Imperial College of Science, Technology and Medicine
Department of Computing

Fast and Robust Methods for Non-rigid Registration of Medical Images

Stefan Philippo Pszczolkowski Parraguez.

A dissertation submitted in part fulfilment of the requirements for the degree of

Doctor of Philosophy

of

Imperial College London

November 2014

Abstract

The automated analysis of medical images plays an increasingly significant part in many clinical applications. Image registration is an important and widely used technique in this context. Examples of its use include, but are not limited to: longitudinal studies, atlas construction, statistical analysis of populations and automatic or semi-automatic parcellation of structures. Although image registration has been subject of active research since the 1990s, it is a challenging topic with many issues that remain to be solved. This thesis seeks to address some of the open challenges of image registration by proposing fast and robust methods based on the widely utilised and well established registration framework of B-spline Free-Form Deformations (FFD).

In this work, a statistical method has been incorporated into the FFD model, in order to obtain a fast learning-based method that produces results that are in accordance with the underlying variability of the population under study. Several comparisons between different statistical analysis methods that can be used in this context are performed. Secondly, a method to improve the convergence of the B-Spline FFD method by learning a gradient projection using principal component analysis and linear regression is proposed. Furthermore, a robust similarity measure is proposed that enables the registration of images affected by intensity inhomogeneities and images with pathologies, e.g. lesions and/or tumours.

All the methods presented in this thesis have been extensively evaluated using both synthetic data and large datasets of real clinical data, such as Magnetic Resonance (MR) images of the brain and heart.

Acknowledgements

First, I would like to thank my supervisor, Daniel Rueckert, for his advice and help during the course of my PhD. It has been an honour to be one of his students. I'm very grateful for his guidance, motivation and infinite patience, especially when things do not go as intended.

I would also like to thank to all the people in the BioMedIA group for providing such a great atmosphere. Special thanks to my friends Andreas Schuh for his patience to explain things I did not understand and Christian Ledig for fruitful discussions, insights and loads of data for my experiments. Many thanks to Ricardo Guerrero for his inconditional friendship and for picking landmarks for me :). Finally, I would like to thank Luis Pizarro for his help during the first year of my PhD and Kevin Keraudren for providing me with figures for this thesis.

Many thanks as well to the National Commission for Scientific and Technological Research of Chile (CONICYT) for the funding they provided.

And last, but not least, I would like to thank my beloved best friend and wife Xaviera (Guatona) for all the love and encouragement, especially during hard times. These 5 years of marriage have been definitely the best of my life so far and I hope they will continue forever.

Dedication

Esta tesis está dedicada a mi amada Guatona y a nuestra Tere.

Declaration of originality

I declare that the work presented in this thesis is my own, unless specifically acknowledged.

Stefan Philippo Pszczolkowski Parraguez.

Copyright Declaration

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The analysis of medical images plays an increasingly significant part in many clinical applications. Patiens are nowadays imaged on a routine basis with a number of different imaging systems. Patients are also monitored over time to assess disease development or response to therapy. However, in order to be able to measure physiological and/or structural changes over time, or to combine the complementary information that the different imaging systems produce, it is necessary to perform *image registration* between the acquired images.

Image registration, also known as *spatial normalisation*, *motion estimation* and *image alignment*, is the process in which corresponding objects, structures or landmarks depicted in two or more images are put into spatial correspondence. In medical image registration, the images can come from different modalities such as X-rays, positron emission tomography (PET), ultrasound (US), magnetic resonance imaging (MRI), computed tomography (CT) or single-photon emission computed tomography (SPECT). A common approach for solving this problem is to consider one of the images as fixed and find the optimal spatial transformation that warps the other image into correspondence with it. The fixed image is referred to as the *target*, *study* or *reference* image, and the warped image is referred to as the *source*, *moving*, or *template* image. Throughout this thesis, we will refer to these images as the *target* and *source* images.

In a clinical context, the need of finding the optimal spatial transformation that places the target and source images into alignment can have several causes. The most common motivations are fusion of images from different modalities, motion between different scans, different patient positioning inside a scanner, disease progression, and the natural anatomical discrepancies between two different individuals. This makes image registration a critical procedure for many clinical workflows that involve medical images, including computer-aided diagnosis, computer-assisted interventions, and therapy planning and monitoring. Further details on these motivations will be given later in this chapter.

1.1 A generic image registration framework

Every image registration method can be decomposed into three main elements, which constitute the *building blocks* of a registration framework: A transformation model, a similarity measure and an optimisation. All these elements will be described in detail in the next chapter.

Transformation models characterise the spatial transformation being sought. An important distinction between *linear* and *non-linear* (also known as *non-rigid*) transformation models has to be made. While linear transformations models describe simple linear mappings such as translations, rotations, scalings and shears, non-rigid transformation models can represent mappings which are much more complex. The similarity measure aims to provide an estimation of how "similar" the images are given the current transformation. This measure can be based, for example, on points or surfaces within the images or, more commonly, on the intensity information. The optimisation seeks the optimal transformation parameters that maximise the similarity measure. These optimisation parameters can be regarded as being continuous or discrete. With this in mind, an image registration algorithm can be described by the interaction of the three building blocks in an iterative fashion, as shown in Figure 1.1.

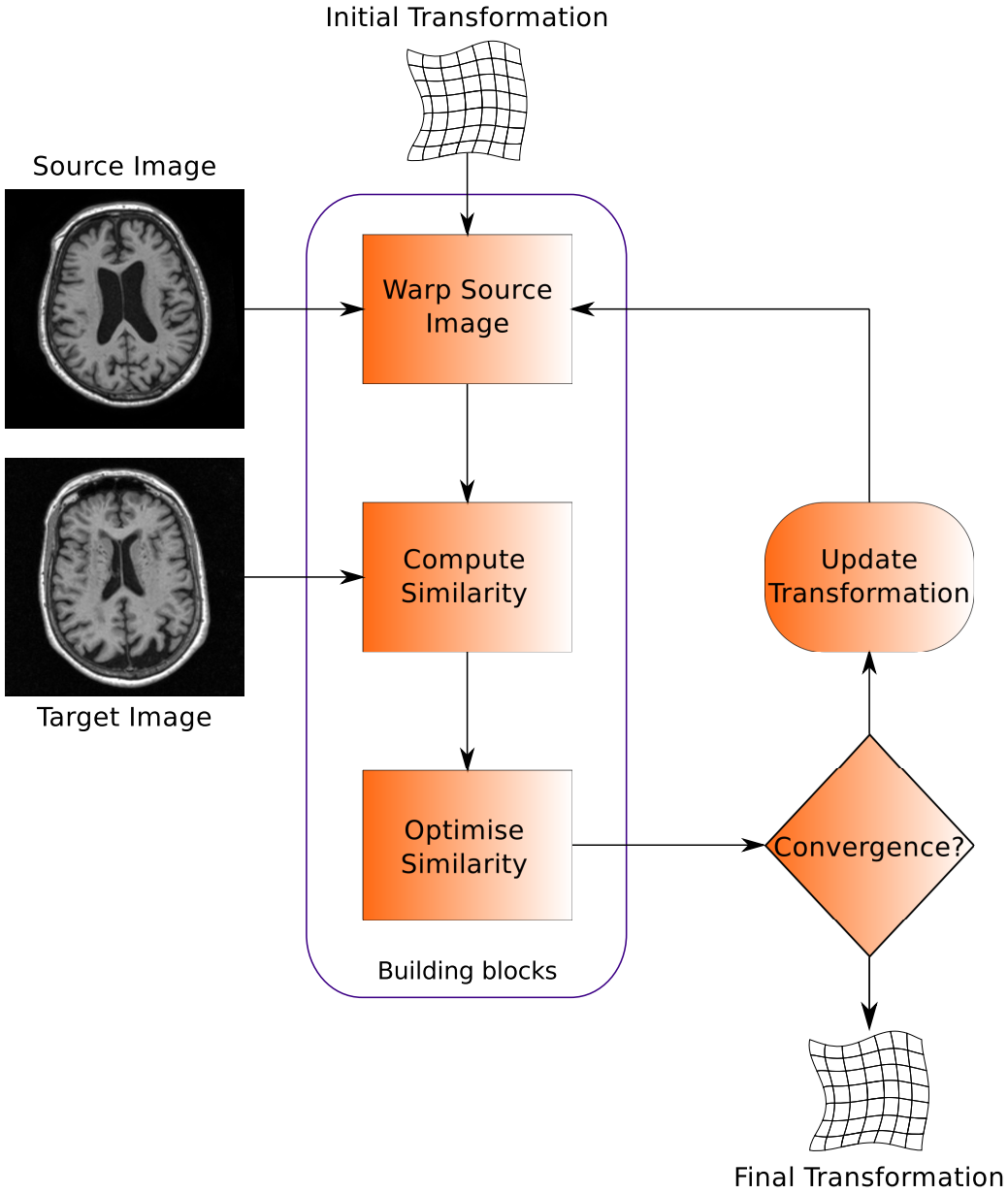


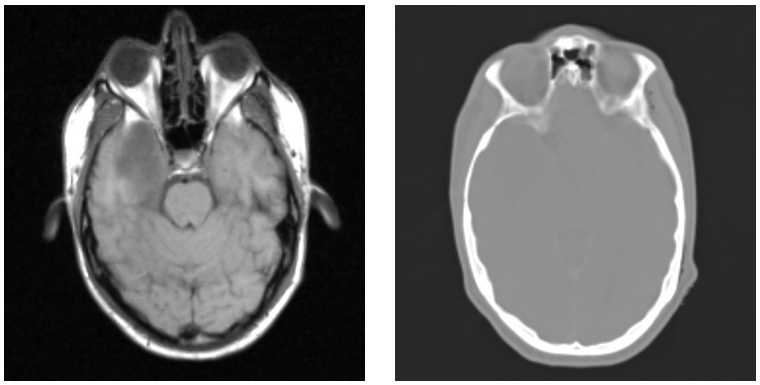
Figure 1.1: Framework of a generic image registration algorithm

1.2 Applications of medical image registration

1.2.1 Multimodal fusion

Different image modalities highlight different aspects of the imaged patient. For example, MR images provide information of the soft tissues, whereas CT images emphasise denser structures such as bones, as can be visualised in Figure 1.2. By fusing images from these two modalities a

combined image is obtained, which is clinically more useful than any of the original images alone. Another example of image fusion is to mix the anatomical information provided by MRI or CT with functional information given by PET or fMRI, in order to have more information to assess or diagnose diseases (see Figure 1.3). In order to be able to fuse two images, they have to be in spatial correspondence. Thus, an image registration process between them is key. Applications of medical image fusion include, but are not limited to: The study of Alzheimer's Disease [20], thermal ablation of prostate cancer [54], head and neck cancer treatment [213, 176, 152], breast cancer diagnosis [23], pancreatic cancer diagnosis [137], lung cancer treatment [109], prostate cancer brachytherapy [3, 159, 208, 207, 117, 138], rectal cancer diagnosis [64, 5], and computer-assisted neurosurgery [153]. We refer the interested reader to a survey of medical image fusion methods recently given by James and Dasarathy [104].



(a) (b)

Figure 1.2: Example of the different appearences of CT and MR images. (a) T2-weighted MR image. (b) CT image.

1.2.2 Longitudinal studies

Longitudinal studies are aimed at the imaging of individuals at different points in time. The image taken at the first point in time is usually called the *baseline* image, and the subsequent scannings are referred to as *follow-up* images. As an example, Figure 1.5 shows baseline and 24-month follow-up MR images of the brain, together with the corresponding voxel-wise measurement of volume change between both time points. By registering images of two different

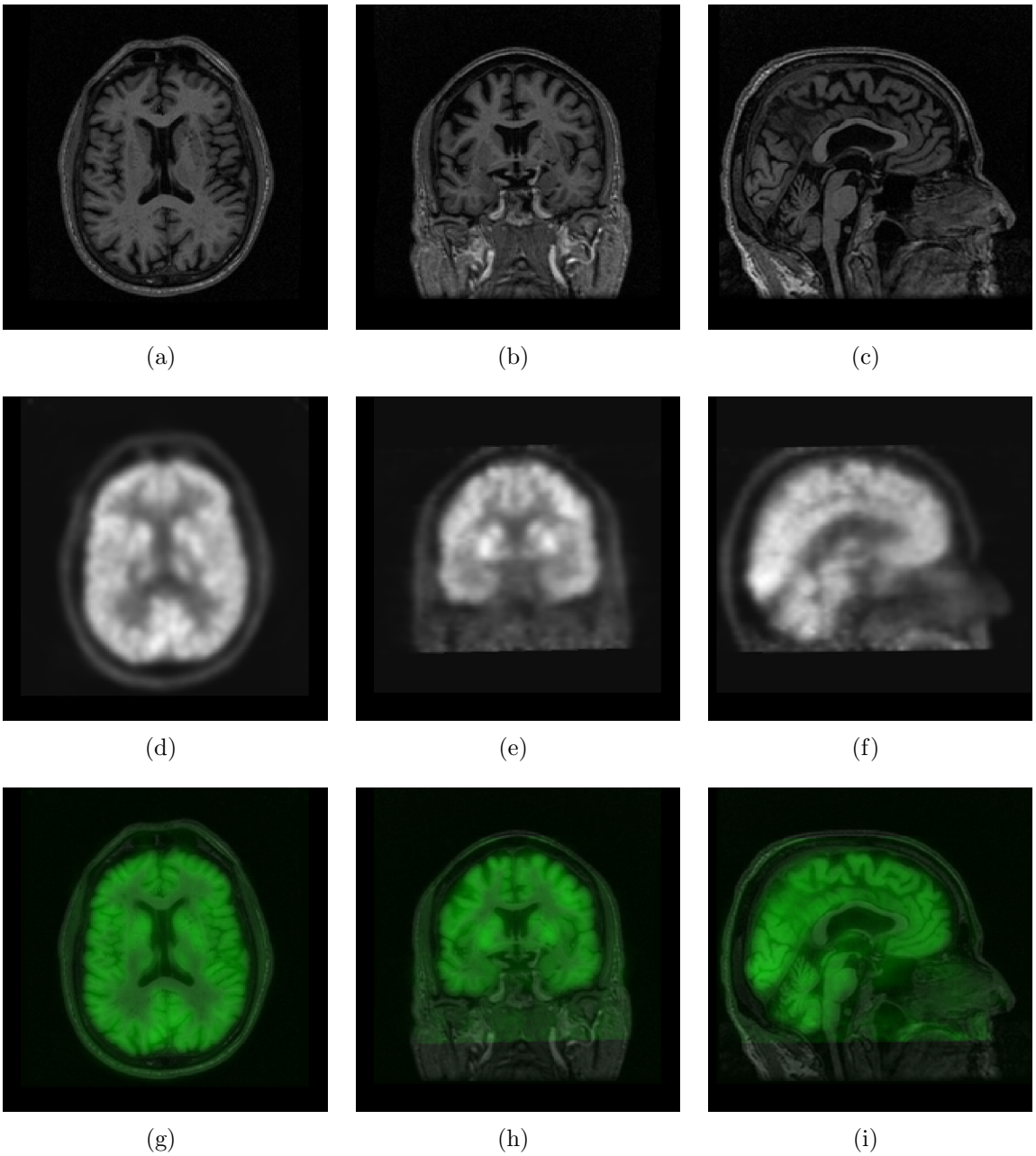


Figure 1.3: Example of image fusion of an MR and a PET image using registration. (a)-(c) MR image. (d)-(f) PET image (g)-(i) Fusion of MR image with PET image in green colour.

timepoints, a quantification of the changes occuring over that period of time can be established. For example, brain atrophy is an important biomarker in neurodegenerative diseases such as Alzheimer's disease [230, 210, 103]. Other examples of clinically important quantification performed with longitudinal studies are tumour progression [6, 72, 240, 229], analysis of the amount and location of heart dyssynchrony [50] and measurement of myocardial mechanics

[236].

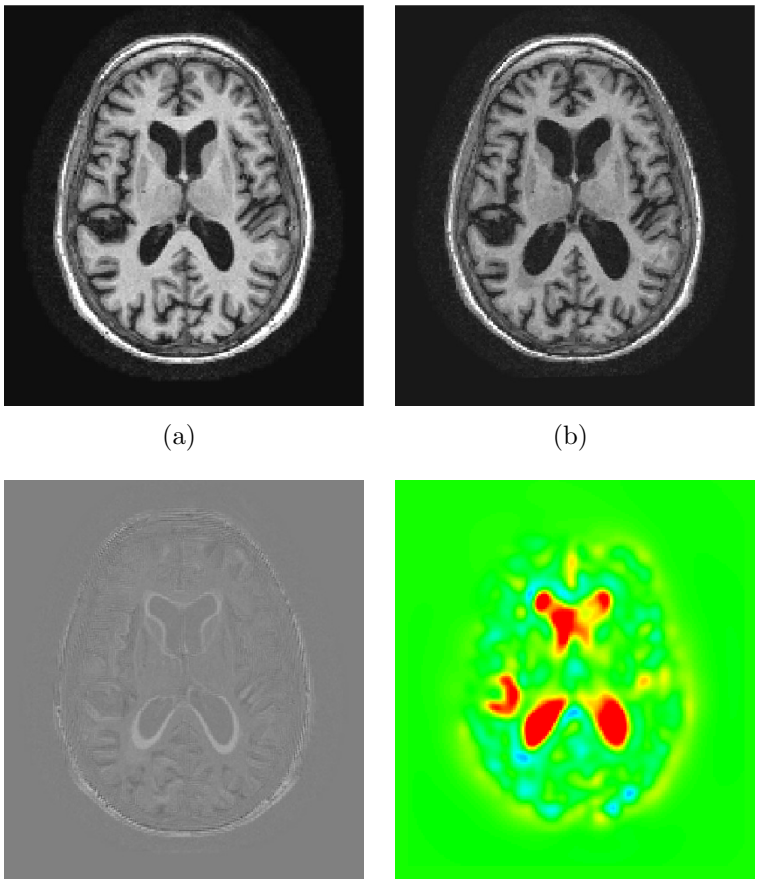
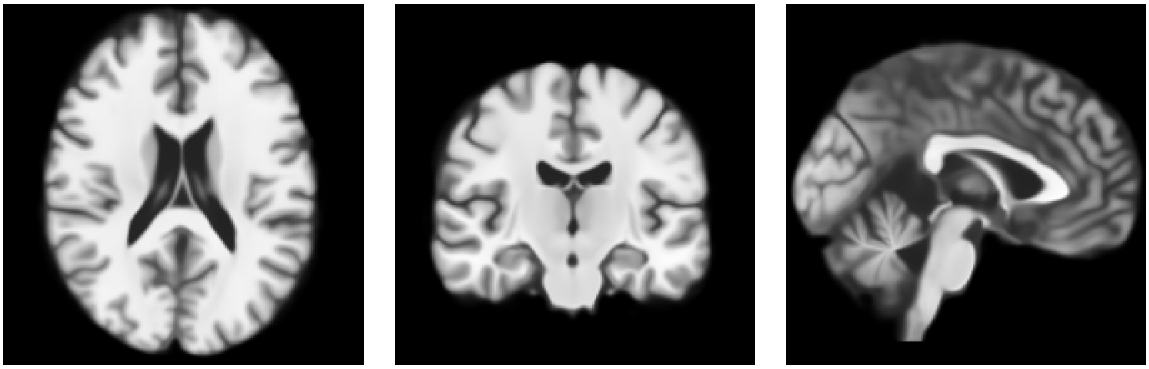


Figure 1.4: Example of a baseline and 24-month follow-up image with their corresponding voxel-wise measurement of volume change. (a) Baseline MR image. (b) Follow-up MR image. (c) Difference image after affine registration. (d) Volume change scaled between 50% contraction (blue) and 50% expansion (red).

1.2.3 Cross-sectional studies

Cross-sectional studies are mainly performed by means of computational anatomy, e.g., for quantitative analysis of the shape variability and morphometry within a population. These procedures usually involve the registration of a large group of subjects to a common template. This allows clinicians and researchers to systematically investigate morphometric differences across population groups and also permits the creation of so-called *templates* that represent the average anatomy of the group. An average template of the brain computed from 820

subjects including patients with Alzheimer's disease (AD), Mild Cognitive Impairment (MCI) and healthy controls is shown in Figure 1.5.



(a) (b) (c)

Figure 1.5: An average atlas of the brain computed from 820 subjects including patients with Alzheimer's disease, Mild Cognitive Impairment and healthy controls. (a) Axial view. (b) Coronal view. (c) Sagital view.

1.2.4 Segmentation through registration

Segmentation of anatomical regions of interest within medical images is of clinical interest. One of the most common ways of automatically segment an image is to register it to an image with pre-computed segmentations (usually carried out manually by an expert) and warp them using the resulting transformations [242, 243]. The images with pre-computed segmentations are usually refer to as *atlases*. Fusing techniques between several registrations are also often utilised to obtain a more refined automatic segmentation [83, 11, 123]. Figure 1.6 shows an example of segmentation fusion over 32 registrations using a simple majority vote scheme.

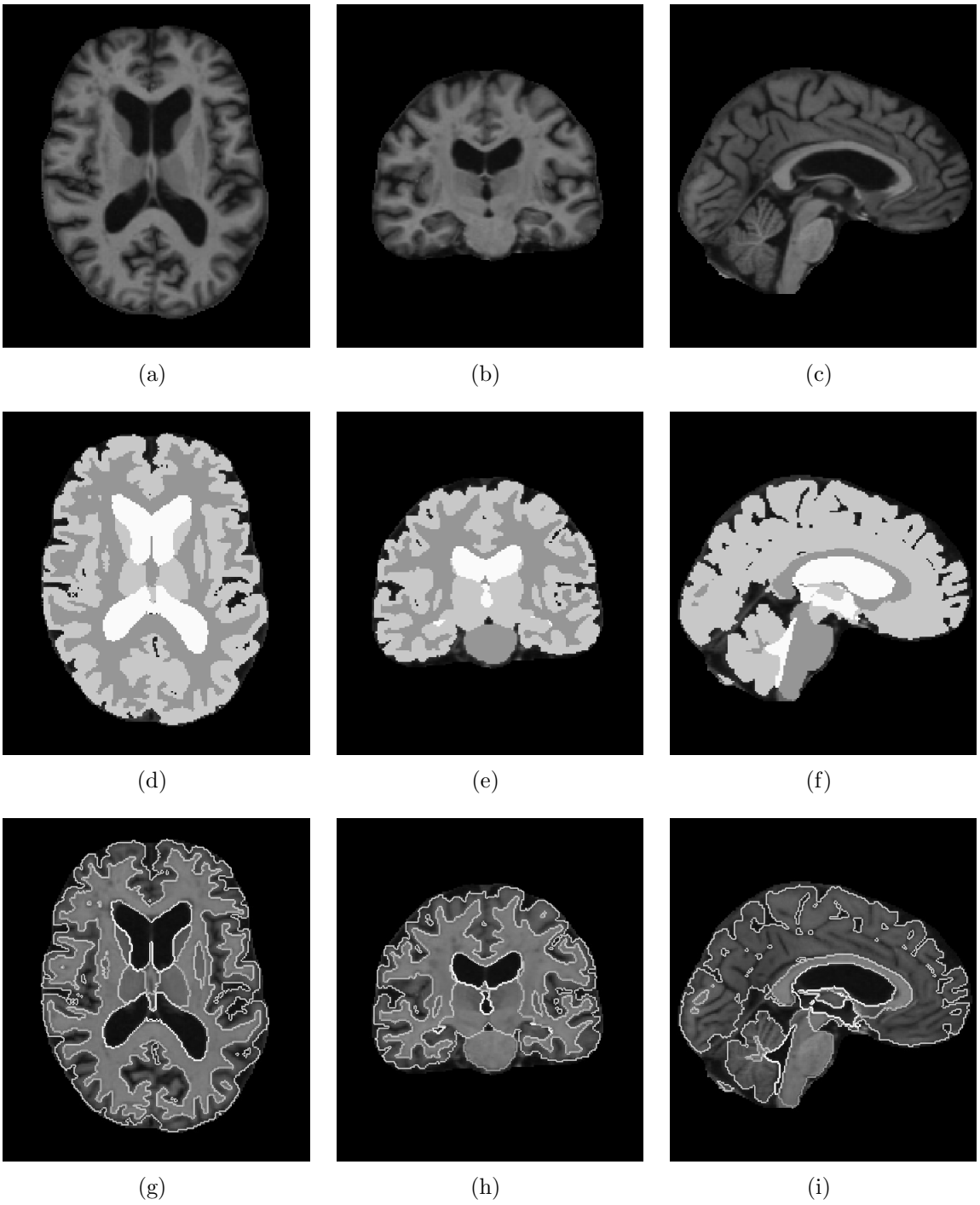


Figure 1.6: Result of atlas fusion over 33 registrations using a simple majority vote scheme. (a)-(c) Original target image. (d)-(f) Original target image with superimposed labels. (g)-(i) Original target image with superimposed label contours.

1.3 Main challenges of medical image registration

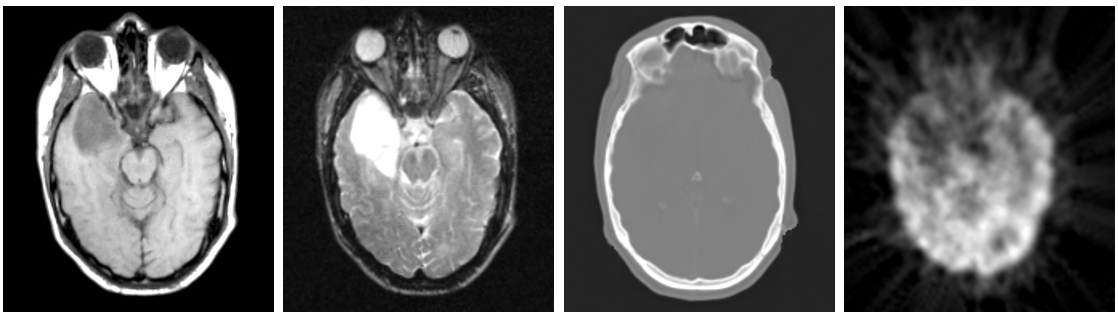
1.3.1 Computational challenges

Most medical applications which involve registration require the estimation of trasformations that are non-rigid. Non-rigid image registration is usually a time-consuming process. Furthermore, the development of medical imaging acquisition techniques in recent years means that nowadays 3D medical images have a number of voxels in the order of tens of millions. This means that the underlying non-rigid transformations may need to be defined by a number of parameters which is at least in the order of hundreds of thousands to produce sensible matchings. There is normally also a tradeoff between accuracy and speed, where a compromise has to be made. This compromise tends to be application-driven. For example, image-guided surgery usually requires real-time registrations, whereas computational anatomy or longitudinal studies can be performed during days or sometimes even weeks without compromising the clinical applicability of the outcomes. Finally, a growing number of registration approaches based on graphical processing units (GPUs) or CPU multicore paralellisation have been proposed [190, 58].

1.3.2 Multimodality challenges

As mentioned before, different modalities in medical imaging are utilised to characterise different aspects of the patient being imaged. Although this opens the possibility to fuse these different types of information, this also poses great challenges from an image registration point of view. This is because images from different modalities are usually acquired with different scanners and thus at different points in time. Hence, the anatomical features of the images might have different spatial arrangements due to motion (e.g. abdominal scans). Furthermore, different modalities show different anatomical or functional properties of the patient being imaged, as shown in Figure 1.7. This makes naïve intensity based registration methods unsuitable, since

the fusion is not clinically relevant if the images are not correctly registered. The rise of multimodal hybrid scanners (e.g. PET-CT or PET-MR scanners) which image the patient using more than one modality at the same time has partially alleviated this problem. However, there are still many clinical centres that rely on more than one scanner for diagnosis.



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

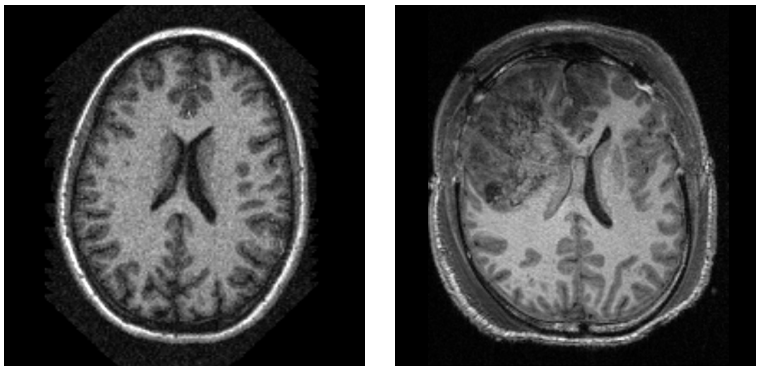
Figure 1.7: Example of the same patient scanned with different image modalities. (a) T1-weighted MR. (b) T2-weighted MR, (c) CT, (d) PET image. Images from the Cancer Imaging Archive (`public.cancerimagingarchive.net`)

1.3.3 Ill-posedness

As previouly mentioned, image registration involves an optimisation on a search space of a dimensionality that can be in the order of hundreds of thousands or even millions, if the transformation to be estimated is non-rigid. This makes non-rigid registration an ill-posed problem in the Hadamard sense [76]. Hadamard states three conditions for a problem to be well-posed: The existence of a solution, the uniqueness of a solution, and the continuous dependency of the solution on the initial conditions (a small change in the initial conditions must lead to a small change in the solution). Non-rigid registration problems usually violate the last two conditions. As a consequence, regularisation terms or models are needed in order to reduce the space of solutions as much as possible and obtain stable results. However, even though a substantial amount of research has been devoted to different regularisation models [177, 132, 180, 201, 214], it still remains an open problem that has to be taken into account when designing registration algorithms.

1.3.4 Ambiguous correspondences

Ambiguous correspondences between two medical images arise when one of them depicts biological features not present in the other. For example, when registering a brain image of a healthy subject with a brain image of a subject with brain pathology, such as lesions or tumours. These ambiguous correspondences can be challenging image registration methods and can lead to these methods performing unexpectedly in those areas. This is especially true for intensity-based registration approaches. Figure 1.8 shows an example of a healthy brain and another of a patient with traumatic brain injury (TBI). Note the presence of a considerable area in the upper-left corner of the TBI subject that has not a clear correspondence in the healthy brain.



(a) (b)

Figure 1.8: Example of ambiguous correspondence between a normal healthy brain and a brain with traumatic brain injury (TBI). Note how the TBI subject's anatomy deviates from that of the healthy subject. (a) Healthy brain. (b) TBI brain. Images from the TBIcare study (`www.tbicare.eu`)

1.3.5 Image acquisition artefacts

All medical image acquisition techniques can produce artefacts, such as noise, motion artefacts and intensity inhomogeneities. As a consequence, image registration techniques must be designed to be as robust as possible to these type of image acquisition artefacts. The following sections review the most common types of artefacts in medical imaging.

Noise

Noise is an inherent artefact in medical imaging. Even though the acquisition parameters of a scanner may be tuned to minimise this artefacts, they are seldom completely removed. Therefore, this issue has to be taken into account when designing image registration methods. Figure 1.9 illustrates an MR image of the brain with a considerable amount of noise artefacts.

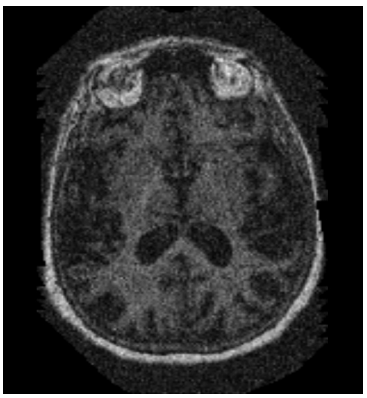


Figure 1.9: Example of noise in an MR image of the brain. Image registration methods, especially gradient-based ones must take these type of artefacts into account.

Motion artefacts

The motion of patients inside the scanner may produce misalignment between acquisition slices, which is usually problematic for registration algorithms. Furthermore, natural motion such as cardiac or respiratory motion may also be troublesome. A rather extreme case is when scanning the fetus inside the womb. This occurs since the fetus is often moving inside the womb while being scanned. Figure 1.10 illustrates this issue. Note the severe slice misalignment in the coronal and sagital views.

Intensity inhomogeneities

Intensity inhomogeneities correspond to a variation in intensity as a result of spatial position. These variations in intensity can usually be modelled as a multiplicative bias field. This type of

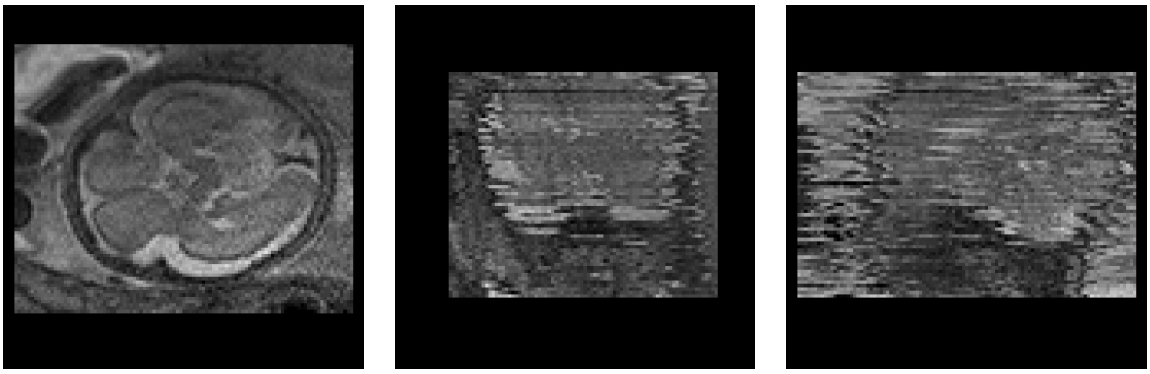
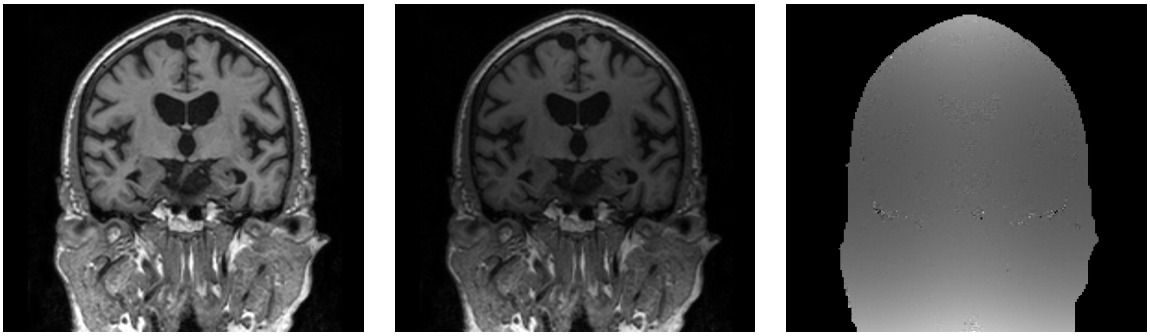


Figure 1.10: Example of motion artefact produced by the motion of the fetus inside the womb. This type of artefact is produced by the movement of the fetus during acquisition. This usually produces severe slice misalignment in the coronal and sagital views.

artefact is often produced by Magnetic Resonance (MR) scanners. The main causes for these artefacts to occur is due to inhomogeneities of the magnetic field of the scanner and the patient's position. Intensity inhomogeneities can hamper the robustness and accuracy of intensity-based registrations considerably, since the intensity profile in the images is not spatially consistent. This issue can be seen in Figure 1.11.



(a) (b) (c)

Figure 1.11: Example of intensity inhomogeneities in MR images. (a) Image without inhomogeneities. (b) Image with inhomogeneities. (c) Bias field.

1.4 Thesis contributions

This thesis aims to provide methods that tackle some of the challenges of image registration presented in the previous section. The main contributions of this thesis are given in chapters 4

through 6 and are focused on providing novel methods for each of the building blocks in the image registration framework, namely transformation model, similarity measure and optimisation. These contributions are as follows:

- The development of a learning-based statistical registration method, which extends a non-rigid transformation model to incorporate information in the form of a **statistical deformation model** (SDM) that accounts for the mean and variability of deformations across a population. The SDM is learned by applying statistical analysis techniques over precomputed training transformation data. The learnt SDM is then utilised to drive the registration in order to obtain deformations that are in accordance with the training population and can be parameterised using significantly fewer degrees of freedom. The proposed registration method is also considerably faster than most standard registration methods. We also show that the SDM can encode information to better register important landmarks. Comparisons between different method to learn the SDMs are also provided. Parts of this chapter were previously reported in [170].

- A learning-based approach to **improve the convergence of the optimisation** of any chosen parametric image registration method is proposed. From a set of training images and their corresponding deformations, the proposed registration method learns offline a projection from the gradient space of the similarity measure to the parameter space of the chosen registration method using PCA dimensionality reduction and standard least squares regression. The proposed learning-based gradients are subsequently used online to approximate the optimisation of the energy functional for unseen images. A state-of-the-art parametric approach is utilised as underlying registration method, but other parametric methods can be used as well. This chapter is based on a previous 2D-only version published in [171].

- A **novel similarity measure** based on normalised gradients for non-rigid registration is proposed. This new similarity is robust in images with intensity inhomogeneities or outliers (e.g. lesions or pathologies). We provide both theoretical and experimental proof

of the robustness and evaluate the approach on manually segmented images, as well as images with real and simulated pathologies, and images with manually annotated landmarks. Compared to commonly used similarity measures, we obtain significant improvements in terms of overlap of anatomical structures for images with intensity inhomogeneities. We also show that the proposed measure is robust to the presence of pathologies such as tumours or lesions in the images being registered. A preliminary version of this work can be found in [172].

1.5 Thesis Overview

Chapter 2 provides a detailed overview of the main components of image registration methods. Examples of how some of these components are designed in different registration approaches are also presented. **Chapter 3** provides an extensive review of the state-of-the-art in non-rigid medical image registration. **Chapter 4** describes a learning-based statistical registration method that requires substantially less degrees of freedom to parameterise deformations than a standard registration method without compromising its accuracy, by constraining the deformation to comply with a statistical deformation model that accounts for the mean and variability of a population of subjects. This method is evaluated on manually segmented and landmark-annotated MR images of the brain, as well as on cardiac MR images. In **Chapter 5**, a learning-based registration method is proposed. This method improves the convergence of standard registration approaches, by learning the projection of the energy gradient with respect to the deformation field into parametric space using PCA dimensionality reduction and standard least squares regression. We demonstrate this method on cardiac MR data as well as on MR images of the brain. **Chapter 6** outlines a similarity measure based on normalised gradient fields. This similarity shows robustness to the presence of intensity inhomogeneities as well as robustness to ambiguous correspondences due to the presence of tumours or lesions in the images. We evaluate the proposed similarity on cardiac data, on manually segmented and landmark-annotated MR images of the brain, on images with simulated tumours and on

images of a patient with traumatic brain injury. Finally, **Chapter 7** gives the conclusions, future work and final remarks of this thesis.

Chapter 2

Background

As previously mentioned, image registration comprises three main aspects or building blocks: (1) A transformation model that maps points in the target image into points in the source image, (2) a measure of how "similar" the two images are given the current transformation and, (3) an optimisation strategy for maximising the similarity functional as a function of the transformation parameters. In this chapter we give a detailed overview of each one of these components of image registration.

2.1 Transformation models

During the process of image registration, the spatial transformation between two images is iteratively estimated. This transformation maps every point $\mathbf{x} = (x, y, z)^\top$ of the target image I_0 to its corresponding point $\mathbf{y} = \mathbf{T}(x, y, z) = (u, v, w)^\top$ in the source image I. This correspondence is constrained to be functionally and/or anatomically meaningful. The mapping $\mathbf{T}$ allows the intensity of the source image at point $\mathbf{y}$ to be estimated (e.g. via interpolation) and associated with the intensity at point $\mathbf{x}$ of the target image. If this process is performed for every voxel in the target image, a warped or transformed version of the source image $I_1^{\mathbf{T}} = I \circ \mathbf{T}$ is obtained. If the spatial transformation is the optimal one, we say that image I is *registered to* image I_0 and

$I_1^{\mathbf{T}}$ is the "registered version" of I. It is important to bear in mind that the point $\mathbf{y}$ does not necessarily coincide with a discrete voxel location. Thus, an interpolation of the source image intensities around $\mathbf{y}$ must be performed in order to obtain the intensity that will be associated to the target point $\mathbf{x}$. An example of a transformation between two MR images of the brain can be seen in Figure 2.1.

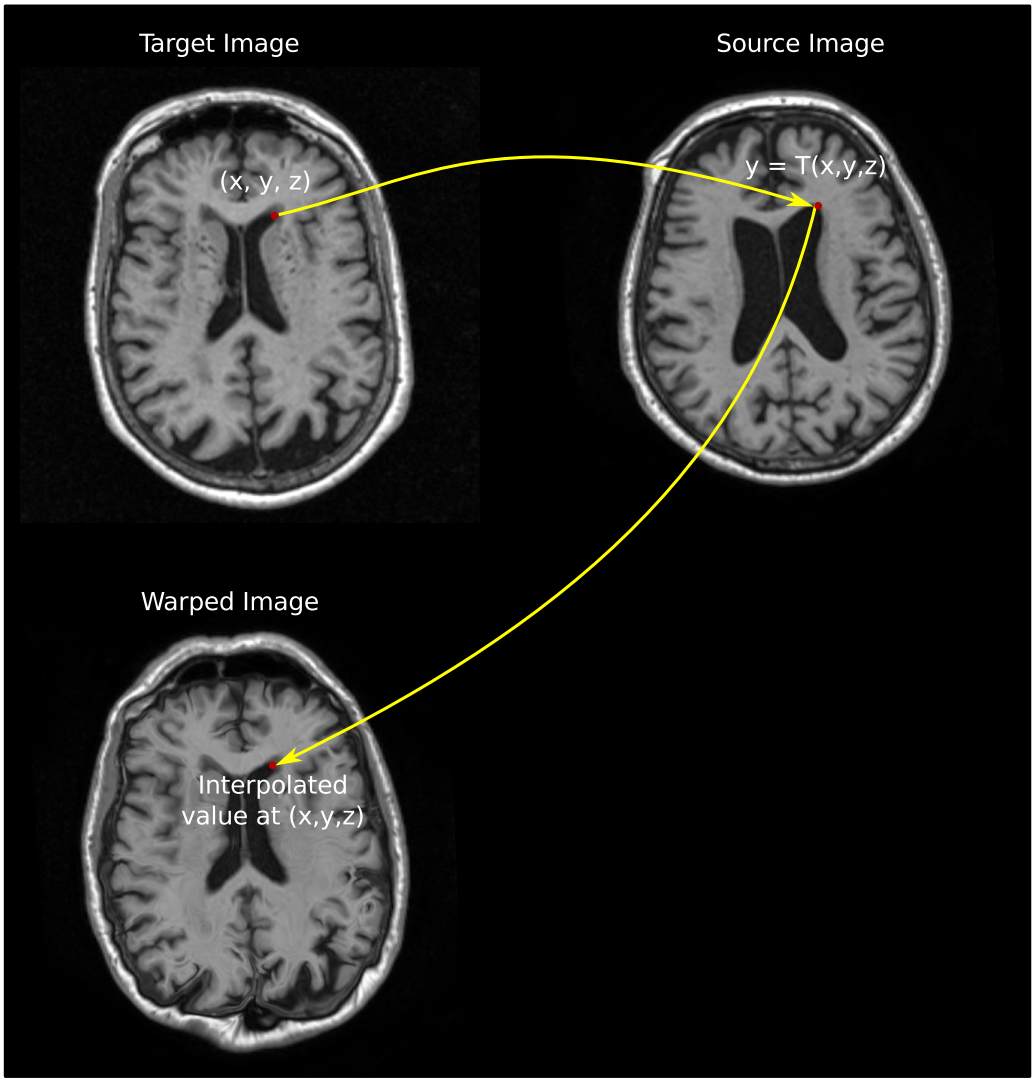
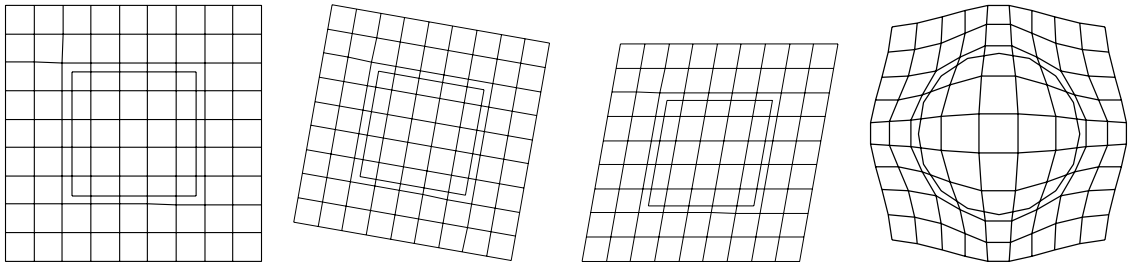


Figure 2.1: A point $\mathbf{x}$ in a target image I_0 is mapped onto its corresponding point $\mathbf{y} = \mathbf{T}(\mathbf{x})$ on a source image I by applying the transformation $\mathbf{T}$. The warped image $I_1^{\mathbf{T}}$ is then obtained by interpolation in the neighbourhood of $\mathbf{y}$. In this example the registration was performed only in the brain area.

In addition to the identity transformation, there are three generic classes of geometric transformations that can be used to represent the underlying correspondences: rigid transformations, affine transformations and non-rigid transformations. Figure 2.2 shows an example of these

transformation types, which will be discussed in more detail in the following.



(a) Identity (b) Rigid (c) Affine (d) Non-rigid

Figure 2.2: Different types of transformations applied to a square. Figure adopted from [77].

2.1.1 Rigid transformations

Rigid transformations can be represented by the application of rotations and translations. Under this type of transformation, lengths and angles are preserved. In the 2D case, images are embedded in a plane, thus the only rotation that has to be considered is the rotation about the axis that is perpendicular to that plane. In the following, we use homogeneous coordinates to represent points, in order to represent transformations using matrix notation. In this type of coordinates, an N-dimensional Euclidean point $\mathbf{a}_{euc} = (a_1, \cdots, a_N)$ is represented by a N+1-dimensional vector $\mathbf{a}_{hom} = (b \cdot a_1, \cdots, b \cdot a_N, b)$, where $b \neq 0$. Thus, $\mathbf{a}_{hom}$ is the *homogeneous coordinate* of $\mathbf{a}_{euc}$. In the following, we adopt the most common choice for b, which is $b = 1$.

An anti-clockwise rotation of the xy-plane about the z-axis by an angle γ is given by:

$$\mathbf{R}_z^{\mathbb{R}^2}(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and translations are represented by a vector that represents a point in $\mathbb{R}^2$:

$$\mathbf{v}^{\mathbb{R}^2} = \begin{pmatrix} t_x \\ t_y \\ 1 \end{pmatrix}$$

Hence, the rigid transformation that maps a point $(x, y, 1)^T$ into a point $(u, v, 1)^T$ can be written in matrix form as

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\gamma & -\sin\gamma & t_x \\ \sin\gamma & \cos\gamma & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ v \\ 1 \end{pmatrix}$$

It can be seen that this transformation can be represented using three parameters: γ, t_x, t_y.

The 3D case of rigid transformations is more complex, since there are a greater number of degrees of freedom for rotations and translations. Rotations can be decomposed into rotations about each of the x-, y- and z-axes, which are called roll, pitch and yaw, respectively; while translations need three parameters to be represented. The rotations about each coordinate axis are given by

$$\mathbf{R}_x^{\mathbb{R}^3}(\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha & 0 \\ 0 & \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R}_y^{\mathbb{R}^3}(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R}_z^{\mathbb{R}^3}(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 & 0 \\ \sin\gamma & \cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and translations are represented by

$$\mathbf{v}^{\mathbb{R}^3} = \begin{pmatrix} t_x \\ t_y \\ t_z \\ 1 \end{pmatrix}$$

Without loss of generality, if we consider that the x-, y- and z-axes as *fixed*, i.e., not affected by rotations, a 3D rotation can be represented by considering a rotation with respect to the z-axis, then a rotation with respect to the y-axis, and finally a rotation with respect to the x-axis

$$\mathbf{R}^{\mathbb{R}^3}(\alpha, \beta, \gamma) = \mathbf{R}_x^{\mathbb{R}^3}(\alpha)\mathbf{R}_y^{\mathbb{R}^3}(\beta)\mathbf{R}_z^{\mathbb{R}^3}(\gamma)$$

Hence, the rigid transformation that maps a point $(x, y, z, 1)^T$ into a point $(u, v, w, 1)^T$ can be written as

$$\begin{pmatrix} u \\ v \\ w \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

where

$$\begin{aligned}
r_{11} &= \cos\beta\cos\gamma \\
r_{12} &= -\cos\alpha\sin\gamma + \sin\alpha\sin\beta\cos\gamma \\
r_{13} &= \sin\alpha\sin\gamma + \cos\alpha\sin\beta\cos\gamma \\
r_{21} &= \cos\beta\sin\gamma \\
r_{22} &= \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma \\
r_{23} &= -\sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma \\
r_{31} &= -\sin\beta \\
r_{32} &= \sin\alpha\cos\beta \\
r_{33} &= \cos\alpha\cos\beta
\end{aligned}$$

It can be seen that this transformation can be represented using 6 parameters: α, β, γ, t_x, t_y, t_z.

There are also other possible parameterisations that can be used to represent spatial rotations. One of them are *quaternions* [78, 150]. The advantages of using quaternions compared to Euler angles (α, β and γ) are that compositions between them are simpler and they do not have the problem of a *gimbal lock* [228] ocurring. A gimbal lock corresponds to the loss of one degree of freedom when one or more of the rotation matrices become the identity transform. For example, $\mathbf{R}_x(\alpha)$ coincides with the identity transform when $\alpha = 0$.

2.1.2 Affine transformations

In addition to rotations and translations, affine transformations account for scalings and shears. Under this type of transformation, parallel lines are preserved, but lengths and angles may vary. As rotations, scalings can be represented as a composition of individual scalings on each axis. The 2D case can be written as

$$\mathbf{S}^{\mathbb{R}^2}(s_x, s_y) = \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and scaling in 3D can be expressed as

$$\mathbf{S}^{\mathbb{R}^3}(s_x, s_y, s_z) = \begin{pmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where s_x, s_y and s_z are the scalings with respect to the x-, y- and z-axes, respectively.

In addition, shears in 2D with respect to the x-axis in the xy-plane can be represented as

$$\mathbf{H}^{\mathbb{R}^2}_{xy-x}(\delta) = \begin{pmatrix} 1 & \tan\delta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that shears with respect to the x-axis in the xy-plane are sufficient to represent any shear in 2D, since a shear with respect to the y-axis in the xy-plane can be computed as:

$$\mathbf{H}^{\mathbb{R}^2}_{xy-y}(\delta) = \mathbf{R}^{\mathbb{R}^2}_{z}(\pi/2)\mathbf{H}^{\mathbb{R}^2}_{xy-x}(-\delta)\mathbf{R}^{\mathbb{R}^2}_{z}(-\pi/2). \tag{2.1}$$

Hence, in general, the affine transformation that maps a 2D point $(x, y, 1)^T$ into a point $(u, v, 1)^T$ can be written as the multiplication of three matrices: one that accounts for rotations and translations, another that accounts for shearing and a last one that accounts for scaling:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\gamma & -\sin\gamma & t_x \\ \sin\gamma & \cos\gamma & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \tan\delta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

thus, it can be represented using 6 parameters: γ, t_x, t_y, δ, s_x, s_y.

In the 3D case, the shears with respect to the x-axis in the xy-plane, with respect to the x-axis in the xz-plane, and with respect to the y-axis in the yz-plane can be respectively represented as

$$\mathbf{H}^{\mathbb{R}^3}_{xy-x}(\delta) = \begin{pmatrix} 1 & \tan\delta & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{H}^{\mathbb{R}^3}_{xz-x}(\mu) = \begin{pmatrix} 1 & 0 & \tan\mu & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{H}^{\mathbb{R}^3}_{yz-y}(\nu) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \tan\nu & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where δ, μ and ν are the *shear angles* in the xy-, xz- and yz-planes, respectively. Analogous to the 2D case, the other possible 3D shears can be computed as:

$$\begin{aligned} \mathbf{H}^{\mathbb{R}^3}_{xy-y}(\delta) &= \mathbf{R}^{\mathbb{R}^3}_{z}(\pi/2)\mathbf{H}^{\mathbb{R}^3}_{xy-x}(-\delta)\mathbf{R}^{\mathbb{R}^3}_{z}(-\pi/2) \\ \mathbf{H}^{\mathbb{R}^3}_{xz-z}(\mu) &= \mathbf{R}^{\mathbb{R}^3}_{y}(\pi/2)\mathbf{H}^{\mathbb{R}^3}_{xz-x}(-\mu)\mathbf{R}^{\mathbb{R}^3}_{y}(-\pi/2) \\ \mathbf{H}^{\mathbb{R}^3}_{yz-y}(\nu) &= \mathbf{R}^{\mathbb{R}^3}_{x}(\pi/2)\mathbf{H}^{\mathbb{R}^3}_{yz-z}(-\nu)\mathbf{R}^{\mathbb{R}^3}_{x}(-\pi/2). \end{aligned} \tag{2.2}$$

Without loss of generality, an arbitrary 3D shear can be represented as a matrix multiplication of 3 individual shears as:

$$\begin{aligned} \mathbf{H}^{\mathbb{R}^3}(\delta,\mu,\nu) &= \mathbf{H}^{\mathbb{R}^3}_{yz-z}(\nu)\mathbf{H}^{\mathbb{R}^3}_{xz-x}(\mu)\mathbf{H}^{\mathbb{R}^3}_{xy-x}(\delta) \\ &= \begin{pmatrix} 1 & \tan\delta & \tan\mu & 0 \\ 0 & 1 & \tan\nu & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned} \tag{2.3}$$

Hence, in general, the affine transformation that maps a 3D point $(x, y, z, 1)^T$ into a point $(u, v, w, 1)^T$ is given by

$$\begin{pmatrix} u \\ v \\ w \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \tan\delta & \tan\mu & 0 \\ 0 & 1 & \tan\nu & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

and can be represented using 12 parameters: γ, α, β, t_x, t_y, t_z, δ, μ, ν, s_x, s_y, s_z. This can be summarised as

$$\begin{pmatrix} u \\ v \\ w \\ 1 \end{pmatrix} = \mathbf{A} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \qquad \text{or} \qquad \mathbf{y} = \mathbf{A}\mathbf{x},$$

with matrix $\mathbf{A}$ containing the 12 parameters of the affine transformation.

2.1.3 Parametric non-linear transformations

In this section, we describe parametric transformation models that deform the images in a non-linear (more commonly known as *non-rigid*) fashion and, hence, are defined by a greater number of parameters or degrees of freedom than rigid or affine transformations. In the case of medical imaging, the need of this type of transformations becomes apparent when considering that the underlying deformation of organs (due to respiration or progression of pathology, for example) normally does not follow a linear pattern, i.e, different regions in the image may deform in different ways. Moreover, this difference in deformation becomes even more apparent when performing registration of images from different subjects.

Thin-plate splines transformations

Thin-plate splines (TPS) are interpolating and smoothing functions which are a generalisation of 1D splines. They were first developed for the interpolation of scattered data [51, 143]. In the

context of medical image registration, the use of TPS was first proposed by Bookstein [26, 27]. Since then many other researchers have also used them [178, 179, 42, 247, 173]. The main idea of TPS is to find a smooth interpolating function that minimises the bending energy applied to a thin plate of metal. A TPS tranformation function for a point $\mathbf{x} = (x, y, z)^\top$ is given by one TPS for each $x-$, $y-$ and $z-$ component

$$T^\xi(\mathbf{x}) = A_0^\xi + A_1^\xi\, x + A_2^\xi\, y + A_3^\xi\, z + \sum_{i=1}^{N} w_i^\xi\, U(|\mathbf{x} - \Phi_i|), \quad \xi \in \{x, y, z\} \tag{2.4}$$

where A_i^ξ correspond to affine transformation coefficients and Φ_i are the locations of N landmarks (or control points) in the target image, at which the radial basis functions (RBFs) U are centered. These locations satisfy the interpolation condition:

$$\mathbf{T}(\Phi_i) = \Phi_i^{'}, \tag{2.5}$$

where $\Phi_i^{'}$ are the locations of the landmarks Φ_i in the source image.

The RBFs U can be written as:

$$U(s) = \begin{cases} |s|^2\, ln(|s|) & \text{in 2D} \\ |s| & \text{in 3D} \end{cases}. \tag{2.6}$$

In 3D, equation (2.5) provides $3N$ interpolation equations U. However, there are $3N + 12$ unknown coefficients ($3N$ non-rigid coefficients w and 12 affine coefficients A). In order to be able to solve the system of linear equations, twelve additional constraints are incorporated as follows:

- The sum of forces applied to the plate per component (non-rigid coefficients) is zero:

$$\sum_{i=1}^{N} w_i^\xi = 0, \qquad \xi \in \{x, y, z\} \tag{2.7}$$

- The moments with respect to each component are zero:

$$\sum_{i=1}^{N} \Phi_i^{\zeta} \, w_i^{\xi} = 0, \qquad \xi \in \{x, y, z\}, \qquad \zeta \in \{x, y, z\}. \tag{2.8}$$

In matrix form, the TPS in 3D can be formulated as:

$$\begin{pmatrix} \mathbf{U} & \mathbf{\Phi} \\ \mathbf{\Phi}^{\top} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{W} \\ \mathbf{A} \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}' \\ \mathbf{0} \end{pmatrix}, \tag{2.9}$$

where $\mathbf{A}$ is the 4×3 matrix of affine coefficients, $\mathbf{W}$ is the $N \times 3$ matrix of non-rigid coefficients, $\mathbf{\Phi}$ is the $N \times 4$ matrix of landmark locations in the target image (with the first column being a column of 1s), $\mathbf{\Phi'}$ is the $N \times 3$ matrix of landmark locations in the source image, and $\mathbf{U}$ is an $N \times N$ matrix such that $u_{ij} = U(|\Phi_i - \Phi_j|)$.

