Multi-Atlas based Segmentation of Multi-Modal Brain Images

by

Keyvan Kasiri

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Acknowledgements

I would like to express my sincere gratitude to my supervisors Professor David Clausi and Professor Paul Fieguth for their guidance, advice, and moral support throughout my Ph.D. studies under their supervision. They have contributed enormously to my growth as a researcher.

I wish to thank my doctoral committee members, Prof. Daniel Stashuk, Prof. Jeff Orchard, Prof. Ed Vrscay, and Prof. Anant Madabhushi for their valuable comments and suggestions.

I wish to acknowledge the University of Waterloo Faculty of Engineering, and the Natural Sciences and Engineering Research Council (NSERC) of Canada for financial support of my research.

I would also like to thank my friends and my colleagues at University of Waterloo for their support during these years.

Finally, I would like to express my deepest gratitude and love to my family for their unconditional love and support. My special gratitude and love goes to my parents, my beloved wife Nazanin, and my dear brother Iman for all their continuous support and encouragement.

Dedication

To my loved ones, my beloved parents, my dear wife, and my brother.

Abstract

Brain image analysis is playing a fundamental role in clinical and population-based epidemiological studies. Several brain disorder studies involve quantitative interpretation of brain scans and particularly require accurate measurement and delineation of tissue volumes in the scans. Automatic segmentation methods have been proposed to provide reliability and accuracy of the labelling as well as performing an automated procedure.

Taking advantage of prior information about the brain's anatomy provided by an atlas as a reference model can help simplify the labelling process. The segmentation in the atlas-based approach will be problematic if the atlas and the target image are not accurately aligned, or if the atlas does not appropriately represent the anatomical structure/region. The accuracy of the segmentation can be improved by utilising a group of atlases. Employing multiple atlases brings about considerable issues in segmenting a new subject's brain image. Registering multiple atlases to the target scan and fusing labels from registered atlases, for a population obtained from different modalities, are challenging tasks: image-intensity comparisons may no longer be valid, since image brightness can have highly differing meanings in different modalities.

The focus is on the problem of multi-modality and methods are designed and developed to deal with this issue specifically in image registration and label fusion. To deal with multi-modal image registration, two independent approaches are followed. First, a similarity measure is proposed based upon comparing the self-similarity of each of the images to be aligned. Second, two methods are proposed to reduce the multi-modal problem to a mono-modal one by constructing representations not relying on the image intensities. Structural representations work on the basis of using un-decimated complex wavelet representation in one method, and modified approach using entropy in the other one. To handle the cross-modality label fusion, a method is proposed to weight atlases based on atlas-target similarity. The atlas-target similarity is measured by scale-based comparison taking advantage of structural features captured from un-decimated complex wavelet coefficients. The proposed methods are assessed using the simulated and real brain data from computed tomography images and different modes of magnetic resonance images. Experimental results reflect the superiority of the proposed methods over the classical and state-of-the art methods.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANN	artificial neural networks
CC	cross correlation
cMI	conditional mutual information
CoCoMI	contextual conditioned mutual information
CR	correlation ratio
CSF	cerebrospinal fluid
CT	computed tomography
DoF	degree of freedom
DT-CWT	dual tree-complex wavelet transform
eSSD	entropy sum of squared differences
FFD	free-form deformation
fMRI	functional magnetic resonance imaging
Gm	gradient magnitude
GM	gray matter
IR	infra-red
LMI	local mutual information
LWV	local weighted voting
MI	mutual information
MIND	modality independent neighbourhood descriptor
MR	magnetic resonance
MRF	Markov random field
MRI	magnetic resonance imaging
MV	majority voting
NCC	normalised cross correlation
NLM	non-local means
NMI	normalised mutual information
PC	phase congruency
PD	proton density
PET	positron emission tomography

RIRE	retrospective image registration evaluation
SAD	sum of absolute differences
SeSaMI	self-similarity α-mutual information
SSD	sum of squared differences
SPECT	single photon emission computed tomography
TPS	thin plate spline
UDWT	undecimated wavelet transform
WM	white matter

List of Symbols

A	atlas
B	B-spline function
D	Dice coefficient
$\mathcal{D}$	pixel descriptor
$\mathcal{D}_{\mathsf{sort}}$	sorted pixel descriptor
D_p	patch distance
$\tilde{D}_p$	sorted patch distance
E	Energy Function
f	pairwise pixel self-similarity function
f_s	complex sinusoid function
f_g	2D Gaussian function
f_R	representation function
F	spatial transformation
$\mathcal{F}$	label fusion function
G	Gaussian kernel
G_x	gradient along x direction
G_y	gradient along y direction
G_m	gradient magnitude
h	weighted pixel information
H	entropy of a random variable
$\tilde{H}$	modified entropy of a random variable
I	image
I_f	fixed image
I_m	moving image
I_T	target image
L	label
m	order of polynomial function
M	number of the most similar pixels
MI	mutual information of two random variables
N	number of pixels

$\mathcal{N}$	spatial neighbourhood
N_A	number of atlases
N_b	number of neighbourhoods
NL	denoised image using non-local means
NMI	normalised mutual information of two random variables
$P_{\mathbf{x}}$	patch centring $\mathbf{x}$
$\tilde{P}_{\mathbf{x}}$	sorted patch centring $\mathbf{x}$
p	probability density function
PC	phase congruency
R_c	complex wavelet representation
R_e	entropy representation
R_{Me}	modified entropy representation
s	Scale
$\mathcal{S}$	pixel self-similarity
SM	pixel similarity
$\mathrm{SM}_{\mathcal{F}}$	similarity in label fusion
T_r	threshold
w	weight
W	weight set
W^{PC}	phase congruency weight
Z_n	normalisation factor
μ	mean of a random variable
σ	standard deviation
α	amplitude of complex wavelet coefficient
ϕ	phase of complex wavelet coefficient
ω	frequency
γ	Gabor filter
γ^e	even-symmetric Gabor filter
γ^o	odd-symmetric Gabor filter
ζ	phase order
ρ	similarity measure
θ	angular orientation

ψ polynomial function
χ self-similarity map construction function
Ω image grid
Γ log-Gabor filter
Υ complex wavelet transform response
Ξ scale-based label fusion function

Chapter 1

Introduction

Brain image analysis is playing a fundamental role in clinical and population-based epidemiological studies. Several brain disorder studies involve quantitative interpretation of brain scans and particularly require accurate measurement and delineation of tissue volumes in the scans [1, 2, 3, 4, 5]. Manual labelling of brain images by human experts is inconsistent and time-consuming, specifically for large datasets [6]. Automatic segmentation methods have been proposed to provide reliability and accuracy of the labelling as well as performing an automated procedure.

Automatic segmentation of brain images is a challenging task due to undesirable artefacts such as noise, partial volume effect or non-uniformity in the intensity of the image. Therefore, using a priori information about the anatomy of the brain, which is provided by a reference image/volume, called an *atlas*, can help simplify this procedure [7]. In the literature, the term 'atlas' is referred to both an intensity image, which is a brain template, or the segmented image, which is the labelled one [7, 8].

In traditional atlas-based segmentation, a target scan is labelled by referring to an atlas where the target is aligned to the atlas using deformable registration and atlas labels are then propagated to the target image space [9]. However, if either the mapping between images is not accurate or the atlas is not anatomically an appropriate representative for a specific structure/region, the segmentation will be problematic. Fig. 1.1 illustrates the process of atlas-based segmentation used for delineation of brain tissues. The atlas-based

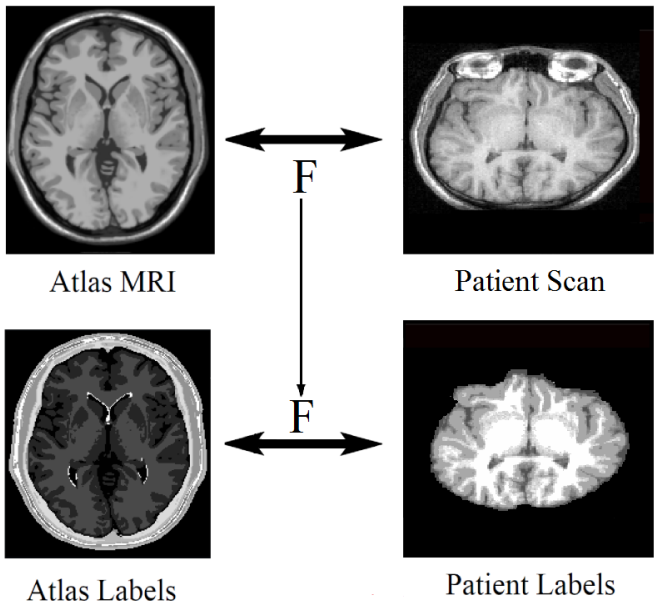


Figure 1.1: Block diagram illustrating the atlas-based segmentation procedure used for brain tissue segmentation. Segmentation is based on registering the atlas to the target patient image and using the resulting spatial transformation F to propagate atlas labels to target space to attain a segmentation.

segmentation is shown as a registration-based segmentation approach where F stands for the spatial transformation between the atlas and the target scan.

The error caused by any single atlas registration will be effectively reduced by using a group of segmented images. There are two different categories of approaches in using multiple atlases for segmentation. In the first class of methods, information from several atlases are combined to create an average or a probabilistic atlas [8, 10, 11]. Then, the constructed atlas is warped to the target image to provide prior information. The second category of methods tries to combine labels from some number of registered atlases [11, 12, 13]; this work has led to an active literature on *multi-atlas* approaches [13, 14].

Figure 1.2: Multi-atlas segmentation approach: The overall block-diagram of multi-atlas segmentation procedure and its major components. Atlas selection is shown in a dashed box as an optional step in the multi-atlas segmentation framework.

1.1 Multi-modal Multi-Atlas Segmentation Problem

The multi-atlas approach takes advantage of more information from multiple atlases and is more robust to anatomical variations than single atlas-based approach [12, 15]. The multi-atlas segmentation approach can be subdivided into several steps. In general, key steps of any multi-atlas segmentation framework are atlas generation, registration, atlas selection, label propagation, and label fusion. These components are generally implemented sequentially in independent steps, however, there are many exceptions to this sequential organization. The overall block-diagram of multi-atlas segmentation procedure and its major components are presented in Fig. 1.2. Here, several already segmented images from different subjects, i.e., atlases, are registered to the patient input image resulting in a set of transformations. A subset of registered atlases may be selected to either reduce the complexity or exclude irrelevant atlases. Atlas labels are required to be propagated to the target space using the obtained transformation. Then, propagated labels are fused for each pixel to form a final segmentation result. Atlas selection is not a necessary step in every multi-atlas segmentation framework, and therefore it is shown in dashed line in this figure.

The multi-atlas approaches are promising, however, these methods remain problematic in those cases where the atlases and the target scan are obtained from different sensors or from different acquisition modalities: image-intensity comparisons may no longer be valid, since image brightness can have highly differing meanings and circumstances in different modes [16, 17]. The goal of this thesis is to focus on the multi-modality issue and design and develop methods to deal with this issue specifically in major steps of the multi-atlas segmentation framework: image registration and label fusion.

1.2 Challenges

As described in Section 1.1, the general form of multi-atlas segmentation framework consists of major steps of atlas generation, registration, label propagation, and label fusion. Since in most cases atlases, i.e., segmented scans, are already available, we skip atlas generation for the rest of thesis. To deal with cross-modality in the multi-atlas segmentation problem, the major components to cope with the issue of intensity variation are registration and label fusion. Thus, the major challenges to address in this problem are

- **Multi-modal registration:** To segment the target image, the atlases, which might exploit multiple imaging modalities, are required to be registered to the target space. The intensity variations across modalities has been an issue in the multi-modal registration. Statistical metrics, such as those based on mutual information (MI), have been proposed in the literature as the solution to address this issue [18, 19, 20]. However, MI-based measures are intrinsically global and therefore may suffer from many false local optima. Moreover, the optimisation of these statistical measures for registration is computationally more complex compared to simple intensity difference metrics [20]. This can be more of a concern when the number of atlases to be registered are increasing in the database [14].

- **Cross-modality label fusion:** A key challenge associated with the multi-atlas approach is label fusion. Most label fusion approaches are limited by the assumption that they depend on the consistency of voxel intensities across different scans. Many label fusion methods, such as majority voting (MV) [13] and weighted voting [21, 22, 23] do not consider image intensities after being warped to the target image. The multi-atlas approaches are promising, however, these methods remain problematic in those cases where the atlases and the target scan are obtained from different acquisition modalities: image-intensity comparisons may no longer be valid, since image brightness can have highly differing meanings and circumstances in different modes [16].

1.3 Objectives and Contribution

The objectives of this thesis target the multi-modal registration and cross-modality label fusion in a multi-atlas segmentation framework. The thesis makes the following contributions:

- Defining a novel similarity measure based on measuring the image self-similarity for registration of multi-modal images, which is described in Chapter 4 and evaluated in Chapter 6,
- Reducing the multi-modal registration problem to a mono-modal one and hence, lowering the complexity of the registration problem by proposing structural representations not relying on the intensity mapping, which is described in Chapter 5 and evaluated in Chapter 6,
- Extending the existing label fusion approach to cross modality multi-atlas segmentation by making cross-modality image comparison based on extracted structural features, which is described and assessed in Chapter 7.

1.4 Thesis Outline

The structure of this thesis closely follows the sequence of mentioned contributions.

In Chapter 2, we present an overview of the atlas-based segmentation and multi-atlas-based approach. A review of methods in image registration and label fusion as two major components of multi-atlas framework is also presented in this chapter.

Chapter 3 states and formulates the problem we are targeting in multi-atlas-based segmentation. Challenges and limitations related to the existing approaches followed by the objectives and contributions in this problem are presented.

Chapter 4 presents a new similarity measure for registering multi-modal images. The concept of self-similarity and measures for multi-modal image registration is presented. Following that, we present the proposed self-similarity measure based on taking the most significant image self-similarities into account.

In Chapter 5, two independent image representations are presented to map multi-modal images into common intensity space. First, complex wavelets is used to present the proposed image representation. Second, independent of the first representation, a modification to the formulation of entropy is applied to build an alternative structural representation.

Experiments to measure the accuracy of multi-modal image registration based on structural representations are presented in Chapter 6. Structural representations in Chapter 5 based on complex wavelets and modified entropy are assessed in the same framework but independent of each other. In the following, employing the self-similarity presented in Chapter 4 is evaluated in the multi-modal image registration framework.

In Chapter 7, the problem of cross-modality label fusion is of focus. The weighted voting label fusion followed by the proposed method for combining labels from multi-modal images are presented. Experiments to evaluate the proposed method comparing with the conventional approach are given later in this chapter.

Chapter 2

Background

This chapter is devoted to reviewing the materials and methods required for the purpose of segmentation of MR images based on using multiple atlases. First, in Section 2.1, a general overview of brain tissue segmentation and different approaches are explained. Second, in Section 2.2, a generic form of atlas-based approach and its components are presented. Third, the multi-atlas-based approach, as a specific case of atlas-based segmentation, its components, and related challenges are presented in Section 2.3. Lastly, the problem of dealing with multiple modalities in this approach is given in Section 2.4.

2.1 Brain Tissue Segmentation

Segmentation is the process of partitioning an image into constituent regions whose elements (pixels in each region) have the same characteristics, such as color, intensity, or texture [6, 7, 24]. Since most studies on medical data highly rely on large datasets, a manual image segmentation approach by a human expert is a time-consuming procedure. Moreover, a manual approach highly depends on intra- and inter-observer variability which results in the degradation of credibility in the segmentation analysis. Therefore, attempts have been made towards an automatic segmentation of medical images to provide a reproducible, accurate, and robust segmentation framework.

Image segmentation, from methods to applications, has been addressed in the liter-

ature [6, 7, 24, 25, 26]. Pham *et al.* categorised segmentation methods into eight main categories of thresholding, region growing, pattern recognition methods, clustering, Markov random field (MRF) model, artificial neural network (ANN) methods, deformable models, atlas-based and other methods [6].

Among them, atlas-guided approaches aim to reduce human interaction and have a fully automatic and accurate segmentation approach. This category of methods, which is described in more detail in Section 2.2, incorporates additional higher level knowledge that can be prior information about the image under consideration or any predefined model [15, 25]. The atlas, which is generally a segmented image, is used as a reference model for the image to be segmented. The simplest atlas-based paradigm finds a one-to-one mapping between the atlas and the image to be segmented. Using the one-to-one mapping, all information available in the atlas is transferred to the target image to help label the image [8]. The typical atlas-based method along with different types of atlases and segmentation strategies are explained in the following.

2.2 Atlas-Based Segmentation

The automatic segmentation of brain images has been always a challenging problem [8, 27, 28, 29, 30]. Therefore, using *a priori* information about the anatomy of the brain can help simplify this procedure. Prior information can be provided by a reference model, called an atlas, which either is a manually segmented version of brain scans where contains label information at specific locations.

In atlas-based segmentation, the segmentation problem turns into a registration one. The atlas, A, is registered to the target patient image, I_T, resulting in a transformation F. Using the transformation F, labels of the atlas, denoted as L, are then propagated to the target image space. However, if either the atlas is not anatomically an appropriate representative for a specific structure/region or there exist labelling errors in the atlas segmentation, the error will be propagated during the registration procedure. In the following, two types of atlases as well as approaches under the atlas-based category are explained in more detail.

2.2.1 Types of Atlases

The construction and application of brain atlases are of great importance in neuroimaging and human brain research [8, 29, 31, 32]. This is due to the need for a standardized template which is the key concept in the field of human brain mapping. Creation of a realistic brain atlas, considering anatomical details and variability, is a time-consuming step. Therefore, many efforts have been recently made to provide this field of research with manually segmented data.

Topological Atlases: The first version of the atlas constructed for human brain research is the topological atlas which, in the literature, is also called the brain template, single-subject, or deterministic atlas. The topological atlas is referred to a volume image chosen from a population of brain scans to represent the whole population in terms of size, shape or intensity. The construction of a template to describe how different parts and structures are organized in the brain is the first step in creation of any probabilistic, region or disease-specific atlases.

The first attempt in creating atlas of the human brain led to the Talairach atlas [31] by which deep brain structures were identified in a space independent from individual differences in the size and overall shape of the brain. Fig. 2.1 shows an example of the deterministic atlas which is a brain template from the BrainWeb simulated brain database [33]. This image indicates the 143^{th} axial slice of one of the twenty anatomical models of 20 normal brains. In each model, a set of "fuzzy" tissue membership volumes is presented. This set consists of different classes of background, cerebrospinal fluid (CSF), gray matter (GM), white matter (WM), fat, muscle, muscle/skin, skull, blood vessels, connective (region around fat), dura matter and bone marrow.

Probabilistic Atlas: The major factor which is not considered in deterministic atlases is the diversity of human brain anatomy. In order to address the anatomical variability in the human brain, a population of brain scans is used to form the brain atlas. This type of atlas is often referred to as population-based, probabilistic, or statistical atlas [8]. In the construction of probabilistic atlases, the population can be subdivided into different groups based on different factors such as age, sex, or handedness. Such a population-based atlas is

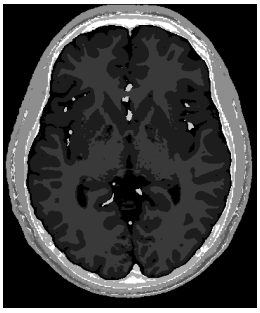


Figure 2.1: An example of deterministic atlas: a slice of a 3D anatomical model of a normal brain from the BrainWeb [33] database. A set of different tissue classes are distinguishable by using different gray-scale values. The gray scale values from dark to bright indicate twelve classes of background, CSF, GM, WM, fat, muscle, muscle/skin, skull, vessels, around fat, dura matter, and bone marrow.

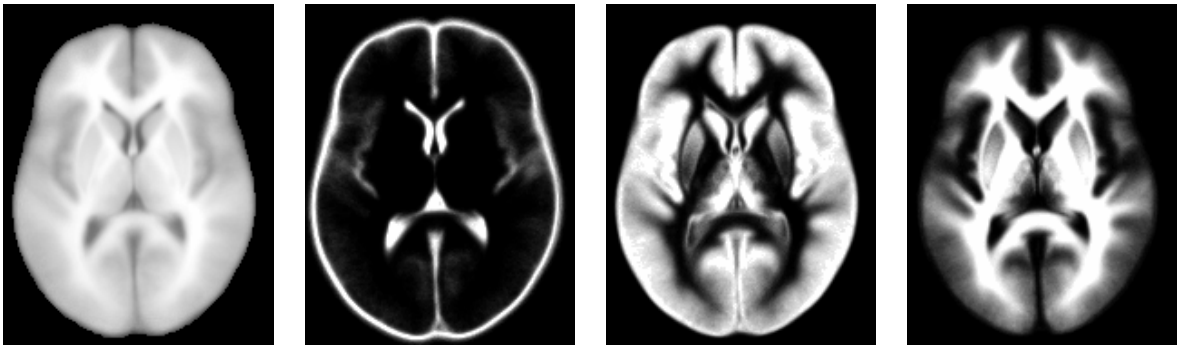
constructed using a set of segmented MRI data sets. For this purpose, all segmented images in the database are registered into a standard space and then the tissue probability of each voxel for a specific structure or region is computed. In Fig. 2.2, a sample probabilistic atlas for brain tissues is shown. This figure shows the 74th axial slice of the ICBM452 [34] atlas from the LONI database [35] which includes T1 mean, WM, GM and CSF probability maps.

