registration and segmentation (ARS) methods.

An ARS approach that alternates between estimating a rigid deformation using Powells’ method and updating the segmentation using iterative conditional modes (ICM) in an MRF was proposed in [Wyatt and Noble, 2003]. Another approach based on MRFs that alternates between solving one MRF to optimize registration parameters and to update segmentation probabilities, and a second MRF to solve the segmentation itself was introduced in [Xiaohua et al., 2004]. A Bayesian framework to alternate between updating the registration and estimating intensity “nuisance” parameters while marginalizing over possible anatomical labels using an EM algorithm was presented in [Pohl et al., 2005]. There have also been studies on alternating registration and segmentation in variational
Formulations. Those also rely on iteratively updating the registration and segmentation solutions, for example, by using gradient descent [Ghosh et al., 2010, Schmidt-Richberg and Ehrhardt, 2012].

In many cases, this process can still be ambiguous, as the segmentation criterion may still find some part of the anatomy that matches the given intensity prior, though not being part of the desired target. This might occur for example in the segmentation of single bones, which unavoidably have similar appearance to other bones. This leads to ambiguity in the joint process, leading to complications such as leakage as illustrated in Fig. 2.1(c). In this paper, we propose the use of automatically-generated auxiliary anatomical labels to alleviate this problem. Such labels are meant to identify voxels with the appearance of the target anatomy in both the atlas and the target images. We show experimentally that such auxiliary labels improve joint registration and segmentation.

### 2.2 Method

Let $X$ be the target and $A$ the atlas image, where both are functions $\Omega \to \mathcal{F}$ from the $D$-dimensional discretized coordinate domain $\Omega \in \mathbb{N}^D$, to image intensities $\mathcal{F}$. A segmentation $S = \{l_0^S, \ldots, l_N^S\}$ is an assignment of one such segmentation label to each pixel of an image. In segmentation, the goal is to find a segmentation $S_X$ of the target image $X$. Without loss of generality, we assume $I^S \in \{l_b, l_f\}$, where $l_b$ labels background and $l_f$ labels the target anatomy. In registration, the goal is to find a transformation $\mathcal{T}$ such that the target image $X$ and the transformed atlas image $A(\mathcal{T})$ are similar according to a selected metric. $\mathcal{T}$ can be a non-rigid deformation represented by a $D$-dimensional displacement vector field $T : \Omega \to \mathbb{R}^D$ so that for each pixel/voxel $i \in \Omega$: $T(i) = i + t_i$, where $T_i$ and $t_i$ are shorthands for the local deformation $T(i)$ and displacement $T(i)$. We will also use this notation to denote pixels in an image, e.g. $X_i, A_i$. Then, $I(\mathcal{T}) = i \to I(i + t_i) \forall i$, indicates a deformed image. Values at non-grid locations are linearly interpolated. Both
finding a displacement vector field $T$ and a segmentation $S_X$ can be defined as separate energy minimization problems as follows:

$$
\hat{S}_X = \arg\min_{S_X} \sum_{i \in \Omega} \left( \psi_i^{\text{seg}}(l^S_i) + \sum_{j \in N(i)} \lambda_{ij} \psi_{ij}^{\text{seg}}(l^S_i, l^S_j) \right), \quad (2.1)
$$

$$
\hat{T} = \arg\min_{T} \sum_{i \in \Omega} \left( \psi_i^{\text{reg}}(l^R_i) + \sum_{j \in N(i)} \lambda_{ij} \psi_{ij}^{\text{reg}}(l^R_i, l^R_j) \right), \quad (2.2)
$$

where $N(i)$ denotes the neighbors of a pixel $i$ and $\psi, \Psi$ are unary and pairwise potential functions, and $\lambda$ are weights. The solutions ensure a smooth labelling of the image and can be found efficiently using graph-based methods such as $\alpha$-expansion graphcuts [Boykov et al., 2001]. We follow standard approaches for both segmentation and registration, using Hounsfield-unit based segmentation potentials as described in [Furnstahl et al., 2008] for bone segmentation in CT, and learned potential functions in the case of brain MR data. For the registration process, we follow the approach of [Glocker et al., 2008], with normalized cross-correlation as the image similarity metric.

### 2.2.1 Joint Registration and Segmentation

As a joint technique, we implement an alternating registration and segmentation (ARS) procedure similar to [Wyatt and Noble, 2003]. In ARS, the estimated solution for one subproblem is used as prior knowledge in the other in an iterative manner.

Using MRF approaches for both segmentation and non-rigid registration, we define an additional energy that links the deformed atlas segmentation and the target segmentation, namely coherence energy $E^{\text{coh}}(S_X, S_A(T))$. We compute this energy using a distance weighted overlap penalty between the target segmentation and the deformed atlas segmentation as follows:

$$
E^{\text{coh}}(S_A, S_X, T) = \sum_i (1 - \delta(l^S_i, S_A(T_i))) \psi^{\text{coh}}(l^R_i, l^S_i) \quad (2.3)
$$

$$
\psi^{\text{coh}}(l^R_i, l^S_i) = \frac{D_{l^S_i}(T_i)}{2\tau^2}, \quad (2.4)
$$

where $\delta$ is the Kronecker-delta, $D_{l^S_i}$ denotes the distance transform of $S_A$ with respect to label $l^S_i$ and $\tau$ is a tolerance parameter, which we decrease after each ARS iteration, starting from an empirically-set tolerance of 16mm as can be seen in Fig. 2.3. This decrease in tolerance was motivated by the fact that the expected difference between deformed atlas segmentation and target segmentation is larger in earlier levels compared to later stages. In experiments, a small tolerance in early levels was observed to undesirably force the target segmentation to be grossly dissimilar from the true segmentation, which in turn
leads to a local optimum in the iterative registration process. This was mitigated by the above-mentioned adaptive setting.

The ARS algorithm then alternates between estimating the registration and the segmentation as follows:

\[
T^{n+1} = \arg \min_{T} E_{\text{reg}}(X, A, T) + \lambda_{\text{coh}} E_{\text{coh}}(S_A, S_X, T) \\
S_{X}^{n+1} = \arg \max_{S_X} E_{\text{seg}}(X, S_X) + \lambda_{\text{coh}} E_{\text{coh}}(S_A, S_X, T^{n+1}),
\]

(2.5)  
(2.6)

where \(\lambda_{\text{coh}}\) weights the influence of the coherence energy, which is included in the MRF based solver as a unary potential. Note that in the first iteration, no estimate of the segmentation is known and a regular registration is computed. We check for convergence by computing the overlap between \(S^n\) and \(S^{n-1}\). The algorithm exits, if this overlap reaches 99% or a maximum of 10 iterations is reached.

### 2.2.2 Auxiliary Labels

Naturally, any criterion similar to \(E_{\text{coh}}(S_X, S_A(T))\) will favor a high overlap between a single-label atlas segmentation and any pixels which satisfy an intensity model for this anatomy. This may lead to a sub-optimal solution in the joint registration and segmentation process, where the atlas segmentation is forced to overlap with non-target anatomy pixels. An example is given in Fig. 2.1, where the mandibular bone is extended into the skull. While this effect may partly be mitigated by strong regularization, this would then not only deteriorate results in regions where high flexibility is necessary, but also could make the algorithm more sensitive to initialization.

We thus propose auxilary anatomical labels in both the atlas and the target segmentation in order to prevent such false positives in ARS segmentation. Such auxiliary labels can be obtained automatically using the following procedure:

1. An automatic binary segmentation \(\bar{S}_A\) of the atlas image is computed using Eq. (2.1).

2. The ground-truth atlas segmentation is dilated and subsequently subtracted from the estimate \(\bar{S}_A\). Such dilatation helps remove minor artefacts near the boundary of the ground-truth segmentation.

3. The auxiliary label \(l_a\) is then assigned to the remaining voxels in \(\bar{S}_A\) before combining it with the ground-truth \(S_A\) to generate an aux-label atlas segmentation.

An example of this process is shown in Fig. 2.2.
2. Auxiliary Anatomical Labels for JRS

Figure 2.2: Pipeline for obtaining the auxiliary-labelled atlas segmentation. The original image (a) is segmented using the segmentation sub-graph yielding a binary segmentation (b). The known ground-truth segmentation (c) is then removed from an eroded version of (b) in order to obtain the second label segmentation, which is subsequently diluted. In a last step, the ground-truth and the second label segmentation are combined, yielding the final aux-label segmentation (d).

2.2.2.1 Coherence Potentials for Auxiliary Labels

While the anatomy represented by the auxiliary label is expected to provide valuable information to an ARS algorithm, it is less reliable in comparison to a manual annotation for two reasons: (i) The auxiliary labelling is obtained automatically and thus prone to errors due to imaging artefacts. (ii) While the primary label was chosen carefully by a medical expert, our automatically labelled structures may not represent anatomies for which correspondences exist across different patients. Therefore, such auxiliary labels are expected inherently to have less coherence between the deformed atlas segmentation and the target segmentation estimate. We accommodate this by augmenting the coherence potential as follows:

\[ \Psi_{\text{coh}}^{\text{aux}}(l^R_i, l^S_i) = \begin{cases} 
\Psi_{\text{coh}}^{\text{aux}}(l^R_i, l^S_i) & \text{if } (l^S_i \equiv l_f) \lor (S_A(T)_i \equiv l_f), \\
\min(1, \Psi_{\text{coh}}^{\text{aux}}(l^R_i, l^S_i)) & \text{otherwise}. 
\end{cases} \]  

This function essentially truncates the coherence potential to a maximum of 1, if neither the target segmentation label nor the deformed atlas segmentation label is the foreground label \( l_f \). This limits the cost of labelling pixels in the target image as background, when the deformed atlas segmentation is labelled as \( l_a \). Analogously, the cost of labelling pixels as auxiliary anatomy, when the deformed atlas segmentation is labelled as background, is then also limited.

2.3 Experimental Evaluation

We evaluated our method on two datasets: a set of 70 mid-saggital slices of 2D brain MR, where the task is to segment the corpora callosa, and a set of 14 3D CTs of the head, where the task is to segment the mandibular bone. The mandibular bone dataset represents a typical clinical problem, and the corpus callosum dataset provides an opportunity to evaluate
2.3. Experimental Evaluation

<table>
<thead>
<tr>
<th>method</th>
<th>2D MRI</th>
<th></th>
<th></th>
<th>3D CT</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DICE</td>
<td>MSD</td>
<td>HD</td>
<td>DICE</td>
<td>MSD</td>
<td>HD</td>
</tr>
<tr>
<td>MRF-Seg.</td>
<td>0.837</td>
<td>0.500</td>
<td>8.34</td>
<td>0.90</td>
<td>3.39</td>
<td>17.00</td>
</tr>
<tr>
<td>MRF-Reg.</td>
<td>0.918</td>
<td>0.093</td>
<td>2.85</td>
<td>0.849</td>
<td>0.32</td>
<td>11.69</td>
</tr>
<tr>
<td>ARS (single-label)</td>
<td>0.918</td>
<td>0.084</td>
<td>2.31</td>
<td>0.924</td>
<td>0.16</td>
<td>9.1</td>
</tr>
<tr>
<td>ARS (aux-label)</td>
<td>0.925</td>
<td>0.064</td>
<td>2.11</td>
<td>0.933</td>
<td>0.13</td>
<td>7.88</td>
</tr>
</tbody>
</table>

Table 2.1: Average segmentation quality measures for the methods on both datasets.

Figure 2.3: Progression of the coherence potential for a given deformation and decreasing tolerance.

our algorithm on a significantly larger dataset and a different modality. Both datasets were rigidly registered to a common reference frame, and gold-standard segmentations have been provided by experts. We report Dice’s coefficient (DICE), Hausdorff distance (HD) and mean surface distance (MSD) as quantitative measures of segmentation accuracy. We optimized the parameters of each method separately using a small subset of the images. Additionally, we report results for pure MRF segmentation (MRF-Seg.), and for pure MRF registration which is our implementation of [Glocker et al., 2008] (MRF-Reg.).

Mandibular Bone Segmentation. A dataset of 15 3D CT scans of the head with 160x160x120 px³ dimensions and 1 mm³ voxel size was used for our validation, with the mandibles in each segmented by an expert. Due to older age group, the amount of teeth varies largely in this set, which in addition to artefacts from metal implants makes both automatic registration and segmentation of the mandibles a challenging task. We performed a leave-one-out (l1o) experiment by using each of the 15 images as the atlas image in turn to segment the remaining 14 samples, resulting in 210 experiments for each method. In Fig. 2.1, we present the average performance values over all experiments for the above-mentioned methods. Note that the pure MRF-based bone segmentation method (MRF-Seg.) is the method presented in [Furnstahl et al., 2008], for which no atlas image is needed and thus only 15 experiments were required for this method. Also, [Furnstahl et al., 2008] merely finds a separate label for each connected bone region, assuming
the target anatomy to be the largest connected region. However, this is not true for the mandible in the head CT scans. Therefore, we had to select the label for the mandible in a manual post-processing step, positively biasing the results in favor of MRF-Seg [Furnstahl et al., 2008].

Corpus Callosum Segmentation. A second experiment was performed using 70 mid-saggital slices of 2D MR images of the brain with 0.33 mm² pixel resolution. After rigid pre-alignment, each image was cropped to a fixed 200x120 px² region of interest. The total number of leave-one-out experiments for each method is then 4830, where we performed the same experiments as above, with the exception of the segmentation potential functions having been replaced by intensity based classifiers. We used Gaussian mixture models for foreground/background with four densities for the segmentation unary potential and random forests [Breiman, 2001] for the pairwise potentials, both trained using the atlas image. The negative logarithm of the classification probabilities were used as unary and pair-wise energies in the segmentation graph. We report the average metrics similarly in Fig. 2.1, and show sample results in Fig. 2.4.

2.3.1 Discussion

For both the corpus callosum and the mandibular bone segmentation tasks, aux-label ARS outperforms single-label ARS significantly ($p \leq 1.7e^{-4}$). This gain comes at some moderate additional cost, as aux-label ARS requires the additional computation of the auxiliary label distance transform in the coherence potential for both the segmentation and the registration steps, and also an additional segmentation label needs to be generated in the segmentation step. We solve the aux-label segmentation problem using the efficient alpha-expansion graph-cuts algorithm [Boykov et al., 2001], which can guarantee that the
solution is within a known factor of the global optimum. The following observations are made when comparing single- and aux-label ARS with respect to changing the free parameter $\lambda_{coh}$. Firstly, the optimal $\lambda_{coh}$ is an order of magnitude larger for the single-label ARS compared to its auxiliary-labelled version. Additionally, the optimal segmentation weights are slightly lower for the single-label ARS, causing it to more strongly rely on the registration information. This is indeed expected, as the segmentation sub-problem is essentially ill-posed when using only one label. In contrast, aux-label ARS is much less susceptible to sub-optimal parametrization in comparison to single-label ARS as can be seen from the narrower range in both Dice’s coefficient and Hausdorff distance. This demonstrates that the aux-label ARS formulation is indeed better suited for joint segmentation and registration problems as in this paper.

2.4 Conclusions

In this work, we have presented a novel approach using automatically generated auxiliary anatomical labels for joint segmentation and atlas registration. A distance-weighted overlap energy was proposed to couple the segmentation and registration process in an iterative fashion. We show experimentally that automatic auxiliary labelling helps improve segmentation quality compared to joint segmentation and registration using single target label alone. Although the presented method focuses on segmentation as a pixel-labelling problem, we expect similar behaviour for other segmentation methods such as active contours. Here, evolving a target contour simultaneously to auxiliary contours may help to avoid poor local optima which can result from initialisation problems or image artefacts.
Simultaneous Segmentation and Multi-Resolution Nonrigid Atlas Registration

In this paper, a novel Markov random field (MRF) based approach is presented for segmenting medical images while simultaneously registering an atlas non-rigidly. In the literature, both segmentation and registration have been studied extensively. For applications that involve both, such as segmentation via atlas-based registration, earlier studies proposed addressing these problems iteratively by feeding the output of each to initialize the other. This scheme, however, cannot guarantee an optimal solution for the combined task at hand, since these two individual problems are then treated separately. In this paper, we formulate simultaneous registration and segmentation (SRS) as a maximum a-posteriori (MAP) problem. We decompose the resulting probabilities such that the MAP inference can be done using MRFs. An efficient hierarchical implementation is employed, allowing coarse-to-fine registration while estimating segmentation at pixel level. The method is evaluated on two clinical datasets: mandibular bone segmentation in 3D CT and corpus callosum segmentation in 2D mid-saggital slices of brain MRI. A video tracking example is also given. Our implementation allows us to directly compare the proposed method with the individual segmentation/registration and the iterative approach using the exact same potential functions. In a leave-one-out evaluation, SRS demonstrated more accurate results in terms of Dice overlap and surface distance metrics for both datasets. We also show quantitatively that the SRS method is less sensitive to the errors in the registration as opposed to the iterative approach.

3. SIMULTANEOUS SEGMENTATION AND ATLAS REGISTRATION

Figure 3.1: A synthetic example comparing SRS to traditional methods. The goal is to segment image $X$, in which the circle represents the anatomy of interest. This can be performed by registering the atlas image $A$ and deforming the known atlas segmentation (Reg.), by directly segmenting $X$ using a learned intensity model (Seg.), or by using the proposed method which combines both traditional approaches (SRS).

3.1 Introduction

In computer vision in general and medical imaging in particular, the segmentation of an image is a common problem that has been studied extensively in the literature. Its applications include organ segmentation for medical planning and analysis [Pham et al., 1998], object tracking [Gu and Lee, 1998] or scene understanding [Li et al., 2009]. Possible approaches can be categorized by the amount of prior knowledge they use, such as intensity-based segmentation on one end of the spectrum and model-based segmentation on the other. Intensity-based segmentation typically involves a local classifier to determine the type of tissue that a pixel most likely belongs to. It can additionally regularize such segmentation, for example using Markov random fields [Boykov and Funke-Lea, 2006]. This method is typically fast, but lacks anatomical correctness as neither absolute nor relative spatial locations (e.g., shapes) are taken into account. On the other hand, model-based segmentation can incorporate such information using, for instance, active shape [Cootes and Taylor, 1992] or appearance models [Cootes et al., 2001]. A review of statistical models for segmentation can be found in [Heimann and Meinzer, 2009]. In order to produce a robust model, all these methods require a large number of training images with ground-truth, which in turn involves the tedious task of numerous manual annotations. Therefore, it is not straight-forward to quickly establish a model-based segmentation for a novel target anatomy.

Due to its weak dependency on manual annotations, many recent studies have focused on registration-based segmentation [Pohl et al., 2006,Glocker et al., 2008,Shekhovtsov et al., 2008,Mahapatra and Sun, 2010], where a single reference (atlas) image is registered to a target image. The resulting transformation is then applied to the labelled atlas, which yields a segmentation of the target image. Whereas various registration algorithms are applicable for this task, state-of-the-art performance has been achieved using Markov random fields (MRFs) on a control grid hierarchy [Glocker et al., 2008]. Methods such as tree-reweighted message passing [Kolmogorov, 2006] or fast primal dual decomposition [Komodakis et al., 2007] provide near-optimal MRF solutions even with a large number of discrete labels and non-convex priors. The solution of these MRFs is optimal with
3.1. Introduction

(a) Multi-res. reg.  (b) Pixel-level seg.

(c) Product-label SRS  (d) Our sum-label SRS

Figure 3.2: Multi-resolution registration (a) typically works on a control grid with increasing resolution to allow robust large-scale deformations. Graph-based segmentation (b) employs a high resolution (pixel-level) graph to capture fine detail. In product-label SRS (c), the registration and segmentation graphs are bound to use the same resolution; meaning the use of either a high-resolution registration control grid (top) hindering efficient large deformations, or a low-resolution segmentation grid (bottom) not being able to resolve small structures. Our proposed sum-label SRS (c) allows for multi-resolution registration concurrently with pixel-level segmentation.
respect to a selected criterion, often defined as a global similarity metric to reduce the difference between the target image and the deformed atlas image. While being widely used in medical imaging, image registration alone cannot solve the segmentation task by itself as it is considered to be an ill-posed problem [Fischer and Modersitzki, 2008]. This is because anatomical correspondences, which are not guaranteed to exist, are computed using surrogate criteria such as intensity similarity. In addition, the problem of non-rigid registration is known to be NP-complete [Keysers and Unger, 2003], thus making approximative algorithms prone to local optima, for example due to poor initialization. Increasing the amount of prior knowledge is a common remedy for such problems, for example by including statistical or multiple atlases [Rohlfing et al., 2005, Heckemann et al., 2006, Glocker et al., 2007].