Rohr et al. [179] extends the interpolating thin-plate splines approach by considering the use of approximating thin-plate splines instead. The use of approximating TBS relaxes the condition of exact landmark matching of interpolating thin-plate splines (equation 2.5), in order to allow the registration to account for landmark localisation errors. In matrix form, the approximating TBS can be formulated as:

$$\begin{pmatrix} \mathbf{U} + \lambda \mathbf{S}^{-1} & \mathbf{\Phi} \\ \mathbf{\Phi}^{\top} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{W} \\ \mathbf{A} \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}' \\ \mathbf{0} \end{pmatrix}. \tag{2.10}$$

Here, λ is a tradeoff parameter between interpolation and approximation of landmark locations and $\mathbf{S} = \text{diag}\{1/\sigma_1^2, \cdots, 1/\sigma_N^2\}$. In this formulation, σ_i represents the uncertainty on the location of landmark i.

A disadvantage of the radial basis function of equation 2.6 is their global support. This means that purely local deformations are not easy to model. To overcome this issue, Fornefett et al.

[60] propose to use the locally supported ψ_{31}-functions of Wendland to define the RBS U:

$$U(s,\rho) = \left(1 - \frac{|s|}{\rho}\right)^4 \left(\frac{4|s|}{\rho} + 1\right). \tag{2.11}$$

Here, ρ is a radius parameter that controls the function support. This parameter can be fixed or adaptively computed for different clusters of landmarks according to cluster shape, as proposed by Shusharina and Sharp [195].

Free-form deformations (FFD)

Free-form deformations (FFD) are a powerful and efficient transformation model introduced in the context of computer graphics by Sederberg and Parry [189]. They are parameterised by a regular lattice of uniformly spaced *control points*. One of its main characteristics is that (in contrast to thin-plate splines) the displacements in each image point can be computed using blending functions with local support, i.e., only a small subset of the control points is needed for the computation. This provides a computationally efficient alternative, even when large number of control points are used to parameterise the transformation.

Let Φ denote a uniform lattice of size $N = n_x \times n_y \times n_z$ and $\mathbf{x} = (x, y, z)$ a point in the target coordinate space. FFD have the following form:

$$\mathbf{T}(\mathbf{x}) = \mathbf{x} + \sum_l \sum_m \sum_n p_{l,m,n}(s,t,u)\Phi_{i+l,j+m,k+n}, \tag{2.12}$$

where $p_{l,m,n}$ corresponds to the chosen blending function and i, j, k are given by

$$i = \lfloor \frac{x}{n_x} \rfloor - 1 \qquad j = \lfloor \frac{y}{n_y} \rfloor - 1 \qquad k = \lfloor \frac{z}{n_z} \rfloor - 1. \tag{2.13}$$

One possible choice of blending function is the Bernstein polynomial as in [189]. In this case, $b_{l,m,n}$ corresponds to the trivariate tensor product of Bernstein polynomials:

$$\mathbf{T}(\mathbf{x}) = \mathbf{x} + \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} P_l(s)P_m(t)P_n(u)\Phi_{i+l,j+m,k+n}, \tag{2.14}$$

where $0 < \{s, t, u\} < 1$ are the coordinates of $\mathbf{x}$ in a local coordinate system on a parallelpiped region [189] and

$$P_r(s) = \binom{3}{r} s^r (1-s)^{3-r} \tag{2.15}$$

are the 3rd order Bernstein polynomials (see Figure 2.3).

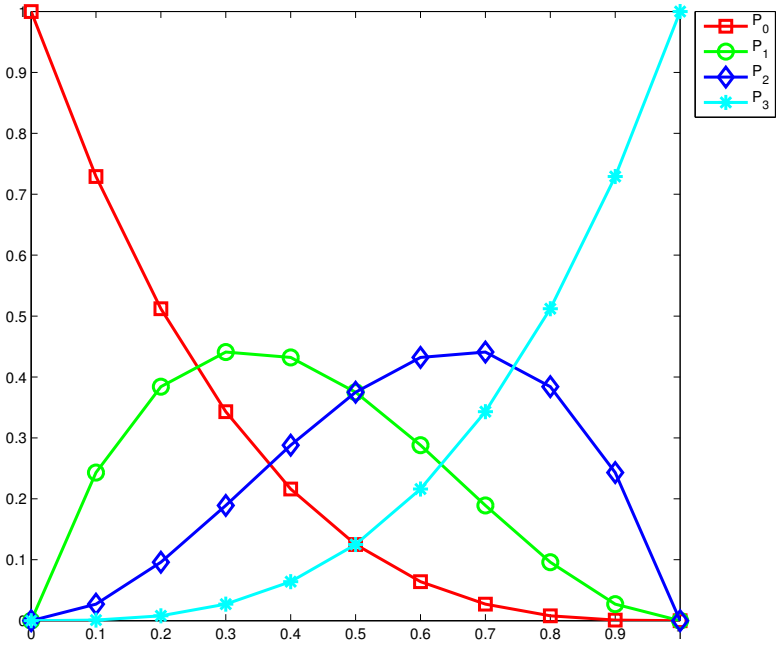


Figure 2.3: 3rd order Bernstein polynomials from equation (2.15).

An alternative choice of blending function are the locally supported cubic B-splines [125]. Here, $b_{l,m,n}$ is expressed as the 3D tensor product of 1D cubic B-splines

$$\mathbf{T}(\mathbf{x}) = \mathbf{x} + \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\Phi_{i+l,j+m,k+n}, \tag{2.16}$$

where $0 < \{u, v, w\} < 1$ are the relative positions of $\mathbf{x}$ in lattice coordinates

$$u = \frac{x}{n_x} - \lfloor\frac{x}{n_x}\rfloor \qquad v = \frac{y}{n_y} - \lfloor\frac{y}{n_y}\rfloor \qquad w = \frac{z}{n_z} - \lfloor\frac{z}{n_z}\rfloor. \tag{2.17}$$

and B_r corresponds to the r-th basis function of the B-spline (see Figure 2.4)

$$\begin{aligned} B_0(u) &= (1-u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6, \end{aligned} \tag{2.18}$$

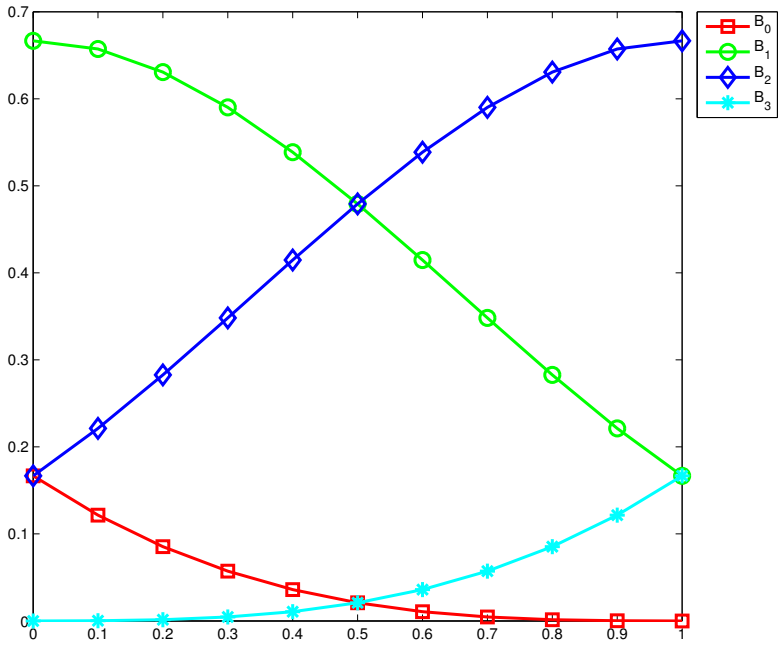


Figure 2.4: B-spline basis functions from equation (2.18).

The use of B-spline blending functions gained significant interest in medical imaging [46, 183, 120, 181, 188]. Furthermore, several extensions to the FFD B-spline model have been proposed [187, 226, 156, 55, 199, 194].

The transformation models in equations (2.14) and (2.16) assume both target and source images to be defined in the same coordinate space. In medical imaging this is usually not the case, hence a pre-registration step using a rigid or affine transformation model is usually carried out to transform the source image into the target coordinate space. To account for this, Rueckert et

al. [183] propose to use a transformation model that is the sum of a global affine tranformation and a local B-spline FFD displacement

$$\mathbf{T}(\mathbf{x}) = \mathbf{T}_{global}(\mathbf{x}) + \mathbf{T}_{local}(\mathbf{x}), \tag{2.19}$$

where

$$\mathbf{T}_{global}(\mathbf{x}) = \mathbf{A}\mathbf{x}, \tag{2.20}$$

and

$$\mathbf{T}_{local}(\mathbf{x}) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\Phi_{i+l,j+m,k+n}. \tag{2.21}$$

The registration proposed in [183] is usually performed in a multiresolution coarse-to-fine fashion. The idea behind this algorithm is to first use an FFD with a large control point spacing to capture global deformations. Then, the resulting FFD mesh is subdivided into an equivalent mesh with half the control point spacing [61] to capture more localised deformations. This process is repeated for a user-specified number of resolution levels. To make the registration faster, an image pyramid is used in which the images are blurred and downsampled according to the resolution level. Figure 2.5 shows an example of a source image being warped into the target coordinate space via a multiresolution B-spline FFD.

It is important to mention that although B-spline blending functions produce a smooth interpolation between control points, they do not prevent tearings and foldings in the transformation. This means that the resulting transformation might not be one-to-one. Therefore, the transformation may not be realistic and thus suitable for medical applications. Hence, regularisation terms like the ones discussed in Section 2.1.5 or hard constraints [181] are additionally needed.

In this section, we described the most common and widely used types of parametric transformations. Other types of parametric transformations include, but are not limited to: discrete cosine transforms [14], wavelet-based transformations [232, 66, 32], Riesz basis of polynomial splines [151, 154, 155], piecewise affine transformations [88, 238, 165, 38, 41, 31] and poly-affine

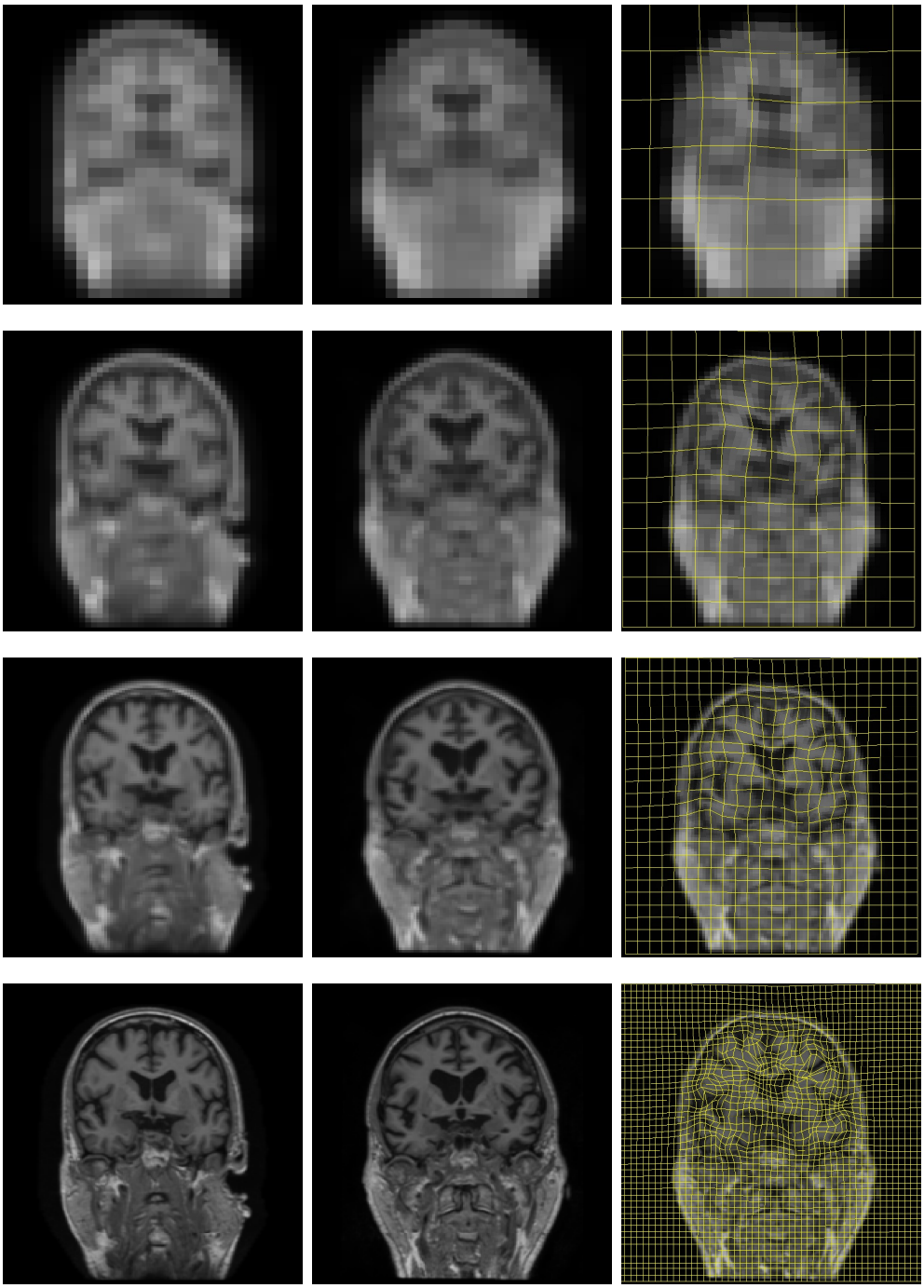


Figure 2.5: Result of non-rigid registration using multilevel B-spline FFDs. Left column: Target image. Centre column: Source image. Right column: Warped image with FFD mesh overlaid. The registration was masked to be performed only in the brain area.

transformations [10, 8].

2.1.4 Non-parametric transformations

Although parametric transformations have the great advantage of requiring a substantially lower number of degrees of freedom, this comes at the cost of having to evaluate the parametric transformation at every voxel of interest. This can be less or more expensive depending on the particular type of parametric transformation that is being used. In addition, parametric deformation models may not always be able to represent the desired transformation accurately. An alternative formulation is to utilise non-parametric non-rigid transformations, which are characterised by a dense set of displacements or velocities, one for every voxel in the target image. As a consequence, the transformation is defined only for the discrete voxel locations. To compute the transformation at locations between voxels, an interpolation scheme is required. A general differentiation can be made for non-parametric transformations, which can be either based on the small or large deformation models.

Small deformation model

In the small deformation framework, the transformation for every voxel is defined as the sum of the identity transformation $Id(\mathbf{x}) = \mathbf{x}$ and a displacement vector $\mathbf{u}(\mathbf{x})$

$$\mathbf{T}(\mathbf{x}) = Id(\mathbf{x}) + \mathbf{u}(\mathbf{x}). \tag{2.22}$$

It is clear that this model is of a highly localised nature, which renders the problem of finding the optimal displacement field an ill-posed problem. In order to constrain the space of possible displacement fields, a regularisation of the displacements is required. Details on a number of different regularisation schemes will be given in section 2.1.5.

An important limitation of this framework is that it is intrinsically limited in its ability to produce large diffeomorphic deformations, i.e., differentiable one-to-one mappings with a differentiable inverse. In this model, the combination of N transformations is performed by adding

the identity transformation and the sum of the individual displacements

$$\mathbf{T}(\mathbf{x}) = Id(\mathbf{x}) + \sum_{i=1}^{N} \mathbf{u}_i(\mathbf{x}). \tag{2.23}$$

However, even if all the individual transformations are diffeomorphic, the resulting combination may no longer be diffeomorphic. Another problem is the estimation of the inverse transformation. In the small deformation model, this estimation is performed by subtracting the deformation from the identity

$$\mathbf{T}^{-1}(\mathbf{x}) = Id(\mathbf{x}) - \mathbf{u}(\mathbf{x}). \tag{2.24}$$

However, this is in general a valid approximation only for very small deformations.

Large deformation model

The large deformation framework serves as an alternative in order to overcome the aforementioned limitations of the small deformation framework. In this framework, transformations are modelled as compositions of other transformations. The idea in the large deformation model is that the set of transformations should form a group under composition. Usually, this set is chosen to be the set of diffeomorphic transformations. The combination of N transformations is thus modelled as:

$$\mathbf{T}(\mathbf{x}) = \mathbf{T}_N \circ \cdots \circ \mathbf{T}_1 \circ \mathbf{x} = (Id + \mathbf{u}_N) \circ \cdots \circ (Id + \mathbf{u}_1) \circ \mathbf{x} = \mathbf{y}. \tag{2.25}$$

Similarly, the inverse transformation is obtained by simple backward composition:

$$\mathbf{T}^{-1}(\mathbf{y}) = \mathbf{T}_1^{-1} \circ \cdots \circ \mathbf{T}_N^{-1} \circ \mathbf{y} = \mathbf{x}. \tag{2.26}$$

One way to characterise the individual composed transformations is through integration of

time-varying velocity fields. When using such model, the transformation $\mathbf{T}(\mathbf{x}) = \mathbf{y}$ is modelled as a time-dependant flow $\mathbf{w}(\mathbf{x}, t)$, where $\mathbf{w}(\mathbf{x}, 0) = \mathbf{x}$ and $\mathbf{w}(\mathbf{x}, 1) = \mathbf{y}$. This flow is defined as the one satisfying the following ordinary differential equation (ODE):

$$\frac{\partial \mathbf{w}(\mathbf{x}, t)}{\partial t} = \mathbf{v}(\mathbf{w}(\mathbf{x}, t), t). \tag{2.27}$$

As mentioned before, the displacements are obtained by integrating over the time-varying velocity field $\mathbf{v}(\mathbf{w}(\mathbf{x}, t), t))$

$$\mathbf{T}(\mathbf{x}) = \int_0^1 \mathbf{v}(\mathbf{w}(\mathbf{x}, t), t) dt. \tag{2.28}$$

A common approach for computing this integral is to use Euler integration

$$\mathbf{w}(\mathbf{x}, t + \delta) = \mathbf{w}(\mathbf{x}, t) + \delta \mathbf{v}(\mathbf{w}(\mathbf{x}, t), t) \tag{2.29}$$

where δ is the time step used for the integration.

One of the main drawbacks of the using time-varying velocity fields is that they are very memory consuming and expensive to compute and evaluate. An alternative is to use *stationary* velocity fields, i.e., velocity fields that do not depend on time. In this case, the ODE becomes:

$$\frac{\partial \mathbf{w}(\mathbf{x}, t)}{\partial t} = \mathbf{v}(\mathbf{w}(\mathbf{x}, t)). \tag{2.30}$$

To model deformations using stationary velocity fields, Arsigny et al. [9] provide a mathematical framework based on Lie Group theory, where the exponential of a (smooth enough) vector field $\exp(\mathbf{V})$ is the flow at time 1 of the stationary ODE $\dot{x} = \mathbf{V}(x)$. Hence, the sought diffeomorphic deformation field $\mathbf{w}(1)$ can be expressed as

$$\mathbf{w}(1) = \exp(\mathbf{v}), \tag{2.31}$$

where $\mathbf{v}$ correspond to the stationary velocity field. A very attractive property of this formula-

tion is that the exponentiation can be very efficiently performed using the *scaling and squaring* algorithm [147, 9] as follows:

$$\begin{aligned} \mathbf{w}(1/2^N) &= Id + \mathbf{v}/2^N \\ \mathbf{w}(2/2^N) &= \mathbf{w}(1/2^N) \circ \mathbf{w}(1/2^N) \\ \mathbf{w}(4/2^N) &= \mathbf{w}(2/2^N) \circ \mathbf{w}(2/2^N) \\ &\vdots \\ \mathbf{w}(1) &= \mathbf{w}(2^{N-1}/2^N) \circ \mathbf{w}(2^{N-1}/2^N). \end{aligned} \tag{2.32}$$

Here, N is chosen so that $\mathbf{v}/2^N$ is close enough to zero, depending on the desired accuracy.

Another appealing property of the framework is that the inverse transformation can be computed in a similar fashion:

$$\begin{aligned} \mathbf{w}(-1/2^N) &= Id - \mathbf{v}/2^N \\ \mathbf{w}(-2/2^N) &= \mathbf{w}(-1/2^N) \circ \mathbf{w}(-1/2^N) \\ \mathbf{w}(-4/2^N) &= \mathbf{w}(-2/2^N) \circ \mathbf{w}(-2/2^N) \\ &\vdots \\ \mathbf{w}(-1) &= \mathbf{w}(-2^{N-1}/2^N) \circ \mathbf{w}(-2^{N-1}/2^N). \end{aligned} \tag{2.33}$$

Similarly to the small deformation framework, a regularisation approach is still needed when using the large deformation model, but this time the regularisation is performed on the velocity field $\mathbf{v}_t$ rather than the displacement field. One of such method is the fluid regularisation described in section 2.1.5.

2.1.5 Regularisation

Regularisation responds to the need of constraining the space of transformations to those that pose certain desired properties, depending on the particular application. In the case of rigid and affine transformations, explicit regularisation is often not performed, due to the very reduced

number of degrees of freedom of these models. However, when using non-rigid models (either parametric or non-parametric) the number of degrees of freedom is increased significantly, making regularisation a key element to ensure the well-behavedness of the registration problem.

Elastic regularisation

Elastic regularisation was first proposed by Broit [30] and then extended to a coarse-to-fine approach by Bajcsy and Kovačič [18]. This model tackles a linear elastostatics problem where the images are viewed as flexible sheets of elastic material such as rubber. In such elastic materials, any external force applied to them is counteracted by an internal force resisting deviations from the equilibrium. The magnitude and direction of the internal forces depend on the properties of the material. The images are deformed until an equilibrium between the external and internal forces is reached. If μ and λ are the elastic constants of the material, the equilibrium solution satisfies the Navier's displacement equation at each point:

$$\mu\nabla^2\mathbf{u} + (\lambda+\mu)\vec{\nabla}(\vec{\nabla}\cdot\mathbf{u}) + \mathbf{f} = 0. \tag{2.34}$$

Here, $\mathbf{u}$ is the displacement field and $\mathbf{f}$ corresponds to the external force used to drive the registration and is chosen to be the gradient of a similarity measure between the target and source images. The PDEs of equation (2.34) can be solved using finite differences. Alternatively, a Bayesian formulation derived from a variational equivalent of the linear elastostatics problem can be used [65].

A further extension to the elastic model was proposed by Davatzikos [45]. In this model, spatially-varying elastic constants of the material are incorporated to allow different anatomical structures to deform differently.

Fluid regularisation

Elastic regularisation penalises large deformations, since the magnitudes of the internal forces increase monotonically with the strain. Furthermore, the elastic model is derived using the small deformation framework described in section 2.1.4. These limitations motivated the use of fluid transformations to accommodate large deformation kinematics [36]. In this approach, images are modelled as a highly viscous fluid allowing to smoothly recover large and small (localised) deformations. Deformations are described using an Eulerian frame of reference, which specifies the time evolution of positions and velocities at fixed points. The equation governing the model is given by

$$\mu\nabla^2\mathbf{v} + (\lambda + \mu)\vec{\nabla}(\vec{\nabla} \cdot \mathbf{v}) + \mathbf{f} = 0 \tag{2.35}$$

where $\mathbf{f}$ corresponds to the force used to drive the registration (as in the elastic case). Note that equation (2.35) resembles equation (2.34), except that λ and μ represent coefficients of viscosity and that a time-varying velocity field is used, as described in section 2.1.4. In the approach of Christensen et al. [36], equation (2.35) is solved using successive overrelaxation (SOR) [202].

The bottleneck of the algorithm is solving equation (2.35). An alternative approach to solve the PDE using a convolution filter was proposed by Bro-Nielsen [29]. This algorithm is reported to be at least an order of magnitude faster, but requires the coefficients of viscosity to be constant over the whole volume, which is not always the case. Hence, when modelling fluids with spatially-varying viscosity (e.g. [129]), conventional numerical schemes like SOR are required.

Diffusion regularisation

Diffusion regularisation is one of the simplest regularisation models [4, 209]. It approximates the energy of a membrane subjected to stretching forces

$$\mathcal{R}_{\text{diffusion}} = \int_{\mathbb{R}^3} \left(\frac{\partial \mathbf{u}}{\partial x}\right)^2 + \left(\frac{\partial \mathbf{u}}{\partial y}\right)^2 + \left(\frac{\partial \mathbf{u}}{\partial z}\right)^2 d\mathbf{x}. \tag{2.36}$$

This regularisation energy is very simple to compute, since it is based solely on the first-order gradients of the displacement $\mathbf{u}$.

Curvature regularisation

Another commonly used regularisation model is curvature regularisation [225]. It is also commonly referred to as *bending energy* [219] or *thin-plate spline regulariser* [26], since it approximates the energy of a thin plate of metal which is subjected to bending deformations [43]. It is given by

$$\begin{aligned}\mathcal{R}_{\text{curvature}} = \int_{\mathbb{R}^3} &\left(\frac{\partial^2 \mathbf{u}}{\partial x^2}\right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial y^2}\right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial z^2}\right)^2 \\ &+ 2\left[\left(\frac{\partial^2 \mathbf{u}}{\partial xy}\right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial xz}\right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial yz}\right)^2\right] d\mathbf{x}.\end{aligned} \tag{2.37}$$

A useful property of curvature regularisation is that it uses only second-order derivatives of $\mathbf{u}$. Therefore, only penalises non-rigid transformations, since it is zero for any rigid or affine transformation.

Volume preserving regularisation

The observation that many tissues in the human body are approximately incompressible for small deformations and short time periods motivated the introduction of volume preserving

regularisation [177]. This regularisation depends on the determinant of the Jacobian matrix of the transformation:

$$J(\mathbf{x}) = \det \mathcal{J}_{\mathbf{T}}(\mathbf{x}) = \det \begin{pmatrix} \frac{\partial \mathbf{T}_x(\mathbf{x})}{\partial x} & \frac{\partial \mathbf{T}_x(\mathbf{x})}{\partial y} & \frac{\partial \mathbf{T}_x(\mathbf{x})}{\partial z} \\ \frac{\partial \mathbf{T}_y(\mathbf{x})}{\partial x} & \frac{\partial \mathbf{T}_y(\mathbf{x})}{\partial y} & \frac{\partial \mathbf{T}_y(\mathbf{x})}{\partial z} \\ \frac{\partial \mathbf{T}_z(\mathbf{x})}{\partial x} & \frac{\partial \mathbf{T}_z(\mathbf{x})}{\partial y} & \frac{\partial \mathbf{T}_z(\mathbf{x})}{\partial z} \end{pmatrix}, \tag{2.38}$$

where $\mathbf{T}(\mathbf{x}) = (u, v, w)$. According to the value of the determinant, it is possible to determine whether the deformation is locally an expansion or contraction:

$$J(\mathbf{x}) = \begin{cases} > 1 & \text{Volume expansion} \\ = 1 & \text{No volume change} \\ < 1 & \text{Volume compression} \end{cases} \tag{2.39}$$

If the value of the determinant is zero, the transformation is no longer a one-to-one diffeomorphic mapping. If the value of the determinant is negative, folding occurs and topology is no longer preserved. The volume preserving regularisation term is given by

$$\mathcal{R}_{\text{volume}} = \int_{\mathbb{R}^3} |\log(J(\mathbf{x}))| \, d\mathbf{x}. \tag{2.40}$$

As pointed out in [177], an alternative volume preserving term is:

$$\mathcal{R}_{\text{volume}} = \int_{\mathbb{R}^3} |J(\mathbf{x}) - 1| \, d\mathbf{x}. \tag{2.41}$$

However, this alternative does not pose the appealing property of the regularisation in Equation (2.40) of allowing to equally penalise compressions and expansions. To verify this property, one can observe that an expansion by a factor of s applied on an object with volume V yields the same regularisation penalty than a compression by the same factor on the same volume, since $|\log(sV)| = |\log(\frac{V}{s})|$.

Rigidity regularisation

Loeckx et al. [132] proposes to use a local rigidity constraint for regularisation. It is based on the observation that for a transformation to be locally rigid at a specific location (e.g. bony structures), the Jacobian matrix at that position must be orthogonal, i.e., the product of the Jacobian matrix with its transpose yields an identity matrix $\mathbf{I}_{3\times3}$. The regularisation constraint is thus formulated as:

$$\mathcal{R}_{\text{rigidity}} = \int_{\mathbb{R}^3} w(\mathbf{x}) \| \mathcal{J}_{\mathbf{T}}(\mathbf{x}) \mathcal{J}_{\mathbf{T}}(\mathbf{x})^{\top} - \mathbf{I}_{3\times3} \|_F \, d\mathbf{x}, \tag{2.42}$$

where $\| \cdot \|_F$ denotes the matrix Frobenius norm. The voxel-wise weighting function $w(\mathbf{x})$ allows different structures to exhibit different amounts of influence on the regularisation. The weighting can be obtained by simple thresholding or more advanced segmentation techniques like level-set segmentation or by performing simultaneous segmentation and registration. A similar rigidity regularisation approach with no spatially adaptive weighting was presented by Ruan et al. [180]. This approach uses the square of the Frobenius norm as penalty function. Finally, in the work by Staring et al. [201] a rigidity penalty term for regularisation is also proposed. This term is the sum of squared affinity, orthogonality and properness conditions.

2.2 Similarity measures

In image registration, a similarity (or dissimilarity) measure produces an estimation of how "similar" the target image and the transformed source image are. The goal of image registration is to find a transformation $\mathbf{T}$ chosen appropriately for the particular application that, when applied to the source image, maximises (or minimises) this similarity.

2.2.1 Point-based similarity measures

In order to be able to use point-based similarity measures, two sets of corresponding points (one for each image) have to be identified. These points may be defined by external markers or landmarks that are introduced to the images during acquisition, i.e., they are physically placed on the object being imaged. Another possibility is to use internal markers or landmarks that describe anatomical features within the images. These internal markers can be manually annotated or automatically detected.

Let $\mathbf{p}_i, \mathbf{q}_i,\ i \in \{1, 2, ..., N\}$ be two sets of N corresponding points located on the target and source images, respectively. Point-based similarities seek to minimise the mean weighted distance between corresponding points

$$S = \frac{1}{N} \sum_{i=1}^{N} w_i \, D(\mathbf{T}(\mathbf{p}_i), \mathbf{q}_i), \tag{2.43}$$

where w_i is a weighting term that represents the degree of confidence on which the corresponding points have been located. D is a distance function, usually $D(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\|^2$, although robust functions can be used as well [95, 33, 2].

2.2.2 Voxel-based similarity measures

One of the main drawbacks of point-based similarities is that internal or external landmarks are not always available. This is because the task of defining these landmarks can be very tedious and time-consuming. Furthermore, automatically annotated landmarks are usually not very reliable, since images commonly have ambiguous information and it is not easy to reliably and accurately detect them. In addition, landmarks are usually very sparsely scattered and do not provide dense spatial information. For these reasons, monomodal voxel-based similarities have been widely used in the context of medical image registration. The idea of these similarities is to use the intesities of every voxel in the target domain Ω to provide a measure of similarity

between the target image I_0 and transformed source image $I_1^{\mathbf{T}} = I \circ \mathbf{T}$.

Sum of squared differences (SSD)

One of the most widely used similarity measures [24, 212, 223] is the squared sum of intensity differences (SSD). For a target image and a source image I, this measure is defined as

$$\mathrm{SSD}(I_0, I_1^{\mathbf{T}}) = \frac{1}{\|\Omega\|} \sum_{\mathbf{x}\in\Omega} (I_0(\mathbf{x}) - I_1^{\mathbf{T}}(\mathbf{x}))^2. \tag{2.44}$$

This measure only includes voxels within the overlap of both images. Note that SSD is actually a measure of *dissimilarity* between the images.

Sum of absolute differences (SAD)

A more robust alternative to SSD is the sum of absolute differences (SAD) [94, 77]. This similarity measure, does not penalise large deviations in intensity as strongly as SSD does. Hence, it exhibits greater robustness against outliers. SAD is defined as

$$\mathrm{SAD}(I_0, I_1^{\mathbf{T}}) = \frac{1}{\|\Omega\|} \sum_{\mathbf{x}\in\Omega} |I_0(\mathbf{x}) - I_1^{\mathbf{T}}(\mathbf{x})|. \tag{2.45}$$

One of the main reasons why this similarity is not extensively used in medical image registration is because it is not a continuously differentiable function. This makes it unsuitable for gradient-based optimisation schemes.

Normalised cross-correlation (NCC)

Another measure commonly used in medical image registration is normalised cross correlation (NCC) [126, 186]. This measure also includes only the voxels that are within the overlap of

both images and is defined as

$$\mathrm{NCC}(I_0, I_1^{\mathbf{T}}) = \frac{\sum_{\mathbf{x}\in\Omega}(I_0(\mathbf{x}) - \bar{I}_0)(I_1^{\mathbf{T}}(\mathbf{x}) - \bar{I_1^{\mathbf{T}}})}{\sqrt{\sum_{\mathbf{x}\in\Omega}(I_0(\mathbf{x}) - \bar{I}_0)^2}\sqrt{\sum_{\mathbf{x}\in\Omega}(I_1^{\mathbf{T}}(\mathbf{x}) - \bar{I_1^{\mathbf{T}}})^2}}, \tag{2.46}$$

where $\bar{I}_0$ and $\bar{I_1^{\mathbf{T}}}$ denote the average intensity of the target and transformed source image, respectively. This measure has values in the $[0, 1]$ range and is invariant to linear intensity scalings.

2.2.3 Multimodal similarities

All the similarity measures described so far are mono-modality measures that assume that the intensities in the target and source images are linearly related. The main drawback of the monomodal voxel-based similarities is that they cannot handle multimodal images, i.e., images taken from different image aquisition systems, e.g. MR, CT and PET. This issue motivated the use of similarity measures that are robust to complex intensity changes, such as local normalised cross-correlation and entropy-based similarity measures derived from information theory.

Local normalised cross-correlation (LNCC)

One such multi-modality similarity measure is local normalised cross-correlation (LNCC) and has been proposed in the work by Avants et al. [16]. In addition to being suitable for multimodal registration, this measure is also robust against slowly varying local intensity inhomogeneities. LNCC is defined as

$$\mathrm{LNCC}(I_0, I_1^{\mathbf{T}}) = \frac{1}{\|\Omega\|}\sum_{\mathbf{x}\in\Omega}\mathrm{NCC}(\mathcal{N}_{I_0}(\mathbf{x}), \mathcal{N}_{I_1^{\mathbf{T}}}(\mathbf{x})), \tag{2.47}$$

where $\mathcal{N}_I(\mathbf{x})$ correspond to a cubic patch from image I and centered at position $\mathbf{x}$.

An alternative way of computing the LNCC measure was proposed by Lorenzi et al. [134]. In

this formulation, the LNCC is given by:

$$\mathrm{LNCC}(I_0, I_1^{\mathbf{T}}) = \sum_{\mathbf{x}\in\Omega} \frac{\overline{I_0(\mathbf{x}) \cdot I_1^{\mathbf{T}}(\mathbf{x})}}{\sqrt{\overline{(I_0(\mathbf{x}))^2} \cdot \overline{(I_1^{\mathbf{T}}(\mathbf{x}))^2}}}, \tag{2.48}$$

where $\overline{I(\mathbf{x})} = \mathbf{G}_\sigma * I(\mathbf{x})$ corresponds to the local mean image defined by convolution with a Gaussian kernel of size σ.

Joint entropy (JE)

Another multimodal similarity measure is the joint entropy of images I_0 and $I_1^{\mathbf{T}}$ [204]. It measures the common or shared information between both images. Maximising this common information aims at maximising the alignment. In order to be able to compute the value of an entropy-based similarity measure, a 2D joint histogram $\mathcal{H}$ has to be constructed. This joint histogram contains the occurence of each pair of intensities at the same locations (intensities from $I_0(\mathbf{x})$ and from $I_1^{\mathbf{T}}(\mathbf{x})$) and is normalised by the total number of occurrences to contain pair-distribution probabilities. The joint histogram can be filled in a naïve way by adding each occurence as a 1 and then normalise. Alternatively, one can use a partial volume approach, where a weight between 0 and 1 is added to the joint histogram according to the transformation [136]. A further possibility is to perform a Parzen window smoothing of the normalised naïvely filled joint histogram [141]. Figure 2.6 shows an example of two smoothed joint histograms.

Let N^{I_0} and $N^{I_1^{\mathbf{T}}}$ be the user-defined number of histogram bins for I_0 and $I_1^{\mathbf{T}}$, respectively. Then, the joint entropy can be computed using the Shannon formula for entropy [191]:

$$H(I_0, I_1^{\mathbf{T}}) = -\sum_{t=1}^{N^{I_0}} \sum_{s=1}^{N^{I_1^{\mathbf{T}}}} P(t,s) \log P(t,s), \tag{2.49}$$

where $P(t,s) = \mathcal{H}(t,s)$ is the probability of an intensity in I_0 coming from histogram row t being aligned with an intensity in $I_1^{\mathbf{T}}$ coming from histogram column s.

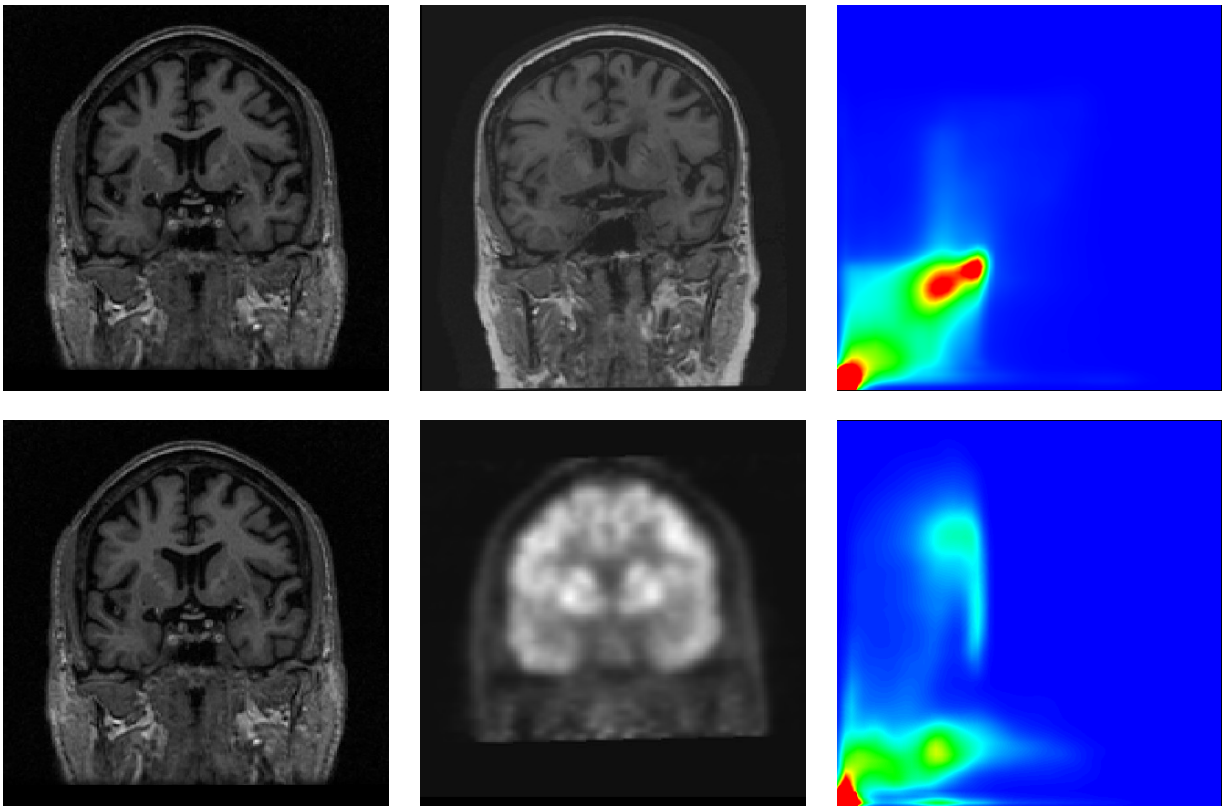


Figure 2.6: Joint histograms in monomodal and multimodal cases. A Parzen window approach [141] has been used to smooth the histograms. Top row: Monomodal MR-MR. Bottom row: Multimodal MR-PET

Mutual information (MI)

The joint entropy measure can yield values which are close to the optimum, even when a very poor alignment between the images is observed. To overcome this issue Maes et al. [136] and Viola and Wells [224] propose the use of mutual information (MI) instead. Mutual information takes into account the joint entropy, but also the marginal entropies for each image. These marginal entropies are given by

$$H(I_0) = -\sum_{t=1}^{N^{I_0}} P(t) \log P(t) \tag{2.50}$$

$$H(I_1^{\mathbf{T}}) = -\sum_{s=1}^{N^{I_1^{\mathbf{T}}}} P(s) \log P(s) \tag{2.51}$$

where $P(t) = \sum_{s=1}^{N^{I_1^{\mathbf{T}}}} P(t,s)$ and $P(s) = \sum_{t=1}^{N^{I_0}} P(t,s)$. Mutual information is then defined as

$$\begin{aligned} \mathrm{MI}(I_0, I_1^{\mathbf{T}}) &= H(I_1^{\mathbf{T}}) + H(I_0) - H(I_0, I_1^{\mathbf{T}}) \\ &= \sum_{t=1}^{N^{I_0}} \sum_{s=1}^{N^{I_1^{\mathbf{T}}}} P(t,s) \log \frac{P(t,s)}{P(t)P(s)} \end{aligned} \tag{2.52}$$

which is the Kullback-Leiber divergence between the joint probablility distribution and the product of the marginal probability distributions.

Normalised mutual information (NMI)

In addition to MI, Studholme et al. [205] proposed the use of normalised mutual information (NMI) as a similarity measure. It has been shown to be more robust than MI with respect to variations in image overlap. NMI is defined as:

$$\mathrm{NMI}(I_0, I_1^{\mathbf{T}}) = \frac{H(I_0) + H(I_1^{\mathbf{T}})}{H(I_0, I_1^{\mathbf{T}})} \tag{2.53}$$

2.2.4 Interpolation

Since a transformation $\mathbf{T}$ estimated via image registration may map a discrete grid location $\mathbf{x}$ in the target image into a non-grid location $\mathbf{T}(\mathbf{x})$ in the source image, an interpolation in the source image must be performed in order to obtain the intensity that will be associated to the target point $\mathbf{x}$ in the warped image to compute the similarity in that point. The most commonly used interpolation methods are:

- *Nearest neighbour:* In this method, the intensity of the pixel in the source image that is nearest to $\mathbf{T}(\mathbf{x})$ is associated with the target point $\mathbf{x}$.
- *Linear:* Here, the intensity associated to the target point $\mathbf{x}$ is obtained by interpolating the intensity values linearly along each axis using a neighbourhood of 8 voxels around $\mathbf{T}(\mathbf{x})$.

(a) Nearest neighbour (b) Linear

Figure 2.7: (a) Nearest neighbour and (b) linear interpolation. Note the sharp nature of the nearest neighbour result against the smoothness provided by linear interpolation.

- *Cubic:* The linear approach is used, but instead of performing linear interpolation along each axis, a cubic polinomial kernel is applied.

Figure 2.7 shows the result of a nearest neighbour and linear interpolation in a brain MR image. For an assessment of these and other interpolation methods see the work of Park et al. [161].

2.3 Optimisation strategies

As mentioned earlier, the aim of image registration is to find the optimal transformation $\mathbf{T}$ that brings the target and source images into alignment. More specifically, the optimal transformation is the one that minimises an energy or objective functional $E(I_0; I; \mathbf{T})$. The most common formulation of the optimisation problem in the context of image registration is

$$\arg\min_{\mathbf{T}} E(I_0; I; \mathbf{T}) = -E_{sim}(I_0; I \circ \mathbf{T}) + \alpha E_{reg}(\mathbf{T}). \tag{2.54}$$

Here, E_{sim} is a similarity measure (e.g. the ones described in section 2.2), E_{reg} is a regularisation term as described in section 2.1.5, and $\alpha \geq 0$ is a tradeoff parameter between the two terms. Note, however, that when using elastic or fluid models, the regularisation is implicit in $\mathbf{T}$, therefore E_{reg} is usually not explicitly utilised in those cases. The optimisation method can treat the parameter (or displacement) space of the transformation as either a continuous or a discrete space. In the following, we describe the most common optimisation approaches for both models.

2.3.1 Continuous optimisation

In the context of image registration, the most widely used optimisation approaches are gradient-based iterative methods. These methods use a general update rule to iteratively recompute the set of parameters or voxel-wise displacements Φ. Upon convergence, the resulting estimate minimises the objective function $E(I_0; I; \mathbf{T})$, either locally or globally, depending on the initial estimate of Φ. The update rule of gradient-based optimisation methods is given by:

$$\Phi_{t+1} = \Phi_t + \eta_t \cdot \mathbf{g}_t(\Phi_t). \tag{2.55}$$

Here, t corresponds to the iteration during the optimisation, $\eta_t > 0$ is the step size and $\mathbf{g}_t(\Phi_t)$ defines the search direction. The parameter η_t can be constant, decrease with each iteration, or be chosen such as it minimises the objective function along $\mathbf{g}_t(\Phi_t)$. In the following, we describe the most commonly used gradient-based optimisation methods utilised for image registration. A distinctive feature of different approaches is the choice of the search direction $\mathbf{g}_t(\Phi_t)$.

Gradient descent

Gradient descent optimises the objective function by following the direction that decreases the energy (its negative gradient). This means the search direction is given by:

$$\mathbf{g}_t(\Phi_t) = -\frac{\partial E}{\partial \Phi_t}. \tag{2.56}$$

Image registration methods that employ this type of optimisation include [24, 49, 105, 183].

Conjugate gradient descent

Conjugate gradient descent aims to improve the convergence rate of standard gradient descent methods. This is achieved by employing information from previous iterations. More specifically, by using a search direction that does not follow the negative gradient, but is conjugate to the previous estimation of $\mathbf{g}_t$. To this end, the conjugate search direction is defined as

$$\mathbf{g}_t(\Phi_t) = -\frac{\partial E}{\partial \Phi_t} + \mu_t \mathbf{g}_{t-1}(\Phi_{t-1}). \tag{2.57}$$

Among the different methods to compute μ_t, one may mention the Polak-Ribière-Polyak formula [167, 168], the Hestenes-Stiefel formula [92], and the Fletcher-Reeves formula [57]. Among image registration methods that rely on conjugate gradients for the optimisation one can cite [169, 67, 144, 107, 216].

Gauss-Newton method

The Gauss-Newton method is specifically tailored for least-squares fitting optimisation problems. This type of optimisation problems are usually seen in registration approaches that optimise SSD as a similarity measure [220, 221, 223, 15]. Contrarily to Newton's or Newton-Raphson methods where the Hessian matrix $\frac{\partial^2 E}{\partial \Phi_t^2}$ is utilised, this method uses an approximation

of the Hessian that ignores derivatives higher than first order

$$\frac{\partial^2 E}{\partial \Phi_t^2} \approx 2J(\Phi_t)^\top J(\Phi_t). \tag{2.58}$$

Here $J(\Phi_t) = \frac{\partial E}{\partial \Phi_t}$. Therefore, the search direction for this method is formulated as

$$\mathbf{g}_t(\Phi_t) = -\left(J(\Phi_t)^\top J(\Phi_t)\right)^{-1} \frac{\partial E}{\partial \Phi_t}. \tag{2.59}$$

Levenberg-Marquardt algorithm

The Levenberg-Marquardt algorithm, also known as *damped least-squares* method is another opimisation approach, which is related to the previously described Gauss-Newton method. For this approach, the search direction is given by

$$\mathbf{g}_t(\Phi_t) = -\left(J(\Phi_t)^\top J(\Phi_t) + \zeta \mathbf{I}\right)^{-1} \frac{\partial E}{\partial \Phi_t}, \tag{2.60}$$

where $\mathbf{I}$ is the identity matrix and ζ is a damping parameter that balances between speed of convergence and stability of the algorithm. Increasing the value of ζ results in a slow down in the speed of convergence, but increases stability. Note also that setting this parameter to zero, results in the Gauss-Newton algorithm. Image registration approaches using this optimisation method can be found in [97, 232, 66, 13].

2.3.2 Discrete optimisation

Starting from the seminal work of Glocker et al. [69, 68, 70], medical image registration techniques using discrete optimisation have gained increased attention in the research community [192, 218, 244, 198, 37, 196, 34, 119, 84, 85, 185]. These techniques are usually formulated using Markov Random Fields (MRF) [112]. In the context of image registration, the energy of an MRF with unary potentials f_p representing the data term and pairwise potentials g_{pq}

representing the regularisation term is defined as

$$E_{\text{MRF}}(\mathbf{l}) = \sum_{p \in V} f_p(l_p) + \sum_{(p,q) \in E} g_{pq}(l_p, l_q). \tag{2.61}$$

Here, V and E are the vertices and edges of an undirected graph $G = (V, E)$ and each random variable $l_p \in \mathbf{l}$ takes its value from a discrete label set L corresponding to a quantised version of the deformation space.

The optimisation of the energy in equation (2.61) can be performed using graph-based methods following the max-flow min-cut principle [59] like the α-expansion algorithm of Boykov et al. [28]. Another approach for discrete MRF optimisation is to utilise Belief Propagation [162] or Loopy Belief Propagation in cases of graphs with loops [62]. Komodakis et al. [115, 116] proposed a linear programming method for MRF optimisation. Other MRF optimisation methods include Fusion Moves [127, 128], Quadratic Pseudo Boolean Optimisation (QPBO) [80] and Higher-order Clique Reduction (HOCR) [101, 102]. Note that none of these optimisation methods require the computation of the derivative of the registration cost function. This is the main advantage of MRF based optimisation methods, since this opens the possibility of using more complex similarity measures that do not need to be differentiable.

2.4 Summary

In this chapter, a description of the components of a generic registration algorithm has been provided. These components are: a transformation model, a similarity measure and an optimisation. In the case of the transformation model, we have introduced both parametric and non-parametric formulations. We have also discussed the main types of similarity measures that can be used to align the images. Finally, we have provided details on the main continuous and discrete optimisation methods for medical image registration.

Different registration approaches use different combinations of transformation model, similarity

measure and optimisation method. The particular choice of which approach to use is rather application specific, since all of them have advantages and disadvantages and perform better in certain contexts than in others. In the case of transformation models, parametric formulations allow to perform fast registration of volumetric images, but at the cost of potentially breaking the topology of the warped images if the deformation to be recovered is large. On the other hand, transformation models based on time-varying or stationary velocity fields can ensure that the mapping is one-to-one and invertible, but at the cost of increased computational time. This makes them unsuitable for real-time or near real-time requirements. Monomodal similarity measures like SSD yield fast registration and are very simple to optimise, but cannot be utilised for registration between images of different modalities such as MR, CT and PET. Finally, continuous optimisation methods usually provide very accurate results, but requires the computation of derivatives of the similarity measure and regularisation term, which eliminates the possibility of using non-differentiable measures. On the other hand, discrete optimisation methods allows to employ any similarity measure, but normally provide lower accuracy than continuous optimisation approaches.

In the next chapter, we focus on a number of important registration methods and new paradigms of this research area.

Chapter 3

Intensity-based non-rigid registration of medical images: Methods

Due to its importance in medical imaging applications, and also for other fields of computer vision, there has been a large amount of research in the area of image alignment and registration. The deformable nature of many of the objects in medical images is the reason for the fact that the vast majority of the registration techniques in this area are intensity-based non-rigid registration techniques. In this chapter, we describe some of the foundational intensity-based non-rigid registration approaches that enabled the development of the field of image registration to what is today. Furthermore, we describe some state-of-the-art registration methods that comply with important properties, such as inverse-consistency or symmetry of the transformations. We also describe registration methods that use attributes as a high dimensional alternative to pure intensity data. Finally, since most of the contributions of this thesis are learning-based approaches, we also describe prior work on this field. Further information on image registration techniques in general can be found on the survey by Zitová et al. [246]. More recently, another survey focused specifically on medical image registration has been presented by Sotiras et al. [197].

3.1 Optical flow estimation

Optical flow is the distribution of displacements in which intensities move between two consecutive frames of an image sequence. The first approach to estimate optical flow was developed by Horn and Schunck [96]. The basic premise of optical flow is the assumption that object intensities $I(\mathbf{x})$ of a point $\mathbf{x} = (x, y, z)$ do not change between frames. In the context of image registration, it means that patterns or structures in one image "move" to form the pattern or structure in the next one. This constant intensity assumption can be formulated as:

$$I_0(x, y, z, t) = I(x + \Delta x, y + \Delta y, z + \Delta z, t + \Delta t), \tag{3.1}$$

where $I_0(x, y, z, t)$ corresponds to the intensity of point (x, y, z) in the target image (at time t), and $I(x + \Delta x, y + \Delta y, z + \Delta z, t + \Delta t)$ corresponds to the intensity of the same point in the source image, after being displaced by $\Delta\mathbf{x} = (\Delta x, \Delta y, \Delta z)$ over Δt unit of time. By using a Taylor expansion of equation (3.1) and ignoring high order terms, it is possible to derive the equation for optical flow (see [96] for details):

$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial z}\frac{\Delta z}{\Delta t} + \frac{\partial I}{\partial t} = 0. \tag{3.2}$$

A more concise representation of equation (3.2) is:

$$\frac{\partial I}{\partial t} + \nabla I \cdot \mathbf{u} = 0. \tag{3.3}$$

Here, $\frac{\partial I}{\partial t}$ is the temporal difference between frames, ∇I is the spatial gradient of the target image, and $\mathbf{u}$ corresponds to the displacement field describing optical flow. Note that equation (3.3) has 3 unknowns, but only one equation is available. This is known as the *aperture problem* of optical flow and can be resolved by adding aditional smoothness constraints on $\mathbf{u}$ [96, 21].

3.2 Demons algorithm

The diffusion transformation model of Thirion [211] is based on on the concept of *demons* introduced by Maxwell to explain the Gibbs paradox of thermodynamics. The main idea is to model the source image as a deformable grid, where the vertices of the grid are particles that can be classified as 'outside' or 'inside'. This grid is allowed to diffuse through the contour of an object in the target image on which several demons are scattered. These demons act as effectors that, using forces inspired from the optical flow equations [96, 21], locally push the grid inside the contour if the corresponding point is labelled 'inside', and outside the contour if it is labelled 'outside'. There are various possibilities on how the demons can be placed. In [211] three alternatives are proposed: (1) placing one demon on each voxel; (2) placing demons on voxels labeled as edge by an edge detector (such as [47, 148]); (3) placing demons along label boundaries of already segmented images.

Pennec et al. [164] shows that the demons formulation can be thought as an approximation of a second order gradient descent using the sum of squared differences as a similarity measure. Vercauteren et al. [223] reformulated the demons model by adapting it to the space of diffeomorphic transformations.

3.3 Large Deformation Diffeomorphic Metric Mapping (LDDMM)

The Large Deformation Diffeomorphic Metric Mapping (LDDMM) is a registration approach proposed by Beg et al. [24]. The transformations in this model are guaranteed to be diffeomorphic. In this setting, the target image I_0 and source image I are registered by generating a diffeomorphic deformation $\mathbf{u}$ using the large-deformation framework of section 2.1.4. The

optimal time-dependant velocity field $\hat{\mathbf{v}}$ is found by solving a variational problem of the form:

$$\hat{\mathbf{v}} = \underset{\mathbf{v}:\dot{u}_t=\mathbf{v}_t(\mathbf{u}_t)}{\arg\max} \int_0^1 \|\mathbf{v}_t\|_V^2 dt + \frac{1}{\sigma^2}\|I_0 \circ \mathbf{u}^{-1} - I\|_{L_2}^2. \tag{3.4}$$

Here, $\mathbf{v}_t$ is the velocity field at time t, $\|\mathbf{v}_t\|_V$ is an appropiate Sobolev norm on the velocity field, and $\|\cdot\|_{L_2}$ corresponds to the L_2 error norm that enforces matching of the images. To ensure that the solution to the underlying ODE is in the space of diffeomorphisms, a differential operator $\mathbf{L}$ is used for regularisation, such that $\|\mathbf{v}_t\|_V = \|\mathbf{L}\mathbf{v}_t\|_{L_2}$. An important property of this framework is that the length of the shortest path (also known as *geodesic path*) which connects the target and source images

$$\inf \int_0^1 \|\mathbf{v}_t\|_V dt \tag{3.5}$$

defines a metric, i.e, gives a notion of distance between the two images.

3.4 DARTEL algorithm

The DARTEL algorithm, which stands for "Diffeomorphic Anatomical Registration using Exponentiated Lie algebra", has been proposed by Ashburner [13]. The algorithm utilises the previously described stationary velocity field framework of Arsigny et al. [9] to generate the deformation, i.e. by means of the scaling and squaring algorithm (see section 2.1.4). The objective function is modelled as the most probable stationary velocity field $\mathbf{v}_i$ given the image data D. This can be formulated using Bayes' theorem:

$$P(\mathbf{v}_i|D) = \frac{P(D|\mathbf{v}_i)P(\mathbf{v}_i)}{P(D)}. \tag{3.6}$$

The term $P(D)$ is ignored since it is constant and the objective is to find the most probable velocity field vectors $\mathbf{v}_i$, not the probability distribution itself. The final objective function is then formulated by taking advantage of the monotonic nature of the logarithm function

(log-likelihoods):

$$-\log P(\mathbf{v}_i|D) = -\log P(\mathbf{v}_i) - \log P(D|\mathbf{v}_i). \tag{3.7}$$

This objective function can be interpreted as a general energy formulation:

$$E(\mathbf{v}_i) = E_R(\mathbf{v}_i) + E_D(\mathbf{v}_i). \tag{3.8}$$

Here, E_D is the similarity (or data) term and E_R corresponds to the regularisation term. In DARTEL, the data term is the sum of squared intensity differences between the target and source images and several alternatives are provided for the regularisation term (membrane energy, bending energy and linear elastic energy). The author also argues that solving the optimisation problem using the maximum a posteriori estimate poses some technical difficulties. Hence, the optimisation strategy of this approach is performed using the Levenberg-Marquardt algorithm (see section 2.3.1).