2.2.2 Segmentation Strategies

The atlas-based segmentation approach tries to deform a brain atlas into a patient's brain scan to create a labelled version of patient's scan. The so-called atlas is a labelled scan which is previously segmented.

To use *a priori* information available in the atlas A, a transformation is required to map the atlas space into target image I_T space which forms a registration problem. Having found the transformation F from atlas space into target space, it is possible to map the reference (atlas) labelled image L to the patient's image (target) space and obtain the labelled version of patient's scan L_T. The labelled volume is defined by $\mathcal{L}$ unique segments:

$$L_T(\mathbf{x}) \in \{1, \ldots, \mathcal{L}\}, \tag{2.1}$$



T1 average CSF GM WM

Figure 2.2: An example of probabilistic atlas: ICBM452 [34] probabilistic atlas showing the average topology of the brain and probabilistic map of CSF, GM, and WM.

where $\mathbf{x}$ is the location in the label map L corresponding to the same location in atlas A.

Label Propagation

Having done the registration step, the easiest and fastest way to do the final labelling process is to propagate atlas labels to the input image space. In typical label propagation, the estimated transformation $\hat{F}$ resulting from the registration step is used to deform the atlas labels, then the labels mapped to the coordinate system of the input image are simply assigned to input image voxels:

$$L_T(\mathbf{x}) = L\big(\hat{F}(\mathbf{x})\big). \tag{2.2}$$

In this way, the labelling error relies on the error that happened at the registration step and the whole segmentation procedure will basically be transformed into a registration problem. Since large anatomical differences will lead to a large registration error, this method is feasible for the cases in which the atlas is sufficiently similar to the input image.

When dealing with intra-subject registration in medical applications, such as registration of multi-modal images for radiotherapy or progression in a specific disease, global rigid registration and affine transformation will perform sufficiently well. Inter-subject registration which highly involves anatomical variations requires high degree of freedom

and therefore more complicated methods, non-rigid registration techniques, are employed. However, the risk of getting stuck in local extrema during the optimization procedure will be increased [8].

Probabilistic Atlas-based Segmentation

Typically, probabilistic atlases are used in a Bayesian framework to maximise the conditional probability of intensities in each class. The classical Bayesian approach for classification is defined by

$$\hat{L}(\mathbf{x}) = \underset{l\in\{1,\cdots,\mathcal{L}\}}{\operatorname{argmax}}\, p\big(L(\mathbf{x}) = l|A(\mathbf{x})\big) = \underset{l\in\{1,\cdots,\mathcal{L}\}}{\operatorname{argmax}}\, p\big(I(\mathbf{x})|L(\mathbf{x}) = l\big) \cdot p(L(\mathbf{x}) = l), \qquad (2.3)$$

where $p\big(I(\mathbf{x})|L(\mathbf{x}) = l\big)$ stands for conditional probability of the voxel intensities given the class label and $p(L(\mathbf{x}) = l)$ represents the label prior. In this approach, class priors are provided by the probabilistic atlas and either parametric or non-parametric methods can be used to estimate the conditional probability.

Multi-Atlas Label Propagation

In a typical label propagation, when the atlas anatomy is far different from the input patient image, the accuracy of the segmentation will decrease. To overcome the registration error and therefore improve the segmentation accuracy, one possible solution is to employ multiple atlases. As was first shown by Heckemann *et al.* [13], as new atlases are taken into consideration, the accuracy of segmentation procedure will increase. Not only is the number of atlases used in the segmentation important to have an acceptable segmentation accuracy, but also the quality of atlases.

The first important issue associated with multi-atlas-based segmentation is the number of atlases and also how to choose them. Atlases should be selected in such a way that maximum anatomical variety in a population of atlases can be achieved. If a large database of atlases is available, the more efficient way will be selecting a subset of atlases which are very close to the input image to be segmented in terms of similarity. Further improvements are achieved by clustering atlases into different classes based on different structures and organs. Atlas ranking is another possibility to deal with using multiple atlases.

Another important issue in multi-atlas-based segmentation is the number of registrations required for segmentation. Typically, all atlases are warped into a common space to reduce the number of registrations and hence reduce the computations. However, the result will always be biased towards the initial selected space. For this reason, groupwise registration techniques are employed to suggest a better way for this problem. These methods try to build an average reference template and register all of available atlases into this common space.

Having aligned all atlases, all deformed labels should be combined in some way. This step can be considered as a specific case of classifier fusion. Weighted voting is the typical way to apply on warped labels which are used both globally and locally.

2.3 Multi-Atlas-Based Segmentation

As described in Section 2.2.2, in multi-atlas-based segmentation approach, each atlas is available and potentially utilised for segmenting the target image. The overall framework of the approach for segmentation of medical images is illustrated in Fig. 2.3. The conventional approach involves registering each atlas A_i, $i = 1, \cdots, N_A$, from a database of N_A atlases, to the target (patient's) image I_T, propagating the atlas labels L_i, $i = 1, \cdots, N_A$, to the target image coordinates, resulting in atlases and labels in the target image coordinates, A'_i and L'_i, and then fusing the propagated labels. This section focuses on registration and label fusion as the main components of multi-atlas-based segmentation procedure.

2.3.1 Image Registration

Image registration, which is also named image matching or alignment, is the process of aligning two or more different images by finding one-to-one spatial correspondence between images [36]. Image alignment, as an image processing step, plays an important role in processing 2D/3D data in a variety of applications including robot vision, remote sensing, and medical imaging [37, 38, 39]. In particular, image registration is considered one of the fundamental problems in processing of medical images. Tracking temporal evolution and change detection, fusing image data, and 3D image construction are some examples medical applications [37, 39].

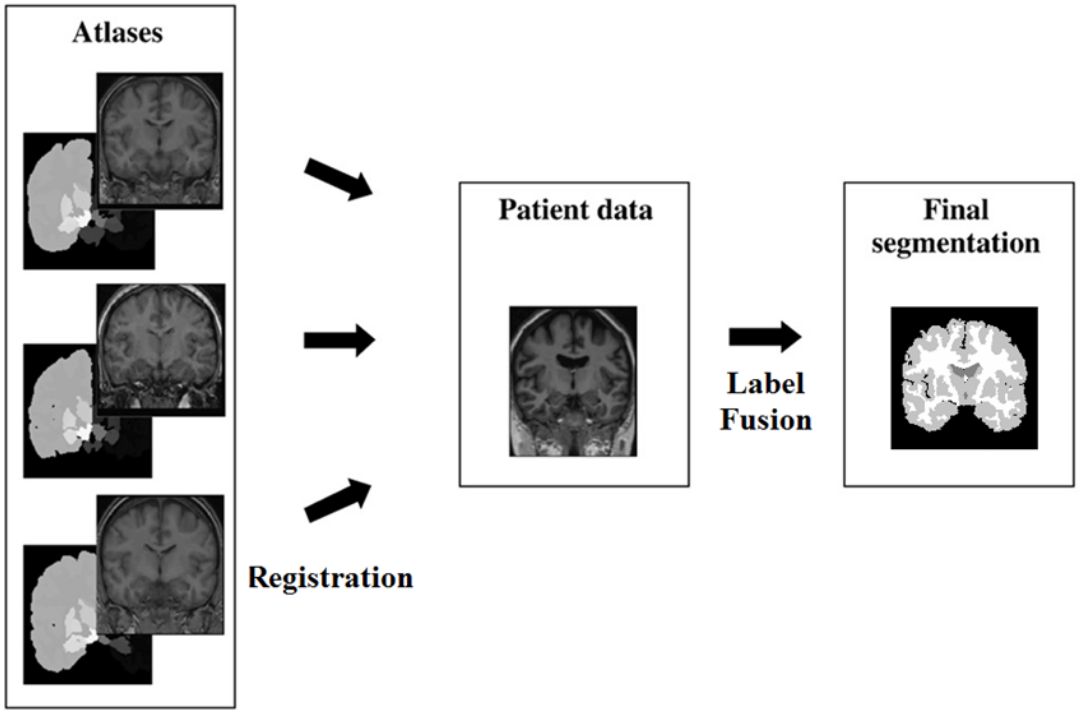


Figure 2.3: Multi-atlas-based segmentation procedure.

The process of registering images in the particular case of medical applications becomes more challenging due to the variety of the imaging modalities and the fact that each modality can deliver the particular type of information [40]. For example, in medical imaging, some modalities provide anatomical information (i.e., computed tomography (CT) and MRI) and some other provide functional information (i.e., positron emission tomography (PET), single photon emission computed tomography (SPECT), and functional MRI (fMRI)) about a specific tissue, structure or organ [41]. The anatomical information provides clinicians with spatial information such as shape, size and spatial relationship between structures and pathology, while the functional information leads clinicians

to studying the relationship between the underlying structure and physiology. Moreover, establishing a model for the relationship between images of human organs or structures is quite difficult, due to the highly complex transformations required.

To overcome the problems and challenges related to registering medical images, different approaches have been proposed in the literature [20, 36, 37, 40, 42]. In this subsection, an overview of the framework for medical image registration and its fundamental components are introduced.

In general, a registration framework involves finding a deformation transform F from a moving image I_m to a fixed image I_f in order to maximise (minimise) an objective (cost) function ρ. The cost function combines a measurement of spatial alignment with a regulariser that quantifies the plausibility of the deformation:

$$\hat{F} = \underset{F}{\operatorname{argmax}}\, \rho\big(I_f, F(I_m)\big) \tag{2.4}$$

Thus, the three main component of registration framework are the deformation model, the objective function, and the optimizer.

Transformation Model

Transformation models are geometric models that establish a one-to-one mapping between the moving I_m and fixed I_f domains. The transformation model used during the registration process relies on the accuracy to be satisfied, the deformation and the images to be registered. These models can be classified into three fundamental categories; *rigid*, *affine*, and *non-rigid* transformations.

Rigid transformation in three dimensions involves three degrees of freedom (DoFs) for rotation and three for translation. Transformation function can be expressed in matrix form as

$$F_{rigid}(x, y, z) = \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}, \tag{2.5}$$

where r_{ij} determine rotations about each coordinate axis and t_x, t_y, and t_z stand for the translation along x, y, and z axes, respectively.

In addition to translation and rotation expressed in rigid transformation, scaling and shearing may be also necessary for aligning images. The matrix form of scaling transformation in a 3D space and a shearing matrix in the (x, y) plane can be expressed in the following way:

$$F_{scale} = \begin{pmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.6}$$

$$F^{xy}_{shear} = \begin{pmatrix} 1 & 0 & h_x & 0 \\ 0 & 1 & h_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{2.7}$$

where s_x, s_y and s_z stand for the scaling in each of the coordinate axes, and h_x, h_y represent the shearing in each of those axes. The overall linear mapping to cover the rigid, shearing, and scaling transformations is affine transformation that can be obtained by multiplying the rigid transformation, scaling and shearing matrices:

$$F_{affine}(x, y, z) = F_{shear} \cdot F_{scale} \cdot F_{rigid} \cdot \begin{bmatrix} x & y & z & 1 \end{bmatrix}^T. \tag{2.8}$$

The resulting transformation provides twelve DoFs specifying translation, rotation, scaling and shearing.

In medical image registration, it is common to use rigid transformations to relate images when registering images of rigid parts of the body such as bones. Rigid models are global in nature and are not able to model local differences between images. Since rigid and affine models are of low complexity, they are often limited to registration of rigid structures and organs or only used as a pre-registration process prior to more complex registration procedures [36]. Since human body organs and structures are mostly deformable structures, *non-rigid* registration approaches are used in medical applications to build flexible elastic models [36, 40].

Basically, two types of deformations are considered in medical image registration: free-form and guided deformations. In free-form deformation models, any kind of deformation is allowed, whereas guided deformations are controlled by a physical model caused by the material properties of the organ or structure [43, 44, 45].

In free-from deformation (FFD) approaches, the registration is mainly performed by defining a grid of control points to determine the deformation between images. For the point located between the grid points, the deformation vector is obtained using any of interpolation methods. The use of B-spline tensor products as the deformation function was first proposed by Rueckert *et al* [45]. If the domain of the image volume is defined as

$$\Omega = \{\mathbf{x} = (x, y, z) | 0 \leq x < X, 0 \leq y < Y, 0 \leq z < Z\}, \tag{2.9}$$

the transformation field by FFD with mesh of control points $\mathbf{d}_{i,j,k}$ with uniform control point spacing δ can be expressed as the 3D tensor product of the 1D cubic B-splines:

$$F(\mathbf{x}) = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \mathbf{d}_{i+l,j+m,k+n} \tag{2.10}$$

where B_l represents the l-th basis function of the B-spline, $i = \lfloor \frac{x}{\delta} \rfloor - 1$, $j = \lfloor \frac{y}{\delta} \rfloor - 1$, $k = \lfloor \frac{z}{\delta} \rfloor - 1$, $u = \frac{x}{\delta} - \lfloor \frac{x}{\delta} \rfloor$, $v = \frac{y}{\delta} - \lfloor \frac{y}{\delta} \rfloor$, and $w = \frac{z}{\delta} - \lfloor \frac{z}{\delta} \rfloor$. This deformation model requires a few degrees of freedom to describe local deformations and can efficiently provide smooth deformations.

Guided deformation models such as elastic models consider objects in the image as elastic solids [46, 47]. Therefore, the model is defined based on internal and external forces applied to the deformation fields. The internal static forces are applied to oppose the deformation, while the external forces caused by similarity metric helps the deformation to fit the configuration. Both forces are applied to deform the image until they reach an equilibrium. Guided deformations are non-parametric models that characterise the deformation at every voxel of the image volume.

Objective Function

The objective function is typically based on either metrics that measure the degree of similarity or the spatial distance between corresponding landmarks to quantify the accuracy of alignment in image registration. In the latter case, the landmarks are manually placed or detected automatically before performing the alignment. Similarity measures can be classified into intensity- and feature-based categories.

Measures based on image *intensity* in image registration [48] are usually based on intensity differences, intensity cross correlation, and information theory [48, 49]. The simplest intensity-based measure is based on sum-of-squared-differences (SSD) between the intensities in I_1 and I_2:

$$\rho_{SSD} = \sum (I_1 - I_2)^2. \tag{2.11}$$

Metrics based on intensity difference are basically assuming the same characteristics for the images to be aligned and restricted to uni-modal image registration. A more general assumption than of having identical modalities is to have a linear relationship between image intensities. In this case, similarity can be measured using normalised cross correlation (NCC) as

$$\rho_{NCC} = \frac{\sum (I_1 - \mu_1)(I_2 - \mu_2)}{\sqrt{\sum (I_1 - \mu_1)^2 \sum (I_2 - \mu_2)^2}} \tag{2.12}$$

where μ_1 and μ_2 are the average pixel intensities in the images I_1 and I_2, respectively. Nevertheless, the NCC is largely restricted to applications in registering mono-modal images.

Information theoretical metrics such as mutual information [20], which are based on Shannon's entropy [50], can be applied to both uni- and multi-modal registration frameworks and measure how well one image is able to explain the other image. Mutual information for two images I_1 and I_2 is defined based on the Shannon entropy as

$$\mathrm{MI}(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2) \tag{2.13}$$

where $H(I_1)$ and $H(I_2)$ represent the entropy of random variables I_1 and I_2, and $H(I_1, I_2)$ stands for the joint entropy of these two random variables. MI can be equivalently expressed as

$$\mathbf{MI}(I_1, I_2) = \sum_i \sum_j p(i,j) \log \frac{p(i,j)}{p(i)p(j)}, \tag{2.14}$$

where $p(i,j)$ is the joint probability distribution function of I_1 and I_2, and $p(i)$ and $p(j)$ are the marginal probability distribution functions of I_1 and I_2 respectively.

Feature-based metrics are usually based on landmarks, salient points, edges, contours, corners and/or surfaces [48, 49]. Distances between the corresponding features are considered as a criterion to measure the alignment. It is required to extract features and estimation of correspondences prior to computing the distance. As an advantage of using feature-based registration is that it can be also used for multi-modal registration. However,

feature-based registration may need a prior segmentation to extract landmarks or features in the images. Furthermore, errors produced during the feature extraction procedure will be propagated into the registration and affect the accuracy of the procedure [36, 40, 42].

Numerical Optimization

The problem of image registration can be expressed as an optimization problem in which the goal is to minimise the cost or maximise the similarity between two images. The method tries to search for the optimum of an objective/cost function in the mapping model. Choosing a global or local optimization technique depends on the form of the objective/cost function, computational complexity, robustness, speed of the algorithm, and the accuracy required for the underlying application [36, 40, 49].

In the case of rigid and affine transformations, there is no constraint as the cost function and the optimisation problem aims to maximise the similarity between images. In non-rigid transformations, the role of the cost function plays the role of regularization or penalty term to constrain the transformation relating both images [36].

A common family of optimisation approaches is based on gradient descent that optimise the objective function by following the negative energy gradient, the direction that decreases the energy. Gradient descent has been utilised to solve various registration problems including the FFD registration algorithm. Conjugate gradient, Gauss-Newton method, stochastic gradient descent, and graph-based methods are the examples of approaches that have been used widely in the application of image processing.

2.3.2 Label Fusion

As described in Section 2.2.2, the key challenge associated with the multi-atlas approach is "label fusion" — the strategy by which atlas labels are combined into a single segmentation [12]. To formulate the problem of label fusion, we consider a set of N_A atlases $\{A_n\}$ with labels $\{L_n\}$, where $n = 1, \cdots, N_A$, and I_T as the target image to be segmented. The label alphabet contains $\mathcal{L}$ unique segments:

$$L_n(\mathbf{x}) \in \{1, \ldots, \mathcal{L}\}, \tag{2.15}$$

where $\mathbf{x}$ denotes the location in the label map L_i corresponding to the i-th atlas. The atlases and the target image are assumed to be aligned using the transformations $\{F_n\}$ corresponding to the $\{A_n\}$ atlases. Given these transformations, each input, whether image or label field, can be transformed to the common space that is the target image space. Thus, $\{A'_n\}$ and $\{L'_n\}$ are the atlases and labels in the target image frame such that

$$A'_n(\mathbf{x}) = A_n\big(F_n(\mathbf{x})\big), \tag{2.16}$$

$$L'_n(\mathbf{x}) = L_n\big(F_n(\mathbf{x})\big). \tag{2.17}$$

A final segmentation result L_T associated with I_T is generated by combining all propagated labels $\{L'_n\}$ using a label fusion method.

Majority Voting

The simplest and most widely used label fusion method is majority voting (MV) [13], which asserts an equal contribution for each atlas. Considering each atlas as a classifier providing class labels, no prior information about each classifier's accuracy is taken into account. In this approach, each voxel is assigned with the label that most classifiers select. Thus, the combination result can be expressed as

$$\hat{L}_T(\mathbf{x}) = \underset{l\in\{1,\cdots,\mathcal{L}\}}{\operatorname{argmax}} \sum_{i=1}^{N_A} L_i^l(\mathbf{x}), \tag{2.18}$$

where $L_i^l(\mathbf{x})$ represents the vote for label l produced by the ith atlas as

$$L_i^l(\mathbf{x}) = \begin{cases} 1 & \text{if } L_i(\mathbf{x}) = l, \\ 0 & \text{otherwise.} \end{cases} \tag{2.19}$$

Weighted Voting

As the image intensity is not taken into account during label fusion, a higher accuracy can be achieved by some form of weighting, based on the similarities between the atlases and the target image.This optimization problem can be solved by simply comparing numbers at each voxel: the fused label of each voxel is computed via a local weighted voting strategy.

The local image likelihood terms serve as weights and the label prior values serve as votes. Therefore, at each voxel, training images that are more similar to the test image at the voxel after registration are weighted more:

$$\hat{L}_T(\mathbf{x}) = \underset{l\in\{1,\cdots,\mathcal{L}\}}{\operatorname{argmax}} \sum_{i=1}^{N_A} w_i(\mathbf{x}) L_i^l(\mathbf{x}), \tag{2.20}$$

where $w_i(\mathbf{x})$ is a local weight assigned to the ith atlas and

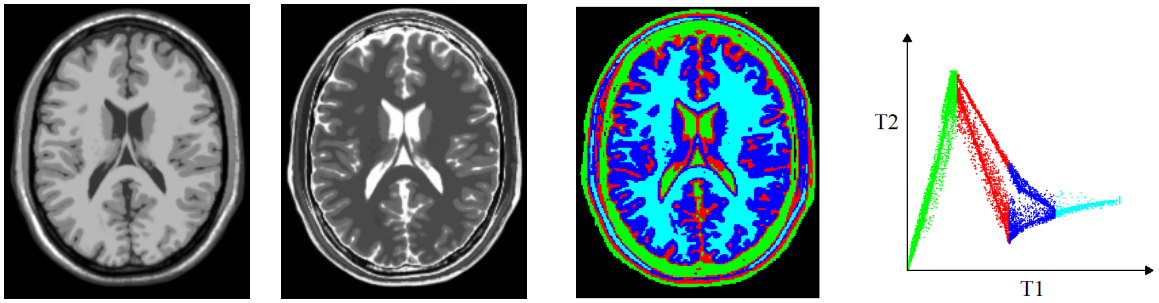
$$\sum_{i=1}^{N_A} w_i(\mathbf{x}) = 1. \tag{2.21}$$

Fixing the weights across all atlases to a constant, $w_i(\mathbf{x}) = C$ ignores the atlas similarities and leads to majority voting. Fixing the weights within a single atlas to a constant, $w_i(\mathbf{x}) = C_i$ globally expresses the similarity between the target and atlas, which models the atlas selection strategy [51, 52].