Joint optimization of the segmentation and registration was first proposed by [Wyatt and Noble, 2003]. In alternating registration and segmentation (ARS), the estimated solution for one subproblem is used as prior knowledge in the other in an iterative manner, whereas simultaneous registration and segmentation (SRS) aims at optimizing both goals at the same time. A comparison of individual and joint approaches is demonstrated on a synthetic example in Fig. 3.1. To the best of our knowledge, almost all approaches published so far in the field have followed the ARS approach; the wording “simultaneous” was used commonly, though it was not used per se as defined above. The method proposed in [Wyatt and Noble, 2003] alternates between estimating a rigid transformation using Powells’ method and updating the segmentation using iterative conditional modes (ICM) in an MRF. [Pohl et al., 2006] proposed a Bayesian framework in which they alternate between updating the registration and estimating intensity “nuisance” parameters while marginalizing over possible anatomical labels using an EM algorithm. [Xiaohua et al., 2004] proposed an ARS approach based on MRFs. Their method alternates between solving one MRF that optimizes registration parameters and updates segmentation probabilities, and a second MRF solving the segmentation. There have been also studies on alternating registration and segmentation in variational formulations, which also rely on iteratively updating the registration and segmentation solutions, for example, by using gradient descent [Wang et al., 2006a, Ghosh et al., 2010, Schmidt-Richberg and Ehrhardt, 2012]. For example, a method which alternates between updating the deformation field using a quasi-Newton optimization method and evolving a contour in the target image while constraining the distance between deformed atlas and said target contour was proposed in [Wang et al., 2006a]. It is also possible to combine registration with model-based segmentation as shown in [Lu and Duncan, 2012], where point-based registration and active-shape segmentation trained from manually annotated images are employed for mutual benefit.

For non-medical applications, optical flow [Horn and Schunck, 1981] has been frequently used for estimating the motion in image sequences or stereo images. Such inter-frame flow estimations may though fail in the presence of occlusions – a problem indeed similar to its medical image registration counterpart when posed as a lack of complete correspon-
3.1. Introduction

Differences between frames. To alleviate this, several studies focused on estimating optical flow while jointly computing a segmentation. For example, [Sudderth and Black, 2012] segments an image sequence into *layers* while estimating flow. Information from such a sequence then helps resolve occlusions. A different technique was presented in [Lei and Yang, 2009], where optical flow is computed on a hierarchy of over-segmentations based on super-pixels. Note that such techniques do not operate on 3D images. Furthermore, they are not designed to accommodate large deformations expected in inter-patient registration. Therefore, those methods are not applicable in our targeted domain.

For *simultaneous* registration and image based segmentation, Markov random field based approaches are attractive due to the following reasons: (i) Image segmentation is naturally a discrete labelling problem, for which MRFs can efficiently find excellent solutions. (ii) Registration can also be defined as a discrete optimization problem [Glocker et al., 2008], which does not rely on gradients and is thus potentially less sensitive to local minima, for example, due to poor initialisation. Two recent methods in the literature have explored such MRF-based simultaneous segmentation and registration: In [Mahapatra and Sun, 2012], intra-patient time-series image registration was shown to be improved by simultaneously estimating a segmentation of target and floating image using a *user-defined ROI*. This work does not use an atlas segmentation and targets only small displacements, which will prevent inter-patient registration. In [Parisot et al., 2012], a healthy brain was registered to one with a tumor while the tumor in the target was being segmented. This method infers tumor tissue from registration non-overlap (tear), and vice versa. Note that, similarly to [Mahapatra and Sun, 2012], this work does not use an atlas segmentation either, as the healthy atlas does not contain a tumor.

Compared to other SRS approaches in the literature, we target inter-patient atlas-based segmentation. Indeed, our SRS technique was developed to address several aspects required for this task, for which other existing approaches are not suitable. The biggest difference of our SRS technique from other approaches is our design to accommodate different levels of detail for the segmentation and registration subproblems concurrently. Without allowing for different levels of detail, an SRS solution will either be unable to robustly model large deformations (e.g., in [Mahapatra and Sun, 2012]) inevitable in inter-atlas registrations; or will lead to very coarse segmentations (e.g., in [Parisot et al., 2012]) unusable for precise delineation. Our method not only can handle both of those at the same time, but it also can do this without increasing the computational cost significantly, due to the treatment of the solution label-space. To clarify such difference, we name the previously proposed techniques *product-label* SRS, as they search for a discrete labelling in the product label space of segmentation and registration labels. Our method can in contrast be called *sum-label* SRS, as its solution label space is the union of segmentation and registration label spaces and therefore the total number of labels is the *sum* of the number of registration and segmentation labels. This fundamental difference is illustrated on MRF-graphs in Fig. 3.2. Different levels of detail are accommodated in this paper by our novel two-layer graph structure connected by inter-layer edges that model a coher-
ence potential. Thanks to our elaborate probabilistic derivation of the underlying SRS problem, we could then soundly define each component of such a two-layer model by a suitable energy term.

Accordingly, the contributions of this paper can be listed as follows: (i.) Our novel simultaneous registration and segmentation method allows pixel-level segmentation with simultaneous multi-level non-rigid registration. (ii.) In contrast to previous SRS-like approaches, our work utilizes a known atlas segmentation. (iii.) Our SRS model is derived through a rigorous Bayesian analysis of the underlying maximum a posteriori problem. (iv.) Experimental comparison between alternating and simultaneous segmentation and registration approaches is presented within the same framework.

This paper is structured as follows. First, we introduce the general concepts and our notation. Next, our decomposition of the underlying MAP criterion is presented. Then, the structure of the resulting MRF is given, followed by our implementation. The paper is concluded with an experimental evaluation on two clinical datasets.

### 3.2 Background and Notation

Let $X$ be the target and $A$ the atlas image, where both are functions $\Omega \to \mathcal{F}$. $\Omega \in \mathbb{N}^D$ is the $D$-dimensional discretized coordinate domain, and $\mathcal{F}$ is an arbitrary feature descriptor, e.g. the image intensity in many cases or a feature vector. A segmentation $S = \{I_0^S, \ldots, I_N^S\}$ is an assignment of one segmentation label $I_s^S$ out of the set of all segmentation labels $L_S$ to each of the $N$ pixels of an image. In a plain segmentation task, the segmentation $S_X$ of the target image $X$ is sought. In contrast, in a plain registration task the goal is to find a transformation $T$ such that the target image $X$ and the transformed atlas image $A(T)$ become similar according to an appropriately selected measure. In our case, $T$ is a non-rigid deformation represented by a $D$-dimensional displacement vector field $T : \Omega \to \mathbb{R}^D$ so that for each pixel/voxel $i \in \Omega$:

$$T(i) = i + T_i. \quad (3.1)$$

$T_i$ and $T_i$ are short-hands for the local deformation $T(i)$ and displacement $T(i)$. We will also use this notation to denote pixels in an image $I$ as $I_i$, e.g. $X_i, A_i$. $I(T) = i \to I(i + T_i) \forall i$, indicates a deformed image. Values at non-grid locations are interpolated.

Both finding a displacement vector field $T$ and finding a segmentation $S_X$ can be defined as separate maximum a-posteriori (MAP) problems as follows:

$$\hat{S}_X = \underset{S_X}{\text{arg max}} p(S_X | X) \quad (3.2)$$

$$\hat{T} = \underset{T}{\text{arg max}} p(T | X, A) \quad (3.3)$$
The solution to each problem can be found by casting the MAP problem into an MRF, for which efficient inference algorithms exist. We will shortly recapitulate this process here as it will also be used for our method.

Assuming Gibbs’ distributions, the MAP estimation can be formulated as the minimization of an energy functional, which is proportional to the negative logarithm of the (non-normalized) probabilities [Li, 1995]:

\[
\hat{S}_X = \arg \min_{S_X} E_{\text{seg}}(S_X, X) \\
\hat{T} = \arg \min_{T} E_{\text{reg}}(T, X, A)
\]

where \( E_{\text{seg}} \) and \( E_{\text{reg}} \) are segmentation and registration costs, respectively. In order to use MRFs for solving the minimization problem, these energies are decomposed into unary (\( \psi \)) and pairwise (\( \Psi \)) potentials over discrete labels as follows:

\[
E_{\text{seg}}(S_X, X) = \sum_{i \in \Omega} \left( \psi_{i}^{\text{seg}}(l_i^S) + \sum_{j \in \mathcal{N}(i)} \lambda_{ij}^{\text{seg}} \Psi_{ij}^{\text{seg}}(l_i^S, l_j^S) \right)
\]

\[
E_{\text{reg}}(T, A, X) = \sum_{i \in \Omega} \left( \psi_{i}^{\text{reg}}(l_i^R) + \sum_{j \in \mathcal{N}(i)} \lambda_{ij}^{\text{reg}} \Psi_{ij}^{\text{reg}}(l_i^R, l_j^R) \right)
\]

where the continuous displacement space is sampled discretely, so that each registration label \( l_i^R \) in the set of all registration labels \( L_R \) maps to a unique displacement vector \( \vec{d}_i \). The unary potentials correspond to a data driven term, representing how likely it is to observe a label at a location given the data at that point. The pairwise potentials correspond to a prior over the entire labelling, which is often implemented as a smoothing over the neighborhood \( \mathcal{N} \) as justified by a first-order Markov assumption. The factors \( \lambda \) provide balancing weights. These are often contrast sensitive to prevent excessive smoothing along image edges. Minimizing (3.6) using graph-cuts [Boykov and Funke-Lea, 2006] is common in medical image segmentation [Boykov and Jolly, 2000, Furnstahl et al., 2008]. However, minimizing (3.7) directly is computationally expensive (due to a large label space) and difficult to regularize meaningfully. Nevertheless, efficient methods have been developed that can result in robust and smooth solutions using (3.7) by finding the displacements \( \vec{d} \) of control points in a multi-resolution cubic B-spline framework [Glocker...
3. SIMULTANEOUS SEGMENTATION AND ATLAS REGISTRATION

et al., 2008]. Such control points are placed on a coarse Cartesian grid, which allows computing the displacement at a position $\vec{x} = (x, y)$ as follows:

$$T(\vec{x}) = \sum_{i=0}^{3} \sum_{m=0}^{3} B_l(u)B_m(v)\vec{d}_{i+l,j+m},$$  \hfill (3.8)

where $i$ and $j$ are the indices of the closest grid control point to $\vec{x}$; $u$ and $v$ are the normalized distances to this control point; and $B_l$ is the $l$th B-spline basis function. The three-dimensional interpolation can be defined analogously.

Given the atlas segmentation $S_A$, one iteration of the alternating approach ARS can be formulated similarly to [Wyatt and Noble, 2003] as:

$$T^{n+1} = \arg \max_T p(T|X, A, S_A, S_X^n)$$  \hfill (3.9)

$$S_X^{n+1} = \arg \max_{S_X} p(S_X|X, A, S_A, T^{n+1}).$$  \hfill (3.10)

where $T^0$ is initialized using plain (often intensity-based) registration and the optimization is terminated upon convergence of these terms. Note that here each maximization problem can be solved using MRFs, iterative methods or a combination thereof, as investigated in different studies in the literature.

3.3 Simultaneous Registration and Segmentation

In simultaneous registration and segmentation (SRS), one seeks for the segmentation $\hat{S}_X$ and the displacement field $\hat{T}$ that maximize the joint posterior probability given the target and atlas image and the atlas segmentation:

$$(\hat{S}_X, \hat{T}) = \arg \max_{S_X,T} p(S_X, T|X, A, S_A).$$  \hfill (3.11)

As this posterior distribution is not trivial to model, let alone devise an algorithm to find the MAP solution, we first decompose this full problem within the Bayesian framework.
3.3. Simultaneous Registration and Segmentation

Figure 3.3: (a) The Bayesian network for (3.16), (b) its moral graph, and (c) factor graph representations. Observed variables are shaded in gray. The factor graph (c) shows how the coherence potential relates the individual registration and segmentation sub-problems.

into smaller sub-problems that are comprehensible, relate to common image analysis tasks (e.g. segmentation and registration), and possibly have well-studied MRF formulations:

\[
\begin{align*}
(\hat{S}_X, \hat{T}) &= \arg \max_{S_X, T} p(S_X, T | X, A, S_A) \\
&= \arg \max_{S_X, T} p(S_X, T | X, A) \cdot p(S_A | X, A, S_X, T) \\
&= \arg \max_{S_X, T} p(S_X | X, A) \cdot p(T | X, A, S_X) \\
&\quad \cdot p(S_A | X, A, S_X, T)
\end{align*}
\]

(3.15)

where the denominator in (3.13) is omitted as it is constant with respect to the parameters optimized.

Note that the above is an exact derivation without any assumptions, yielding three probabilistic factors that represent the main building blocks of our method. It is now observable that maximizing this SRS criterion corresponds to maximizing the product of a segmentation posterior, a registration posterior, and a term that expresses the likelihood of observing an atlas segmentation given the remaining variables and estimates. Below, these three terms in (3.15) are further analyzed individually and our assumptions of certain (conditional) independencies are introduced with justifications, where necessary.

**Segmentation Posterior (a)** is the posterior probability of the segmentation \(S_X\) given the target and atlas image intensities. Given \(X\), it is trivial that \(S_X\) is conditionally independent of \(A\). This simplifies the term (a) to \(p(S_X | X)\), which is indeed the original MAP segmentation criterion given in (3.3).
3. SIMULTANEOUS SEGMENTATION AND ATLAS REGISTRATION

**Registration Posterior (b)** is the posterior probability of the deformation $T$ given the target and atlas intensities, as well as the target segmentation $S_X$. Here, we introduce the assumption that $T$ is independent of $S_X$, given $A$ and $X$. Note that this effectively restricts the use of smoothing the deformation differently at object boundaries defined by $S_X$. Nevertheless, such smoothing is already not feasible at coarse levels of deformation we employ. The final term is then identical to the MAP criterion given for image registration in (3.3).

**Coherence (c)** is the probability of observing $S_A$, given $S_X, T, X, A$. This term thus controls the similarity of the (deformed) atlas segmentation and the target segmentation estimates. It can be simplified by assuming that $S_A$ is independent of $X, A$, given $S_X$ and $T$. It is important to note that this term motivates the modelling of uncertainty in the SRS process, since it expresses a probabilistic relation between $S_X$ and $S_A(T)$ rather than assuming $S_X = S_A(T)$.

The observations above then lead to the following final MAP formulation for simultaneous segmentation and registration:

$$
(\hat{S}_X, \hat{T}) = \arg \max_{S_X, T} \underbrace{p(S_X|X)}_{\text{segmentation}} \underbrace{p(T|A, X)}_{\text{registration}} \underbrace{p(S_A|S_X, T)}_{\text{coherence}},
$$

where the segmentation $S_X$ and the registration $T$ are connected (indirectly) through the coherence factor $p(S_A|S_X, T)$. This is visualized formally in Fig. 3.3, where the Bayesian network defined by (3.16) and its conversion to a factor graph are shown. This definition has the advantage that both the registration and the segmentation appear in a common, well-studied form and thus both can be estimated using well-known, state-of-the-art techniques. In addition, the coherence term can be implemented efficiently as we will show below.

### 3.4 Implementation using Markov Random Fields

In order to solve the MAP problem, we cast it into an MRF, for which efficient inference algorithms exist. We will first motivate and describe the structure of the MRF before detailing the implementation of the potential functions and its optimization.
3.4. IMPLEMENTATION USING MARKOV RANDOM FIELDS

Figure 3.4: Schematic of a local pixel neighborhood in the SRS MRF defined on the pixel level. Observed variables are shaded in gray, and the factors (potential functions) are colored according to their semantics.

3.4.1 Two-Layer Graph

In order to cast the problem in the form of a first order MRF, the posterior estimation in (3.11) needs to be decomposed into unary and binary terms. Commonly, a Gibbs distribution is assumed as follows:

\[
p(S_X, T | X, A, S_A) = \frac{\exp \left( -E_{SRS}(S_X, T, X, A, S_A) \right)}{Z} \tag{3.17}
\]

We now use our decomposition (3.16) to show the structure of the defined SRS energy \( E_{SRS} \). We assume Gibbs distributions for each individual term. We also use the fact that the product of Gibbs distributions is again a Gibbs distribution [Koller and Friedman, 2009, Proposition 4.7, pg. 134], and omit normalization factors \( Z \) which are constant with regard to the optimization:

\[
\text{arg max}_{S_X, T} p(S_X, T | X, A, S_A) \tag{3.18}
\]

\[
= \text{arg max}_{S_X, T} p(S_X | X)p(T | A, X)p(S_A | S_X, T) \tag{3.19}
\]

\[
= \text{arg min}_{S_X, T} \left( \theta_{\text{seg}} E_{\text{seg}}(S_X, X) + \theta_{\text{reg}} E_{\text{reg}}(T, X, A)
+ \theta_{\text{coh}} E_{\text{coh}}(S_A, S_X, T) \right), \tag{3.20}
\]

where the parameters \( \theta_{\{\}} \) are weighting parameters to compensate the dropped normalization terms. This results in a two-layer graph, where one layer is a segmentation graph, and the other layer is a registration graph. These layers are connected by coherence edges which are motivated as follows.
The coherence potential couples the registration and segmentation potential by introducing the likelihood of observing an atlas segmentation given registration and target segmentation estimates. Intuitively, it shall yield a high probability, when the deformed atlas segmentation \( S_A(T) \) can be well explained by the target segmentation \( S_X \). The likelihoods can be computed locally analogously to the unary potentials in (3.6) and (3.7). Then, the coherence term in (3.16) can be rewritten as:

\[
p(S_A|S_X, T) = p(S_A(T)|S_X) = \prod_{i \in \Omega} p_i(S_A(T_i)|l_i^S).
\]

This allows modelling the coherence for each pixel individually, and to convert it into an energy similar to (3.6):

\[
E_{coh}(S_A, T, S_X) = \sum_{i \in \Omega} \Psi_{coh}^{ij}(l_R^i, l_S^i),
\]

where \( \Psi_{coh}^{ij} \propto -\log(p_i(S_A(T_i)|l_i^S)) \) is a pairwise potential function over the joint label space of the segmentation and registration labels which will be defined in Sec. 3.4.3.

This definition of the coherence potentials along with the segmentation and registration energies defined in (3.6,3.7) entails an MRF on the pixel level which is shown in Fig. 3.4. As mentioned earlier, this will make the method either computationally expensive or limit the possible range of deformations. We thus propose the following novel multi-resolution framework for SRS.

### 3.4.2 SRS using a Multi-Resolution Framework

Despite a coarse-to-fine approach for registration and different grid resolutions/iterations thereby employed, we perform segmentation at a fine level at all times. Accordingly, the segmentation energy \( E_{seg} \) is implemented as given in (3.6) at pixel resolution similarly to conventional MRF-based segmentation methods. For the registration, however, we use a multi-resolution registration approach [Glocker et al., 2008], where only the displacements of a coarse (Cartesian) control grid are estimated iteratively. Similarly to [Glocker et al., 2008], such grid displacements are applied to the full-resolution image at each iteration using B-spline transform as in (3.8). Then, the \( E_{reg} \) in MRF can be defined similarly to (3.7), but now on a graph \( G_{reg} \) with nodes representing grid corner displacements as follows:

\[
E_{reg}(T, A, X) = \sum_{i \in G_{reg}} \left( \psi_{i}^{reg}(l_R^i) + \sum_{j \in N(i)} \Psi_{ij}^{reg}(l_R^i, l_R^j) \right)
\]
where registration unaries $\psi_{i,reg}$ at grid locations $i$ are computed by back-projecting the cost (i.e. similarity metric for a given grid node displacement) from non-grid locations, while $\Psi_{reg}$ ensures a smooth deformation field. We utilize a similar approach when inferring pixel-level segmentation effects of coarse registration-grid displacements, and vice versa. The coherence potential $E_{coh}$ is indeed what models such mutual effects and thus links the registration and segmentation sub-graphs. Formally, displacements $\tilde{d}_k$ between grid locations $i \in \mathcal{G}_{reg}$ can be interpolated from grid displacements $\tilde{d}_i$ as:

$$\tilde{d}_k = \sum_{i \in \mathcal{G}_{reg}} \eta(k, i) \cdot \tilde{d}_i$$  \hspace{1cm} (3.25)

where $\eta$ defines the interpolation weight of $i$ at $k$. Consequently, the coherence energy uses such displacements interpolated at pixel-level as:

$$E_{coh} = \sum_{i \in \mathcal{G}_{reg}} \sum_{k \in \Omega} \eta(k, i) \Psi_{ik}^{coh}(t^R_i, t^S_k),$$  \hspace{1cm} (3.26)

Each segmentation-graph node is connected by coherence edges to all registration-graph nodes within such interpolation kernel $\eta$. In other words, every non-zero $\eta$ weight will yield a graph edge for coherence. For instance, a linear interpolation will require 8 edges per voxel in 3D. For a minimal graph structure, we thus chose $\eta$ to be nearest-neighbour interpolation, leading to the two-layer graph structure seen in Fig. 3.5.
Algorithm 1 Outline of the SRS algorithm.