The work of Hernandez et al. [90, 91], is based on an essentially similar concept. However, the authors derive the Euler-Lagrange equations asociated with the minimisation of a different energy in an inverse consistent variational formulation. Another difference is that integration is performed using a semi-Lagrangian algorithm [200].

3.5 Inverse-consistent or symmetric approaches

Christensen and Johnson [35] present an inverse consistent approach based on the small deformation framework. In this approach, the forward transformation $\mathbf{T}_{\text{forw}}$ and the backward transformation $\mathbf{T}_{\text{back}}$ between the target image I_0 and the source image I are optimised simultaneously. In order to achieve inverse consistency, the authors propose to optimise an energy function with three terms: The first term correspond to the data term which enforces that both the forward and backward transformations warp the corresponding images is such way that the

SSD between them is minimised in both cases. This term is given by:

$$E_{\text{data}} = \sum_{\mathbf{x}\in\Omega}(I_0(\mathbf{x}) - I(\mathbf{T}_{\text{forw}}(\mathbf{x})))^2 + \sum_{\mathbf{x}\in\Omega}(I_0(\mathbf{T}_{\text{back}}(\mathbf{x})) - I(\mathbf{x}))^2. \tag{3.9}$$

The second term enforces the forward and backward transformations to be the inverse of each other:

$$E_{\text{IC}} = \sum_{\mathbf{x}\in\Omega} \|\mathbf{T}_{\text{forw}}(\mathbf{x}) - \mathbf{T}^{-1}_{\text{back}}(\mathbf{x})\|^2 + \sum_{\mathbf{x}\in\Omega} \|\mathbf{T}_{\text{back}}(\mathbf{x}) - \mathbf{T}^{-1}_{\text{forw}}(\mathbf{x})\|^2. \tag{3.10}$$

Here, the inverse transformations are computed using a discrete approximation method that assumes that the transformations have a positive Jacobian over the entire space Ω. The final term is a linear elastic regularisation constraint that enforces the transformations to be diffeomorphisms:

$$E_{\text{reg}} = \sum_{\mathbf{x}\in\Omega} \|\mathbf{L}\mathbf{u}_{\text{forw}}(\mathbf{x})\|^2 + \sum_{\mathbf{x}\in\Omega} \|\mathbf{L}\mathbf{u}_{\text{back}}(\mathbf{x})\|^2. \tag{3.11}$$

Here, $\mathbf{L}$ is a differential linear elasticity operator acting on the displacement fields $\mathbf{u}_{\text{forw}}$ and $\mathbf{u}_{\text{back}}$. A fourth energy term, which can be used either to replace the regularisation term or in addition of the three previous terms, is also proposed. It prevents both forward and backward transformations to produce extreme local compression (Jacobian values close to zero) or extreme local expansion (very large Jacobian values):

$$E_{\text{jac}} = \sum_{\mathbf{x}\in\Omega} \mathcal{J}(\mathbf{T}_{\text{forw}}(\mathbf{x}))^{-2} + \mathcal{J}(\mathbf{T}_{\text{back}}(\mathbf{x}))^{-2} + \mathcal{J}(\mathbf{T}_{\text{forw}}(\mathbf{x}))^2 + \mathcal{J}(\mathbf{T}_{\text{back}}(\mathbf{x}))^2. \tag{3.12}$$

Here, $\mathcal{J}(\mathbf{T}(\mathbf{x}))$ corresponds to the local Jacobian value of transformation $\mathbf{T}$ at point $\mathbf{x}$.

The SyN registration method is presented in the work by Avants et al. [16]. This approach is based on a previous formulation using time-varying velocity fields [17]. This formulation provides symmetric diffeomorphisms, i.e., diffeomorphisms where the geodesic path connecting the target image and the source image is the same as the geodesic path connecting the source and target. Hence, the integration is performed "half way" (up to $t = 0.5$) of this geodesic path

using a greedy algorithm. The energy formulation for this approach is as follows:

$$E_{\text{sym}} = \inf_{\mathbf{u}_{\text{forw}}} \inf_{\mathbf{u}_{\text{back}}} \int_0^{0.5} \left(\|\mathbf{v}_{\text{forw}}(\mathbf{x},t)\|_L^2 + \|\mathbf{v}_{\text{back}}(\mathbf{x},t)\|_L^2 \right) dt + \int_\Omega LNCC(I_0(\mathbf{u}_{\text{forw}}^{-1}(\mathbf{x},0.5)), I(\mathbf{u}_{\text{back}}^{-1}(\mathbf{x},0.5))). \quad (3.13)$$

Here, $\|\cdot\|_L$ induces regularity on the velocity field via a linear differential operator and $LNCC$ is the local normalised cross correlation similarity measure described in section 2.2.3. An important property is that, unlike the inverse consistent formulation of [35], in this framework the deformation field inversions are guaranteed to have sub-voxel accuracy. The final deformations are then given by:

$$\begin{aligned} \mathbf{u}_{\text{forw}}(1) &= \mathbf{u}_{\text{back}}^{-1}(0.5) \circ \mathbf{u}_{\text{forw}}(0.5) \\ \mathbf{u}_{\text{back}}(1) &= \mathbf{u}_{\text{forw}}^{-1}(0.5) \circ \mathbf{u}_{\text{back}}(0.5). \end{aligned} \quad (3.14)$$

Figure 3.1 provides an overview of the SyN registration framework.

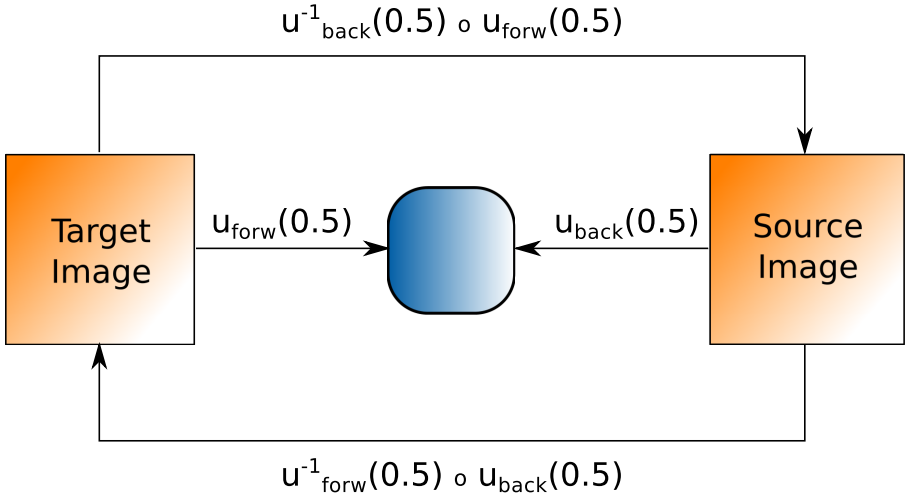


Figure 3.1: SyN registration framework.

A symmetric log-domain diffeomorphic registration approach is proposed in [222]. In this work the authors extend their diffeomorphic Demons approach of [221], where the update step $\mathbf{s}$ is a velocity field that is mapped into the space of diffeomorphisms via the exponential map. Hence, the update step is $\mathbf{u} \leftarrow \mathbf{u} \circ \exp(\mathbf{s})$. In the proposed log-domain approach, the deformation field $\mathbf{u}$ is also the result of the exponential mapping of a velocity field, yielding an update step of

the form $\mathbf{u} = \exp(\mathbf{v}) \leftarrow \exp(\mathbf{v}) \circ \exp(\mathbf{s})$. To find the smooth velocity field $Z(\mathbf{v}, \mathbf{s})$ such that $\exp(Z(\mathbf{v}, \mathbf{s})) \approx \exp(\mathbf{v}) \circ \exp(\mathbf{s})$ the authors use the Baker-Campbell-Hausdorff approximation:

$$Z(\mathbf{v}, \mathbf{s}) \approx \mathbf{v} + \mathbf{s} + \frac{1}{2}[\mathbf{v}, \mathbf{s}] + \frac{1}{12}[\mathbf{v}, [\mathbf{v}, \mathbf{s}]]. \tag{3.15}$$

Here, $[\cdot, \cdot]$ corresponds to the Lie bracket operator that, given two velocity fields, yields a third velocity field defined for each point $\mathbf{x}$ such that:

$$[\mathbf{v}, \mathbf{s}](\mathbf{x}) = \mathcal{J}_{\mathbf{v}}(\mathbf{x}) \cdot \mathbf{v}(\mathbf{x}) - \mathcal{J}_{\mathbf{s}}(\mathbf{x}) \cdot \mathbf{s}(\mathbf{x}). \tag{3.16}$$

Here, $\mathcal{J}_{\mathbf{v}}$ is the Jacobian matrix of the field $\mathbf{v}$. The advantage of using a log-domain approach is that the inverse of a transformation can be easily computed as $\mathbf{u}^{-1} = \exp(-\mathbf{v})$, as discussed in Section 2.1.4. Taking this into account, the authors also propose a symmetric extension that optimises the sums of squared differences obtained by forward and backward (inverse) transformation:

$$E_{\text{symSSD}} = \text{SSD}(I_0, I, \mathbf{u}) + \text{SSD}(I, I_0, \mathbf{u}^{-1}). \tag{3.17}$$

In each optimisation step, the forward and backward update steps in log-domain are independently obtained from the respective demon forces and then (thanks to the log-domain representation) projected onto the space of symmetric transformations by simply averaging them. This yields the log-domain update step that is finally used to update the transformation.

A symmetric approach using free-form deformations is presented in the work by Modat et al. [145]. For inverse consistency, the authors optimise both the forward and backward transformations simultaneously using an inverse consistency penalty which is similar to the one used in [35]. It is based on the norm of the *composition* of the forward and backward transformations:

$$E_{\text{IC}} = \sum_{\mathbf{x} \in \Omega} \|\mathbf{T}_{\text{forw}}(\mathbf{T}_{\text{back}}(\mathbf{x}))\|^2 + \sum_{\mathbf{x} \in \Omega} \|\mathbf{T}_{\text{back}}(\mathbf{T}_{\text{forw}}(\mathbf{x}))\|^2. \tag{3.18}$$

As a similarity measure, the authors choose to use a symmetric approach similar to the one in

[222], but using normalised mutual information instead:

$$E_{\text{symNMI}} = \sum_{\mathbf{x}\in\Omega} \text{NMI}(I_0, I \circ \mathbf{T}_{\text{forw}}) + \text{NMI}(I, I_0 \circ \mathbf{T}_{\text{back}}). \tag{3.19}$$

Further curvature and volume-preserving regularisation penalties complete the proposed energy model.

The LCC-Demons registration framework has been proposed by Lorenzi et al. [134]. It extends the log-Demons approach of [222] by replacing the SSD similarity measure by a symmetric version of the LNCC similarity of section 2.2.3 which is robust to additive and multiplicative intensity inhomogeneities:

$$E_{\text{symLNCC}} = \sum_{\mathbf{x}\in\Omega} \frac{\overline{(I_0 \circ \exp(-\frac{\mathbf{v}}{2}))(\mathbf{x})(I \circ \exp(\frac{\mathbf{v}}{2}))(\mathbf{x})}}{\sqrt{\overline{\left[(I_0 \circ \exp(-\frac{\mathbf{v}}{2}))(\mathbf{x})\right]^2}\ \overline{\left[(I \circ \exp(\frac{\mathbf{v}}{2}))(\mathbf{x})\right]^2}}}. \tag{3.20}$$

Here, $\overline{I} = \mathcal{G}_\sigma \star I$ is the local mean image of image I defined by Gaussian convolution $\mathcal{G}_\sigma$ with kernel size σ.

3.6 Attribute-matching registration approaches

The HAMMER algorithm [193], which stands for "Hierarchical Attribute Matching Mechanism for Elastic Registration" is an attribute-matching registration method for MR images of the brain. In this method, attribute vectors $\mathbf{a}$ are defined as the concatenation of two values $\mathbf{a}_1$, $\mathbf{a}_2$ and a vector $\mathbf{a}_3$:

$$\mathbf{a}(\mathbf{x}) = (\mathbf{a}_1(\mathbf{x}), \mathbf{a}_2(\mathbf{x}), \mathbf{a}_3^1(\mathbf{x}), \cdots, \mathbf{a}_3^K(\mathbf{x}))^\top. \tag{3.21}$$

Here, $\mathbf{a}_1(\mathbf{x})$ defines the edge-type of voxel $\mathbf{x}$. It takes one of seven possible values: no edge or any of the six combinations of boundaries between grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF). $\mathbf{a}_2(\mathbf{x})$ corresponds to the image intensity value normalised between 0 and 1. Finally, $\mathbf{a}_3(\mathbf{x})$ is a vector of $K = 13 \times 3 \times L$ values representing 13 rotation-invariant

moments computed from the zeroth-, second- and third-order 3-D regular moments [130], for each GM, WM and CSF tissue label volume in L different scales (spherical neighbourhoods) of voxel $\mathbf{x}$. The energy model is a function of the similarity between the target attribute vector $\mathbf{a}^T$ at voxel $\mathbf{x}$ and the source attribute vector $\mathbf{a}^S$ at voxel $\mathbf{y} = \mathbf{T}(\mathbf{x})$ plus a weighted bending energy regularisation term. This attribute similarity is defined as

$$S(\mathbf{a}^T(\mathbf{x}), \mathbf{a}^S(\mathbf{y})) = \begin{cases} 0 & \mathbf{a}_1^T(\mathbf{x}) \neq \mathbf{a}_1^S(\mathbf{y}) \\ C(\mathbf{a}^T(\mathbf{x}), \mathbf{a}^S(\mathbf{y})) & \text{otherwise} \end{cases}, \tag{3.22}$$

where

$$C(\mathbf{a}^T(\mathbf{x}), \mathbf{a}^S(\mathbf{y})) = (1 - |\mathbf{a}_2^T(\mathbf{x}) - \mathbf{a}_2^S(\mathbf{y})|) \cdot \prod_{i=1}^{K} (1 - |\mathbf{a}_3^{T^i}(\mathbf{x}) - \mathbf{a}_3^{S^i}(\mathbf{y})|). \tag{3.23}$$

To reduce the risk of getting trapped in local minima, the algorithm reduces the complexity of the similarity by finding *driving voxels* in both images based on their attribute vectors. Then, displacements are found by hierarchically matching driving voxels in the target and the source images according to how similar their attribute vectors are, and interpolating outside the driving voxel positions using a Gaussian kernel. The standard deviation of the Gaussian kernel is chosen so that every non-driving voxel depends on the distance from only one driving voxel.

In the paper by Ou et al. [158], the authors propose an alternative attribute matching image registration approach named "Deformable Registration via Attribute Matching and Mutual-Saliency Weighting" (DRAMMS). This approach utilises multi-scale and multi-orientation 3D Gabor attributes $\mathbf{a}^T$ and $\mathbf{a}^S$ extracted form each voxel in the target and source images, respectively. The energy model comprises a mutual saliency term and an attribute matching term plus a weighted regularisation term:

$$E_{\text{DRAMMS}} = \sum_{\mathbf{x} \in \Omega} \left(\text{ms}(\mathbf{x}, \mathbf{T}(\mathbf{x})) \cdot \frac{1}{D} \|\hat{\mathbf{a}}^T(\mathbf{x}) - \hat{\mathbf{a}}^S(\mathbf{T}(\mathbf{x}))\|^2 \right) + \alpha \, \text{R}(\mathbf{T}). \tag{3.24}$$

Here, D is the dimensionality of the attribute vectors. The mutual-saliency term is given by

$$\mathrm{ms}(\mathbf{x}, \mathbf{T}(\mathbf{x})) = \frac{\mathrm{MEAN}_{\mathbf{y}\in \mathrm{CN}(\mathbf{T}(\mathbf{x}))}[\mathrm{sim}(\mathbf{a}^T(\mathbf{x}), \mathbf{a}^S(\mathbf{y}))]}{\mathrm{MEAN}_{\mathbf{y}\in \mathrm{PN}(\mathbf{T}(\mathbf{x}))}[\mathrm{sim}(\mathbf{a}^T(\mathbf{x}), \mathbf{a}^S(\mathbf{y}))]}, \tag{3.25}$$

where the matching similarity $\mathrm{sim}(\cdot, \cdot)$ is given by

$$\mathrm{sim}(\mathbf{a}^T(\mathbf{x}), \mathbf{a}^S(\mathbf{y})) = \frac{1}{1 + \frac{1}{D}\|\mathbf{a}^T(\mathbf{x}) - \mathbf{a}^S(\mathbf{y})\|^2} \in [0, 1]. \tag{3.26}$$

$CN(\mathbf{x})$ corresponds to an immediate neighbourhood of voxel $\mathbf{x}$ and $PN(\mathbf{x})$ corresponds to a peripherical area further away from it. The idea behind this is that two matching voxels have high mutual saliency if they are similar to each other, but are not similar to anything else in the vicinity. In contrast to HAMMER [193] where all attributes are used, DRAMMS only considers the vectors of *optimal* attributes $\hat{\mathbf{a}}$. To select this optimal attribute set, the target image space Ω^T is partitioned into J regular regions Ω_j^T. A training voxel pair $(\hat{\mathbf{p}}, \hat{\mathbf{q}})$ is selected for each region, so that $\hat{\mathbf{p}}$ is in that region Ω_j^T, $\hat{\mathbf{q}}$ is anywhere in the source image space Ω^S and both maximise the following similarity based on the full attribute vectors $\tilde{\mathbf{a}}$:

$$(\hat{\mathbf{p}}_j, \hat{\mathbf{q}}_j) = \underset{\substack{\mathbf{p}\in\Omega_j^T \\ \mathbf{q}\in\Omega^S}}{\arg\max} \left[\mathrm{sim}(\tilde{\mathbf{a}}^T(\mathbf{p}), \tilde{\mathbf{a}}^S(\mathbf{q})) \cdot \mathrm{ms}(\mathbf{p}, \mathbf{q})\right]. \tag{3.27}$$

The optimal attributes are then selected from the full set of attributes by an iterative backward elimination and forward inclusion strategy trained using the voxel pairs $(\hat{\mathbf{p}}_j, \hat{\mathbf{q}}_j)$ such as:

$$(\hat{\mathbf{a}}^T, \hat{\mathbf{a}}^S) = \underset{\substack{\mathbf{a}_T^T \\ \mathbf{a}^S}}{\arg\max} \sum_{j=1}^{J} \left[\mathrm{sim}(\mathbf{a}^T(\hat{\mathbf{p}}_j), \mathbf{a}^S(\hat{\mathbf{q}}_j)) \cdot \mathrm{ms}(\hat{\mathbf{p}}_j, \hat{\mathbf{q}}_j)\right]. \tag{3.28}$$

3.7 Learning-based registration approaches

A learning-based method for non-rigid registration of multi-modal 2D images is presented in Guetter et al. [73]. The main idea consists on adding an additional term to a mutual information

(MI)-based energy function with regularisation:

$$E = -\alpha \, \mathrm{MI}(I_0, I \circ \mathbf{T}) + (1 - \alpha) \, \mathrm{KL}(I_0, I \circ \mathbf{T}) + \lambda \, \mathrm{R}(\mathbf{T}) \qquad \alpha \in [0, 1]. \tag{3.29}$$

The additional learning-based term $\mathrm{KL}(\cdot, \cdot)$ accounts for the Kullback-Leibler divergence of the observed joint intensity distribution for the target and source images, and a joint intensity distribution learnt from pre-aligned training images:

$$\mathrm{KL}(I_0, I \circ \mathbf{T}) = \sum_{t=1}^{N^{I_0}} \sum_{s=1}^{N^{I \circ \mathbf{T}}} P^o(t, s) \log \frac{P^o(t, s)}{P^l(t, s)}. \tag{3.30}$$

Here, P^o corresponds to the joint probability of the observed data under registration and P^l is the joint probability drawn from the learnt probability distribution. This learnt distribution steers the registration proccess and can be acquired in several ways, e.g., using clinical expert knowledge or hybrid scanner data. For improved robustness, a density distribution representing the mean prior information of the pre-aligned training images can be used.

In the work by Tang et al. [206], the RABBIT algorithm (which stands for "Rapid Alignment of Brains by Building Intermediate Templates") is presented. In this approach, registration is performed with the help of intermediate templates generated by a statistical deformation model learnt from deformation fields warping training source images to a common template, which are then analysed using PCA. The generation of these intermediate templates is performed by interpreting the statistical deformation model as a multidimensional Gaussian distribution where each dimension is an eigenvector of the covariance matrix. In this way, the coefficients of the statistical deformation model for the P eigenvectors with highest corresponding eigenvalues c_i $i = 1, \cdots, P$ can be found using a uniform coefficient sampling method proposed in [122]:

$$c_i = \sqrt{2} \cdot \mathrm{erf}^{-1}(2h_j - 1) \qquad j = 1, \cdots, n. \tag{3.31}$$

Here, h_j is the set of coefficients uniformly distributed in the interval $[0, 1)$ and $\mathrm{erf}^{-1}(\cdot)$ is the inverse of the Gaussian Error Function. Once the n^P intermediate templates are constructed

offline, an unseen image is registered to the common template by finding the closest intemediate template (using sum of squared intensity differences) to initialise the statistical model parameters and further optimised using Powell's optimisation algorithm iteratively. This yields an image that is closer to the common template. These two close images are further registered using a standard registration method (the HAMMER registration algorithm in the case of this work), resulting in the final transformation.

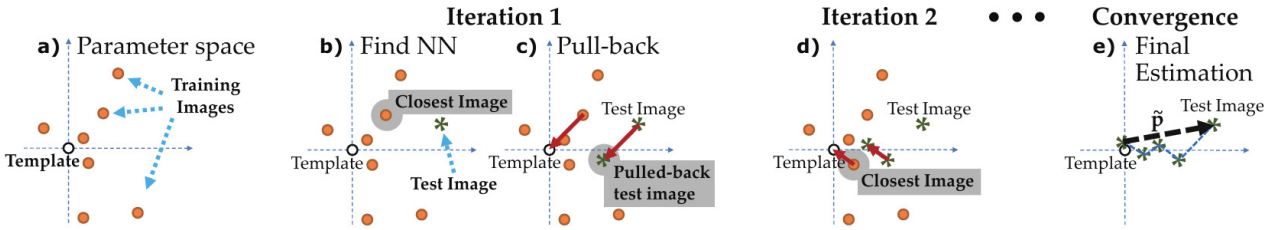


Figure 3.2: Algorithm for the estimation of the transformation parameters using the *pull-back* operation. Figure adopted from [212].

A different approach for the estimation of image distortion has been proposed by Tian et al. [212]. In this work, the authors introduce the concept of a *pull-back* operation, which is used to estimate the parameters of a dense deformation field that aligns the target (undistorted) and source (distorted) images. The nearest neighbour of the source image in parameter space from a set of several training image samples with known deformation parameters is used to pull-back the source image, i.e., the parameters of this nearest neightbour are applied on the image, yielding a less distorted image. Then, the process is iterated until convergence, which is proven to result in the global optimum being estimated. The algorithm is outlined in Figure 3.2. To counteract the effects of possible error accumulation, a cummulative parameter estimation is used to pull-back the original source image on each iteration. A key feature of this method is that the chosen nearest neighbour does not need to be in the neighbourhood of the source image (is not ϵ-close) in parameter space. Hence, only a number of training images that grows logarithmically with the desired accuracy is needed, thus breaking the *curse of dimensionality* [140].

Another registration method that learns the appearance-deformation correlation is proposed by Kim et al. [111]. The basic idea is to learn a PCA statistical deformation model from

deformation fields warping training source images to a common template, as in RABBIT [206]. Then, different deformation fields are generated by randomly perturbing the coefficients of the statistical model. These are then inverted to generate additional training images. From this augmented set of training images, a number of signature images are obtained. These signature images capture the boundaries between grey matter, white matter and cerebrospinal fluid labels to represent the shape variations of the training images (although a standard edge detector can be used as well). By using these signature images, a PCA signature model is constructed and applied to them, yielding low-dimensional signature vectors. Finally, a support vector regression approach is used in order to relate each of the coefficients of the statistical deformation model to the set of low-dimensional signature vectors. When registering an unseen image to the common template, the PCA signature model is utilised to obtain the low-dimensional representation of this unseen image. Then each coefficient of the statistical deformation model is estimated using the corresponding regressor and the computed low-dimensional vector. This allows to obtain the deformation that is applied to the image, yielding an intermediate template which is much closer to the common template. Similar to RABBIT, the deformation is further optimised by means of a standard registration method between these two close images.

Loeckx et al. [131] presents a method for registration of 2D thorax X-ray radiography images of the same subject. In this approach, a B-spline FFD transformation model is utilised. A statistical deformation model trained from previous registrations of a set of similar pairs of follow-up images of several subjects is used to constrain the B-spline FFD transformation parameters. In general, to construct a statistical deformation model, the transformation parameters Φ_i of the N training registrations are vectorised and assembled together to form a matrix $\mathbf{\Phi} = [\Phi_1|\cdots|\Phi_N]$ and PCA is applied to $\mathbf{\Phi}$. In this method, the affine translation and scaling components of the transformation are explicitly removed from matrix $\mathbf{\Phi}$:

$$\mathbf{\Phi}' = \mathbf{\Phi} - \mathbf{\Phi}_{\text{aff}}\mathbf{\Phi}_{\text{aff}}^{\top}\mathbf{\Phi}. \tag{3.32}$$

Here $\mathbf{\Phi}_{\text{aff}} = [\mathbf{t}_x|\mathbf{t}_y|\mathbf{s}]$ contains the affine components. Then, PCA is performed on $\mathbf{\Phi}'$ yielding

the statistical deformation model basis $\mathbf{B}'$, which is augmented with the affine components to yield the final basis matrix $\mathbf{B}$:

$$\mathbf{B} = [\mathbf{t}_x|\mathbf{t}_y|\mathbf{s}|\mathbf{B}']. \tag{3.33}$$

Once the statistical deformation model is learnt, the transformation parameters for any new pair of follow-up images is given by:

$$\Phi_{\text{new}} = \overline{\Phi} + \mathbf{B}\mathbf{w}. \tag{3.34}$$

Here, $\overline{\Phi}$ is the mean of the training transformation parameters and $\mathbf{w}$ is the parameter vector that has to be optimised. The number of elements in $\mathbf{w}$ corresponds to the number of principal components retained in the PCA step plus three to account for the affine components.

The GRAM framework, which stands for "Geodesic Registration on Anatomical Manifolds", has been proposed by Hamm et al. [79]. This method allows large deformation registrations over a dataset of images using a learnt manifold of anatomical variation over this dataset. The main advantage is that the learnt manifold restricts deformations to represent biologically plausible geodesic paths of registration, in contrast to the much broader analytical manifold of diffeomorphisms. To construct the empirical manifold, coarse diffeomorphic registrations are performed between all N images $I_1, \cdots, I_N$ in the dataset. Then, their pairwise distances are computed as:

$$d_{ij} = w \cdot \text{SSD}(I_i, I_j(\mathbf{T}_{ij})) + (1 - w) \cdot \text{HE}(\mathbf{T}_{ij}). \tag{3.35}$$

Here, w is a weighting term and $\text{HE}(\mathbf{T}_{ij})$ corresponds to the harmonic energy of the transformation, i.e., the mean Frobenius norm of the Jacobians. The next step is to construct a k-NN graph based on the distances and find the shortest paths between all pairs (e.g., by using Dijkstra's or Floyd-Warshall's algorithm). The length g_{ij} of a geodesic path is then the sum of the edge lengths along that path. The next step is to define the "template" image. In this work the template image is the median image I_t such that $t = \arg\min_i \sum_j g_{ij}$. The large deformations from the template image to all other images are then computed by a recursive

composition of the small deformations from its edges along the geodesic path. From this point, any unseen image I_* is coarsely registered to all the images in the dataset, generating N new transformations $\mathbf{T}_{*i}$. Finally, the transformation $\mathbf{T}_{*j}$ for which d_{*j} is minimal is composed with $\mathbf{T}_{jt}$ to obtain the large deformation between the unseen image and the template.

3.8 Summary

Table 3.1: Summary of the image-based non-rigid registration methods described in this chapter.

Type	**Methods**
Foundational	Optical flow [96]; Demons [211]; LDDMM [24]; DARTEL [13].
Inverse-consistent/symmetric	Christensen et al. [35]; SyN [16]; Log-domain [222]; Modat et al. [145]; LCC-Demons [134].
Attribute-matching	HAMMER [193]; DRAMMS [158].
Learning-based	Guetter et al. [73]; RABBIT [206]; Tian et al. [212]; Kim et al. [111]; 2D Statistical Deformation Model [131]; GRAM [79].

In this chapter, a number of important registration methods have been described (see Table 3.1). We have focused on the main classical approaches that aided in the development of intensity-based non-rigid image registration, such as optical flow estimation, Demons algorithm, LDDMM, and DARTEL. We also focused on inverse consistent and symmetric approaches, which pose important properties for many applications that require inversion of the transformations. Attribute-matching aproaches were also described to show that high dimensional data can be used as an alternative to pure intensity data. Finally, we outlined a number of learning-based approaches since they closely relate to most of the methods outlined in this thesis. The different nature and properties of all methods have been outlined here. This demonstrates that image registration is usually an application-specific task and that no approach is general enough to work well on every possible scenario. This makes medical image registration a fruitful area for research, but on the same time a very challenging one. In the next three chapters, a detailed description of the main contributions of this thesis is provided.

Chapter 4

Non-rigid image registration using Statistical Deformation Models

This chapter is based on:

- Stefan Pszczolkowski, Luis Pizarro, Ricardo Guerrero, and Daniel Rueckert. "Nonrigid free-form registration using landmark-based statistical deformation models." In Proceedings of SPIE, vol. 8314, p. 831418. 2012.

4.1 Introduction

One of the most clinically important applications of image registration are cohort studies. Since the anatomical variability within the population in these studies can be potentially very large, non-rigid registrations to a common template space have to be performed in order to be able to subsequently conduct voxel-based or statistical comparisons within that common space. However, standard non-rigid image registration problems usually involve a large amount of degrees of freedom. This amount is normally in the order of hundreds of thousands or even millions. Therefore, it is necessary to further impose certain constraints to these registrations so as to produce results that are coherent with respect to the group in study.

In this chapter, a new registration approach referred to as *statistically-based FFD registration (SFFD)* is proposed. This registration method is a modification of a well-known free-form deformations (FFD) approach [183]. Within this new framework any new unseen subject can be registered in a very efficient way by means of a *statistical deformation model* (SDM) learnt a priori from N known deformations of similar images to a given common template space. These SDMs capture the variability of the known deformations and constrain the registration procedure to produce only statistically likely warps. Also, the number of parameters that must be optimised is dramatically reduced. Only a single-resolution optimisation is needed to account for coarse and fine local displacements, in contrast to the multi-resolution strategy commonly employed by the FFD-based registration. Moreover, in case where anatomical landmarks are available, we propose to encode them within the statistical framework, in order to provide a better alignment of the anatomical structures described by these landmarks.

4.1.1 Related work

A related approach has been previously proposed by Loeckx et al. [131]: Here, a statistical deformation model is trained using Principal Component Analysis (PCA) for the registration of thorax X-ray radiography images of the same subject for temporal substraction. In contrast

to this work, the proposed method uses deformations to train the SDM that are obtained automatically via registration, rather than (semi-)manually. Furthermore, we investigate different statistical learning approaches beyond PCA. Another important difference is the ability of this approach to perform intra- and inter-subject registration on more complex images, such as 3D brain MR images.

Other previous contributions on the use of statistics for medical image registration include the work by Wang and Staib [227], where statistical information of boundary points is incorporated into a Bayesian formulation of the objective function, which is subsequently used for elastic and fluid registrations. Benameur et al. [25] use a statistical deformable template to constrain a set of admissible deformations for the registration of scoliotic vertebrae. Xue et al. [234] performs PCA on each band of wavelets coefficients of voxel-wise deformations to obtain an SDM that permits the construction of a statistically-constrained voxel-wise registration framework. Finally, PCA models are used to generate intermediate templates for faster registration in the methods by Kim et al. [110] and Tang et al. [206].

Manually determined anatomical landmarks or salient points automatically detected within the images via feature extraction algorithms can be used to drive the registration process to produce deformations that align them as accurately as possible. In Johnson and Christensen [105] images are registered by iteratively aligning landmark positions and image intensities away from these locations. A similar approach is presented by Fischer and Modersitzki [56], but in this case the landmark locations are used as a constraint for the energy function. Hartkens et al. [82] uses point and surface information to correct for large differences between pre- and post-resection images. Rohr et al. [179] proposes a landmark-based elastic registration algorithm where thin-plate splines resulting from a minimising functional allow to weight the landmarks according to their localisation uncertainty. Finally, Pennec et al. [163] present an approach to extract features like crest lines and extremal points and perform rigid registration between them.

4.2 Methods

4.2.1 Statistical model construction

To construct the statistical deformation model (SDM) that is subsequently used in the proposed approach, we register a population of N subjects $S_1, \ldots, S_N$ to a common reference template using the standard B-spline free-form deformation (FFD) model of Rueckert et al. [183], which is described in section 2.1.3. The P control point values of the resulting free-form deformations (FFDs) are vectorised to form a matrix $\mathbf{\Phi} = [\Phi_1| \cdots |\Phi_N] \in \mathbb{R}^{P \times N}$. Generalising the approach of Rueckert et al. [182] and assuming that matrix $\mathbf{\Phi}$ has been mean centered, the proposed method seeks to find a factorisation such that

$$\mathbf{\Phi} \approx \mathbf{B}\mathbf{W}^\top, \tag{4.1}$$

Here, $\mathbf{B} \in \mathbb{R}^{P \times C}$ is the statistical model encoding matrix we are interested in, and $\mathbf{W} \in \mathbb{R}^{N \times C}$ just acts as a matrix whose rows correspond to vectors of weighting factors. Note there is a free parameter $C \leq N$. This parameter is user-defined and controls the number of basis vectors used to describe the statistical deformation model. As a result, C also controls the number of parameters that are required to instantiate the SDM. The factorisation approach is illustrated in Figure 4.1. As we later show in this chapter, the purpose of this factorisation model in the context of image registration is twofold: (1) capture the variability of the known training deformations with a reduced number C of basis vectors, and (2) constrain the subsequent statistical registration to produce only statistically likely types of mappings.

In the approach by Rueckert et al. [182] principal component analysis (PCA) is utilised, albeit only for statistical analysis, not for registration. We explore two further techniques, namely independent component analysis (ICA) and semi-Nonnegative matrix factorisation (SNMF). All three factorisation methods are described in the following section, with emphasis on how they are used in the proposed framework.

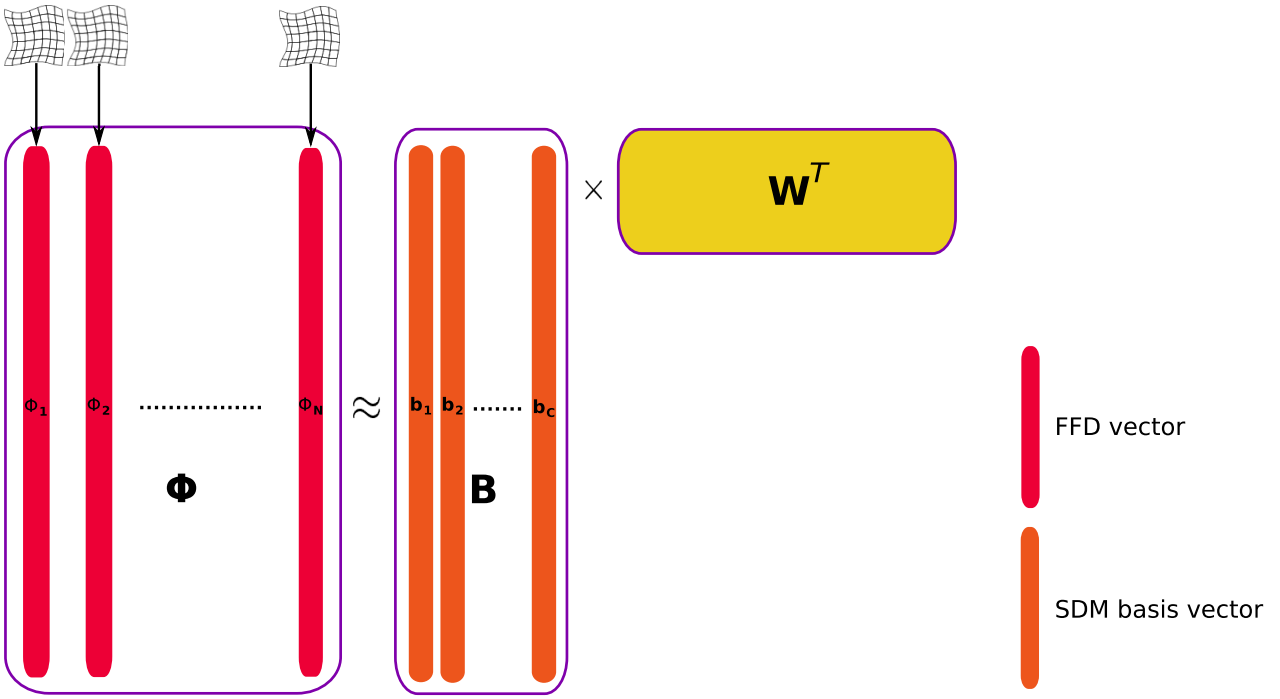


Figure 4.1: Factorisation approach to learn the statistical deformation model encoded in matrix **B**.

4.2.2 Statistical factorisation techniques

Principal Component Analysis (PCA)

Principal component analysis (PCA) [106] is utilised to decompose multidimensional input data, which is assumed to be mean centered, into a set of C orthogonal components sorted according to the amount of variance in the data that they explain. In other words, it finds the matrix $\mathbf{B}$ which corresponds to the eigendecomposition of the covariance matrix $\mathbf{S}$ of the input data $\mathbf{S} = cov(\mathbf{\Phi})$. Hence, each of the C columns $\mathbf{b}_i$, $i = \{1, .., C\}$ of $\mathbf{B}$ is an eigenvector that satisfies

$$\mathbf{S}\mathbf{b}_i = \lambda_i \mathbf{b}_i \tag{4.2}$$

with corresponding eigenvalue λ_i such that $\lambda_i \geq \lambda_{i+1}$. The columns of $\mathbf{B}$ (eigenvectors) are thus sorted by the amount of variance that is explained (eigenvalues). Therefore, it is possible to control the amount of variance that is taken into consideration by choosing different values for C (the higher the value of C, the higher the amount of variance that is explained).

The covariance matrix $\mathbf{S}$ is a $P \times P$ matrix. This means that computing the eigendecomposition of this matrix is an intractable task if P is large (typically, $P \gg N$). Luckily, the eigendecomposition can still be computed by solving on a much smaller eigenvector problem for a $N \times N$ matrix [40]:

$$\mathbf{S}' = \frac{1}{N}\mathbf{\Phi}^\top \mathbf{\Phi}. \tag{4.3}$$

From the eigendecomposition of $\mathbf{S}'$, eigenvalues λ_i' and eigenvectors $\mathbf{b}_i'$ are obtained. Then, λ_i and $\mathbf{B}$ are computed as [40]:

$$\begin{aligned} \lambda_i &= \lambda_i' \\ b_{ij} &= \frac{1}{\sqrt{C \cdot \lambda_j}}[\mathbf{\Phi}\mathbf{B}']_{ij}, \end{aligned} \tag{4.4}$$

where $\mathbf{B}' = [\mathbf{b}_1'| \cdots |\mathbf{b}_C']$.

Independent Component Analysis (ICA)

The method of independent component analysis (ICA) [108, 39] separates multidimensional input data, which is assumed to be mean centered, into components that are maximally independent, where different measures of independence can be utilised. ICA is normally used for the separation of mixed signals, a process known as blind source separation. From the point of view of matrix factorisation, it finds a matrix $\mathbf{B}$ such that

$$\mathbf{\Phi} \approx \mathbf{B}\mathbf{W}^\top. \tag{4.5}$$

where $\mathbf{W}$ is an orthogonal weighting matrix and the rows of $\mathbf{B}$ are maximally independent. During the process of ICA the data $\mathbf{\Phi}$ has to be whitened (i.e., made unit variance and uncorrelated). To whiten the data, PCA is performed on it, yielding a matrix of eigenvectors $\mathbf{U}$ and a diagonal matrix $\mathbf{\Sigma}$ with the corresponding eigenvalues in the main diagonal. The whitened data $\tilde{\mathbf{\Phi}}$ is then obtained as follows:

$$\tilde{\mathbf{\Phi}}^\top = \mathbf{\Sigma}^{-1/2}\mathbf{U}^\top \mathbf{\Phi}^\top. \tag{4.6}$$

The PCA step in the whitening process is often utilised to reduce the dimensionality of the data, in order to perform ICA on simplified data. If the whitening process reduces the dimensionality to C_{white} dimensions, then C independent components can be obtained from the ICA algorithm, where $1 \leq C \leq C_{white} \leq N$.

As previoulsy mentioned, the goal of ICA is that the rows of $\mathbf{B}$ are maximally independent. Hence, each observation of the whitened data $\tilde{\phi}_{ij} = \sum_{k=1}^{C} b_{ik} w_{jk}$ is a weighted sum of maximally independent variables $b_{i,1\ldots C}$. One way of achieving statistical independence for a set of random variables is by minimising their (pairwise) mutual information. The motivation for minimising mutual information comes from the observation that two variables are independent if information about the value of one of them does not contain any information about the value of the other. This means that their mutual information is zero. Another approach to achieve statistical independence is to maximise non-Gaussianity. The fact that non-Gaussianity is a measure for independence arises from the Central Limit Theorem which states that the distribution of a sum of independent random variables converges towards a Gaussian distribution. Moreover, in the papers of Hyvärinen [99] and Hyvärinen and Oja [100], the authors show that minimising the mutual information of the resulting components is (roughly) equivalent to maximising their negentropy $J(\cdot)$, which is a measure of non-Gaussianity:

$$J(b_{i,1\ldots C}) = H(b^{gauss}_{i,1\ldots C}) - H(b_{i,1\ldots C}). \tag{4.7}$$

Here, $b^{gauss}_{i,1\ldots C}$ is a random variable from a Gaussian distribution with the same mean and variance as the distribution of $b_{i,1\ldots C}$, and $H(\mathbf{x})$ is the entropy of the (multidimensional) random variable $\mathbf{x}$:

$$H(\mathbf{x}) = -\int_u f_{\mathbf{x}}(u) \log(f_{\mathbf{x}}(u)) du. \tag{4.8}$$

Here, $f_{\mathbf{x}}(u)$ is the probability density function of $\mathbf{x}$. Negentropy can be considered as a measure of non-Gaussianity since a Gaussian variable has the largest entropy among all random variables of equal variance [44, 160]. Hence $J(b_{i,1\ldots C})$ is strictly non-negative, and reaches a minimum of zero when $b_{i,1\ldots C}$ has a Gaussian distribution.

Algorithm 1 FastICA algorithm with simultaneous update. $\mathbf{1}_N$ is a column vector whose N entries are equal to one, $Avg(\mathbf{X})$ corresponds to averaging over the columns of $\mathbf{X}$, and $\otimes$ is the Hadamard product.

Input: C Number of desired components
Input: $\mathbf{\Phi}^\top \in \mathbb{R}^{N \times P}$ Matrix containing the mean centered data in row-layout
Output: $\mathbf{B}^\top \in \mathbb{R}^{C \times P}$ Matrix encoding the statistical deformation model

Obtain $\mathbf{U}$ and $\mathbf{\Sigma}$ by performing PCA on $\mathbf{\Phi}$ retaining $C_{white} = N$ eigenvectors and eigenvalues

$\mathbf{X}_{whitening} = \mathbf{\Sigma}^{-1/2}\mathbf{U}^\top$
$\tilde{\mathbf{\Phi}}^\top = \mathbf{X}_{whitening}\mathbf{\Phi}^\top$
Randomly initialise orthonormal matrix $\mathbf{A} \in \mathbb{R}^{N \times C}$
while not converged **do**
  $\mathbf{A} \leftarrow \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1/2}$
  $\mathbf{A} \leftarrow \frac{1}{P} \cdot \tilde{\mathbf{\Phi}}^\top \cdot g(\tilde{\mathbf{\Phi}}\mathbf{A}) - \left[\mathbf{1}_N \cdot Avg(g'(\tilde{\mathbf{\Phi}}\mathbf{A}))\right] \otimes \mathbf{A}$
end while
$\mathbf{W}^\top \leftarrow \mathbf{A}^\top\mathbf{X}_{whitening}$
return $\mathbf{B}^\top = \mathbf{W}^\top\mathbf{\Phi}^\top$

The authors in [99, 100] also propose one of the most popular algorithm for ICA: The FastICA algorithm. This algorithm has two variants: One where each column of $\mathbf{W}$ is estimated one-by-one, and another where all columns are simultaneously estimated. We utilise a version based on the second variant, which is described in Algorithm 1, including the preprocessing and postprocessing steps. Note that the algorithm involves the use of a function $g(\cdot)$ and its derivative $g'(\cdot)$. This function is utilised to obtain an approximation of negentropy, since it is difficult to compute in practice. The authors in [99, 100] state that a good function for general-purpose FastICA is $g(x) = \tanh(x)$, with $g'(x) = 1 - \tanh^2(x)$. This is also the $g(\cdot)$ and $g'(\cdot)$ we use in this thesis.

Semi-Nonnegative Matrix Factorisation (SNMF)

Non-negative matrix factorisation (NMF) [124] is a matrix decomposition method that assumes that the input data and resulting factors have only non-negative entries. These constraints cause the decomposition to produce non-holistic parts-based representations, since only additive, but no substractive combinations are allowed. In order to expand the suitability of NMF to cases

where the input data is not strictly non-negative (as is the case with our application), Ding et al. [48] proposed an NMF variant called semi-Nonnegative matrix factorisation (SNMF). This algorithm seeks to find a factorisation of the input data $\mathbf{\Phi}$

$$\mathbf{\Phi} \approx \mathbf{B}\mathbf{W}^\top, \tag{4.9}$$

constraining only $\mathbf{W}$ to have non-negative entries, while leaving $\mathbf{B}$ unconstrained. The SNMF algorithm can also be motivated from a soft-clustering point of view: Here, the columns of $\mathbf{B}$ represent cluster centroids and $\mathbf{W}$ contains the soft membership indicators for each column of $\mathbf{\Phi}$. In contrast to PCA and ICA, SNMF does not pose any assumption on the data being mean centered and is thus usually run on the original data (with no prior mean centering). The procedure starts by running a hard-clustering k-means algorithm [81] with $k = C$ over the input data $\mathbf{\Phi}$. This yields an indicator matrix $\mathbf{W}$, where $w_{ik} = 1$ if Φ_i belongs to cluster k. Then, $\mathbf{B}$ and $\mathbf{W}$ are alternately updated in an iterative fashion until convergence

$$\begin{aligned} &(1) \quad \mathbf{B} = \mathbf{\Phi}\mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1} \\ &(2) \quad w_{ik} \leftarrow w_{ik}\sqrt{\frac{(\mathbf{\Phi}^\top\mathbf{B})^+_{ik} + \left[\mathbf{W}(\mathbf{B}^\top\mathbf{B})^-\right]_{ik}}{(\mathbf{\Phi}^\top\mathbf{B})^-_{ik} + \left[\mathbf{W}(\mathbf{B}^\top\mathbf{B})^+\right]_{ik}}}. \end{aligned} \tag{4.10}$$

Here, $\mathbf{A}^+_{ik} = (|\mathbf{A}_{ik}| + \mathbf{A}_{ik})/2$ and $\mathbf{A}^-_{ik} = (|\mathbf{A}_{ik}| - \mathbf{A}_{ik})/2$. After the process finishes, we apply a Gram-Schmidt process [93] on the resulting matrix $\mathbf{B}$ as a postprocessing step to obtain orthogonal bases.

4.2.3 Statistically-based FFD registration (SFFD)

Once the SDM matrix $\mathbf{B}$ is obtained, the proposed approach seeks to register a new unseen subject $S^* \notin \{S_1, \ldots, S_N\}$ by introducing a modification to the classical FFD transformation model of equation (2.19):

$$\mathbf{T}^{\text{SFFD}}(\mathbf{x}) = \mathbf{T}_{global}(\mathbf{x}) + \mathbf{T}^{\text{SFFD}}_{local}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{T}^{\text{SFFD}}_{local}(\mathbf{x}) \tag{4.11}$$

Here, $\mathbf{T}^{\text{SFFD}}_{local}(\mathbf{x})$ are the local displacements for S^* given by:

$$\mathbf{T}^{\text{SFFD}}_{local}(\mathbf{x}) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\Phi^{(\mathbf{w})}_{i+l,j+m,k+n}, \tag{4.12}$$

where $\Phi^{(\mathbf{w})}$ are the FFD control point values parameterising the transformation computed from the SDM:

$$\Phi^{(\mathbf{w})} = \mathbf{B}\mathbf{w} + \overline{\Phi}. \tag{4.13}$$

Here, $\overline{\Phi}$ is the arithmetic mean of the control point values of the training FFDs and $\mathbf{w} \in \mathbb{R}^{C\times 1}$ is a vector containing the actual parameterisation of the transformation. The mean control point vector $\overline{\Phi}$ is added back because we assume mean centered data. This effectively models the bias of the transformation to the common template. One way of avoiding this bias is to use groupwise registration. Note that registering S^* directly with the standard FFD registration method involves optimising P variables, in contrast of the C variables of the proposed method. Thus, since in most cases $C \ll P$, this method allows for a substantial reduction in the number of degrees of freedom in the optimisation. A summary of the proposed framework is depicted on Figure 4.2.

4.2.4 Landmark-based FFD registration (LFFD)

The statistical registration assumes that a set of N pre-computed FFDs are available. As already discussed, these examples can be obtained via the standard FFD registration algorithm. However, if *a priori* information about the position of anatomical landmarks is available, it is posible to incorporate this information into the statistical deformation model. For this purpose, we propose an alternative approach to obtain the pre-computed FFDs. This approach can be realised as the standard FFD optimising an energy functional as the one in equation (2.54)

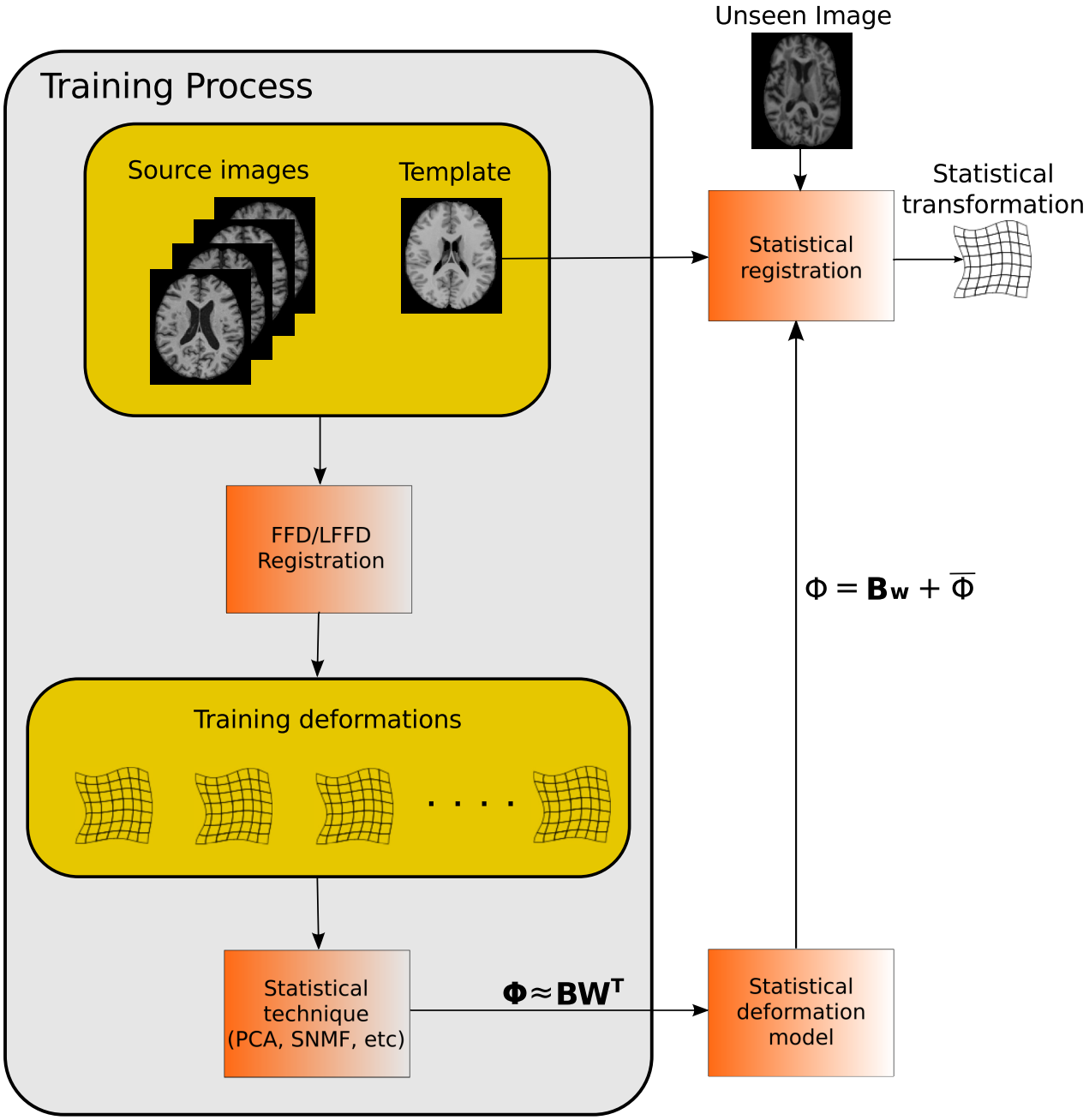


Figure 4.2: Proposed statistical framework. Note that the training process needs to be performed only once in order to be able to statistically register any number of unseen images.

with an additional term that accounts for the alignment of the landmarks:

$$E_{\text{LFFD}}(I_0; I; l^{I_0}; l^I; \Phi) = E_{\text{FFD}}(I_0; I; \Phi) - \gamma\, E_{landmark}(l^{I_0}; l^I; \Phi). \tag{4.14}$$

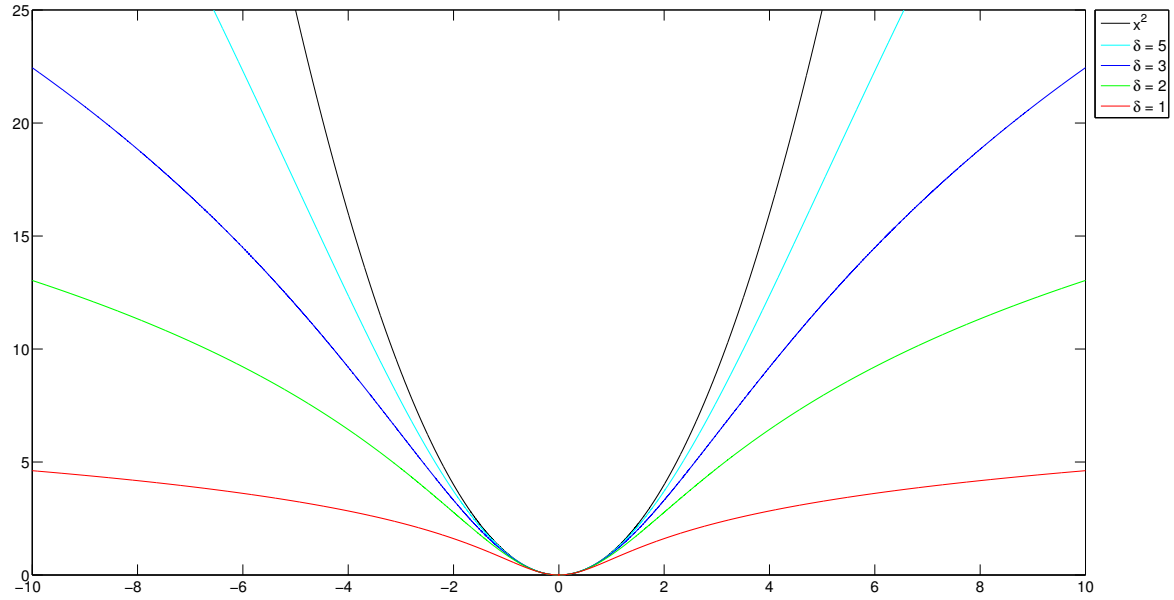


Figure 4.3: Cauchy function $\Psi(x^2) = \delta^2 \log(1 + (x^2/\delta^2))$ with different values of the parameter δ togheter with $\Psi(x) = x^2$ (black). Note how large errors are less severely penalised. Note also the inflection poins in $x = -\delta$ and $x = \delta$.

Here, $E_{\text{FFD}}(I_0; I; \Phi)$ is the standard FFD energy functional defined in equation 2.54 (as a function of the parameters Φ of the transformation), γ is a tradeoff parameter and $E_{landmark}$ is defined as

$$E_{landmark}(l^{I_0}; l^I; \Phi) := \frac{1}{L} \sum_{i=1}^{L} \Psi \left(\| l_i^I - \mathbf{T}_\Phi(l_i^{I_0}) \|^2 \right). \tag{4.15}$$

Here, L is the number of landmarks, l_i^I is the location of the i-th landmark in the source image I, and $\mathbf{T}_\Phi(l_i^{I_0})$ is the location of the i-th landmark in the reference template I_0 mapped onto I. We employ a robust function Ψ to control the penalisation of large landmark discrepancies [98]. We choose the Cauchy function $\Psi(x^2) = \delta^2 \log(1 + (x^2/\delta^2))$ [95]. The parameter δ represents an inflection point on the curve from which large deviations (outliers) are less strongly penalised. Figure 4.3 shows a plot of Ψ for different values of the parameter δ and $\Psi(x^2) = x^2$.

We refer the interested reader to appendix A for a detailed derivation of the gradient of the energy function in equation 4.14.