Global label fusion approaches perform generally better than single atlas-based segmentation. However, as weights are assigned globally, it is impossible for the atlases to have higher contribution in the areas where the registration performs successfully, even if the registration was inaccurate in the rest of the image.

2.4 Problem of Multi-Modality

In medical image analysis, multiple modalities of the same subject or organ provide complementary information that is very important for medical diagnosis and computer-aided surgery [53]. In a multi-atlas-based segmentation problem, of particular interest is dealing with atlases acquired from different sensors, imaging protocols, or modalities [17]. Another scenario could be the cross-modality segmentation of a patient's image with the single-mode atlas database. In either cases, both the registration and label fusion steps would be challenging since image-intensity comparisons may no longer be valid across different modalities [16]. This section reviews the multi-modality challenge and approaches dealing with cases in multi-modal image registration and label fusion.



(a) T1 mode (b) T2 mode (c) labelled anatomy (d) joint histogram

Figure 2.4: Different parts of the images can have different intensity relations in multi-modal images. Perfectly aligned slices in T1 (a) and T2 (b) from simulated BrainWeb [33] database are shown. The brain anatomy in different colors is described in (c). Image (d) is the joint histogram of (a) and (b). Images (c) and (d) show how the brain anatomy relates to the joint histogram by mapping pixel intensities from T1 to T2.

2.4.1 Multi-Modal Image Registration

A key component in every image registration tool is defining a way of measuring the similarity of images to be aligned. As described in Section 2.3.1, for images captured from the same modality, classical similarity measures, such as SSD and cross-correlation coefficient (CC), assume a linear relationship between intensities of the corresponding pixels across the whole image domain. This assumption will not be valid for images obtained from different modalities or imaging sensor types [53]. Since different physical phenomena are measured in different imaging systems, no functional relation between the image intensities can be defined to map the corresponding elements from one image to another. As shown in Fig. 2.4 illustrates how the intensities in two modes of MR brain images are related. Perfectly aligned slices of T1 and T2 modes are shown along with the segmented anatomical parts corresponding to the joint histogram of those images. The joint histogram shows the simultaneous occurrences of intensities between the two images. In Fig. 2.4(c) and (d), the intensity of different tissues are related differently in the two modes.

Traditionally, multi-modal image registration employs mutual information, which uses the statistical dependency of the intensity values between images for evaluating the registration results [20]. Mutual information has been first introduced for rigid alignment of

multi-modal images [18] and later used for deformable registration [45].

In calculating MI, in Eq. 2.13, for measuring image similarity, changing the overlap between two images during the registration process affects the MI value, therefore, normalised mutual information (NMI) has been introduced to cope with this issue [54]. A direct approach to normalisation is presented to evaluate the ratio of the joint and marginal entropies

$$\mathrm{NMI}(I_1, I_2) = \frac{H(I_1) + H(I_2)}{H(I_1, I_2)}. \tag{2.22}$$

A major drawback of mutual information and its variants for image registration is that they do not take spatial information into account. For those cases in which the intensity relations are not spatially invariant or there is a complex intensity relationship, MI-based approaches may suffer from local maxima and an incorrect global maximum problem [55]. Further works have been proposed to overcome this problem by integrating spatial and contextual information in the MI formulation in expense of higher computational time and complexity [56, 57, 58, 59].

Structural information has been also used in the literature of multi-modality problem for improving the robustness of similarity measures to image intensity variations [60, 61, 62, 63, 64]. Thus, the multi-modal registration problem will be transformed to registering two image representations using a simple intensity-based similarity/dissimilarity measure. The registration problem formulated in Eq. 4.1 will be changed into

$$\hat{F} = \underset{F}{\operatorname{argmax}}\, \rho\big(R_f, F(R_m)\big), \tag{2.23}$$

where R_f and R_m are the image representation of the fixed image I_f and moving I_m, respectively. The challenge is still how to find a mapping function that transforms image intensities from different modalities into a new intensity space, so that all images can share similar features in the new space.

2.4.2 Multi-Modal Label Fusion

The multi-atlas approaches are promising compared to single atlas-based segmentation [14]; however, these methods remain problematic in those cases where the atlases and the target

scan are obtained from different sensors or from different acquisition modalities: measuring intensity-based proximity may no longer be valid, since image brightness can have highly differing meanings and circumstances in different modes [16].

Many label fusion methods have been introduced in the medical atlas literature [22]. As described in Section 2.3.2, the simplest and most widely used one is MV [13], which asserts an equal contribution for each atlas. As the image intensity is not taken into account during label fusion, a higher accuracy can be achieved by some form of weighting, based on the similarities between the atlases and the target image. Weighting strategies can be applied in both global and local forms [65, 66], where local weighted voting (LWV) outperforms global strategies when dealing with high contrast anatomical structures [21, 22, 23].

Most label fusion approaches are limited by the assumption that they depend on the consistency of voxel intensities across different scans. In these cases, approaches based on MI do help [67] by assigning weights to atlas labels based on the similarity between the target and the atlases. Thus, the weights in Eq. 7.3 will be defined by

$$w_i(\mathbf{x}) = \mathrm{MI}(A_i', I_T). \tag{2.24}$$

However the inherent non-locality in MI make it problematic for local weighted label fusion. This issue will be highlighted when atlases and target image are acquired with different modalities [16, 21].

2.5 Summary

This chapter provided a review of the background required for brain image segmentation in a multi-atlas-based framework. The brain image segmentation in the context of atlas-based segmentation as a registration-based method, the advantage of using prior knowledge available in atlases, and the issue regarding the atlas-target registration were discussed. The multi-atlas-based segmentation framework, which aims to cope with the basic atlas-target registration problem, was reviewed. As described in this chapter, the key steps in performing the multi-atlas segmentation are the image registration and label fusion. Due to the growth of atlas databases and availability of scans from different modalities, multi-atlas approaches are required to deal deal with multi-modality issue. Multi-modal

registration of brain scans and cross-modal combination of labels from registered atlases are the remaining challenges in multi-atlas problem.

Chapter 3

Problem Formulation

This chapter formulates the problem of multi-atlas-based segmentation and states the motivation, limitations, and the objectives to contribute to the conventional framework. An overview of the problem, the general framework, and its components are given in Section 3.1. Section 3.2 overviews the existing limitations and challenges of the multi-atlas segmentation framework. To address these limitations, the objectives, which are pursued in the following chapters, are introduced in Section 3.3.

3.1 Overview of the Problem

As described in Section 2.3, a general multi-atlas segmentation framework consists of two major components, image registration and label fusion. Fig. 3.1 shows the block diagram of the general multi-atlas-based segmentation framework, in which $\{A_n\}$, $\{L_n\}$, and I_T respectively represent the set of N_A atlases, the labels corresponding to these atlases, and the target image. In the first stage, the atlases are all warped to the target image resulting in the inferred transformations $\{F_n\}$. Given these transformations, each input, whether image or label field, can be transformed to the common reference of the target space. Thus $\{A'_n\}$ and $\{L'_n\}$ are the atlases and labels in the common reference frame. All warped labels are then combined together to form the final segmentation L_T based on information obtained from warped atlases and the target image.

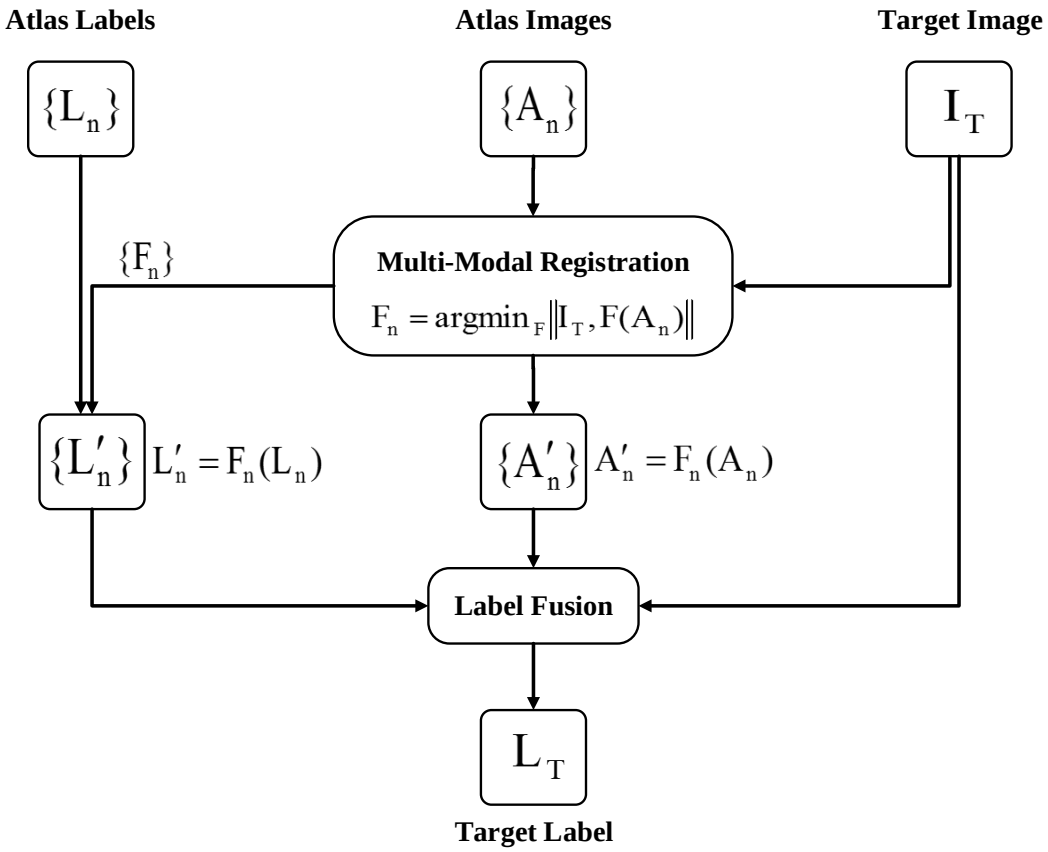


Figure 3.1: Block-diagram of the multi-atlas-based segmentation framework.

In this general framework, the problem is how to perform each of the blocks 'Multi-Modal Registration' and 'Label Fusion' to attain accurate segmentation of the target image. Performing an accurate registration of atlases to the target image and propagating the atlas labels to the target space is crucial for the next step which is the label fusion. The registration is generally defined as an optimisation problem to find the optimal transformation F which maximises the similarity ρ between the moving image I_m and a fixed image I_f:

$$\hat{F} = \underset{F}{\operatorname{argmax}}\, \rho\big(I_f, F(I_m)\big). \tag{3.1}$$

In the context of multi-atlas segmentation problem, I_m and I_f are A_n and I_T. Given the atlases aligned with the target image, accurate segmentation of the target image requires

a method of combining labels from multiple atlases in the database:

$$L_T = \mathcal{F}(L'_n, A'_n), \tag{3.2}$$

where n is the atlas index, $\mathcal{F}$ represents the fusion method, and

$$A'_n = A_n\big(F_n(\mathbf{x})\big), L'_n = L_n\big(F_n(\mathbf{x})\big). \tag{3.3}$$

In the following, the limitations related to the problem of multi-atlas segmentation are reviewed.

3.2 Existing Limitations

As described in Section 2.4, the general multi-atlas segmentation approach is limited to mono-modal cases. From the discussion in Chapter 1, Section 2.3, and Section 2.4, the cross-modality multi-atlas segmentation has brought major challenges regarding the multi-modality problem that can be summarised in the multi-modal image registration and cross-modality label fusion.

The first major challenge in cross-modality multi-atlas segmentation is to register multiple atlases from different modalities. Conventional multi-modal registration methods use the statistical dependency of the intensity values between images for evaluating the alignment accuracy. When the image intensity relations are not spatially invariant or there is a complex intensity relationship, these measures may suffer from local maxima and an incorrect global maximum problem. Performing the registration framework based on employing similarity measures robust to complex intensity relationships requires more complicated procedures, specifically in the optimization step. The amount of computation will increase at least linearly with the number of atlases in the database [11].

Cross-modality label fusion is the second major challenge in the multi-atlas segmentation problem. Existing label combination strategies either use only atlas labels independent of image intensities or rely on the intensity similarity of each atlas to the target volume. While existing label fusion methods can achieve very good segmentation accuracy for images captured from the same modality, extending them for those cases in which the atlases and the target image are in different intensity mappings is challenging: image brightness can have highly differing meanings and circumstances in different modes.

3.3 Objectives

The objectives introduced in Section 1.3 are listed below for reference and the details are presented in Sections 3.3.1, 3.3.2, and 3.3.3.

- Defining a new similarity measure ρ for multi-modal image registration in Eq. 3.1
- Reducing the multi-modal registration problem in Eq. 3.1 to a mono-modal problem
  - Create a structural representation R not relying on the intensity of the images to be aligned (I_m and I_f)
  - Reduce the complexity of the registration problem
- Extending the label fusion problem in Eq. 3.2 to cross modality multi-atlas segmentation
  - Extract structural features not depending on the intensity of atlases $\{A_n\}$
  - Define a measure $\rho_{\mathcal{F}}$ to make a cross-modality comparison

3.3.1 Defining a Similarity Measure for Multi-Modal Image Registration

Section 2.3.1 presents a general framework and components for registering two images, in either the same or different intensity mappings. To deal with complex intensity relationship in multi-modal images, one should define an appropriate similarity measure in 3.1 which is robust to those intensity variations. The objective is to define a similarity measure independent of image intensity based on assessing the image self-similarity $\mathcal{S}$ — the similarity of a pixel to other pixels in an image:

$$\mathcal{S}(I, \mathbf{x}) = f\big(I(x), I(\mathbf{x} + \Delta\mathbf{x})\big), \quad \mathbf{x} + \Delta\mathbf{x} \in \mathcal{N}(\mathbf{x}), \tag{3.4}$$

where f reflects the pairwise similarity between the pixels $\mathbf{x}$ and $\mathbf{x} + \Delta\mathbf{x}$ in an image I, while $\mathcal{N}(\mathbf{x})$ specifies a neighbourhood around $\mathbf{x}$. The similarity measure in Eq. 3.4 can be calculated by comparing the self-similarities in each of the images to be aligned:

$$\rho(I_1, I_2) = \Psi\big(\mathcal{S}(I_1, \mathbf{x}), \mathcal{S}(I_2, \mathbf{x})\big), \forall \mathbf{x}, \tag{3.5}$$

where $\rho(I_1, I_2)$ measures the proximity between two images I_1 and I_2 and Ψ denotes a function to compare two self-similarities. Chapter 4 provides the proposed approach for measuring the similarity based on image self-similarity. The proposed approach will be evaluated in a registration framework in Chapter 6.

3.3.2 Reducing the Multi-Modal Image Registration

For the cases where images are from different modalities, defining the objective function in Eq. 3.1 to measure the image similarity is a challenging part of the problem. Here, the goal is to count on structural features, which are invariant to image intensity in different modalities, instead of intensity relationship. We aim to find a new structural representation, R, of different modalities, which will be a common intensity space for images of different modalities and can reduce the problem of multi-modal registration to a mono-modal one, so that a simple measure can effectively be employed to assess the degree of alignment. Reducing the multi-modal problem will result in using simple L_1 or L_2 distance metrics that are computationally less expensive than statistical or structural similarity measures. For the representation R, the registration problem stated in Eq. 3.1 will be reformulated as

$$\hat{F} = \underset{F}{\operatorname{argmax}}\, \rho\big(R_f, F(R_m)\big), \tag{3.6}$$

where R_f and R_m stand for the representation of images I_f and I_m, respectively. This objective and details about presenting two structural representations are pursued in Chapter 5, Sections 5.2 and 5.3. Structural representation will be employed in a registration framework and the accuracy of alignment is assessed in Chapter 6. The structural representations proposed in Sections 5.2 and 5.3 are presented respectively by Kasiri *et al.* [68] and Kasiri *et al.* [69].

3.3.3 Extending the Problem to Cross Modality Multi-Atlas Segmentation

The problem of label fusion and its conventional solutions are discussed in Section 2.3.2 and is formulated in Eq. 3.2. The goal is to design a label combination method $\mathcal{F}$ to form

a final segmentation result L_T, with the assigned labels on the basis of the similarity of the transformed atlases $\{A'_n\}$ and the target I_T. In the weighted voting equation

$$\hat{L}_T(\mathbf{x}) = \underset{l\in\{1,\cdots,\mathcal{L}\}}{\operatorname{argmax}} \sum_i w_i(\mathbf{x}) L_i^l(\mathbf{x}), \tag{3.7}$$

the labels from each atlas are weighted relying on how the similarity of each atlas' structures to the ones from the target image. The weighting approach can be either global, which makes it an atlas ranking approach, or local. The set of weights $W(\mathbf{x}) = \{w_i(\mathbf{x})\}_{i=1}^{i=N_A}$ for a location $\mathbf{x}$ in the target image can locally be assigned as

$$W(\mathbf{x}) = \Big\{ w_i(\mathbf{x}); w_i(\mathbf{x}) = \rho_{\mathcal{F}}\big(A'_i(\mathbf{x}), I_T(\mathbf{x})\big) \Big\}, \tag{3.8}$$

where $\rho_{\mathcal{F}}(I_1, I_2)$ measures the similarity of two images I_1 and I_2 in the label fusion framework. Details about the label fusion paradigm, how to extract structural features, and measuring the similarity of structures in images are given in Chapter 7 and has been also presented by Kasiri *et al.* [70].

Chapter 4

Similarity Measure

This chapter describes the overall design of the proposed similarity measure for multi-modal image registration. An introduction to the problem of assessing cross-modal similarity in medical images is presented. An overview of the multi-modal similarity measures, specifically related works based on mutual information, is presented to illustrate the challenges and issues that need to be addressed in designing a similarity measure. Following the described methods and issues, a new similarity measure is proposed based on the concept of self-similarity, the proximity of patches within an image, motivated by the assumption that similar structures are more probable to undergo similar intensity transformations[1].

4.1 Introduction

In multi-modal image registration, a challenge is to deal with the large spectrum of intensity variations originating from illumination changes, inhomogeneities, or simply imaging modalities. Since different physical phenomena are measured in different imaging systems, no functional relation between the image intensities can be defined to map the corresponding elements from one image to another. To deal with this issue, one should define an appropriate similarity/dissimilarity measure which is robust to those intensity variations.

Conventional multi-modal approaches tend to assess the accuracy of the alignment by measuring a similarity based on statistical dependency of the intensity values between

[1]Some text and materials in this chapter have been accepted for publication [71, 72].

images. Traditionally, mutual information and its variants such as normalized mutual information (NMI) [18, 19, 20] are used to measure the statistical dependency by assuming a functional or statistical relationship between image intensities [53]. However, these measures do not consider local structures and would be problematic in those cases with complex and spatially dependent intensity relations [55, 73]. Conditioning MI calculation on the spatial information [57, 56, 74], measuring patch similarities [58, 59], estimating local entropies and aligning the structural representations [75] are some examples of taking local contextual information into account for registering multi-modal images.

In this chapter, we propose a self-similarity measure based on estimating the similarity of a point in an image to other points in the same image. A similarity map for the image is made from the pixel similarities measured based on the patch-based estimation of mutual information. The similarities corresponding to each pixel are ranked and the higher ones are considered to describe the pixel of interest. Having a pixel descriptor, independent of pixel values, will allow us to measure the similarity of two images with different intensity mappings.

4.2 Related Research

As described in Chapter 2, the registration of a moving image I_m to a fixed image I_f is formulated as

$$\hat{F} = \underset{F}{\operatorname{argmax}}\, \rho\big(I_f, F(I_m)\big), \tag{4.1}$$

where $I_m, I_f : \Omega \longrightarrow I$, ρ stands for the similarity measure to assess the degree of alignment, and F represents the spatial transformation. Dissimilarity measures such as sum of squared differences (SSD) take their minimum when the images are aligned, therefore, the negative of dissimilarity measure is used as the similarity in the Eq. 4.1. In the following, an overview of measuring cross-modal similarity is described.

4.2.1 Mutual Information

As described in Section 2.4.1, mutual information is the traditional measure to evaluate the similarity of images obtained from different imaging sensors by measuring the statistical

dependency of images to be aligned. Mutual information for two images I_1 and I_2 is defined based on the Shannon entropy as

$$\mathrm{MI}(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2) \tag{4.2}$$

where $H(I_1)$ and $H(I_2)$ represent the entropy of random variables I_1 and I_2, and $H(I_1, I_2)$ stands for the joint entropy of these two random variables.

A major drawback of mutual information and its variants for image registration is that they do not take spatial information into account. This drawback can degrade the quality of registration when there is an intensity distortion such as a non-stationary bias field in an MR image [76].

4.2.2 Local Mutual Information

To overcome the problem related to non-locality of MI, one approach is to take spatial information into account and integrate it in the joint and marginal histogram computation. One approach is to use spatial kernels as box filters to implement the localised mutual information (LMI) [56]. In LMI, the average of MI computed over multiple local neighbourhoods is returned as the similarity measure:

$$\mathrm{LMI}(I_m, I_f; \Omega) = \frac{1}{N_b} \sum_{i=1}^{N_b} \mathrm{MI}(I_m, I_f; \mathcal{N}(\mathbf{x}_i)). \tag{4.3}$$

where $\mathcal{N}(\mathbf{x}_i) \subset \Omega$ is the spatial neighbourhood for pixel i and N_b stands for the number of neighbourhoods.