**Data:** $S_A, A, X$

**Result:** $\hat{S}_X, \hat{T}$

$n_{\text{max}} \leftarrow 5$

**forall the grid levels do**

$n \leftarrow 0$

$\hat{T} \leftarrow \bar{T}$

$d_{\text{max}} \leftarrow 0.4 \cdot \text{gridSpacing}$

**while** $(n < 2) \lor (E^n_{\text{srs}} - E^{n-1}_{\text{srs}} > \epsilon) \land (n < n_{\text{max}})$ **do**

$n \leftarrow n + 1$

$E^n_{\text{srs}} \leftarrow \text{solveMRF}(S_A, X, A, \hat{T}, d_{\text{max}})$

$\hat{S}_X, \hat{T}_{\text{low}} \leftarrow \text{getMRFSolution()}$

$\hat{T} \leftarrow \text{bSplineInterpolate}(\hat{T}_{\text{low}})$

$\hat{T} \leftarrow \hat{T} \circ \bar{T}$

$d_{\text{max}} \leftarrow 0.5 \cdot d_{\text{max}}$

**end**

Increase grid resolution by halving gridSpacing

**end**

---

Figure 3.6: Progression of the SRS algorithm on a 2D MRI corpus callosum segmentation task. Each column shows the initialization and the solution of each level of the MR hierarchy. The first row shows the (deformed) atlas image $A$ with a control point grid overlay, the second the (deformed) atlas segmentation $S_A(T)$, the third row the distance transform $D$ used for the coherence potential, and finally the last row shows the target segmentation estimate $S_X$. Note that $D$ is shown as a logarithmically scaled version of $p_i(S_A(T_i)|l^S_i)$ for better contrast.
3.4.3 Algorithm

Having defined the MRF structure above, we now summarize our algorithm next. As a multi-resolution approach, it solves a sequence of individual MRFs with different registration control grid resolutions. The deformations from each step are then composed to initialize the next step. This process is outlined in Alg. 1.

We start with a coarse displacement control grid and interpolate the costs of labelling pixels at sub-grid level. We use a B-spline interpolation to interpolate the full deformation once the displacements of grid control points are optimized [Szeliski and Coughlan, 1997]. This interpolation is then used for the composition of the deformations at full resolution as follows:

\[
\hat{T} = T \circ \hat{T}, \quad \hat{T}(i) = T(\hat{T}(i)) + \hat{T}(i).
\] (3.27)

For efficiency reasons, we always keep the last deformation in memory at this full resolution, as we would need to interpolate multiple times in the registration unary and coherence potentials. The coarsest control grid divides the shortest edge of the image by 3 and tries to keep square (or cubic in 3D) grid cells. This allows capturing large deformations in registration. Similarly to [Glocker et al., 2008], the control grid resolution is doubled at each level, where the displacement parameters are optimized iteratively for each resolution until the convergence of the MRF energy. Such inner iterations of SRS, labelled \( n \) in Alg. 1, allows us to find precise displacements by re-sampling the displacement label space. The maximum sampled displacement at each inner iteration is set to be 40% of the control grid spacing. This guarantees that deformations are diffeomorphic when B-spline interpolating them at non-grid locations [Rueckert et al., 2006]. At each iteration, the maximum displacement \( d_{\text{max}} \) is then halved to increase the precision of registration estimates (down to sub-pixel displacements) using only a limited number of displacement labels for an efficient MRF implementation. We sample the displacements uniformly to a \( 9 \times 9 \) grid in 2D, while restricting the sampling to the principal axes in 3D for increased efficiency. For the registration similarity metric, both target and atlas images are downsampled by a factor linked to the grid resolution, after smoothing with a Gaussian kernel to avoid aliasing. The kernel size of the Gaussian was set to \( \sigma = 0.5 * (\Delta_{\text{target}} - \Delta_{\text{source}}) \), where \( \Delta_{\text{target}} \) is the target spacing, and \( \Delta_{\text{source}} \) is the spacing of the original image.

3.4.3.1 Potential Functions

In this section, we present the individual potential functions used in our approach. We demonstrate the flexibility of our method by using domain-specific segmentation potentials, which only need to satisfy general properties as imposed by the used MRF optimizer. We will explain the coherence potential in detail, while only briefly summarizing the remaining functions as they are standard methods.
Coherence Pairwise \( \Psi_{ij}^{coh} \): As explained in (3.22), the coherence potential is based on pixel-wise comparisons between the deformed atlas segmentation and the target segmentation estimate. We model the probability \( p_i(S_A(T_i)|l_i^S) \) as a Gaussian based on the distance of a point \( i \) and the nearest point in \( S_A(T) \) with the segmentation label \( l_i^S \). Let \( D_{l_i^S} \) denote the distance transform of \( S_A \) with respect to label \( l_i^S \). Then, we compute the probability as:

\[
p_i(S_A(T_i)|l_i^S) \propto \exp\left(-\frac{D_{l_i^S}(T_i)^2}{2\tau^2}\right),
\]

(3.28)

where \( \tau \) is a tolerance parameter, which we set based on multi-resolution level. This can be seen in the third row of Fig. 3.6, where we started with a tolerance of 16 mm which is halved at every subsequent level of hierarchy. This is motivated by the fact that the expected difference between deformed atlas segmentation and target segmentation is larger in earlier levels compared to later stages. In experiments, a small tolerance in early levels was observed to force the target segmentation to be grossly dissimilar to the true segmentation, leading to a local optimum in the iterative registration process since the deformed atlas segmentation has not yet aligned properly. This is mitigated by the adaptive tolerance setting proposed before. Taking the negative logarithm of (3.28), this then results in the following potential function:

\[
\Psi_{ij}^{coh}(l_i^R, l_i^S) = \frac{D_{l_i^S}(T_i)^2}{2\tau^2},
\]

(3.29)

We threshold the distance transforms to be \( \geq 0 \), so that the potential is zero when the label of the deformed atlas segmentation is equal to the segmentation label, i.e. \( S_A(T_i) = l_i^S \).

Segmentation Unary \( \psi_i^{seg} \): Depending on the image domain, we use a different segmentation potential. For CT bone images, we use a threshold-based potential using the known Hounsfield units (HU) for cortical bone (>500) and air (<-300) as presented in [Furnstahl et al., 2008]. For the brain MR images, we train a Gaussian mixture model (GMM) with two densities per class on the atlas (image+segmentation) and use the negative logarithm as the potential.

Segmentation Pairwise \( \Psi_{ij}^{seg} \): Similar to the segmentation unaries, we also use domain specific pairwise segmentation potential. For the CT images, we also use the potential of [Furnstahl et al., 2008], which is based on the contrast of a sheetness filtered version of the CT image [Descoteaux et al., 2006]. For the MR images, we train a random forest (RF) classifier [Breiman, 2001] using each pair of neighboring atlas image pixels as input.
and their respective segmentation labels as class labels. Then, the pairwise potential is selected as the negative logarithm of such classifier probability:

\[
\Psi_{i,j}^{\text{seg}}(l_i^S, l_j^S, X_i, X_j) = -\log p(l_i^S = l_j^S \mid X_i - X_j) (1 - \delta(l_i^S, l_j^S)).
\]

The term \((1 - \delta(l_i^S, l_j^S))\) makes sure that the potential function is zero for equal labels in order to satisfy the metric properties necessary for using alpha-expansion graph-cuts.

**Registration Unary** \(\psi_{i,j}^{\text{reg}}\): For the unary registration potentials, we use normalized cross correlation (NCC) between atlas and target image pair, as evaluated on the patches that are centered at the grid control points extending up to the neighboring grid points.

**Registration Pairwise** \(\Psi_{i,j}^{\text{reg}}\): The pairwise registration potentials are computed using the squared Euclidean distance of the displacement of neighboring grid control points.

We optimize each resulting MRF using alpha-expansion graph-cuts [Boykov et al., 2001]. This algorithm iteratively solves binary graph-cuts [Boykov and Jolly, 2000], where in each iteration one label is chosen and its region expanded by the algorithm. The method is very fast and guaranteed to converge to a solution within a known factor of the global optimum. We used an implementation that allows us to use sparse label sets [Veksler and Delong, 2013]. This is essential for our MRF, as the nodes from the registration and segmentation subgraph have different sets of labels. Note that in each expansion move, either a segmentation or a registration label is expanded. In our two-layer graph, this means that only the labels of nodes in one such layer are being changed, while the labels of all nodes of the respective other subgraph are fixed. Therefore, the cost of the coherence pairwise potentials can be added to the unary potentials as done in the implementation [Veksler and Delong, 2013]. As the pairwise potentials of both the registration and segmentation subgraph are metrics, the requirements for \(\alpha\)-expansion graph-cuts are met. To test whether this approach still yields good solutions, we compared \(\alpha\)-expansion graph-cuts with tree-reweighted message passing (TRW-S, [Kolmogorov, 2006]), which is a much slower discrete optimization technique with the advantage of giving a lower bound of the energy. This allows assessing the quality of the solution. It can be seen in Fig. 3.7 that \(\alpha\)-expansion graph-cuts quickly finds a minimum (2784) which is very close to the lower bound (2768) found by TRW-S after several iterations. Note that TRW-S had not converged even after over 40 minutes of running, shown in this figure.

### 3.4.4 Implementation of ARS

In order to show the advantages of optimizing for the segmentation and registration simultaneously as opposed to iteratively, we also implemented an ARS approach. Using
the ARS formulation in (3.9), and the same assumption as for SRS (cf Sec. 3.3), we obtain the following iterative MAP problems:

\[
T^{n+1} = \arg \max_T p(T|X, A)p(S_A|T, S^n_X)
\]

\[
S^{n+1}_X = \arg \max_{S_X} p(S_X|X)p(S_A|T^{n+1}, S_X).
\]

This formulation entails separate segmentation and registration MRFs, where the coherence potentials (3.26) are added to the unary potentials. This allows utilizing the exact same potential functions as for the SRS to also iteratively estimate the parameters. Note that an initialization step is needed, where we start by computing the registration without a segmentation estimate. We check for convergence by computing the overlap between \(S^n\) and \(S^{n-1}\) and stop if this overlap reaches 99% or a maximum of 10 iterations is reached. To enforce convergence, we decrease the tolerance parameter \(\tau\) after every ARS iteration, analogously to the tolerance decrease in SRS.

### 3.5 Experimental Evaluation

We evaluated our method on two datasets of different anatomy, modality and dimensionality. A summary is given in Tab. 3.1. The mandibular bone dataset represents a typical clinical problem, and the corpus callosum dataset provides an opportunity to evaluate our algorithm on a significantly larger dataset and a different modality. Both datasets were rigidly registered to a common reference frame, and gold-standard segmentations have been provided by experts. In order to assess both volumetric and surface errors, the fol-
Table 3.1: Evaluation datasets. The corpus callosum images are obtained by extracting a fixed ROI from rigidly aligned mid-saggital slices. Inside the ROI the size and location variability of the CC is high, as can be seen from the examples in Fig. 3.11. The head CTs are taken from cadavers of elderly people, and thus have a large variability with regard to the number of teeth.

<table>
<thead>
<tr>
<th></th>
<th>Corpus callosum</th>
<th>Mandibular bone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality</td>
<td>MR</td>
<td>CT</td>
</tr>
<tr>
<td># samples</td>
<td>70</td>
<td>15</td>
</tr>
<tr>
<td>Dimensionality</td>
<td>120 × 200</td>
<td>160 × 160 × 120</td>
</tr>
<tr>
<td>Spacing</td>
<td>0.3 mm</td>
<td>1 mm</td>
</tr>
</tbody>
</table>

Table 3.2: Parameters for the different experiments. We fixed \( \theta_{\text{reg}} = 1 \) and set the remaining parameters relative to this.

<table>
<thead>
<tr>
<th></th>
<th>ARS</th>
<th>SRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda_{\text{seg}} )</td>
<td>0.2 ( 1e^{-7} )</td>
<td>0.2 ( 1e^{-7} )</td>
</tr>
<tr>
<td>( \lambda_{\text{reg}} )</td>
<td>30 ( 1e^{-7} )</td>
<td>0.2 ( 1e^{-7} )</td>
</tr>
<tr>
<td>( \theta_{\text{seg}} )</td>
<td>0.03</td>
<td>0.2</td>
</tr>
<tr>
<td>( \theta_{\text{reg}} )</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>( \theta_{\text{coh}} )</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

following three metrics were used: (i) Dice’s coefficient (DICE) [Dice, 1945] measures the volumetric overlap between the estimated and the ground-truth segmentations; (ii) Hausdorff distance (HD) [Huttenlocher et al., 1993] measures the maximum surface-to-surface distance (worst-case error); (iii) Mean surface distance (MSD) is the average distance from each point on segmented surface to the other surface (computed both directions to avoid bias).

The parameters \( \theta \) have been optimized for each method separately using a small subset of the images, considering the DICE metric of the segmentation outcome as the optimization goal. The parameters for the experiments presented in this paper can be seen in Tab. 3.2. The interpolation \( \eta \) for coherence potentials described in Sec. 3.4.2 was chosen to be nearest-neighbour interpolation in this work for a minimal and hence efficient two-layer MRF graph structure. Furthermore, no significant improvement was observed in our initial experiments using, for instance, linear interpolation instead.

We start our experimental evaluation with the observation that most segmentation unary potentials are not anatomy specific, but rather sensitive to an intensity range shared among multiple anatomic structures, such as bones. For plain segmentation, this leads to necessary post-processing steps to separate anatomies and even select the target anatomy manually in case no automatic criterion such as the largest connected region exists [Furnstahl et al., 2008]. For any joint estimation of segmentation and registration, the implications are more severe as commonly only the target anatomy is labelled in the atlas [Gass et al.,
Table 3.3: Average segmentation quality metrics in the 3D CT mandibular bone segmentation task, for the evaluation of $S_A(\hat{T})$, i.e. the atlas segmentation deformed by estimated registration, and $\hat{S}_X$, i.e. the estimated segmentation.

<table>
<thead>
<tr>
<th>Method</th>
<th>$S_A(\hat{T})$ evaluation</th>
<th>$\hat{S}_X$ evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRF-Seg</td>
<td>-</td>
<td>0.900</td>
</tr>
<tr>
<td>MRF-Reg</td>
<td>0.849</td>
<td>0.32</td>
</tr>
<tr>
<td>ARS</td>
<td>0.864</td>
<td>0.27</td>
</tr>
<tr>
<td>SRS</td>
<td>0.898</td>
<td>0.16</td>
</tr>
<tr>
<td>Multi-atlas fusion</td>
<td>0.920</td>
<td>0.13</td>
</tr>
</tbody>
</table>

2014a. Since the segmentation unaries will impose penalties on not labelling a pixel according to its intensity profile, but any kind of coherence potential will penalize different labels in deformed atlas segmentation and target segmentation labelling, a distorted result may be obtained as illustrated in Fig. 3.8(left). It can be seen that the algorithm tries to maximize the overlap between bone and the atlas segmentation. While this effect can partly be countered by strong regularization, this will deteriorate results in regions where high flexibility is necessary, and also probably make the algorithm more sensitive with regard to initialization. We instead use an auxiliary label in both atlas and target segmentation label space, which is assigned to any pixel/voxel which meets the segmentation unary criterion but is not part of the target anatomy. For the atlas, this labelling can be obtained automatically and leads to much improved results as shown in Fig. 3.8(right). We utilize this multi-label scheme for all following ARS/SRS experiments.

Note that in our ARS and SRS implementations, the deformed atlas segmentation and the direct segmentation estimate are not forced to be equivalent. Although we consider the direct segmentation result $\hat{S}_X$ to be the main outcome of our algorithm, we also evaluate the quality of the estimated deformation by reporting segmentation overlap for the deformed atlas segmentation $S_A(\hat{T})$. As this is merely a surrogate measure for the registration quality [Rohlfing, 2012], we also report target registration errors (TRE) using manually placed landmarks in our 2D MR dataset. We also report results for only using the segmentation MRF (MRF-Seg), and only the original registration approach in [Glocker et al., 2008] (MRF-Reg). To compare our method to non-MRF based approaches, we also include results for the 2D dataset using the demons registration algorithm (Demons-Reg) [Vercauteren et al., 2009]. Lastly, we compare our method to a multi-atlas fusion approach described in [Heckemann et al., 2006]. Here, each image was segmented by locally adaptive fusion of the MRF-Reg deformed segmentations from all remaining images in the dataset.

Mandibular Bone Segmentation. We performed a leave-one-out (l1o) experiment by using each of the 15 images as the atlas image in turn to segment the remaining 14 samples, resulting in 210 experiments for each of the methods to be compared. In Tab. 3.3, we
Figure 3.8: Deformed segmentation $S_A(T)$ (top row) and segmentation estimate $S_X$ (bottom row) for single-label SRS (left) and multi-label SRS (right). Using a single label, the algorithm tries to maximize the overlap between any bone and the atlas segmentation. This is resolved by using an auxiliary label for non-target bone.

We present average performance values over all experiments for both ARS and SRS, and also compare it to the performance of plain intensity based registration using MRFs [Glocker et al., 2008] and to the bone segmentation method presented by [Furnstahl et al., 2008]. Note that for the latter, no atlas image is needed yielding only 15 experiments for this method. A separate label is found for each connected region and it is assumed that the target anatomy is the largest connected region. This does not hold for the mandible in the head CT scans. Therefore, we had to select the label for the mandible in a manual post-processing step. Since ARS and SRS return estimates for both the registration and segmentation parameters, we give their results separately. Recall that the coherence potential does not enforce the deformed atlas segmentation and the estimated segmentation to be equivalent, but rather utilizes a probabilistic approach which is dependent on parameterization. It can be observed, that the optimal parameterization allows SRS to not only obtain the best performance in term of the segmentation quality with regard to all evaluation metrics, but also improve the similarity of the deformed atlas segmentation to the ground-truth over both plain registration and ARS. We show qualitative results in Fig. 3.9, where typical behaviour for the individual approaches can be observed. The intensity based segmentation gives very precise bone boundaries in areas of strong contrast, but cannot differentiate between teeth and bone. Leakage occurs at the condyle-skull interface due to weak contrast. Plain registration shows an anatomically generally correct segmentation with one notable error at the left condyle, and relatively low precision at bone boundaries. ARS can be seen to follow the accuracy of the segmentation at bone boundaries, but also shows errors in areas which are difficult to register. The SRS result is the best, showing only minor errors at the condyles and the teeth area. On the mandibu-
Figure 3.9: Sample segmentations using the different approaches. The surfaces are colored with the local distances between estimated and ground-truth segmentation surfaces.

Figure 3.9: Sample segmentations using the different approaches. The surfaces are colored with the local distances between estimated and ground-truth segmentation surfaces.

lar bone segmentation task, SRS also performs better than the multi-atlas fusion method, despite the latter using 13 atlases per segmentation whereas our method using a single atlas.

Corpus Callosum Segmentation. We performed the same set of experiments on the CC dataset as we have on the mandibular bones, with the exception of exchanged segmentation potential functions as explained in Sec. 3.4.3.1. The total number of pairwise evaluations in the l1o scheme is 4830, for which we report the average metrics in Tab. 3.4. It can be observed that again SRS gives the best results both in terms of segmentation and deformed atlas segmentation quality. The improvement over the registration baseline is not as large compared to the mandibular bones, as the registration quality was already good to begin with. It is interesting, however, that even if the segmentation itself gives very weak results due to ambiguities in the intensities, the SRS method can utilize that knowledge to improve on both baselines. In addition to the MRF registration we also report error metrics using the common demons approach for registration [Vercauteren et al., 2009]. Overall, the results are worse compared to the MRF registration, which is partly
Table 3.4: Average segmentation quality metrics in the 2D MR corpus callosum segmentation task, for the evaluation of $S_A(\hat{T})$, i.e. the atlas segmentation deformed by estimated registration, and $\hat{S}_X$, i.e. the estimated segmentation.

<table>
<thead>
<tr>
<th>Method</th>
<th>DICE</th>
<th>MSD</th>
<th>HD</th>
<th>DICE</th>
<th>MSD</th>
<th>HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRF-Seg</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.837</td>
<td>0.500</td>
<td>8.34</td>
</tr>
<tr>
<td>Demons-Reg</td>
<td>0.907</td>
<td>0.321</td>
<td>3.30</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MRF-Reg</td>
<td>0.918</td>
<td>0.093</td>
<td>2.85</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ARS</td>
<td>0.923</td>
<td>0.086</td>
<td>2.68</td>
<td>0.929</td>
<td>0.082</td>
<td>2.45</td>
</tr>
<tr>
<td>SRS</td>
<td>0.932</td>
<td>0.062</td>
<td>2.19</td>
<td>0.935</td>
<td>0.061</td>
<td>2.08</td>
</tr>
</tbody>
</table>

Table 3.5: Landmark registration error (mm) on the CC. Each column contains the average distance for one of the five geometric and the manually obtained fornix landmark over all atlas-target image pairs for each method. 1=anterior point of CC; 2=topmost point of CC; 3=posterior point of CC; 4=inferior tip of splenum; 5=posterior tip genu.