Algorithm 2 Conjugate gradient optimisation employed on all registrations.

```
max_step ← voxel_size (in mm)
min_step ← 0.01 × max_step
repeat
  Compute conjugate gradient using Polak-Ribière-Polyak formula [167, 168]
  step_size ← max_step
  repeat
    Make line step according to step_size and evaluate energy value
    if energy improvement then
      step_size ← 1.1 × step_size
      if step_size > max_step then
        step_size ← max_step
      end if
    else
      step_size ← 0.5 × step_size
    end if
  until step_size < min_step
until no energy improvement
```

4.3 Results

In all experiments presented in this section, we compare the standard B-spline FFD registration method and the statistical training approaches described in the previous section: Principal Component Analysis (PCA), Independent Component Analysis (ICA), Semi-Nonnegative Matrix Factorisation (SNMF). We also perform the comparisons using Semi-Nonnegative Matrix Factorisation performed on zero-centered data (SNMF-C). For all training and testing registrations, we optimise normalised mutual information (NMI) as similarity measure and thin-plate bending energy as regulariser with a weight of $\alpha = 0.0002$ for brain data and $\alpha = 0.002$ for cardiac data using conjugate gradient descent and line search as summarised in Algorithm 2. For the landmark-based registrations, we set the robust function parameter δ to 3mm. Finally, all the reported p-values are obtained using paired Wilcoxon signed rank tests.

4.3.1 Data

In this section we describe the datasets employed to test the performance of the proposed method. For convenience, the data is also summarised in table 4.1.

Table 4.1: Summary of the datasets utilised to test the proposed method.

Dataset	Number of Images	Modality	Organ	Annotation	Size	Resolution (mm)
ADNI	820	MR	Brain	20 landmarks	$182 \times 218 \times 182$	$1 \times 1 \times 1$
OASIS	30	MR	Brain	134 structures	$256 \times 256 \times Z, Z \in [261, 320]$	$1 \times 1 \times 1$
Cardiac	10 sequences $\times$ 30 frames	MR	Heart	3 structures	$N \times N, N = 240, 256, 288$	1.2308×1.2308

The first set of image data corresponds to 820 preprocessed 1.5 Tesla T1-weighted images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [149] with $L = 20$ landmarks per subject, as defined in [182] (see Figure 4.4 and Table 4.2). All the images have $182 \times 218 \times 182$ voxels with an isotropic resolution of 1mm. They are also skull stripped and affinely aligned to the Montreal Neurological Institute (MNI) template [71]. Exemplary subjects can be seen in Figure 4.5.

Table 4.2: Summary of the 20 landmarks utilised for the landmark-based registrations as defined in [182].

Landmark ID	Anatomical landmark
1	Splenium of corpus callosum (outer aspect)
2	Splenium of corpus callosum (inferior tip)
3	Splenium of corpus callosum (inner aspect)
4	Genu of corpus callosum (outer aspect)
5	Genu of corpus callosum (inner aspect)
6	Superior aspect of the pons
7	Inferior aspect of the pons
8	Superior aspect of the cerebellum
9	Fourth ventricle
10	Putamen posterior (left)
11	Putamen anterior (left)
12	Putamen posterior (right)
13	Putamen anterior (right)
14	Anterior commissure
15	Posterior commissure
16	Inferior aspect of the cerebellum
17	Anterior tip of lateral ventricle (left)
18	Anterior tip of lateral ventricle (right)
19	Inferior tip of lateral ventricle (left)
20	Inferior tip of lateral ventricle (right)

Secondly, we utilise 35 preprocessed T1-weighted MR brain images originating from the OASIS database [139], which have been manually segmented by experts[1] into 138 anatomical structures.

[1] provided by Neuromorphometrics, Inc. under academic subscription. (`www.neuromorphometrics.com`)

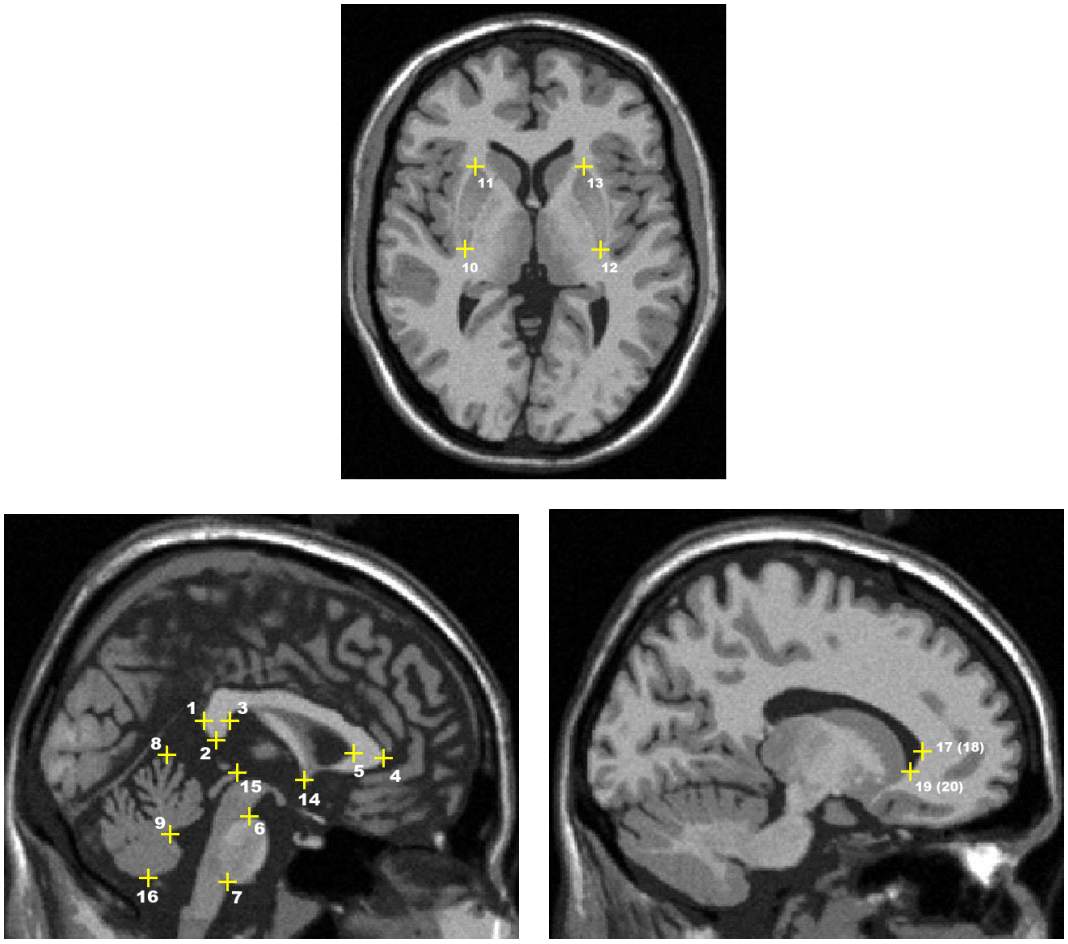


Figure 4.4: Distribution of the 20 landmarks defined in [182]. See table 4.2 for details on the landmark IDs.

Four structures are ignored because they do not appear consistently across the dataset (see Table 4.3 for the final list of structures). From this set of images, 5 duplicates (follow-up versions of images already in the dataset) were discarded, leaving a final set of 30 images. The images are skull stripped, with an image size of $256 \times 256 \times Z$ voxels with Z ranging between 261 and 320, and isotropic resolution of 1mm. Figure 4.6 illustrates some example subjects with their anatomical segmentations as coloured contours.

Finally, an image set of $3 + t$-dimensional 1.5T Philips Achieva SSFP MR cardiac sequences from 10 subjects is used. Each sequence covers one complete cardiac cycle over 30 frames, the first frame being where the end of diastole occurs. We extract one mid-ventricular short-axis slice to produce $2 + t$-dimensional sequences. Additionally, both end-diastole and end-systole frames of each subject are manually segmented into 3 structures: right ventricle cavity (RVC),

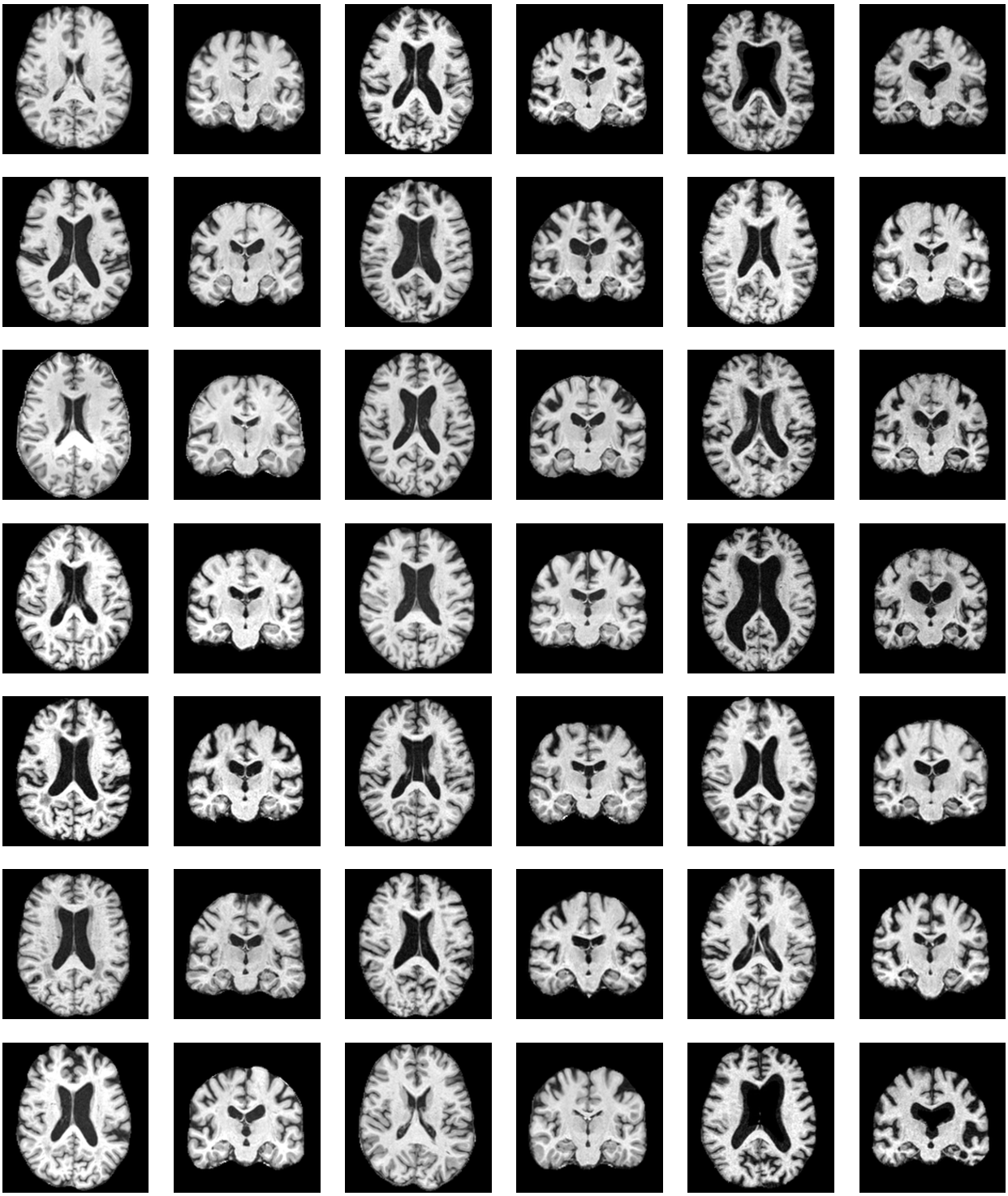


Figure 4.5: Example subjects from the ADNI dataset. The first and second columns depict control subjects (CN) in axial and coronal view, respectively. The third and fourth columns depict subjects with mild cognitive impairment (MCI) in axial and coronal view, respectively. The fifth and sixth columns depict subjects with Alzheimer's Disease (AD) in axial and coronal view, respectively.

Table 4.3: Summary of the 134 structures used in this thesis for the OASIS data. The original structure list includes 4 more labels, which we ignore because they do not appear consistently across the dataset.

Structure ID	Label	Structure ID	Label
1	3rd Ventricle	68	Left lateral orbital gyrus
2	4th Ventricle	69	Right middle cingulate gyrus
3	Right Accumbens Area	70	Left middle cingulate gyrus
4	Left Accumbens Area	71	Right medial frontal cortex
5	Right Amygdala	72	Left medial frontal cortex
6	Left Amygdala	73	Right middle frontal gyrus
7	Brain Stem	74	Left middle frontal gyrus
8	Right Caudate	75	Right middle occipital gyrus
9	Left Caudate	76	Left middle occipital gyrus
10	Right Cerebellum Exterior	77	Right medial orbital gyrus
11	Left Cerebellum Exterior	78	Left medial orbital gyrus
12	Right Cerebellum White Matter	79	Right postcentral gyrus medial segment
13	Left Cerebellum White Matter	80	Left postcentral gyrus medial segment
14	Right Cerebral White Matter	81	Right precentral gyrus medial segment
15	Left Cerebral White Matter	82	Left precentral gyrus medial segment
16	Cerebrospinal Fluid	83	Right superior frontal gyrus medial segment
17	Right Hippocampus	84	Left superior frontal gyrus medial segment
18	Left Hippocampus	85	Right middle temporal gyrus
19	Right Inf Lateral Ventricle	86	Left middle temporal gyrus
20	Left Inf Lateral Ventricle	87	Right occipital pole
21	Right Lateral Ventricle	88	Left occipital pole
22	Left Lateral Ventricle	89	Right occipital fusiform gyrus
23	Right Pallidum	90	Left occipital fusiform gyrus
24	Left Pallidum	91	Right opercular part of the inferior frontal gyrus
25	Right Putamen	92	Left opercular part of the inferior frontal gyrus
26	Left Putamen	93	Right orbital part of the inferior frontal gyrus
27	Right Thalamus Proper	94	Left orbital part of the inferior frontal gyrus
28	Left Thalamus Proper	95	Right posterior cingulate gyrus
29	Right Ventral DC	96	Left posterior cingulate gyrus
30	Left Ventral DC	97	Right precuneus
31	Optic Chiasm	98	Left precuneus
32	Cerebellar Vermal Lobules I-V	99	Right parahippocampal gyrus
33	Cerebellar Vermal Lobules VI-VII	100	Left parahippocampal gyrus
34	Cerebellar Vermal Lobules VIII-X	101	Right posterior insula
35	Left Basal Forebrain	102	Left posterior insula
36	Right Basal Forebrain	103	Right parietal operculum
37	Right anterior cingulate gyrus	104	Left parietal operculum
38	Left anterior cingulate gyrus	105	Right postcentral gyrus
39	Right anterior insula	106	Left postcentral gyrus
40	Left anterior insula	107	Right posterior orbital gyrus
41	Right anterior orbital gyrus	108	Left posterior orbital gyrus
42	Left anterior orbital gyrus	109	Right planum polare
43	Right angular gyrus	110	Left planum polare
44	Left angular gyrus	111	Right precentral gyrus
45	Right calcarine cortex	112	Left precentral gyrus
46	Left calcarine cortex	113	Right planum temporale
47	Right central operculum	114	Left planum temporale
48	Left central operculum	115	Right subcallosal area
49	Right cuneus	116	Left subcallosal area
50	Left cuneus	117	Right superior frontal gyrus
51	Right entorhinal area	118	Left superior frontal gyrus
52	Left entorhinal area	119	Right supplementary motor cortex
53	Right frontal operculum	120	Left supplementary motor cortex
54	Left frontal operculum	121	Right supramarginal gyrus
55	Right frontal pole	122	Left supramarginal gyrus
56	Left frontal pole	123	Right superior occipital gyrus
57	Right fusiform gyrus	124	Left superior occipital gyrus
58	Left fusiform gyrus	125	Right superior parietal lobule
59	Right gyrus rectus	126	Left superior parietal lobule
60	Left gyrus rectus	127	Right superior temporal gyrus
61	Right inferior occipital gyrus	128	Left superior temporal gyrus
62	Left inferior occipital gyrus	129	Right temporal pole
63	Right inferior temporal gyrus	130	Left temporal pole
64	Left inferior temporal gyrus	131	Right triangular part of the inferior frontal gyrus
65	Right lingual gyrus	132	Left triangular part of the inferior frontal gyrus
66	Left lingual gyrus	133	Right transverse temporal gyrus
67	Right lateral orbital gyrus	134	Left transverse temporal gyrus

left ventricle myocardium (LVM) and left ventricle cavity (LVC). The image sizes are $N \times N$ pixels with N being 240, 256 or 288. All images have an isotropic resolution of 1.2308mm. The end diastole and end systole frames with anatomical segmentations can be seen in Figure 4.7.

For the ADNI and OASIS brain data, we use the MNI single-subject T1-atlas [142] as common

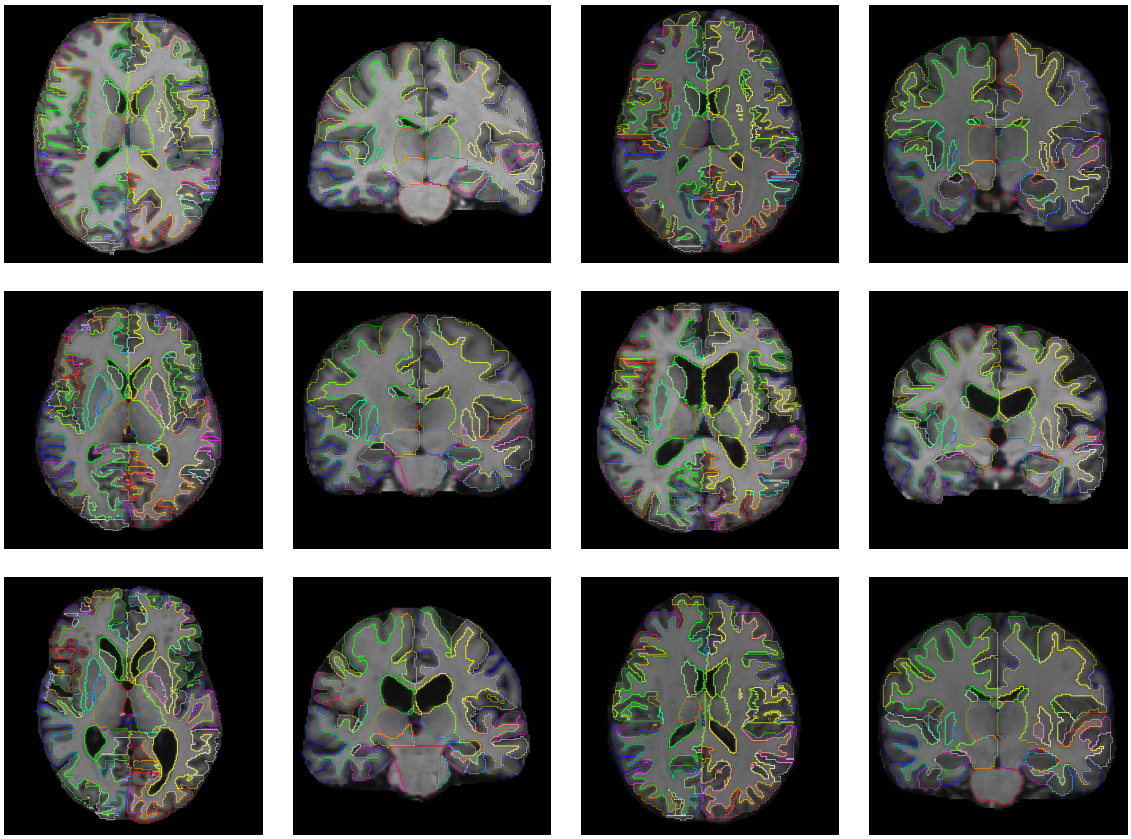


Figure 4.6: Example subjects from the OASIS dataset in axial and coronal view. Coloured contours represent the anatomical segmentations.

reference template. In the case of the cardiac data, we utilise the first frame (end diastole) of the corresponding sequence, cropped to depict only a rectangular region around the heart. All FFD registrations where performed with control point spacings of 20, 10, and 5mm. Thus, all the statistical registrations are performed with a single resolution level using a control point spacing of 5mm.

4.3.2 Visualisation of SDMs

To visualise the SDM, we show the first 3 components of SDMs trained with different methods: Figure 4.8 shows the magnitude of the components at each FFD control point (middle slice) obtained by applying the statistical approaches to the registration results of 820 brain images from the ADNI database. Figure 4.9 also shows the magnitude of the first 3 components

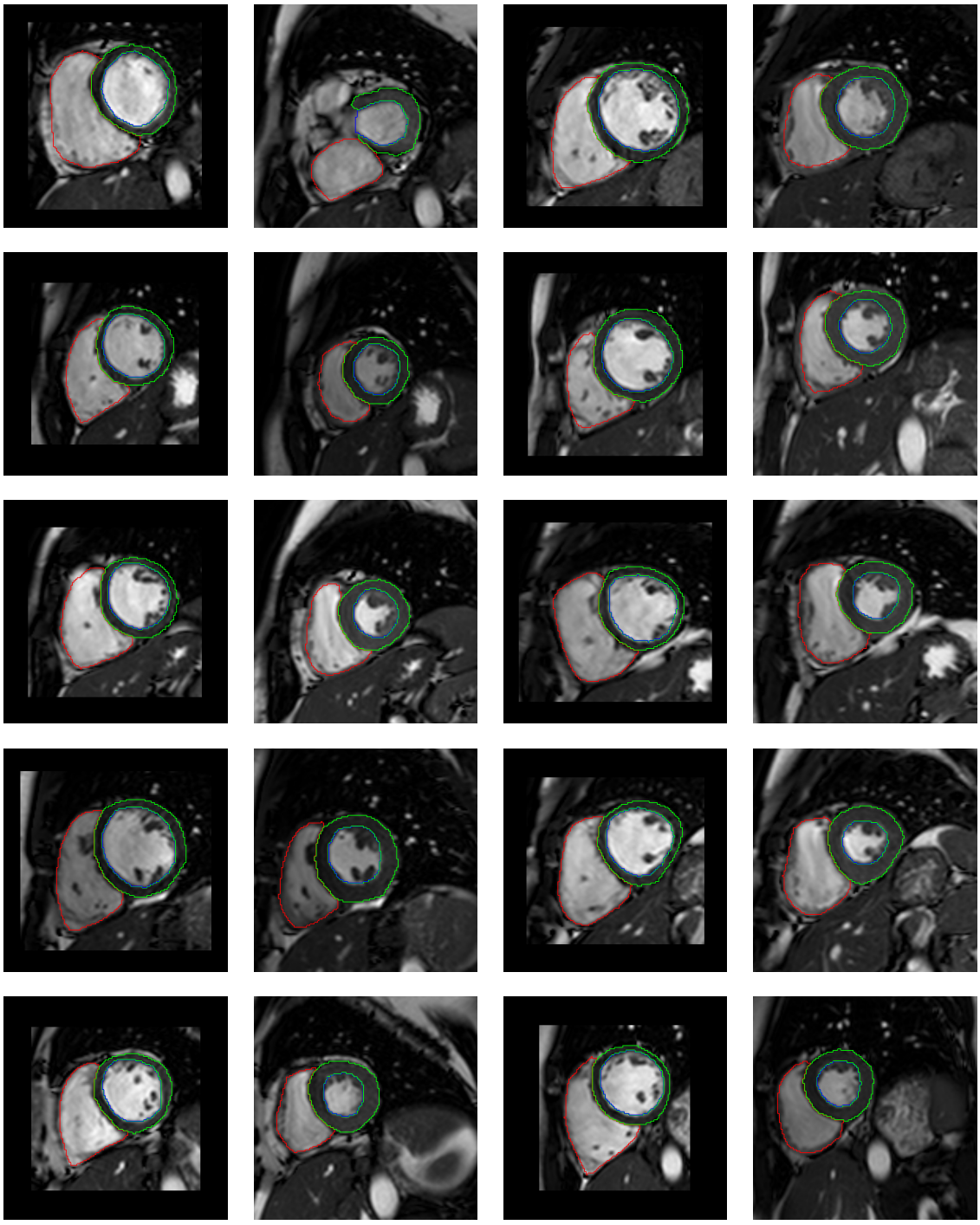


Figure 4.7: End diastole and end-systole frames of all 10 subjects from the cardiac dataset with anatomical segmentations. Coloured contours represent the anatomical segmentations. First and third columns depict end-diastole frames. Second and fourth columns depict end-systole frames

of resulting SDMs, but for registrations over a 2D cardiac sequence. Note that for the ADNI data, the first component accounts for a significant proportion of the variability in the ventricles

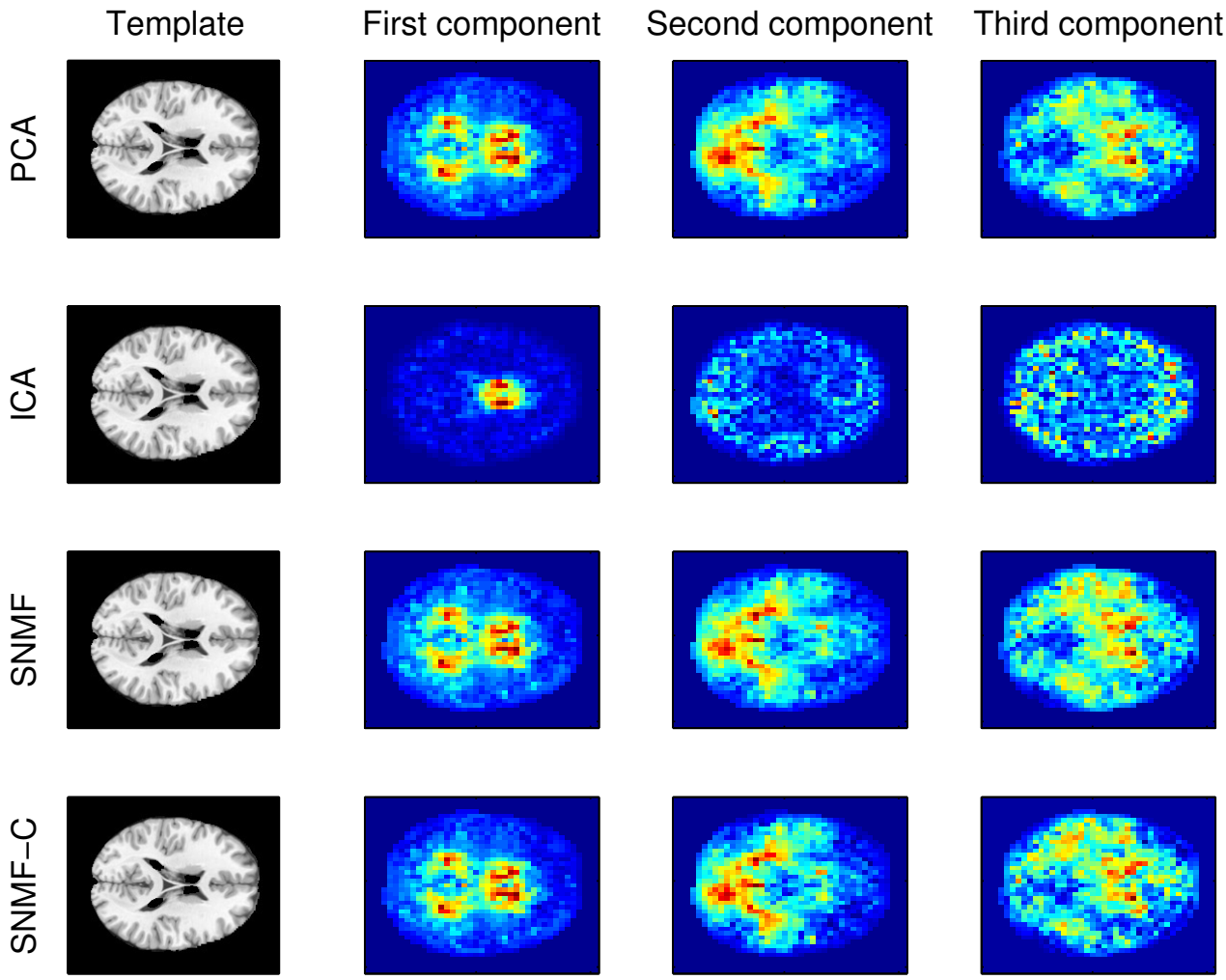


Figure 4.8: Visualisation of the first 3 components in FFD control point space obtained by different statistical training approaches using the registration results of 820 images from ADNI data (see section 4.3.1).

(which is a very important characteristic in ADNI data), while the other two components aid in the alignment of other subcortical areas. In the case of cardiac data, the first two components mainly account for radial contraction. The third component principally accounts for twisting motion.

4.3.3 Compactness of the SDM

An important feature that can be used to assess the quality of statistical deformation models is their compactness, i.e., how much of the variance of the training data can be explained as

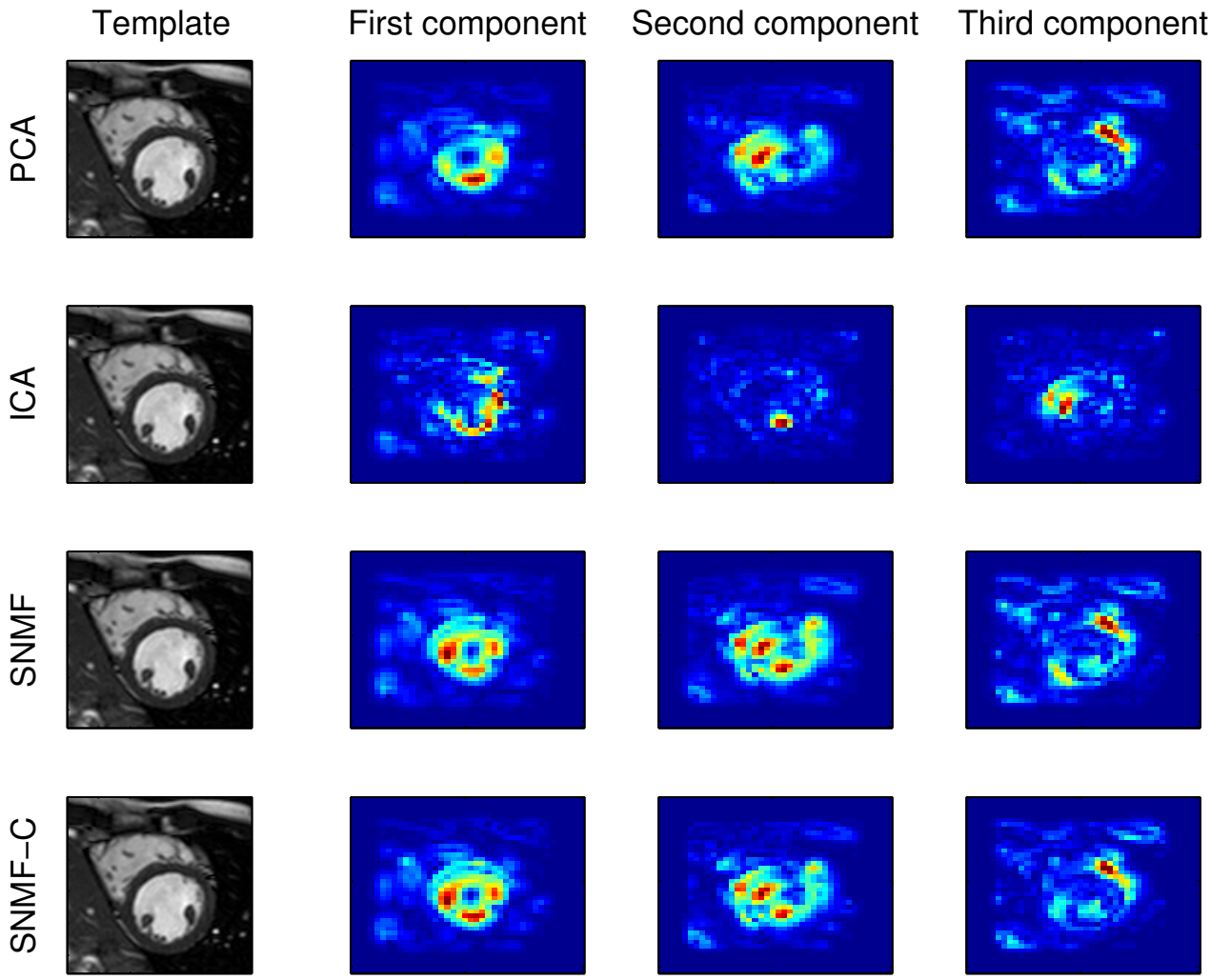


Figure 4.9: Visualisation of the first 3 components in FFD control point space obtained by different statistical training approaches using the registration results of 28 frames from a cardiac sequence (see section 4.3.1).

a function of the number of components retained in the matrix $\mathbf{B}$. We investigate this aspect using the 820 registration results of all ADNI images as training data with four different values for the landmark energy weight γ. Since of the four statistical training approaches, only PCA is designed to maximise the explained variance directly, we perform this analysis on PCA only. If the eigenvalues λ_i are sorted in descending order into a vector λ^{sort}, the percentage of explained variability V obtained by keeping C out of N principal components can be computed as

$$V = 100 \times \frac{\sum_{i=1}^{C} \lambda_i^{sort}}{\sum_{i=1}^{N} \lambda_i^{sort}}. \tag{4.16}$$

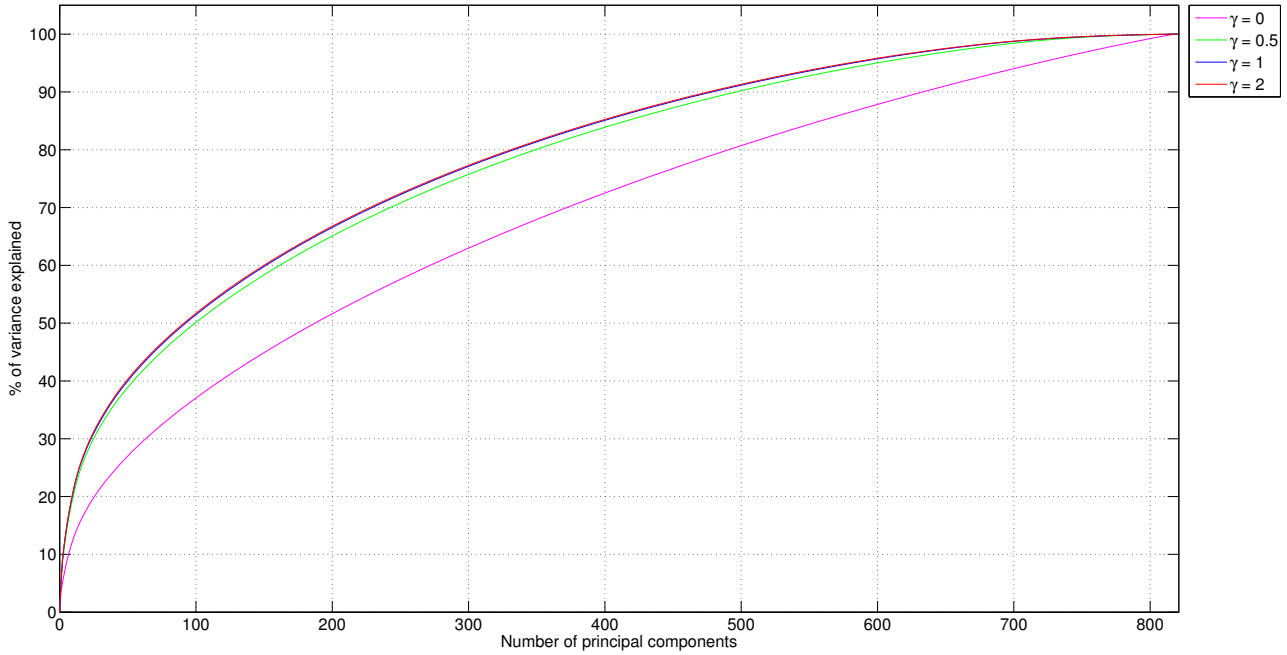


Figure 4.10: Percentage of explained variance using 820 training registrations for ADNI with different values of the landmark weight γ as a function of the number of principal components retained in PCA.

The results are shown in Figure 4.10. The first aspect that becomes apparent is that the training deformations have a great amount of variability since more than 600 components are needed to account for 95% of the variance in all cases. Moreover, even if only 70% of the variance is to be explained, more than 220 components are still needed. Another interesting observation is that the more we enforce landmark alignment, the higher the compactness of the statistical model. This means that the training registrations are less variable with increasing landmark weight. Finally, we observe that the compactness of the PCA statistical deformation model using $\gamma = 1$ and $\gamma = 2$ is very similar. We believe that this is due to the fact that using $\gamma = 1$ is enough to enforce all the landmarks to get very closely aligned. Hence, increasing the value of this parameter does not produce a different result, as there is no room for further landmark alignment.

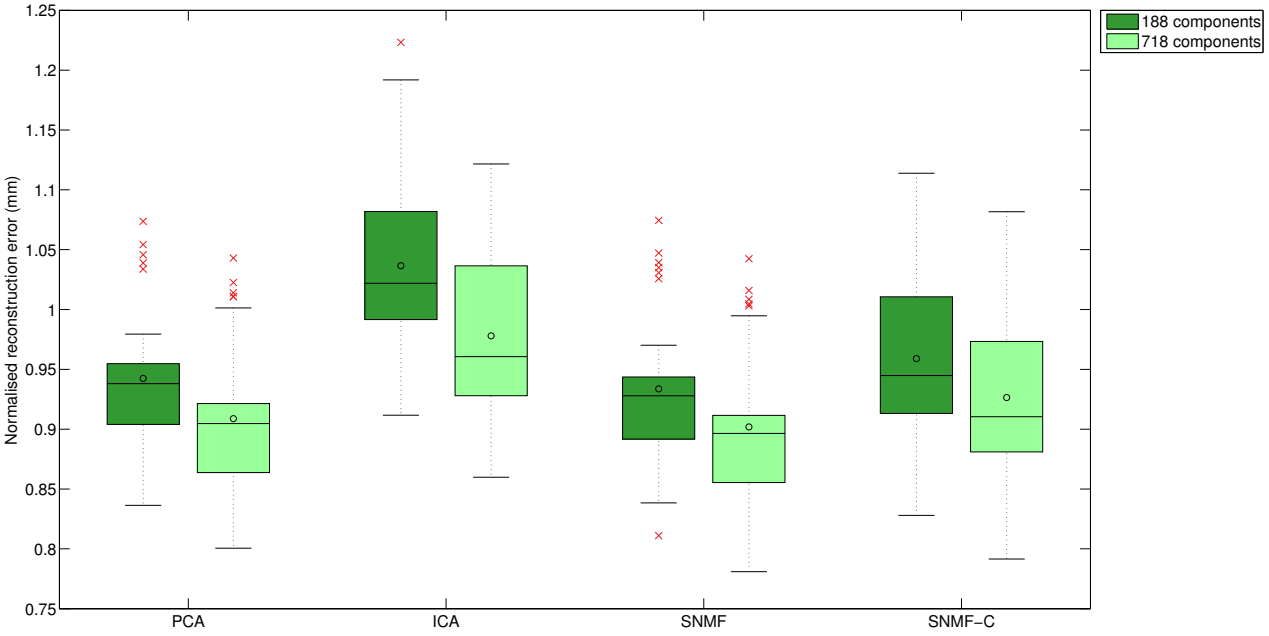


Figure 4.11: Representation error over OASIS data for each of the four statistical training approaches trained with ADNI data and using 188 and 718 components (50% and 95% of variance, respectively).

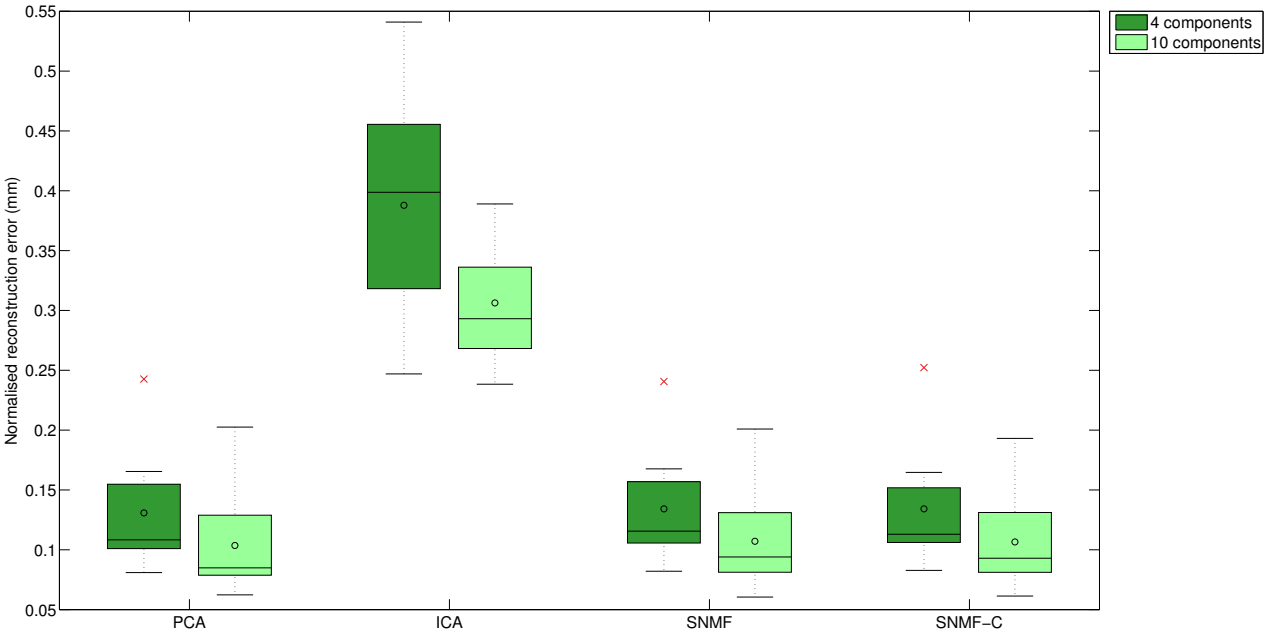


Figure 4.12: Representation error over cardiac data for each of the four statistical training approaches using 4 and 10 components (90% and 98% of variance, respectively).

4.3.4 Reconstruction capability of the SDM

Another key aspect of a statistical deformation model is its ability to represent unseen data. This representation can be obtained by projecting the high-dimensional parameterisation (control point values) of the unseen testing registration Φ_{true} into the low-dimensional parameter vector $\mathbf{w}$

$$\mathbf{w} = \mathbf{B}^{\top}(\Phi_{true} - \overline{\Phi}). \tag{4.17}$$

We can now compute the estimated representation Φ_{estim} of the unseen data by combining equations 4.13 and 4.17:

$$\Phi_{estim} = \mathbf{B}\left(\mathbf{B}^{\top}(\Phi_{true} - \overline{\Phi})\right) + \overline{\Phi}. \tag{4.18}$$

Finally, the normalised reconstruction error e of the SDM is given by

$$e = \frac{1}{P}\|\Phi_{true} - \Phi_{estim}\|_{L_1}. \tag{4.19}$$

We investigate the ability of statistical deformation models trained with ADNI brain MR images to represent testing registrations results from the OASIS brain MR image dataset, using 188 and 718 components, which account for approximately 50% and 95% of the training set variability, respectively. The results are shown in Figure 4.11. We observe that, as expected, the representation error decreases significantly for all four methods when using more components ($p < 10^{-5}$). Another observation that can be made is that ICA appears to have less reconstruction capabilities than the other statistical methods, which have similar reconstruction errors between them. We also study the reconstruction ability of statistical deformation models trained using the cardiac data. For each sequence, we train the statistical approaches using all 28 frames which do not correspond to end-systole (sources), by registering them to the end-diastole frame (target). We utilise 4 and 10 components, which account for approximately 90% and 98% of variability among the training registrations, respectively. Figure 4.12 shows these results. A similar picture to that of the brain MR images is observed: Increasing the number of components decreases the error significantly ($p = 0.002, 0.004, 0.002, 0.002$ for PCA,

ICA, SNMF and SNMF-C, respectively), and ICA appears to have much less reconstruction capabilities than the other statistical methods. What is different for the cardiac MR images is that the errors themselves are much lower, with a difference of almost an order of magnitude for PCA, SNMF and SNMF-C. This is expected due to two reasons: (1) Almost all the cardiac variablity can be explained by just 3 components, since cardiac motion is essentially described by a radial contraction, a longitudinal contraction and a twisting motion. Deformations across brain MR images, on the other hand, have a much less predictable deformation pattern. (2), the cardiac deformations are obtained from images of the cardiac cycle of the same subject and therefore represent an intra-subject registration. On the other hand, the brain deformations correspond to results of inter-subject registrations.

4.3.5 Statistical registration results

Table 4.4: Number of components utilised for each of the 5 ADNI image groups and values of the landmark weight γ in the training set.

	Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**
$\gamma = 0$	578	579	579	578	578
$\gamma = 0.5$	487	490	487	486	488
$\gamma = 1$	474	476	474	470	476
$\gamma = 2$	472	472	470	472	473

We first perform both standard and statistical FFD registration for each subject from the ADNI dataset. The subjects from the ADNI datasets are randomly separated once into 5 groups of 164 images. We use 4 groups (656 images) to learn $\mathbf{B}$ and the remaining group is used for the testing phase. Thus, each image in a group is registered using a statistical deformation model trained on the images of the other 4 groups. The number of components we retain correspond to the amount needed to account for 95% of the training data variability using PCA. Table 4.4 provides details on the number of retained components for each group and landmark weight γ in the training set.

We compare both standard and statistical FFD models by measuring the mean target registration error, i.e., how well landmarks are aligned after registration. Note that no landmark

Table 4.5: Mean (Median) target registration errors and standard deviation over all images for the proposed statistical registration method (SFFD) using different statistical techniques, trained with examples from landmark registration using different values of the landmark weight γ, compared to the standard FFD approach. Since the image have isotropic voxel size of 1mm, the errors can be considered to be either in mm or in voxels. Significant differences ($p < 10^{-5}$) of ICA, SNMF and SNMF-C with respect to PCA are marked in bold.

	PCA	**ICA**	**SNMF**	**SNMF-C**
$\gamma = 0$	$2.98(2.95) \pm 0.37$	$\mathbf{3.07(3.05) \pm 0.38}$	$\mathbf{3.04(3.01) \pm 0.41}$	$2.98(2.96) \pm 0.37$
$\gamma = 0.5$	$2.74(2.70) \pm 0.36$	$2.77(2.72) \pm 0.39$	$\mathbf{2.86(2.83) \pm 0.39}$	$2.74(2.70) \pm 0.36$
$\gamma = 1$	$2.68(2.66) \pm 0.35$	$\mathbf{2.77(2.74) \pm 0.39}$	$\mathbf{2.85(2.82) \pm 0.39}$	$2.68(2.65) \pm 0.36$
$\gamma = 2$	$2.65(2.61) \pm 0.35$	$\mathbf{2.76(2.72) \pm 0.38}$	$\mathbf{2.76(2.74) \pm 0.39}$	$\mathbf{2.64(2.61) \pm 0.35}$
FFD	$2.93(2.87) \pm 0.48$			

information is explicitly used to drive the statistical FFD registration. Table 4.5 shows the results for different values of the trade-off parameter γ. We observe than PCA and SNMF-C produce the best results, with almost no difference between the two. ICA and SNMF on the other hand have slightly lower performance than PCA. If we focus on the best overall results (PCA), it can be observed that the FFD method produces significantly more accurate landmark alignment ($p < 10^{-5}$) than statistical registration when there is no utilisation of the landmark information (i.e. $\gamma = 0$). On the contrary, increasing the value of γ does lead the statistical registration to yield significantly better landmark alignments than standard FFD ($p < 10^{-5}$ for $\gamma \in \{0.5, 1, 2\}$). Thus, our method embeds both the variability across the population and the landmark information. Note also that the standard deviation of errors also tends to decrease with increasing γ. This supports the intuition that a greater weight on the landmark penalty should tend to better align them and, hence, reduce the variability.

We also look at the individual alignments of each of the 20 landmarks for different values of the landmark energy weight γ. In this case, we only use PCA to learn the statistical deformation model. These results are shown in Figure 4.13. It can be observed that for 16 out of 20 landmarks we obtain lower registration error when using a statistical model trained with landmark alignment enforcement via the weight γ, albeit this error does not always decrease monotonically with γ. This, toghether with the results shown in table 4.5, confirms that the statistical deformation model has a good degree of learning capabilities in terms of how to

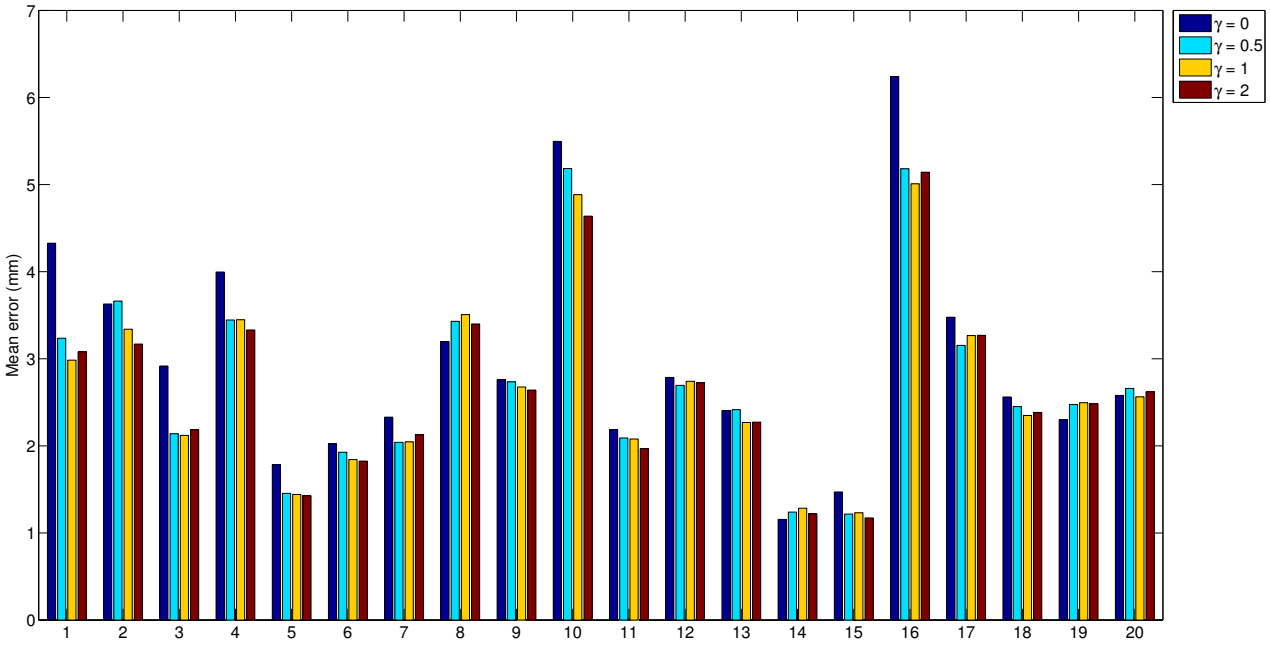


Figure 4.13: Mean errors over all ADNI images for each individual landmark using the statistical registration trained with PCA. We show results for different values of the landmark penalty weight γ

better align the landmarks.

Visual results of registration between the template image and a subject's image from the ADNI dataset are given in Figure 4.14. We utilise both FFD and the proposed statistical method with no landmark information, i.e., $\gamma = 0$. We observe that the FFD-based method appears more accurately aligned than the statistical method. However, the proposed method is capable of providing a good alignment, with the added benefit that we can also enforce landmark alignment.