4.2.3 Conditioned Mutual Information

To deal with the sensitivity of MI to intensity non-uniformities, Studholme *et al.* [73] introduced a third channel to the joint histogram containing the regional label. Conditioning MI upon pixel locations was integrated into the MI formulation known as conditional mutual information (cMI) [57]. In this method, one dimension is added to both marginal and joint histograms representing the location of intensity pairs:

$$\mathrm{cMI}(I_m, I_f|\mathbf{x}) = H(I_m|\mathbf{x}) + H(I_f|\mathbf{x}) - H(I_m, I_f|\mathbf{x}) \tag{4.4}$$

cMI was shown to be effective in lowering the negative effect of bias fields and yields a higher registration accuracy. The drawbacks of this approach is still the difficulty of populating the 3D histogram to compute the similarity measure.

4.2.4 Self-Similarity Measures

The principle of self-similarity, which has first been proposed as non-local means for image denoising [77], is based on looking at similar image patches across an image. To obtain a denoised pixel, a weighted average of intensities from all other pixels in the image is computed. The distance between the patch surrounding the pixel of interest and all other patches are used as the weight in averaging. In medical image registration, self-similarity is used to measure the similarity of multi-modal images based on the assumption that internal pixel-to-pixel relationships are similar in different modalities.

Modality Independent Neighbourhood Descriptor

Self-similarity for the purpose of registration has been first used in the non-local shape descriptor [78]. Later, Heinrich *et al.* [79] proposed the modality independent neighbourhood descriptor (MIND) based on the idea of non-local means filtering. In this method, the similarity of every image patch to its neighbours is measured by taking a sum of squared distances (SSD) followed by an exponential function to transform SSD distances to a set of multi-dimensional normalised weights that are the descriptor elements. MIND is robust to the non-functional intensity relations, noise, and bias fields. Mathematically, MIND is defined by measuring the Euclidean patch distance D_p between the locations $\mathbf{x}$ and $\mathbf{x}+\Delta\mathbf{x}$ and a variance estimate V which is the mean of the patch distances within a neighbourhood:

$$\mathrm{MIND}(I,\mathbf{x},\Delta\mathbf{x}) = \frac{1}{Z_n}\exp\left(-\frac{D_p(I,\mathbf{x},\mathbf{x}+\Delta\mathbf{x})}{V(I,\mathbf{x})}\right), \tag{4.5}$$

where $\Delta\mathbf{x}$ is restricted to a spatial search region and Z_n is a normalisation constant. The resulting descriptor has the dimension of the patch size. The similarity measure is then defined by averaging the SSD of MIND descriptors over different $\Delta\mathbf{x}$. So large neighbourhoods as the spatial search region will lead to further computational burden in performing the registration.

Contextual Conditioned Mutual Information

The self-similarity α-MI (SeSaMI) proposed by Rivaz *et al.* [59] uses local structural information in a graph-based implementation of mutual information for non-rigid image registration. Using the α-entropy, a generalization of Shannon entropy, α-MI is calculated on multiple features of intensities and their gradients. The SeSaMI is a rotation invariant measure which is also robust to bias fields.

In another work proposed by Rivaz *et al.* [58], the contextual conditioned mutual information (CoCoMI) is proposed based on conditioning the estimation of MI on similar structures. The idea behind this method is based on the limitation in calculating MI, which is considering only the intensity values of corresponding pixels and not of neighbourhoods and therefore, losing contextual information. CoCoMI is formulated as

$$\mathrm{CoCoMI}(I_m, I_f; \Omega) = \frac{1}{N} \sum_{j=1}^{N} \mathrm{MI}(I_m, I_f; \mathcal{M}_j) \tag{4.6}$$

where $\mathcal{M}_j$ is the similarity map corresponding to pixel j. The similarity map of a pixel is defined as the set of pixels whose small neighbouring patches are similar to the one surrounding the pixel of interest. So for every pixel j, the similarity map $\mathcal{M}_j$ is obtained containing the pixels with the smallest dissimilarity to the pixel j. The MI-based similarity is computed based upon the pixels in the similarity map for each of the N pixels and the average result is returned as the similarity measure.

4.3 Sorted Self-Similarity

In this section, a self-similarity measure for multi-modal registration is proposed based on creating a descriptor independent of intensity mapping. A self-similarity map is constructed for each pixel of an image and unlike the similarity measure based on MIND descriptor, the patch relationship is defined based on sorted intensity values in the patch. The pixels with higher similarities with the pixel of interest are marked to transmit the significant information about that pixel. Therefore, all the pixel relationships will no longer be taken into account and, as a result, the amount of computation will be significantly reduced.

4.3.1 Motivation

As mentioned in Section 4.2.4, the motivation behind the self-similarity comes from the non-local means (NLM) method for image denoising. The NLM approach seeks similar patches across a noisy image to reduce the pixel noise in the image. The noise-free pixel is estimated as a weighted average of all other pixels in the image where the weights are based on calculating the Euclidean distance between the patch surrounding the pixel of interest, and all other patches in the image. As the distance between patches increases, the weight decreases. In general form, the denoised pixel $NL(i, I)$ in an image I is calculated as

$$NL(i, I) = \sum_{j \in \Omega} w(i, j) I(j), \tag{4.7}$$

where $w(i, j)$ is based on the normalised Euclidean distance between the patches surrounding pixels i and j. To simplify this approach, similar patches within a smaller non-local region are only considered, therefore in Eq. 4.7, $j \in \Omega$ will change to $j \in \mathcal{N}(i)$, where $\mathcal{N}(i)$ is neighbourhood of i [80].

4.3.2 Patch Similarity

Similar to the non-local means in Eq. 4.7, the self-similarity of an image is calculated by measuring the pairwise similarity/dissimilarity between patches surrounding the pixels of interest, where the pairwise similarity/dissimilarity can be interpreted as the weights $w(i, j)$ between pixels i and j. The straightforward choice of a distance measure $D_p(\mathbf{x}_1, \mathbf{x}_2)$ between two pixels $\mathbf{x}_1$ and $\mathbf{x}_2$ is the SSD of all pixels between the two patches $P_{\mathbf{x}_1}$ and $P_{\mathbf{x}_2}$ centred at pixels $\mathbf{x}_1$ and $\mathbf{x}_2$,

$$D_p(I, \mathbf{x}_1, \mathbf{x}_2) = \sum_{\Delta\mathbf{x} \in \mathcal{N}_p} \big(I(\mathbf{x}_1 + \Delta\mathbf{x}) - I(\mathbf{x}_2 + \Delta\mathbf{x})\big)^2, \tag{4.8}$$

where $\mathcal{N}_p \subset \Omega$ is the neighbourhood of central pixels in the patches $P_{\mathbf{x}_1}$ and $P_{\mathbf{x}_2}$.

The issue with using the simple SSD for measuring the patch dissimilarity is that it is not rotation-invariant, which might be a restriction for those cases where strong rotations exist. To cope with the rotational deformations, one can use measures that are invariant to rotation. One approach is to calculate the statistical dependency between patches as a

measure of patch proximity. Mutual information can be employed to measure the similarity between patches $P_{\mathbf{x}_1}$ and $P_{\mathbf{x}_2}$ as

$$\mathrm{MI}(P_{\mathbf{x}_1}, P_{\mathbf{x}_2}) = \mathrm{H}(P_{\mathbf{x}_1}) + \mathrm{H}(P_{\mathbf{x}_2}) - \mathrm{H}(P_{\mathbf{x}_1}, P_{\mathbf{x}_2}), \tag{4.9}$$

where $\mathrm{H}(P_{\mathbf{x}_1})$ and $\mathrm{H}(P_{\mathbf{x}_2})$ denote the entropy of intensities in $P_{\mathbf{x}_1}$ and $P_{\mathbf{x}_2}$, and $\mathrm{H}(P_{\mathbf{x}_1}, P_{\mathbf{x}_2})$ is the joint entropy of these two patches. Although MI provides a good measure of similarity of signals, it forced further loads to computations of the procedure compared to calculating distance-based dissimilarities. The marginal and joint histogram of patches have to be estimated for a large number of pixel comparisons. To reduce the computations of the MI calculation, we propose to use an intensity based patch-comparison which is computationally efficient and yields a rotation invariant measure. The patch comparison is based on the idea of sorted random projection designed for texture classification [81]. Sorting ignores the ordering of elements in the patch $P_{\mathbf{x}}$ and clearly yields a rotation invariant output $\tilde{P}_{\mathbf{x}}$:

$$\tilde{P}_{\mathbf{x}} = \mathsf{sort}(P_{\mathbf{x}}). \tag{4.10}$$

The dissimilarity between two patches P_{x_1} and P_{x_2} can be obtained by measuring the Euclidean distance between $\tilde{P}_{\mathbf{x}_1}$ and $\tilde{P}_{x_2}$ according to Eq. 4.8:

$$\tilde{D}_p(I, \mathbf{x}_1, \mathbf{x}_2) = \sum_{\Delta\mathbf{x}} \big(\tilde{P}_{\mathbf{x}_1}(\Delta\mathbf{x}) - \tilde{P}_{\mathbf{x}_2}(\Delta\mathbf{x})\big)^2. \tag{4.11}$$

Given the patch dissimilarity measurement, we are able to form a descriptor for each pixel $\mathbf{x}$ defined based on the pixel dissimilarity to all other pixels $\mathbf{x}_i$ in the r-distance neighbourhood of $\mathbf{x}$ in the image. Therefore, the descriptor $\mathcal{D}$ at pixel $\mathbf{x}$ is constructed based on the patch distance measured in Eq. 4.11 such that

$$\mathcal{D}(\mathbf{x}, i) = \tilde{D}_p(I, \mathbf{x}, \mathbf{x}_i), \quad \mathbf{x}_i \in \mathcal{N}_r(\mathbf{x}), \tag{4.12}$$

where $\mathcal{N}_r(\mathbf{x})$ represents the r-distance neighbourhood of pixel $\mathbf{x}$. Fig. 4.1 shows the self-similarity measurement for a pixel in the three MR modes: T1, T2, and PD. The neighbourhood is shown by a red box which specifies the spatial search region of the central pixel. Patches with size 11×11 are used to compute the patch dissimilarities. This figure illustrates three different intensity mappings in which a pixel will have similar intensity-relationship with its surrounding pixels using the proposed self-similarity measure.

4.3.3 Patch Selection

At this step, the objective is to find similar structures in the image by choosing the most similar pixels to the pixel of interest. Therefore, the M pixels in the neighbourhood $\mathcal{N}_r(\mathbf{x})$ with the lowest dissimilarity to the pixel of interest $\mathbf{x}$ are identified and selected to carry the most significant information about self-similarity:

$$\mathcal{D}_{\mathsf{sort}}(\mathbf{x}) = \underset{i}{\mathsf{sort}}\left(\mathcal{D}(\mathbf{x}, i)\right), \tag{4.13}$$

$$\mathcal{S}(I, \mathbf{x}) = \chi\left(\mathcal{D}_{\mathsf{sort}}(\mathbf{x}), M\right), \tag{4.14}$$

where χ picks the first M elements in $\mathcal{D}_{\mathsf{sort}}(\mathbf{x})$ and returns the indices of those pixels in the self-similarity map $\mathcal{S}(I, \mathbf{x})$. By applying an ascending sort operation to the representation $\mathcal{D}$ at pixel $\mathbf{x}$ and picking the first M elements, we try to only consider the M most similar patches to $P_{\mathbf{x}}$ and reduce the number of pixels required to describe the pixel $\mathbf{x}$ and carry self-similarity information.

To determine M corresponding to the pixel $\mathbf{x}$, we look at the average dissimilarity of that pixel to all other pixels in the spatial search region $\mathcal{N}_r(\mathbf{x})$. The dissimilarity values less than this average value are considered to represent the most significant ones. For pixel of interest $\mathbf{x}_i$, the number of most significant dissimilarities $M(\mathbf{x}_i)$ are obtained as

$$M(\mathbf{x}_i) = \left|\{\mathcal{D}_{\mathsf{sort}}(\mathbf{x}_i, k); \mathcal{D}_{\mathsf{sort}}(\mathbf{x}_i, k) < \bar{\mathcal{D}}_{\mathsf{sort}}(\mathbf{x}_i)\}_{k=1}^{k=N}\right|, \tag{4.15}$$

where $\bar{\mathcal{D}}_{\mathsf{sort}}(\mathbf{x}_i)$ is the average of the elements in $\mathcal{D}_{\mathsf{sort}}(\mathbf{x}_i)$, $|\cdot|$ reflects the cardinality of a set, and N denotes the number of pixels contributing to the similarity measure. To have a unified M for all of the N pixels, the average of $M(\mathbf{x}_i)$ over i is used to set the number of most significant patches:

$$\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M(\mathbf{x}_i). \tag{4.16}$$

By choosing the M most significant elements of R, we will be able to extend the search region as far as the registration performance allows.

Algorithm 1 Outline of the proposed self-similarity approach.

(1) Select N random samples over the image to calculate the overall similarity measure.
(2) Obtain patch similarity $\tilde{D}_p$ in a neighbourhood $\mathcal{N}_r$ (Eq. 4.11).
(3) Construct a representation $\mathcal{S}$ for each of the N pixels by choosing the most significant patch similarities (Eq. 4.12–Eq. 4.14).
(4) Compare pixel self-similarities in I_m and I_f to form a similarity matrix SM (Eq. 4.18).
(5) Average the similarity matrix SM to form the scalar similarity measure (Eq. 4.19).

4.3.4 Multi-Modal Similarity Measure

At this stage, it is required to compare the self-similarity maps obtained from the moving image I_m and the fixed image I_f using a function Ψ and find the similarity measure ρ:

$$\rho(I_m, I_f) = \Psi\big(\mathcal{S}(I_m, \mathbf{x}), \mathcal{S}(I_f, \mathbf{x})\big), \forall \mathbf{x}. \tag{4.17}$$

As described in Section 4.3.3, the self-similarity of each pixel can be obtained using the set of equations from 4.10 to 4.14. The self-similarity, which can be considered as a pixel descriptor, is obtained for pixel $\mathbf{x}$ in each of the moving and fixed images and the result is compared by employing mutual information:

$$\mathrm{SM}(I_m, I_f; \mathbf{x}) = \mathrm{MI}\big(\mathcal{S}(I_m, \mathbf{x}), \mathcal{S}(I_f, \mathbf{x})\big). \tag{4.18}$$

The self-similarity is measured for N randomly selected pixels in each of the moving and fixed images. As the number of pixels increases, a better estimation of image similarities will be attained. To attain a scalar as the similarity measure required for the optimisation in Eq. 4.1, SM is averaged over all N pixels as

$$\rho(I_m, I_f; \Omega) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{SM}(I_m, I_f; \mathbf{x}_i). \tag{4.19}$$

The overall step-by-step algorithm for obtaining the similarity measure is summarised in Algorithm 1.

4.4 Summary

In this chapter, we have focused on the similarity measure for multi-modal image registration. A review of the classical multi-modal similarity measures along with the challenges

regarding the non-locality was presented. An overview of using the self-similarity in recent literature was presented to address the issues related the classical approaches. In this line of research, we have presented a similarity measure based on assessing the self-similarity of images to be aligned. The self-similarity is measured in a patch-based paradigm where each pixel in the image was described by the pixel similarity to the most similar pixels in a neighbourhood. By employing the sorting operation the ordering of patch pixels were ignored and thus the a rotation invariant descriptor was obtained. Unlike the common multi-modal registration techniques, such as mutual-information, that utilise statistical dependency, the new measure is able to take the internal structural relationship into account.

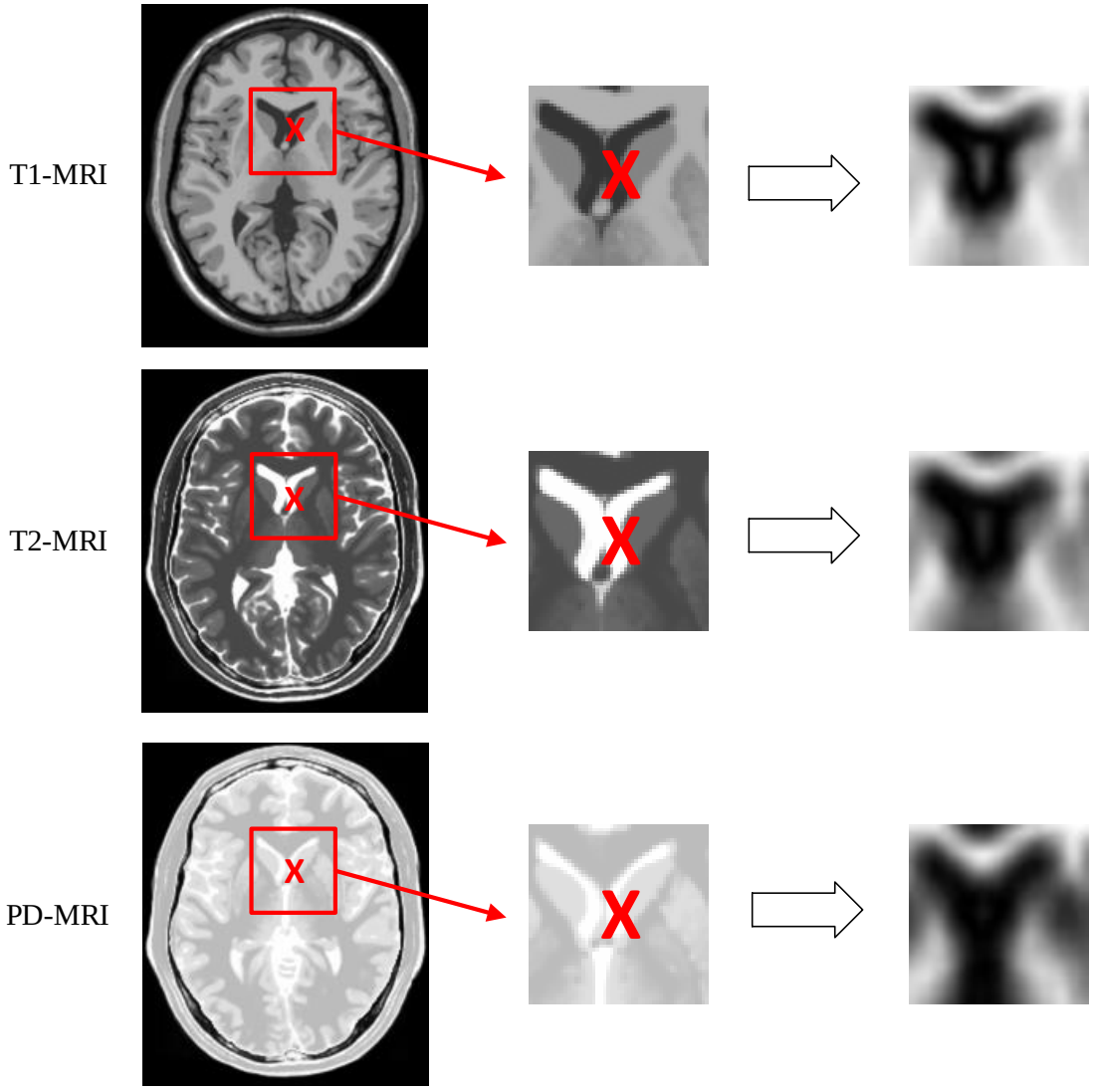


Figure 4.1: Self-similarity in different modes of MR images: The dissimilarity of a pixel $\mathbf{x}$ and its neighbouring pixels $\mathbf{x}_i \in \mathcal{N}_r(\mathbf{x})$ is measured to provide the pixel descriptor $\mathcal{D}(\mathbf{x}, i)$. The pixel $\mathbf{x}$ and its neighbourhood $\mathcal{N}_r(\mathbf{x})$ are specified by a red X in a red box. In the resulting pixel descriptor, darker areas show more similar pixels to the pixel of interest.

Chapter 5

Structural Representation

This chapter describes in detail the overall design of structural image representation to evaluate the similarity of multi-modal images. The concept of modality independent representation based on structural information is explained in Section 5.1. In Section 5.2, an overview of the image representation based on complex phase and amplitude using complex wavelet transform is presented. An image representation based on a combination of complex wavelet representation and gradient information is proposed for the application of multi-modal image registration. Independent of the complex wavelet representation, Section 5.3 presents the entropy-based structural representation, and the issues regarding the image entropy. A new approach is proposed based on a modification of entropy image representation to better represent the structures in the image. The main contributions in this chapter are: 1) the introduction of a new structural representation based on a combination of complex wavelet and gradient information to improve the representation of structural characteristics as described in Section 5.2.3, and 2) the modification of structural representation based on image entropy to improve the response sensitivity to local structures, as described in Section 5.3.3[1].

[1]Some text and materials in this chapter have been previously published [68, 69].

5.1 Modality Independent Image Representation

Structural information has been used in the literature of multi-modal registration problem for improving the robustness of similarity measures to image intensity variations [60, 61, 62, 82, 83]. The structural information are the image characteristics, such as edges and corners, that are intensity-independent and similar at different modalities of the same scene.