<table>
<thead>
<tr>
<th>Landmark</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>4.65</td>
<td>5.85</td>
<td>5.68</td>
<td>5.97</td>
<td>5.79</td>
<td>5.74</td>
</tr>
<tr>
<td>MRF-Reg</td>
<td>1.48</td>
<td>1.96</td>
<td>4.14</td>
<td>2.30</td>
<td>3.23</td>
<td>2.81</td>
</tr>
<tr>
<td>ARS</td>
<td>1.47</td>
<td>1.91</td>
<td>4.11</td>
<td>2.24</td>
<td>3.20</td>
<td>2.79</td>
</tr>
<tr>
<td>SRS</td>
<td>1.43</td>
<td>1.83</td>
<td>3.95</td>
<td>2.02</td>
<td>3.04</td>
<td>2.69</td>
</tr>
</tbody>
</table>

due to the fact that demons is dependent on the quality of a mandatory affine initialization step. Qualitative results are shown in Fig. 3.11. We also evaluated the registration quality in a principal fashion by estimating target registration error (TRE) from landmarks. We used the approach presented in [Ozdemir et al., 2007] to automatically obtain geometric landmarks. The results for each method are given in Tab. 3.5, where it can be seen that SRS gives a slight but consistent improvement over the remaining single-atlas methods. In contrast to the previous 3D example, the multi-atlas fusion herein using 69 atlases outperforms the SRS method that uses a single atlas.

**Video Tracking Example.** The presented SRS approach is indeed applicable in other image processing settings as well. We demonstrate this on a facial-feature tracking example in Fig. 3.12. In the first frame of this 900x900 px² 24 frames/s RGB sequence, the eyebrows were manually annotated. From those, a Random Forests intensity classifier was trained with 100 trees and the negative logarithm of the normalized classification scores were used as segmentation unary potentials. An auxiliary label was obtained automatically as described in Sec. 3.5. Fig. 3.12 shows selected frames from this sequence overlaid with results from: (i) segmentation using only such unary potentials, (ii) initial frame segmentation propagated via registrations between consecutive frames, and (iii)
SRS applied between consecutive frames. Note that SRS uses the same learned classifier above, along with the segmentation of the previous frame as a reference in segmenting the next. The MRF-Seg results show that the information from the segmentation alone is insufficient to segment the structure and MRF-Reg shows the accumulation of errors through imprecise registrations. The target was segmented best in all frames using the proposed SRS approach with a maximal surface distance of 6 px in the last frame, compared to 9.5 px for MRF-REG.

Figure 3.10: Distribution of the results for mandibular bone (top) and CC (bottom) 110 experiments. We show the percentiles for each method, indicating the overlap in dice below which a certain percentage of the experiments fall.
3.6 Discussion

Our experiments demonstrate that SRS provides significantly better ($p \leq 1.24 \times 10^{-5}$ for all experiments and quality measures) results in comparison to ARS. An insight in the distribution of the errors can be obtained by inspecting Fig. 3.10, where it can be seen that especially on the mandible segmentation task the lower percentiles improve significantly over ARS and registration. It is notable that the correlation between the registration and ARS dice is significantly higher (0.74/0.86 for mandibles and CC, respectively) compared to the correlation of registration and SRS (0.4/0.66). This shows that SRS can leverage more information from the segmentation sub-problem, while ARS is more dependent on the registration results. Note that this correlations correspond to the peak performance of each algorithm in terms of parametrization. Therefore, while it may be possible to make ARS also less reliant on the registration, this will likely deteriorate the overall segmentation quality. Similarly to other non-rigid registration methods, both ARS and our SRS techniques are also initialized through a rigid pre-registration step, to minimize distances between anatomical structures that are to be non-rigidly registered. Accordingly,
Figure 3.12: Facial feature tracking example in a video sequence. The methods were applied on RGB images, which are shown in chroma here to overlay labels on top in color. Manual segmentation was provided in the initial frame for segmentation model training and propagation via registrations. The auxiliary label is colored blue, while left/right eyebrow are red/green. Segmentation cannot discriminate between features of similar appearance (top row). Sequential registration may accumulate errors (center row). SRS simultaneously segments and tracks the eyebrows correctly over the entire sequence (bottom row).
3.6. Discussion

the capture range of such a deformable registration can be perceived as the likelihood of reaching a satisfactory local optimum. This is indeed what we believe as one of the major reasons behind why SRS provides a significant improvement over ARS. Even though both methods have the same theoretical deformation range, the distance-based segmentation information provided to SRS by its coherence potential is seen to improve the fidelity of such achieved local optimum.

Formally, the computational complexity of our SRS implementation can be divided into three parts: (i) computation of registration unary potentials, and (ii) computation of distance transforms, (iii) solution of MRF using alpha-expansion graph-cuts. An upper bound for the first is $O(rg_a)$, where $r=|L_R|$ is the number of registration labels, $g$ is the number of nodes in the registration graph $G$ and $a$ is the size of the local support area defined by $\eta$. The computation time of the registration unaries was reduced considerably by using efficient similarity metrics such as local normalized cross correlation (LNCC) [Cachier et al., 2003], which uses convolution with Gaussian kernels. The distance transform of each segmentation label can be computed in linear time with respect to the number of pixels $N$ in the image, resulting in a complexity of $O(Ns)$, where $s=|L_S|$ is the number of segmentation labels. \(\alpha\)-expansion in general needs $O(lBi_\alpha)$ run-time, where $l$ is the number of labels, $B$ is the complexity of a binary graph-cut, and $i_\alpha$ is the number of required alpha-expansion iterations (typically about three). Since the binary graphs to be solved in SRS differ depending on whether a segmentation or a registration label is expanded, we describe the complexity as $O(i_\alpha(sB_S + rB_R))$, where subscripts $R$ and $S$ denote the registration and segmentation sub-domains, respectively. $B_S$ is $O(N^2C_S)$, where $C_S$ is the cost of the minimal cut and $B_R$ is analogously $O(g^2C_R)$. This needs to be solved multiple times for each multi-resolution level, where the last level will have the highest complexity. Therefore, an upper bound to the complexity of SRS is given by

$$O(m_{\text{max}}(rg_{\text{max}}a_{\text{max}} + s + i_\alpha(sN^2C_S + rg^2C_R)))$$

where $m$ is the number of multi-resolution levels, $i_{\text{max}}$ is the maximum number of iterations at that level, and $g_{\text{max}}$ and $a_{\text{max}}$ are the maximal sizes of, respectively, registration graph and local support area at the highest multi-resolution level.

Although neither [Mahapatra and Sun, 2012] nor [Parisot et al., 2012] performs SRS using an atlas segmentation, they both use discrete MRFs to solve simultaneous segmentation and registration problems. Therefore, below we compare our method to those from an algorithmic viewpoint, where the main distinction is their use of the full product label space $L_T \times L_S$. While this may facilitate a joint regularization of segmentation and registration [Mahapatra and Sun, 2012] or an adaptive thresholding of registration similarity [Parisot et al., 2012], it raises two main difficulties. First is the substantially larger label space. For example, dense displacement sampling in 2D with 7 samples per axis and ten segmentation labels yields $7^2 \cdot 10 = 490$ labels in the product label space, while only $7^2 + 10 = 59$ labels in the sum label space used by our SRS implementation. Since the complexity of discrete optimization algorithms depends on the number of labels
3. SIMULTANEOUS SEGMENTATION AND ATLAS REGISTRATION

(e.g. alpha expansions complexity is linear with the number of labels), this greatly effects computational efficiency. Second is the use of one single grid for both segmentation and registration. For the task of [Mahapatra and Sun, 2012] to improve fine registration of time-series images, a high resolution (pixel-level) control grid is used, requiring over 5 min per iteration for our 2D images, compared to ≈1 min runtime for our entire segmentation/registration. To avoid tearing, hard-constraints were used in [Mahapatra and Sun, 2012] such that the difference between neighboring displacements should be smaller than \(\sqrt{3}\) pixel. However, not only that folding can still occur when deformations from succeeding iterations are composed, but also a dense registration label space is required to allow sufficient deformation flexibility, which in turn would yield an infeasible number of labels for our application. For instance, for the same displacement range (60 pixel or 20 mm) as in our 2D experiments, \((2 \times 60 + 1)^2 = 14641\) registration labels would be needed. In contrast, [Parisot et al., 2012] uses the multi-resolution hierarchy of [Glocker et al., 2008], where segmentation labels are assigned at the same resolution as the registration grid. Therefore, at initial coarse levels to allow for large displacements, the coarse segmentation resolution does not allow small structures to be detected. We implemented this in our algorithm by fixing the segmentation grid resolution to the registration grid resolution. On the 2D dataset, this yielded a DICE coefficient of 0.8 which is significantly worse than any reported result in Tab. 3.4.

SRS using a fixed number of iterations is also faster compared to ARS. Indeed, the full SRS optimization runs under 2 min in 2D and 50 min in 3D on average with an ITK implementation in C++ on an Intel Core i7 1.6GHz processor. In comparison, one ARS iteration takes 0.5 min in 2D, 13 min in 3D and does often not converge in 10 iterations, causing it to be slower on average. Still, such run-time requirements can be seen as a limiting factor of our SRS method. However, other methods to achieve similar performance may require similar run-time in the form of multiple non-rigid registrations, on top of additional manual labor to generate a larger amount of prior knowledge. This is demonstrated in our comparison of SRS to multi-atlas fusion, which is the state-of-the-art in image segmentation [Rohlfing et al., 2005, Heckemann et al., 2006]. While multi-atlas fusion outperforms SRS on the 2D MRI data, 69 atlas segmentations were fused to segment one image. This amount of prior knowledge is hard to obtain in clinical scenarios. In contrast, on the 3D CT bone segmentation task, the amount of prior knowledge is more limited (14 fused segmentations per target). Here, it can be seen that SRS outperforms multi-atlas fusion with regard to volume overlap and mean surface distance, while being slightly less robust in terms of Hausdorff distance. This can be attributed to the difficult registration procedure caused by different anatomies and artifacts caused by dental implants on the one side, and intensity based bone segmentation being a well defined task on the other side. While multi-atlas fusion can only rely on the difficult registrations, SRS is able to successfully combine segmentation and registration information, which yields very good results despite using an order of magnitude less prior knowledge.
3.7 Conclusions

In this work, we have presented a novel approach for simultaneous registration and segmentation (SRS). We have employed a Bayesian derivation to obtain a formulation of the simultaneous model, which can be solved using discrete MRFs on a control grid hierarchy. The solution for each MRF is found efficiently by using alpha-expansion graph-cuts. Making use of multi-resolution non-rigid registration in conjunction with pixel-level segmentation, our method allows a precise, robust segmentation of target images using only a single atlas image. We show experimentally that SRS is more accurate and less dependent on the registration in comparison to an iterative optimization using the same criterion. Theoretically, our method estimates all parameters simultaneously using the proposed formulation in this paper. However, memory requirements make a hierarchical displacement refinement approach necessary, which is solved using a sequence of MRFs. In the future, we plan to address this implementation constraint. We also envision increased performance by using classifiers which can make use of context, e.g. local shape, contrast and texture.
A semi-supervised segmentation method using a single atlas is presented in this paper. Traditional atlas-based segmentation suffers from either a strong bias towards the selected atlas or the need for manual effort to create multiple atlas images. Similar to semi-supervised learning in computer vision, we study a method which exploits information contained in a set of unlabelled images by mutually registering them non-rigidly and propagating the single atlas segmentation over multiple such registration paths to each target. These multiple segmentation hypotheses are then fused by local weighting based on registration similarity. Our results on two datasets of different anatomies and image modalities, corpus callosum MR and mandible CT images, show a significant improvement in segmentation accuracy compared to traditional single atlas based segmentation. We also show that the bias towards the selected atlas is minimized using our method. Additionally, we devise a method for the selection of intermediate targets used for propagation, in order to reduce the number of necessary inter-target registrations without loss of final segmentation accuracy.

4.1 Introduction

Image segmentation is an essential problem in medical image processing. Among automatic segmentation methods, the amount of prior information needed is a major distinguishing characteristic of different approaches. It is desirable to limit the constraints posed by prior knowledge both to retain generalizability and to reduce the effort required
to acquire the information needed. Arguably, intensity-based approaches are among the methods that require the least amount of prior information. However, these methods are often susceptible to imaging artifacts, ambiguous intensities, and low contrast low signal-to-noise ratio imaging modalities, since no prior knowledge of shape or pose is assumed. Examples of such methods are active contours [Kass and Witkin, 1988] and MRF-based segmentation [Furnstahl et al., 2008, Boykov and Funke-Lea, 2006]. A straight-forward method for incorporating anatomical knowledge into automatic segmentation is the registration of an atlas image with known segmentation to a target image, namely *atlas based segmentation* [Glocker et al., 2008, Rueckert et al., 2006]. Then, the resulting transformation can be applied to the labelled atlas, yielding a segmentation of the target image. As the registration is ill-posed due to ambiguous and non-convex criteria [Fischer and Modersitzki, 2008], only approximate solutions can be achieved and these are influenced strongly by the choice of the atlas image [Rohlfing et al., 2005]. To remedy this, the use of multiple atlases is a common approach [Rohlfing et al., 2005, Heimann and Meinzer, 2009, Isgum et al., 2009, van Rikxoort et al., 2010, Jia et al., 2010, Rueckert and Frangi, 2001]: The information contained in such set of atlases can be used to create an average atlas [Jia et al., 2010]; to train statistical models of shape [Heimann and Meinzer, 2009] or deformations [Rueckert and Frangi, 2001]; or to register them individually to the target and fuse these segmentations afterward [Heckemann et al., 2006]. While all these methods improve segmentation accuracy compared to single atlas based segmentation, several studies indicate that the latter fusion of multiple deformed atlas segmentations is superior to registering an average atlas [Rohlfing et al., 2005, Isgum et al., 2009]. In contrast to increasing the amount of manual annotation, harnessing information from unlabelled data has been a major research focus in computer vision. To this end, semi-supervised learning [Chapelle et al., 2006] is an established framework which enables tasks like image classification in large and diverse databases [Wang et al., 2006b, Guillaumin et al., 2010]. For medical image segmentation, the situation is similar: large amounts of raw data are readily available while annotated data (atlases) are scarce.

In this paper, we study a method that propagates atlas segmentation labels via a graph of inter-target registrations. This is inspired by label propagation in semi-supervised learning [Zhu and Ghahramani, 2002] and similar to a method presented in [Heckemann et al., 2006], which used *indirect propagation* of atlas labels to validate multi-atlas segmentation results. In our framework, the traditional single atlas segmentation can be seen as direct, or *zero-hop* propagation. Multiple segmentation hypotheses per target can be obtained by allowing more *hops* via other target images, generating a different segmentation hypothesis for each path from the atlas to the target (c.f. Fig. 4.1b), which are then fused. In this paper, different strategies are studied for fusing such propagated labels for the segmentation of each target. Experiments with different anatomies and image modalities show significant improvement over the traditional atlas-based segmentation. In order for the method to scale successfully to larger datasets, we also investigate methods to reduce the
4.2 Label Propagation

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

$$S^1_{m,n} = T_{m,n}(S^0_m).$$  \hfill (4.1)

Figure 4.1: Illustration of standard atlas based segmentation (left) and our proposed method (right). In the latter, additional segmentation hypotheses are created by deforming the atlas segmentation multiple times along each path to each target. In a last step, these hypotheses are then fused to create the final segmentations.
These hypotheses must then be fused: We use a function $F$ to first generate a spatial segmentation probability map $\hat{S}_n = \Omega \rightarrow [0, 1]$ and subsequently binarize this using thresholding. The said probability map is generated as a weighted average of the zero- and one-hop segmentation hypotheses:

$$\hat{S}_n = F\left(S_0^n,\{S_1^{m,n}\mid m \neq n}\right) = \frac{1}{\sum \lambda} \left(\lambda_0^n S_0^n + \sum_{m \neq n} \lambda_1^{m,n} S_1^{m,n}\right). \quad (4.2)$$

We propose and evaluate two different strategies for the choice of weights $\lambda$:

**Global Similarity Weighting (GSW):** Assuming correlation between segmentation accuracy and a normalized post-registration similarity $f_G$, the latter can be used as a scalar weight. Using the zero-hop deformed atlas image $A_0^n = T_{A,n}(A)$, the zero-hop weight is then $\lambda_0^n = f_G(X_n, A_0^n)$. The atlas image is propagated analogously to the atlas segmentation along each path, i.e. $A_1^{m,n} = T_{m,n}(A_0^m)$, which leads to one-hop weights as follows:

$$\lambda_1^{m,n} = f_G(X_n, A_1^{m,n}) . \quad (4.3)$$

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

$$\lambda_1^{m,n}(p) = f_L(X_n(p), A_1^{m,n}(p)), \forall p \in \Omega . \quad (4.4)$$

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

### 4.3 Compact Graphs

To use all one-hop segmentation hypotheses, all graph connections should be computed requiring $N^2$ registrations. With large datasets this may easily become computationally challenging. Below, different methods are proposed for selecting a target image subset $X$, called *support-samples*, that will act as intermediate nodes via which the atlas segmentation is propagated using (4.1). Reducing the size $K = |X|$ then decrease the number of edges in $G$ and hence the number of necessary inter-target registrations. A natural choice is to sort nodes by their GSW-based zero-hop weights $\lambda_0^n$, as this corresponds to image similarity to deformed atlas, and to use the highest ranked images as support samples. We call this GSW-based ranking. This scheme makes two assumptions: first, that it is utmost important to propagate ‘good’ zero-hop segmentations; and second, that the quality of these segmentations can be assessed reliably using the similarity function $f$. Using our
label propagation framework, we propose the following two additional ranking criteria for selecting support samples.

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

$$\text{ARE}(m) = 1 - \text{Dice}(S_A, S_{m,A}^1),$$ (4.5)

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

$$\text{ARE-G}(S_A, \{m_1, \ldots, m_K\}) = 1 - \text{Dice}(S_A, F(S_{m_1,A}^1, \ldots, S_{m_K,A}^1)),$$ (4.6)

where $F$ is the fusion function in (4.2). As it is not feasible to evaluate all $\binom{N}{K}$ $K$-sized sets of support samples, we rely on a greedy scheme where we pick the first support sample based on its individual ARE, and iteratively add support samples that reduce ARE-G the most.

### 4.4 Results and Discussion

We evaluated our method using a set of 70 mid-saggital slices of MR brain scans containing the corpus callosum and a set of 15 3D CT scans of the head. Both datasets were rigidly pre-aligned. In the MR dataset a fixed region of interest containing the corpus callosum was cropped out in all images. The images are 120x200 pixels with 0.3 mm spacing in the MR dataset and 160x160x129 voxels with 1 mm spacing in the CT dataset. We used the Dice coefficient to measure volume overlap, the Hausdorff distance (HD) for estimating maximum surface-to-surface distance, and the mean surface distance (MSD) as an additional distance based metric. In both datasets we performed a leave-one-out (l1o) evaluation scheme, using each image as atlas in turn to segment all remaining images. We used our own implementation of the MRF-based registration in [Glocker et al., 2008] as the registration method of choice in our experiments. We used normalized cross
correlation (NCC) as the registration similarity criterion throughout our experiments, and accordingly defined GSW weights using \( f_0 = \frac{1 - \text{NCC}^2}{2} \). Since NCC is not suitable as a pointwise metric for LAW, we used the following intensity difference based radial basis function

\[
\lambda_{m,n}^l(p) = \exp \left( -\frac{|X_n(p) - A_{m,n}(p)|}{\sigma^2} \right)
\]

where \( \sigma^2 \) is the intensity variance over all images. Sample results are given in Fig. 4.2.

**Quantitative results** can be found in Tab. 4.1. For both datasets, the improvement over traditional atlas based segmentation is significant, especially considering the distance based metrics which are outperformed by \( \approx 35% \) (HD) and \( \approx 60% \) (MSD). The CT images also show a strong improvement in Dice similarity metric, and LAW expectedly performed superior to GSW. GSW only slightly outperforms an uniform weighting, which results in a simple max-voting scheme. We also compared our method to a recent group-wise registration approach (ABSORB [Jia et al., 2010]) a 3D implementation of which is publicly available. This method uses the atlas as a reference image, on which all test images are aligned iteratively to improve a mean image. Similarly to ours, this method also utilizes image information contained in the entire dataset. However, unlike the probabilistic map in our method, ABSORB generates a single binary segmentation per target and thus suffers from a bias towards the atlas similarly to the traditional atlas based segmentation. This is seen in the results as it is outperformed by our proposed method. In order to estimate an upper bound for the expected performance of our method, we also computed multi-atlas segmentation of each target by using all remaining images and their ground-truth segmentations as multiple atlases. We used the same registration method and LAW weighting to achieve comparable results. As seen in Tab. 4.1, even though our method did not reach the performance of such multi-atlas segmentation, it performed remarkably close to it while using orders of magnitude less prior knowledge.
4.4. RESULTS AND DISCUSSION

Table 4.1: Mean segmentation accuracy from leave-one-out evaluation on the MR (2D) and CT (3D) datasets.