We also explore the utilisation of statistical deformation models trained with ADNI data for statistical registration of images from the OASIS database. In this experiment, we first make use of the FFD registrations with no landmark information (i.e., $\gamma = 0$). To evaluate the performance of the proposed statistical registration, we use it to register all images in the OASIS data to the common template and compute all the mean pairwise similarity indices of the propagated labels for each image. We perform this experiment in two settings: In the first setting we retain 188 componens for each statistical model. This corresponds to the

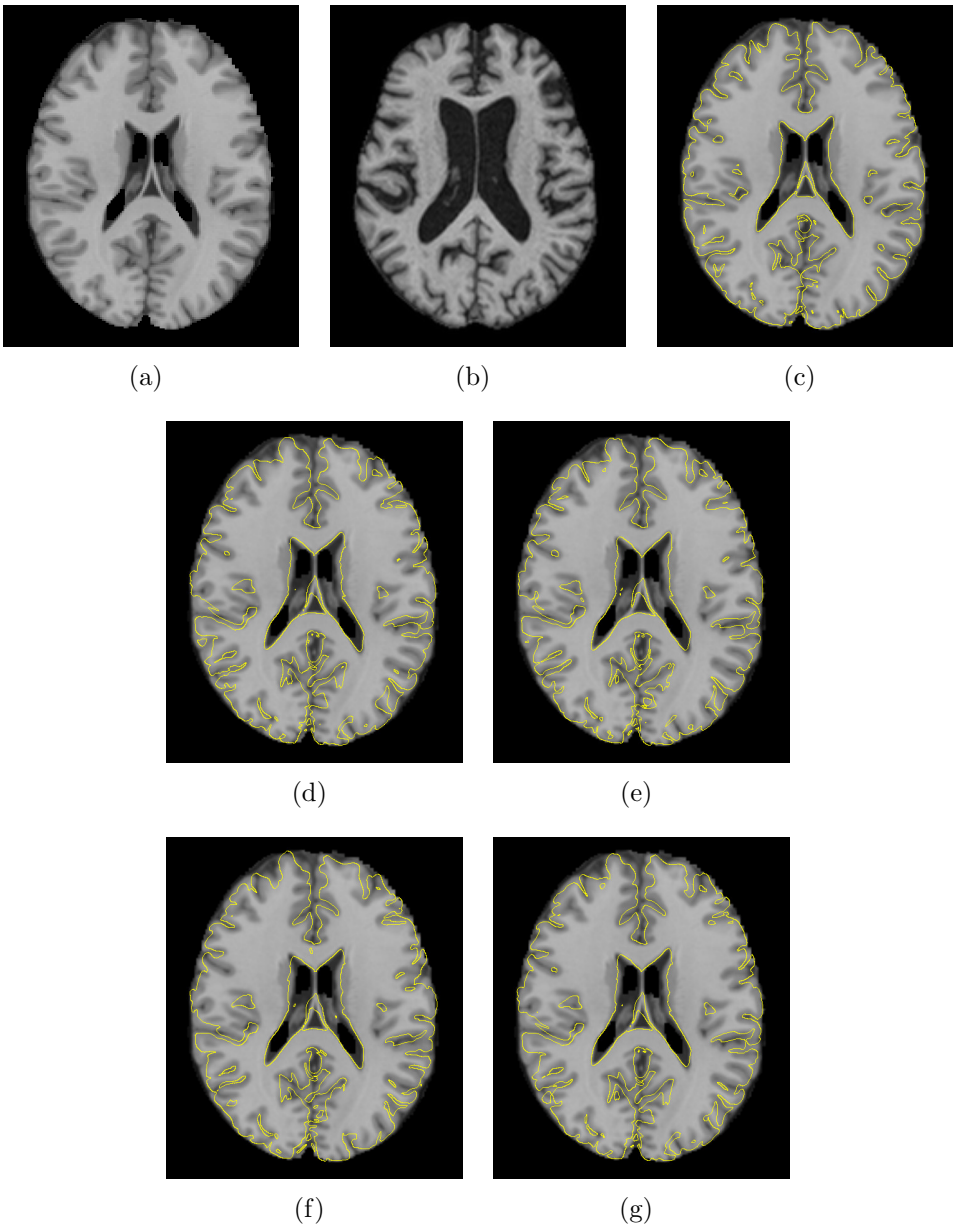


Figure 4.14: Visual comparison of the proposed statistical registration for ADNI data trained with no landmark information ($\gamma = 0$) against FFD registration (axial view). (a): Target image. (b): Source image. (c): Transformed source isolines using FFD registration result overlayed on the target image. (d): Transformed source isolines using statistical PCA registration result overlayed on the target image. (e): Transformed source isolines using statistical ICA registration result overlayed on the target image. (f): Transformed source isolines using statistical SNMF registration result overlayed on the target image. (g): Transformed source isolines using statistical SNMF-C registration result overlayed on the target image.

number of principal components that explain 50% of the variability of the deformations in the training set from ADNI. In the second setting, we utilise 718 components per statistical training approach, which correspond to 95% of the variability. Figures 4.15 and 4.16 show results for

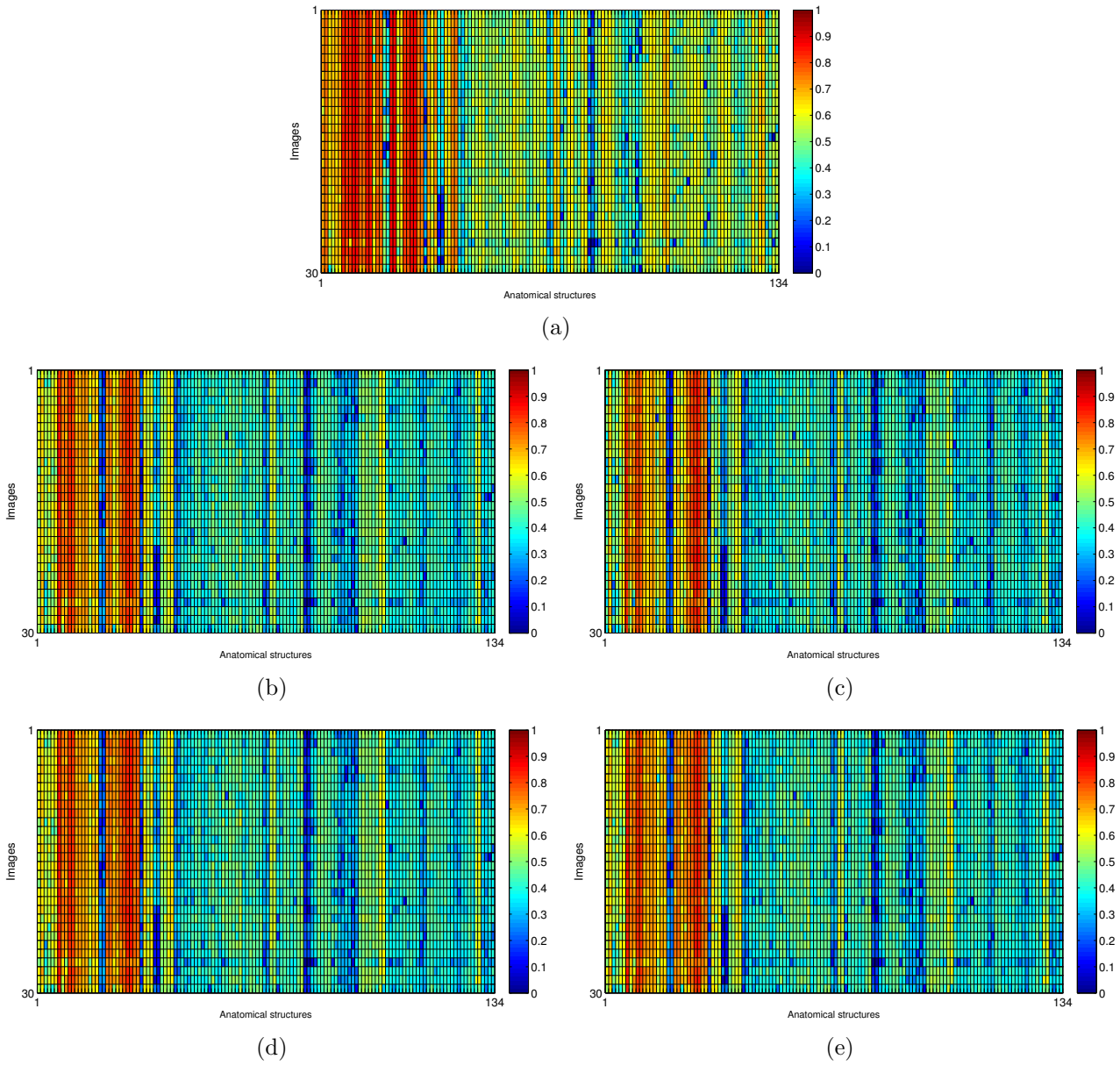


Figure 4.15: Mean pairwise similarity index per image per label for the OASIS data using 188 components (i.e., 50% of ADNI's results' variance explained by PCA), compared to standard FFD. (a) Standard FFD (b) Trained with PCA. (c) Trained with ICA. (d) Trained with SNMF. (e) Trained with SNMF-C.

the first and second settings, respectively. Each row corresponds to one image and each column correspond to one of the 134 structure labels. It is possible to observe that for some structures the performance is very poor for both standard FFD and statistical registrations. This is due to these structures being relatively small and hard to segment (e.g. Right Inf Lateral Ventricle (ID 19) and Right Inf Lateral Ventricle (ID 20)), or due to the structure being isolated and not well

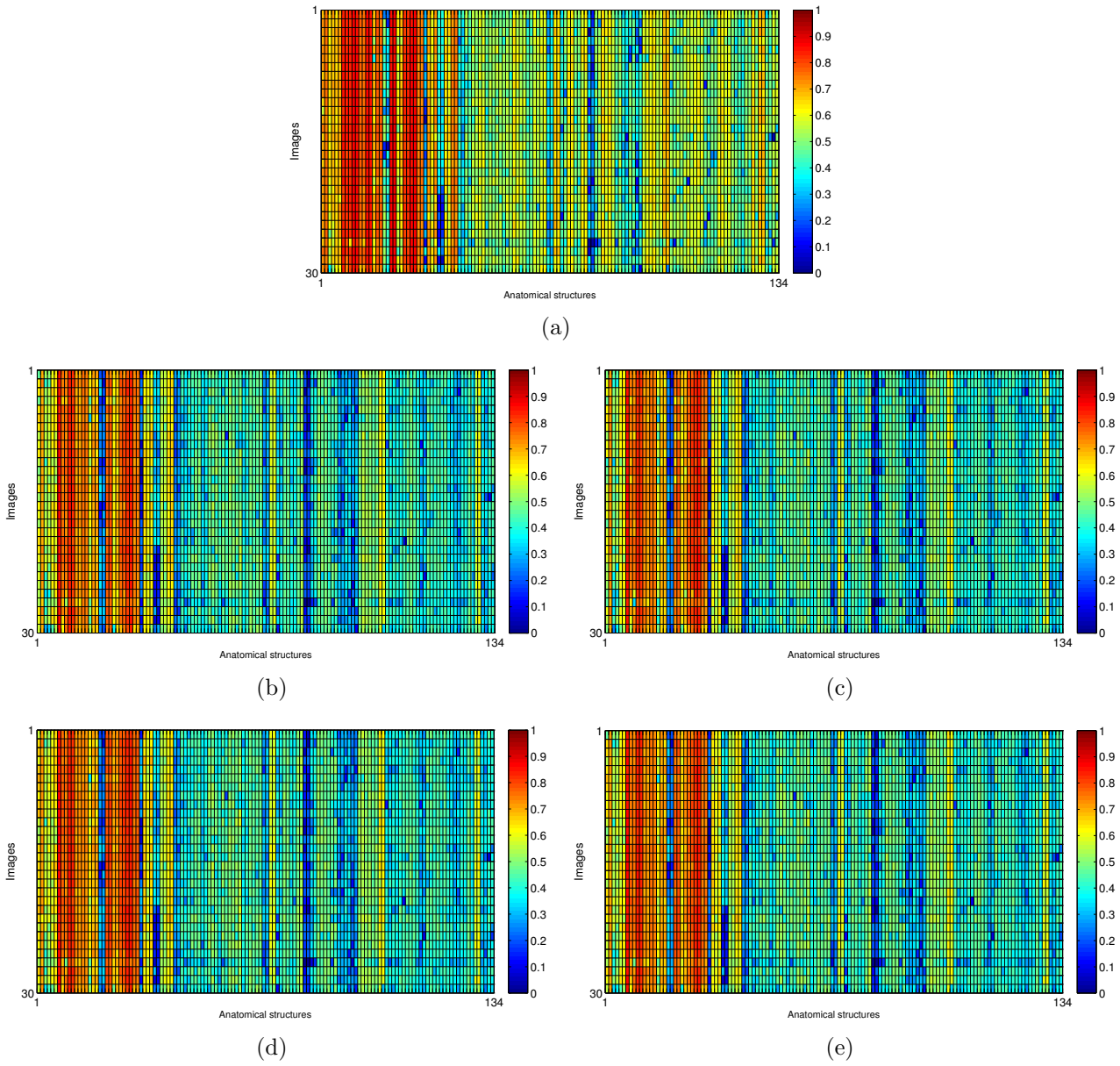


Figure 4.16: Mean pairwise similarity index per image per label for the OASIS data using 718 components (i.e., 95% of ADNI's results' variance explained by PCA), compared to standard FFD. (a) Standard FFD (b) Trained with PCA. (c) Trained with ICA. (d) Trained with SNMF. (e) Trained with SNMF-C.

defined by intensity features, like the Optic Chiasm (ID 31). Another important observation is that, when compared to the standard FFD registration, the proposed method have comparable results in subcortical structures (IDs 1 to 36). However, it performs considerably worse in cortical areas (IDs 37 to 134). This can be explained by the fact that the cortical anatomy is extremely variable across different subjects and therefore very hard to match. This means that

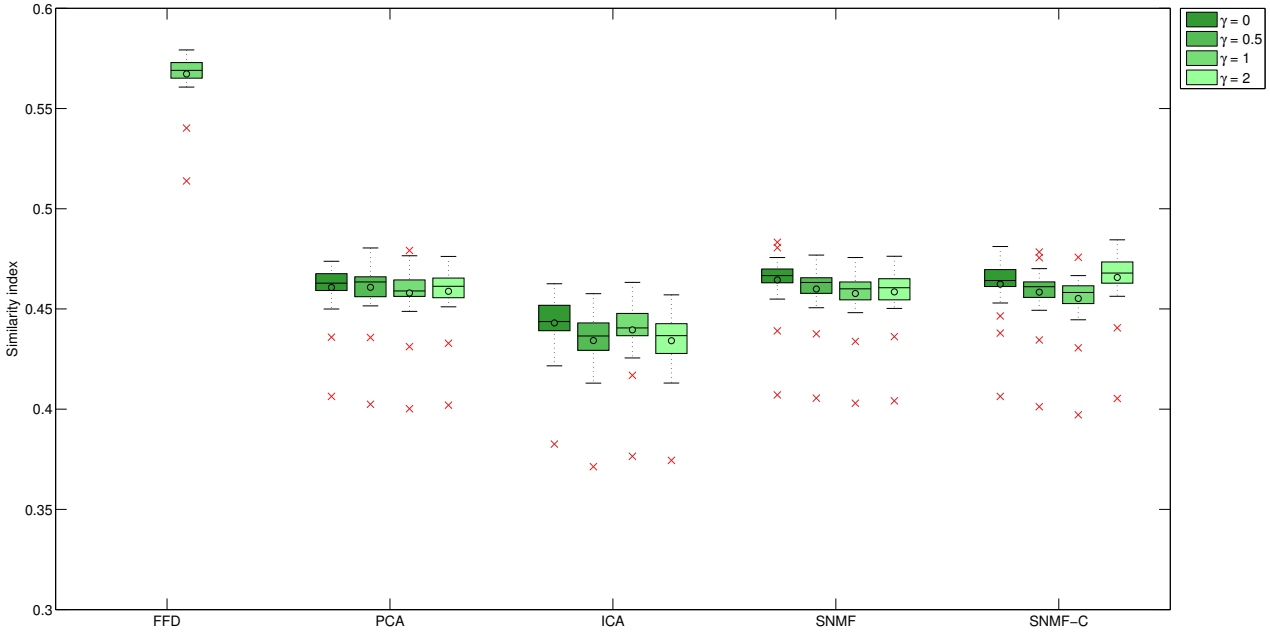


Figure 4.17: Mean pairwise similarity index for each statistical training approach using statistical registrations on the OASIS data trained keeping 188 components and employing different values of the landmark weight γ. FFD result is also given for comparison.

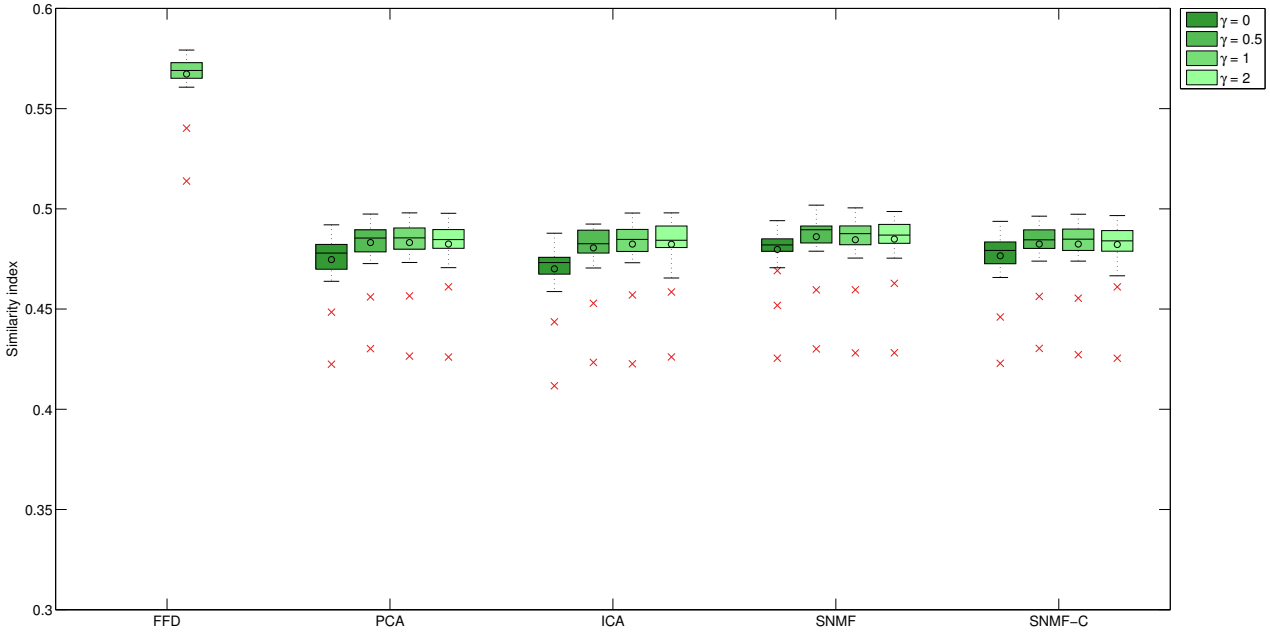


Figure 4.18: Mean pairwise similarity index for each statistical training approach using statistical registrations on the OASIS data trained keeping 718 components and employing different values of the landmark weight γ. FFD result is also given for comparison.

the trained statistical deformation models are not able to represent the deformation pattern in those areas in such a way that it could be explained with a limited number of parameters.

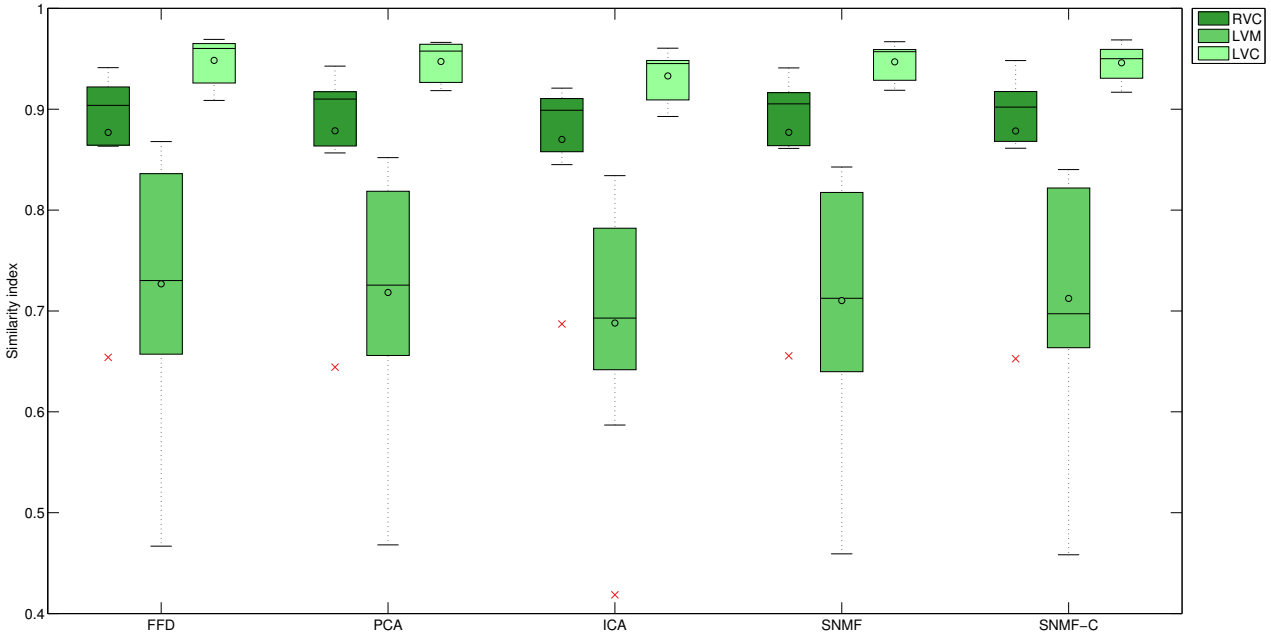


Figure 4.19: Results on cardiac data measuring Right ventricle cavity (RVC), Left ventricle myocardium (LVM) and Left ventricle cavity (LVC) similarity indices for each of the four statistical training approaches using 4 components. FFD result is also given for comparison.

We perform the previous experiment a second time, but with different values of the landmark energy weight γ. The results for the two settings (with 188 and with 718 components) are illustrated in Figures 4.17 and 4.18, respectively. Of the four statistical training approaches, ICA performs worse than the other appoaches in both settings when no landmark information is encoded in the SDMs. The remaining methods performing very similarly, although with relatively poor overall results when compared to FFD. This is mainly due to the fact that cortical areas are not well registered. This means that for 98 out of the 134 labels, we obtained similarity indices of around 0.5 or less, which greatly biases the evaluation. We also observe that the number of components used for the statistical registration does have an important impact, with a difference of more than 1.5% between the two settings. Moreover, in the first setting (188 components), the landmark penalty weight does not have great influence on the final similarity index of the results. We believe that this is because using less components means that there is much less landmark information encoded in the model. In contrast to this, in the second setting (718 components), a much stronger influence of enforcing landmark alignment can be observed in the final similarity indices.

The effectiveness of the proposed framework in cardiac MR images is also investigated. For each sequence, we train the statistical approaches using all 28 frames which do not correspond to end-systole (sources), by registering them to the end-diastole frame (target). We keep 4 components on all experiments, which account for approximately 90% of the variability. Using the resulting statistical deformation models for each subject, we register the end-systole frame to the end-diastole frame using the proposed registration method and measure the similarity index after propagating the three cardiac labels. Figure 4.19 shows the overlap results across the 10 subjects. We observe results that are very comparable to FFD for all three labels. Additionally, all four statistical methods perform similarly well in the left and right cavity areas. However, they tend to have more difficulty aligning the myocardium, albeit this is also true for FFD. Comparatively, PCA is the statistical method that performs the best and, as in the case of brain MR images, ICA performs worst. We also perform the same experiment using 10 components. Nevertheless the improvement in the reconstruction capability by using additional components (Figure 4.12) was not sufficient to produce any noticeable difference in the results.

A visual example is given in Figure 4.20 using a cardiac MR image of a subject with noticeable difference between the end systole and end diastole frames. We observe that the results for FFD and statistical registration have no important visual difference, except for the proposed method trained with ICA. These results are in concordance with the fact that ICA has a much poorer representation capability than PCA, SNMF and SNMF-C.

4.3.6 Runtime

Finally, we measured the runtimes of the FFD registration and proposed registration trained with PCA for all 30 OASIS images. Statistics of these runtimes are summarised in table 4.6. All registrations were performed using a multithreaded implementation of the algorithms on a standard PC desktop with 8 cores (3.40GHz) and 16GB of RAM. We observe that the proposed statistical method is up to 6 times faster than the original FFD implementation. This

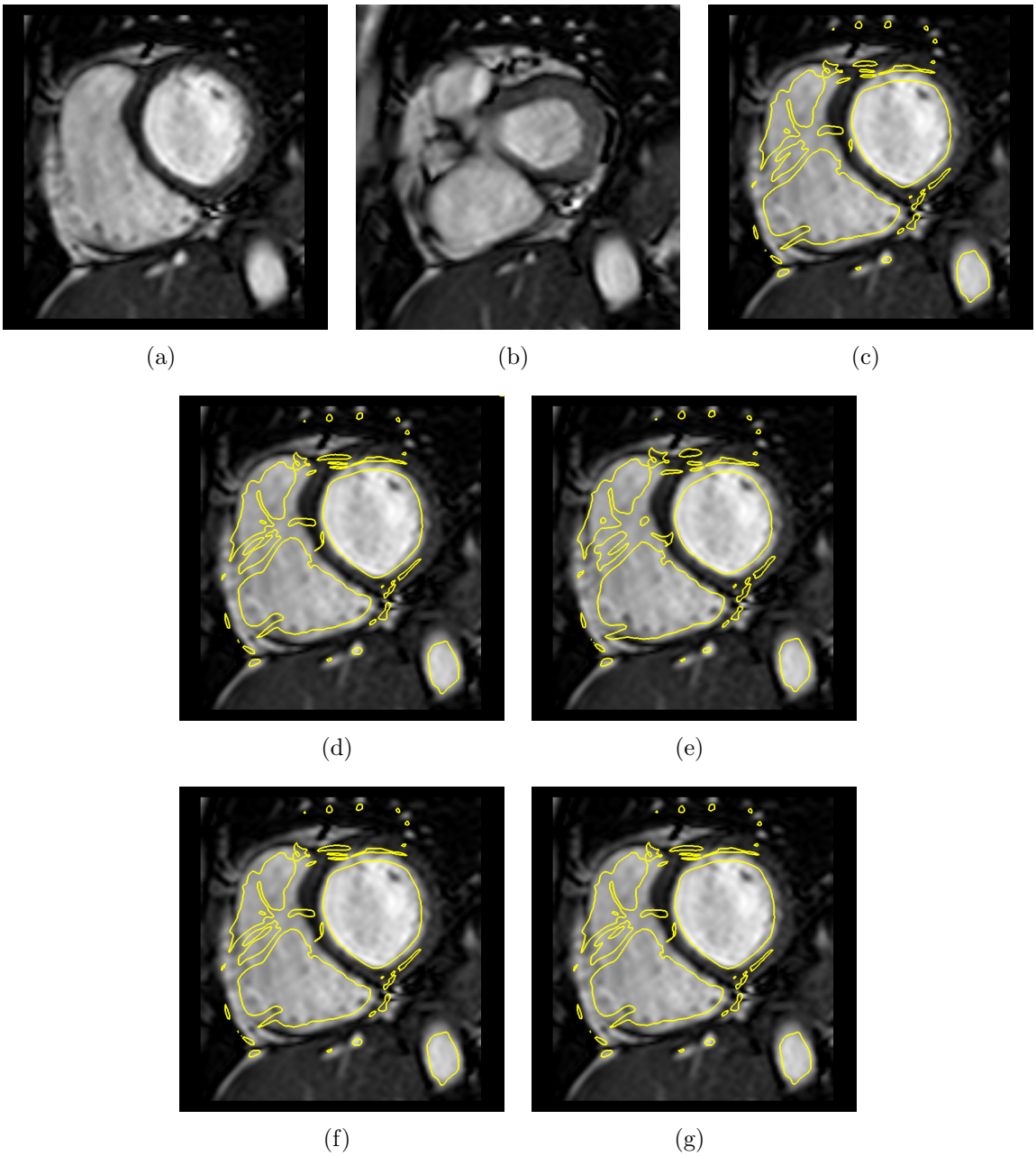


Figure 4.20: Visual comparison of the proposed statistical registration for one of the cardiac sequences against FFD registration. (a): Target image. (b): Source image. (c): Transformed source isolines using FFD registration result overlayed on the target image. (d): Transformed source isolines using statistical PCA registration result overlayed on the target image. (e): Transformed source isolines using statistical ICA registration result overlayed on the target image. (f): Transformed source isolines using statistical SNMF registration result overlayed on the target image. (g): Transformed source isolines using statistical SNMF-C registration result overlayed on the target image.

is due to two reasons: Firstly, the statistical registration is performed using only one resolution level, since the SDM encodes the whole multiresolution pyramid of the training registrations.

Table 4.6: Mean (Median) registration runtimes and standard deviation over all 30 registrations for the standard FFD method and the proposed statistical FFD method trained with PCA for different values of γ.

Method	**Runtime (seconds)**
Original FFD	$344.7(308.9) \pm 158.7$
Statistical FFD ($\gamma = 0$)	$81.4(53.9) \pm 41.3$
Statistical FFD ($\gamma = 0.5$)	$234.4(218.6) \pm 54.9$
Statistical FFD ($\gamma = 1$)	$67.7(59.8) \pm 44.3$
Statistical FFD ($\gamma = 2$)	$54.0(49.6) \pm 7.3$

Secondly, in our proposed statistical registration the number of degrees of freedom to optimise is substantially less than that of the standard FFD, allowing the registration to converge with fewer iterations.

4.4 Summary

In this chapter, we introduced a registration approach based on statistical deformation models learnt from a population of registered subjects. These models are employed as prior information knowledge to guide the alignment of a new subject to a common reference template. Furthermore, the learning stage of our framework can take advantage of annotated information to enforce the alignment of certain anatomical structures. We focused on annotations of anatomical landmark, although other types of annotations such as segmentations in the form of label maps can be used as well. If a set of deformations registering a group of subjects to a common template is already available, any subsequent registration of an unseen image to that template space can be performed via the proposed framework. Our results suggest that the proposed statistical registration method can provide registration results comparable to the standard FFD-based approach, especially for cardiac MR images. On the other hand, when using brain data, it is important to consider the fact that the set of subjects employed to train the statistical model might be very variable, which can have a significant impact in the performance of a statistical registration. It is important to mention, however, that this problem can be alleviated by using the proposed approach as an initialisation to a standard

FFD approach, as proposed by Onofrey et al. [157]. The number of training samples is also a crucial factor to consider when designing a statistical registration approach. This number has to be large enough in order to allow the statistical model to learn sufficient information to drive the registration. This is especially true when the variability of training data is large. Finally, a relevant feature of the proposed framework is that it has a much lower computational cost than standard FFD. It also uses substantially less degrees of freedom, and also constrains the set of possible transformations to be in statistical concordance with the training data. In conclusion, the proposed registration framework is particularly useful when the underlying transformations can be characterised by a reduced number of degrees of freedom, making the use of complex high-dimensional models such as standard B-spline FFD unjustified. In the next chapter, we further explore the learning capabilities of PCA in order to devise a registration method that is able to improve the convergence of the optimisation of any chosen parametric registration method.

Chapter 5

Gradient projection learning for parametric non-rigid image registration

This chapter is based on:

- Stefan Pszczolkowski, Luis Pizarro, Declan P. O'Regan, and Daniel Rueckert. "Gradient Projection Learning for Parametric Nonrigid Registration." In Machine Learning in Medical Imaging, pp. 226-233. Springer Berlin Heidelberg, 2012.

5.1 Introduction

The large anatomical variability across subjects in a population makes non-rigid image registration techniques prone to inaccuracies. Moreover, the cost functions typically used for non-rigid alignment are highly non-convex, which increases the chances that the optimisation of the registration may be trapped in local minima. In this chapter, we propose a new learning-based estimate of the gradient to improve the convergence of the optimisation of any chosen parametric energy-based image registration method. This computed gradient is shown to yield better registration accuracy than standard analytical methods, and aids the optimisation method to potentially avoid early local minima. From a set of training images and their corresponding deformations, the proposed method learns offline a projection from the gradient of the similarity measure in voxel space to the parameter space of the chosen registration method using PCA dimensionality reduction and standard least squares regression. The learnt gradient is subsequently used online to approximate the optimisation of the energy functional for unseen images. We employ the B-spline FFD approach [183] as underlying registration method, but any other parametric method can be used instead. We perform experiments on 3D brain MRI data and $2D + t$ cardiac MRI sequences to demonstrate that our approach improves the accuracy of the chosen registration method.

5.1.1 Related work

Learning-based image registration techniques have captured the interest of many researchers in the last few years. A popular approach is to capture the statistics of deformations by applying PCA over each band of wavelet coefficients [234] or over the control point values of B-splines that provide a parametric representation of the deformation fields [131, 182, 206, 170]. In [212], the parameters of the deformation are estimated by a nearest neighbour search over training images that have been generated according to a special criterion that breaks the curse of dimensionality. A low dimensional representation of images, with maximally discriminative power is

obtained in [22] by combining generative and discriminative objective functions in a constrained optimisation problem. The work in [231] presents a method where features are extracted from regions obtained by adaptively partitioning brain images. The statistics of deformation fields are used to robustly place the control points that parameterise the deformations. In [1], a partial least squares approach is employed to relate cardiac deformation due to respiration with surface intensity traces. Finally, in [111], support vector regression is utilised to estimate the principal modes of a deformation model of the brain given low dimensionality image features.

One of the main issues of image registration techniques is their high computational cost. Recently, a new type of methods for improving optimisation convergence have been developed, albeit these methods are not learning-based. These type of schemes are the so-called *preconditioning schemes*, where the image gradient is scaled differently for different areas of the image. The main contributions on this kind of approaches are [245, 113]. The difference between these two approaches is that in [113], the preconditioner is thought to work specificly for sum of squared differences, while the precontitioning scheme by [245] works for any similarity measure.

5.2 Methods

5.2.1 Gradient projections

Non-parametric methods, e.g. [96, 18, 36, 211, 221], estimate a dense (voxel-wise) displacement field $\mathbf{u}$ that aligns the target and source images. This flow $\mathbf{u}$ is obtained as the solution of an energy functional. As previously mentioned (see Section 2.3.1), a standard gradient-based scheme updates the solution with a step size $\eta > 0$,

$$\mathbf{u}_{t+1} = \mathbf{u}_t - \eta_t \cdot \mathbf{g}_t(\mathbf{u}_t). \tag{5.1}$$

Here, t corresponds to the iteration number and $\mathbf{g}(\mathbf{u})$ is the gradient of the energy functional of equation (2.54). When using steepest gradient descent as the optimisation method, $\mathbf{g}(\mathbf{u})$ corresponds to

$$\mathbf{g}(\mathbf{u}) = -\frac{\partial E}{\partial \mathbf{u}} = -\frac{\partial E_{sim}}{\partial \mathbf{u}} - \alpha \frac{\partial E_{reg}}{\partial \mathbf{u}}. \tag{5.2}$$

The gradients are computed according to the specific choice of similarity and regularisation terms. For our later developments, we call the term $\frac{\partial E_{sim}}{\partial \mathbf{u}}$ *similarity gradient image* (SGI).

In parametric registration approaches, such as the B-spline based free-form deformation (FFD) registration algorithm [183], the unknown deformation field is parameterised by $N = n_x \times n_y \times n_z$ control points. In order optimise the parameters (control point values) Φ^ξ, with $\xi \in \{x, y, z\}$, it is necessary to compute the energy gradient in parametric space rather than voxel space. Thus, the energy gradient is calculated with respect to the control point values, by taking the SGI and regularisation terms, and projecting them from voxel space to parameter space by means of the chain rule:

$$\mathbf{g}^\xi(\mathbf{u}^\xi) = -\frac{\partial E}{\partial \Phi^\xi} = -\frac{\partial E_{sim}}{\partial \mathbf{u}^\xi} \cdot \frac{\partial \mathbf{u}^\xi}{\partial \Phi^\xi} - \alpha \frac{\partial E_{reg}}{\partial \mathbf{u}^\xi} \cdot \frac{\partial \mathbf{u}^\xi}{\partial \Phi^\xi}. \tag{5.3}$$

For the B-spline FFD approach, the projection term $\frac{\partial \mathbf{u}^\xi}{\partial \Phi^\xi}$ is given by the tensor product of the 1D cubic B-splines

$$\frac{\partial \mathbf{u}^\xi}{\partial \Phi^\xi_i} = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \tag{5.4}$$

where u, v, w correspond to relative positions in control point space. Finally, the update step for the transformation parameters is given by

$$\Phi^\xi_{i_{t+1}} = \Phi^\xi_{i_t} - \eta_t \cdot \frac{\partial E}{\partial \Phi^\xi_{i_t}}. \tag{5.5}$$

5.2.2 Learning the projection

We introduce a learning-based method (LB-FFD) in order to project the SGI from the gradient space of the energy functional to parameter space, in such a way that it avoids local minima. In our setting, the projection is not constrained to be computed using the B-spline tensor product. Instead, it can be estimated as follows.

Given M training SGIs based on any similarity metric (e.g. normalised mutual information (NMI), sum of squared differences (SSD) or normalised cross correlation (NCC)), the first step of our method is to extract patches $\mathbf{p}_{i,j}^{\xi} \in \mathbb{R}^{s_x \cdot s_y \cdot s_z \times 1}$ from all control point locations $i = 1 \dots N$ and training SGIs $\left(\frac{\partial E_{sim}}{\partial \mathbf{u}}\right)_j$, $j = 1 \dots M$:

$$\mathbf{p}_{i,j}^{\xi} = \mathbb{P}_i^s \left(\frac{\partial E_{sim}}{\partial \mathbf{u}}\right)_j \tag{5.6}$$

where $\mathbb{P}_i^s$ is an operator that extracts a patch of size $s = s_x \cdot s_y \cdot s_z$ centered on the location of the control point i in the ξ component of the SGI. The motivation of using these patches comes from the fact that the B-spline tensor model has local support, i.e., only the voxels of the SGI within a neighbourhood of a control point have to be considered to perform the corresponding projections. The patch sizes in each direction s_x, s_y and s_z may take any arbitrary value. In this work, we choose them such that they cover the support region of the B-spline basis centered at each control point.

Since the patches $\mathbf{p}_{i,j}^{\xi}$ are of high dimensionality, a PCA dimensionality reduction step is performed for each control point separately, using the M patches centered on that control point. More specifically, for each control point i, we take all the patches $\mathbf{p}_{i,j}^{\xi}$, $j = 1 \cdots M$ in the training set that are centered on it, and assemble them into a matrix $\mathbf{P}^{\xi} = [\mathbf{p}_{i,1}^{\xi} | \cdots | \mathbf{p}_{i,M}^{\xi}]$. We then compute the mean patch $\overline{\mathbf{p}_i^{\xi}} = \sum_{j=1}^{M} \mathbf{p}_{i,j}^{\xi}$ and the covariance matrices $\mathbf{S}_i^{\xi}$

$$\mathbf{S}_i^{\xi} = \frac{1}{M-1} \sum_{j=1}^{M} (\mathbf{p}_{i,j}^{\xi} - \overline{\mathbf{p}_i^{\xi}})(\mathbf{p}_{i,j}^{\xi} - \overline{\mathbf{p}_i^{\xi}})^{\top}. \tag{5.7}$$

An eigendecomposition of $\mathbf{S}_i^{\xi}$ is subsequently performed, yielding matrices $\mathbf{B}_i^{\xi} \in \mathbb{R}^{s_x \cdot s_y \cdot s_z \times C}$, which contain the C eigenvectors of $\mathbf{S}_i^{\xi}$ with the highest eigenvalues. From this point we can compute low dimensionality patches $\mathbf{p}_{i,j}^{\xi^{low}} \in \mathbb{R}^{C \times 1}$ by projecting the original patches $\mathbf{p}_{i,j}^{\xi}$ into the low dimensional space using the PCA projection rule:

$$\mathbf{p}_{i,j}^{\xi^{low}} = \mathbf{B}_i^{\xi^{\top}} \cdot (\mathbf{p}_{i,j}^{\xi} - \overline{\mathbf{p}_i^{\xi}}) \tag{5.8}$$

Finally, for each training SGI, a non-rigid FFD registration is performed between the target and source images that define it, yielding "optimal" training FFD control point values $\hat{\phi}_{i,j}^{\xi}$. We regard $\hat{\phi}_{i,j}^{\xi}$ as a good approximation of the direction of the similarity gradient term $\frac{\partial E_{sim}}{\partial \Phi_i^{\xi}}$, plus an error ε. This means that

$$\begin{aligned} \hat{\phi}_{i,j}^{\xi} &\approx \kappa \cdot \tfrac{\partial E_{sim}}{\partial \Phi_i^{\xi}} + \varepsilon \\ &\approx \kappa \cdot \tfrac{\partial E_{sim}}{\partial \mathbf{u}} \tfrac{\partial \mathbf{u}}{\partial \Phi_i^{\xi}} + \varepsilon. \end{aligned} \tag{5.9}$$

Here, κ is a positive constant. Given that we have the training FFD control point values $\hat{\phi}_{i,j}^{\xi}$ and the low dimensionality SGI patches representing $\frac{\partial E_{sim}}{\partial \mathbf{u}}$, we can learn the gradient projections $\hat{\beta}_i^{\xi} \approx \kappa \frac{\partial \mathbf{u}}{\partial \Phi_i^{\xi}}$ using standard least squares regression over each of the x-, y- and z- components of the training control point values separately

$$\hat{\beta}_i^{\xi} = (\mathbf{X}^{\top}\mathbf{X})^{-1}\mathbf{X}^{\top}\hat{\phi}_{i,1\cdots M}^{\xi}. \tag{5.10}$$

Here, $\mathbf{X} \in \mathbb{R}^{M \times C+1}$ is a matrix containing ones in the first column and the low dimensionality patches spanning the C remaining columns in each row. To this end, the learnt coefficients $\hat{\beta}_i^{\xi}$ satisfy

$$\hat{\phi}_{i,j}^{\xi} \approx \left(\hat{\beta}_i^{\xi}\right)_0 + \sum_{k=1}^{C} \left(\hat{\beta}_i^{\xi}\right)_k \cdot \left(\mathbf{p}_{i,j}^{low}\right)_k. \tag{5.11}$$

An important disadvantage of the previously described approach for learning the projection is that is necessary to perform PCA for each control point on each SGI image during training. This is very computationally expensive, as the total number of control points can reach several

millions just by using a few dozen training SGIs in 3D (the number of control points in 2D would be considerably less). To alleviate this problem, we propose to learn the projection by performing only three PCA operations, each one over the x-, y- and z-components of all possible SGI patches combined. We then regress all x-, y- and z-related low dimensionality patches with the x-, y- and z-components of their respective training control point value. Hence, only three regression calculations are needed. To account for this new setting, equation (5.11) can be rewritten as

$$\hat{\phi}_l^{\xi} \approx \left(\hat{\beta}^{\xi}\right)_0 + \sum_{k=1}^{C} \left(\hat{\beta}^{\xi}\right)_k \cdot \left(\mathbf{p}_l^{low}\right)_k^{\xi}, \qquad l = 1 \cdots M \times N. \tag{5.12}$$

5.2.3 Registration of unseen images

Once the learning procedure described in the previous section is performed, it is possible to register an unseen image to any of the target images used to produce the training SGIs, e.g. the ones depicted in Figure 5.1. For this purpose, we devise an optimisation scheme similar to (5.5) which uses a mixture of our learning-based gradient and the standard analytical gradient of equation (5.3)

$$\begin{aligned} \Phi_{t+1}^{\xi} &= \Phi_t^{\xi} + \eta \cdot \left[(1-\omega) \cdot (\mathbf{g}_{\text{ffd}})_t^{\xi} + \omega \cdot (\mathbf{g}_{\text{lb}})_t^{\xi}\right] \\ (\mathbf{g}_{\text{lb}})_t^{\xi} &:= \left(\hat{\beta}^{\xi}\right)_0 + \textstyle\sum_{k=1}^{C} \left(\hat{\beta}^{\xi}\right)_k \cdot (\mathbf{p}_l^{low^*})_k^{\xi} \\ (\mathbf{g}_{\text{ffd}})_t^{\xi} &:= \tfrac{\partial E}{\partial \Phi^{\xi}}. \end{aligned} \tag{5.13}$$

Here, $0 \leq \omega \leq 1$ is a parameter that controls the gradient mixture. The unseen low dimensional patch $(\mathbf{p}_l^{low^*})_k^{\xi}$ is computed by projecting the corresponding high dimensionality patch $(\mathbf{p}^*)^{\xi}$, taken from the SGI between the target and the current transformed source, using the PCA projection rule

$$(\mathbf{p}_l^{low^*})_k^{\xi} = \mathbf{B}^{\xi^{\top}} \cdot \left((\mathbf{p}^*)^{\xi} - \overline{\mathbf{p}^{\xi}}\right). \tag{5.14}$$

Note that $\overline{\mathbf{p}^{\xi}}$ and $\mathbf{B}^{\xi}$ are computed analogously to $\overline{\mathbf{p}_i^{\xi}}$ and $\mathbf{B}_i^{\xi}$. The only difference is that they are obtained using all the training patches taken from every control point. A summary of the

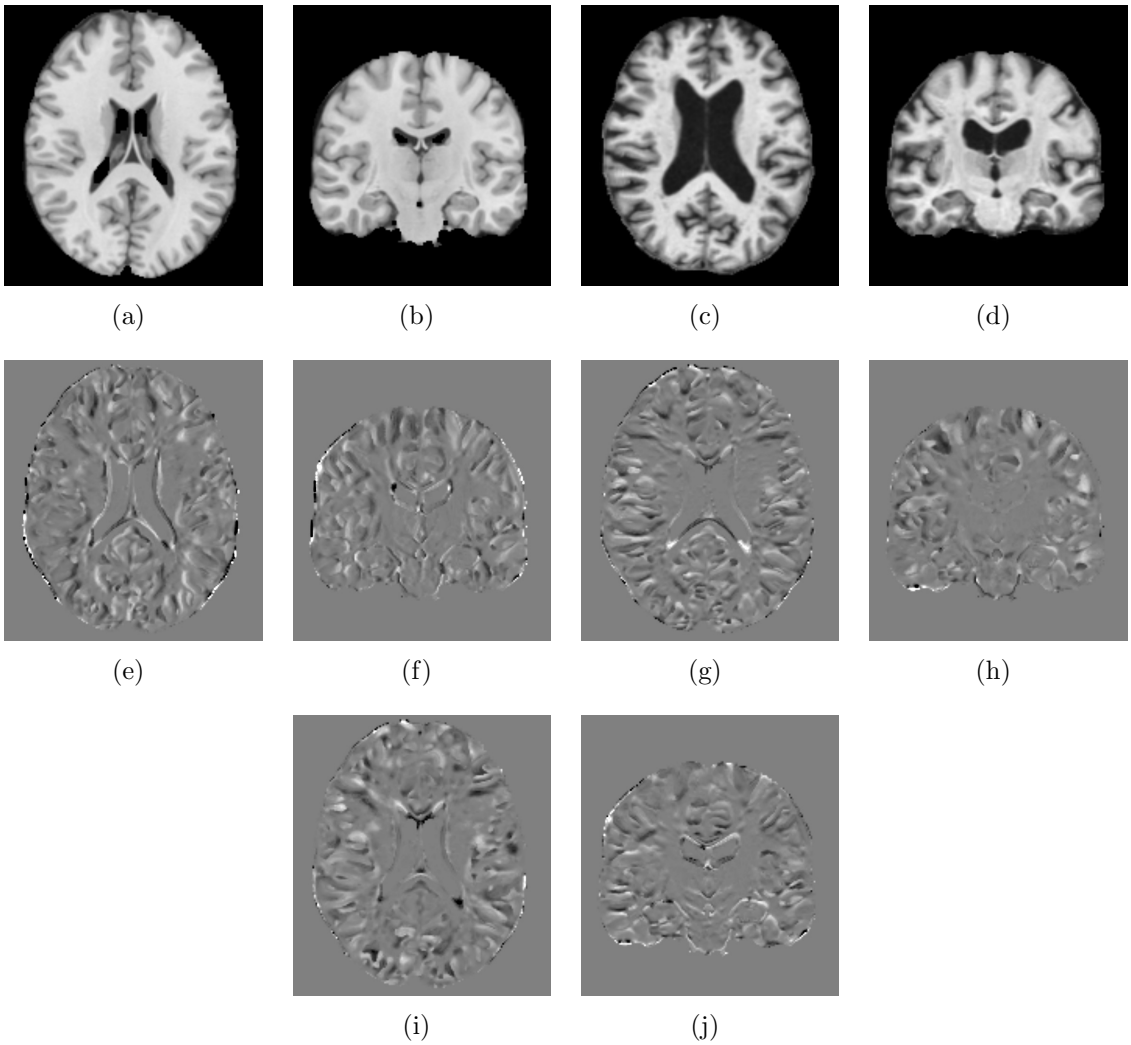


Figure 5.1: One example of the brain data used and its similarity gradient in all x-, y- and z-directions. (a)-(b): Axial and coronal views of target image. (c)-(d): Axial and coronal views of source image. (e)-(f): Axial and coronal views of SGI in x-direction. (g)-(h): Axial and coronal views of SGI in y-direction. (i)-(j): Axial and coronal views of SGI in z-direction.

proposed framework is depicted on Figure 5.2.

5.2.4 Multi-resolution framework

The procedure described in section 5.2.2 is only valid for a single-resolution registration framework. In order to extend it to be able to perform in a multi-resolution framework with L levels starting from the coarsest level L down to the finest level 1, some considerations have to be made. For level L, the method requires no change, but for all other levels l with $1 \leq l \leq L-1$,

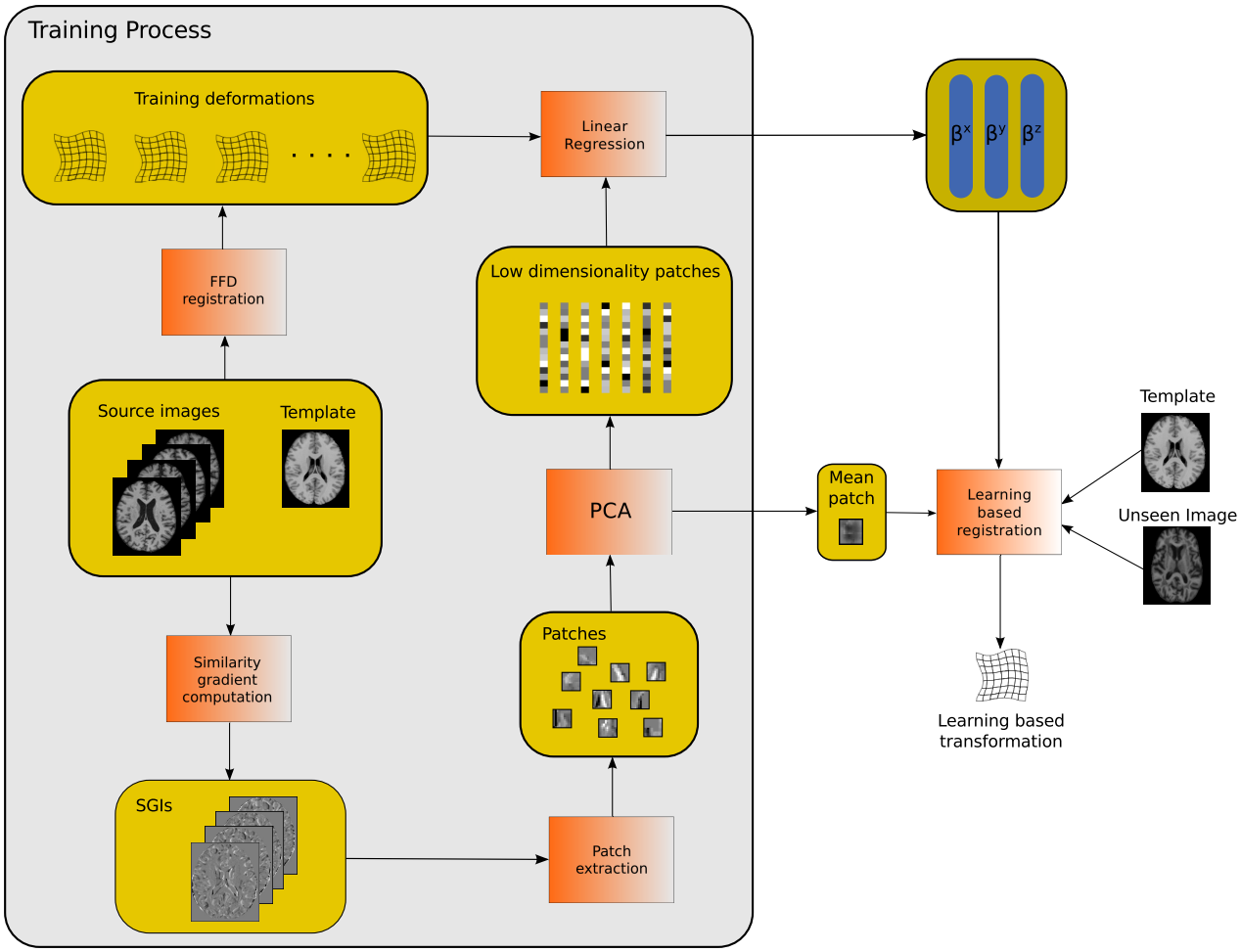


Figure 5.2: Proposed learning based framework. Note that the training process needs to be performed only once in order to be able to subsequently register any number of unseen images.

the source images used to generate the training SGIs are first deformed according to the FFD deformation field obtained down to level $l+1$ (while the target remains the same), and the FFD control point values used for regression are the difference between the FFD control point values down to level l and the FFD control point values down to level $l+1$.

5.3 Results

As previously mentioned, we utilise the B-spline FFD approach [183] as underlying parametric method for our approach. All registrations (both training and testing) where performed with control point spacings of 20, 10, and 5mm. We use Algorithm 2 of Chapter 4 for the opti-

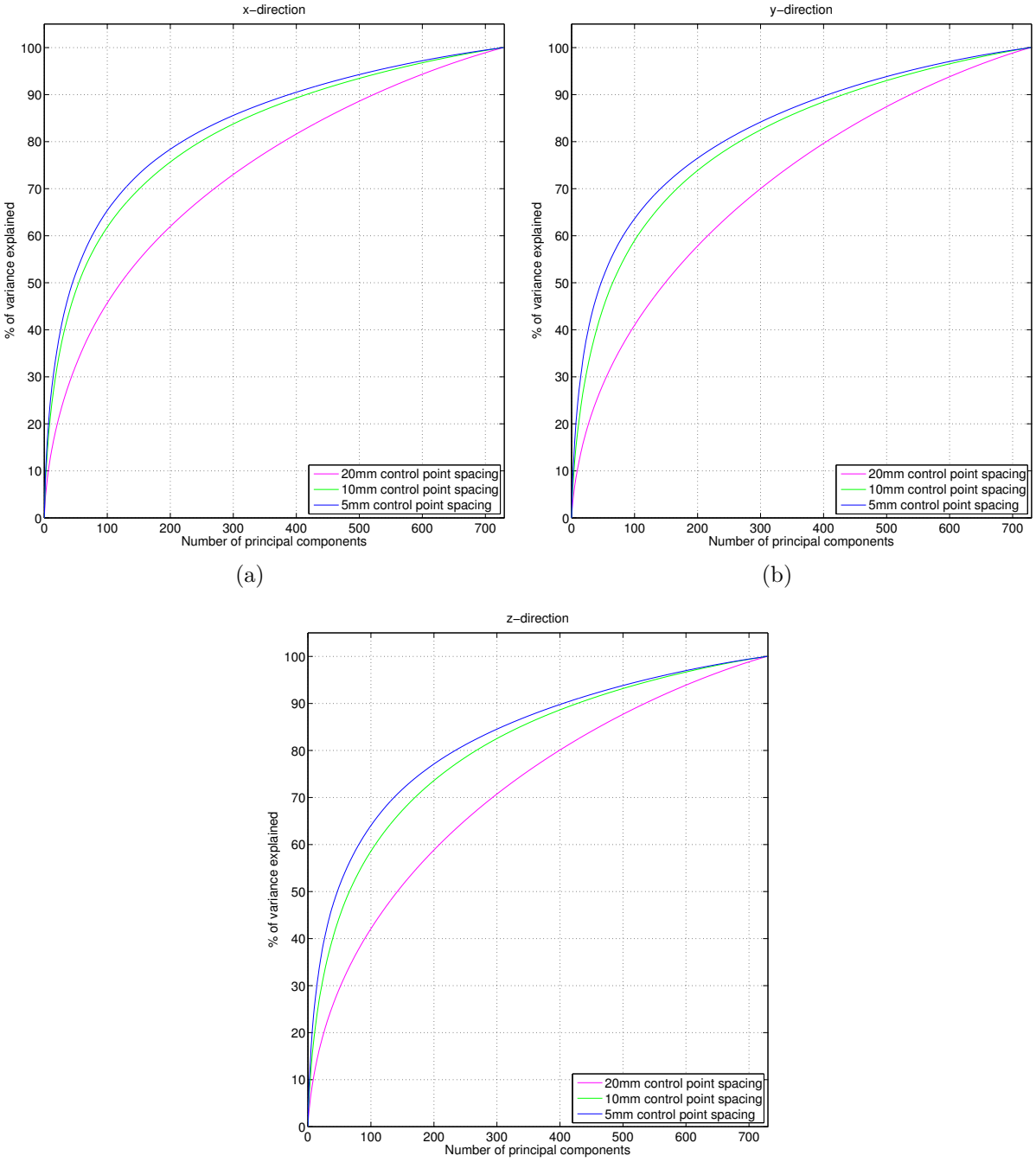


Figure 5.3: Percentage of explained variance as a function of the number of PCA components for the three utilised resolution levels. (a) Explained variance in x-direction. (b) Explained variance in y-direction. (c) Explained variance in z-direction.

misation. In the case of testing registrations, the gradient that is conjugated is the complete mixture $(1-\omega)\cdot(\mathbf{g}_{\text{ffd}})+\omega\cdot(\mathbf{g}_{\text{lb}})$. In all the conducted experiments, we utilise normalised mutual information (NMI) as similarity measure and the thin-plate bending energy of the deformation

field with a weight of $\alpha = 0.002$ as a regularisation term. We also learn the low dimisionality patches using PCA, retaining the number of components necessary to account for 98% of the total SGI patch variance. The patch sizes needed to cover the B-spline support regions are $9 \times 9 \times 9$ for the brain images used in our experiments and 9×9 for the cardiac ones. Finally, all the reported p-values are obtained using paired Wilcoxon signed rank tests.

5.3.1 Data

We employ an image set of $2+t$-dimensional 1.5T Philips Achieva SSFP MR cardiac sequences from 10 subjects and 35 preprocessed T1-weighted MR brain images originating from the OASIS database [139]. See Section 4.3.1 from Chapter 4 for a detailed description of these datasets.

5.3.2 Compactness of the patch data

We investigate the compactness of the patch data obtained from the OASIS brain dataset. Compactness determines how much of the total variance of the data can be explained as a function of the number of retained PCA components. We perform the analysis in all three x-, y- and z- directions and with patches taken for all three resolution levels (20mm, 10mm and 5mm control point spacing). Compactness of the patch data is an important aspect, since it describes the amount of dimensionality reduction achieved by the PCA step. The results are summarised in Figure 5.3. We observe that the finer the resolution level, the more compact the patches are. This is expected because on coarser resolution levels, the patches are taken from SGIs generated with downsampled versions of the target and source images, increasing their variability. Another important observation is that the dimensionality reduction in our approach (retaining 98% of the variability) is very limited, since the low dimensionality patches are described using at least 600 parameters.

5.3.3 Appearance of the learnt coefficients

We study the appearance of the projections utilised over high-dimentional patch data, compared to the B-spline projections of the standard FFD approach. Since the learnt coefficients $\hat{\beta}^{\xi}$ are trained using low-dimensional patches, the projections π^{ξ} are given by projecting these coefficients (ignoring the first one) into the high-dimensional space:

$$\pi^{\xi} = \left(\hat{\beta}^{\xi}\right)_{1\cdots C} \cdot \mathbf{B}^{\xi\top}. \tag{5.15}$$

Figure 5.4 provides a visualisation of the projections π^{ξ}, alongside the B-spline projections of standard FFD for three different cardiac subject. We show these results on cardiac data since they are 2D images and thus simpler to visualise. The main fact that becomes apparent is that the learning-based projections do not resemble a symmetric kernel, but a rather random one. We also observe that the projections look different for different subjects. These observations mean that the learned projections are specific to the particular data used for training. Another observation is that the learned projections may have both positive and negative coefficients, in contrast to the strictly non-negative coefficients of the B-spline kernel.

5.3.4 Learning-based registration results

In our experiments, we compare the registration performance of four different settings: The first setting corresponds to the standard B-spline FFD algorithm ($\omega = 0$). In the second setting, we register images using a purely learning-based gradient ($\omega = 1$). The third setting is a mixture of FFD and learning-based gradients ($\omega = 0.5$). In the final setting, the gradients are hierarchically combined by setting the value of ω to 1 at the beginning. The registration then proceeds until convergence. Subsequently, this valued is halved and the registration continues again until convergence. The algorithm continues by halving the value of ω until a new value does not produce any improvement in the line search, or until $\omega < 0.001$.

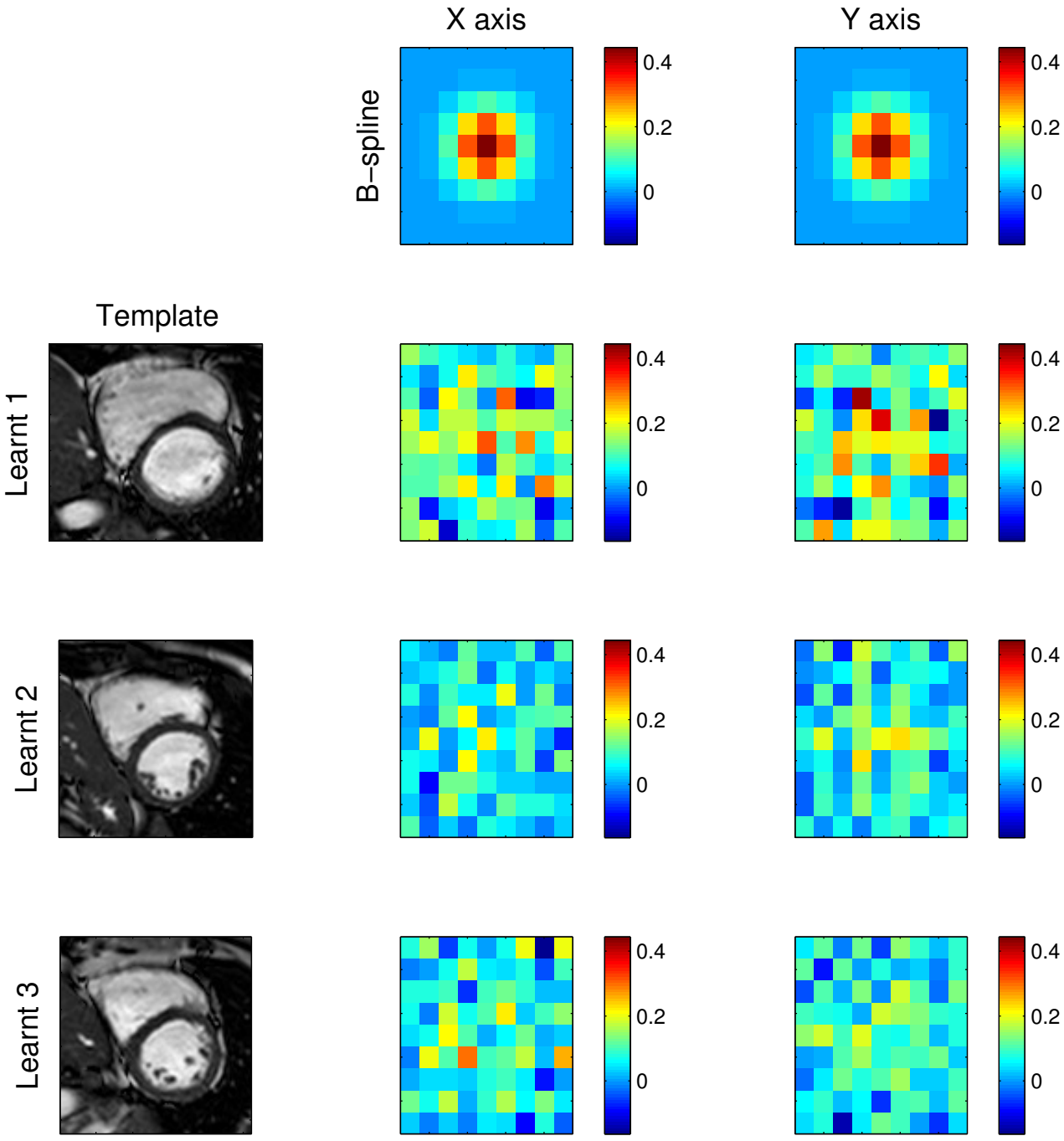


Figure 5.4: Visualization of the 2D projection factors π^{ξ} applied over patches taken from 3 subjects of the cardiac data.