The combination of edge orientation information and intensity information in an entropy-based objective function was utilised for registering images captured from different sensors, such as visible and infra-red (IR) images [61]. De Nigris *et al.* [82] proposed a registration method based on the alignment of gradient orientations with minimal uncertainty. Later, a multi-resolution approach was proposed based on employing the dual-tree complex wavelet transform (DT-CWT) to align IR and visible images [60]. In this approach, accurate estimation of registration in finer levels is obtained using edge information in coarser levels. Cross-correlation and mutual information are used to measure the similarity in the coarser and finer levels, respectively. Complex phase order has been used as a similarity measure in registering MR with CT images in [62]. Feature-level information fusion method based on Gabor wavelets transformation and independent component analysis (ICA) has been used in inter-subject multi-channel registration by Li, *et al.* [83] to combine the complementary information that characterize tissue types in different modalities.

Registration methods based on the scale-space representations try to analyse an image at various resolutions [84, 85, 86]. Texture features obtained from different scales of resolution can reveal similar structural attributes between the images to be aligned. Scale-based registration for studying multiple sclerosis in MR images was presented based on the local scale value assigned to each voxel [84]. This scale value for a voxel of interest was defined locally as the radius of the largest ball centred at that voxel with homogeneous intensities. In another work by Saha [85], a local morphometric parameter called tensor scale was presented to attain a unified representation of size, orientation, and anisotropy. A multi-scale representation for multi-modal registration has been proposed by Li, *et al.* [86] that works on the basis of applying the ICA at textures extracted from each length scale, spectrally embedding the ICA components, and identifying and combining the optimal length scales using MI to perform the registration.

Structural information is utilized to transform images from different modalities to a common mode and therefore transform the multi-modal problem to a mono-modal registration. Therefore, the multi-modal registration problem will be

$$\hat{F} = \underset{F}{\operatorname{argmax}}\, \rho\big(R_f, F(R_m)\big), \tag{5.1}$$

where R_f and R_m are respectively the image representation for I_f and I_m. Reducing the multi-modal problem to a mono-modal one results in using simple L_1 or L_2 distance metrics that are computationally less expensive than statistical or structural similarity measures. Usage of gradient intensity, ridge, and estimation of cross correlating gradient directions are examples of creating a structural representation of input images for registration [64]. Structural representation based on entropy images followed by measuring SSD has been proposed [63].

For images being represented with the same intensity values, sum of absolute differences (SAD) or SSD can be good choices for the distance measure. Registration of images with complex intensity relationships requires more complicated similarity/dissimilarity measures. Correlation coefficient, correlation ratio (CR), and mutual information are widely used in this case [53]. The objective is to find structural representations of multi-modal images, R, that are invariant to the image intensity. Therefore, simple measures based on intensity difference can be used to assess the image similarity.

5.2 Complex Wavelet Representation

Traditional wavelets became very conventional tools in image processing, however, they are shift variant transforms and suffer from a poor resolution in orientation [87]. Alternative multi-resolution transforms with better orientation representations have been proposed that fix the shift invariance problem by being over-complete [87, 88, 89]. Among them, Gabor transform as a band-pass multi-resolution transform provides localised frequency and orientation representation and is widely used for image feature extraction and texture analysis. Complex-valued Gabor filters have gained considerable attention in texture representation and discrimination since they can well approximate characteristics of receptive fields in human visual system [90, 91].

Gabor texture features have been used successfully for registering both mono-modal and multi-modal images as they are capable of extracting information across different scales and orientations. Gabor filters are capable of capturing local edge and texture information and create local frequency representations from images [92]. Ou *et al.* employed Gabor filters in deformable image registration, in which the filter responses were used to build the pixel descriptor [93]. Gabor filter responses have been also used to transform images of different modalities to a common space [92, 94]. These image representations in a common space are robust to contrast variations and edge magnitude.

In the following, details about the complex wavelet representation, its characteristics and limitations, along with the proposed image representation are introduced.

5.2.1 Complex Amplitude and Phase

The general complex representation of an image I based on an over-complete wavelet at scale s and orientation θ can be formulated as

$$\Upsilon_{s,\theta}(\mathbf{x}) = \alpha_{s,\theta}(\mathbf{x}) \exp\big(j\phi_{s,\theta}(\mathbf{x})\big), \tag{5.2}$$

where $\alpha_{s,\theta}(\mathbf{x})$ and $\phi_{s,\theta}(\mathbf{x})$ are the amplitude and phase of the complex wavelet coefficients at location $\mathbf{x}$.

One of the most popular complex wavelet transforms is the Gabor complex wavelet which has been used widely for extracting features from images [87, 90, 95]. The impulse response of a Gabor filter can be viewed as a sinusoidal wave plane modulated by a Gaussian envelope. For a pixel coordinate $\mathbf{x} = [x\ y]^T$ and particular frequency $\boldsymbol{\omega_0} = [\omega_{x_0}\ \omega_{y_0}]$, the impulse response of a Gabor filter $\gamma(x, y)$ is given by

$$\gamma(x, y) = f_s(x, y) f_g(x, y), \tag{5.3}$$

where $f_s(x, y)$ is a complex sinusoid known as a carrier and $f_g(x, y)$ is a 2D Gaussian function as

$$f_s(x, y) = \exp\big(-2\pi j(\omega_{x_0} x + \omega_{y_0} y)\big), \tag{5.4}$$

$$f_g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right], \tag{5.5}$$

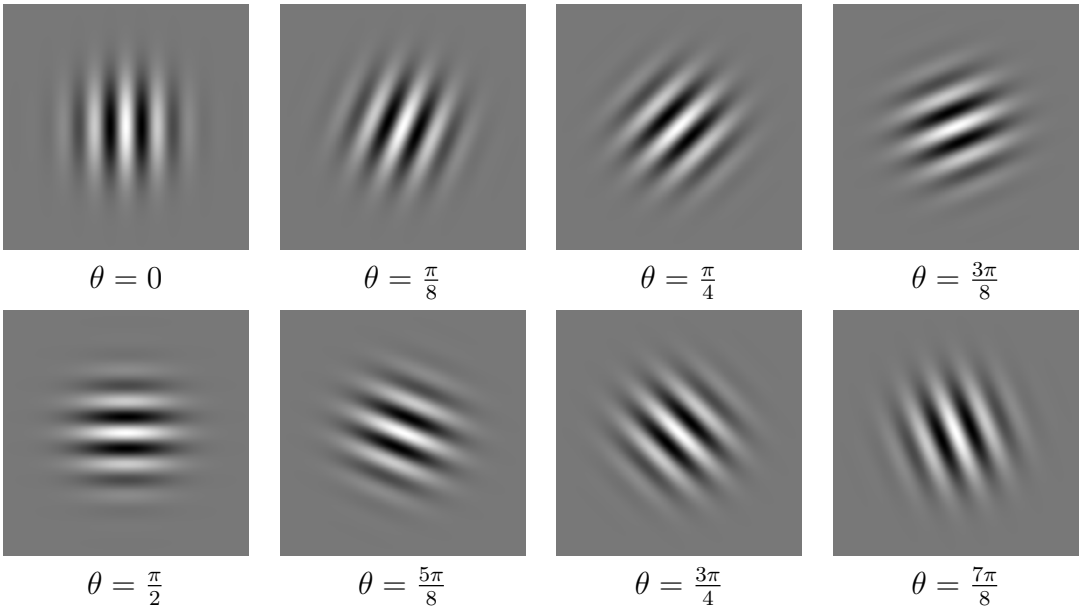


Figure 5.1: 2D Gabor complex wavelets in spatial domain with different orientations: the even symmetric component of the Gabor filters are shown when $\theta \in [0, \pi]$.

where (σ_x, σ_y) specifies the spread of the Gaussian envelope.

The orientation of complex Gabor filter is determined by the center frequencies ω_{x_0} and ω_{y_0}. Fig. 5.1 illustrates eight different orientations of a Gabor filter in spatial domain. In this figure, the even symmetric component of the Gabor filters are shown when the orientation θ varies in the range $[0, \pi]$.

One of the benefits of the complex Gabor filter is that it can reach the optimal compromise between the localisation in the spatial and frequency domains, meaning that any arbitrary bandwidth used to construct the filter can be optimised with minimal spatial extent. However, Gabor filters are restricted to a non-zero mean for bandwidths over one octave and the response of the filter will depend upon the mean value of the signal [90]. For this reason, Gabor complex wavelets are limited to bandwidths below one octave and as a consequence lead to an inefficient representation of a signal with broad spectral information. To address this problem while maintaining the optimal spatial-frequency resolution, one effecting approach is to use Log-Gabor complex wavelet transform [90]. The Log-Gabor

transform in the frequency domain under the polar coordinate can be expressed as

$$\Gamma(\omega,\theta) = \exp\Bigg[-\frac{\big(\log(\omega/\omega_0)\big)^2}{2\big(\log(\sigma_\omega/\omega_0)\big)^2}\Bigg]\exp\Bigg[-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\Bigg], \tag{5.6}$$

where (ω,θ) show the polar coordinates, the (ω_0,θ_0) are the coordinates of the center of the filter, and $(\sigma_\omega,\sigma_\theta)$ determine the bandwidths in f and θ. It can be seen that the DC component of the Log-Gabor filter approaches zero value.

The amplitude $\alpha_{s,\theta}(\mathbf{x})$ and phase $\phi_{s,\theta}(\mathbf{x})$ in Eq. 5.2 for the Log-Gabor complex wavelet $\gamma_{s,\theta}(\mathbf{x})$ are specified using the odd-symmetric $\gamma^o_{s,\theta}(\mathbf{x})$ and even-symmetric $\gamma^e_{s,\theta}(\mathbf{x})$ pairs at scale s and orientation θ:

$$\alpha_{s,\theta}(\mathbf{x}) = \sqrt{\big(I(\mathbf{x}) * \gamma^e_{s,\theta}(\mathbf{x})\big)^2 + \big(I(\mathbf{x}) * \gamma^o_{s,\theta}(\mathbf{x})\big)^2}, \tag{5.7}$$

$$\phi_{s,\theta}(\mathbf{x}) = \tan^{-1}\Big(\frac{I(\mathbf{x}) * \gamma^e_{s,\theta}(\mathbf{x})}{I(\mathbf{x}) * \gamma^o_{s,\theta}(\mathbf{x})}\Big), \tag{5.8}$$

where $*$ denotes the convolution operator.

5.2.2 Phase Congruency

One of the first complex wavelet representations of images was designed by Kovesi based on the congruency of Fourier components rather than the intensity gradient in edges [96, 97]. Based on this phase congruency (PC), the feature is perceived at any angle where the Fourier components are maximally in phase. Fig. 5.2 presents a clear edge in a square wave and its Fourier components which are all in phase. Physiological and psychological evidences also confirm that the phase congruency is able to provide a simple model to imitate the human visual system for detecting and identifying edge and corner features in an image [98].

Based on the definition by Kovesi [96], the phase congruency of an image is computed using an over-complete Log-Gabor complex wavelet transform as

$$PC_1(\mathbf{x}) = \max_{\theta\in[0,2\pi]} \frac{\sum_s \alpha_s \cos(\phi_s(\mathbf{x}) - \theta)}{\sum_s \alpha_s + \epsilon}, \tag{5.9}$$

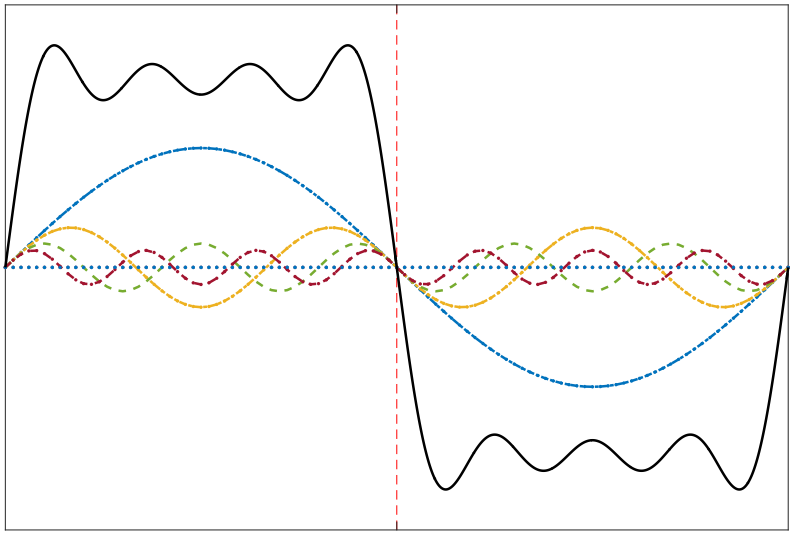


Figure 5.2: Fourier components of a step in a square wave: Fourier components and the approximated signal based on the first five terms of the Fourier series are presented respectively by the dashed color lines and a solid black line. The phase congruency of all components can be seen at the edge specified by the vertical red dashed line.

where s is the wavelet scale and ϵ is a small constant used to avoid division by zero. The value θ that maximises Eq. 5.9 is the amplitude weighted mean phase across all scales ($\theta = \bar{\phi}(\mathbf{x})$). As an alternative to this formulation, maximum phase congruency can be found by looking at the peaks in the local energy function [99]. The local energy function $E(\mathbf{x})$ at location $\mathbf{x}$ is defined as

$$E(\mathbf{x}) = \sqrt{M_o^2(\mathbf{x}) + M_e^2(\mathbf{x})}, \tag{5.10}$$

where

$$M_e(\mathbf{x}) = \sum_s I(\mathbf{x}) * \gamma_s^e(\mathbf{x}), \tag{5.11}$$

and $M_o(\mathbf{x})$ is computed as

$$M_o(\mathbf{x}) = \sum_s I(\mathbf{x}) * \gamma_s^o(\mathbf{x}). \tag{5.12}$$

Therefore, the phase congruency will be

$$PC_2(\mathbf{x}) = \frac{E(\mathbf{x})}{\epsilon + \sum_s \alpha_s(\mathbf{x})}. \tag{5.13}$$

The ratio in Eq. 5.13 equals one if all the Fourier components are in phase and takes its minimum of zero when there is no phase coherence.

To increase the robustness of the representation to the low level image noise and improve the localisation of structural information, a modified formulation for phase congruency was proposed by Kovesi [97]:

$$PC_3(\mathbf{x}) = \frac{\sum_s W^{PC}(\mathbf{x}) \lfloor \alpha_s(\mathbf{x}) \cos\big(\Delta\phi_s(\mathbf{x})\big) - |\sin\big(\Delta\phi_s(\mathbf{x})\big)| - T_r) \rfloor}{\sum_s \alpha_s(\mathbf{x}) + \epsilon} \tag{5.14}$$

$$\Delta\phi_s(\mathbf{x}) = \phi_s(\mathbf{x}) - \bar{\phi}(\mathbf{x}). \tag{5.15}$$

In Eq. 5.14, $\Delta\phi_s(\mathbf{x})$ is the phase deviations from the mean at scale s, the threshold T_r is to eliminate the energy values that are estimated as the noise influence, and $\lfloor \cdot \rfloor$ denotes a truncation operator that sets all the enclosed negative quantities to zero. $W^{PC}(\mathbf{x})$ is a weighting function that is constructed to decay the filter response where its spread is narrow.

To combine data from several orientations, one should note that each orientation should contribute to the final representation in proportion to the energy of that orientation and the normalization will be based on the total energy over all orientations and scales. This produces the following equation for the phase congruency based on a filter applied on scales s and orientation θ:

$$PC_3(\mathbf{x}) = \frac{\sum_\theta \sum_s W_\theta^{PC}(\mathbf{x}) \lfloor \alpha_{s,\theta}(\mathbf{x}) \cos\big(\Delta\phi_{s,\theta}(\mathbf{x})\big) - |\sin\big(\Delta\phi_{s,\theta}(\mathbf{x})\big)| - T_r) \rfloor}{\sum_\theta \sum_s \alpha_{s,\theta}(\mathbf{x}) + \epsilon}. \tag{5.16}$$

5.2.3 Representation Based on Complex Wavelets

An important issue in the design of the complex phase representation is related to dealing with images with poor structural contrast. Images captured from some certain imaging modalities, such as PD mode in MR imaging, do not provide enough sharpness where the structures exist. The poor contrast may cause difficulties in extracting and distinguishing fine structural details that can be an important issue in measuring the detailed structural dissimilarity between two images in an alignment procedure. Fig. 5.3 illustrates how the complex wavelet representation Eq. 5.16 behave in different conditions of imaging modalities. Three modes of MR imaging, T1, T2, and PD modes from the RIRE database [100]

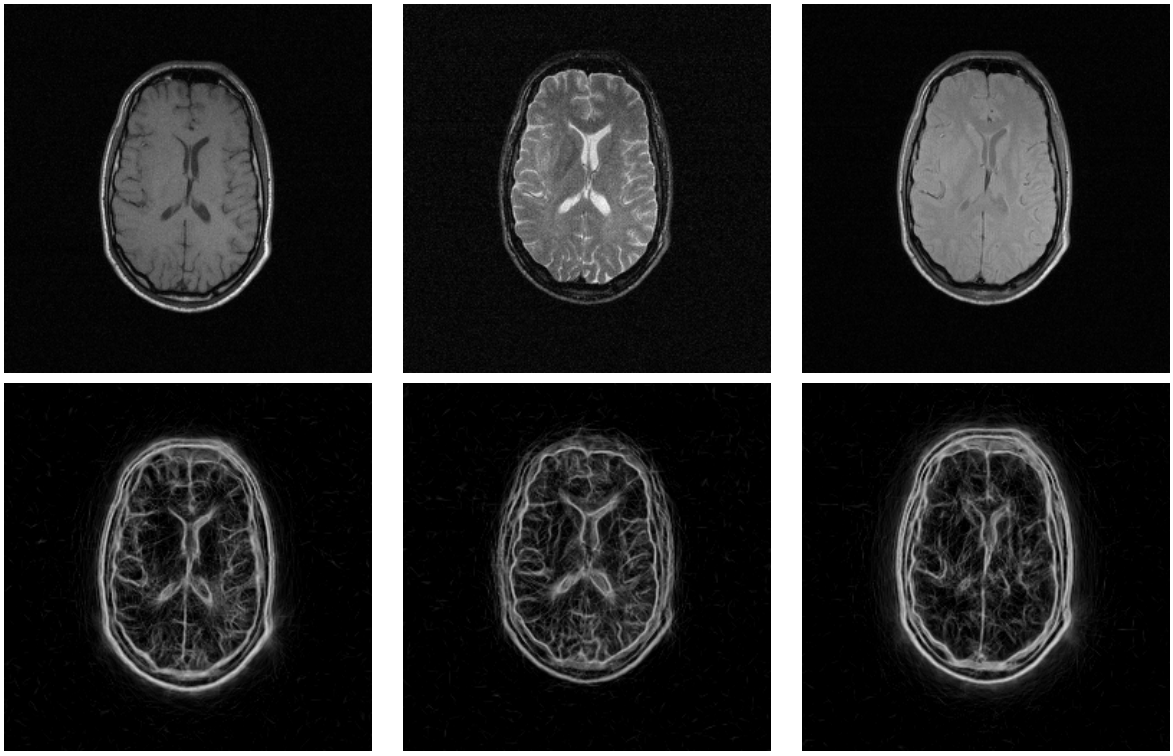


Figure 5.3: Complex wavelet representation for images with different structural contrast: The top row shows the original MR images in T1, T2, and PD modes from RIRE database [100] and the second row shows the PC computed for the three modes. The complex wavelet representation by phase congruency in Eq. 5.16 yields a poor representation of details with images having low structural contrast, which is particularly the issue in the PD mode compared to other two modes.

are shown along with the corresponding structural representation. As can be seen, the fine details in structures are poorly represented when different structures in the original image are presented in a low contrast. As the structural contrast is decreasing in a mode, the representation will not be able to distinguish the edges between tissues and regions. This issue is more clear in the PD mode of MRI particularly in the regions distinguishing the gray and white matter.

One approach to address the issues associated with the poor structural contrast is to

increase the response sensitivity of the representation to structural characteristics. The approach is to force more emphasis on the finer level of details in the image and integrate the results with the features captured by complex wavelet transform. Aside from phase congruency, which is used to extract highly informative features from the image, the gradient of the image is utilised as the secondary feature to encode contrast information. The traditional method to extract edge information from an image is to compute the image gradient [24], which can be expressed in the form of convolution masks. Here, the common Sobel operator [24] is used to extract the gradient

$$\begin{aligned} G_x(\mathbf{x}) &= \tfrac{1}{4}\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * I(\mathbf{x}) \\ G_y(\mathbf{x}) &= \tfrac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I(\mathbf{x}), \end{aligned} \tag{5.17}$$

where G_x and G_y are the partial derivatives along the x and y directions. Then, the gradient magnitude is defined as

$$G_m(\mathbf{x}) = \sqrt{G_x^2(\mathbf{x}) + G_y^2(\mathbf{x})}. \tag{5.18}$$

The final stage of extracting structural features is to combine features captured by complex wavelet representation with gradient-based information. After applying intensity normalisation on PC and gradient magnitude, a combination strategy in the following generic form can be used

$$R_c(\mathbf{x}) = \varphi\Big(\varphi_1\big(PC(\mathbf{x})\big), \varphi_2\big(G_m(\mathbf{x})\big)\Big), \tag{5.19}$$

where φ_1, φ_2, φ, and R_c are respectively the function applied on the phase congruency, gradient magnitude of the image, fusion function, and the resulting image representation.