<table>
<thead>
<tr>
<th>Method</th>
<th>DICE HD</th>
<th>MSD</th>
<th>DICE HD</th>
<th>MSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-atlas [Glocker et al., 2008]</td>
<td>0.926</td>
<td>2.45</td>
<td>0.088</td>
<td>0.828</td>
</tr>
<tr>
<td>Propagation-maxvote</td>
<td>0.944</td>
<td>1.50</td>
<td>0.027</td>
<td>0.852</td>
</tr>
<tr>
<td>Propagation-GSW</td>
<td>0.944</td>
<td>1.49</td>
<td>0.027</td>
<td>0.859</td>
</tr>
<tr>
<td>Propagation-LAW</td>
<td><strong>0.946</strong></td>
<td>1.46</td>
<td>0.025</td>
<td><strong>0.886</strong></td>
</tr>
<tr>
<td>ABSORB [Jia et al., 2010]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.698</td>
</tr>
<tr>
<td>Multi-atlas-LAW</td>
<td>0.965</td>
<td>1.06</td>
<td>0.014</td>
<td>0.917</td>
</tr>
</tbody>
</table>

In Fig. 4.3, the Dice measure is plotted for each atlas/target pair of our l1o-experiments. For both datasets, it is seen that using traditional atlas based segmentation some atlases lead to sub-optimal segmentations. Using our method, however, these sub-optimal pairs seen as ‘speckles’ disappear, indicating an improved performance for arbitrary atlas selection.

Support Sample Reduction: We first evaluated a random selection of \( K \) support samples for each target and repeated this 10 times each. As shown in Fig. 4.4(left), Dice metric improves rapidly until reaching 20 support samples, which is also the point at which the standard deviation becomes negligible. We then used the GSW-based ranking, which did not provide any improvement over random draws. Individual ARE based ranking expectedly outperformed random draws, and the group-wise ARE ranking was superior to all other support sample selection methods. Interestingly, the results indicate the presence of support sample subsets that can perform better in comparison to using all the samples (the full graph). We also analyzed the progression of error during the expansion of the support-sample subset. As seen in Fig. 4.4(right), error using ARE-G based ranking starts
deteriorating beyond a certain number of subset size. This number is also near the optimal Dice metric shown in Fig. 4.4(left) as the posterior target segmentation accuracy. Based on this observation, we propose to use the deflecting point (minimum value) of ARE-G to determine optimal graph size. Note that individual ARE does not exhibit such a behaviour. Accordingly, we have repeated the l1o-experiments on the MR dataset using only 20 support samples selected by ARE-G ranking. This yielded mean Dice of 94% and mean HD of 1.56 mm, which are nearly identical to the results using all target images while requiring less than three times the number of intermediate target images and their registrations.

Discussion: Our method was shown to increase segmentation accuracy substantially compared to the standard atlas based segmentation. This can be attributed to the boosting nature of the approach, which can be seen as creating and fusing multiple weak classifiers in a semi-supervised manner. Our results being close to that of multi-atlas segmentation, we conclude that a substantial amount of the information contained in a set of atlas images is indeed available in the support (target) images, and that is the information leveraged by our method. Note that this contradicts partially with the findings of [Heckemann et al., 2006], which concludes that the major benefit of multi-atlas segmentation is due to the increase in anatomical variation in the available ground-truth. We believe that this difference in findings might be due to different registration algorithms and datasets, which will be important to explore in the future work. Additionally, it will be interesting to explore whether a principled, probabilistic approach can be employed to also take into account the improved segmentations in an iterative manner. A similar method was proposed in [Riklin-Raviv et al., 2010] for images aligned to a single template, whereas we aim to include information from inter-target registrations as well. For the CT images, we used the probabilistic segmentation output as shape prior in an MRF-based segmen-
4.5 Conclusions

We presented a novel method that augments single-atlas based segmentation using multiple segmentation hypotheses for each target obtained by propagating atlas segmentation along different paths. This outperforms both the traditional single-atlas based registration and group-wise registration. We also demonstrated that a smaller set of support samples providing complementary information can be found automatically. Using such reduced set of support samples both decreases the computational complexity of the method and improves the results as redundant and possibly detrimental information is then discarded.
5

Registration Fusion using Markov Random Fields

Image registration is a ubiquitous technique in medical imaging. However, finding correspondences reliably between images is a difficult task since the registration problem is ill-posed and registration algorithms are only capable of finding local optima. This makes it challenging to find a suitable registration method and parametrization for a specific application. To alleviate such problems, multiple registrations can be fused which is typically done by weighted averaging, which is sensitive to outliers and can not guarantee that registrations improve. In contrast, in this work we present a Markov random field based technique which fuses registrations by explicitly minimizing local dissimilarities of deformed source and target image, while penalizing non-smooth deformations. We additionally propose a registration propagation technique which combines multiple registration hypotheses which are obtained from different indirect paths in a set of mutually registered images. Our fused registrations are experimentally shown to improve pair-wise correspondences in terms of average deformation error (ADE) and target registration error (TRE) as well as improving post-registration segmentation overlap.

5.1 Introduction

Image registration is at the core of a multitude of medical image analysis techniques. It enables ubiquitous methods such as atlas generation, atlas-based segmentation, multi-modal information fusion, pre-operative planning and intra-operative guidance. High registration reliability and accuracy are critical for the success of any applied method. However, it is well known that the registration problem is ill-posed [Fischer and Modersitzki, 2008] when anatomical correspondences are established using image similarity. Artefacts from the imaging process (noise, bias) and from discretizing the continuous spatial domain into

pixels/voxels further complicate the image registration process. Numerous techniques have been developed to cope with these challenges. Firstmost, registration algorithms typically include regularization terms that penalize anatomically unreasonable deformations such as due to folding or tearing. For example, Markov random field (MRF) based registration penalizes deformations with large Euclidean distance between displacements of neighboring points [Glocker et al., 2008]. Other approaches include Gaussian smoothing of velocity and update fields [Thirion, 1998], or they restrict the type of deformation to be diffeomorphic (smoothly invertable) [Vercauteren et al., 2009]. A comprehensive overview of registration regularization is given in [Cachier and Ayache, 2001]. Higher-level prior knowledge can also be used to regularize registration algorithm. For example, it is reasonable to expect that the outcome of registering image A to B should be the same as inverting the registration from B to A. Such registration symmetry has been successfully enforced for example in the demons [Vercauteren et al., 2008] and SyN registration algorithms [Avants et al., 2008a]. One key challenge of regularized registration is that the resulting elastic matching problem is NP-complete [Keysers and Unger, 2003] and therefore solutions can be found only approximatively, typically following the gradient of the objective function. To avoid poor local optima in such optimizations, most non-rigid registration methods require a suitable initialization from another robust rigid or affine registration algorithm. Multi-resolution pyramids are frequently used to prevent algorithms from falling into local minima due to fine detail in early stages of the registration [Rueckert et al., 1999]. Despite these advances, medical image registration is still an open problem.

It is known from information theory that many weak information sources can be combined into strong information for example via boosting [Schapire, 1990]. This is reflected by the trend in medical image segmentation, where state-of-the art results are achieved by fusing segmentation hypotheses obtained from multiple atlases [Heckemann et al., 2006]. Optical flow can be computed by fusing multiple hypotheses using a mixed continuous-discrete approach [Lempitsky et al., 2008]. There also exist studies that present methods to fuse multiple hypotheses for landmark detection [Iglesias and Karssemeijer, 2009, Santemiz et al., 2013], which can be seen as a sparse registration problem. Some of the fusion methods are specific to the chosen detection method, for example choosing landmarks based on the cumulative score of multiple support vector machines that are independently trained [Santemiz et al., 2013]. A more general technique was presented in [Iglesias and Karssemeijer, 2009], where multiple landmark hypotheses obtained from different atlases were fused using similarity weighted averaging. Such averaging can be also applied to dense registration fusion, but has two drawbacks: First, averaging can give no guarantee that the result improves the registration fidelity, or that it even improves any image-to-image similarity. Second, the average may easily result in anatomically unreasonable deformations since there is no straight-forward way to regularize the result.

In this paper, we formulate the registration fusion as an optimization problem where the objective function is designed such that post-registration image similarity is maximized.
5.2. Registration Fusion using an MRF

Let image $X$ be a function that maps points in the $D$-dimensional spatial domain $\Omega$ to a space $F$ of image intensities, e.g. CT Hounsfield units. A non-rigid deformation $T$ is a mapping from $\Omega$ to $\Omega$ which is based on a displacement field $D$ such that $T(p) = p + D(p)$ for all points $p$ in $\Omega$. Note that both $X$ and $D$ are commonly defined on a discrete (Cartesian) regular grid, where non-grid values can be obtained by interpolation. In this notation, deforming an image is a function composition $X \circ T = X(T)$. A set of registrations $R(X, Y) = \{T_1, \ldots, T_N\}$ may contain $N$ different registrations between images $X$ and $Y$. The goal of registration fusion is then to combine such registrations into a new registration estimate. Ideally, such fused registrations should maximize post-registration image similarity while penalizing anatomically unreasonable deformations.
We propose a discrete optimization method which locally selects one displacement of \(N\) hypotheses based on local image similarity and global registration smoothness (Fig. 5.1). To this end, we define a discrete label domain \(L = \{1, \ldots, N\}\), such that each label \(l\) uniquely refers to a registration \(T_l \in \mathcal{R}(X, Y)\). Then, a labelling \(L : \Omega \to L\) assigns a label \(l_p\) to every pixel/voxel \(p\) of an image. Finally, such labelling defines the corresponding registration \(T_L\) as follows:

\[
T_L : \Omega \to \Omega : p \to T_{l_p}(p).
\] (5.1)

This allows us to define the discrete optimization problem as energy minimization in a first-order Markov random field (MRF), for which efficient solvers exist. The MRF energy is commonly defined as a sum of potential functions defined on a graph which typically is the lattice of pixels/voxels in computer vision. The goal of the MRF is then to assign a label \(l_p\) to each graph node \(p\) such that a energy-criterion is minimized. The energy that we minimize for our registration fusion problem can then be formalized as follows:

\[
\hat{L} = \arg\min_{L'} \sum_{p \in \Omega} \left( V_p(l_p) + \lambda \sum_{q \in \mathcal{N}(p)} V_{pq}(l_p, l_q) \right),
\] (5.2)

where the unary term \(V_p(l_p)\) is the data fitness at point \(p\) for label \(l_p\), and \(V_{pq}\) is a pairwise energy term which allows implementing prior knowledge about the spatial smoothness of the solution, and \(\lambda\) is the weighting for the latter.

**Unary Term** \(V_p\). This term penalizes choosing labels that result in low similarity \(\Phi\) of image \(X\) and deformed image \(Y(T_{l_p})\) at location \(p\) as follows:

\[
V_p(l_p) = 1 - \Phi(X, Y(T_{l_p}), p).
\] (5.3)

We use local normalized cross correlation (LNCC) [Cachier et al., 2003] as a robust similarity metric, which was successfully utilized earlier to locally rank segmentation hypotheses in [Cardoso et al., 2012]. LNCC is smooth, and can be efficiently computed using convolutions with Gaussian kernels of size \(\sigma\). From the LNCC metric, we compute the similarity

\[
\Phi(X, Y(T_{l_p}), p) = \left(1 + \frac{\text{LNCC}_\sigma(X, Y(T_{l_p}), p)}{2}\right)^\gamma,
\] (5.4)

which normalizes LNCC to the range \([0, 1]\). \(\gamma\) is used to scale the similarity such that contributions from individual registrations are well spread [Iglesias and Karssemeijer, 2009]. An example of such unary potentials is shown in Fig. 5.2.
5.2. Registration Fusion using an MRF

Figure 5.2: Unary potential computation for one label $l$, $\sigma=4$ and $\gamma=10$.

**Pair-wise Term** $V_{pq}$. We penalize non-smooth deformations based on squared Euclidean distance between the displacements of neighboring points:

$$V_{pq}(l_p, l_q) = ||\mathcal{T}_{l_p}(p) - \mathcal{T}_{l_q}(q)||^2.$$  \hfill (5.5)

**Solving the MRF.** Although the number of labels is typically relatively small, the underlying MRF graph can become huge, which makes solving (5.2) on the lattice of image pixels prohibitive. Since registrations can be represented on a coarse grid of control points [Rueckert et al., 1999], we define our MRF over a coarse graph $\mathcal{G}$, and interpolate displacements at non-grid locations using B-splines after $\hat{R}$ is estimated. Since there is no guarantee that the pairwise potential in (5.5) satisfies any metricity or submodularity criteria, it is not possible to use common discrete optimization methods such as $\alpha$-expansion [Kolmogorov and Zabih, 2004]. We thus use tree-reweighted message passing TRW-S, which allows using arbitrary pairwise potentials [Kolmogorov, 2006] and additionally gives a lower bound on the energy which can be used to assess the quality of the solution. If the energy of the solution is equal to the lower bound, TRW-S arrives at the globally optimal solution.

5.2.1 MR\textsubscript{egF}use with Folding Removal (MR\textsubscript{egF}useR)

Since labels and hence deformations are selected in a locally discrete fashion, displacement continuity at label seams cannot be guaranteed. We therefore propose the following post-processing step to remove folding if the minimum of the Jacobian determinant, $\min(|J|)$, of a deformation field $\hat{T}$ is smaller than zero: $\hat{T}$ is smoothed with $K$ Gaussian kernels each of size $\sigma_k = 0.5 \cdot 2^k$ mm, resulting in $K$ deformation fields $\hat{T}_{\sigma_k}$. $K$ is increased sequentially until $\hat{T}_{\sigma_K}$ has a positive $\min(|J|)$. We then fuse the resulting set of regis-
Figure 5.3: Generating registration hypotheses by composing deformations along indirect paths in a population-based setting (c.f. Sec. 5.2.2).

\[
\{\hat{T}, \hat{T}_{\sigma_1}, \ldots, \hat{T}_{\sigma_K}\}
\]

using MR\textsubscript{eg}Fuse as before. If the \(\min(|J|)\) is still negative, we repeat the fusion step with an increased pair-wise potential weight \(\lambda \leftarrow 2\lambda\) until folding is removed. To avoid over-smoothing the result and to increase computational efficiency, we update any labels for folding removal solely for nodes that are in the vicinity (3\(\sigma\)) of negative Jacobian determinants.

5.2.2 Pair-wise and Population-Based Registration Fusion

For a given pair of images, the proposed MR\textsubscript{eg}Fuse can fuse multiple registration hypotheses obtained from running different registration algorithms or parametrizations thereof. This can, for example, be useful when a universally optimal registration algorithm/parametrization is difficult to find or does not exist.

MR\textsubscript{eg}Fuse can also be used when a population of images is to be mutually registered. Here, it is possible to obtain multiple registration hypotheses by composing deformations along indirect paths between images. This principle was used successfully to generate and fuse segmentation hypotheses in [Heckemann et al., 2006, Gass et al., 2012]. Formally, if a registration \(T_{ij}\) registers images \(X_i\) and \(X_j\), an indirect registration \(T_{ikj}\) is the composition of \(T_{ik}\) and \(T_{kj}\), i.e. \(T_{ikj} = T_{kj} \circ T_{ik}\). The set of registration hypotheses is then \(\mathcal{R}(X_i, X_j) = \{T_{ij}, T_{ik}, T_{ilj}, \ldots\}\), which we illustrate in Fig. 5.3. For a set of \(N\) images, it is therefore possible to obtain \(N-1\) registration hypotheses per image pair without explicitly re-computing registrations. It is also possible to iteratively improve the results by using the fused pairwise registrations themselves to compute new hypotheses, which can then again be fused. We perform this iterative improvement until the average post-registration image similarity stops improving significantly.
5.3 Experimental Results

We evaluate our method in two experiments on two datasets. First, we fuse registrations obtained from running different registration algorithms with several different parametrizations each. Then, we evaluate the performance of MR\textsubscript{reg}\textsubscript{Fuse} in a group-wise registration approach.

**Data Sets.** To evaluate the accuracy of dense registrations, we generated a dataset of medical images with known dense correspondences as follows: We used 19 mid-saggital slices of brain MRI with $481 \times 374 \text{px}^2$ resolution and 0.3 mm spacing. We first registered one randomly chosen image to the remaining 18 images using Markov-random field (MRF) based registration [Glocker et al., 2008]. The computed registrations were then used to deform the source image, and the 18 deformed images plus the source image were used as a new dataset. This dataset then exhibits true anatomical variability, and ground-truth registrations can be inferred from the deformation process.

The second dataset consists of 15 clinical 3D CT scans of the head of different individuals, with $160 \times 160 \times 129 \text{px}^3$ resolution and unit spacing. The presence and the number of teeth vary substantially in this dataset. To evaluate registration accuracy, landmarks were placed at anatomically identifiable locations on the jawbone and the skull for all images. Additionally, the jawbone was manually segmented by experts, which allows us to also report post-registration segmentation overlap for this dataset.

**Pair-wise Registration Fusion.** In this experiment, we evaluated the fusion of multiple independently computed pair-wise registrations. This is a relevant scenario, as it is frequently not known a-priori which registration algorithm with which parametrization would be best suited for a given registration problem. We registered each image pair of the

---

### Table 5.1: Results using independently computed registrations and their fusion. Different parameter sets $p$ were obtained for each registration method, and the resulting registrations then fused using globally and locally weighted averaging (GW A and LW A) and our proposed MR\textsubscript{reg}\textsubscript{Fuse} and MR\textsubscript{reg}\textsubscript{Fuse}R. mJD denotes the minimum Jacobian determinant of all pair-wise registrations, while ADE and NCC are averaged over the set.

<table>
<thead>
<tr>
<th>Parameter Sets</th>
<th>Demons</th>
<th>MRF-reg</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{d_1}$</td>
<td>$p_{d_2}$</td>
<td>$p_{d_3}$</td>
<td>$p_{d_4}$</td>
</tr>
<tr>
<td>ADE</td>
<td>4.00</td>
<td>4.44</td>
<td>3.50</td>
</tr>
<tr>
<td>NCC</td>
<td>0.77</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td>mJD</td>
<td>0.64</td>
<td>-165</td>
<td>0.05</td>
</tr>
</tbody>
</table>
REGISTRATION FUSION USING MARKOV RANDOM FIELDS

Figure 5.4: Example fusion result for independently computed registrations for the 2D MR image pair shown in Fig. 5.2. The small images on the left side show the local deformation error magnitude of the computed registration with regard to the known ground-truth (log-scale). The frame of each image is colored such that the resulting MRF labelling (center) locally selects one of the registrations based on its color. The right-most image then shows the deformation error magnitudes of the fused registration.

2D MR dataset multiple times using different parametrizations of well-known demons- and MRF-based registrations [Vercauteren et al., 2009, Glocker et al., 2008]. For each image pair, we fuse all available registrations using globally and locally weighted averaging (GWA and LWA) in addition to the proposed MR\text{Fuse}. For each method we report the average (dense) deformation error (ADE), average post-registration image similarity (NCC), and the minimum of the local determinants of the Jacobian, which indicates folding [Christensen et al., 1995]. The free parameters were optimized for each fusion method separately, yielding $\gamma=20$ for GWA, $\gamma=16$ and $\sigma=8$ for LWA, and $\gamma=10$, $\sigma=4$ and $\lambda=0.02$ for our MR\text{Fuse}. For our MRF fusion, we used MRFs with 4 px spacing, resulting in control grids of size $121 \times 94$ in 2D. Results are reported in Tab. 5.1. MR\text{Fuse} is seen to achieve lower ADE compared to both averaging-based registration fusion techniques, meanwhile leading to less folding compared to locally weighted averaging. Note that globally weighted averaging leads to no folding overall, as the initial registrations are diffeomorphic. However, it does not yield much improvement over the best individual registration. On the other hand, LWA, which is seen to improve ADE over all input registrations, meanwhile causes strong image folding as spatial smoothness of the deformations is not considered. MR\text{Fuse}, in contrast, leverages such smoothness information to find an optimal partitioning of input registrations. In Tab. 5.1, it is also seen that the proposed MR\text{FuseR} is able to remove folding completely, as shown exemplary in Fig. 5.5, without deteriorating registration accuracy as indicated by ADE and average NCC. Jacobian determinants before and after two such sample folding removals are seen in Fig. 5.5.