We perform registrations using each of the four experimental settings using the MNI single-subject T1-atlas [142] as target and the 30 OASIS images as source. The learning-based gradients for each registration are trained using patches from SGIs generated using the target image and all other 29 images of the OASIS dataset. For evaluation, we compute all the pairwise

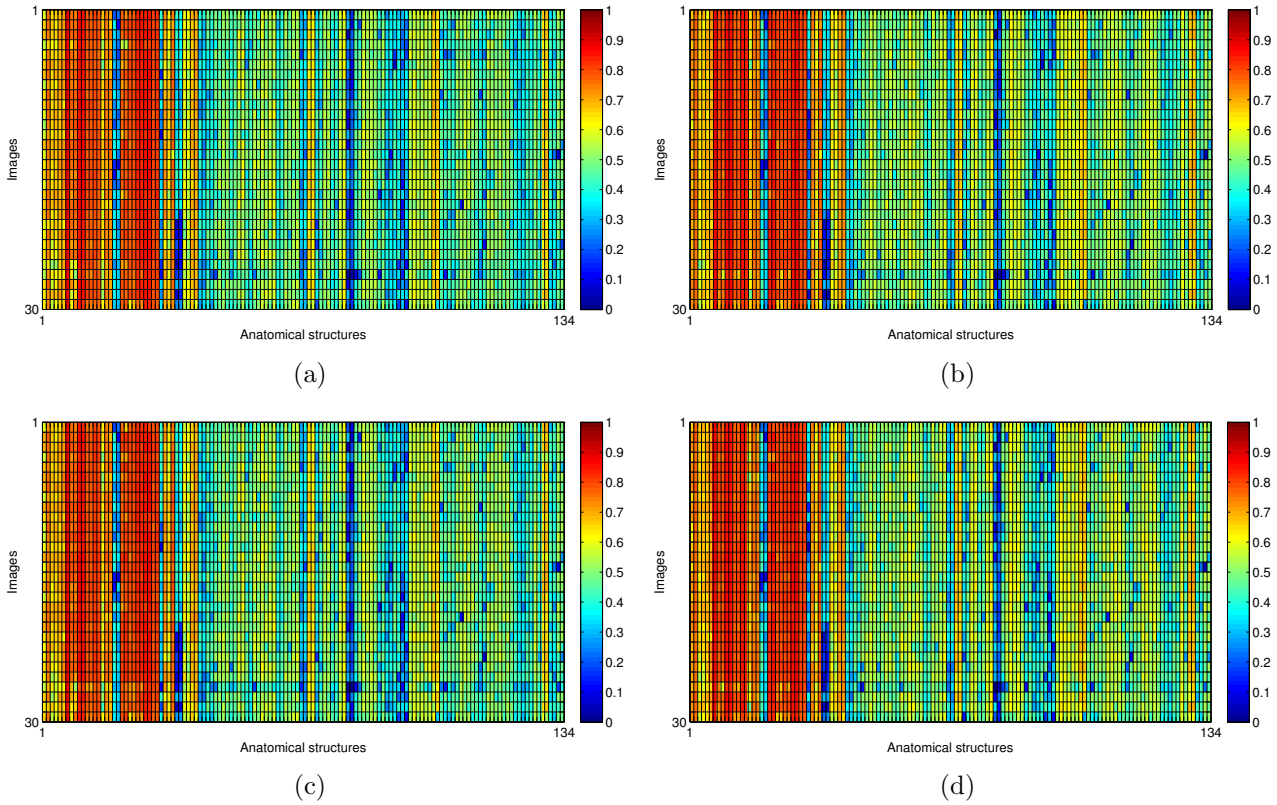


Figure 5.5: Mean pairwise similarity index per image per label for the OASIS data. We show results for each of the four experimental settings. (a) FFD (b) Learning based. (c) Mixed. (d) Hierarchically mixed.

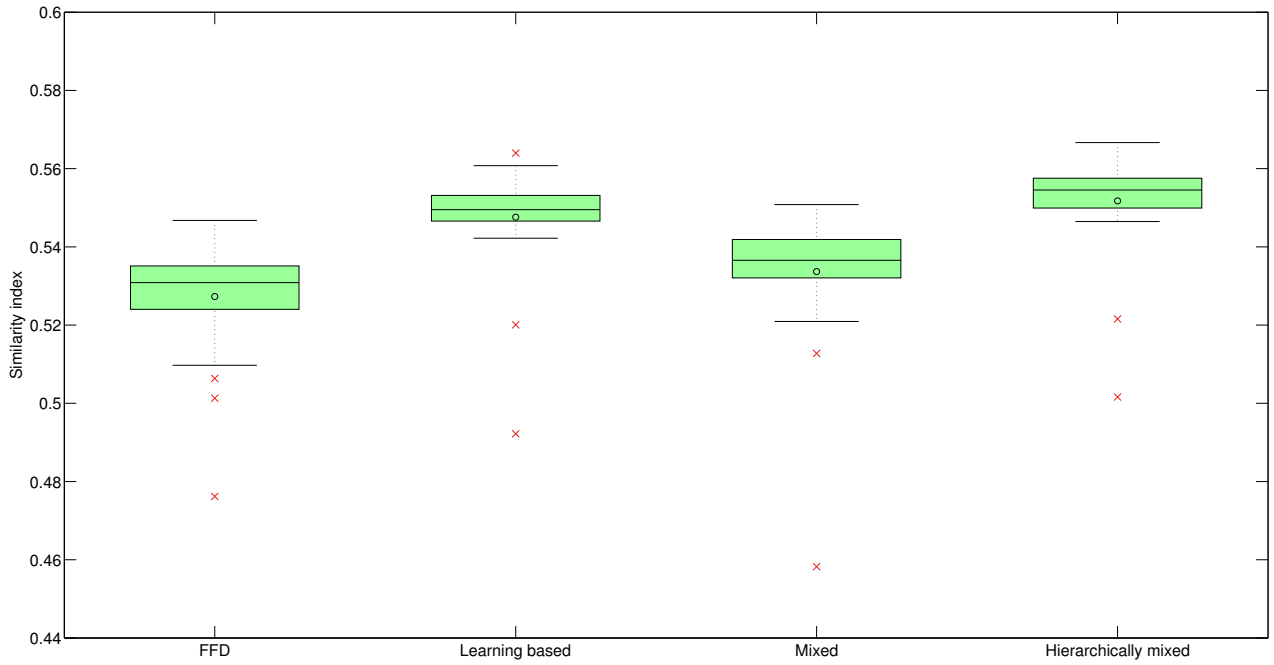


Figure 5.6: Results on OASIS brain data for each of the four experimental settings.

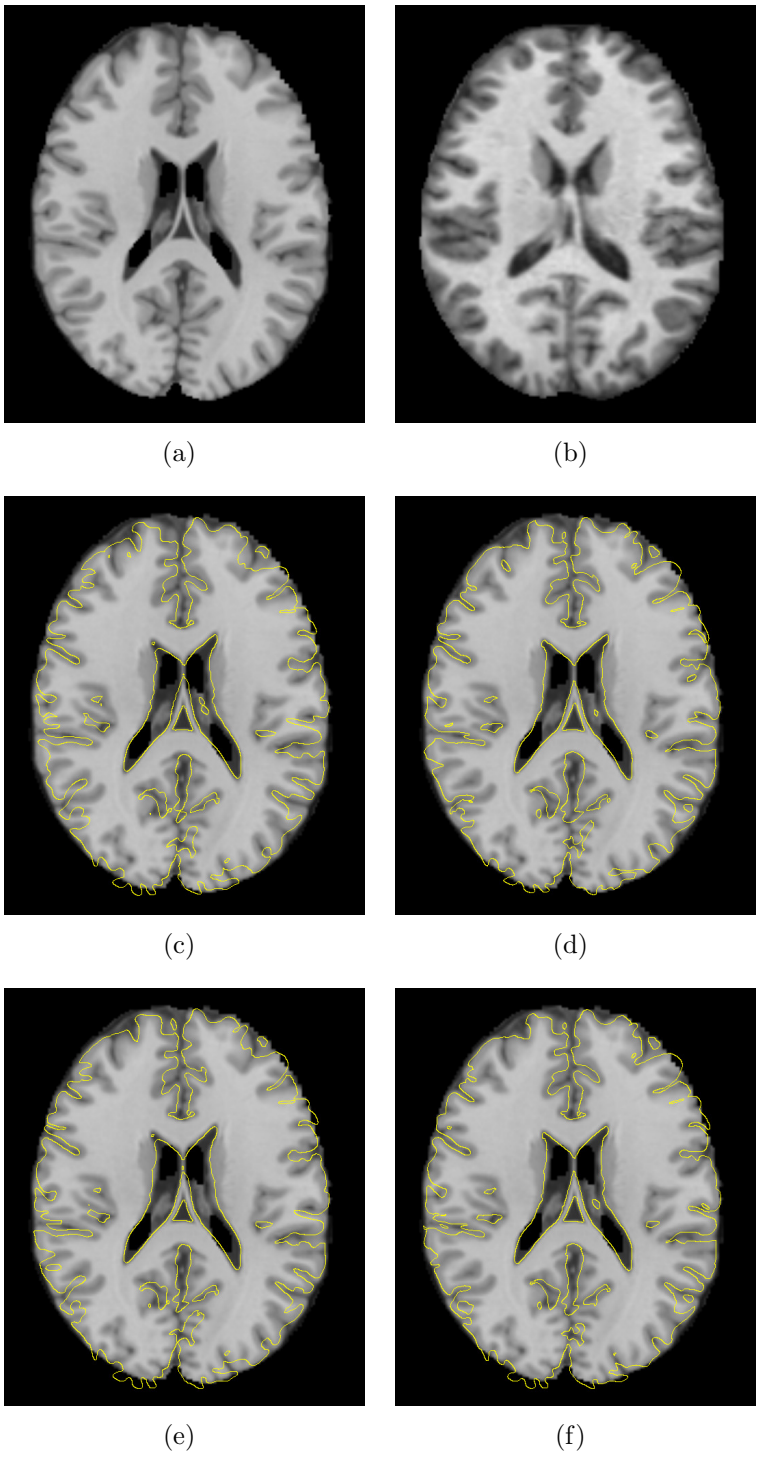


Figure 5.7: Visual comparison of the proposed registration for OASIS data against FFD registration (axial view). (a): Target image. (b): Source image. (c): Transformed source isolines using FFD registration result overlayed on the target image. (d): Transformed source isolines using pure learning based registration result overlayed on the target image. (e): Transformed source isolines using mixed registration result overlayed on the target image. (f): Transformed source isolines using hierarchically mixed registration result overlayed on the target image.

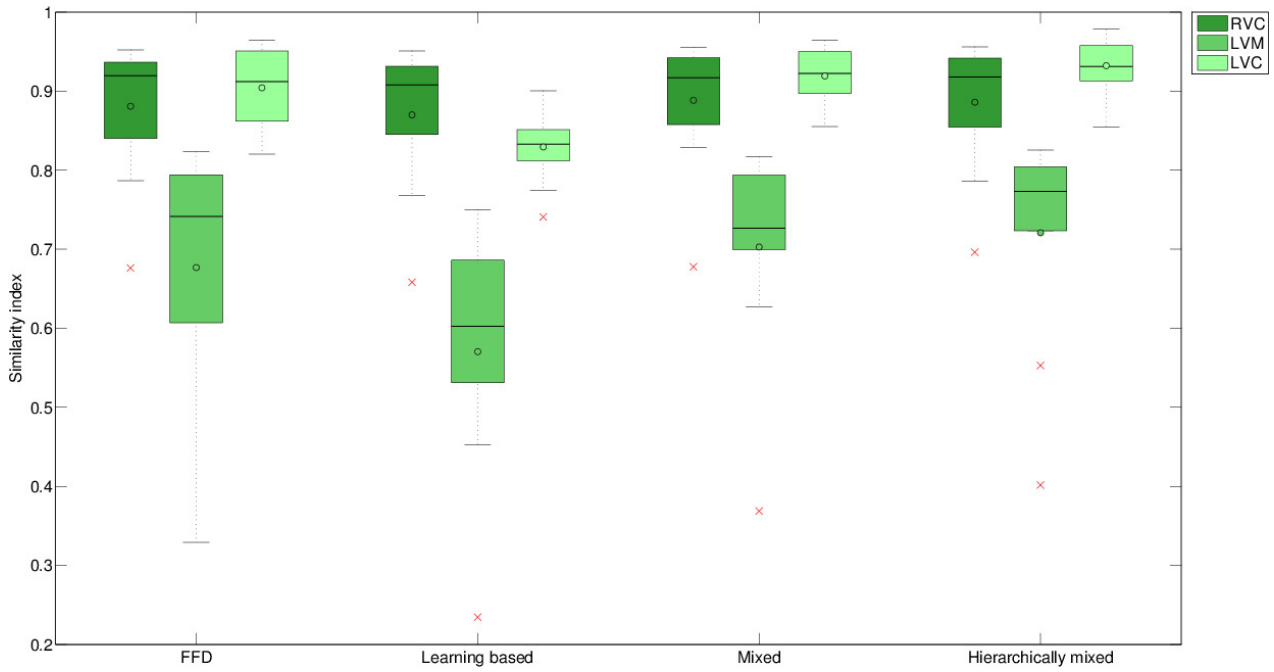


Figure 5.8: Results on cardiac data measuring Right ventricle cavity (RVC), Left ventricle myocardium (LVM) and Left ventricle cavity (LVC) similarity indices. We show results for each of the four experimental settings.

similarity indices of 134 labels propagated into the target space. These results are shown in Figure 5.6. We observe that the proposed learning-based gradient yields an improvement on the overlaps over the standard FFD approach. Furthermore, this is observed in all the settings where a learning-based gradient is involved. Moreover, the improvements are statistically significant at a 5% significance level ($p = 0.0132$ for mixed and $p < 10^{-5}$ for learning-based and hierarchically-mixed). These results suggest that the proposed method is able to converge to a better solution, as it does not get prematurely trapped in a local minima. Further visualisation of the mean overlaps for each of the anatomical labes and each image in the OASIS dataset can be seen in Figure 5.5.

Visual results of registration between the template image and a subject's image from the OASIS dataset are given in Figure 5.7. We observe that the purely learning-based hierarchical gradients yield better visual results than standard FFD registration and the mixed gradient, especially in the ventricles.

We also perform registrations over the cardiac MR dataset according to all four experimental

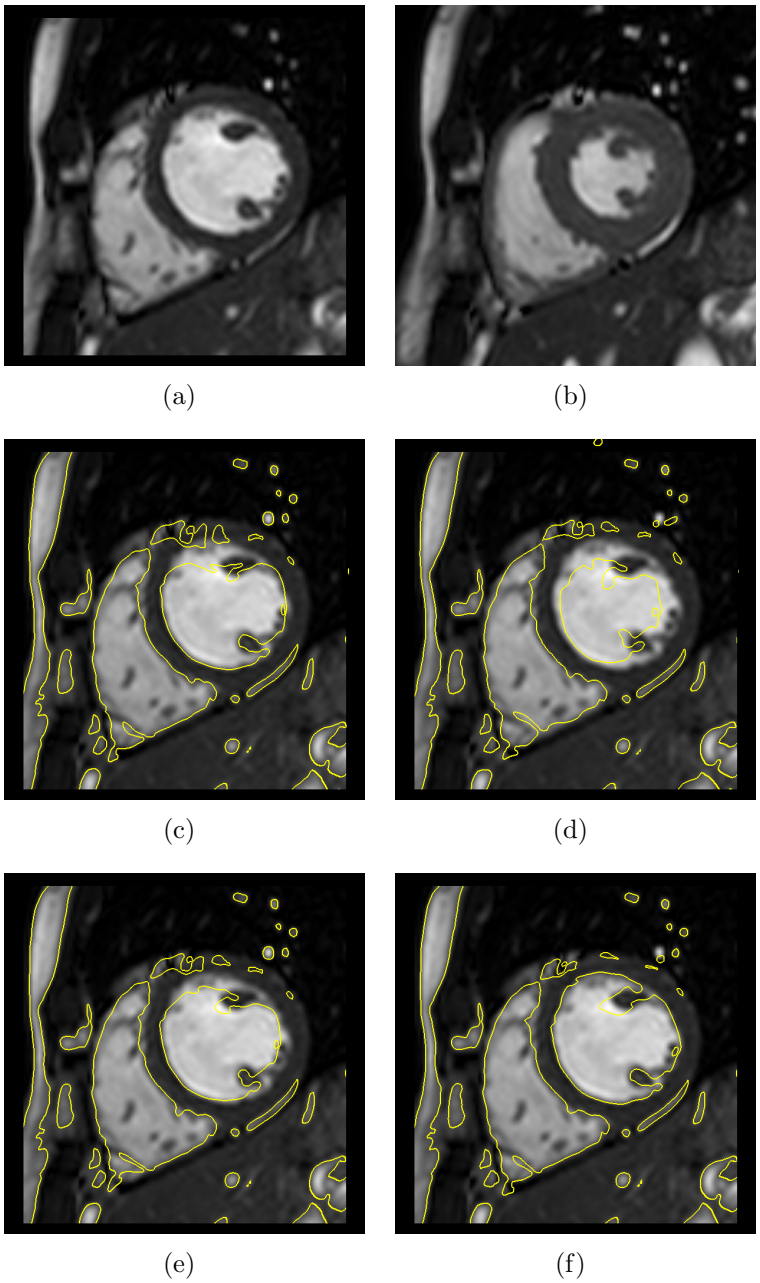


Figure 5.9: Visual comparison of the proposed registration for one of the cardiac sequences against FFD registration. (a): Target image. (b): Source image. (c): Transformed source isolines using FFD registration result overlayed on the target image. (d): Transformed source isolines using pure learning based registration result overlayed on the target image. (e): Transformed source isolines using mixed registration result overlayed on the target image. (f): Transformed source isolines using hierarchically mixed registration result overlayed on the target image.

settings. We use the first frame (end diastole) of the corresponding sequence as target image and the end systole frame of the same sequence as source. Hence, a total of 10 registrations

are performed on each setting. The learning-based gradients for each registration are trained using patches from SGIs generated using the target image and all other 28 frames of the corresponding sequence. As there are segmentations of the right ventricle cavity, left ventricle myocardium and left ventricle cavity available for both end diastole and end systole frames, we measure the similarity index between the target and the propagated source labels. Figure 5.8 shows the overlap results. The first observation that we make is that all four settings yield very similar results for the right ventricle, with differences that are not statistically significant ($p = 0.084, 0.2324, 0.375$ for learning-based, mixed and hierarchically-mixed compared to FFD, respectively). For the two labels on the left ventricle, the settings with mixed gradients yield the best results, with overlaps which are more consistent, especially for the myocardium, although these improvements are significant only for the hierarchical setting on a 5% significance level ($p = 0.0488, 0.0273$ for the left myocardium and left ventricle, respectively). However, purely learning-based registrations perform much worse for these two labels, although this is only statistically significant for the left ventricle cavity ($p = 0.0059$). This may be explained by the fact that the cardiac frames are in 2D, hence a much lower number of patches can be extracted to learn the projection.

Visual results for one of the subjects in the cardiac MR dataset is given in Figure 5.9. It is possible to visually confirm the previous result that all registration approaches have good alignment capabilities in the right ventricle cavity. We also observe that the pure learning based and standard FFD registrations are limited in its ability to register the cavity of the left ventricle accurately. This is not the case for the proposed hierarchical mixing, which visually appears to have very good registration capabilities for all structures.

5.3.5 Runtime

Finally, we measure the mean runtime of registrations over the OASIS dataset for each of the four experimental settings. The registrations were run on a standard PC desktop with 3.40GHz core frequency (they run only on one core) and 16GB of RAM. These results are

Table 5.1: Mean (Median) registration runtimes and standard deviation over all 30 registrations for the standard FFD method and the proposed statistical FFD method trained with PCA for different values of γ.

Setting	**Runtime (seconds)**
FFD	$311.3(300.3) \pm 155.3$
Learning based	$173.5(172.9) \pm 14.6$
Mixed	$324.5(311.3) \pm 155.1$
Hierarchically mixed	$477.3(483.4) \pm 26.7$

summarised in table 5.1. It can be observed that the purely learning-based registration is significantly faster than standard FFD. Furthermore, this speedup comes together with a better registration accuracy, as previously reported. The mixed gradients setting yields registrations that are comparable to standard FFD in terms of runtime. On the other hand, registrations using hierarchically mixed gradients perform substantially slower. This is expected, since the hierarchical mixing involves several serial optimisations.

5.4 Summary

We developed a new general learning-based non-rigid registration approach to improve the convergence of the optimisation of any chosen parametric energy-based image registration method. A projection from the gradient space of the energy functional to the parameter space is learnt offline from reduced-dimensionality data and subsequently used online to approximate the optimisation of the energy functional for unseen images. This computed gradient is able to yield better registration accuracy than standard analytical methods, by aiding the optimisation method to potentially avoid early local minima. Our preliminary results from experiments on brain MR image data and MR cardiac sequences show that the proposed gradient can be utilised in conjunction with the standard analytical gradient in order to improve the registration accuracy. If utilised in isolation, the proposed gradient can also significantly decrease registration runtimes and provide a better registration accuracy than FFD, if a sufficient amount of training data is available.

The training scheme of the proposed approach may be limited in the sense that SGI patches are taken from all control points. This can potentially restrict the number of training SGI that can be used, as the patch matrix may occupy a non-negligible amount of memory. For example, when using 1000 3D training images, the matrix takes approximately 300GB of memory using double precision, which is a memory capacity seldom found on standard desktop computers. As future work, we will investigate how to restrict the number of patches taken from each SGI without compromising the learning outcomes. Another possible avenue of further development is to replace the PCA dimensionality reduction and standard regression steps by one Partial Least Squares regression step using the original high-dimensional patches. In the next chapter, we propose a novel robust similarity measure for image registration.

Chapter 6

A robust similarity measure for image registration

This chapter is based on:

- Stefan Pszczolkowski, Stefanos Zafeiriou, Christian Ledig and Daniel Rueckert. "A Robust Similarity Measure for Non-rigid Image Registration with Outliers." In Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on, pp. 568-571. IEEE, 2014.

6.1 Introduction

Most of the research in the context of medical image registration focuses on alignment where correspondences can be established everywhere in the images. However, standard registration methods usually fail to perform well when the assumption of correspondence locally breaks down and does not hold. For example, the spatial correspondence between an image of a healthy subject and an image of a subject with a tumor is not defined in the pathological area. As a consequence, registrations using standard similarity measures may produce unsatisfactory results in those cases. In addition, MR imaging can introduce intensity inhomogeneities into images as a result of inhomogeneities in the magnetic field of the MR scanner. This also poses a problem for registration, since this means that the intensities of the images being registered are not spatially consistent, impeding the registration results. Therefore, there is a significant need for improved robustness of registrations involving images with ambiguous correspondences and/or intensity inhomogeneities. To the best of our knowledge, there is no previously proposed similarity measure for image registration that is robust to both imaging artefacts such as intensity inhomogeneities caused by bias fields and outliers in the images, e.g., in form of pathology. To address this chalenge, we utilise a simple, but effective similarity measure based on the angle between gradient orientations, which are obtained from the normalised image gradients. A similar approach has been recently successfully applied for the robust affine alignment of facial images [217] and shown to be robust towards occlusions and changes in illumination. Specifically, we employ this similarity measure within a widely and successfully used non-rigid registration framework based on free-form deformations (FFD) [183]. We provide both theoretical and experimental evidence of its robustness and evaluate on manually annotated MR images, comparing the proposed similarity measure to other similarity measures such as normalised mutual information and an alternative similarity measure based on normalised gradients. We also confirm robustness of the proposed similarity measure on simulated pathological imaging data from a tumour database and on real brain MR images from patients with traumatic brain injury.

6.1.1 Related work

In medical imaging, several methods have been proposed for registration of images with ambiguous correspondences, focusing on robustness [175], tumour models [237], resection cavity and recurrence models [118], or Bayesian models [75]. However, all these methods require prior knowledge in order to aid in the alignment of the regions where the spatial correspondences are not clear. Additionally, a number of methods have been proposed to reduce the effect of intensity inhomogeneities in the registration. They can be based on local similarity measures [89, 7, 203, 114, 235, 133, 241], Markov Random Fields [233, 52, 239], or correction of intensity distortions [63, 146]. Moreover, a number of robust similarity measures have also recently been presented in the computer vision community [19, 53, 217, 135].

One of the earliest approaches using normalised image gradients in the field of medical image registration was proposed by Pluim et al. [166]. In this work, normalised mutual information (NMI) [205] is weighted voxelwise by the normalised image gradients in order to incorporate spatial information. After this initial work, the first similarity measure based solely on normalised gradients was proposed by Haber et al. [74]. Since its introduction, this similarity measure has been successfully utilised for CT liver registration [87], ultrasound-CT liver registration [121], tracking in fluoroscopy images [174], and optical flow based tracking [86]. However, as we show in this chapter, this similarity measure is less robust to image inhomogeneities and may be affected when gross outliers, such as lesions or tumours, are present in the images.

6.2 Methods

6.2.1 Proposed similarity measure

In the paper by Haber et al. [74], the authors use the observation that a target and a source image come into alignment when the square of the cosine of the angle between the target and warped source gradient orientations is maximised. In contrast, we propose to adopt the

similarity measure introduced by Tzimiropoulos et al. [217], which corresponds to only the cosine (not squared) between gradient orientations, and introduce it into the problem of medical image registration. Hence, we propose to utilise the following similarity measure as energy functional:

$$E_{similarity}(I_0; I; \mathbf{T}) = -\frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \cos \alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})). \tag{6.1}$$

Here, Ω is the set of indices corresponding to the target image voxels, $\alpha(\cdot, \cdot)$ is the angle between two gradient orientations, $\mathbf{T}$ is the current spatial transformation, and $I_1 = I \circ \mathbf{T}$ denotes the warped source image. The proposed energy term in Equation (6.1) can be expressed in terms of the dot product $\langle \cdot, \cdot \rangle$ between gradients

$$E_{similarity}(I_0; I; \mathbf{T}) = -\frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \frac{\langle \nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}) \rangle}{||\nabla I_0(\mathbf{x})|| \, ||\nabla I_1(\mathbf{x})||}. \tag{6.2}$$

As we later show, the minor difference between the proposed similarity and that of Haber et al. has an important impact in the ability of the similarity measure to deal with ambiguous corespondences.

We refer the interested reader to appendix A for a detailed derivation of the gradient for both the proposed similarity measure and the similarity measure proposed by Haber et al. [74].

6.2.2 Numerical stability

As discussed in [74], normalised gradient field are not differentiable in homogeneous areas because the terms in the denominator become zero. We thus compute the energy term using regularised normalised gradient fields as presented in [184]:

$$E_{similarity}(I_0; I; \mathbf{T}) = -\frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \frac{\langle \nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}) \rangle_{\varrho, \tau}}{||\nabla I_0(\mathbf{x})||_{\varrho} \, ||\nabla I_1(\mathbf{x})||_{\tau}}, \tag{6.3}$$

where $\langle \cdot, \cdot \rangle_{\varrho,\tau} = \langle \cdot, \cdot \rangle + \varrho\tau$ and $|| \cdot ||_* = \sqrt{\langle \cdot, \cdot \rangle_{*,*}}$. In other words, this normalisation scheme corresponds to a reformulation of the image gradients as:

$$\begin{aligned} \nabla I_0 &\equiv \left(\tfrac{\partial I_0}{\partial x}, \tfrac{\partial I_0}{\partial y}, \tfrac{\partial I_0}{\partial z}, \varrho \right)^\top \\ \nabla I &\equiv \left(\tfrac{\partial I}{\partial x}, \tfrac{\partial I}{\partial y}, \tfrac{\partial I}{\partial z}, \tau \right)^\top \end{aligned} . \tag{6.4}$$

In this work, ϱ and τ are not user-defined parameters as in [184]. Instead, they are computed following a choice based on total variation [12]:

$$\varrho = \frac{\eta}{V_{I_0}} \sum_{\mathbf{x} \in \Omega_{I_0}} |\nabla I_0(\mathbf{x})|, \qquad \tau = \frac{\eta}{V_I} \sum_{\mathbf{x} \in \Omega_I} |\nabla I(\mathbf{x})|. \tag{6.5}$$

Here, V_* is the volume of interest in the image domain Ω_* and $\eta > 0$ is a parameter to allow noise filtering. Note that the value of η (and consequently the values of ϱ and τ), sets a threshold that separates small gradients (noise) from the strong gradients that drive the registration.

6.2.3 Robustness of the proposed similarity measure

Robustness against intensity inhomogeneities

A significant advantage of normalised gradient-based methods is their invariance towards low frequency intensity changes, as we now demonstrate. Consider an image signal M with no intensity inhomogeneities and a multiplicative, non-negative bias field F which is assumed to be smooth, i.e., constant in the small neighborhood $\mathcal{N}(\mathbf{x}) = (\Delta x, \Delta y, \Delta z)$. This means $F(\mathbf{p}) \approx F(\mathbf{x}), \forall \mathbf{p} \in \mathcal{N}(\mathbf{x})$. We have for Δx:

$$\begin{aligned} I_{\text{BIAS}}(\mathbf{x}) &= M(\mathbf{x})F(\mathbf{x}) + \varepsilon \\ I_{\text{BIAS}}(\mathbf{x} + \Delta x) &= M(\mathbf{x} + \Delta x)F(\mathbf{x} + \Delta x) + \varepsilon. \end{aligned} \tag{6.6}$$

Here, ε is an additive noise term that does not depend on the spatial variable $\mathbf{x}$. Given that $F(\mathbf{x})$ is constant within the neighborhood, we have:

$$\begin{aligned}\frac{\partial I_{\text{BIAS}}(\mathbf{x})}{\partial x} &= \lim_{\Delta x \to 0} \frac{I_{\text{BIAS}}(\mathbf{x}+\Delta x) - I_{\text{BIAS}}(\mathbf{x})}{\Delta x} \\ &\approx \lim_{\Delta x \to 0} \frac{F(\mathbf{x})(M(\mathbf{x}+\Delta x) - M(\mathbf{x}))}{\Delta x} \\ &= F(\mathbf{x}) \frac{\partial M(\mathbf{x})}{\partial x}.\end{aligned} \tag{6.7}$$

Using this result, we now show that the proposed cost function is indeed robust to any locally constant bias field using the normalisation scheme in (6.3). If the contributions of ϱ and τ are disregarded, we have:

$$\frac{\nabla^x I_{\text{BIAS}}(\mathbf{x})}{||\nabla I_{\text{BIAS}}(\mathbf{x})||} = \frac{\frac{\partial I_{\text{BIAS}}(\mathbf{x})}{\partial x}}{\sqrt{\left(\frac{\partial I_{\text{BIAS}}(\mathbf{x})}{\partial x}\right)^2 + \left(\frac{\partial I_{\text{BIAS}}(\mathbf{x})}{\partial y}\right)^2 + \left(\frac{\partial I_{\text{BIAS}}(\mathbf{x})}{\partial z}\right)^2}}. \tag{6.8}$$

By using equation (6.7) we obtain

$$\frac{\nabla^x I_{\text{BIAS}}(\mathbf{x})}{||\nabla I_{\text{BIAS}}(\mathbf{x})||} = \frac{F(\mathbf{x}) \frac{\partial M(\mathbf{x})}{\partial x}}{\sqrt{\left(F(\mathbf{x}) \frac{\partial M(\mathbf{x})}{\partial x}\right)^2 + \left(F(\mathbf{x}) \frac{\partial M(\mathbf{x})}{\partial y}\right)^2 + \left(F(\mathbf{x}) \frac{\partial M(\mathbf{x})}{\partial z}\right)^2}}. \tag{6.9}$$

Here we observe that $F(\mathbf{x})$ vanishes, yielding

$$\begin{aligned}\frac{\nabla^x I_{\text{BIAS}}(\mathbf{x})}{||\nabla I_{\text{BIAS}}(\mathbf{x})||} &= \frac{\frac{\partial M(\mathbf{x})}{\partial x}}{\sqrt{\left(\frac{\partial M(\mathbf{x})}{\partial x}\right)^2 + \left(\frac{\partial M(\mathbf{x})}{\partial y}\right)^2 + \left(\frac{\partial M(\mathbf{x})}{\partial z}\right)^2}}. \\ &= \frac{\nabla^x M(\mathbf{x})}{||\nabla M(\mathbf{x})||}\end{aligned} \tag{6.10}$$

Equations (6.6)-(6.10) are analogous for Δy and Δz. This leads to

$$\frac{\nabla I_{\text{BIAS}}(\mathbf{x})}{||\nabla I_{\text{BIAS}}(\mathbf{x})||} = \frac{\nabla M(\mathbf{x})}{||\nabla M(\mathbf{x})||}, \tag{6.11}$$

demonstrating the invariance of normalised gradient-based similarity measures with respect to F.

Robustness against ambiguous correspondences

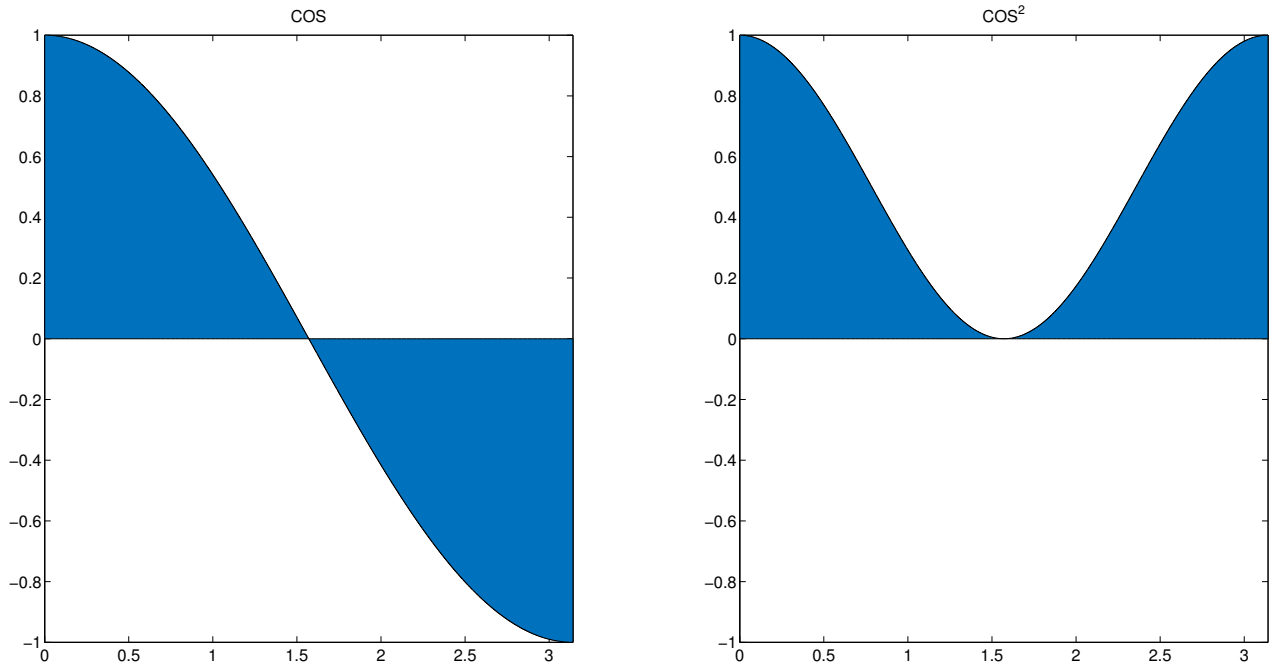


Figure 6.1: Integrals for both cosine and cosine squared functions in the $[0, \pi]$ range. Note that the mean of the cosine is zero in that range, which is not the case for cosine squared.

As we later show in the results, the similarity measure presented in [74], is not robust against ambiguous spatial correspondences. Consider a region Ω_0 in the target image with no such correspondences in the source image. Intuitively, the angles form by the normalised gradients are random and thus uniformely distributed in the $[0, \pi]$ range. Hence, the sum of the values for the inner product of the normalised gradients taken from Ω_0 can be approximated by the integral in that range. As shown in Figure 6.1, the integral for cosine squared similarity measure is strictly positive. Thus, $\sum_{\mathbf{x}\in\Omega_0} \cos^2 \alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})) \gg 0$. Consequently, the total cost function can be arbitrarily biased by the presence of outliers producing ambiguous correspondences. In contrast, the integral for cosine similarity measure is exactly zero. As a consequence, $\sum_{\mathbf{x}\in\Omega_0} \cos \alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})) \approx 0$, which means that the presence of outliers does not bias the proposed similarity measure.

6.3 Results

As previously mentioned, we incorporate the proposed similarity measure into a B-spline FFD approach [183]. For comparison, we also incorporate the cosine squared similarity measure of Haber et al. [74] and normalised mutual information (NMI) [205] into our framework. In all the conducted experiments, we utilise the thin-plate bending energy of the deformation field with a weight of $\alpha = 0.002$ as a regularisation term and optimise using conjugate gradient descent (see Algorithm 2 of Chapter 4). We use the same regularisation weight for every similarity since the empirical range of values that they take using our experimental image datasets are of very similar width for all of them. Finally, all the reported p-values are obtained using paired Wilcoxon signed rank tests.

6.3.1 Data

We employ an image set of $2+t$-dimensional 1.5T Philips Achieva SSFP MR cardiac sequences from 10 subjects. We also use 35 preprocessed T1-weighted MR brain images originating from the OASIS database [139] and 820 preprocessed 1.5 Tesla T1-weighted images from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [149]. See Section 4.3.1 from Chapter 4 for a detailed description of these datasets.

Additionally, we use a dataset of 10 simulated MR images of the brain depicting tumours. These images are taken from the BraTS MICCAI 2012 challenge [1]. Half of these images show high grade gliomas and the other half has low grade ones. The images are labelled into white matter (WM), gray matter (GM), cerebrospinal fluid (CSF) and 2 further labels for the tumour areas. All the images are skull stripped, and have $256 \times 181 \times 256$ voxels with an isotropic resolution of 1mm. A visualisation of all the subjects from this dataset is provided in Figure 6.2.

Finally, a pair of MR brain images of a patient with traumatic brain injury (TBI) is also

[1] `http://www2.imm.dtu.dk/projects/BRATS2012/`

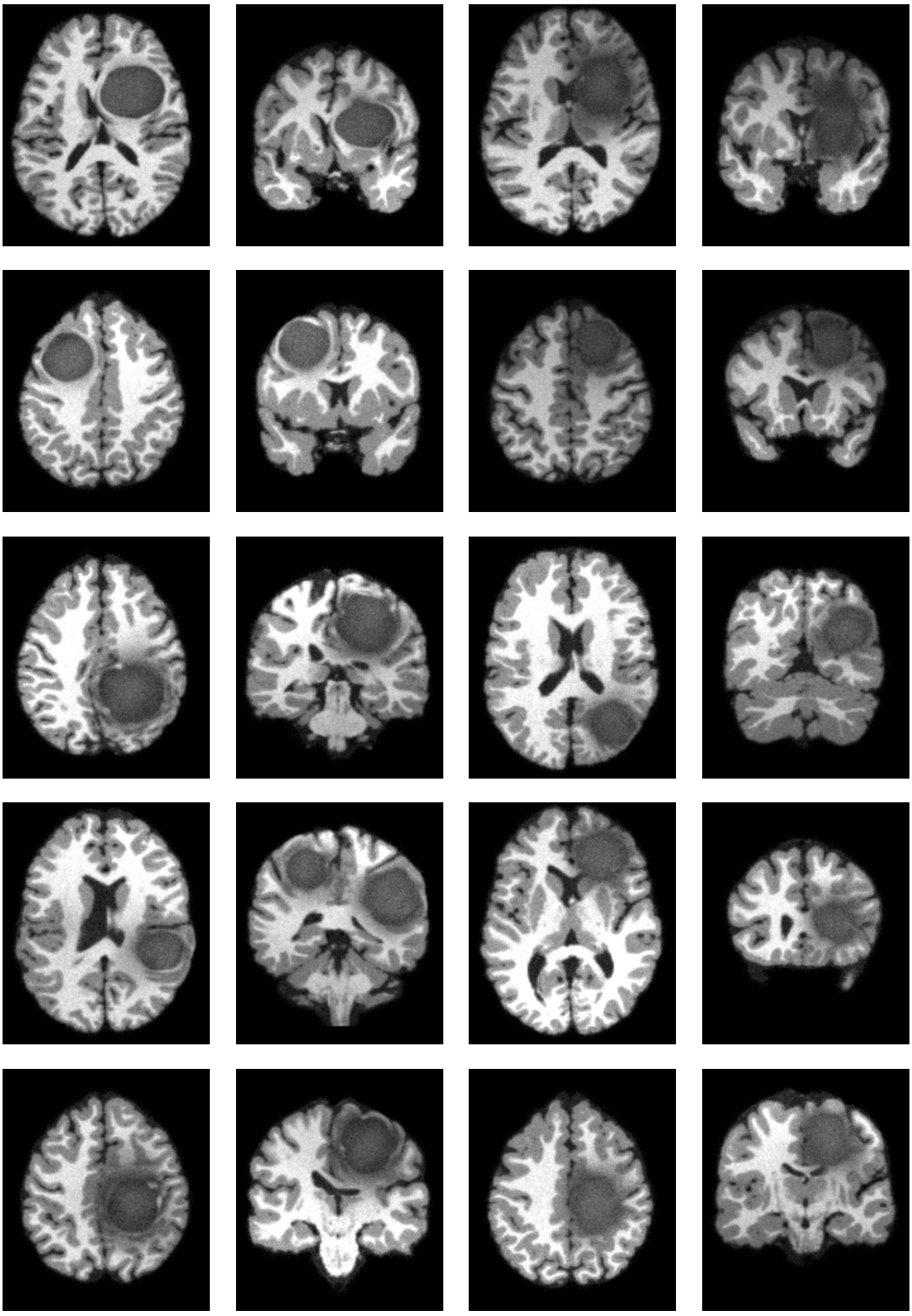
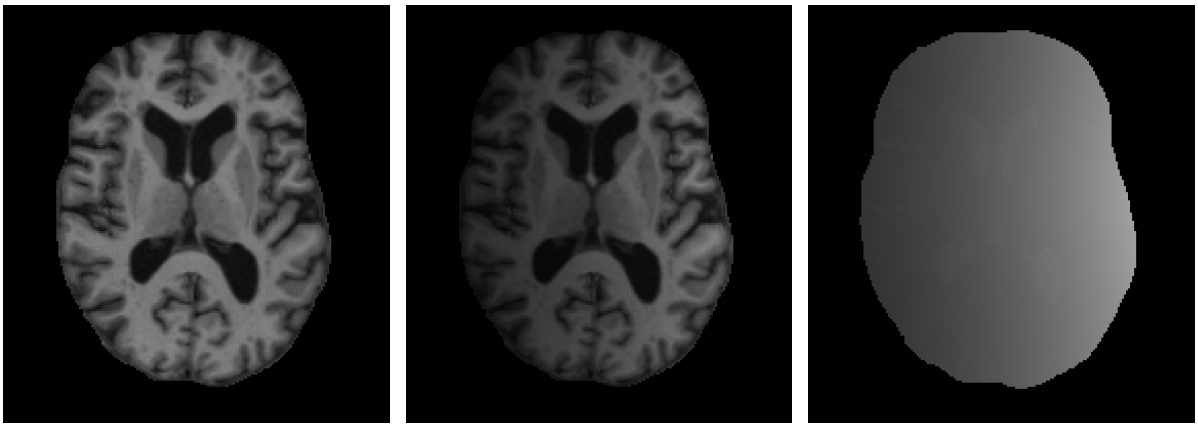


Figure 6.2: BraTS subject in axial and coronal view. First and second columns depict subjects with high-grade gliomas. Third and fourth columns depict subjects with low-grade gliomas.

Table 6.1: Summary of the datasets utilised to thest the proposed method.

Dataset	Number of Images	Modality	Organ	Annotation	Size	Resolution (mm)
BraTS	10	MR	Brain	5 structures	$256 \times 181 \times 256$	$1 \times 1 \times 1$
TBI	2	MR	Brain	N/A	$176 \times 240 \times 256$	$1 \times 1 \times 1$



(a) (b) (c)

Figure 6.3: Axial view of a T1-weighted brain images utilised for intensity inhomogeneity simulation. (a): Original. (b): With simulated bias field applied. (c): Bias field

used. One image corresponds to a baseline scan and the other is the followup scan taken after 4 months and 19 days at Turku University Hospital. Both images have $176 \times 240 \times 256$ voxels with an isotropic resolution of 1mm. These images are later shown in Figure 6.9. For convenience, we also summarise the datasets used exclusively in this Chapter in table 6.1.

6.3.2 MR images with intensity inhomogeneities

Here we evaluate the performance of our proposed similarity measure against intensity inhomogeneities. This relaxes the necessity of an explicit intensity correction step in the registration pipeline (e.g. [215]), which can be time consuming and a potential source of errors, especially for non-brain images.

Bias field simulation

To introduce intensity inhomogeneities into the images, we simulate several two-dimensional complex-valued MRI sensitivity maps using a MATLAB tool[2]. For each image, we simulate the effect of 8 uniformly placed coils. Then, we randomly select one of the 8 generated sensitivity maps as the final map S for the image. Since the sensivity maps are two-dimensional, we apply them to every 2D slice of the image along the Z-axis in a weighted fashion. The idea behind this weighting is to apply the simulated 2D bias field along the z-direction with peak strength in the middle slice and gradually decaying towards the extreme slices. Hence, if we denote the original image as M, then the simulated image with intensity inhomogeneities I is constructed according to

$$I(\cdot,\cdot,z) = int\left[w(z) \otimes \|S(\cdot,\cdot,z)\| \otimes M(\cdot,\cdot,z)\right] \qquad z \in [1, N_z], \tag{6.12}$$

where N_z corresponds to the number of image slices along the z-direction, $\otimes$ is the voxelwise multiplication and $w(z)$ is given by

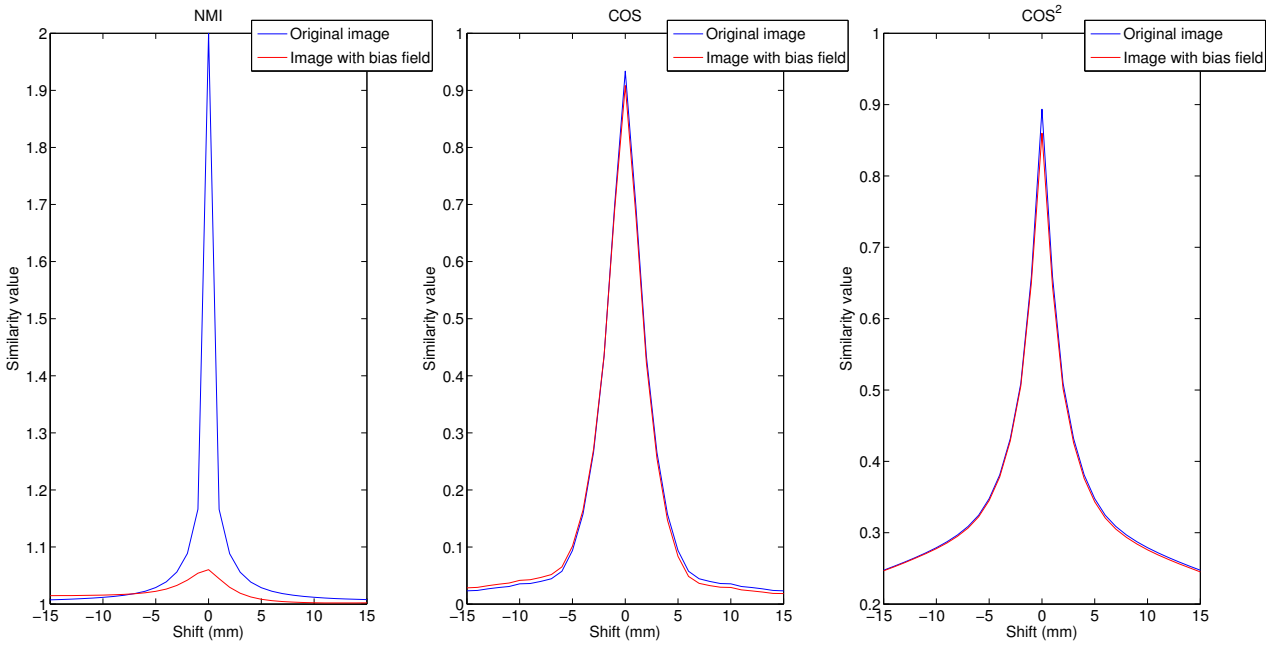
$$w(z) = 1 + 10 \cdot \mathcal{G}\left(\frac{N_z - 1}{2}, \sigma\right). \tag{6.13}$$

Here, $\mathcal{G}(\mu, \sigma)$ is the Gaussian distribution function with mean μ and standard deviation σ. We use a factor of 10 in order to amplify the range of the Gaussian function. For all the simulations we use $\sigma = 0.15 \cdot (N_z - 1)$. As an example, we show in Figure 6.3 the middle slice of an image before and after applying the simulated bias on it and the resulting bias field.

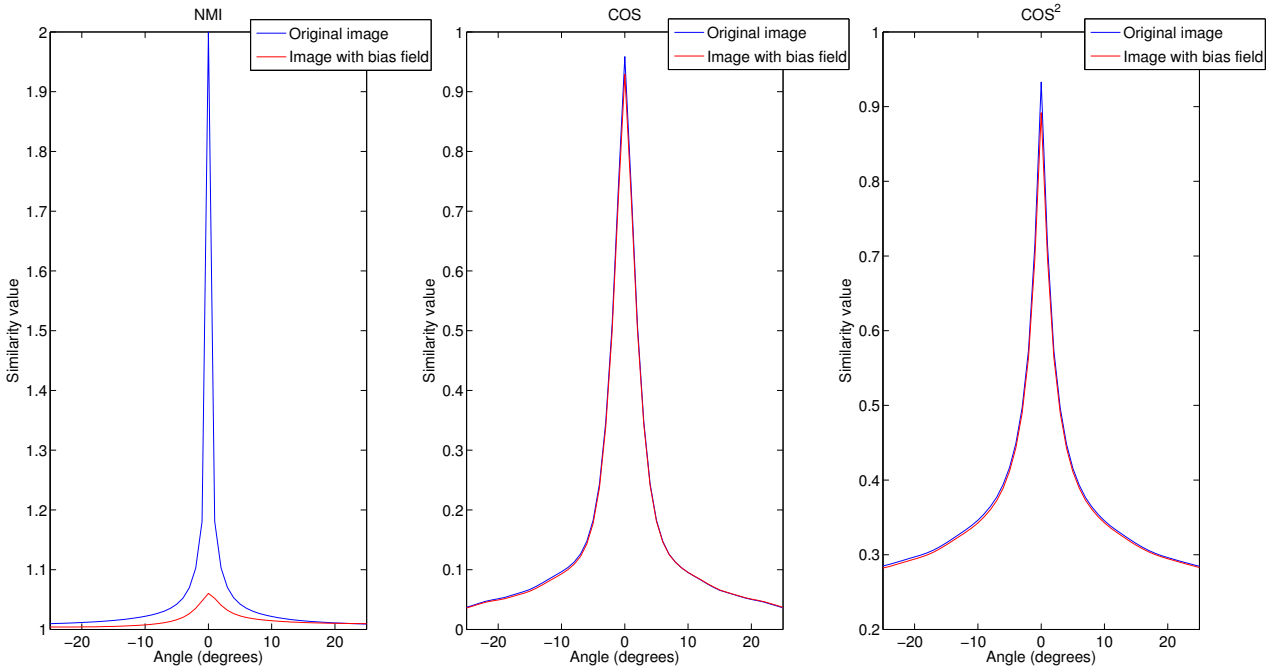
Behaviour of the similarity measures for rigid transformations

We study the behaviour of the proposed similarity measure, cosine squared by Haber et al. and normalised mutual information (NMI) with respect to translations and rotations for a 3D MR

[2] `bigwww.epfl.ch/algorithms/mri-reconstruction`



(a)



(b)

Figure 6.4: Effect of (a) shift and (b) rotation when using the original image and the image with bias field as target. We observe that the values of proposed similarity measures are almost not affected by the presence of intensity inhomogeneities

image of a brain. The target image corresponds to either the original image or the image with simulated inhomogeneities. The source image corresponds to the translated/rotated original image (with no intensity inhomogeneities). Figure 6.4 shows how each of the similarity measures

behave as a function of the rigid transformations. We observe that all three similarity measures achieve a peak when there is no translation/rotation. However in contrast to NMI, the values of the normalised gradient field based similarity measures are hardly affected by the presence of intensity inhomogeneities.

Non-rigid registration

To evaluate the proposed similarity measure for non-rigid registration, we perform the $30 \times 29 = 870$ pairwise registrations with control point spacings of 20, 10, 5 and 2.5mm, using the original images from the OASIS dataset. We subsequently introduce different smooth intensity inhomogeneities individually to all the images according to the procedure described in section 6.3.2 and repeat the registrations again using the original images as target and the affected ones as source.

We compare the gradient-based similarity measures with noise parameter η set to 0.1 against NMI in their ability to produce a deformation field able to accurately propagate the manual segmentation labels. We measure the registration accuracy using the similarity index (SI), both for the original images and the images with bias field applied. We compute the mean and standard deviation of the SI values calculated on the propagated and reference labels for all 870 propagations. We differentiate between the 98 cortical and 36 subcortical labels. The results are shown in Figure 6.5. We observe that NMI performs well when there are no intensity inhomogeneities in the images. On the contrary, it is severely affected by the presence of intensity inhomogeneities. Conversely, both gradient-based similarity measures show similar performance for registrations with and without intensity inhomogeneities, demonstrating their robustness. Nevertheless, the proposed similarity measure performs slightly better that cosine squared, and the differences are statistically significant ($p < 10^{-5}$). To complement the analysis, we show the same results for each of the images and each label in Figure 6.6.

It is important to note that in the case where no intensity inhomogeneities are present, the proposed similarity measure performs not as well as NMI ($p < 10^{-5}$). We observe that in the

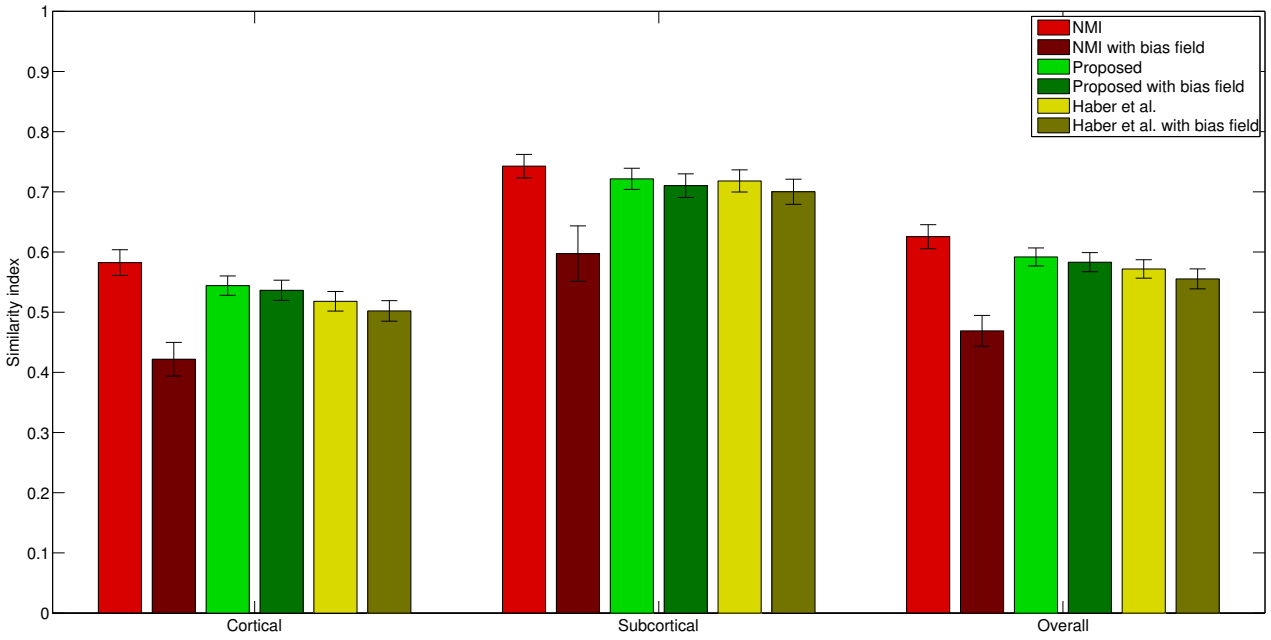


Figure 6.5: Mean similarity index and standard deviation over cortical and subcortical labels for all $30 \times 29 = 870$ registrations.

particular case of MR brain images, the discrimination between noise- and structure-related gradients is very challenging, especially in cortical areas, as can be deducted from the results shown in Figure 6.6-(a), -(c) and -(e).

6.3.3 MR images with pathologies

Registration of images depicting pathology is a challenging task, since the images may exhibit strong structural differences that cannot always be matched. Here, we show that our similarity measure is capable of handling images with areas of ambiguous spatial correspondences, e.g., areas of pathology, without any prior knowledge nor any subsequent correction step.