Since images have different intensity mappings, the edge information obtained by gradient magnitude may be different in terms of contrast and brightness. Therefore, after having edges extracted, a step of intensity normalization followed by histogram equalization can help to equalise the edge representation [24]. The result of histogram equalization will be an image, named $\widetilde{G_m}$, which can be calculated for each intensity value $G_m(\mathbf{x})$.

The goal is to fuse structures extracted by PC and edge information in gradient image in such a way that pixel locations with high edge information will be strengthened in the PC image. Therefore, the combination strategy is proposed to be in the following format:

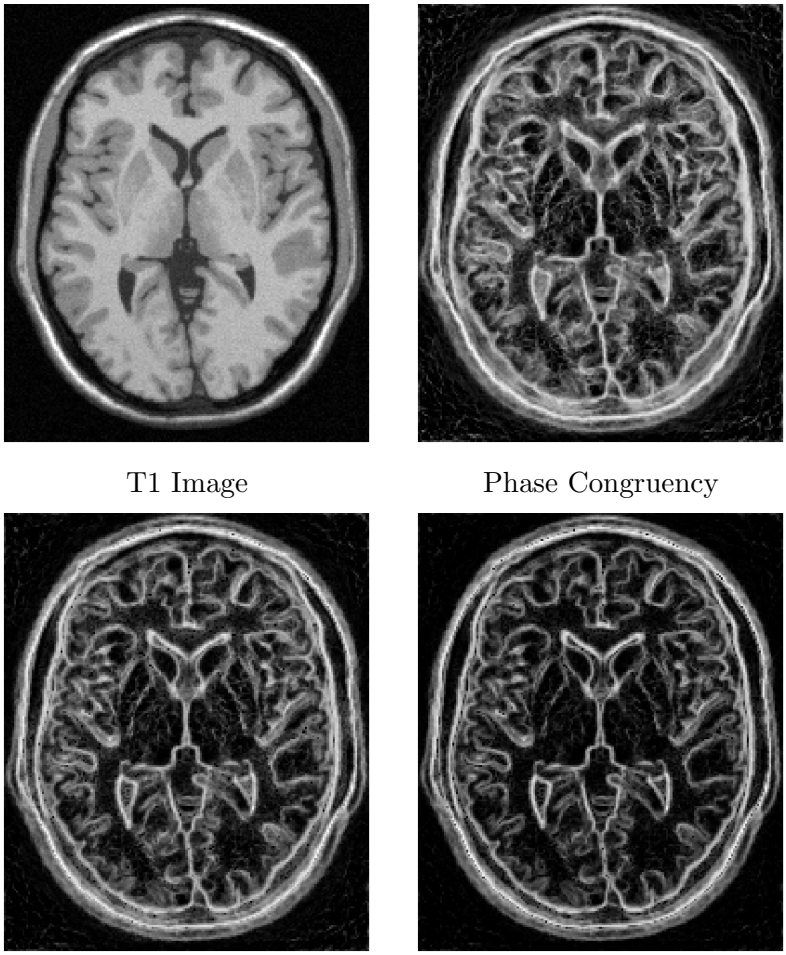
$$R_c(\mathbf{x}) = \widetilde{G_m}^a(\mathbf{x}) \cdot PC^b(\mathbf{x}), \tag{5.20}$$

where $0 \leq \widetilde{G_m}(\mathbf{x}) \leq 1$, $0 \leq PC(\mathbf{x}) \leq 1$, and (a,b) are constant parameters that are used to adjust the importance of phase congruency and edge information. One can control the contribution of PC and gradient magnitude in the resulting structural representation by adjusting factors a and b. Fig. 5.4 shows the result of applying gradient magnitude on the PC result for a T1 brain slice from BrainWeb in two different cases with $(a = 0.5, b = 1)$ and $(a = 1, b = 1)$. As can be seen in this figure, with $a < 1$, more edge information as well as more blurry and noisy effects will be preserved.

Fig. 5.5 shows the resulting structural representation for a slice of BrainWeb MR data in three modes of T1, T2, and PD using the proposed representation. The parameters in this test are set to $a = 0.5$ and $b = 1$. As is shown in this figure, significant edge information which is common in all modalities is preserved and the intensity information which is not consistence across modalities is ignored.

5.3 Entropy-based Representation

As discussed in Section 5.1, to reduce the multi-modal image registration problem to a mono-modal one, an image representation is required to be independent of intensities for encoding the image. Section 5.2 discussed about using complex wavelets to construct a form of image representation. As mentioned in Section 5, various methods have been proposed in the recent literature to transform the problem of multi-modality into mono-modal registration. Employing image entropy is one of the recent methods that work successfully in structural representation for multi-modal image registration [63]. In this section, another image representation, independent of the method in Section 5.2, is proposed based on measuring the local entropy to measure the local information content that is invariant to intensity. Entropy-based representation is constructed by utilizing a modified version of entropy images in a patch-based manner. Fig. 5.6 illustrates the overall procedure of constructing the structural representation using the entropy images.

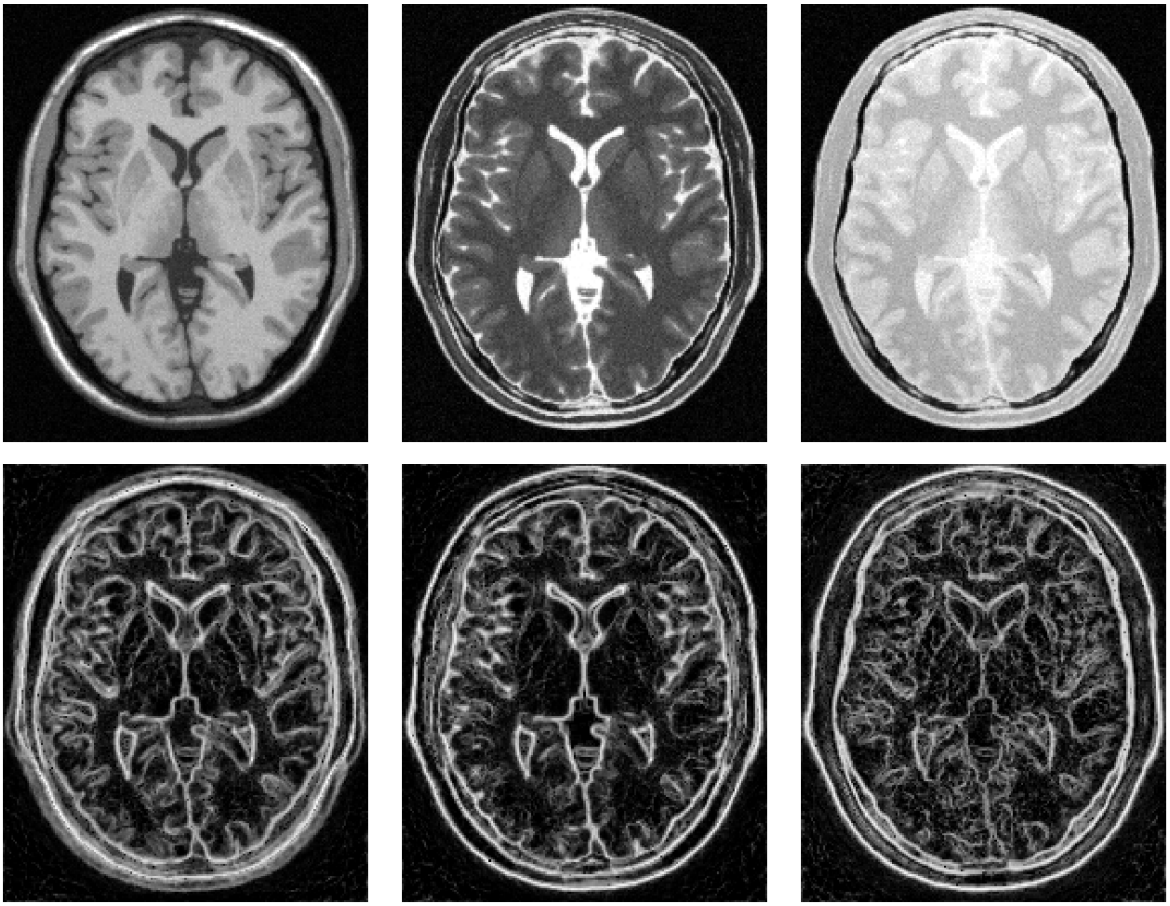


T1 Image | Phase Congruency

Fused Image ($a = 0.5$, $b = 1$) | Fused Image ($a = 1$, $b = 1$)

Figure 5.4: Effect of applying gradient magnitude on PC for a slice of T1 brain MR image. The combination is performed using Eq. 5.20 and the results for two different a values ($a = 0.5$ and $a = 1$) are compared. For lower a value ($a = 0.5$), more edge information as well as more blurry and noisy effects will be preserved.

The information required for constructing the representation are captured from patches. Consider patches $P_{\mathbf{x}}$ defined on the local neighbourhood $\mathcal{N}(\mathbf{x})$ centred at $\mathbf{x}$. The objective is to find a mapping $f_R : P_{\mathbf{x}} \longrightarrow R(\mathbf{x})$ such that $R(\mathbf{x})$ represents the pixel $\mathbf{x}$ based on the information in the surrounding neighbourhood $\mathcal{N}(\mathbf{x})$. The function f is desired to be



T1 T2 PD

Figure 5.5: Structural representation for different MR modes based on a combination of phase congruency and gradient information. A slice of brain scans in T1, T2, and PD modes and the corresponding structural representations are shown in the first and second rows respectively. Significant edge information which is common in all modalities is preserved and the intensity information which is not consistent across modalities is ignored.

defined in a way that it could meet the following requirements:

- Similar patches should lead to similar representations

$$\|P_1 - P_2\|_I < \epsilon \Rightarrow \|f_R(P_1) - f_R(P_2)\|_I < \epsilon'. \tag{5.21}$$

The criteria to choose ϵ and ϵ' rely on the definition of the distance norm $\|\cdot\|_I$

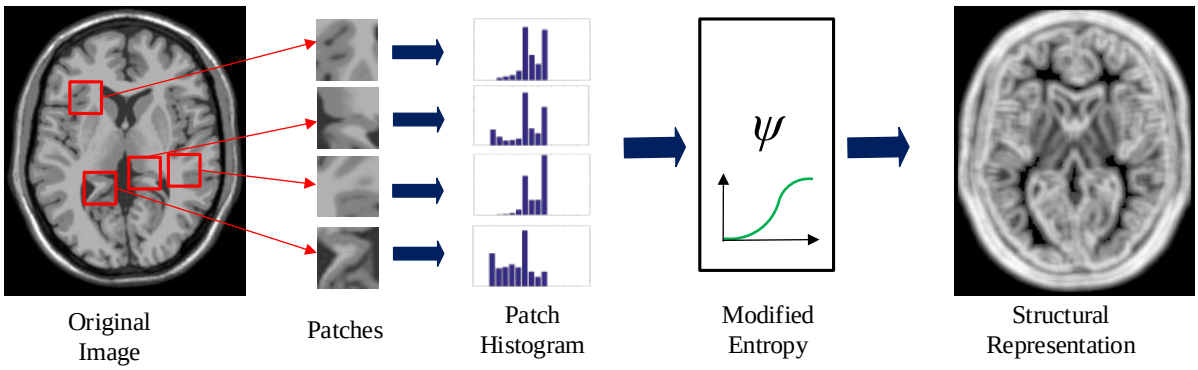


Figure 5.6: Overview of the modified entropy approach for constructing the structural representation: Patch-based calculation image histogram followed by a modified version of entropy results in the structural representation.

to determine the patch dissimilarity. Here, the patch dissimilarity is based on the intensity-based comparison between patches.

- Patches with the same structures should lead to similar representations

$$\|P_1 - P_2\|_S < \epsilon \Rightarrow \|f_R(P_1) - f_R(P_2)\|_I < \epsilon'. \tag{5.22}$$

The norm $\|\cdot\|_S$ here represents the dissimilarity based on structural comparison.

- Different patches should lead to different representations

$$\|P_1 - P_2\|_S > \tau \Rightarrow \|f_R(P_1) - f_R(P_2)\|_I > \tau'. \tag{5.23}$$

In other words, when the patch dissimilarity exceeds a specified threshold τ, the dissimilarity between the representations is expected to be greater than a certain level τ'.

5.3.1 Entropy Image

Wachinger *et al.* in [63] presented to use image entropy as the structural representation for registration of multi-modal images. To form the image representations, the idea is to

extract structural information of each patch based on the amount of information content in the patch. The bound for the amount of information in the patch $P_{\mathbf{x}}$ can be represented by Shannon's entropy which is defined as

$$H(P_{\mathbf{x}}) = -\sum_{\mathbf{x}\in\mathcal{N}(\mathbf{x})} p(I = I(\mathbf{x})) \log \big(p(I = I(\mathbf{x}))\big), \tag{5.24}$$

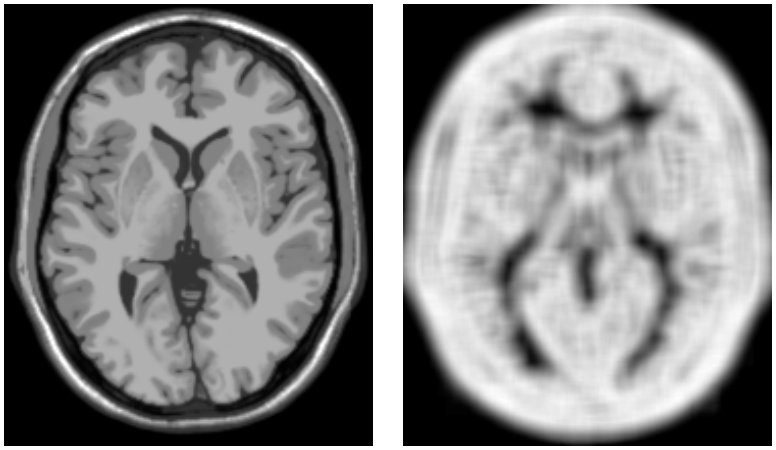
where the random variable I takes the pixel intensity values in $\mathcal{N}(\mathbf{x})$ with possible values in $\mathcal{I}$ characterized by the patch histogram p. Calculating the entropy on the image grid Ω results in an image representation R_e

$$R_e(\mathbf{x}) = H(P_{\mathbf{x}}). \tag{5.25}$$

To obtain the patch histogram p, Parzen windowing method for the non-parametric PDF estimation is used that yields a better estimation for small number of samples in the smaller patch sizes. Based on the entropy representation, as the variation in the patch intensity increases, the representation reflects higher entropy and a higher value will be assigned to the centre of the patch. Fig. 5.7 presents an example of patch-based entropy representation for a brain scan obtained from the BrainWeb database [33] while the patch size is chosen to be 11×11. Patches with different structures are shown to illustrate that patches with higher intensity variation will take higher entropy value to represent the patch structures.

5.3.2 Problem of Distinctiveness

The entropy is able to reflect the information about the patch as a representation for the pixel centring the patch. According to criteria explained for having the representation, we can see that the first requirement is fulfilled, since small changes in the patches lead to small changes in the entropy as well. The second requirement guarantees the same structures to have the same representations. This requirement is also satisfied since the difference in the intensity mapping of the images will result in a permutation in the histogram bins which does not affect the entropy value. However, the third requirement is not fulfilled since it is possible that patches with different structures can end up with the same histogram and therefore the same entropy value. This concept is shown in Fig. 5.8, in which patches encoded in the same intensity mappings but with different structure take the same value as entropy.



Original image Entropy image

$H(P_{\mathbf{x_1}}) = 5.72 \quad H(P_{\mathbf{x_2}}) = 5.21 \quad H(P_{\mathbf{x_3}}) = 3.84$

Figure 5.7: Entropy as a representation for image structures: The first row shows the resulting entropy representation of a T1 weighted MR image from the BrainWeb database [33]. The second row illustrates that higher variations in the patch intensity results in higher entropy values.

Weighting patch histogram based on spatial information forces a constraint in the calculation of patch entropy resulting in differentiating different patches with the same information content. A Gaussian weighting kernel defined as follows is employed for this purpose

$$G(\mathbf{x}) = G_\sigma(\|\mathbf{x} - \mathbf{x_0}\|), \tag{5.26}$$

where $G(\mathbf{x})$ is centred at $\mathbf{x_0}$ with variance σ. Therefore, the entropy for the patch P_x will be modified to

$$\tilde{H}\big(I(P_\mathbf{x})\big) = -\sum_{\mathbf{x}\in P_\mathbf{x}} G(\mathbf{x}) p\big(I = I(\mathbf{x})\big) \log\Big(p\big(I = I(\mathbf{x})\big)\Big). \tag{5.27}$$

The discrimination between patches is not optimal since we are not assigning a unique

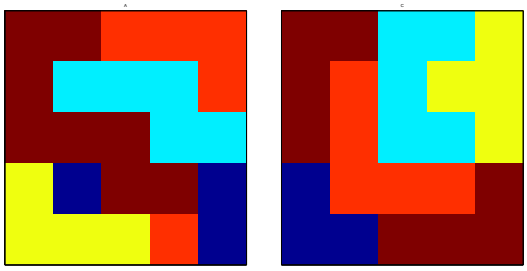


Figure 5.8: Problem of distinctiveness for entropy-based image representation: two sample patches with different structures have the same entropy ($H = 2.24$) and are represented with the same value.

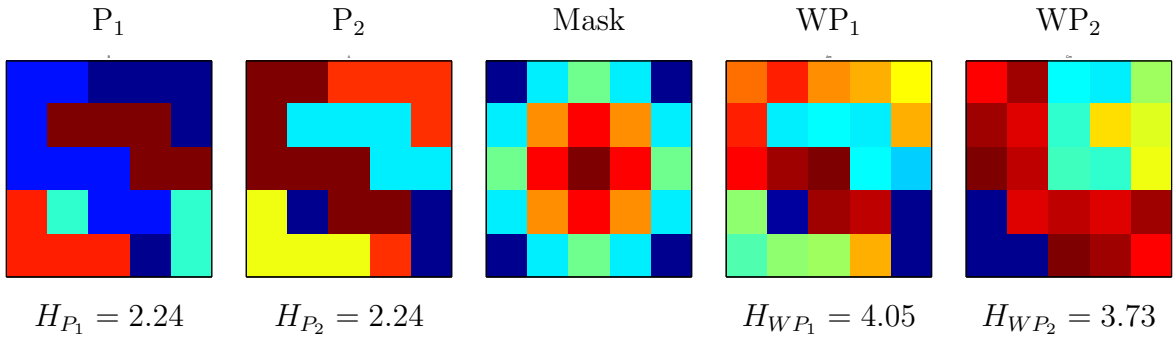


Figure 5.9: Applying a location dependent weighting to differentiate patches with different structures and the same entropy: P_1 and P_2, with the same structure and entropy, are encoded in two different intensity mappings. Applying a Gaussian kernel (Mask) to the patches results in WP_1 and WP_2 with different entropy values.

weight at each patch location. However, conditioning the histogram on the spatial information helps to reduce the number of different structures with the same entropy. Fig. 5.9 shows how weighting the patch histogram by using a Gaussian mask helps to differentiate patches with different structures and the same entropy. In this figure, patches P_1 and P_2, which have the same structure but are encoded in two different intensity mappings, take the same value as entropy value of $H = 2.24$. Patches WP_1 and WP_2 are the weighted patches corresponding to P_1 and P_2 that can be differentiated by two different entropy values of $H_{WP_1} = 4.05$ and $H_{WP_2} = 3.73$.

5.3.3 Modified Entropy Representation

Patch information is mainly concentrated on structures and edges, whereas smooth areas contain less information in the patch. Edges, corners, and generally important structures are mostly pixels with lower probability and smooth areas are represented with the higher probability values in the patch histogram. We propose to focus on structures and highlight the pixels with higher uncertainty while decreasing the contribution of those pixels in the patch that are located in the smooth areas.