Population-Based Registration Fusion. We also evaluated the performance of MR\text{Fuse} when registration hypotheses are obtained from indirect paths in a population of images. For the 2D MR dataset, we chose the best performing parametrizations from the previous experiment. For the fusion, we then used the same parameters as in the
5.3. Experimental Results

Figure 5.5: Two registration examples showing the Jacobian determinant values for MRegFuse and MRegFuseR, with little folding (left) and substantial folding (right). The left example also corresponds to the images/results presented both in Figs. 5.2 and 5.4. In both examples, MRegFuseR is seen to completely remove folding.

Table 5.2: Fusing registrations obtained by automatic hypothesis generation in sets of mutually registered images. mJD denotes the minimum Jacobian determinant of all pairwise registrations, while ADE, NCC, TRE and Dice are averaged over the set.

<table>
<thead>
<tr>
<th></th>
<th>2D MRI</th>
<th></th>
<th>3D CT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Demons</td>
<td>MRF</td>
<td>Demons</td>
<td>MRF</td>
</tr>
<tr>
<td>Registration</td>
<td>ADE</td>
<td>NCC</td>
<td>mJD</td>
<td>ADE</td>
</tr>
<tr>
<td>LWA</td>
<td>3.5</td>
<td>0.84</td>
<td>0.05</td>
<td>1.92</td>
</tr>
<tr>
<td>+ MRegFuse</td>
<td>2.34</td>
<td>0.94</td>
<td>-65.1</td>
<td>1.83</td>
</tr>
<tr>
<td>+ MRegFuseR</td>
<td>0.78</td>
<td>0.98</td>
<td>-2.29</td>
<td>0.94</td>
</tr>
</tbody>
</table>

experiment above, except for \( \lambda \), which was changed to 0.002 in the 3D experiment to account for the larger physical image spacing. The MRF grid-size was the same as above for the 2D data and was set to \( 40 \times 40 \times 33 \) in 3D using the same grid spacing of 4 px. Results on both datasets are reported in Tab. 5.2. MRegFuse results in lower ADE and TRE compared to averaging-based registration fusion, while generating less folding. For the 3D CT dataset, we also report pairwise post-registration segmentation overlap which was computed using Dice’s coefficient. We also compare the results to a typical group-wise registration method, in which all images of a set are iteratively registered to an evolving mean image using residual complexity [Myronenko and Song, 2010, Myronenko, 2013]. From the registrations to this mean, we compute pair-wise registrations by inversion and composition. The resulting ADE is then 1.84 mm for the 2D MR dataset, and TRE is 8.1 mm in 3D CT, which are inferior to MRegFuse results. MRegFuseR is seen to successfully remove folding.
Figure 5.6: Left: MR\textsubscript{Fuse} with different regularization weights $\lambda$ on the pair-wise fusion experiment for the 2D MR dataset. Right: Histogram of local Jacobian determinants for LWA, MR\textsubscript{Fuse} and MR\textsubscript{FuseR} for all pair-wise registrations on the 2D data.

5.3.1 Discussion

MR\textsubscript{Fuse} can be seen to result in lower ADE/TRE in all experiments compared to averaging-based fusion. The latter can only improve results when local weights are used, which cannot guarantee that the resulting deformation is smooth even when using large-kernel LNCC weighting. In contrast, the optimization of MR\textsubscript{Fuse} allows for a trade-off between smooth deformations and local similarity as shown in Fig. 5.6(left).

We are also very interested in exploring the robustness of MR\textsubscript{Fuse} with regard to the number and quality of registrations to be fused. As seen in our experimental validation, MR\textsubscript{Fuse} is more robust with regard to poor input registrations in comparison to averaging-based fusion. One downside of MR\textsubscript{Fuse} is that it cannot achieve a better performance locally than the best of its the input registrations due to its selection-based approach. However, the inputs can be augmented with additional registration hypotheses such as the (weighted) mean of the registrations.

Our MRF fusion is also fast and memory efficient. Average run-time of the MRF-solver for fusing 14 registrations is 30.5 seconds in 3D, while requiring less than 800 MB memory. The local similarity computation is also efficient. The entire process of generating indirect deformations in the group-wise experiment, computing all local similarities, and fusion using the MRF was completed in under two minutes in all cases. Consequently, finer resolution grids or more registrations/parametrizations can be fused in reasonable timeframes, which we will explore in the future.

The employed iterative TRW-S optimizer was stopped when the relative change of the lower bound was small ($1 \times 10^{-7}$) or a maximum number of iterations (1000) was reached. We observed that TRW-S was able to find the globally optimal solution in about 69% of the cases for the 2D pair-wise fusion experiment, and the final energy being within 0.5% of the lower bound in all remaining cases of that experiment. For the 3D experiments, the final energy was within 0.03% of the lower bound in all cases.
As can be seen from the experimental results, the fused registrations still exhibit folding as indicated by the negative Jacobian determinants. However, both the minima presented in the tables, as well as the distribution of all such values that can be seen in Fig. 5.6(right) show that the folding is less severe compared to LWA-fusion. Additionally, the proposed \( \text{MR}_{\text{regFuse}} \) algorithm can be used to remove folding by fusing the original \( \text{MR}_{\text{regFuse}} \) result with smoothed versions of itself. This process is seen to retain the high registration accuracy of \( \text{MR}_{\text{regFuse}} \), even with potential (minor) ADE/TRE improvements as seen in the population-based demons experiments in Tab. 5.2. Using the described kernel pyramid, the folding removal in 3D examples needed on average \( K=4.4 \) additional smoothed deformation fields and 12 iterations to remove folding entirely. Since the number of graph labels during folding removal is significantly smaller compared to the \( \text{MR}_{\text{regFuse}} \) step, the random field is solved much quicker at 1.5 s on average. The folding removal adds less than one minute on average to the registration fusion process. Potentially, such folding removal can be applied as a post-processing step to any other registration method that may yield non-positive Jacobian determinants.

## 5.4 Conclusions

We have presented a novel algorithm for registration fusion, which locally selects displacements from multiple registration hypotheses. An MRF is used to find an optimal solution in terms of local post-registration similarity and registration smoothness. This algorithm can furthermore be used to automatically remove inevitable folding at label seams. The proposed \( \text{MR}_{\text{regFuse}} \) algorithm has been shown to be able to fuse registrations from random parametrizations of different registration algorithms such that the fused registrations exhibit higher accuracy than any of the input registrations. This is a relevant scenario for many image registration applications, where it is typically challenging to choose and parametrize a registration method correctly. We have also demonstrated that registration hypotheses can be obtained automatically within a set of mutually registered images, which allows improving the average pair-wise registration accuracy by using the proposed registration fusion.
Consistency-Based Rectification of Non-Rigid Registrations

In this paper we present a novel technique to rectify non-rigid registrations by improving their group-wise consistency, which is a widely-used unsupervised measure to assess pair-wise registration quality. While pair-wise registration methods cannot guarantee any group-wise consistency, group-wise approaches typically enforce perfect consistency by registering all images to a common reference. However, errors in individual registrations to the reference then propagate, distorting the mean and accumulating in the pair-wise registrations inferred via the reference. Furthermore, the assumption that perfect correspondences exist is not always true, e.g. for inter-patient registration. The proposed consistency-based registration rectification (CBRR) method addresses these issues by minimizing the group-wise inconsistency of all pair-wise registrations using a regularized least-squares algorithm. The regularization controls the adherence to the original registration, which is additionally weighted by the local post-registration similarity. This allows CBRR to adaptively improve consistency while locally preserving accurate pair-wise registrations. We show that the resulting registrations are not only more consistent, but also have lower average transformation error when compared to known transformations in simulated data. On clinical data, we show improvements of up to 50% TRE in breathing motion estimation from 4D MRI, and improvements in atlas-based segmentation quality of up to 65% in terms of mean surface distance in 3D CT. Such improvement was observed consistently using different registration algorithms, dimensionality (2D/3D) and modalities (MRI/CT).

6.1 Introduction

Image registration is an essential tool in medical image analysis. It enables applications such as atlas-based segmentation [Rohlfing et al., 2005], statistical model building, and automatic landmark detection. In many of these tasks, high registration accuracy is critical, but difficult to achieve since the registration problem is known to be ill-posed [Fischer and Modersitzki, 2008] with anatomical correspondences estimated merely using visual similarity. Regularization terms based on global smoothness assumptions are commonly utilized to facilitate anatomically reasonable transformations. However, the resulting elastic matching problem is very difficult, and its approximations lead to registration algorithms which are only locally optimal. Techniques such as multi-resolution registration are widely employed to reduce the likelihood of poor local optima [Lee et al., 1997].

There have been several studies that aimed at detecting registration inaccuracies or errors posterior to registration. Some of these were based on intensity measures, e.g. using the multiple Gaussian state-space [Crum et al., 2004], voxel-statistics based on active appearance models from registered images [Schestowitz et al., 2006] or supervised learning [Muenzing et al., 2012b, Lotfi et al., 2013]. However, it was shown that such voxel similarity metrics are indeed not suitable indicators of registration quality due to homogeneous tissue regions, partial volume effects, and anatomies of similar appearance [Crum et al., 2003, Rohlfing, 2012]. While independent labelling of landmarks or anatomic regions can be used to measure registration fidelity, this requires invaluable time and effort of trained medical personnel, which an automatic registration is supposed to avoid in the first place. To estimate the uncertainty in single pair-wise registrations, different methods such as the Cramér-Rao bound [Robinson and Milanfar, 2004], bootstrap resampling [Kybic, 2010] or Markov Chain Monte-Carlo (MCMC) [Risholm et al., 2013] have been proposed, often increasing the complexity of a registration algorithm by orders of magnitude. Recently, such uncertainty estimations have been utilized to improve registrations by adaptively estimating an optimal regularization weight [Simpson et al., 2013] or by re-registering uncertain regions in a boosting framework [Muenzing et al., 2012a].

Moving from pairs of images to groups, it is possible to investigate the consistency of multiple registrations as a measure of registration fidelity [Pennec and Thirion, 1997, Woods et al., 1998, Holden et al., 2000, Christensen and Johnson, 2001, Rohlfing, 2012, Datteri and Dawant, 2012]. The most common application of this has been the evaluation and comparison of registration algorithms by computing the residual norm of inconsistencies in pair-wise registrations [Woods et al., 1998, Rohlfing, 2012]. Recently, a method was proposed to exploit redundancy in multiple registration circles for estimating the spatial location and magnitude of errors in pair-wise registrations [Datteri and Dawant, 2012]. Using a simplified model for the accumulation of registration errors, this method can estimate a dimensionless measure for local registration error, which was shown therein to correlate with true errors. A similar method was also used to predict the segmentation...
quality for multi-atlas segmentations [Datteri et al., 2011]. This was also shown in [Goksel et al., 2013] to be able to leverage additional information from unlabelled images.

In contrast to registration error detection, registration consistency can also directly be incorporated into a registration algorithm. Several algorithms optimize a symmetric energy term, which adheres to the inverse consistency described in [Christensen and Johnson, 2001]. Noteable examples are symmetric log-domain demons [Vercauteren et al., 2008] and symmetric normalization [Avants et al., 2008b]. Group-wise registration can naturally incorporate consistency criteria as well. Here, images are commonly registered to a reference image [Joshi et al., 2004] and pair-wise registrations can then be computed, if the individual registrations are invertible [Skrinjar et al., 2008]. To avoid bias to a specific image, an averaged intensity image is used as the group mean, on which all other images are then registered. Using the registered images, a new (improved) mean is then computed, and the process is repeated iteratively. the group mean image is typically used as a reference, which is then refined iteratively. In order to ensure that the group mean is in a natural coordinate system, the sum of the registrations can additionally be constrained to be zero [Bhatia and Hajnal, 2004]. A transitive registration method for triplets of manifolds was presented in [Geng et al., 2005]. The described algorithm was extended to larger groups of images by a hierarchical clustering approach, where first the images within a cluster were registered transitively and the cluster means were registered to each other [Geng, 2007].

In this paper, we introduce a novel method that rectifies a given set of pairwise transformations within a group of images. Importantly, our algorithm allows for rectifying non-rigid registrations, in contrast to a previous method for rigid registrations [Glatard et al., 2006]. Our method extends an approach [Datteri and Dawant, 2012] presented earlier, which computes dimensionless estimates of non-rigid registration errors by optimizing a surrogate model of inconsistency in mutually registered sets of images. In contrast, our work directly minimizes actual registration inconsistencies within registrations. This then estimates error vectors which can be used to rectify the given registrations which is not possible by [Datteri and Dawant, 2012]. In comparison to group-wise registration, one major advantage of our approach is that it can be used as a post-processing technique to any given pair-wise registration method, necessitating neither a reference image nor any guarantees on the invertibility of given transformations. Accordingly, it can easily be integrated in any existing work-flow that is tailored to a specific clinical problem. Additionally, our approach treats consistency as a criterion that is encouraged rather than enforced, as implicitly stated by group-wise registration (c.f. Fig. 6.1). This enables its application in scenarios where perfect correspondences are not guaranteed and may not exist, e.g. in inter-patient registration.

We compute our consistency based registration rectification (CBRR) by solving a series of linear least-squares problems, for which efficient large-scale algorithms exist. In combination with a coarse grid, this allows CBRR to run on large datasets. Preliminary results of this work have been presented in [Gass et al., 2014c]. A main contribution of the cur-
6. Consistency-Based Registration Rectification

Figure 6.1: (a) Group-wise registration to the mean image (center), implicitly implying complete consistency via the inferred inter-image registrations (dashed). CBRR utilises the information from the complete set of pair-wise registrations (b). In this set, it aims at reducing the registration inconsistency by modifying pair-wise transformations such that the differences between direct and indirect transformations are minimised (c). This is done simultaneously for all pair-wise transformations using all possible registration triplets. This allows us to exploit the redundant information from multiple loops for each transformation (d).
rent paper is an improved formulation of the CBRR problem which (i) requires only half of the free variables compared to [Gass et al., 2014c]; (ii) requires fewer weighting parameters; and (iii) allows for an iterative rectification process yielding better accuracy for employed approximations and thereby improving the resulting registrations. We present a thorough evaluation of CBRR on fully synthetic and clinical data. Consistency is shown to improve on datasets of 2D and 3D images, while average transformation error, target registration error, and segmentation overlap improve substantially, regardless of initial registration algorithm and image modality. Such measures are also shown to outperform a common group-wise registration algorithm.

This paper is structured as follows: First, we introduce the notation used in this paper, along with a presentation of common concepts regarding consistency. Subsequently, we describe our CBRR algorithm before presenting experimental results and a discussion thereof.

6.2 Notation and Background

An image \( X \) is a function that maps points in the \( D \)-dimensional spatial domain \( \Omega \) to a space \( I \) of image intensities, e.g., CT Hounsfield units. A non-rigid transformation \( T \) is a mapping from \( \Omega \) to \( \Omega \). \( T \) is based on a displacement field \( D \) such that \( T(p) = p + D(p) \) for all points \( p \) in \( \Omega \). Note that both \( X \) and \( D \) are commonly defined on discrete (Cartesian) regular grids, where non-grid values can be obtained by interpolation. In this notation, deforming an image is a function composition \( X \circ T = X(T) \). Composing two transformations is also a function composition: \( (T_2 \circ T_1)(p) = p + D_1(p) + D_2(T_1(p)) \). \( \circ \) denotes the composition operator, read from right to left. Therefore, \( T_2 \circ T_1 \) means transformation \( T_2 \) applied after \( T_1 \).


\[
\mathcal{R} = \{ T_{ij} | i \neq j; i, j \in \{1, \cdots, N\} \}
\]

between images of a set \( X = \{X_1, \cdots, X_N\} \). \( T_{ij} \) then denotes the transformation of image \( X_i \) to \( X_j \). As originally introduced in [Woods et al., 1998], a perfect registration should lead to consistent correspondences over the entire set of images. Inverse consistency as given in [Rohlfing, 2012, Christensen and Johnson, 2001] stipulates that registering \( X_i \) to \( X_j \) should result in the same transformation as the inversion of the transformation from registering \( X_j \) to \( X_i \). Formally, it requires \( T_{ij} \circ T_{ji} = I \), where \( I \) is the identity transform. Another form of consistency is based on registration circuits [Holden
et al., 2000], where any composition of transformations in a circle should result in the identity transform; for instance, for a circle of three transformations:

$$T_{ij} \circ T_{jk} \circ T_{ki} = I.$$

(6.1)

Transitivity has also been used for consistency evaluation [Christensen and Johnson, 2003]. Intuitively, transitivity entails that composing two transformations from registering $X_i$ to $X_j$ and $X_j$ to $X_k$ should result in the same transformation as registering $X_i$ to $X_k$:

$$T_{ik} = T_{ij} \circ T_{jk}.$$  

(6.2)

Note that transitivity and circuit consistency follow from each other when the transformations are symmetric. A general definition of consistency is the following: Let $\Phi^n_{ij}$ be the set of all paths of length $n$ such that each path $p \in \Phi^n_{ij}$ is a composition of $n$ transformations $\{T_1, T_2, \cdots, T_n\}$ that registers $X_i$ to $X_j$. Then, the PTS $R$ is $(m,n)$-consistent if $p = q$ for all transformations $p \in \Phi^m_{ij}$ and $q \in \Phi^n_{ij}$. In this notation, inverse consistency can be denoted as $(2,0)$-consistency, circuit consistency as $(3,0)$-consistency for circuits of length 3 and transitivity as $(2,1)$-consistency. Since each pair $p$ and $q$ share the same start and end points, we will henceforth refer to each such pair as a loop. Then, the inconsistency can be quantified using the residual vector fields $r_{pq} = p - q$ for $p \in \Phi^m_{ij}$ and $q \in \Phi^n_{ij}$. In this notation the metric for algorithmic efficacy of a transformation can be computed as the mean over the norms of all $r_{pq}$ [Song et al., 2010].

### 6.3 Method

In this section, we introduce our algorithm for Consistency-Based Registration Rectification (CBRR). The goal of the method is to improve pair-wise registrations by minimising an inconsistency criterion $C$ defined as the average norm of the inconsistency residuals $r$. Note that consistency is only a necessary condition for correct registrations. This can intuitively be observed when considering the set of identity transforms, which are perfectly consistent but are unlikely to register images properly. This effect can also partly be seen in Fig. 6.2, where several loops with low inconsistency but substantial registration errors can be observed. We therefore additionally constrain the solution of CBRR by penalizing large deviations from the input PTS $R$ using a function $F$. This then makes use of the fact that a priori PTS $R$ contains useful information.

We define our CBRR algorithm as the minimisation of the inconsistency criterion $C$ for a given input PTS $R$ obtained using an arbitrary registration method as follows:

$$R' = \arg \min_{\hat{R}} C(\hat{R}) + \lambda F(R, \hat{R}),$$

(6.3)
6.3. Method

Figure 6.2: Inconsistency and target registration error (TRE) for registration loops in 4D liver MRI sequences. Each point indicates the inconsistency of one loop of three registrations and the average TRE of those three registrations, and the blue line is a least-square fit with \( r=0.77 \) denoting the Pearson correlation. For more details on the data and registration algorithms please see Sec. 6.4.4.

where \( R' \) is the rectified PTS and \( \lambda \) is the weight controlling its adherence to the prior transformations.

### 6.3.1 CBRR Algorithm: Consistency

CBRR minimises the aggregate inconsistency that is induced by all loops in the set of pair-wise transformations:

\[
C(\hat{R}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{p \in \Phi_{ij}^{m}} \sum_{q \in \Phi_{ij}^{n}} ||p - q||^2.
\]  

(6.4)

A desirable property of such inconsistency criterion is redundancy, which means that each pair-wise registration occurs in the residual of multiple loops. This is an important cue for determining the cause of inconsistencies in a registration. For example, consider a PTS \( R \) where only one transformation \( T_{ij} \) has errors which cause inconsistencies. Any consistency loop containing \( T_{ij} \) will therefore have a non-negative residual \( r \), the cause of which cannot easily be located using only information from said loops alone. However, since every transformation other than \( T_{ij} \) will also contribute to multiple loops that all have zero residuals, it is thus inferable that the cause of inconsistency originates from \( T_{ij} \). It is observable that such redundancy is created for any \( m \) and \( n \) when \( m + n > 2 \). However, the number of summands in (6.4) increases exponentially with \( m \) and \( n \). Furthermore, each transformation composition, especially in longer paths, will require interpolation and therefore loses accuracy. This motivates our choice of consistency criterion as (2, 1)-
Figure 6.3: Local normalized cross correlation (LNCC) weight used for the CBRR regularizer. Shown are target image, deformed source, and LNCC weight computed with $\sigma = 4$ mm and $\gamma = 10$ for sample pairs from the MRI (top) and CT (bottom) datasets.

**consistency** (transitivity), where each individual transformation $T_{ij}$ occurs in $3(N - 2)$ different loops and the number of transformation compositions is minimal. Then, $C$ becomes:

$$C(\hat{R}) = \frac{1}{Z_C} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{p \in \Omega} ||\hat{T}_{ij} \circ \hat{T}_{jk}(p) - \hat{T}_{ik}(p)||^2,$$

(6.5)

where $Z_C = MN(N - 1)(N - 2)$ is a normalization constant (the number of all summands, where $M$ is the total number of all points $p$). This definition also minimizes the number of non-linear terms in the form of transformation compositions, the importance of which will be explained in Sec. 6.3.3.