Affine registration

For a quantitative evaluation, a labelled image of a healthy subject is registered to all 10 images in our BraTS dataset using NMI and both normalised gradient field based similarity measures. The registrations using either the cosine or cosine squared similarity measure were run in two

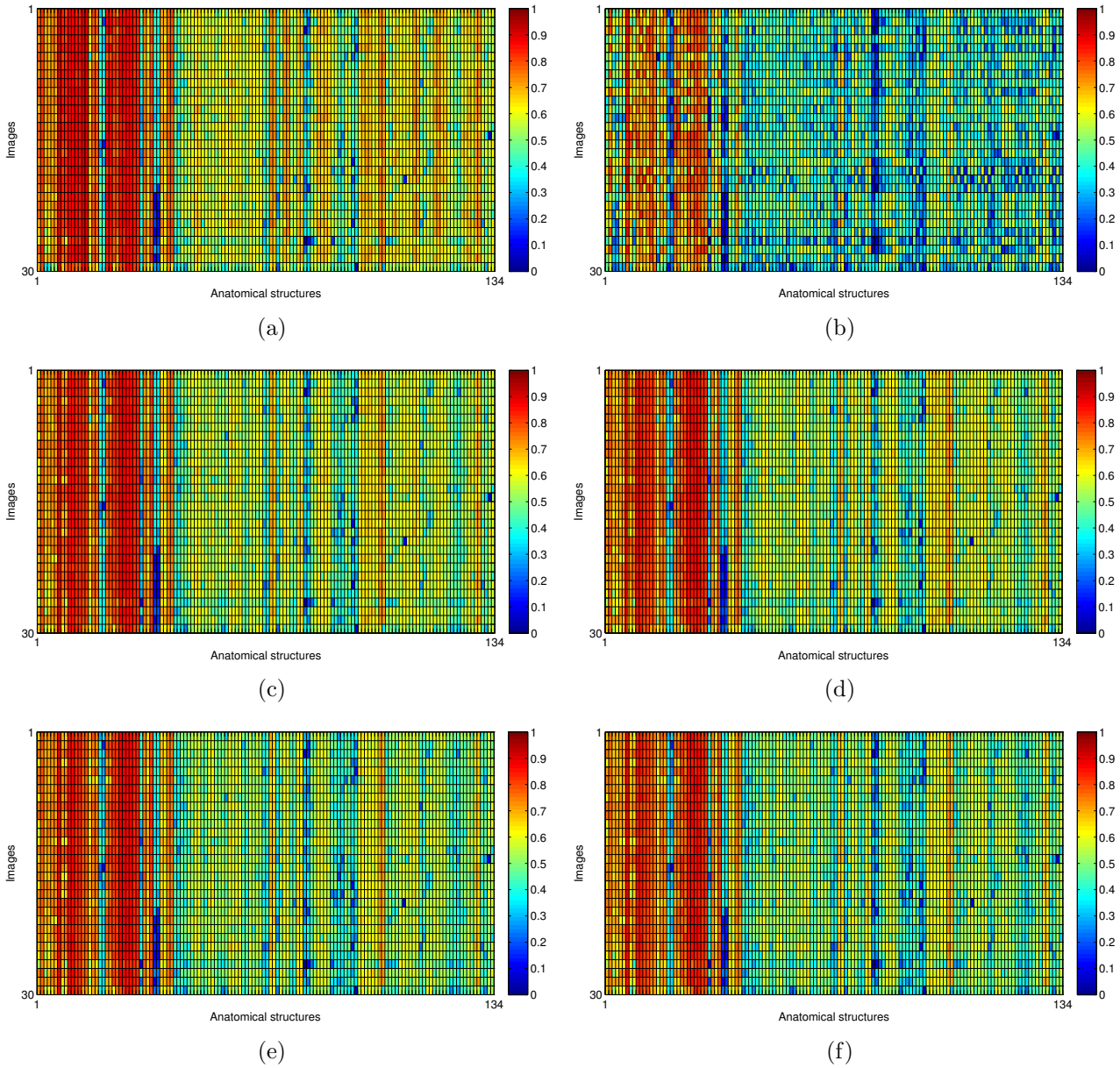


Figure 6.6: Mean pairwise similarity index per image per label for the OASIS data. (a) NMI (b) NMI with bias field. (c) Proposed similarity. (d) Proposed similarity with bias field. (e) Haber et al. [74]. (f) Haber et al. [74] with bias field.

settings. In the first setting, the noise parameter η is set to 0.1. In the second, we set the value of η to 1. We measure registration accuracy using SI over three labels, namely WM, GM and CSF. We ignore the two available tumour labels as there is no equivalent in the healthy scan. A good overlap for non-tumour labels is an indicator that the similarity measure is not biased by the presence of pathology.

Table 6.2: **Images with pathology:** Mean (Median) overlap measures and standard deviation for white matter (WM), grey matter (GM) and ventricular cerebrospinal fluid (CSF) labels propagated using **affine** registration.

		WM	**GM**	**CSF**	**Overall**
NMI		61.0(61.0) ± 0.01	64.2(64.6) ± 0.01	57.3(54.5) ± 0.08	60.8(60.1) ± 0.03
Proposed similarity	($\eta = 0.1$)	62.1(62.0) ± 0.00	64.2(64.6) ± 0.01	62.1(61.5) ± 0.06	62.8(62.7) ± 0.02
Haber et al. [74]	($\eta = 0.1$)	61.3(61.2) ± 0.01	64.1(64.4) ± 0.01	58.0(56.1) ± 0.07	61.1(60.6) ± 0.03
Proposed similarity	($\eta = 1$)	62.6(62.6) ± 0.01	64.8(65.1) ± 0.01	64.7(64.8) ± 0.05	64.0(64.1) ± 0.02
Haber et al. [74]	($\eta = 1$)	62.1(61.9) ± 0.01	64.7(64.9) ± 0.01	61.9(61.8) ± 0.07	62.9(63.0) ± 0.03

Table 6.3: **Images with pathology:** Mean (Median) overlap measures and standard deviation for white matter (WM), grey matter (GM) and ventricular cerebrospinal fluid (CSF) labels propagated using **non-rigid** registration.

		WM	**GM**	**CSF**	**Overall**
NMI		78.1(78.1) ± 0.03	79.0(78.0) ± 0.03	88.4(88.9) ± 0.02	81.8(81.3) ± 0.02
Proposed similarity	($\eta = 0.1$)	78.1(78.1) ± 0.00	76.4(76.6) ± 0.01	83.4(84.5) ± 0.03	79.3(79.9) ± 0.01
Haber et al. [74]	($\eta = 0.1$)	74.2(74.4) ± 0.01	75.1(75.0) ± 0.01	80.9(82.2) ± 0.03	76.8(77.2) ± 0.02
Proposed similarity	($\eta = 1$)	80.9(80.8) ± 0.00	79.1(79.3) ± 0.01	88.1(89.1) ± 0.03	82.7(83.3) ± 0.01
Haber et al. [74]	($\eta = 1$)	81.2(81.2) ± 0.00	79.6(79.7) ± 0.01	88.4(89.0) ± 0.03	83.1(83.7) ± 0.01

Table 6.2 shows the overlaps obtained using affine registration. We observe the best overall alignment for the proposed similarity measure, thus demonstating increased robustness against the presence of tumours. When comparing the results of the proposed similarity measure with either NMI or cosine squared similarity measure, we observe that the improvements are statistically significant with 99% confidence ($p < 0.01$), both with $\eta = 0.1$ and $\eta = 1$.

Non-rigid registration

To evaluate the proposed similarity in the context of non-rigid registration of pathological images, we first utilise the BraTS dataset and the same two settings as for affine registration. Overlap results for non-rigid registration are shown in Table 6.3. It can be observed that for the case where $\eta = 0.1$, the proposed similarity measure yields a worse alignment than NMI ($p = 0.0098$). However, when compared to the cosine squared similarity, a much better result is obtained for the proposed similarity measure ($p = 0.002$). Further visual results for this comparison are shown in Figure 6.7. The main areas where the registration using cosine squared similarity measure is affected by the tumour presence are highlighted by a red ellipse.

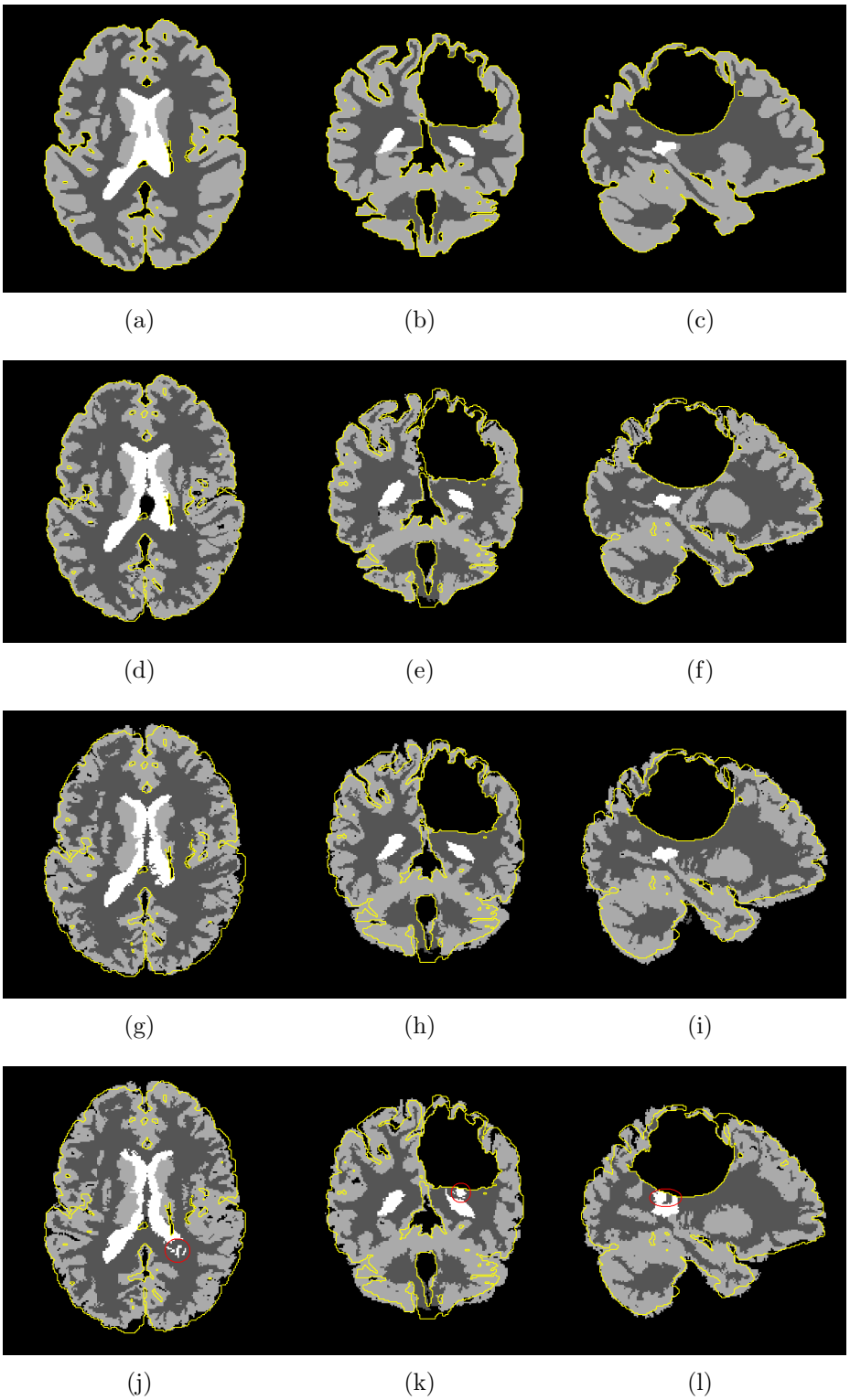
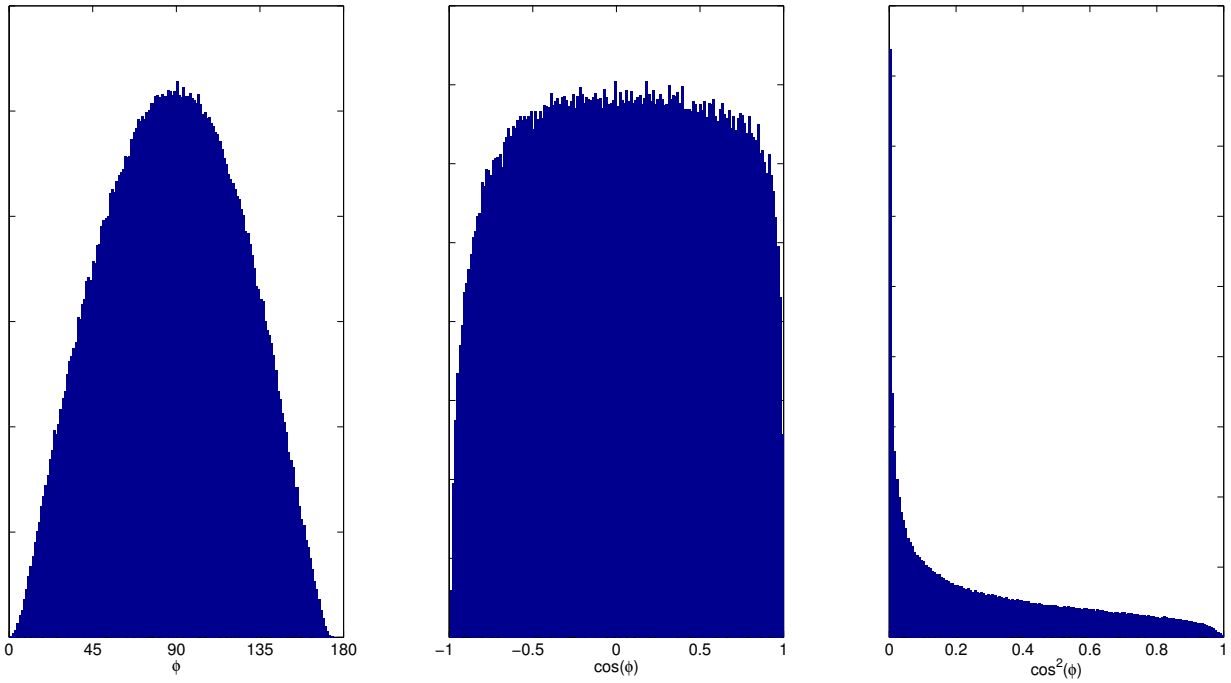


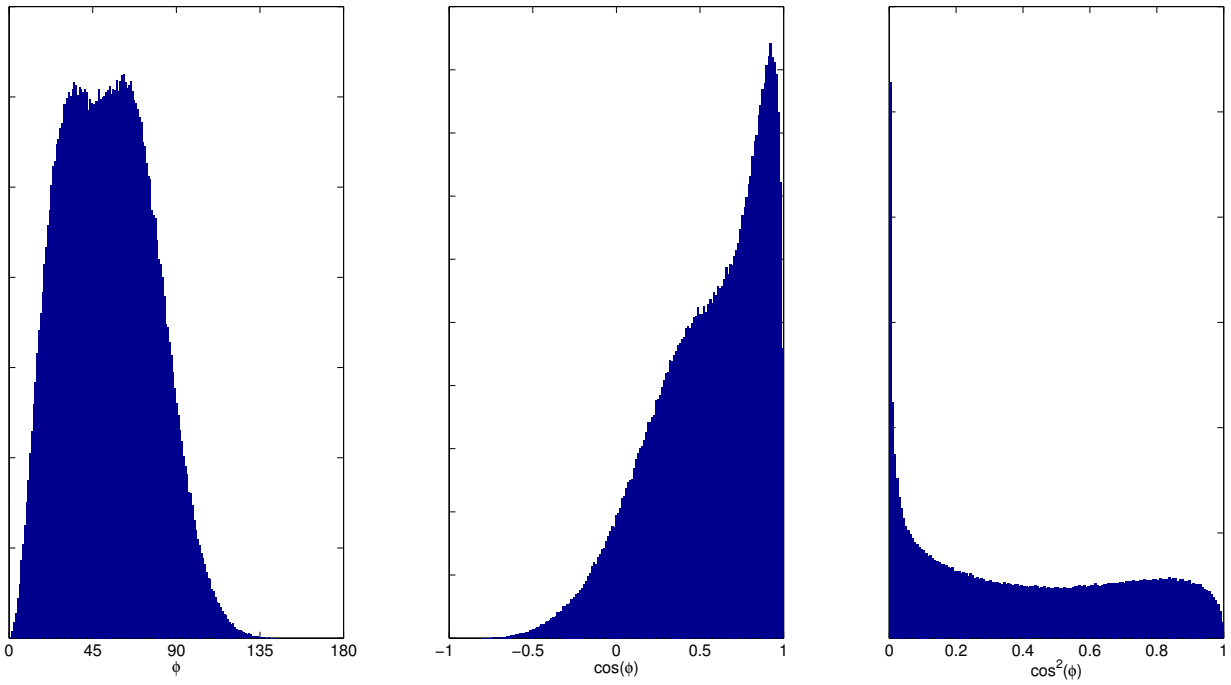
Figure 6.7: Reference and propagated labels using $\eta = 0.1$. (a)-(c): Reference. (d)-(f): Propagated using NMI. (g)-(i): Propagated using proposed similarity. (j)-(l): Propagated using Haber et al. [74]. Boundaries of the tumours and image are provided for visualisation.

On the other hand, when setting η to 1 the proposed similarity yields greater similarity indices than NMI, although this is not statistically significant ($p = 0.193$). The indices slightly less than those of the cosine squared similarity measure, and are also statistically significant ($p = 0.002$). The observation that the proposed similarity performs worse than cosine squared similarity measure in this case is explained by the fact that increasing values of η reduce the relative effect of the gradient components, thus hampering the cancellation property of the cosine similarity in areas of ambiguous correspondences. This effect is further supported by the experimental evidence shown in Figure 6.8. We observe that for $\eta = 0.1$, the histogram of values for $\cos(\phi)$ in the tumour areas is roughly symmetric within the range $[-1, 1]$, with a mean value of 0.031. On the other hand, the histogram for $\cos^2(\phi)$ has strictly positive values, with a mean of 0.296 hence biasing the energy computation. In the case of $\eta = 1$, the angles ϕ are skewed towards zero due to the fact that the values of the gradients become comparably much smaller than the values of ϱ and τ. This causes the mean of the histogram of $\cos(\phi)$ to become even greater than that of the histogram of $\cos^2(\phi)$. It is hence important to stress that when appropiate values of the parameter η are used (e.g. 0.1 or less) the robustness of the proposed similarity measure is met and the performance is good. In contrast, the performance of the cosine squared similarity measure is heavily affected by the change of this parameter.

In addition, we perform registrations using NMI, the proposed similarity measure and cosine squared similarity measure on TBI data. We utilise the baseline image as target and the followup image as source. For the gradient-based similarity measures we set η to 0.1. Visual results are given in Figure 6.9. The main observation is that, in contrast to NMI and the cosine squared similarity measure, the proposed measure is able to recover most of the underlying changes in shape within the pathology area (as pointed out with a red arrow). This is because the registration inside the area of pathology affect the value of both NMI and cosine squared similarity measure significantly, causing the optimisation to converge very quickly. On the other hand, the proposed similarity is almost unaffected by the forces in the area of pathology, allowing the optimisation to continue until good alignment is achieved.



(a)



(b)

Figure 6.8: Angle ϕ and histograms of $cos\phi$ and $cos^2\phi$ using (a) $\eta = 0.1$ and (b) $\eta = 1$ between a healthy subject and the BraTS simulated images in the tumour areas. The means are 0.031 and 0.296 respectively for $\eta = 0.1$ and 0.548 and 0.410 respectively for $\eta = 1$.

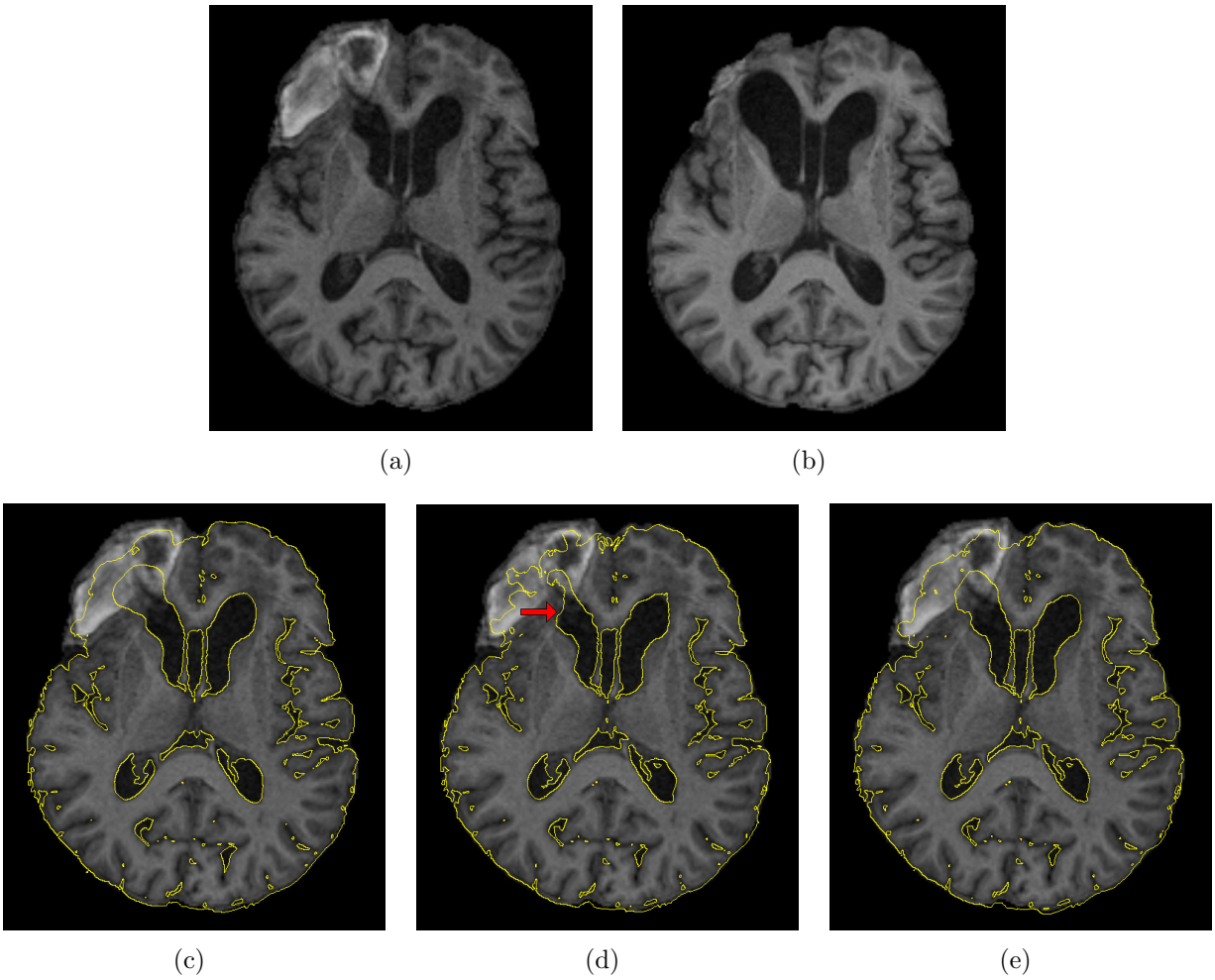


Figure 6.9: Visual comparison of the proposed similarity, normalised mutual information and cosine squared for TBI data. (a): Baseline image. (b): Followup image. (c): Transformed followup isolines using NMI registration result overlayed on the baseline image. (d): Transformed followup isolines using cosine registration result overlayed on the baseline image. (e): Transformed followup isolines using cosine squared registration result overlayed on the baseline image.

6.3.4 MR images with no outliers

We also evaluate the performance of NMI, the proposed similarity measure and cosine squared similarity measure using images from the ADNI dataset. We register all 820 images to the MNI single-subject T1-atlas [142] with control point spacings of 20, 10, 5 and 2.5mm. We measure the target registration error at 20 landmarks. The results are shown in Table 6.4. We can observe that the normalised gradient field based similarity measures perform significantly better than NMI in terms of the alignment of the landmarks ($p < 10^{-5}$ in both cases). The

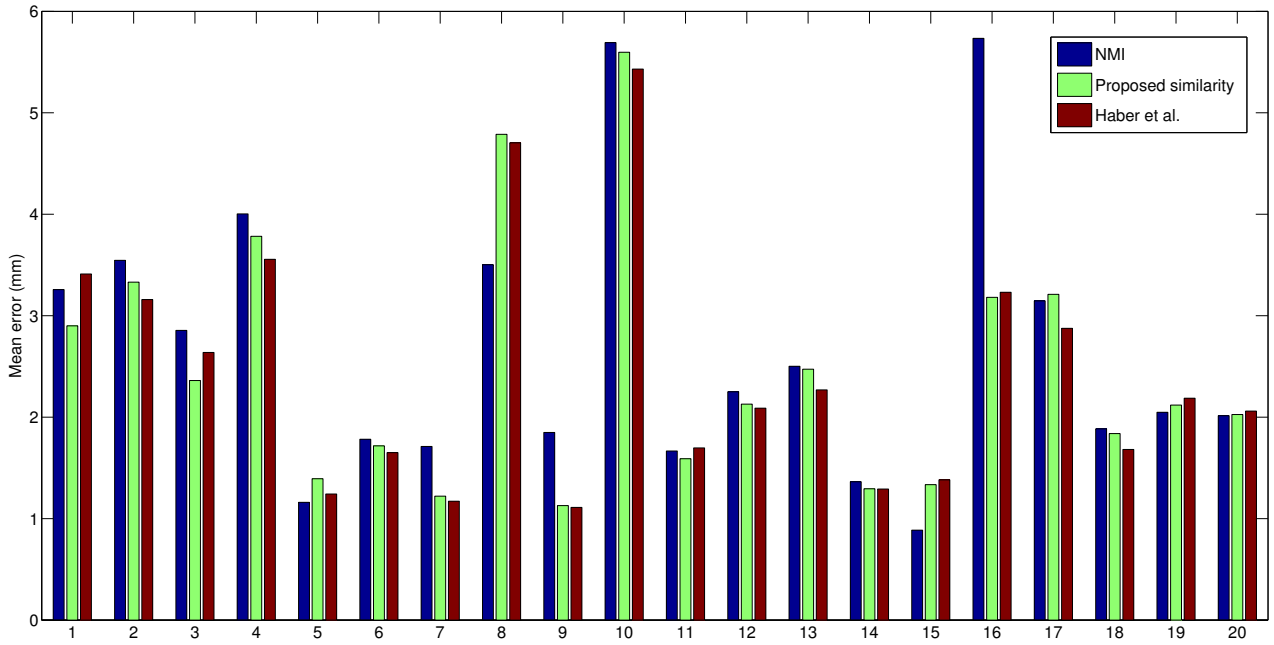


Figure 6.10: Mean errors over all ADNI images for each individual landmark. We show results for normalised mutual information, the proposed similarity and the similarity by Haber et al. [74].

cosine squared similarity performs slightly better than the proposed similarity, with statistically significant differences ($p < 10^{-5}$). To further investigate this result, we assess the errors for each landmark individually. These errors are show in Figure 6.10. It can be observed that the normalised gradient field based similarity measures have little difference in general. On the other hand, normalised mutual information only performs noticeably better than the other similarity measures for landmark number 8 (superior aspect of the cerebellum) with a difference in performance of more than 1mm. This is explained by the fact that this landmark is located in an area where the folia of the cerebellum are more visible and thus image intensity gradients are more noisy.

Table 6.4: Mean (Median) target registration errors and standard deviation over all ADNI images for the proposed similarity, the similarity by Haber et al. [74] and normalised mutual information. Since the image have isotropic voxel size of 1mm, the errors can be considered to be either in mm or in voxels.

	Target registration error
NMI	$2.64(2.54) \pm 0.55$
Proposed similarity	$2.47(2.42) \pm 0.42$
Haber et al. [74]	$2.44(2.34) \pm 0.47$

Finally, we assess the behaviour of the proposed similarity in cardiac MR images. We register the end-systole frame to the end-diastole frame of each subject with control point spacings of 20, 10, and 5mm using normalised mutual information, the proposed similarity measure and the cosine squared similarity measure. We employ a noise parameter $\eta = 0.1$ for both normalised gradient field similarity measures. To evaluate the quality of the registration, we measure similarity indices after propagating the three cardiac labels (see Figure 6.11 for the results). The first observation that we make is that both normalised gradient based similarity measures perform comparatively well in the right ventricle with respect to normalised mutual information, with the proposed similarity measure being more consistent than cosine squared. Nonetheless, these similarity measures have a considerably worse performance in both the cavity and myocardium of the left ventricle. To illustrate the reason behind this issue, we show a visual example of registration results in Figure 6.12. We can infer from the figure that the main problem is the fact that papillary muscles in the target image produces strong gradients that confuse both cosine-based similarity measures, causing them to align the boundary of the myocardium to these muscles. Furthermore, the size of these muscles is not sufficient as to allow the proposed similarity measure to detect them as potential outliers. Finally, intensity gradients produced by the blood flow within the cavity of the left ventricle seem to have an important influence in the cosine squared similarity measure. However, this is not the case for the proposed similarity measure.

6.4 Summary

In this chapter, we have proposed a similarity measure for medical image registration that is robust towards bias fields and outliers in form of pathologies. We demonstrated the effectiveness and robustness of our similarity measure on MR brain images with simulated bias fields and on images with both simulated and real pathology, showing improved robustness compared to NMI and the cosine squared similarity measure of Haber et al. [74], especially for affine registration and when using appropiate values for the noise parameter η. We also evaluated

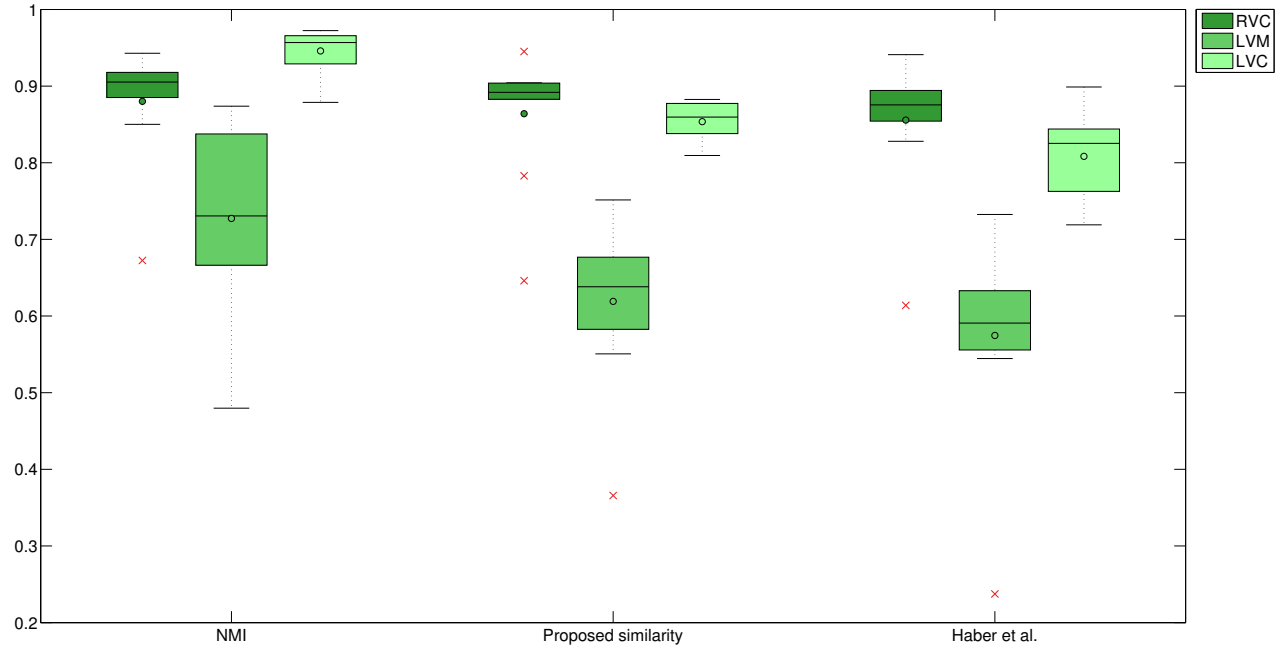


Figure 6.11: Results on cardiac data measuring Right ventricle cavity (RVC), Left ventricle myocardium (LVM) and Left ventricle cavity (LVC) similarity indices. We show results for normalised mutual information, the proposed similarity and the similarity by Haber et al. [74].

the proposed similarity measure in landmark-annotated images, showing that the proposed gradient-based similarity and the gradient-based cosine squared similarity measure show better performance than of normalised mutual information in terms of target registration error. Finally, an evaluation in cardiac MR images shows that the presence of the papillary muscle and the gradients produced by the flow of blood inside the heart cavities can affect the performance of gradient-based similarity measures and, hence, normalised mutual information seem more appropiate in that case.

The main contribution of this chapter is that our similarity measure relaxes the need for using preprocessing steps like bias field correction, which can be time consuming and prone to errors. This is especially true for images depicting organs different from the brain, since most intensity inhomogeneity correction methods are tailored toward particular images (e.g. images of the brain). Morover, the proposed similarity measure can also be utilised to register images in the presence of pathologies, without the need for relying on any particular deformation model and without requiring segmentations of the outliers.

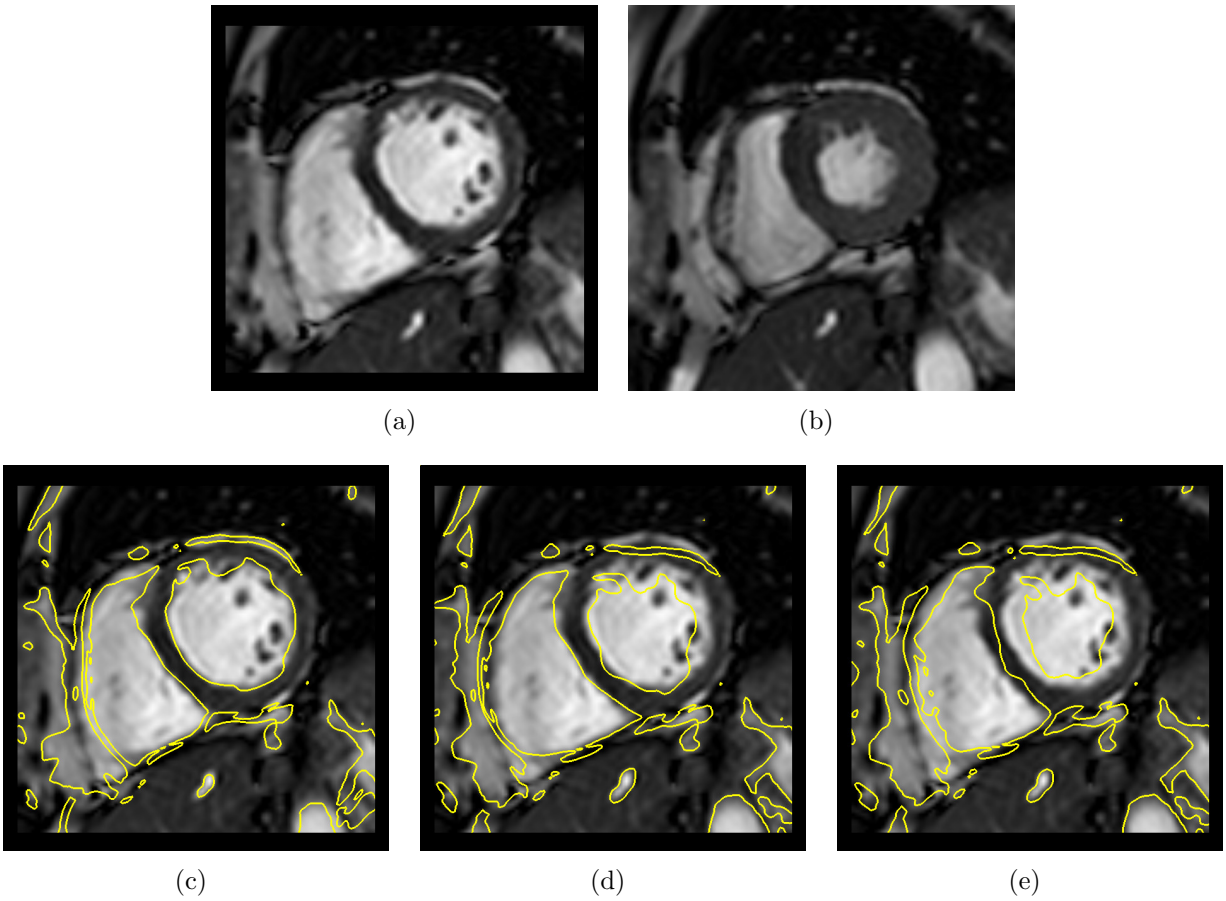


Figure 6.12: Visual comparison of the proposed similarity, normalised mutual information and cosine squared for cardiac data. (a): Target image. (b): Source image. (c): Transformed source isolines using NMI registration result overlayed on the target image. (d): Transformed source isolines using cosine registration result overlayed on the target image. (e): Transformed source isolines using cosine squared registration result overlayed on the target image.

As future work, we plan to investigate possible extensions to the proposed method, in order to be able to deal with multimodal registrations tasks such as T1-T2 MRI registration. Another important avenue of work is to investigate about alternative ways of utilising normalised gradient fields in order to reduce or eliminate the influence of the noise parameter η in the robustness against ambiguous correspondences. Preliminary results show that computing the cube of cosine (cos^3) from the normalised gradient field might lead to improved registration results with high levels of noise filtering (e.g. $\eta = 1$), without compromising the ability of being unbiased with respect to these kind of unpredictable correspondences. This is mainly because cos^3 tends to have a more symmetric histogram than cos.

Chapter 7

Conclusion

7.1 Summary

This thesis has proposed new methods that tackle some of the challenges of medical image registration. All of the proposed approaches are based upon the free-form deformation (FFD) framework of Rueckert et al. [183]. We focused particularly on providing techniques that are not only comparably accurate to state-of-the-art techniques, but are also robust and computationally efficient. The evaluation of the described methods has been carried out on several datasets of 3D MR images of the brain, including images annotated with anatomical landmarks, images with segmentations of anatomical structures, images of patients with traumatic brain injury and synthetic images depicting brain tumours. We also demonstrate the proposed approaches on a dataset of $2D + t$ MR images of the heart.

7.2 Contributions

Chapter 4 presents a learning-based statistical registration method, which extends the FFD transformation model to incorporate information in the form of a statistical deformation model.

This SDM is learnt a priori from a number of known deformations of medical images of a population to a given common template space. It also accounts for the mean and variability of deformations across this population and constrains the set of possible transformations to be in statistical concordance with the training data. The proposed registration method is also considerably faster than most standard registration methods, since only a single-resolution optimisation is needed to account for coarse and fine local displacements. This is in contrast to the multi-resolution strategy employed by the FFD-based registration. Another important feature of the proposed registration method is that the transformations are parameterised by significantly fewer degrees of freedom than standard FFD transformations. Finally, the learning stage of our framework can take advantage of annotated information to enforce plausible alignment of certain anatomical structures. We focused on annotations of anatomical landmarks, although other type of annotations such as segmentations in the form of label maps can be used as well. Experiments on 3D MR brain data and $2D + t$ MR sequences of the heart show that the proposed approach is able to perform with a similar accuracy as the standard FFD approach, but with lower computational cost.

In chapter 5, a learning-based approach to improve the convergence of the optimisation of any chosen parametric image registration method is proposed. This method computes a novel learning-based estimate of the gradient that is able to yield better registration accuracy than standard analytical methods, by aiding the optimisation method to avoid local minima. This is achieved by learning offline a projection from the gradient of the similarity measure in voxel space to the parameter space of the chosen registration method using PCA dimensionality reduction and standard least squares regression. The learnt projection is subsequently used online to approximate the optimisation of the energy functional for unseen images. We employ the B-spline FFD approach as underlying registration method, but any other parametric registration method can be used instead. We perform experiments on 3D brain MRI and $2D + t$ cardiac MRI sequences to show that our approach can improve the accuracy of the FFD registration method.

Chapter 6 introduces a novel similarity measure based on normalised gradients for non-rigid registration. The main contribution is that the proposed measure, relaxes the need of using preprocessing steps like bias field correction, which can be time consuming and prone to errors. This is especially true for images depicting organs different from the brain, since most intensity inhomogeneity correction methods are tailored to the intensity profiles of that particular organ. Moreover, the proposed similarity measure can also be utilised to register images in the presence of pathologies, without the need to rely on any particluar deformation model and without requiring segmentations of the outliers. We provide both theoretical and experimental proof of the robustness using 3D MR images of the brain, both on healthy subjects and subjects suffering from traumatic brain injury or brain tumours.

7.3 Limitations and future work

There are a number of possible avenues to explore in extending the proposed methods and tackling their limitations. For example, an extension to chapter 4 can be the use of cortical surface information or anatomical segmentations to learn a better statistical deformation model for brain data. This may help overcome the limitation that the cortical anatomy is very variable across different subjects. Another possible development is the use of sparsity constraints when learning the SDM in order to be able to characterise the way in which different regions of the training data are deformed. This would enable the analysis of deformations in a non-holistic manner, which can potentially help to determine disease progression for specific areas in the images. A limitation of the method proposed in chapter 5 is that, even though the learning process must be performed only once, it is very memory consuming. This is due to the fact that we extract a large number of training patches from several images, and these patches must all be aggregated into a matrix to perform PCA. This may restrict the number of training images that can be used (around 300GB may be needed for 1000 3D training images). One way of alleviateing this problem is by taking only a subset of the available patches. This subset may be defined such as it does not contain outliers or is defined only for salient regions of the

training data. However, care must be taken in order to not reduce the amount of patches to a point of having insufficient training information to learn sensible projections. One of the main limitations of the novel similarity measure of chapter 6 is that it is not designed to work in a muti-modal context. A possible way of mitigating this problem is to investigate the feasibility of a preprocessing step in either of the images being registerd in order to ensure that gradients in corresponding areas point in the same direction. Another limitation is that ability of the proposed similarity to be robust to ambiguous correspondences is hampered when the value of the noise parameter increases. This is problematic since usually large values of this noise parameter are needed to render gradient field similarities insensitive to noise. Preliminary results show that computing the cube of cosine (cos^3) from the normalised gradient field might lead to improved registration results with high levels of noise filtering, without compromising the ability of being unbiased with respect to these kind of unpredictable correspondences.

Bibliography

[1] N. Ablitt, J. Gao, J. Keegan, L. Stegger, D.N. Firmin, and G.Z. Yang. Predictive cardiac motion modeling and correction with partial least squares regression. *IEEE Trans. Med. Imaging*, 23(10):1315–24, 2004. 108

[2] R. Acar and C.R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Probl.*, 1217, 1994. 42

[3] R.J. Amdur, D. Gladstone, K.A. Leopold, and R.D. Harris. Prostate seed implant quality assessment using MR and CT image fusion. *Int. J. Radiat. Oncol. Biol. Phys.*, 43(1): 67–72, 1999. 4

[4] Y. Amit, U. Grenander, and M. Piccioni. Structural image restoration through deformable templates. *J. Am. Stat. Assoc.*, 86(414):376–387, 1991. 39

[5] C. Anderson, M. Koshy, C. Staley, N. Esiashvili, S. Ghavidel, Z. Fowler, T. Fox, F. Esteves, J. Landry, and K. Godette. PET-CT fusion in radiation management of patients with anorectal tumors. *Int. J. Radiat. Oncol. Biol. Phys.*, 69(1):155–62, 2007. 4

[6] E.D. Angelini, J. Delon, A.B. Bah, L. Capelle, and E. Mandonnet. Differential MRI analysis for quantification of low grade glioma growth. *Med. Image Anal.*, 16(1):114–26, 2012. 5

[7] B.A. Ardekani, S. Guckemus, A. Bachman, M.J. Hoptman, M. Wojtaszek, and J. Nierenberg. Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans. *J. Neurosci. Methods*, 142(1):67–76, 2005. 128

[8] V. Arsigny and O. Commowick. A fast and log-euclidean polyaffine framework for locally linear registration. *J. Math. Imaging Vis.*, 33(2):222–238, 2009. 32

[9] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-Euclidean framework for statistics on diffeomorphisms. *Med. image Comput. Comput. Interv.*, 9(Pt 1):924–31, 2006. 35, 36, 57

[10] V. Arsigny, X. Pennec, and N. Ayache. Polyrigid and polyaffine transformations: a novel geometrical tool to deal with non-rigid deformations - application to the registration of histological slices. *Med. Image Anal.*, 9(6):507–23, 2005. 32

[11] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano. Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans. Med. Imaging*, 28(8):1266–77, 2009. 7

[12] U.M. Ascher, E. Haber, and H. Huang. On effective methods for implicit piecewise smooth surface recovery. *SIAM J. Sci. Comput.*, 28:1–20, 2006. 130

[13] J. Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1): 95–113, 2007. 51, 57, 69

[14] J. Ashburner and K.J. Friston. Nonlinear spatial normalization using basis functions. *Hum. Brain Mapp.*, 266:254–266, 1999. 31

[15] J. Ashburner and K.J. Friston. Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *Neuroimage*, 55(3):954–67, 2011. 50

[16] B.B. Avants, C.L. Epstein, M. Grossman, and J.C. Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.*, 12(1):26–41, 2008. 44, 59, 69

[17] B.B. Avants, P.T. Schoenemann, and J.C. Gee. Landmark and intensity-driven Lagrangian frame diffeomorphic image registration: Application to structurally and func-

tionally based inter-species comparison. *Med. Image Anal. Spec. issue WBIR 2003*, 10: 397–412, 2006. 59

[18] R.K. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Comput. vision, Graph. image Process.*, 1(0224):1–2, 1989. 37, 108

[19] S. Baker, R. Gross, T. Ishikawa, and I. Matthews. Lucas-kanade 20 years on: A unifying framework: Part 2. *Carnegie Mellon Univ. Tech. Rep.*, 2003. 128

[20] V. Barra and J. Boire. A general framework for the fusion of anatomical and functional medical images. *Neuroimage*, 13(3):410–24, 2001. 4

[21] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *Int. J. Comput. Vis.*, 1994. 55, 56

[22] N.K. Batmanghelich, B. Taskar, and C. Davatzikos. Generative-discriminative basis learning for medical imaging. *IEEE Trans. Med. Imaging*, 31(1):51–69, 2012. 108

[23] K.G. Baum, K. Rafferty, M. Helguera, D.H. Feiglin, and A. Krol. Investigation of PET/MRI image fusion schemes for enhanced breast cancer diagnosis. *IEEE Nucl. Sci. Symp. Conf. Rec.*, pages 3774–3780, 2007. 4

[24] M.F. Beg, M.I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.*, 61(2):139–157, 2005. 43, 50, 56, 69

[25] S. Benameur, M. Mignotte, S. Parent, H. Labelle, W. Skalli, and J. de Guise. 3D/2D registration and segmentation of scoliotic vertebrae using statistical models. In *Comput. Med. Imaging Graph.*, volume 27, pages 321–337, 2003. 72

[26] F.L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1989. 26, 39

[27] F.L. Bookstein. Thin-plate splines and the atlas problem for biomedical images. *Inf. Process. Med. Imaging*, 1991. 26

[28] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001. 52

[29] M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. *Vis. Biomed. Comput.*, 1996. 38

[30] C. Broit. *Optimal registration of deformed images.* Ph.d. thesis, University of Pennsylvania, 1981. 37

[31] C. Buerger, T. Schaeffter, and A.P. King. Hierarchical adaptive local affine registration for fast and robust respiratory motion estimation. *Med. Image Anal.*, 15(4):551–64, 2011. 31

[32] P. Cathier. Iconic feature registration with sparse wavelet coefficients. In *Med. image Comput. Comput. Interv.*, pages 694–701, 2006. 31

[33] P. Charbonnier. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Int. Conf. Image Process.*, number 3, pages 168–172, 1994. 42

[34] A.S. Chowdhury, R. Roy, S.Kr. Bose, F. Khalifa, A. Elnakib, and A. El-Baz. Non-rigid biomedical image registration using graph cuts with a novel data term. In *Int. Symp. Biomed. Imaging*, pages 446–449, 2012. 51

[35] G.E. Christensen and H.J. Johnson. Consistent image registration. *IEEE Trans. Med. Imaging*, 20(7):568–82, 2001. 58, 60, 61, 69

[36] G.E. Christensen, R.D. Rabbitt, and M.I. Miller. Deformable templates using large deformation kinematics. *IEEE Image Process.*, 5(10):1435–1447, 1996. 38, 108

[37] D. Cobzas and A. Sen. Random walks for deformable image registration. In *Med. image Comput. Comput. Interv.*, pages 557–565, 2011. 51

[38] O. Commowick, V. Arsigny, A. Isambert, J. Costa, F. Dhermain, F. Bidault, P-Y. Bondiau, N. Ayache, and G. Malandain. An efficient locally affine framework for the smooth registration of anatomical structures. *Med. Image Anal.*, 12(4):427–41, 2008. 31

[39] P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3): 287–314, 1994. 75

[40] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models-their training and application. *Comput. Vis. image Underst.*, 61(1):38–59, 1995. 75

[41] T.F. Cootes, C.J. Twining, V.S. Petrović, K.O. Babalola, and C.J. Taylor. Computing accurate correspondences across groups of images. *Pattern Anal. Mach. Intell.*, 1(1):1–12, 2010. 31

[42] M.M. Coselmon, J.M. Balter, D.L. McShan, and M.L. Kessler. Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines. *Med. Phys.*, 31(11):2942, 2004. 26

[43] R. Courant and D. Hilbert. *Methods of mathematical physics*. 1966. 39

[44] T.M. Cover and J.A. Thomas. *Elements of information theory*. 1991. 76

[45] C. Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *Comput. Vis. image Underst.*, 66(2):207–22, 1997. 37

[46] J. Declerck, J. Feldmar, M.L. Goris, and F. Betting. Automatic registration and alignment on a template of cardiac stress and rest reoriented SPECT images. *IEEE Trans. Med. Imaging*, 16(6):727–37, 1997. 30

[47] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. J. Comput. Vis.*, 187:167–187, 1987. 56

[48] C. Ding, T. Li, and M.I. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1):45–55, 2010. 78

[49] M. Droske and M. Rumpf. A variational approach to nonrigid morphological image registration. *SIAM J. Appl. Math.*, 64(2):668–687, 2004. 50

[50] N. Duchateau, M. De Craene, G. Piella, E. Silva, A. Doltra, M. Sitges, B.H. Bijnens, and A.F. Frangi. A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. *Med. Image Anal.*, 15(3):316–28, 2011. 5

[51] J. Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. *Constr. theory Funct. Sev. Var.*, 1977. 25

[52] A. El-Baz, A. Farag, G. Gimel'farb, and A.E. Abdel-Hakim. Image alignment using learning prior appearance model. In *Int. Conf. Image Process.*, pages 341–344, 2006. 128

[53] G.D. Evangelidis and E.Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(10):1858–65, 2008. 128

[54] B. Fei, Z. Lee, D.T. Boll, and J.L. Duerk. Image registration and fusion for interventional MRI guided thermal ablation of the prostate cancer. *Med. image Comput. Comput. Interv.*, pages 364–372, 2003. 4

[55] W. Feng, S.J. Reeves, and T.S. Denney. A new consistent image registration formulation with a B-spline deformation model. In *Int. Symp. Biomed. Imaging*, pages 979–982, 2009. 30

[56] B. Fischer and J. Modersitzki. Combination of automatic non-rigid and landmark based registration: the best of both worlds. In Milan Sonka and J. Michael Fitzpatrick, editors, *Proc. SPIE Med. Imaging*, volume 5032, pages 1037–1048, 2003. 72

[57] R. Fletcher and C.M. Reeves. Function minimization by conjugate gradients. *Comput. J.*, 7(2):149–154, 1964. 50

[58] O. Fluck, C. Vetter, W. Wein, A. Kamen, B. Preim, and R. Westermann. A survey of medical image registration on graphics hardware. *Comput. Methods Programs Biomed.*, 104(3):45–57, 2011. 9

[59] L.R. Ford and D.R. Fulkerson. *Flows in networks*. Princeton U. Press, 1962. 52

[60] M. Fornefett, K. Rohr, and H.S. Stiehl. Radial basis functions with compact support for elastic registration of medical images. *Image Vis. Comput.*, 19(1-2):87–96, 2001. 28

[61] D.R. Forsey and R.H. Bartels. Hierarchical B-spline refinement. *ACM SIGGRAPH Comput. Graph.*, 22(4):205–212, 1988. 31

[62] B.J. Frey and D.J.C. MacKay. A revolution: Belief propagation in graphs with cycles. *Adv. Neural Inf. Process. Syst.*, 1998. 52

[63] K. Friston, J. Ashburner, C.D. Frith, J.-B. Poline, J.D. Heather, and R.S.J. Frackowiak. Spatial registration and normalization of images. *Hum. Brain Mapp.*, (2):165–189, 1995. 128

[64] H. Fukunaga, M. Sekimoto, M. Ikeda, I. Higuchi, M. Yasui, I. Seshimo, O. Takayama, H. Yamamoto, M. Ohue, M. Tatsumi, J. Hatazawa, M. Ikenaga, T. Nishimura, and M. Monden. Fusion image of positron emission tomography and computed tomography for the diagnosis of local recurrence of rectal cancer. *Ann. Surg. Oncol.*, 12(7):561–9, 2005. 4

[65] J.C. Gee, D.R. Haynor, L.L. Briquer, and R.K. Bajcsy. Advances in elastic matching theory and its implementation. In *Jt. Conf. Comput. Vision, Virtual Real. Robot. Med. Med. Robot. Comput. Surg.*, 1997. 37

[66] S. Gefen, O. Tretiak, and J. Nissanov. Elastic 3-D alignment of rat brain histological images. *IEEE Trans. Med. Imaging*, 22(11):1480–1489, 2003. 31, 51

[67] J. Glaunès, M. Vaillant, and M.I. Miller. Landmark matching via large deformation diffeomorphisms on the sphere. *J. Math. Imaging Vis.*, pages 179–200, 2004. 50

[68] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios. Dense image registration through MRFs and efficient linear programming. *Med. Image Anal.*, 12(6):731–41, 2008. 51

[69] B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: continuous deformations meet efficient optimal linear programming. In *Inf. Process. Med. Imaging*, volume 20, pages 408–20, 2007. ISSN 1011-2499. 51

[70] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios. Deformable medical image registration: setting the state of the art with discrete methods. *Annu. Rev. Biomed. Eng.*, 13:219–44, 2011. 51

[71] G. Grabner, A.L. Janke, M.M. Budge, D. Smith, J. Pruessner, and D.L. Collins. Symmetric Atlasing and Model Based Segmentation : An Application to the Hippocampus in Older Adults. In *Med. image Comput. Comput. Interv.*, pages 58–66, 2006. 83

[72] G. Grabner, I. Nöbauer, K. Elandt, C. Kronnerwetter, A. Woehrer, C. Marosi, D. Prayer, S. Trattnig, and M. Preusser. Longitudinal brain imaging of five malignant glioma patients treated with bevacizumab using susceptibility-weighted magnetic resonance imaging at 7T. *Magn. Reson. Imaging*, 30(1):139–47, 2012. 5

[73] C. Guetter, C. Xu, F. Sauer, and J. Hornegger. Learning based non-rigid multi-modal image registration using Kullback-Leibler divergence. In *Med. image Comput. Comput. Interv.*, volume 8, pages 255–62, 2005. 64, 69

[74] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. In *Med. image Comput. Comput. Interv.*, volume 46, pages 292–9, 2006. xiv, xxii, xxiii, 128, 129, 132, 133, 140, 141, 142, 146, 147, 148

[75] M. Hachama, A. Desolneux, and F.J.P. Richard. Bayesian Technique for Image Classifying Registration. *IEEE Trans. image Process.*, 21(9):4080–4091, 2012. 128

[76] J Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princet. Univ. Bull.*, 1902. 10

[77] J.V. Hajnal, D.L.G. Hill, and D.J. Hawkes. *Medical image registration*. CRC, 2001. xvi, 19, 43

[78] W.R. Hamilton. II. On quaternions; or on a new system of imaginaries in algebra. *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, 25(163):10–13, 1844. 22

[79] J. Hamm, D.H. Ye, R. Verma, and C. Davatzikos. GRAM: A framework for geodesic registration on anatomical manifolds. *Med. Image Anal.*, 14(5):633–42, 2010. 68, 69

[80] P.L. Hammer, P. Hansen, and B. Simeone. Roof duality, complementation and persistency in quadratic 01 optimization. *Math. Program.*, 28:121–155, 1984. 52

[81] J.A. Hartigan and M.A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Appl. Stat.*, 28(1):100–108, 1979. 78

[82] T. Hartkens and D.L.G. Hill. Using points and surfaces to improve voxel-based non-rigid registration. In *Med. image Comput. Comput. Interv.*, pages 565–572, 2002. 72

[83] R.A. Heckemann, J.V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage*, 33(1):115–26, 2006. 7

[84] M.P. Heinrich, M. Jenkinson, J.M. Brady, and J. Schnabel. Globally optimal deformable registration on a minimum spanning tree using dense displacement sampling. *Med. image Comput. Comput. Interv.*, 15(Pt 3):115–22, 2012. 51

[85] M.P. Heinrich, M. Jenkinson, J.M. Brady, and J. Schnabel. MRF-based deformable registration and ventilation estimation of lung CT. *IEEE Trans. Med. Imaging*, 32(7): 1239–48, 2013. 51

[86] M.P. Heinrich, J. Schnabel, and F.V. Gleeson. Non-rigid multimodal medical image registration using optical flow and gradient orientation. *Proc. Med. Image Underst. Anal.*, pages 1–5, 2010. 128

[87] S. Heldmann and S. Zidowitz. Elastic registration of multiphase CT images of liver. In David R. Haynor and Sébastien Ourselin, editors, *Proc. SPIE Med. Imaging*, 2009. 128

[88] P. Hellier, C. Barillot, E. Mémin, and P. Pérez. Hierarchical estimation of a dense deformation field for 3-D robust registration. *IEEE Trans. Med. Imaging*, 20(5):388–402, 2001. 31

[89] G. Hermosillo and O. Faugeras. Dense image matching with global and local statistical criteria: a variational approach. In *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pages 73–78, 2001. 128