For calculating the patch entropy in Eq. 5.27, the weighted pixel information is defined as

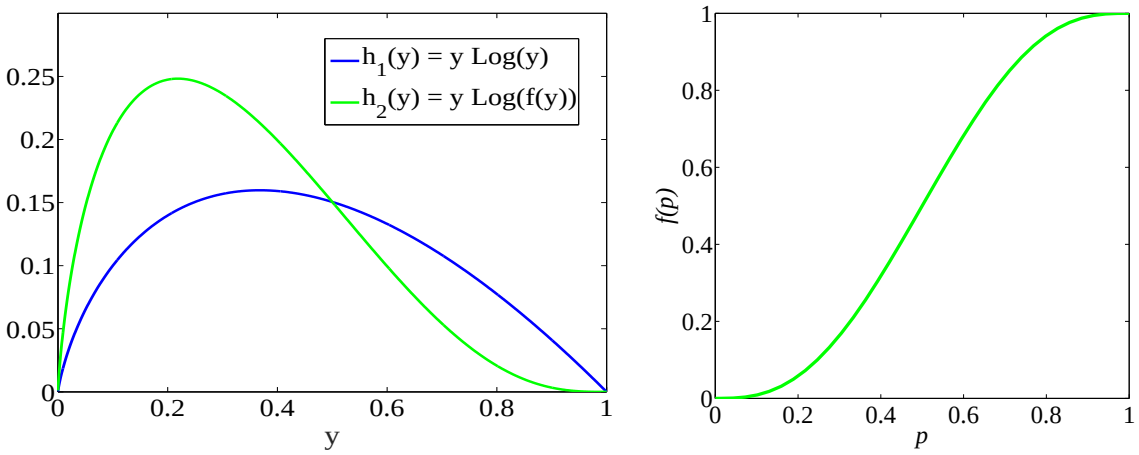
$$h(y) = -y \log(y), \tag{5.28}$$

where $y = p\big(I = I(\mathbf{x})\big)$. In Fig. 5.10(a), $h(y)$ is shown by the blue curve. When y represents the histogram for the patch intensity values, smoother areas will take larger values of y, and edges and structures will take smaller ones. To lessen the contribution of smoother areas and highlight edges and structures, one way is to use a function ψ to map the probability values of the patch histogram such that $\psi(y) > y$ for larger ys, and $\psi(y) < y$ for small ys. Therefore, the weighted pixel information in Eq. 5.28 will be modified to

$$h(y) = -y \log(\psi(y)). \tag{5.29}$$

An example of function f is shown in Fig. 5.10(b). The green curve in Fig. 5.10(a) is the result of applying such function on the patch histogram. As is illustrated in this figure, applying ψ increases the contribution of pixels with lower probability and highly weakens the pixel contribution in the smooth areas compared to calculating the conventional entropy. Having these characteristics for the function $\psi(.)$, it should be an ascending function defined in the range of $[0, 1]$ with lower derivatives on the two endpoints of the range $[-1, 1]$ and a linear behaviour in the middle of the range. The function ψ, which is able to satisfy those characteristics, can simply be chosen as an m–th order polynomial function with symmetry property:

$$\psi(y) = \sum_{i=0}^{m} a_i y^i. \tag{5.30}$$



(a) Weighted pixel information (b) Polynomial function ψ

Figure 5.10: Applying function f on the patch histogram. (a) Weighted pixel-information before and after applying the function ψ on the patch histogram. Applying f makes the curve tilt towards the vertical axis and highly attenuates its value around $y = 1$, where we have higher intensity probabilities. (b) Function ψ to apply on the patch histogram, which has almost linear behaviour around center and a smooth slope around boundaries.

.

As an example of such function, we chose a polynomial function with order $m = 5$. The resulting polynomial function, which is shown in Fig. 5.10(b), will be:

$$\psi(y) = 6y^5 - 15y^4 + 10y^3. \tag{5.31}$$

Finally, the modified entropy with respect to $P_{\mathbf{x}}$ will be calculated by applying the proposed function ψ and weighting kernel G as

$$\tilde{H}\big(I(P_{\mathbf{x}})\big) = -\sum_{\mathbf{x}\in P_{\mathbf{x}}} G(\mathbf{x}) p\big(I = I(\mathbf{x})\big) \log\left(\psi\Big(p\big(I = I(\mathbf{x})\big)\Big)\right), \tag{5.32}$$

which is proposed as the new representation, $R_{M_e}(\mathbf{x})$, for the pixel located at $\mathbf{x}$

$$R_{M_e}(\mathbf{x}) = \tilde{H}\big(I(P_{\mathbf{x}})\big). \tag{5.33}$$

Fig. 5.11 shows the resulting structural representation of different MR modes for a slice of a brain scan from simulated BrainWeb MR data [33]. As indicated in this figure,

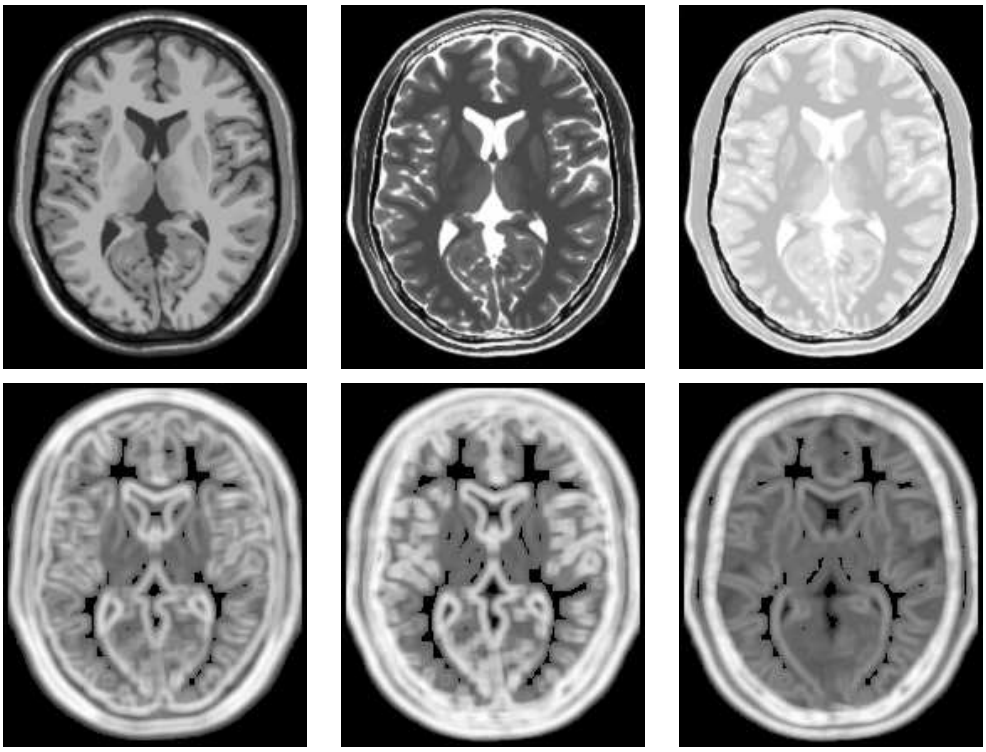


Figure 5.11: Structural representation for different MR modes. The first row shows a slice of brain scans in T1, T2, and PD modes from BrainWeb database. Second row shows the structural representations R_{M_e} associated with the first row images.

structural representation is capable of changing the problem of multi-modal registration to a mono-modal one by applying the mono-modal distance metrics on the representation of images.

5.4 Summary

In this chapter, two structural representations for registering multi-modal images were proposed. The proposed methods were designed to reduce the multi-modal problem to a mono-modal one and by representing images from multiple modalities in a new intensity mapping, so that a mono-modal registration framework can be employed for the alignment.

The first proposed approach extracts structural features based on information from over-complete complex wavelet transform along with gradient magnitude of images. Gradient information was integrated with the complex wavelet response to make an emphasis on the finer level of details. A combination strategy was designed to fuse the information captured by the phase congruency and the gradient magnitude.

The second proposed approach introduced a structural representation which was generated in a patch-based framework by measuring the information content in the patches. The conventional entropy representation was modified to increase the sensitivity of the representation to important structures in the image. Since entropy cannot provide a distinct representation for each structure, a weighting mask was used to condition the measurement on the spatial information. The modification in measuring the patch entropy was designed to decrease the contribution of smooth areas and highlight the edges in the entropy measurement. The proposed approaches, which are aimed to transform the multi-modal registration problem to a mono-modal problem, will be assessed in Chapter 6 in a framework for registering images from different modalities.

Chapter 6

Multi-Modal Image Registration

This chapter presents the results of performance evaluation for the similarity measure proposed in Chapter 4 and structural representations proposed in Chapter 5. Proposed methods are employed in separate frameworks of registering multi-modal images. Brain scans from CT and MR images are used for the assessment. Rigid and non-rigid deformations on both simulated and real brain scans are considered to assess the proposed methods[1].

6.1 Introduction

As discussed in Section 4.1, the registration problem is formulated as

$$\hat{F} = \underset{F}{\operatorname{argmax}}\, \rho\big(I_f, F(I_m)\big), \tag{6.1}$$

where I_m and I_f are the moving and fixed images. The objective is to find a transformation F that maximises the similarity ρ between I_f and transformed I_m. Based upon the problem description in Chapter 3, the focus is on registering images from multiple modalities. This problem was tackled from two different points of view.

First, in Chapter 4, a similarity measure was proposed to assess the degree of alignment for multi-modal image registration. The proposed similarity measure works based on the

[1]Some text and materials in this chapter have been previously published [68, 69] or accepted for publication [71, 72].

assumption that internal pixel-to-pixel relationships are similar in different modalities. The internal similarity, known as image self-similarity, is measured for each of the images to be aligned and compared to form the similarity measure in Eq. 6.1. The self-similarity of an image is estimated by assessing the proximity of image pixels in a patch-based paradigm.

In the second way of tackling the registration problem, two approaches of structural representation were proposed in Chapter 5 to reduce the multi-modal problem to a mono-modal one. The first approach, in Section 5.2, makes use of a combination of gradient information and undecimated wavelet complex representation to extract structural features of images and represents an intensity-independent representation. As an alternative way of constructing structural representation, the second approach was presented in Section 5.3 based on using localised entropy in images. A modified entropy formulation was proposed to extract structural information from images of multiple modalities.

Experiments have been designed to assess the accuracy of multi-modal registration for the proposed methods. In the experiments, the registration accuracy is quantitatively assessed by the average pixel displacement, which measures the Euclidean distance between the pixel positions in the transformed image and their corresponding positions in the ground truth [101]:

$$\tau = \frac{1}{|\Omega|} \sum_{i=1}^{|\Omega|} (\mathbf{x}_i - \mathbf{x}'_i)^2, \tag{6.2}$$

where $\mathbf{x}_i$ and $\mathbf{x}'_i$ are respectively the position of the i-th pixel defined on the image grid Ω in the ground truth and aligned images.

In this chapter, the methods proposed in the previous chapters are used in a framework of multi-modal registration and experiments in both rigid and non-rigid registrations are preformed to evaluate the performance of the methods. Registering multi-modal images from CT and different MR modes are employed and the registration accuracy are quantitatively evaluated using the measure τ in Eq. 6.2.

6.2 Experimental Data

In order to evaluate the performance of the proposed similarity measure in Chapter 4 and the structural representations presented in Chapter 5, the registration procedure is

performed in independent experiments conducted on simulated and real brain scans. The performance of each registration method is evaluated by comparing the estimated transformations to the gold standard transformations. The gold standard transformation is obtained by artificially deforming the image. This difference of deformations by the artificial deformation and the estimated deformation by the registration method is quantified using the average pixel displacement, which is defined as the distance of each pixel position from its true position in the gold standard and averaged over all pixels employed in the registration.

Simulated Data: Simulated scans are obtained from the BrainWeb simulated brain database [33] containing a set of realistic MR brain volumes produced by an MRI simulator. 3D MR scans are provided in T1, T2, and PD modes at a resolution of 1mm^3 with different levels of noise and intensity non-uniformity.

Real Data: Real data are from the Retrospective Image Registration Evaluation (RIRE) [100] real database. The RIRE database provides real brain scans in different modalities of T1/T2/PD-weighted MR, PET, and CT scans. The ground truth alignment is also provided in this database.

6.3 Self-similarity measure

The self-similarity measure proposed in Chapter 4 is used in a registration framework to assess the multi-modal registration accuracy. According to Section 4.3.4, the similarity SM corresponding to every pixel $\mathbf{x}$ in the transformed moving image FI_m and the fixed image I_f is measured given the self-similarities $\mathcal{S}(I_m, \mathbf{x})$ and $\mathcal{S}(I_f, \mathbf{x})$ as

$$\text{SM}(I_m, I_f; \mathbf{x}) = \text{MI}\big(\mathcal{S}(I_m, \mathbf{x}), \mathcal{S}(I_f, \mathbf{x})\big), \tag{6.3}$$

where MI is used to compare the self-similarity of the two images. Parzen windowing [102] is used to estimate the intensity histogram in the MI calculation. The self-similarity $\mathcal{S}$ of an image at pixel $\mathbf{x}$ is obtained based on patch-based comparing of pixel $\mathbf{x}$ and other pixels in neighbourhood $\mathcal{N}_r(\mathbf{x})$. The patch-based comparison was suggested to be either

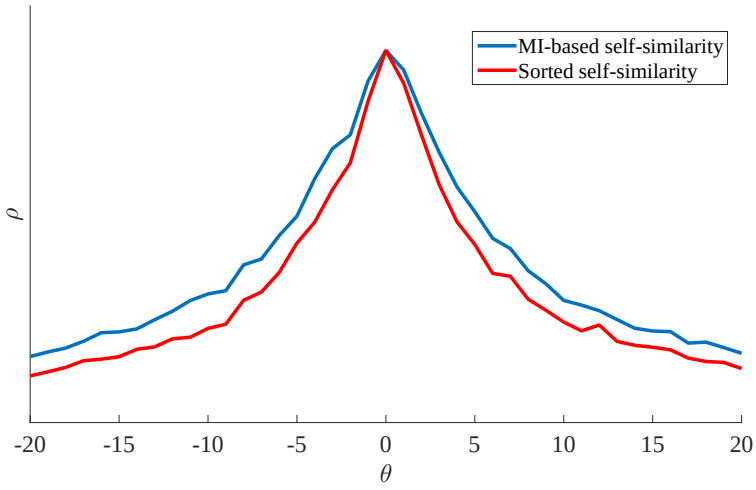


Figure 6.1: Comparing the usage of MI and sorted patch intensity comparison in measuring self-similarity: similarity is measured for a pair of T1-T2 MR images from BrainWeb database when one image rotated by θ.

based on measuring MI of patches or the SSD of sorted patches $\tilde{P}$ as described in Eq. 4.9 to Eq. 4.11. Fig. 6.1 describes a simple test to show how the two approach of patch-comparison can detect rotational deformations. The similarity for a 2D T1-T2 comparison is measured when one image is taking rotations in the range [-20°, 20°]. As can be seen, both approaches lead to correct detection of rotations and both take their maximum at $\theta = 0$. The difference is that using sorted patches results in a slightly more sensitivity to rotational deformations, while the usage of MI brings about capturing a slightly wider range of deformations. For the sake of simplicity of sorting operation and its sensitivity to rotation, the sorted patch intensity comparison is considered for the rest of simulations.

The similarity in Eq. 6.3 is measured for N randomly selected pixels and averaged to yield the scalar similarity measure:

$$\rho(I_m, I_f; \Omega) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{SM}(I_m, I_f; \mathbf{x}_i). \tag{6.4}$$

In the experiments, $N = 10^4$ voxels are used to estimate the similarity between the fixed image and the transformed image. The similarity measure in Eq. 6.4 is used for both rigid and non-rigid registration of brain scans.

To evaluate the performance of the proposed similarity measure, it is compared with the multi-modal registration based on MI as the similarity measure [19] and registration based on MIND descriptor [79]. Both rigid and deformable registration scenarios are considered for the evaluation procedure. For the MIND method, the parameters are set to the defaults as suggested in [79]: a Gaussian weighting $\sigma = 0.5$ with a corresponding patch size $3 \times 3 \times 3$, the and search region within six pixel neighbourhood for the pixel of interest. In the proposed method, the patch size and number of bins in the histogram are empirically chosen to be $7 \times 7 \times 7$ voxels and 64 bins. We also limit the self-similarity to the neighbourhood with radius of 25 pixels.

Experiments are conducted on the BrainWeb simulated database and RIRE real database. In the following experiments, scans with 3% noise and 20% intensity non-uniformity are chosen to include the effect of noise and bias field in the experiments. Brain scans that are used from the BrainWeb and RIRE datasets are in different MR modes of T1, T2, and PD.

6.3.1 Rigid Registration

For rigid registration, the configuration is $11 \times 11 \times 11$ for 3D patches, 64 bins and Parzen-window estimation [102] for MI calculation in Eq. 6.3.

Translation and rotation are examined on 3D data in two separate experiments by generating 50 random transformations for each case. First, translation is chosen in the range of $[-20, 20]$ mm with no rotation. In the second experiment, we have maximum rotation of $\pm 20^\circ$ with zero translation. The average results of rigid registration for random transformations in terms of average displacement τ in mm are illustrated in Table 6.1 for BrainWeb and in Table 6.2 for RIRE data.

Table 6.1 reports the accuracy for registration of BrainWeb data with rigid deformations (rotation and translation). Different configurations with MR modalities are examined. As is shown in Table 6.1, the proposed method shows a substantial improvement over the conventional MI-based registration for all rotational and translational deformations. Comparing to MIND, in both translations and rotations, promising improvements have been achieved, specifically for rotational deformations improvements were considerable.

Table 6.1: Multi-modal rigid registration (translation and rotation) using the self-similarity measure for BrainWeb dataset. Registration errors are represented in average pixel displacement τ.

	Similarity	T1-T2	T1-PD	T2-PD
Rotation	MI	1.87	1.54	1.12
	MIND	1.15	1.32	1.03
	Proposed	**0.61**	**0.82**	**0.69**
Translation	MI	1.87	1.32	1.11
	MIND	1.43	0.87	0.77
	Proposed	**1.33**	**0.78**	**0.56**

Table 6.2: Multi-modal rigid registration (translation and rotation) using the self-similarity measure for RIRE dataset. Registration errors are represented in average pixel displacement τ.

	Similarity	T1-T2	T1-PD	T2-PD	T1-CT
Rotation	MI	3.82	2.34	2.73	4.61
	MIND	2.87	2.13	2.85	3.92
	Proposed	**2.08**	**1.74**	**2.21**	**3.86**
Translation	MI	2.94	2.12	2.04	3.86
	MIND	2.29	1.67	1.71	**2.88**
	Proposed	**2.17**	**1.54**	**1.56**	2.97

The same experiment has been performed for the real RIRE dataset. Results are shown in Table 6.2 with different configurations with MR modalities and CT scans. As is shown, the proposed method outperforms the conventional MI-based registration for all cases of this experiment. Comparing to MIND, the proposed method shows a significant improvement, especially for the rotational transformation. The results for the translational transformation are still promising and only in two cases of T1-PD and T1-CT the MIND can achieve a better accuracy.

Overall, it can be deduced from the results from both simulated and real data that

Table 6.3: Multi-modal deformable registration using the self-similarity measure for RIRE dataset. Registration errors are represented in average pixel displacement τ.

Similarity	T1-T2	T1-PD	T2-PD	T1-CT
MI	2.87	3.12	3.54	**5.93**
MIND	2.04	2.41	2.73	6.72
Proposed	**1.91**	**2.24**	**2.61**	7.85

the proposed self-similarity measure is more robust in rigid transformations especially in rotation, since the self-similarity is independent of pixel ordering in patch-based comparison and does not rely on the arrangement of the pixels in the patch.

6.3.2 Non-Rigid Registration

For deformable registration, we used artificial deformations by the thin-plate spline (TPS) [103] to generate a set of randomly deformed training data. The deformation field is normalised to limit the maximum displacement to 20mm. The registration is modelled by the FFD with three hierarchical levels of B-spline control points [45]. The optimisation is performed by the gradient descent optimization method to iteratively update the transformation parameters. The results of deformable registration in multi-modal cases are shown in Table 6.3. Similar to experiments in Section 6.3.1, the performance of the proposed method is compared with the MIND and MI-based registration. The results in this table are obtained by averaging the alignment error for 20 random deformations.

As is shown in Table 6.3, the proposed similarity measure achieves a better performance in T1-PD and T2-PD registration compared to both MIND and MI-based registration. The registration with CT is more challenging due to the significant differences between MR and CT images.

6.4 Structural Representation for Image Registration

This section presents the results of registering multi-modal images using the structural representations proposed in Chapter 5. The structural representation have been proposed

to reduce the multi-modal registration problem to a mono-modal one, so that a simple SSD measure can be used in the optimisation framework. Thus, given the representations R_f and R_m for I_f and I_m respectively, the registration problem turns into

$$\hat{F} = \underset{F}{\operatorname{argmax}}\, \rho\big(R_f, F(R_m)\big). \tag{6.5}$$

Two approaches were proposed to transform the images into representatives independent of image intensities. The first proposed approach in Section 5.2 works based on a combination of gradient information and complex wavelet transform and the second one presents a new representation by applying a modified entropy on the images. In the following, experimental results regarding each of the two methods are presented.

6.4.1 Complex Phase and Gradient Information

The method presented in Section 5.2 is assessed based on the multi-modal brain scans. The proposed method, which is the result of complex-wavelet representation and gradient information, is evaluated using brain scans from T1, T2, and PD modes generated using the BrainWeb simulator. To assess the method, we used MR scans with noise level of 3%, 5%, and 7%, and intensity non-uniformity (INU) of 20% and 40%. The noise level is specified by a number representing the percent ratio of the standard deviation of the white Gaussian noise versus the signal. The intensity inhomogeneity level is presented by the scaled range of field values over the brain area. The structural features are extracted using log-Gabor transform in 4 scales and 6 orientations, with wavelengths of 3, 9, 27, and 81 pixels to keep bandwidths of two octaves.

To investigate the performance of the proposed complex wavelet representation, the similarity measures based on phase congruency (PC), gradient magnitude (GM), and the proposed method (PC-GM) are shown in Fig. 6.2. The image dissimilarity is measured by measuring the SSD of structural representation in each case over rotations in the range $[-40°, 40°]$. As is shown, the dissimilarity measure using the proposed representation performs correctly and takes its minimum at $\theta = 0$. The behaviour over the changes in θ is smooth and not far from the response from gradient magnitude or PC. Depending the parameters α and β, the response of the proposed method may change.