### 6.3.2 CBRR Algorithm: Prior

We assume that the initial PTS $\mathcal{R}$ was computed properly and therefore carries usable information. This motivates us to penalise large deviations of CBRR solution $\mathcal{R}'$ from the input PTS. This approach also enables us to avoid degenerate solutions such as the set of identity transforms. To include similarity-based information in the process, we allow larger deviations from the initial transformation in cases of locally low post-registration similarity. Intuitively, this allows CBRR to limit any negative impact on local registra-
tion accuracy while improving consistency. We then define such penalty function \( \mathcal{F} \) as follows:

\[
\mathcal{F}(\hat{\mathcal{R}}, \mathcal{R}) = 1/Z_F \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{p \in \Omega} \Psi_{ij}(p) ||\hat{T}_{ij}(p) - T_{ij}(p)||^2,
\]

(6.6)

where \( p \in \Omega \) is a (discrete) point, \( Z_F = MN(N-1) \) is a normalization constant, and \( \Psi_{ij} \) is the local similarity between the transformed image \( \mathcal{X}_i(\mathcal{T}_{ij}) \) and \( \mathcal{X}_j \).

We use local normalized cross correlation (LNCC) [Cachier et al., 2003] as the local similarity metric, which was successfully utilized to locally rank segmentation hypotheses in [Cardoso et al., 2012]. LNCC can be computed efficiently as follows:

\[
\text{LNCC}(X_i, X_j, p) = \frac{\langle X_i, X_j \rangle(p)}{\sigma_{X_i}(p) \sigma_{X_j}(p)}
\]

(6.7)

\[
\langle X_i, X_j \rangle(p) = \frac{X_i \cdot X_j(p) - \overline{X_i}(p) \cdot \overline{X_j}(p)}{\overline{X}} = G_{\sigma_G} \ast X
\]

(6.8)

\[
\sigma^2_{X_i}(p) = X^2_i(p) - \overline{X_i^2}(p)
\]

(6.9)

where \( \ast \) is the convolution operator and \( G_{\sigma_G} \) is a Gaussian kernel with standard deviation \( \sigma_G \). From the LNCC metric, we compute the weights

\[
\Psi_{ij} = \left( 1.0 - \frac{\text{LNCC}(X_i(\mathcal{T}_{ij}), X_j)}{2} \right)^\gamma,
\]

(6.11)

which normalizes LNCC to the range \([0, 1]\) and subsequently applies a contrast scaling using \( \gamma \). The resulting local weights are shown on an example in Fig. 6.3.

### 6.3.3 Casting into a System of Linear Equations

We find a solution for (6.3) by casting it as a linear least-squares optimization problem, for which efficient large-scale algorithms exist. The only non-linear term is the transformation composition in (6.5), which we approximate as follows:

\[
\hat{T}_{ij} \circ \hat{T}_{jk} - \hat{T}_{ik} = \hat{D}_{ij}(\hat{T}_{jk}) + \hat{D}_{jk} - \hat{D}_{ik}
\]

(6.12)

\[
\approx \hat{D}_{ij}(\mathcal{T}_{jk}) + \hat{D}_{jk} - \hat{D}_{ik},
\]

(6.13)

by substituting the unknown true transformation \( \hat{T}_{jk} \) in (6.12) with the observed transformation \( \mathcal{T}_{jk} \) when it is used to deform the displacement field \( \hat{D}_{ij} \).
This allows us to rewrite the minimization problem in the form of \( x' = \min_{x} |Ax - b|^2 \), where \( x \) is the column vector of all grid-point displacement variables for all transformations in \( \hat{R} \). Then, \( x \) has \( MN(N-1) \) entries and has the form

\[
x = \left[ \hat{D}_{1,2}(p_1), \ldots, \hat{D}_{ij}(p_m), \ldots, \hat{D}_{N-1,N}(p_M) \right]^T
\]  

(6.14)

The matrix \( A \) and the vector \( b \) can be decomposed into two parts, \( A_C \) and \( b_C \) for the consistency criterion and \( A_F \) and \( b_F \) for the regularization term:

\[
x' = \min_{x} \left| \begin{bmatrix} A_C \\ A_F \end{bmatrix} x - \begin{bmatrix} b_C \\ b_F \end{bmatrix} \right|^2.
\]  

(6.15)

Each row of \( A_C \) consists of 0s and ±1s such that multiplication with \( x \) yields one of the summands of the consistency criterion. The locations of the nonzero entries correspond to the locations of the respective displacements in the vector \( x \) and can be calculated with the help of an indexing function \( \text{idx} \) that returns the index of displacement \( p \) in transformation \( T_{ij} \) as follows:

\[
\text{idx} : \mathbb{N}^3 \rightarrow \mathbb{N} : \text{idx}(i, j, p) \rightarrow \{1, \ldots, MN(N-1)\}
\]

(6.16)

Then, each row \( r \) of \( A_C \) corresponding to a consistency term \( \hat{D}_{ij}(T_{jk}(p)) + \hat{D}_{jk}(p) - \hat{D}_{ik}(p) \) has the following entries in columns \( c \):

\[
A_C(r, c) = \begin{cases} 
1 & \text{if } c = \text{idx}(i, j, T_{jk}(p)) \text{ or } \text{idx}(j, k, (p)) \\
-1 & \text{if } c = \text{idx}(i, k, (p)) \\
0 & \text{otherwise}
\end{cases}
\]  

(6.17)

All corresponding entries in \( b_C \) are then zero, since we want the inconsistency to vanish. When \( T_{jk}(p) \) is not on a grid point, we use linear interpolation:

\[
\hat{D}_{ij}(T_{jk}(p)) + \hat{D}_{jk}(p) - \hat{D}_{ik}(p) \approx \left( \sum_{p' \in \mathcal{N}(T_{jk}(p))} \nu(p') \cdot \hat{D}_{ij}(p') \right) + \hat{D}_{jk}(p) - \hat{D}_{ik}(p),
\]  

(6.18)

where \( \mathcal{N} \) returns the grid neighborhood of a non-grid point, and \( \nu \) computes the interpolation weights. Thus, such consistency-expressing rows in \( A_C \) can contain up to 10 non-zero elements in 3D, where the first non-grid point is interpolated from 8 neighboring grid points.

\( A_F \) is a square, diagonal matrix such that the multiplication of one row of \( A_F \) with \( x \) and subtracting the corresponding entry of \( b_F \) yields one of the prior terms \( \lambda \Psi_{ij}(p)\hat{D}_{ij}(p) - \lambda \Psi_{ij}(p)D_{ij}(p) \) being minimized. In the same notation as above, this can be expressed as:
6.3. Method

\[ A_F(r, c) = \begin{cases} \lambda \Psi_{ij}(p), & \text{if } r = c \\ 0, & \text{otherwise.} \end{cases} \quad (6.19) \]

\[ b_F(c) = \lambda \Psi_{ij}(p) \mathcal{D}_{ij}(p) \quad (6.20) \]

with \( c = \text{idx}(i, j, p) \).

6.3.4 Iterative Refinement

In order to improve the approximation in (6.13), we employ a fixed point iterative update scheme. In each iteration \( t \), a new PTS \( \mathcal{R}^t \) is computed, where \( \mathcal{R}^0 \) is the CBRR initialisation from a prior registration. The estimate of the last iteration is then used for the approximation in (6.13) as follows:

\[ \mathcal{T}_{ij}^t \circ \mathcal{T}_{jk}^t - \mathcal{T}_{ik}^t \approx \mathcal{D}_{ij}^t (\mathcal{T}_{jk}^{t-1}) + \mathcal{D}_{jk}^t - \mathcal{D}_{ik}^t. \quad (6.22) \]

This allows the approximation to become more accurate as the difference between \( \mathcal{T}^{t-1} \) and \( \mathcal{T}^t \) becomes smaller. The iterative process can also be used to update the prior term. Since we assume that the input registration contains errors, adhering too strongly to it might deteriorate the results of the algorithm. A straightforward way to check for improvements during the iterations is to compare the local post-registration similarities. We therefore re-compute the local similarity weights \( \Psi^t \) after every iteration, and keep track of the locally best encountered value along with the corresponding displacement:

\[ (\Psi^{\text{best}}, \mathcal{D}^{\text{best}})_{ijp} = \begin{cases} (\Psi^t, \mathcal{D}^t)_{ijp}, & \text{if } \Psi^t_{ij}(p) > \Psi^{\text{best}}_{ij}(p) \\ (\Psi^{\text{best}}, \mathcal{D}^{\text{best}})_{ijp}, & \text{otherwise.} \end{cases} \quad (6.23) \]

where we moved the sub-indices outside of the tuples for readability. Then, we change the prior terms in (6.6) to penalize differences from these current best estimates:

\[ \mathcal{F}(\hat{\mathcal{R}}, \mathcal{R}^{\text{best}}) = \frac{1}{Z_F} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{p \in \Omega} \Psi^{\text{best}}_{ij}(p) ||\hat{\mathcal{D}}_{ij}(p) - \mathcal{D}^{\text{best}}_{ij}(p)||^2. \quad (6.24) \]

6.3.5 Implementation

The number of equations grows with the cube of the number of images, \( N \), and linearly with the number of pixels, \( M \). However, efficient algorithms for solving large-scale linear
least-squares problems exist. We chose the Matlab minFunc\(^1\) package for its computation and memory efficiency, particularly when dealing with large sparse matrices. In particular, we use the provided limited-memory BFGS solver with a fixed number of 100 iterations. Furthermore, it can be observed that the rows of linear equations corresponding to each principal displacement axis are independent from other axes. Consequently, such equations are separable and each displacement direction can be solved individually, reducing the problem size to half in 2D and a third in 3D. Note that these separately solved individual axes still interact with each other in the iterative scheme when the transformation from the previous iteration is used to index current displacements.

Despite the abovementioned implementation optimizations to reduce memory complexity, solving (6.3) at pixel level is prohibitive mostly due to memory demand. We therefore solve our CBRR problem on a coarse grid for efficiency and then interpolate the results. To this end, \(R\) and \(\Psi\) need to be downsampled, leading to information loss, which we aim to minimise by using the following full-resolution registration estimates. Let \(R_{\text{low}}\) be the downsampled original PTS. The rectified PTS \(R'_{\text{low}}\) will then be at the same resolution, and can be used to estimate the update field \(\Delta_{\text{low}} = R_{\text{low}} - R'_{\text{low}}\). We then upsample \(\Delta_{\text{low}}\) to the original resolution, and subsequently obtain \(R' = R - \Delta\). This yields registration estimates with the local details preserved, while the local similarities \(\Psi\) are being computed from the full-resolution registrations.

The overall algorithm then works as shown in Alg. 2. Convergence is achieved when the relative change of residual inconsistency is smaller than 1%. The remaining free parameters consist of the regularizer weight \(\lambda\), the width of LNCC window \(\sigma\), and the exponent \(\gamma\).

### 6.4 Experimental Results

We evaluated our method in three experiments. We first present a controlled environment using synthetic data, where known ground-truths exist in order to better demonstrate our method. Subsequently, we present our results on two clinical datasets.

#### 6.4.1 Synthetic Images

We first evaluated our method using a fully synthetic setup where images, registrations and registration errors were controlled. Nine 199×199 px\(^2\) images were created by randomly deforming a reference image and adding 50% SNR noise. We used three different sets of random deformations with varying complexity: To generate smooth, large scale deformations we randomly sampled b-Spline control grid parameters of an 8x8 grid. To

\(^1\)http://www.di.ens.fr/~mschmidt/Software/minFunc.html, 2012
Algorithm 2 Outline of the proposed CBRR algorithm.

Data: \( \mathcal{R}, \mathcal{X}, \lambda, \gamma, \sigma \)

Result: \( \mathcal{R}' \)

\[
t \leftarrow 0 \\
\mathcal{R}_0 \leftarrow \mathcal{R}
\]

\textbf{while not-converged and } t < t_{\text{max}} \textbf{ do}

\hspace{1em} \text{Compute } \Psi^t \text{ using (6.7)-(6.11)}

\hspace{1em} \text{Update } \Psi^{\text{best}} \text{ and } D^{\text{best}} \text{ according to (6.23)}

\hspace{1em} \text{Downsample } \mathcal{R}^t \text{ and } \Psi^t

\hspace{1em} \textbf{forall the axes } d \text{ do}

\hspace{2em} \text{Form matrix } A_d \text{ and vector } b_d \text{ using (6.17)-(6.24)}

\hspace{2em} \mathcal{R}_d^t = \min \left| A_d x - b_d \right|^2

\hspace{1em} \text{end}

\hspace{1em} \mathcal{R}^t \leftarrow \bigcup_d \mathcal{R}_d^t

\hspace{1em} \text{Upsample } \mathcal{R}^t

\hspace{1em} t \leftarrow t + 1

\textbf{end}

\[
\mathcal{R}' \leftarrow \mathcal{R}^t
\]

ensure diffeomorphisms, the parameters were uniformly sampled from \([-0.4s, \ldots, 0.4s]\), where \(s\) is the control grid spacing [Rueckert et al., 2006]. More complex transformations were generated by using the same process on 20x20 and 50x50 b-Spline control grids. We then created three sets of ground-truth transformations, the first using the transformations from the 8x8 grid, the second by composing the transformations form the 20x20 grids onto the first, and the last by composing the transformations from the 50x50 grid onto the second. Pairwise registrations were then analytically computed by inversion and composition as follows: Let \(T_{ri}\) denote the transformation from reference \(r\) to an image \(i\). The pairwise transformation \(T_{ij}\) can then be computed as:

\[
T_{ij} = T_{rj} \circ T_{ri}^{-1}
\]

as described in [Skrinjar et al., 2008]. We then randomly generated errors for each pairwise transformation as follows: Of a 20x20 b-Spline grid, each control point has a varying likelihood \(f\) between 1\% and 10\% to have an error that was uniformly sampled from a variable range \(r \in \{1, \ldots, 9\}\), and then added to the ground-truth transformations. Each set of parameters \(f\) and \(r\) then thus creates a separate PTS, on which we ran CBRR with different grid resolutions. We then compute the relative improvement of the average transformation error (ATE) which is the mean of the euclidean distance between ground-truth displacements and error displacements as \(1.0 - \frac{\text{ATE}_{\text{err}}}{\text{ATE}_{\text{max}}},\) where \(\text{ATE}_{\text{err}}\) is the ATE of the input PTS. The resulting improvements can be seen in Fig. 6.4. It can be observed that CBRR can correct a substantial fraction of errors, especially for frequent and large magnitude errors. However, infrequent and small errors can not be corrected well, which is likely due to CBRR’s relatively coarser grid resolution and additionally due to the
This results in small, new errors, which on average can be larger than the errors of the original input registration, even though they are very small in absolute terms. CBRR, especially when using a coarse grid, is also challenged by complex underlying true transformations. In those cases, however, a substantial fraction of large deformation errors can still be corrected, and the performance can be improved using higher resolution grids.

### 6.4.2 2D MR Images

For the second experiment, we simulated medical images with known correspondences. We used 19 mid-sagittal slices of brain MRI with $481 \times 374$ px$^2$ resolution and isotropic spacing of 0.3 mm. In order to obtain a dataset with existing and known dense corre-
Figure 6.5: Data generation for the 2D MR experiment. First, image $X_1$ is registered to the remaining images $X_i$, resulting in transformations $T_{1,i}$ (top left). Then, $X_1$ is deformed using the estimated transformation and 50% SNR Gaussian noise ($\zeta$) is added, generating new images $X'_i$ (top right). Pairwise transformations $T_{i,j}$ between all $X'_i$ are known analytically through the registrations from $X_1$. The image set $\{X_1, X'_2, \ldots, X'_N\}$ is then used as the experimental dataset and the known dense correspondences $T_{i,j}$ as the ground-truth in evaluation.

Corresponences, we generated such data using Markov-random field (MRF) based registration [Glocker et al., 2008] as shown in Fig. 6.5. We then used the well-known diffeomorphic demons algorithm [Vercauteren et al., 2009] and a MRF-based algorithm [Glocker et al., 2008] to mutually re-register all 19 images, resulting in 342 registrations for each method. We solved CBRR separately for both of these sets, using a coarse grid with 2.4 mm spacing (downsampling the input PTS by a factor of 8).

We first evaluate different choices of $\lambda$ and $\gamma$, the results of which are presented in Fig. 6.6(a). $\sigma$ was fixed to 4mm. While a small improvement can even be seen without using the local weights ($\gamma = 0$), the performance of CBRR improves substantially for appropriate choices for those parameters ($\gamma = 10, \lambda = 0.5$), which we use throughout all remaining experiments. Note that in the limit cases, CBRR either finds the identity transforms (without any regularization) or the input transforms. We further evaluate how the number of images affect CBRR performance, where results can be seen in Fig. 6.6(b). A convergence pattern can be observed from about 8 images onward.
Figure 6.6: Relative ATE improvement using CBRR with an increasing number of 2D MR images. Each experiment is repeated 10 times with randomly selected images. The standard deviation of the results is shown by the error bars.

<table>
<thead>
<tr>
<th>Demons</th>
<th>ATE</th>
<th>ρ</th>
<th>C</th>
<th>MRF</th>
<th>ATE</th>
<th>ρ</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registration</td>
<td>3.50</td>
<td>2.53</td>
<td>1.92</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ LWA</td>
<td>2.34</td>
<td>0.60</td>
<td>1.78</td>
<td>1.83</td>
<td>0.69</td>
<td>1.87</td>
<td></td>
</tr>
<tr>
<td>+ CLERC [Gass et al., 2014c]</td>
<td>2.66</td>
<td>0.74</td>
<td>0.86</td>
<td>1.28</td>
<td>0.85</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>+ CBRR</td>
<td>1.64</td>
<td>0.84</td>
<td>0.35</td>
<td>0.88</td>
<td>0.91</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>+ AQUIRC [Datteri and Dawant, 2012]</td>
<td>-</td>
<td>0.47</td>
<td>-</td>
<td>0.64</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.1: Left: Average transformation error (ATE), inconsistency C and pearson correlation ρ of estimated and true registration error magnitude for the 2D MRI experiments using simulated transformations before and after using CBRR. We also compare CBRR to locally averaged weighting (LWA), an earlier version of this work (CLERC) presented in [Gass et al., 2014c], and AQUIRC [Datteri and Dawant, 2012], which computes a dimensionless estimate of registration error which can be correlated with the true error.

Quantitative results by comparing with the analytically computed ground-truth are given in Tab. 6.1. It is observed, that ATE improves by over 45% while the inconsistency C is reduced considerably. We show one example case for each registration method in Fig. 6.8(left), where CBRR is seen to improve the original registration in both cases.

We compare CBRR to straight-forward locally-weighted averaging (LWA), which fuses all indirect transformations $T_{ik} \circ T_{kj}$ and the direct transformation $T_{ij}$. The local weights are computed using the same LNCC technique as in CBRR, but parameters have been
tuned separately ($\sigma=8, \gamma=30$) to not put the LWA method at a disadvantage. We also compared CBRR to CLERC (Correcting Local Errors using Registration Consistency) [Gass et al., 2014c], which is a preliminary version of the technique presented in this paper. The main difference is that CLERC uses no iterative updates and a different numerical solver. This difference can partly be observed in Fig. 6.7, where the CLERC results shown in Tab. 6.1 are similar to CBRR after the first iteration, with remaining differences caused by a different numerical solver. CBRR then typically converges after about 5-6 iterations. As shown in Tab. 6.1, CBRR outperforms CLERC and LWA in all metrics for all experiments. We elaborate on the differences between CLERC and CBRR later in Sec. 6.4.5. We also compare our method to our own implementation of AQUIRC [Datteri and Dawant, 2012], which computes a dimensionless measure of local registration error magnitude which can be correlated with the true local error magnitudes. It is observed that the magnitude of the updates computed by CBRR correlate substantially stronger with the true error magnitudes in comparison to AQUIRC.

### 6.4.3 3D CT Segmentation

The next experiment was performed using a clinical dataset of 15 3D CT head scans from different individuals, with $160 \times 160 \times 129$ pixel$^3$ resolution and 1 mm spacing. All images were mutually registered using demons-, MRF- and ANTs [Avants et al., 2009]-based registrations, yielding a total of 210 transformations for each registration algorithm. Post-registration TRE was computed using 12 manually placed landmarks, which were placed at anatomically identifiable locations on the jawbone and the skull for all images. In addition to TRE, we also evaluated anatomical overlap of manually segmented mandibles using Dice’s coefficient, mean surface distance (MSD) and Hausdorff distance (HD). Note that, in contrast to the above experiments, in this current experiment with
Figure 6.8: CBRR of MRF (top) and demons (bottom) registration on the 2D MRI data. Shown are from left to right: target image, deformed source, true registration error, error of rectified registration and the source image deformed by the rectified registration. The hue of the color in the error visualizations indicates its orientation, while the saturation corresponds to the error magnitude.