[90] M. Hernandez, M.N. Bossa, and S. Olmos. Registration of anatomical images using geodesic paths of diffeomorphisms parameterized with stationary vector fields. *Int. Conf. Comput. Vis.*, pages 1–8, 2007. 58

[91] M. Hernandez, M.N. Bossa, and S. Olmos. Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *Int. J. Comput. Vis.*, 85(3):291–306, 2009. 58

[92] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand. (1934).*, 49(6):409, 1952. 50

[93] W. Hoffmann. Iterative algorithms for Gram-Schmidt orthogonalization. *Computing*, 41 (4):335–348, 1989. 78

[94] C.K. Hoh, M. Dahlbom, G. Harris, Y. Choi, R.A. Hawkins, M.E. Phelps, and J. Maddahi. Automated iterative three-dimensional registration of positron emission tomography images. *J. Nucl. Med.*, 34(11):2009–18, 1993. 43

[95] P.W. Holland and R.E. Welsch. Robust regression using iteratively reweighted least-squares. *Commun. Stat. - Theory Methods*, 6(9):813–827, 1977. 42, 81

[96] B.K.P. Horn and B.G. Schunck. Determining optical flow. *Artif. Intell.*, 17(1-3):185–203, 1981. 55, 56, 69, 108

[97] W.M. Hsu, J.F. Hughes, and H. Kaufman. Direct manipulation of free-form deformations. *ACM SIGGRAPH Comput. Graph.*, 26(2):177–184, 1992. 51

[98] P. Huber and E. Ronchetti. *Robust statistics*. John Wiley & Sons, 1981. 81

[99] A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. neural networks*, 10(3):626–34, 1999. 76, 77

[100] A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–30, 2000. ISSN 0893-6080. 76, 77

[101] H. Ishikawa. Higher-order clique reduction in binary graph cut. *Conf. Comput. Vis. Pattern Recognit.*, pages 2993–3000, 2009. 52

[102] H. Ishikawa. Transformation of general binary MRF minimization to the first-order case. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(6):1234–49, 2011. 52

[103] C.R. Jack, H.J. Wiste, P. Vemuri, S.D. Weigand, M.L. Senjem, G. Zeng, M.A. Bernstein, J.L. Gunter, V.S. Pankratz, P.S. Aisen, M.W. Weiner, R.C. Petersen, L.M. Shaw, J.Q. Trojanowski, and D.S. Knopman. Brain beta-amyloid measures and magnetic resonance imaging atrophy both predict time-to-progression from mild cognitive impairment to Alzheimer's disease. *Brain*, 133(11):3336–48, 2010. 5

[104] A.P. James and B.V. Dasarathy. Medical image fusion: A survey of the state of the art. *Inf. Fusion*, 19:4–19, 2014. 4

[105] H.J. Johnson and G.E. Christensen. Consistent landmark and intensity-based image registration. *IEEE Trans. Med. Imaging*, 21(5):450–61, 2002. 50, 72

[106] I. Jolliffe. *Principal Component Analysis*. John Wiley & Sons, 2005. 74

[107] A.A. Joshi, D.W. Shattuck, P.M. Thompson, and R.M. Leahy. Surface-Constrained Volumetric Brain Registration Using Harmonic Mappings. *IEEE Trans. Med. Imaging*, 26 (12):1657–1669, 2007. 50

[108] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991. 75

[109] S. Katyal, E.L. Kramer, and M.E. Noz. Fusion of immunoscintigraphy single photon emission computed tomography (SPECT) with CT of the chest in patients with non-small cell lung cancer. *Cancer Res.*, 1995. 4

[110] M.J. Kim, M.H. Kim, and D. Shen. Learning-based deformation estimation for fast non-rigid registration. *Conf. Comput. Vis. Pattern Recognit.*, 2008. 72

[111] M.J. Kim, G. Wu, P. Yap, and D. Shen. A general fast registration framework by learning deformation-appearance correlation. *IEEE Trans. image Process.*, 21(4):1823–33, 2012. 66, 69, 108

[112] R. Kindermann and J.L. Snell. *Markov random fields and their applications*. 1980. 51

[113] S. Klein, M. Staring, P. Andersson, and J.P.W. Pluim. Preconditioned stochastic gradient descent optimisation for monomodal image registration. In *Med. image Comput. Comput. Interv.*, number 1, pages 549–556, 2011. 108

[114] S. Klein, U.A. van der Heide, I.M. Lips, M. van Vulpen, M. Staring, and J.P. W. Pluim. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Med. Phys.*, 35(4):1407, 2008. 128

[115] N. Komodakis and G. Tziritas. Approximate labeling via graph cuts based on linear programming. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(8):1436–53, 2007. 52

[116] N. Komodakis, G. Tziritas, and N. Paragios. Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies. *Comput. Vis. image Underst.*, 112(1):14–29, 2008. 52

[117] M. Krengli, A. Ballarè, B. Cannillo, M. Rudoni, E. Kocjancic, G. Loi, M. Brambilla, E. Inglese, and B. Frea. Potential advantage of studying the lymphatic drainage by

sentinel node technique and SPECT-CT image fusion for pelvic irradiation of prostate cancer. *Int. J. Radiat. Oncol. Biol. Phys.*, 66(4):1100–4, 2006. 4

[118] D. Kwon, M. Niethammer, H. Akbari, M. Bilello, C. Davatzikos, and K.M. Pohl. PORTR: Pre-operative and post-recurrence brain tumor registration. *IEEE Trans. Med. Imaging*, 33(3):651–667, 2014. 128

[119] D. Kwon, I.D. Yun, K.M. Pohl, C. Davatzikos, and S.U. Lee. Nonrigid volume registration using second-order MRF model. In *Int. Symp. Biomed. Imaging*, number 2, pages 708–711, 2012. 51

[120] J. Kybic and M. Unser. Fast parametric elastic image registration. *IEEE Trans. image Process.*, 12(11):1427–42, 2003. 30

[121] T. Lange, N. Papenberg, S. Heldmann, J. Modersitzki, B. Fischer, H. Lamecker, and P.M. Schlag. 3D ultrasound-CT registration of the liver using combined landmark-intensity information. *Int. J. Comput. Assist. Radiol. Surg.*, 4(1):79–88, 2009. 128

[122] M. Lázaro-Gredilla, J. Quiñonero Candela, and A. Figueiras-Vidal. Sparse spectral sampling gaussian processes. Technical report, Microsoft Research, 2007. 65

[123] C. Ledig, R. Wolz, P. Aljabar, J. Lötjönen, R.A. Heckemann, A. Hammers, and D. Rueckert. Multi-class brain segmentation using atlas propagation and EM-based refinement. In *Int. Symp. Biomed. Imaging*, pages 896–899, 2012. 7

[124] D.D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(October):788–791, 1999. 77

[125] S.Y. Lee, G. Wolberg, and S.Y. Shin. Scattered data interpolation with multilevel B-splines. *IEEE Trans. Vis. Comput. Graph.*, 3(3):228–244, 1997. 29

[126] L. Lemieux, R. Jagoe, D.R. Fish, N.D. Kitchen, and D.G.T. Thomas. A patienttocomputedtomography image registration method based on digitally reconstructed radiographs. *Med. Phys.*, 21(11):1749, 1994. 43

[127] V. Lempitsky, C. Rother, and A. Blake. Logcut-efficient graph cut optimization for markov random fields. *Int. Conf. Comput. Vis.*, pages 1–8, 2007. 52

[128] V. Lempitsky, C. Rother, S. Roth, and A. Blake. Fusion moves for Markov random field optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(8):1392–405, 2010. 52

[129] H. Lester, S.R. Arridge, and K.M. Jansons. Non-linear registration with the variable viscosity fluid algorithm. In *Inf. Process. Med. Imaging*, pages 238–251, 1999. 38

[130] C.H. Lo and H.S. Don. 3-D moment forms: their construction and application to object identification and positioning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(10):1053–1064, 1989. 63

[131] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens. Temporal subtraction of thorax CR images using a statistical deformation model. *IEEE Trans. Med. Imaging*, 22(11): 1490–1504, 2003. 67, 69, 71, 107

[132] D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens. Nonrigid image registration using free-form deformations with a local rigidity constraint. In *Med. image Comput. Comput. Interv.*, pages 639–646, 2004. 10, 41

[133] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens. Nonrigid image registration using conditional mutual information. *IEEE Trans. Med. Imaging*, 29(1): 19–29, 2010. 128

[134] M. Lorenzi, N. Ayache, G.B. Frisoni, and X. Pennec. LCC-Demons: a robust and accurate symmetric diffeomorphic registration algorithm. *Neuroimage*, 81:470–83, 2013. 44, 62, 69

[135] S. Lucey, R. Navarathna, A.B. Ashraf, and S. Sridharan. Fourier Lucas-Kanade Algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2012. 128

[136] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging*, 16(2):187–98, 1997. 45, 46

[137] A. Malesci, L. Balzarini, A. Chiti, and G. Lucignani. Pancreatic cancer or chronic pancreatitis? An answer from PET/MRI image fusion. *Eur. J. Nucl. Med. Mol. Imaging*, 31 (9):1352, 2004. 4

[138] K.L. Maletz, R.D. Ennis, J. Ostenson, A. Pevsner, A. Kagen, and I. Wernick. Comparison of CT and MR-CT fusion for prostate post-implant dosimetry. *Int. J. Radiat. Oncol. Biol. Phys.*, 82(5):1912–7, 2012. 4

[139] D.S. Marcus, T.H. Wang, J. Parker, J.G. Csernansky, J.C. Morris, and R.L. Buckner. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.*, 19(9):1498–507, 2007. 83, 116, 133

[140] R.B. Marimont and M.B. Shapiro. Nearest neighbour searches and the curse of dimensionality. *IMA J. Appl. Math.*, pages 59–70, 1979. 66

[141] D. Mattes, D.R. Haynor, H. Vesselle, T.K. Lewellen, and W. Eubank. PET-CT image registration in the chest using free-form deformations. *IEEE Trans. Med. Imaging*, 22(1): 120–8, 2003. xvii, 45, 46

[142] J. Mazziotta, A.W. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R.P. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, D.L. Collins, P.M. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima, and B. Mazoyer. A probabilistic atlas and reference system for the human brain. *Philos. Trans. R. Soc. London*, 356(1412):1293–322, 2001. 86, 118, 145

[143] J. Meinguet. Multivariate interpolation at arbitrary points made simple. *Zeitschrift für Angew. Math. und Phys.*, 30(4), 1979. 25

[144] M.I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *Int. J. Comput. Vis.*, 41:61–84, 2001. 50

[145] M. Modat, M.J. Cardoso, P. Daga, D. Cash, N.C. Fox, and S. Ourselin. Inverse-consistent symmetric free form deformation. *Int. Symp. Biomed. Imaging*, pages 79–88, 2012. 61, 69

[146] J. Modersitzki and S. Wirtz. Combining homogenization and registration. In *Biomed. Image Regist.*, pages 257–263, 2006. 128

[147] C. Moler and C. Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.*, 45(1):3–49, 2003. 36

[148] O. Monga, R. Deriche, and J.M. Rocchisani. 3D edge detection using recursive filtering: application to scanner images. *CVGIP Image Underst.*, 53(1):76–87, 1991. 56

[149] S.G. Mueller, M.W. Weiner, L.J. Thal, R.C. Petersen, C.R. Jack, W. Jagust, J.Q. Trojanowski, A.W. Toga, and L. Beckett. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's Dement.*, 1(1):55–66, 2005. 83, 133

[150] R Mukundan. Quaternions: From classical mechanics to computer graphics, and beyond. In *Proc. Asian Technol. Conf. Math.*, 2002. 22

[151] O. Musse, F. Heitz, and J.-P. Armspach. Topology preserving deformable image matching using constrained hierarchical parametric models. *IEEE Trans. image Process.*, 10(7): 1081–93, 2001. 31

[152] Y. Nakamoto, K. Tamai, T. Saga, T. Higashi, T. Hara, T. Suga, T. Koyama, and K. Togashi. Clinical value of image fusion from MR and PET in patients with head and neck cancer. *Mol. imaging Biol.*, 11(1):46–53, 2009. 4

[153] S.F. Nemec, M.A. Donat, S. Mehrain, K. Friedrich, C. Krestan, C. Matula, H. Imhof, and C. Czerny. CT-MR image data fusion for computer assisted navigated neurosurgery of temporal bone tumors. *Eur. J. Radiol.*, 62(2):192–8, 2007. 4

[154] V. Noblet, C. Heinrich, F. Heitz, and J.-P. Armspach. 3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization. *IEEE Trans. Image Process.*, 14(5):553–566, 2005. 31

[155] V. Noblet, C. Heinrich, F. Heitz, and J.-P. Armspach. Retrospective evaluation of a topology preserving non-rigid registration method. *Med. Image Anal.*, 10(3):366–84, 2006. 31

[156] V. Noblet, C. Heinrich, F. Heitz, and J.-P. Armspach. Symmetric nonrigid image registration: application to average brain templates construction. In *Med. image Comput. Comput. Interv.*, volume 11, pages 897–904, 2008. 30

[157] J.A. Onofrey, L.H. Staib, and X. Papademetris. Fast nonrigid image registration using statistical deformation models learned from richly-annotated data. In *Int. Symp. Biomed. Imaging*, pages 580–583, 2013. 105

[158] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Med. Image Anal.*, 15(4):622–39, 2011. ISSN 1361-8423. 63, 69

[159] N. Papanikolaou, D. Gearheart, T. Bolek, A. Meigooni, D. Meigooni, and M. Mohiuddin. A volumetric and dosimetric study of LDR brachytherapy prostate implants based on image fusion of ultrasound and computed tomography. *IEEE Int. Conf. Eng. Med. Biol. Soc.*, 4:2769–2770, 2000. 4

[160] A. Papoulis and S.U Pillai. *Probability, random variables, and stochastic processes*. 2002. 76

[161] S.H. Park and H.J. Sung. Assessment of image registration interpolation methods for pressure-sensitive paint measurements. In *Int. Symp. Flow Vis.*, pages 1–7, 2004. 48

[162] J Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. 1988. 52

[163] X. Pennec, N. Ayache, and J.P. Thirion. Landmark-based registration using features identified through differential geometry. *Handb. Med. Imaging*, pages 499–513, 2000. 72

[164] X. Pennec, P. Cachier, and N. Ayache. Understanding the demon's algorithm: 3D non-rigid registration by gradient descent. In *Med. image Comput. Comput. Interv.*, pages 597–606, 1999. 56

[165] A. Pitiot, E. Bardinet, P.M. Thompson, and G. Malandain. Piecewise affine registration of biological images for volume reconstruction. *Med. Image Anal.*, 10(3):465–83, 2006. 31

[166] J.P.W. Pluim, J.B. Maintz, and M. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Med. Imaging*, 19 (8):809–14, 2000. 128

[167] E. Polak and G. Ribière. Note sur la convergence de méthodes de directions conjuguées. *Rev. française dinformatique Rech. opérationnelle, série rouge*, 3(1):35–43, 1969. 50, 82

[168] B.T. Polyak. The conjugate gradient method in extremal problems. *USSR Comput. Math. Math. Phys.*, 9(4):94–112, 1969. 50, 82

[169] G. Postelnicu, L. Zollei, and B. Fischl. Combined volumetric and surface registration. *IEEE Trans. Med. Imaging*, 28(4):508–522, 2009. 50

[170] S. Pszczolkowski, L. Pizarro, R. Guerrero, and D. Rueckert. Nonrigid free-form registration using landmark-based statistical deformation models. In David R. Haynor and Sébastien Ourselin, editors, *Proc. SPIE Med. Imaging*, 2012. 14, 107

[171] S. Pszczolkowski, L. Pizarro, D.P. O'Regan, and D. Rueckert. Gradient projection learning for parametric nonrigid registration. *Proc. MICCAI Work. Mach. Learn. Med. Imaging*, (1):1–8, 2012. 14

[172] S. Pszczolkowski, S. Zafeiriou, C. Ledig, and D. Rueckert. A robust similarity measure for nonrigid image registration with outliers. In *Int. Symp. Biomed. Imaging*, number 2, pages 568–571, 2014. 15

[173] Z. Qiu, H. Tang, and D. Tian. Non-rigid Medical Image Registration Based on the Thin-Plate Spline Algorithm. In *World Congr. Comput. Sci. Inf. Eng.*, pages 522–527, 2009. 26

[174] R. Reinartz, B. Platel, T. Boselie, H. Van Mameren, H. Van Santbrink, and B.T.H. Romeny. Cervical vertebrae tracking in video-fluoroscopy using the normalized gradient field. In *Med. image Comput. Comput. Interv.*, volume 12, pages 524–31, 2009. 128

[175] F.J.P. Richard. A new approach for the registration of images with inconsistent differences. In *Int. Conf. Pattern Recognit.*, volume 5, pages 5–8, 2004. 128

[176] A.C. Riegel, A.M. Berson, S. Destian, T. Ng, L.B. Tena, R.J. Mitnick, and P.S. Wong. Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion. *Int. J. Radiat. Oncol. Biol. Phys.*, 65(3):726–32, 2006. 4

[177] T. Rohlfing, C.R. Maurer, D.A. Bluemke, and M.A. Jacobs. Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE Trans. Med. Imaging*, 22(6):730–41, 2003. 10, 40

[178] K. Rohr, H.S. Stiehl, and R. Sprengel. Point-based elastic registration of medical image data using approximating thin-plate splines. *Vis. Biomed. Comput.*, (1), 1996. 26

[179] K. Rohr, H.S. Stiehl, R. Sprengel, T.M. Buzug, J. Weese, and M.H. Kuhn. Landmark-based elastic registration using approximating thin-plate splines. *IEEE Trans. Med. Imaging*, 20(6):526–34, 2001. 26, 27, 72

[180] D. Ruan, J.A. Fessler, M. Roberson, J.M. Balter, and M.L. Kessler. Nonrigid registration using regularization that accomodates local tissue rigidity. In Joseph M. Reinhardt and Josien P. W. Pluim, editors, *Proc. SPIE Med. Imaging*, volume 6144, pages 614412–614412–9, 2006. 10, 41

[181] D. Rueckert, P. Aljabar, R.A. Heckemann, J.V. Hajnal, and A. Hammers. Diffeomorphic registration using B-splines. In *Med. image Comput. Comput. Interv.*, volume 9, pages 702–9, 2006. 30, 31

[182] D. Rueckert, A.F. Frangi, and J. Schnabel. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Trans. Med. Imaging*, 22(8):1014–25, 2003. xiii, xvii, 73, 83, 84, 107

[183] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging*, 18(8):712–21, 1999. 30, 31, 50, 71, 73, 107, 109, 114, 127, 133, 150, 181

[184] J. Rühaak, L. Konig, M. Hallmann, N. Papenberg, S. Heldmann, H. Schumacher, and B. Fischer. A Fully Parallel Algorithm For Multimodal Image Registration Using Normalized Gradient Fields. In *Int. Symp. Biomed. Imaging*, pages 572–575, 2013. 129, 130

[185] N. Samavati, M. Velec, and K. Brock. A hybrid biomechanical intensity based deformable image registration of lung 4DCT. In Sebastien Ourselin and Martin A. Styner, editors, *Proc. SPIE Med. Imaging*, volume 9034, page 90343J, 2014. 51

[186] J.N. Sarvaiya, S. Patnaik, and S. Bombaywala. Image registration by template matching using normalized cross-correlation. In *Int. Conf. Adv. Comput. Control. Telecommun. Technol.*, number 3, pages 819–822. Ieee, December 2009. 43

[187] J. Schnabel, D. Rueckert, M. Quist, J.M. Blackall, A.D. Castellano-Smith, T. Hartkens, G.P. Penney, W.A. Hall, H. Liu, C.L. Truwit, F.A. Gerristen, D.L.G. Hill, and D.J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Med. image Comput. Comput. Interv.*, pages 573–581, 2001. 30

[188] M. Sdika. A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization. *IEEE Trans. Med. Imaging*, 27(2):271–281, 2008. 30

[189] T.W. Sederberg and S.R. Parry. Free-form deformation of solid geometric models. *ACM SIGGRAPH Comput. Graph.*, 20(4):151–160, 1986. 28, 29

[190] R. Shams, P. Sadeghi, R.A. Kennedy, and R.I. Hartley. A survey of medical image registration on multicore and the GPU. *Signal Process. multicore platforms*, (March), 2010. 9

[191] C.E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423,623–656, 1948. 45

[192] A. Shekhovtsov, I. Kovtun, and V. Hlaváč. Efficient MRF deformation model for non-rigid image matching. *Comput. Vis. image Underst.*, 112(1):91–99, 2008. 51

[193] D. Shen and C. Davatzikos. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. Med. Imaging*, 21(11):1421–39, 2002. 62, 64, 69

[194] W. Shi, X. Zhuang, L. Pizarro, W. Bai, H. Wang, K.P. Tung, P.J. Edwards, and D. Rueckert. Registration using sparse free-form deformations. *Med. image Comput. Comput. Interv.*, 15:659–66, 2012. 30

[195] N. Shusharina and G. Sharp. Image registration using radial basis functions with adaptive radius. *Med. Phys.*, 39(11):6542–9, 2012. 28

[196] R.W.K. So, T.W.H. Tang, and A.C.S. Chung. Non-rigid image registration of brain magnetic resonance images using graph-cuts. *Pattern Recognit.*, 44(10-11):2450–2467, 2011. 51

[197] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: a survey. *IEEE Trans. Med. Imaging*, 32(7):1153–90, 2013. 54

[198] A. Sotiras, Y. Ou, B. Glocker, C. Davatzikos, and N. Paragios. Simultaneous geometric-iconic registration. In *Med. image Comput. Comput. Interv.*, volume 13, pages 676–83, 2010. 51

[199] A. Sotiras and N. Paragios. Discrete symmetric image registration. In *Int. Symp. Biomed. Imaging*, pages 342–345, 2012. 30

[200] A. Staniforth and J. Côté. Semi-Lagrangian integration schemes for atmospheric models: a review. *Mon. Weather Rev.*, 1991. 58

[201] M. Staring, S. Klein, and J.P.W. Pluim. A rigidity penalty term for nonrigid registration. *Med. Phys.*, 34(11):4098, 2007. 10, 41

[202] J.C. Strikwerda. *Finite difference schemes and partial differential equations*. 1989. 38

[203] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas. Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Trans. Med. Imaging*, 25(5):626–39, 2006. 128

[204] C. Studholme, D.L.G. Hill, and D.J. Hawkes. Multiresolution voxel similarity measures for MR-PET registration. In *Inf. Process. Med. Imaging*, volume 252, 1995. 45

[205] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.*, 32(1):71–86, 1999. 47, 128, 133

[206] S. Tang, Y. Fan, G. Wu, M.J. Kim, and D. Shen. RABBIT: rapid alignment of brains by building intermediate templates. *Neuroimage*, 47(4):1277–87, 2009. 65, 67, 69, 72, 107

[207] D. Taussky, L. Austen, A. Toi, I. Yeung, T. Williams, S. Pearson, M. McLean, G. Pond, and J. Crook. Sequential evaluation of prostate edema after permanent seed prostate brachytherapy using CT-MRI fusion. *Int. J. Radiat. Oncol. Biol. Phys.*, 62(4):974–80, 2005. 4

[208] L.M. Taylor, J.D. Beaty, J.D. Enderle, and M.A. Escabi. Design of a simple ultrasound/CT fusion image fusion solution for the evaluation of prostate seed brachytherapy. *IEEE Annu. Northeast Bioeng. Conf. Des.*, pages 57–58, 2001. 4

[209] R.H. Taylor, S. Lavallee, G. Burdea, and R. Mösges. Computer-integrated surgery. Technical report, MIT Press, 1995. 39

[210] S.J. Teipel, C. Born, M. Ewers, A.L.W. Bokde, M.F. Reiser, H.J. Möller, and H. Hampel. Multivariate deformation-based analysis of brain atrophy to predict Alzheimer's disease in mild cognitive impairment. *Neuroimage*, 38(1):13–24, 2007. 5

[211] J.P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med. Image Anal.*, 2(3):243–260, 1998. 56, 69, 108

[212] Y. Tian and S.G. Narasimhan. A globally optimal data-driven approach for image distortion estimation. In *Conf. Comput. Vis. Pattern Recognit.*, pages 1277–1284, 2010. xvii, 43, 66, 69, 107

[213] N. Tomura, O. Watanabe, K. Omachi, I. Sakuma, S. Takahashi, T. Otani, H. Kidani, and J. Watarai. Image fusion of thallium-201 SPECT and MR imaging for the assessment of recurrent head and neck tumors following flap reconstructive surgery. *Eur. Radiol.*, 14 (7):1249–54, 2004. 4

[214] N.J. Tustison and B.B. Avants. Explicit B-spline regularization in diffeomorphic image registration. *Front. Neuroinform.*, 7(December):39, 2013. 10

[215] N.J. Tustison, B.B. Avants, P. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, and J.C. Gee. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging*, 29(6):1310–20, 2010. 135

[216] N.J. Tustison, B.B. Avants, and J.C. Gee. Directly manipulated free-form deformation image registration. *IEEE Trans. image Process.*, 18(3):624–35, 2009. 50

[217] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. Robust and efficient parametric face alignment. In *Int. Conf. Comput. Vis.*, 2011. 127, 128, 129

[218] M. van Stralen and J.P.W. Pluim. Optimal discrete multi-resolution deformable image registration. In *Int. Symp. Biomed. Imaging*, pages 947–950, 2009. 51

[219] L.J. van Vliet and P.W. Verbeek. Curvature and bending energy in digitized 2D and 3D images. In *Proc. Scand. Conf. Image Anal.*, volume 2, pages 1403–1410, 1993. 39

[220] T. Vercauteren, X. Pennec, E. Malis, A. Perchant, and N. Ayache. Insight into efficient image registration techniques and the demons algorithm. *Inf. Process. Med. Imaging*, pages 495–506, 2007. 50

[221] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Non-parametric diffeomorphic image registration with the demons algorithm. In *Med. image Comput. Comput. Interv.*, volume 10, pages 319–26, 2007. 50, 60, 108

[222] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Symmetric log-domain diffeomorphic Registration: a demons-based approach. *Med. image Comput. Comput. Interv.*, 11(Pt 1):754–61, 2008. 60, 62, 69

[223] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. Diffeomorphic demons: efficient non-parametric image registration. *Neuroimage*, 45(1 Suppl):S61–72, 2009. 43, 50, 56

[224] P. Viola and W.M. Wells. Alignment by maximization of mutual information. *Int. J. Comput. Vis.*, 24(2):1–29, 1997. 46

[225] G. Wahba. *Spline models for observational data*. Siam, 59 edition, 1990. 39

[226] J. Wang and T. Jiang. Nonrigid registration of brain MRI using NURBS. *Pattern Recognit. Lett.*, 28(2):214–223, 2007. 30

[227] Y. Wang and L.H. Staib. Physical model-based non-rigid registration incorporating statistical shape information. *Med. Image Anal.*, 4(1):7–20, 2000. 72

[228] A. Watt and M. Watt. *Advanced animation and rendering techniques*. 1992. 22

[229] L. Weizman, L.B. Sira, L. Joskowicz, D.L. Rubin, K.W. Yeom, S. Constantini, B. Shofty, and D.B. Bashat. Semiautomatic segmentation and follow-up of multicomponent low-grade tumors in longitudinal brain MRI studies. *Med. Phys.*, 41(5):052303, 2014. 5

[230] J.L. Whitwell, S.D. Weigand, M.M. Shiung, B.F. Boeve, T.J. Ferman, G.E. Smith, D.S. Knopman, R.C. Petersen, E.E. Benarroch, K.A. Josephs, and C.R. Jack. Focal atrophy in

dementia with Lewy bodies on MRI: a distinct pattern from Alzheimer's disease. *Brain*, 130(Pt 3):708–19, 2007. 5

[231] G. Wu, F. Qi, and D. Shen. Learning best features and deformation statistics for hierarchical registration of MR brain images. *Inf. Process. Med. Imaging*, 20:160–71, 2007. 108

[232] Y.T. Wu, T. Kanade, C.C. Li, and J. Cohn. Image registration using wavelet-based motion model. *Int. J. Comput. Vis.*, 38(2):129–152, 2000. 31, 51

[233] P.P. Wyatt and J.A. Noble. MAP MRF joint segmentation and registration. In *Med. image Comput. Comput. Interv.*, pages 580–587, 2002. 128

[234] Z. Xue, D. Shen, and C. Davatzikos. Statistical representation of high-dimensional deformation fields with application to statistically constrained 3D warping. *Med. Image Anal.*, 10(5):740–51, 2006. 72, 107

[235] Z. Yi and S. Soatto. Nonrigid registration combining global and local statistics. In *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pages 2200–2207, 2009. 128

[236] A.A. Young and A.F. Frangi. Computational cardiac atlases: from patient to population and back. *Exp. Physiol.*, 94(5):578–96, 2009. 6

[237] E.I. Zacharaki, D. Shen, S.K. Lee, and C. Davatzikos. ORBIT: a multiresolution framework for deformable registration of brain tumor images. *IEEE Trans. Med. Imaging*, 27 (8):1003–17, 2008. 128

[238] H. Zhang, P.A. Yushkevich, D.C. Alexander, and J.C. Gee. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Med. Image Anal.*, 10 (5):764–85, 2006. 31

[239] G. Zheng and X. Zhang. A unifying MAP-MRF framework for deriving new point similarity measures for intensity-based 2D-3D registration. In *Int. Conf. Pattern Recognit.*, number c, pages 2–6, 2006. 128

[240] H. Zhou, M. Chen, and D. Zhao. Longitudinal MRI evaluation of intracranial development and vascular characteristics of breast cancer brain metastases in a mouse model. *PLoS One*, 8(4):e62238, 2013. 5

[241] X. Zhuang, S.R. Arridge, D.J. Hawkes, and S. Ourselin. A nonrigid registration framework using spatially encoded mutual information and free-form deformations. *IEEE Trans. Med. Imaging*, (c):1–10, 2011. 128

[242] X. Zhuang, D.J. Hawkes, W.R. Crum, R. Boubertakh, S. Uribe, D. Atkinson, P. Batchelor, T. Schaeffter, R.S. Razavi, and D.L.G. Hill. Robust registration between cardiac MRI images and atlas for segmentation propagation. In Joseph M. Reinhardt and Josien P. W. Pluim, editors, *Proc. SPIE Med. Imaging*, volume 6914, pages 691408–691408–11, 2008. 7

[243] X. Zhuang, K.S. Rhode, R.S. Razavi, D.J. Hawkes, and S. Ourselin. A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Trans. Med. Imaging*, 29(9):1612–25, 2010. 7

[244] D. Zikic, B. Glocker, O. Kutter, M. Groher, N. Komodakis, A. Kamen, N. Paragios, and N. Navab. Linear intensity-based image registration by Markov random fields and discrete optimization. *Med. Image Anal.*, 14(4):550–62, 2010. 51

[245] D. Zikic, A. Kamen, and N. Navab. Natural gradients for deformable registration. In *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pages 2847–2854, 2010. 108

[246] B. Zitová and J. Flusser. Image registration methods: a survey. *Image Vis. Comput.*, 21 (11):977–1000, 2003. 54

[247] G. Zou, J. Hua, and O. Muzik. Non-rigid surface registration using spherical thin-plate splines. In *Med. image Comput. Comput. Interv.*, pages 367–374, 2007. 26

Appendix A

Derivations

A.1 Gradient of Statistical Transformation Model

Considering that the proposed statistical transformation model of Equation (4.12) is based on the free-form deformation (FFD) model of Rueckert et al. [183] (Equation (2.21)), we look into the gradient of the FFD transformation model first. In this model, the gradient of the energy functional with respect to the transformation parameters (control point values) Φ is computed by means of the chain rule as:

$$\frac{\partial E(\mathbf{x})}{\partial \Phi} = \frac{\partial E(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})} \frac{\partial \mathbf{T}(\mathbf{x})}{\partial \Phi}, \tag{A.1}$$

where $\frac{\partial E(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})}$ is the standard voxelwise gradient of the energy functional and

$$\frac{\partial T^{\xi}(\mathbf{x})}{\partial \Phi_i} = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w). \tag{A.2}$$

Here, B_r, u, v, w have the same meaning as in Equation (2.16). Now, in order to compute the gradient for the statistical transformation model, i.e, the gradient with respect to the statistical parameterisation $\mathbf{w}$, we can simply apply the chain rule once again. This yields

$$\frac{\partial E(\mathbf{x})}{\partial \mathbf{w}} = \frac{\partial E(\mathbf{x})}{\partial \Phi} \frac{\partial \Phi}{\partial \mathbf{w}}. \tag{A.3}$$

Since we model the control point value parameterisation as $\Phi \equiv \Phi^{(\mathbf{w})} = \mathbf{B}\mathbf{w} + \overline{\Phi}$, we have:

$$\frac{\partial \Phi}{\partial \mathbf{w}} = \mathbf{B}. \tag{A.4}$$

By combining equations (A.1), (A.3) and (A.4), we obtain the final gradient of the energy functional with respect to the statistical parameterisation:

$$\frac{\partial E(\mathbf{x})}{\partial \mathbf{w}} = \frac{\partial E(\mathbf{x})}{\partial \mathbf{T}(\mathbf{x})} \frac{\partial \mathbf{T}(\mathbf{x})}{\partial \Phi} \mathbf{B}. \tag{A.5}$$

A.2 Gradient of Landmark Alignment term

Recall the landmark alignment term, which is given by

$$E_{landmark}(l^{I_0}; l^I; \mathbf{T}_\Phi) = \frac{1}{L} \sum_{i=1}^{L} \Psi \left(\| l_i^I - \mathbf{T}_\Phi(l_i^{I_0}) \|^2 \right). \tag{A.6}$$

We also have that the derivative of this energy term has to be computed by following the chain rule:

$$\frac{\partial E_{landmark}}{\partial \Phi} = \frac{\partial E_{landmark}}{\partial \mathbf{T}_\Phi(l^{I_0})} \frac{\partial \mathbf{T}_\Phi(l^{I_0})}{\partial \Phi}. \tag{A.7}$$

The first term of Equation (A.7) is given by:

$$\frac{\partial E_{landmark}}{\partial \mathbf{T}_\Phi(l^{I_0})} = -2 \sum_{i=1}^{L} \Psi' \left(\| l_i^I - \mathbf{T}_\Phi(l_i^{I_0}) \|^2 \right) (l_i^I - \mathbf{T}_\Phi(l_i^{I_0})), \tag{A.8}$$

where $\Psi'(x^2) = \frac{1}{1+x^2/\delta^2}$. And the second term of Equation (A.7) is expressed as:

$$\frac{\partial \mathbf{T}_\Phi(l^{I_0})}{\partial \Phi} = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w), \tag{A.9}$$

where, B_r, u, v, w have the same meaning as in Equation (2.16), except in that u, v, w are relative to the target landmark positions l^{I_0} instead of voxel positions.

A.3 Gradient of Cosine Similarities

Since cosine similarities aim to match vector fields rather than intensities, it is necessary to utilise vector (matrix) calculus to derive the gradients. In the following derivations, the Numerator-layout notation of matrix calculus is used.

A.3.1 Cosine similarity

Let $\mathbf{x} = (x, y, z)^\top$ be a 3D point in space and $\mathbf{y} = (u, v, w)^\top = \mathbf{T}(\mathbf{x})$ be its transformed location. Given the target image I_0 and the transformed source image $I_1 = I \circ \mathbf{T}$ the cosine similarity corresponds to:

$$\cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})) = \frac{\nabla I_0(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})^\top}{\|\nabla I_0(\mathbf{x})\| \|\nabla I_1(\mathbf{x})\|}. \tag{A.10}$$

By using the quotient rule, the partial derivative of the similarity with respect to a transformation parameter Φ_i is given by:

$$\begin{aligned}\frac{\partial \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}))}{\partial \Phi_i} = &\frac{\|\nabla I_0(\mathbf{x})\| \|\nabla I_1(\mathbf{x})\| \cdot \frac{\partial(\nabla I_0(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})^\top)}{\partial \Phi_i}}{\|\nabla I_0(\mathbf{x})\|^2 \|\nabla I_1(\mathbf{x})\|^2} \\ &- \frac{(\nabla I_0(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})^\top) \cdot \frac{\partial(\|\nabla I_0(\mathbf{x})\| \|\nabla I_1(\mathbf{x})\|)}{\partial \Phi_i}}{\|\nabla I_0(\mathbf{x})\|^2 \|\nabla I_1(\mathbf{x})\|^2}.\end{aligned} \tag{A.11}$$

Considering the fact that the gradient of the target image $\nabla I_0(\mathbf{x})$ does not depend on the transformation parameters, we have

$$\frac{\partial(\nabla I_0(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})^\top)}{\partial \Phi_i} = \nabla I_0(\mathbf{x}) \cdot \left[\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}\right]^\top \tag{A.12}$$

$$\frac{\partial(\|\nabla I_0(\mathbf{x})\| \|\nabla I_1(\mathbf{x})\|)}{\partial \Phi_i} = \|\nabla I_0(\mathbf{x})\| \cdot \frac{\partial \|\nabla I_1(\mathbf{x})\|}{\partial \Phi_i}. \tag{A.13}$$

Hence, we can rewrite equation (A.11) as

$$\frac{\partial \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}))}{\partial \Phi_i} = \frac{\nabla I_0(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|} \cdot \left[\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}\right]^\top - \frac{\nabla I_0(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})^\top}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|^2} \cdot \frac{\partial \|\nabla I_1(\mathbf{x})\|}{\partial \Phi_i} \tag{A.14}$$

Now, we focus on $\frac{\partial \|\nabla I_1(\mathbf{x})\|}{\partial \Phi_i}$. We have:

$$\|\nabla I_1(\mathbf{x})\| = \sqrt{\nabla I_1(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})^\top}, \tag{A.15}$$

thus

$$\begin{aligned}
\frac{\partial \|\nabla I_1(\mathbf{x})\|}{\partial \Phi_i} &= \frac{1}{2\|\nabla I_1(\mathbf{x})\|} \cdot \frac{\partial(\nabla I_1(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})^\top)}{\partial \Phi_i} \\
&= \frac{1}{2\|\nabla I_1(\mathbf{x})\|} \cdot \left(2 \cdot \nabla I_1(\mathbf{x}) \cdot \left[\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}\right]^\top\right) \\
&= \frac{\nabla I_1(\mathbf{x})}{\|\nabla I_1(\mathbf{x})\|} \cdot \left[\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}\right]^\top .
\end{aligned} \tag{A.16}$$

By replacing equation (A.16) into equation (A.14), we realise that a common term $\left[\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}\right]^\top$ appears, which we can factorise yielding:

$$\begin{aligned}
\frac{\partial \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}))}{\partial \Phi_i} &= \left(\frac{\nabla I_0(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|} - \frac{\nabla I_0(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|^2} \cdot \frac{\nabla I_1(\mathbf{x})}{\|\nabla I_1(\mathbf{x})\|}\right) \cdot \left[\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}\right]^\top \\
&= \left(\frac{\nabla I_0(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|} - \frac{\nabla I_0(\mathbf{x}) \cdot \nabla I_1(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|} \cdot \frac{\nabla I_1(\mathbf{x})}{\|\nabla I_1(\mathbf{x})\|^2}\right) \cdot \left[\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}\right]^\top \\
&= \left(\frac{\nabla I_0(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|} - \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})) \cdot \frac{\nabla I_1(\mathbf{x})}{\|\nabla I_1(\mathbf{x})\|^2}\right) \cdot \left[\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}\right]^\top .
\end{aligned} \tag{A.17}$$

Now we focus in the common term $\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i}$. This term is given by:

$$\frac{\partial(\nabla I_1(\mathbf{x}))}{\partial \Phi_i} \overset{I_1 = I \circ \mathbf{T}}{=} \frac{\partial(\nabla I(\mathbf{T}(\mathbf{x})))}{\partial \Phi_i}$$

$$\overset{\mathbf{y}=\mathbf{T}(\mathbf{x})}{=} \frac{\partial(\nabla_{\mathbf{x}} I(\mathbf{y}))}{\partial \Phi_i}$$

$$= \frac{\partial}{\partial \Phi_i}\left(\frac{dI(\mathbf{y})}{d\mathbf{x}}\right)$$

$$= \frac{\partial}{\partial \Phi_i}\left(\frac{dI(\mathbf{y})}{d\mathbf{y}} \cdot \frac{d\mathbf{y}}{d\mathbf{x}}\right)$$

$$= \frac{\partial}{\partial \Phi_i}\left(\frac{dI(\mathbf{y})}{d\mathbf{y}}\right) \cdot \frac{d\mathbf{y}}{d\mathbf{x}} + \frac{dI(\mathbf{y})}{d\mathbf{y}} \cdot \frac{\partial}{\partial \Phi_i}\left(\frac{d\mathbf{y}}{d\mathbf{x}}\right)$$

$$= \frac{\partial^2 I}{d\mathbf{y} \partial \Phi_i} \cdot \frac{d\mathbf{y}}{d\mathbf{x}} + \frac{dI(\mathbf{y})}{d\mathbf{y}} \cdot \frac{\partial^2 \mathbf{y}}{d\mathbf{x} \partial \Phi_i}$$

$$= \frac{\partial \mathbf{y}}{\partial \Phi_i} \cdot \left[\frac{d^2 I(\mathbf{y})}{d\mathbf{y}^2}\right]^\top \cdot \frac{d\mathbf{y}}{d\mathbf{x}} + \frac{dI(\mathbf{y})}{d\mathbf{y}} \cdot \frac{\partial^2 \mathbf{y}}{d\mathbf{x} \partial \Phi_i}. \tag{A.18}$$

Here, $\frac{dI(\mathbf{y})}{d\mathbf{y}}$ is the transformed gradient of the source image I, $\frac{d^2 I(\mathbf{y})}{d\mathbf{y}^2}$ is the transformed Hessian of the source image I and $\frac{d\mathbf{y}}{d\mathbf{x}}$ is the Jacobian matrix of the transformation. The remaining terms $\frac{\partial \mathbf{y}}{\partial \Phi_i}$ and $\frac{\partial^2 \mathbf{y}}{d\mathbf{x} \partial \Phi_i}$ depend upon the particular transformation model in use (rigid, affine, Thin-plate splines, B-spline FFD, etc). By combining equations (A.17) and (A.18), we obtain the final expression for the partial derivative of the similarity with respect to a transformation parameter

Φ_i:

$$\begin{aligned}\frac{\partial \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}))}{\partial \Phi_i} &= \left(\frac{\nabla I_0(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|} - \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})) \cdot \frac{\nabla I_1(\mathbf{x})}{\|\nabla I_1(\mathbf{x})\|^2}\right) \\ &\quad \cdot \left[\frac{\partial \mathbf{y}}{\partial \Phi_i} \cdot \left[\frac{d^2 I(\mathbf{y})}{d\mathbf{y}^2}\right]^\top \cdot \frac{d\mathbf{y}}{d\mathbf{x}} + \frac{dI(\mathbf{y})}{d\mathbf{y}} \cdot \frac{\partial^2 \mathbf{y}}{d\mathbf{x}\partial\Phi_i}\right]^\top \\ &= \left(\frac{\nabla I_0(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|} - \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})) \cdot \frac{\nabla I_1(\mathbf{x})}{\|\nabla I_1(\mathbf{x})\|^2}\right) \\ &\quad \cdot \left(\left[\frac{d\mathbf{y}}{d\mathbf{x}}\right]^\top \cdot \frac{d^2 I(\mathbf{y})}{d\mathbf{y}^2} \cdot \left[\frac{\partial \mathbf{y}}{\partial \Phi_i}\right]^\top + \left[\frac{\partial^2 \mathbf{y}}{d\mathbf{x}\partial\Phi_i}\right]^\top \cdot \left[\frac{dI(\mathbf{y})}{d\mathbf{y}}\right]^\top\right) \end{aligned} \tag{A.19}$$

To simplify the gradient computation, we regard the term $\frac{\partial^2 \mathbf{y}}{d\mathbf{x}\partial\Phi_i}$ as a zero matrix. This also has benefits in terms of memory savings and computation time. Hence, our final gradient is given by:

$$\begin{aligned}\frac{\partial \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}))}{\partial \Phi_i} &\approx \left(\frac{\nabla I_0(\mathbf{x})}{\|\nabla I_0(\mathbf{x})\|\|\nabla I_1(\mathbf{x})\|} - \cos\alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})) \cdot \frac{\nabla I_1(\mathbf{x})}{\|\nabla I_1(\mathbf{x})\|^2}\right) \\ &\quad \cdot \left(\left[\frac{d\mathbf{y}}{d\mathbf{x}}\right]^\top \cdot \frac{d^2 I(\mathbf{y})}{d\mathbf{y}^2} \cdot \left[\frac{\partial \mathbf{y}}{\partial \Phi_i}\right]^\top\right) \end{aligned} \tag{A.20}$$

To show that the gradient in Equation (A.20) is a good approximation of the gradient in Equation (A.19) we registered two images from the ADNI database using the proposed similarity measure with both the exact and approximate gradient computations. The results in axial, coronal and sagital views are shown in figures A.1, A.2 and A.3, respectively. We observe the results are extremely similar between the two approaches, demonstrating that the approximate gradient is a suitable option for registration. The multi-threaded registration using the exact gradient took 5 minutes and 23 seconds, while the approximate version took 3 minutes and 19

seconds, both on a standard PC desktop with 8 cores (3.40GHz) and 16GB of RAM.

A.3.2 Cosine squared similarity

The gradient of the cosine squared similarity with respect to the transformation parameters corresponds to

$$\frac{\partial \cos^2 \alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}))}{\partial \Phi_i} = 2 \cdot \cos \alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x})) \cdot \frac{\partial \cos \alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}))}{\partial \Phi_i}, \tag{A.21}$$

where $\frac{\partial \cos \alpha(\nabla I_0(\mathbf{x}), \nabla I_1(\mathbf{x}))}{\partial \Phi_i}$ is given by equation (A.20).

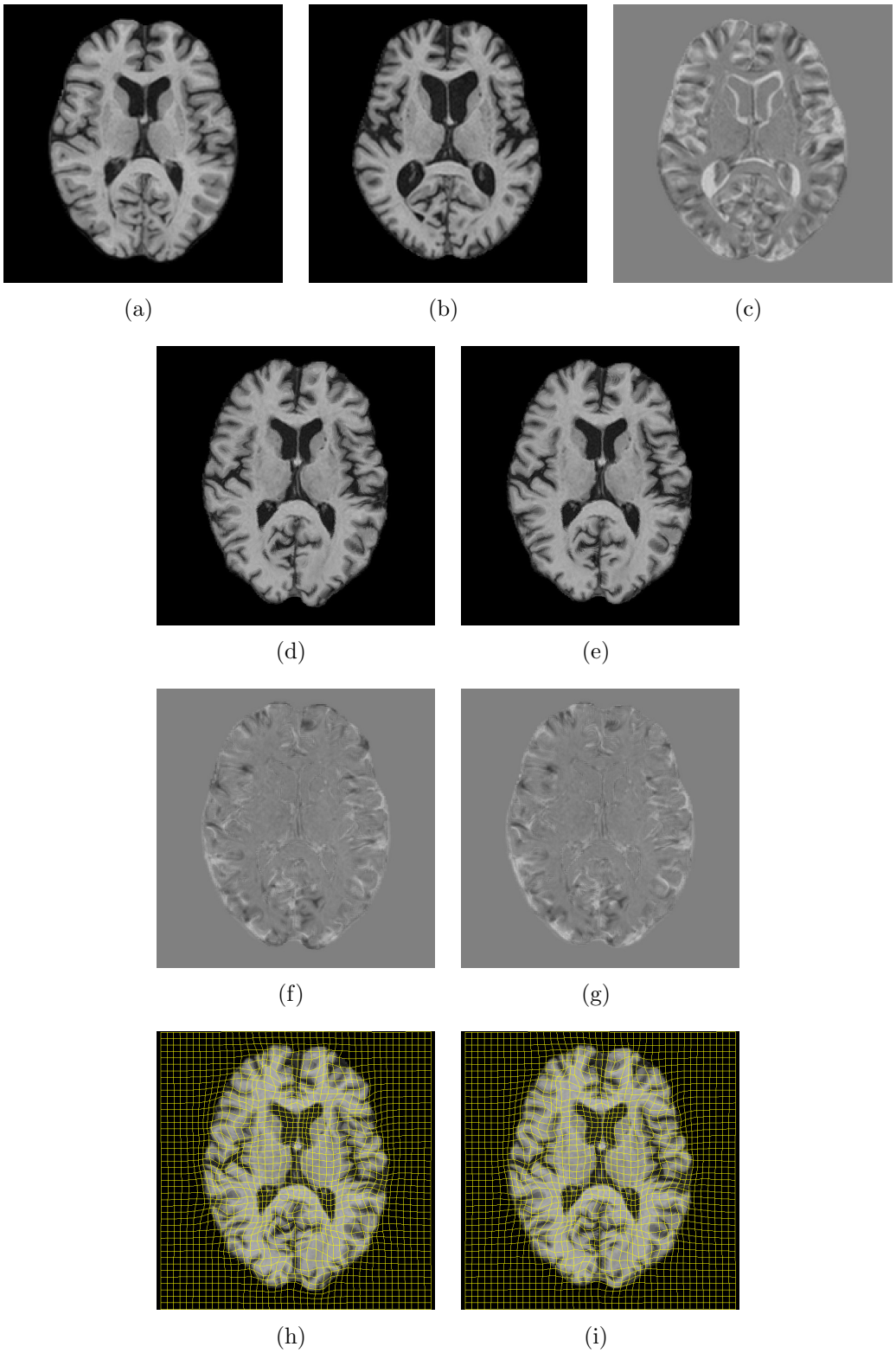


Figure A.1: Registration results using both exact and approximate gradients (axial view). (a): Target image. (b): Source image. (c): Difference image. (d): Warped source using exact gradient. (e): Warped source using approximate gradient. (f): Difference image using exact gradient. (g): Difference image using approximate gradient. (h): Resulting FFD using exact gradient. (i): Resulting FFD using approximate gradient.

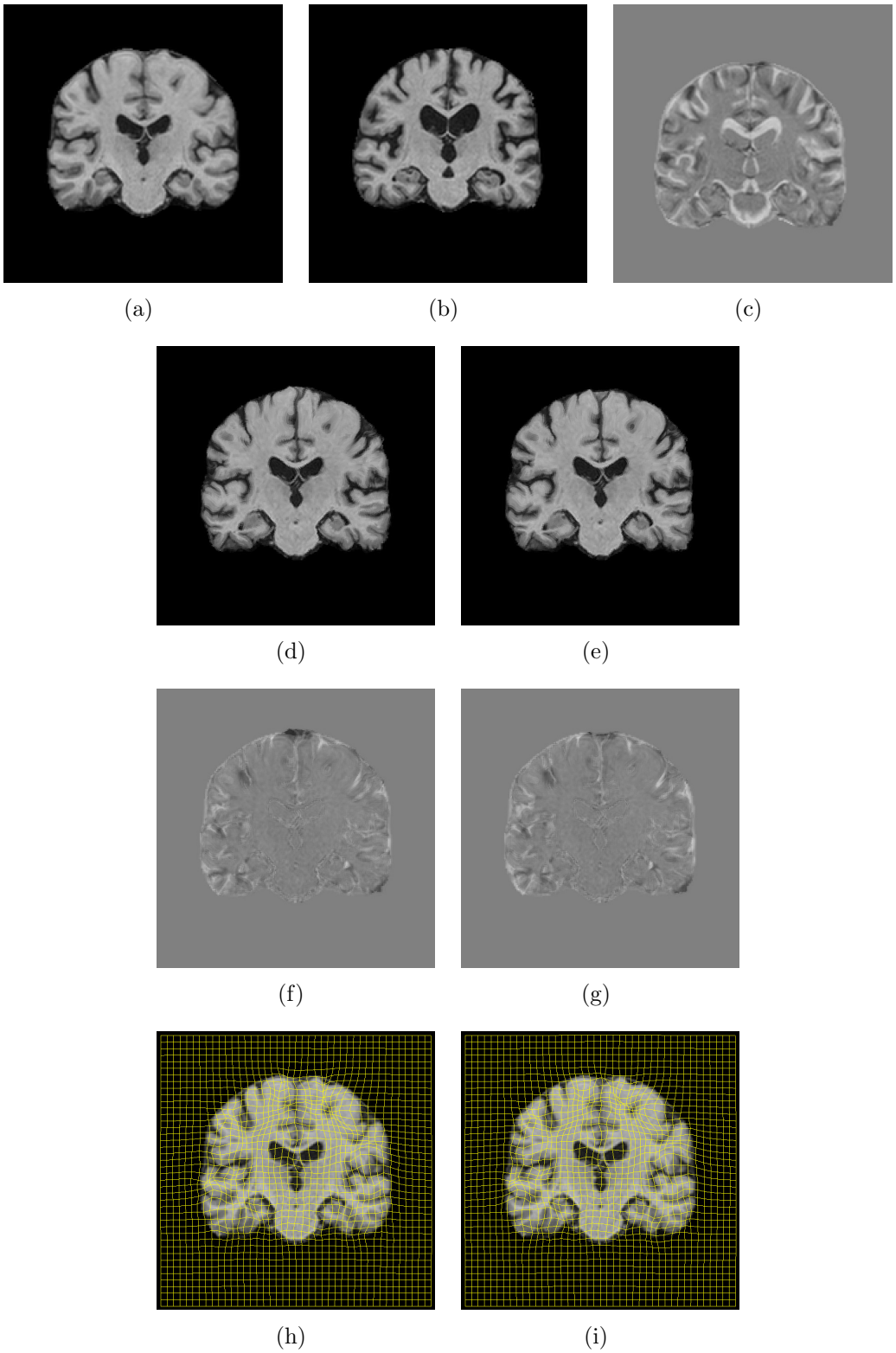


Figure A.2: Registration results using both exact and approximate gradients (coronal view). (a): Target image. (b): Source image. (c): Difference image. (d): Warped source using exact gradient. (e): Warped source using approximate gradient. (f): Difference image using exact gradient. (g): Difference image using approximate gradient. (h): Resulting FFD using exact gradient. (i): Resulting FFD using approximate gradient.

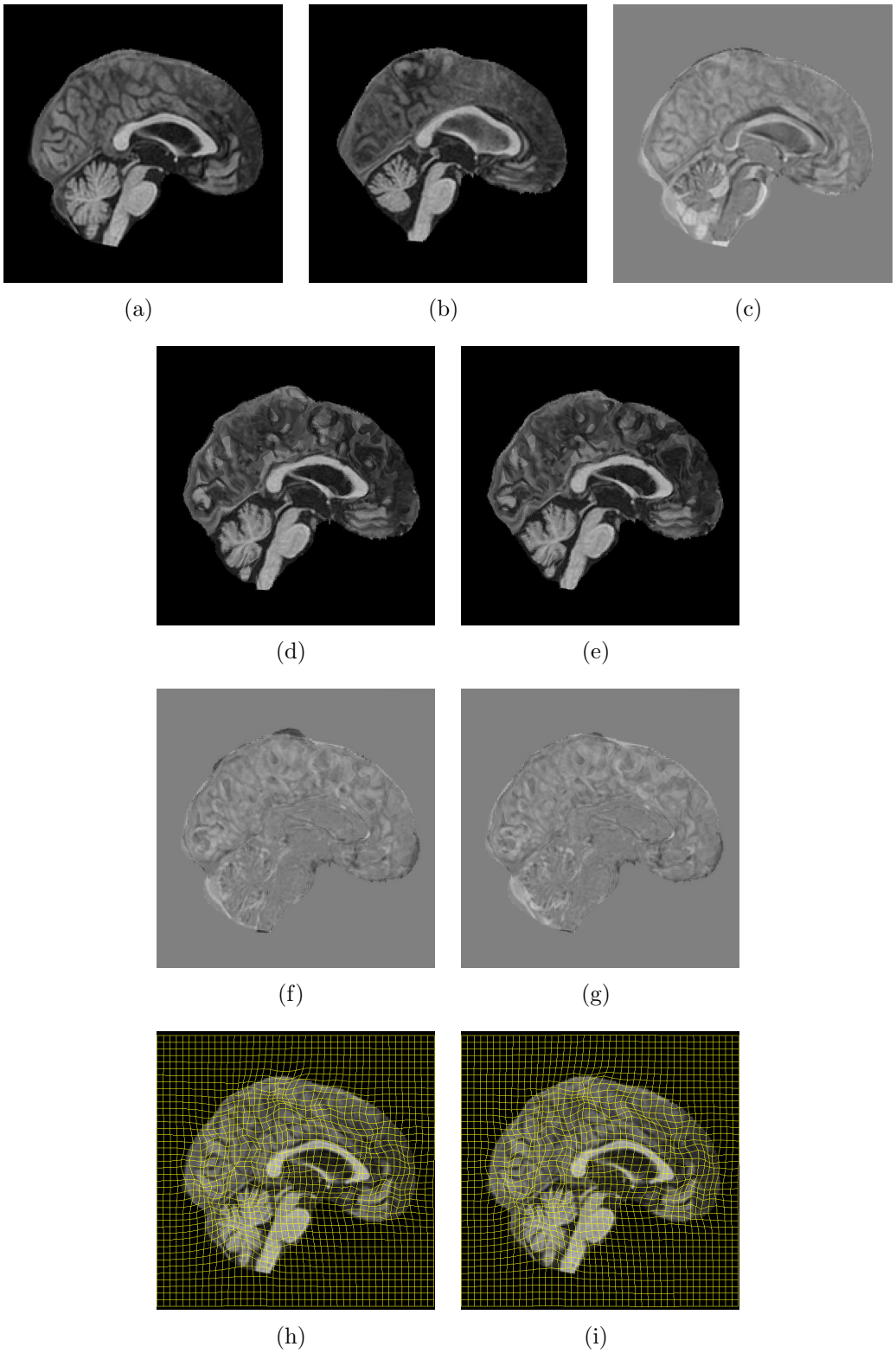


Figure A.3: Registration results using both exact and approximate gradients (sagital view). (a): Target image. (b): Source image. (c): Difference image. (d): Warped source using exact gradient. (e): Warped source using approximate gradient. (f): Difference image using exact gradient. (g): Difference image using approximate gradient. (h): Resulting FFD using exact gradient. (i): Resulting FFD using approximate gradient.