Figure 6.9: Surface distance of ground-truth jawbone segmentation to MRF-registered source (left) and subsequent CBRR (right). Best viewed in color.

actual inter-patient registrations, neither consistency nor dense correspondences can be guaranteed. Indeed, in these 3D CTs the presence and number of teeth vary substantially among the scanned individuals. We solved CBRR using a coarse grid with 8 mm spacing, using the same parameters as the 2D MRI experiment above ($\lambda=0.5$, $\sigma=4$ mm, $\gamma=10$). Results are given in Tab. 6.2, in which a significant ($\rho<5\%$) improvement in both TRE and segmentation metrics using CBRR can be observed. A larger improvement is observed for the demons registrations which has a poorer initial performance. We show an example of corresponding post-registration and post-CBRR segmentations in Fig. 6.9, and the distributions of segmentation overlap before and after CBRR can be seen in Fig. 6.10. Here, it is clearly observable that CBRR substantially improves particularly the worst-case segmentation quality induced by the input PTS, while best-case dice overlap can be slightly reduced. In the case of the demons registration algorithm, segmentation quality is substantially improved for all registrations.
### 6.4. Experimental Results

#### 3D CT

<table>
<thead>
<tr>
<th>Method</th>
<th>TRE</th>
<th>Dice</th>
<th>MSD</th>
<th>HD</th>
<th>C</th>
<th>Total run-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>reg-demons</td>
<td>9.25 ± 4.22</td>
<td>0.65 ± 0.15</td>
<td>1.27 ± 1.19</td>
<td>17.86 ± 8.83</td>
<td>8.14 ± 0.15</td>
<td>140h</td>
</tr>
<tr>
<td>+ CBRR</td>
<td>7.16 ± 2.19</td>
<td>0.75 ± 0.09</td>
<td>0.58 ± 0.36</td>
<td>14 ± 5.65</td>
<td>1.07 ± 0.15</td>
<td>+2.5h</td>
</tr>
<tr>
<td>reg-MRF</td>
<td>5.25 ± 1.95</td>
<td>0.84 ± 0.06</td>
<td>0.34 ± 0.29</td>
<td>11.43 ± 4.86</td>
<td>5.67 ± 0.58</td>
<td>39h</td>
</tr>
<tr>
<td>+ CBRR</td>
<td>4.66 ± 1.20</td>
<td>0.86 ± 0.03</td>
<td>0.22 ± 0.10</td>
<td>9.72 ± 3.65</td>
<td>1.12 ± 0.36</td>
<td>+2.5h</td>
</tr>
<tr>
<td>reg-ANTs</td>
<td>6.02 ± 3.35</td>
<td>0.85 ± 0.11</td>
<td>0.49 ± 0.68</td>
<td>12.27 ± 5.99</td>
<td>6.43 ± 1.72</td>
<td>105h</td>
</tr>
<tr>
<td>+ CBRR</td>
<td>4.70 ± 1.60</td>
<td>0.89 ± 0.04</td>
<td>0.17 ± 0.12</td>
<td>8.79 ± 3.48</td>
<td>1.13 ± 0.58</td>
<td>+2.5h</td>
</tr>
<tr>
<td>group-wise reg.</td>
<td>8.1 ± 2.86</td>
<td>0.75 ± 0.10</td>
<td>0.97 ± 0.61</td>
<td>22.64 ± 4.30</td>
<td>0.0 ± 0.36</td>
<td>26h</td>
</tr>
</tbody>
</table>

Table 6.2: Quantitative evaluation of CBRR and comparison to group-wise registration on the 3D CT dataset. Mean TRE, mean surface distance (MSD), Hausdorff distance (HD) and inconsistency $C$ are given in mm.

![Figure 6.10: Segmentation overlap percentiles on the 3D CT dataset. Each datapoint shows the maximum Dice coefficient for a given percentage of the 210 total pair-wise registrations.](image)

We also compared CBRR to a traditional group-wise registration method. We used the medical image registration toolkit (MIRT), which is publicly available at [Myronenko, 2013]. It implements a well-known group-wise registration method in which the images are iteratively registered to an evolving group mean. We used the default parametrization, which utilizes residual complexity [Myronenko and Song, 2010] as similarity measure. The results for this comparison is given in Tab. 6.2. While the group-wise registration outperforms the plain pair-wise demons algorithm, using CBRR is seen to yield the best result in all cases and for all evaluation metrics.
6.4.4 4D Liver MRI Motion Estimation

We also evaluated our method on motion estimation from 4D MRI. Images of a 4D sequence are typically registered to one reference frame, e.g. exhale, and pair-wise transformations and subsequently an estimate of the motion can be inferred from those [Skrinjar et al., 2008]. Other studies have shown that it is beneficial to initialize such registrations with the transformation of the previous image in the sequence [Tanner et al., 2013]. Using CBRR, we compute all pair-wise registrations for all images in a sequence, which then includes both the direct registrations between the reference frame and the remaining frames, as well as the sequential registrations. CBRR can then update the resulting PTS, where consistent correspondences are guaranteed to exist. We use 4 sequences of the abdomen for our evaluation, which were recorded using the technique described in [von Siebenthal et al., 2007]. Each sequence contains 11 breathing phases (images) with a spatial resolution of $1.37 \times 1.37 \times 4$ mm$^3$, about 370 ms temporal resolution and a mean image size of $164 \times 189 \times 23$ voxels. 52 landmarks were placed manually at anatomically identifiable positions such as ribs, vertebrae and vessel bifurcations in the liver, kidneys and lungs, for one exhale and two inhale images in each sequence. Registration quality can then be assessed in terms of mean target registration error (TRE) between the ground-truth and transformed landmarks for all image pairs where landmarks are available. We used two registration methods for each sequence: ANTs [Avants et al., 2009] as it delivers state-of-the-art results on many public datasets, and a parametric total variation (PTV) registration method which was specifically developed considering sliding motion that occur during breathing in abdominal imaging [Vishnevskiy et al., 2014]. Fig. 6.11 shows one CBRR loop of exhale-inhale-inhale using PTV, and the resulting dense inconsistency and extrapolated TRE before and after CBRR, which was run on deformations that were downsampled by a factor of 4. The grid spacing was kept as isotropic as possible, resulting in a spacing of 5.48 mm$^3$. CBRR parameters were the same as for the other experiments with $\gamma = 10, \sigma = 4$, and $\lambda = 0.5$. Tab. 6.3 shows quantitative results, where a substantial and significant ($\rho < 5\%$) improvement in TRE can be observed. Interestingly, the CBRR TRE results of both registration methods are very similar to each other, despite substantial difference in input TRE. Fig. 6.12 visualizes the change in registration error per landmark for both registration methods before and after using CBRR. It can be observed that especially the worst-case performance is improved considerably, reducing the maximum TRE by up to 60%.

6.4.5 Discussion

We experimentally demonstrated that CBRR improves (2,1)-consistency on a variety of different input data and using different pair-wise registration methods. In addition to the consistency improvement, CBRR was shown to yield improved transformations in terms of average transformation error, TRE, and segmentation overlap. As can be seen
6.4. Experimental Results

Figure 6.11: Inconsistency and TRE of one loop of sequence 1 before and after using CBRR for the PTV registration method. Mid-sagittal slices are shown for MRIs, inconsistency, and TRE extrapolated using thin-plate splines. CBRR was computed on a longer sequence using 11 images and 110 registrations in total.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>TRE \text{exhale,inhale} 1</th>
<th>TRE \text{inhale} 1</th>
<th>TRE \text{inhale} 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before CBRR</td>
<td>![Before CBRR Image]</td>
<td>![Before CBRR Image]</td>
<td>![Before CBRR Image]</td>
</tr>
<tr>
<td>After CBRR</td>
<td>![After CBRR Image]</td>
<td>![After CBRR Image]</td>
<td>![After CBRR Image]</td>
</tr>
</tbody>
</table>

Table 6.3: Quantitative evaluation of CBRR on the 4D liver MRI sequences. The average run time is the average time required to process one of the sequences with 11 breathing phases of 3D MR.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>TRE per phase [mm]</th>
<th>Average TRE [mm]</th>
<th>Total mean TRE [mm]</th>
<th>Average run-time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10.99 ±6.58</td>
<td>3.91±1.92</td>
<td>5.64±3.29</td>
<td>11.12±6.26</td>
</tr>
<tr>
<td>2</td>
<td>3.75±5.50</td>
<td>1.49±1.36</td>
<td>1.26±1.24</td>
<td>6.30±6.70</td>
</tr>
<tr>
<td>3</td>
<td>2.31±2.83</td>
<td>1.42±1.17</td>
<td>1.13±0.92</td>
<td>2.27±1.85</td>
</tr>
<tr>
<td>4</td>
<td>2.50±3.51</td>
<td>1.58±1.85</td>
<td>1.18±1.02</td>
<td>2.89±3.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.89±3.28</td>
<td>1.81±1.91</td>
<td>1.12±0.77</td>
<td>2.20±1.89</td>
</tr>
<tr>
<td></td>
<td>2.04±2.70</td>
<td>1.68±1.60</td>
<td>1.81±1.60</td>
<td>1.81±1.60</td>
</tr>
<tr>
<td></td>
<td>2.27±1.85</td>
<td>1.47±1.41</td>
<td>1.10±0.77</td>
<td>2.20±1.89</td>
</tr>
<tr>
<td></td>
<td>1.81±1.60</td>
<td>1.47±1.41</td>
<td>1.10±0.77</td>
<td>2.20±1.89</td>
</tr>
</tbody>
</table>

in Fig. 6.10, such improvements mostly affect poorer pair-wise registrations (lower percentile) in case of a relatively well-performing initial registration, e.g. using MRFs. In contrast, if the initial registration is poorer as in demons registration algorithm, the Dice overlap increases among all but the top percentiles. This is attributed to the higher ini-
Figure 6.12: TRE before and after CBRR on the 4D Liver MRI sequences. Each point shows the error for a single landmark before (x-axis) and after (y-axis) CBRR. Note that points below the $y = x$ line indicate a reduction in error using CBRR.

Figure 6.13: Mid-coronal slices of the mean images of the 3D CT dataset for (a) group-wise and (b) CBRR registration. For CBRR, one image was chosen as reference to which all images were aligned for averaging. We deformed the group-wise mean to the same image to allow a better comparison.

Potential inconsistency of poorer registrations, which yields more space for improvements via registration rectification.

Besides an improvement in pair-wise alignment, the results of CBRR are particularly interesting for group-wise methods such as atlas generation. As shown in Tab. 6.2, CBRR
can out-perform group-wise registration quantitatively. A typical qualitative result of group-wise registration is the resulting mean image, which represents the center of the distribution. Here, a sharp mean image is an indicator for the successful alignment of all images to such mean. We present mean images using registrations from CBRR and group-wise registrations in Fig. 6.13. Since CBRR does not return the center of the distribution directly, the images are in the reference space of one randomly chosen image for better comparison. It can be seen that the CBRR mean is much sharper than the group-wise mean, indicating that group-wise registration could not align the images satisfactorily in this task.

Note that CBRR yields satisfactory results in all tests with an uniform parametrization. Note that further experiment-specific improvement is possible, in particular by adjusting $\lambda$ to incorporate prior knowledge on the expected error. Different choices of the local similarity metric can also be beneficial, especially for multi-modal data.

CBRR can also be interpreted as a detection method for inconsistency-based registration errors. Based on the difference between the estimated and the original registrations, one can estimate local error magnitudes. These estimations can be evaluated through ground-truth local-error magnitudes, for experiments for which those are known. This allows for a comparison with AQUIRC [Datteri and Dawant, 2012], which computes a dimension-less measure of local error magnitude based on (3,0)-consistency. We used our own implementation of their technique, the results of which were presented in Tab. 6.1. Note that CBRR could also be adapted to utilize (3,0)-consistency, but would require an additional approximation (c.f. Sec. 6.3.3) for the additional transformation composition, which would make the implementation more complicated and the solution less robust. The error magnitudes estimated by CBRR correlate much stronger with the true error magnitudes in comparison to AQUIRC, in addition to surpassing our earlier algorithm CLERC [Gass et al., 2014c]. Additionally, both CLERC and CBRR estimate actual error vectors in contrast to the dimension-less measure being reported by AQUIRC. The difference in performance between CBRR and CLERC can mainly be explained through the additional iterative refinement in CBRR, which improves the approximation in (6.12). An additional benefit of CBRR is that it requires to only estimate registration variables, in contrast to registration and error variables in CLERC. This makes it computationally more efficient and easier to parametrize since one less parameter is required.

Still, the bottleneck of the proposed CBRR algorithm is its large memory requirement since all the displacement variables are estimated simultaneously. This contrasts the AQUIRC approach [Datteri and Dawant, 2012], which can solve for the error at every spatial location independently. In our experiments, the memory utilization reached up to 15 GB for the 3D CT segmentation experiment. The main factor for this is the number of non-zero entries in $A$, which is proportional to $N^3M$, where $M$ denotes the number of coarse grid points. Note that we keep only the downsampled pair-wise registrations in memory, while re-loading the full resolution registrations only on demand. We implemented our CBRR prototype in C++/ITK. Constructed systems of equations are solved
using the minFunc solver called via the Matlab engine interface for C++. The equation system itself is not particularly large in memory, requiring about 6GB including all auxiliary variables for the 15 3D CT image problem. However, a significant memory overhead is needed in Matlab for passing and solving the problem. Therefore, using a dedicated C++ solver might be one approach to reducing memory demands. Additionally, our experiments showed that satisfactory results can be already achieved with a limited number of images (c.f. Fig. 6.6). For instance, reducing the number of images from 15 to 7 in the 3D case would reduce the memory requirements by a factor of 8. One advantage of CBRR is the speed with which all pair-wise registrations are updated. One full iteration, which updates 210 3D registrations, takes about 30 minutes, including the computation of pair-wise local similarities. This is about 80 times as fast as computing all pair-wise registrations using MRFs. Since the proposed approach is likely to be used in an offline fashion, it would be reasonable to trade off run-time for decreased memory usage, thereby also allowing for more images and/or higher resolution transformations.

A strong point of CBRR is that it is independent of the used registration algorithm, and imposes no restrictions on the spatial smoothness of the transformations. This enables its application also in challenging scenarios as in registration with sliding motion, as can be seen from the results on the 4D MRI experiment, where substantial improvements are seen both for a smooth (ANTs) and for a sliding-motion enabled registration method.

6.5 Conclusions

We presented a novel population-based method for refining pair-wise registrations by reducing their inconsistency. By formulating the task as a linear least-squares problem, we obtain an efficient algorithm that is able to resolve such inconsistencies in groups of up to 15 3D images simultaneously, which is, to the best of our knowledge, not possible with any other technique. In contrast to group-wise registration, our method is arguably better suited for inter-patient registration, as correspondences need not be perfect across all images. Instead, a trade-off is achieved, balancing group-wise consistency and pair-wise visual similarity. Experimentally, our approach was shown to be able to not only increase consistency, but also improve segmentation overlap in atlas-based segmentation, and target registration error in both inter- and intra-patient registration, where the latter is an important problem e.g. for breathing motion estimation.
Summary and Perspectives

In this thesis, several methods for medical image segmentation using weak priors have been presented. Each chapter was published separately in peer-reviewed journals and conference proceedings. Since each article, and therefore chapter, already contains a specific discussion of the presented algorithm and perspectives, this summary focuses on more generic observations and implications.

In Chap. 2, an iterative algorithm for joint segmentation and registration has been investigated that utilizes the given segmentation of an atlas and penalizes deviations between deformed atlas segmentation and an intensity-based segmentation of the target image using a distance function. This allows the optimization to avoid poor local optima in early stages of the iterative registration. Still, this definition of the joint segmentation and registration problem is inherently ill-posed as such since it relies on the overlap between the deformed atlas segmentation and any voxel which matches the intensity model of the target anatomy. In medical imaging, such intensity models are often not organ-, but rather tissue-specific, e.g. the well-known threshold in Hounsfield units for the bone. To create a more robust optimization problem, it has been proposed in this chapter to use an auxiliary segmentation label in both atlas and target, which accounts for any structure that fits the intensity model but does not cover the target anatomy.

Chap. 3 extends on the previous chapter by developing a simultaneous non-rigid registration and segmentation (SRS) method, in contrast to many studies in the literature that utilize alternating optimization schemes. The goal function is obtained through a rigorous analysis of the underlying MAP problem, and a novel two-layer Markov random field is proposed which allows for estimating registration and segmentation variables at different levels of detail. As one of the first studies in the literature, the presented approach allows for the direct comparison of alternating and simultaneous solvers for this problem, and it is demonstrated that SRS achieves superior segmentation fidelity in terms of both overlap- and distance based criteria. Such improvements over standard registration-based segmentation and the alternating approach have been mainly observed in the lower percentiles with regard to pair-wise post registration segmentation overlap in a leave-one-out study. This demonstrates the increased robustness with regard to atlas-target mismatch or
poor local optima. While achieving the best segmentation quality on the 3D CT dataset of all algorithms presented in this thesis, the presented method has two drawbacks: (i) the registration algorithm uses to a discrete optimization scheme, and (which might not be preferable for a specific application, or available due to patents on the optimization techniques) (ii) there are eight free parameters which need to be set specifically for each task.

The algorithms presented in the remaining chapters of this thesis address these challenges by utilizing semi-supervised techniques which can leverage information from sets of unsegmented images. Crucially, they all act as post-processing given a set of non-rigid registrations as input. This addresses challenge (i) of the previous section since it allows for using the result of any available registration algorithm. This can be important for clinical applications where personnel might be familiar with or constrained to a specific implementation in their workflow. Furthermore, it can be observed that the algorithms in the following chapters use fewer parameters (two in Chap. 4 and three in chapters 5 and 6) in addition to those used by the registration. Those parameters generalize quite well and can be applied with little adjustments when dealing with different segmentation tasks and image modalities. This is a major benefit which makes it much easier, in comparison to SRS, to use those semi-supervised methods out-of-the-box.

To differentiate further, the algorithms presented in chapters 4 and 5, respectively, combine automatically generated segmentation and registration hypotheses. In both cases, it is assumed that the intermediate registration steps between atlas and target may contain information which complements the direct registration of atlas and target. This is consistent with the literature, where it was frequently observed that a certain fit between atlas and target images is necessary to obtain satisfactory segmentation performance when using conventional atlas-based segmentation. This challenge can at least partly be overcome by the proposed methods, as also observed from the segmentation results which exhibit increased robustness with regard to atlas selection (c.f. Fig. 4.3a).

The algorithm presented in Chap. 6 aims at a similar goal compared to the previous two chapters, but replaces the heuristic fusion of segmentation/registration hypotheses with a more principled goal function: By optimizing the consistency of the set of pair-wise registrations in a least-squares optimization framework, information from all pair-wise registrations are combined in order to update all pair-wise registrations simultaneously. Interestingly, this approach can not surpass the registration and segmentation fidelity of the algorithms presented in chapters 4 and 5. This can be attributed to the cubic complexity of CBRR with regard to the number of images, in contrast to the quadratic complexity of the other two methods. This necessitates a coarser control grid resolution (8 px in CBRR vs. 4 px in MRsegFuse), which prohibits an explicit computation of high-frequency deformation detail. However, the registrations computed by CBRR have the benefit of being optimized for registration consistency, which can be useful in certain scenarios as discussed in the following.
The algorithms presented in this thesis have been demonstrated to improve the fidelity of registration-based segmentation methods using weak priors. It is important to consider the different use-cases for such automatically generated segmentations. While they can be used directly, for example to guide dose estimation in radiotherapy, a set of segmented images can also be used to generate strong prior knowledge, for example in the form of a statistical shape model. The use of such automatically generated segmentations in contrast to manual, gold-standard segmentations for the generation of strong prior knowledge has not yet been studied comprehensively and shall be a topic for further research. For the purpose of generating such models with strong prior knowledge, it is especially interesting to consider other additional information besides the segmentation results that is generated by the proposed algorithms. For example, the consistent registrations of CBRR can contribute to solving the challenging correspondence problem in statistical shape model generation from binary segmentations.

In the future, it will be interesting to further investigate both weak prior combinations and semi-supervised segmentation methods. As demonstrated in this thesis, it can be beneficial to formulate target criteria which are as similar to the intended goal as possible. Also, the exploration of correspondences between unsegmented images appears promising for the advancement of segmentation algorithms. While often being computationally challenging, pair-wise registrations between such unsegmented images may carry valuable information which have not yet been widely used. In this thesis, this approach has been demonstrated to benefit even small sets of images with strong variability, where common group-wise methods do not perform satisfactorily.
Bibliography