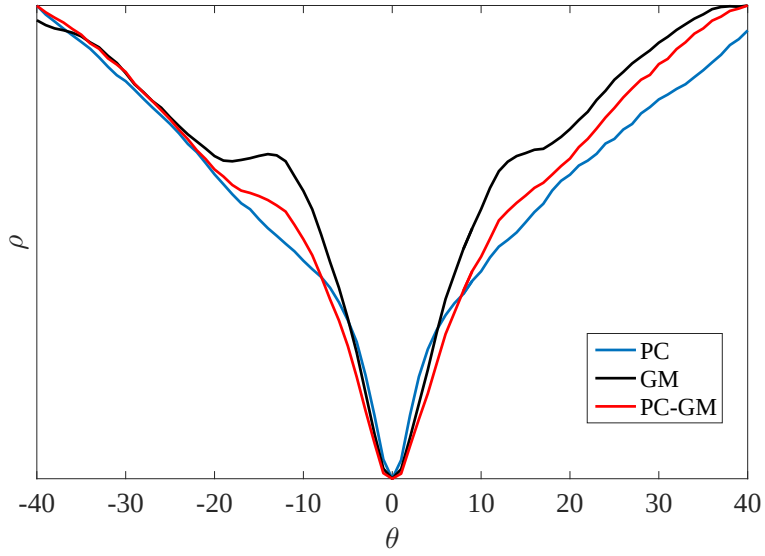
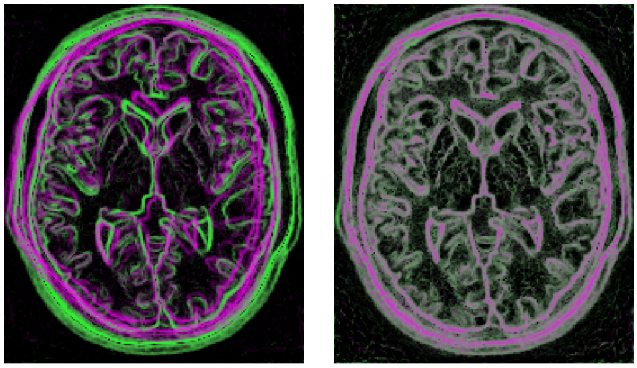


Figure 6.2: Similarity plots for BrainWeb dataset when one image is deformed by rotation in the range $[-40°, 40°]$

To assess the performance of the method over random non-rigid deformations, a set of training data was generated using artificial deformations generated by TPS. We compared our approach with the conventional multi-modal registration method based on using mutual information as the similarity measure.

In order to qualitatively assess the performance of the proposed method, the result of multi-modal registration for two different modalities is shown in Fig. 6.3. For this figure, we have selected the 75th slice of brain scan in PD and T1 modes of MR imaging generated by BrainWeb simulator with 3% noise and 20% intensity non-uniformity level. The T1 image is considered as the fixed image and the slice in PD mode is deformed using the TPS to generate the test moving image. Features extracted from both moving and fixed images, before and after being aligned, are shown in this figure. Features are shown in different colors, so that the alignment can be compared before and after applying the registration.

Quantitative results for registering multi-modal images with different levels of noise and intensity non-uniformities are shown in Table 6.4 for T1-T2, T1-PD, and T2-PD registration. Quantities in this table are obtained by averaging the results of registering 20



Before Registration After Registration

Figure 6.3: Cross-modal registration using the proposed method based on complex wavelet representation: A PD slice (red) is registered to a T1 slice (green) for a sample slice from BrainWeb database with 3% noise and 20% INU. Features of the two images are shown before and after registration to illustrate the degree of alignment.

randomly deformed images to a fixed image. The performance of the registration by the proposed method is compared to the conventional MI-based multi-modal registration. As can be seen, as the noise and intensity non-uniformity level increase, the performance of the registration method is degraded in all three cases. In case of T1-T2 registration, for 7% noise and 20% intensity non-uniformity, the proposed method and MI-based registration method perform almost the same. For T1-PD and T2-PD cases, because of poor contrast representation of PD mode compared to other modes, the registration accuracy is seen to be lowered. Specifically, at 7% noise and 20% INU, MI-based registration performs better than the proposed method. As the non-uniformity increases, the proposed method is shown to be more accurate than the MI-based method. This is due to the fact that MI is highly sensitive to non-uniformity in image intensity. However, the overall performance of the proposed registration method, which is illustrated as the average over all noise and INU levels, demonstrates higher accuracy compared to the conventional MI-based registration method.

Table 6.4: Quantitative comparison of registration errors (in mm) obtained by MI and the proposed complex wavelet representation method (Proposed) from BrainWeb with different levels of noise and INU.

	Method	Noise and INU level (in percent)						Average
		3,20	5,20	7,20	3,20	5,20	7,20	
T1-T2	MI	1.74	2.13	3.07	2.34	3.81	5.11	3.03
	Proposed	**1.11**	**1.89**	**3.05**	**1.27**	**2.32**	**3.46**	**2.18**
T1-PD	MI	1.97	2.85	**4.21**	3.63	5.64	7.21	3.19
	Proposed	**1.59**	**2.13**	4.28	**1.93**	**3.14**	**5.03**	**3.02**
T2-PD	MI	2.14	3.48	**5.63**	4.83	6.94	8.12	4.97
	Proposed	**1.23**	**2.74**	5.94	**2.39**	**4.03**	**5.84**	**3.69**

6.4.2 Modified Entropy Image

This section focuses on the structural representation, proposed in Section 5.3, based on applying a modification in entropy formulation to increase the sensitivity of dissimilarity measure to finer structures. In order to evaluate the performance of the proposed method, experiments are again conducted on the BrainWeb and RIRE data that are provided by ground truth alignment. In the following experiments, T1, T2, and PD modes of MR scans from BrainWeb dataset and real brain scans T1, T2, PD, and CT from the RIRE dataset are used.

The proposed method, which is represented as 'Proposed' in the following tables, is compared with the MI-based registration [19] and SSD on entropy images (eSSD) [63]. The optimization for the rigid registration is carried out by MATLAB tools based on gradient descent optimizer for the SSD based mono-modal, and one-plus-one evolutionary optimizer for the MI-based multi-modal registration. Both rigid and deformable registration scenarios are considered for the evaluation procedure. The deformable registration is performed by FFD. In our simulations, the patch size and number of bins in the histogram are empirically chosen to be 7×7 pixels and 64 bins.

Table 6.5: Multi-modal rigid registration (translation and rotation) using modified entropy for BrainWeb dataset: Registration errors are represented in average pixel displacement τ.

	Similarity	T1-T2	T1-PD	T2-PD
Rotation	MI	0.63	0.76	0.35
	eSSD	0.85	0.54	0.14
	Proposed	**0.54**	**0.38**	**0.08**
Translation	MI	0.41	0.52	0.32
	eSSD	0.72	0.64	0.18
	Proposed	**0.37**	**0.48**	**0.14**

Rigid Registration

For rigid registration, the proposed method is evaluated by comparing the alignment result with the ones using MI and eSSD. Fig. 6.4, shows the behaviour of the multi-modal similarity/dissimilarity measures when one image is rotated by $\theta \in [-40°, 40°]$. The plots are obtained from different combination of MR modes from BrainWeb scans. In general, the proposed method and eSSD have the same behaviour when θ changes and in terms of smoothness, the proposed method does not force more cost compared to the eSSD.

Quantitative assessment is performed by measuring the displacement error in both cases of having rotation and translation in separate experiments. Experiments are conducted when translation is in the range of $[-20, 20]$ mm with $0°$ rotation, and in maximum rotation of $\pm 20°$ with zero translation. Table 6.5 and Table 6.6 report the average results for BrainWeb and RIRE datasets, respectively. The experiments have been carried out for 50 times over different rotations and translations and the results are reported in terms of average displacement τ in mm.

Quantitative results on the BrainWeb dataset show that all three methods result in comparable alignment accuracy, however the proposed method shows its superiority over the other two methods. On the real RIRE dataset, the proposed method performs significantly better than MI-based registration and could improve the results of eSSD as well. Despite the increase in the registration error for CT-T1 alignment, the improvement for

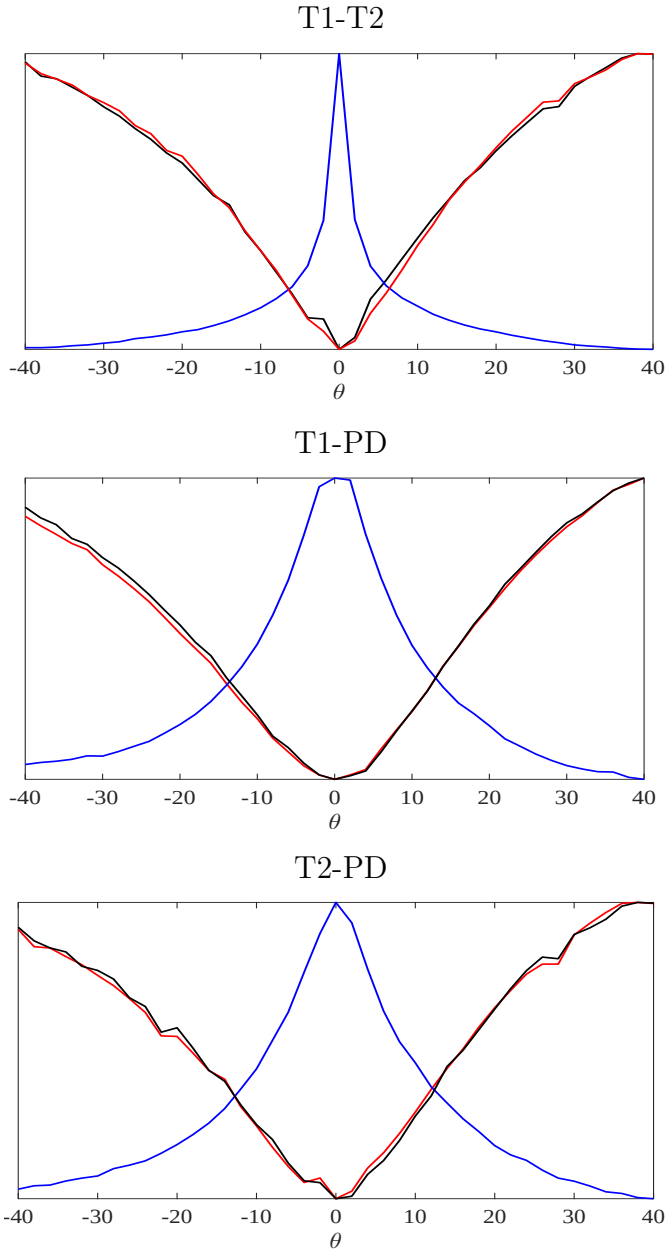


Figure 6.4: Similarity plots for BrainWeb dataset when one image is deformed by rotation in the range $[-40°, 40°]$ (black: modified entropy, red: eSSD, blue:MI)

Table 6.6: Multi-modal rigid registration (translation and rotation) using modified entropy for RIRE dataset: Registration errors are represented in average pixel displacement τ.

	Similarity	T1-T2	T1-PD	T2-PD	T1-CT
Rotation	MI	3.02	1.14	2.74	3.62
	eSSD	2.03	0.83	2.34	2.87
	Proposed	**1.74**	**0.61**	**2.13**	**2.64**
Translation	MI	1.58	0.87	1.93	2.53
	eSSD	0.35	0.44	0.98	1.73
	Proposed	**0.28**	**0.33**	**0.71**	**1.69**

Table 6.7: Multi-modal deformable registration using modified entropy for RIRE dataset. Registration errors are represented in average pixel displacement.

Similarity	T1-T2	T1-PD	T2-PD	T1-CT
MI	1.23	1.47	1.87	**2.15**
eSSD	0.67	0.61	0.55	7.32
Proposed	**0.61**	**0.58**	**0.41**	5.43

the both MR and CT data is still considerable.

Non-rigid Registration

For deformable registration, a set of training data was generated from the dataset using artificial deformations by the thin-plate spline. The deformation field is normalized such that the maximum displacement is limited to 20 mm. The results of deformable registration is given in Table 6.7 for different combinations of image modalities. Similar to Table 6.1 and Table 6.2, the proposed method is compared with eSSD and MI-based registration results. Quantities in this table are obtained by averaging the results of aligning 20 randomly deformed images to a fixed image.

As can be seen, the proposed method in most cases outperforms the eSSD and MI-based registration. Since the proposed method tends to extract structural features and

structural features are mainly located in the rigid body of the image, the improvement in the alignment accuracy for the rigid registration is more significant. It can be seen that for non-rigid registration, the proposed method leads to considerable improvement over the MI. The results show a slight improvement over eSSD, however, the method is not able to outperform the MI method in the T1-CT registration.

6.5 Discussion

Three different registration approaches, two based on structural representation and the other one based on self-similarity measurement, have been evaluated in this chapter. The average displacement error is measured to assess the accuracy of each method on real and simulated data. An average pixel displacement of zero represents perfect registration, and a large average pixel displacement indicates poor registration performance. If the average pixel displacement obtained from each of the methods in registering real data is greater than 3 pixels, then the performance of the registration method is considered to be failed [104].

Looking at the results from registering simulated and real brain images, we can deduce the following points. First, in all experiments, registering different modes of MR images is performed successfully, when comparing to the traditional registration method based on mutual information. Wavelet-based registration performs promising in registering T1 to T2 modes of MRI, comparing to other combinations, which means that low contrast PD mode with the poor edge representation cannot yield good accuracy compared to T1-T2 registration. Among all three methods, registration based on modified entropy seems to perform more robust on registering images from different combination of MR modes.

Second, in all experiments, registering MRI T1 scan to CT scans are problematic and the proposed methods fail to attain acceptable alignment accuracy. Comparing the proposed methods based on self-similarity measure and modified entropy to registering based on MI as the similarity measure, MI can overcome the proposed methods specifically in non-rigid registration real brain images. The key issue in this case is that the MI-based registration performs globally on the image and the proposed methods are local. Since, for the CT scan that mainly contains rigid structures and not much of fine details of other tissues, global measurement can perform better. Performing a hierarchical framework to

Table 6.8: Comparison of computation time in seconds for different registration approaches in non-rigid registration of T1-T2 3D MR brain images.

Method	Time (sec)
MI	287
MIND	524
Proposed self-similarity	407
eSSD	83
Proposed modified entropy	112
Proposed wavelet-based	168

start with global alignment and leading to local warping could offer more in case of MR-CT registration.

To evaluate the three proposed registration methods in terms of computation time, an experiment has been performed to register a set of 3D MR scans from T1 to T2 mode from the RIRE dataset in a non-rigid framework. The running time for the methods that have been used as in the previous comparison has been recorded. Table 6.8 illustrates the running time for the non-rigid registration based on mutual information (MI), MIND self-similarity method (MIND), proposed self-similarity, structural representation based on entropy and SSD comparison (eSSD), proposed modified entropy, and proposed wavelet-based registration. As can be seen, eSSD, proposed modified entropy, and proposed wavelet-based method, which are all based structural representation, have the lowest computation time and the MIND method has the highest one. This table demonstrates that registration based on structural representation and using a simple intensity-based dissimilarity measure increases the speed of the registration procedure significantly. The proposed self-similarity measure is also compared to the MIND self-similarity approach and shows faster performance due to using lower number of pixel-similarities in the descriptor.

6.6 Summary

We presented the results of registration assessment for the methods presented in Chapter 4 and Chapter 5. Evaluations are performed on simulated and real brain data from CT scans and T1,T2, and PD modes of MR images. The registration is performed in both rigid and non-rigid frameworks and the results are shown in terms of average pixel displacement from the true pixel position. The methods are compared to the registration methods from the literature. Mutual information is used as the classical method of registering multi-modal images and MIND as the state-of-the-art method for self-similarity measurement. Results are obtained from independent experiments for each of the proposed methods. Overall, based on the results presented in this chapter, the proposed methods can outperform the conventional mutual information-based and the state of the art in terms of overall accuracy. In terms of computation time, the methods based on structural representation performs highly faster that the ones based on self-similarity. The running time for the proposed self-similarity approach is less than the state-of-the-art MIND method, due to employing smaller sets of pixels in the self-similarity map.

Chapter 7

Label Fusion

This chapter describes in detail the overall problem of cross modality label combination in multi-atlas segmentation problems. The problem of label fusion in multi-atlas-based segmentation framework, related issues, and challenges are explained in Section 7.1. Section 7.2 presents the weighted voting strategy which is the conventional fusion approach. However, weighted label fusion performed either globally or locally relies on the intensity consistency across images. To address this issue, the problem of multi-modality in fusing atlas labels and the proposed method for cross modality label fusion are presented in Section 7.3. The proposed method is presented based on assessing the structural similarity across different modalities instead of intensity based comparison. The performance of the method is evaluated in Section 7.4 in a procedure of segmenting brain tissues in MR images given a multi-modal brain atlas database[1].

7.1 Introduction

As described in Section 2.3 a major component in the multi-atlas framework is "label fusion" by which atlas labels are combined to form a single segmentation for a target image [12, 13]. According to description of overall multi-atlas-based segmentation framework which is presented in Chapter 3 and Fig. 3.1, a final segmentation result L_T is generated

[1]Some text and materials in this chapter have been previously published [70].

by combining all propagated labels, $\{L'_n\}$ using a label fusion method. Fig. 7.1 reviews the multi-atlas segmentation framework with the focus on label fusion.

Many label fusion methods have been introduced in the medical atlas literature [22]. Majority voting (MV) as the simplest and most widely used fusion method assumes each atlas contributes to the target labels equally [13]. As the image intensity is not taken into account during label fusion, a higher accuracy can be achieved by some form of weighting, based on the similarities between the atlases and the target image. Weighting strategies including both global and local forms [65, 66], where local weighted voting (LWV) outperforms global strategies when dealing with high contrast anatomical structures [21, 22, 23]. Many label fusion methods, such as MV, do not consider image intensities after being warped to the target image. If we do consider the image intensities and give higher weights to those more similar atlases, whether globally or locally, we obtain improvements in segmentation accuracy [21, 65, 105].

The multi-atlas approaches are promising, however these methods remain problematic in those cases where the atlases and the target scan are obtained from different sensors or from different acquisition modalities: image-intensity comparisons may no longer be valid, since image brightness can have highly differing meanings and circumstances in different modes [16]. Most label fusion approaches are limited by the assumption that they depend on the consistency of voxel intensities across different MRI scans. In these cases, approaches based on mutual information do help [56, 67, 106], however its inherent non-locality make it problematic for local weighted label fusion. This issue will be highlighted when atlases and target image are acquired with different modalities [16, 21].

Relying on the similarity between intensity values of the atlases and target scan is often problematic in medical imaging — in particular when the atlases and target image are obtained via different sensor types or imaging protocols. In [17], a generative probabilistic model is proposed that yields an algorithm for solving the atlas-to-target registrations and label fusion steps simultaneously. This model exploits the consistency of voxel intensities within the target scan to drive the registration and label fusion instead of intensity similarity, hence the atlases and target image can be of different modalities. The method is based on exploiting the consistency of voxel intensities within the segmentation regions, as well as their relation with the propagated labels.

To focus on the process of label fusion in this chapter, the multi-atlas segmentation

framework is presented in Fig. 7.1. We seek to develop a cross-modality label fusion weighted on the basis of the similarity of the transformed atlases $\{A'_n\}$ and the target image I_T. The goal is to measure the atlas-target similarities $\mathrm{SM}_{\mathcal{F}}$ and weight the contribution of atlases' label map $\{L'_n\}$ to construct the final target segmentation L_T. The design of similarity measure relies on the structural relationships of the atlases and the target and based on the scale-based features extracted from an undecimated wavelet transform (UDWT).

7.2 Weighted Label Voting

The label fusion problem in a multi-atlas segmentation can be inferred from a maximum-a-posteriori (MAP) estimation framework [21]:

$$\hat{L}_T(\mathbf{x}) = \underset{l\in\{1,\cdots,\mathcal{L}\}}{\operatorname{argmax}} \sum_{n=1}^{N_A} p\big(L'_T(\mathbf{x}) = l|L_n\big)p\big(I_T(\mathbf{x})|A'_n\big), \tag{7.1}$$

where $p\big(L'_T(\mathbf{x}) = l|L_n\big)$ is the label prior value and $p\big(I_T(\mathbf{x})|A'_n\big)$ is the probability that relates the n-th atlas to the target image which can be interpreted an assigned weight to the n-th vote [107].

Traditional majority voting produces the final segmentation, L_T, by assuming that different atlases provide equal registration quality and no prior knowledge about the accuracy of the labels of each atlas as a classifier labels is used. It is assumed that $p(I_T(\mathbf{x})|A'_n) = C$, where C is a constant and reduces the Eq. 7.1 to

$$\hat{L}_T(\mathbf{x}) = \underset{l\in\{1,\cdots,\mathcal{L}\}}{\operatorname{argmax}} \sum_{n=1}^{N_A} p\big(L_T(\mathbf{x}) = l|L'_n\big). \tag{7.2}$$

Typically, for deterministic atlases, discrete values of 0 and 1 are used instead of $p\big(L_T(\mathbf{x}) = l|L'_n\big)$. As mentioned above, $p\big(I_T(\mathbf{x})|A'_n\big)$ gives a hint of the relation between two images which has been interpreted in the literature as the image likelihood and is quantified by measuring the image similarity [21, 107, 108]. Thus, the target label map in Eq. 7.1 is estimated by weighting the label prior and assigning greater weights to warped atlases

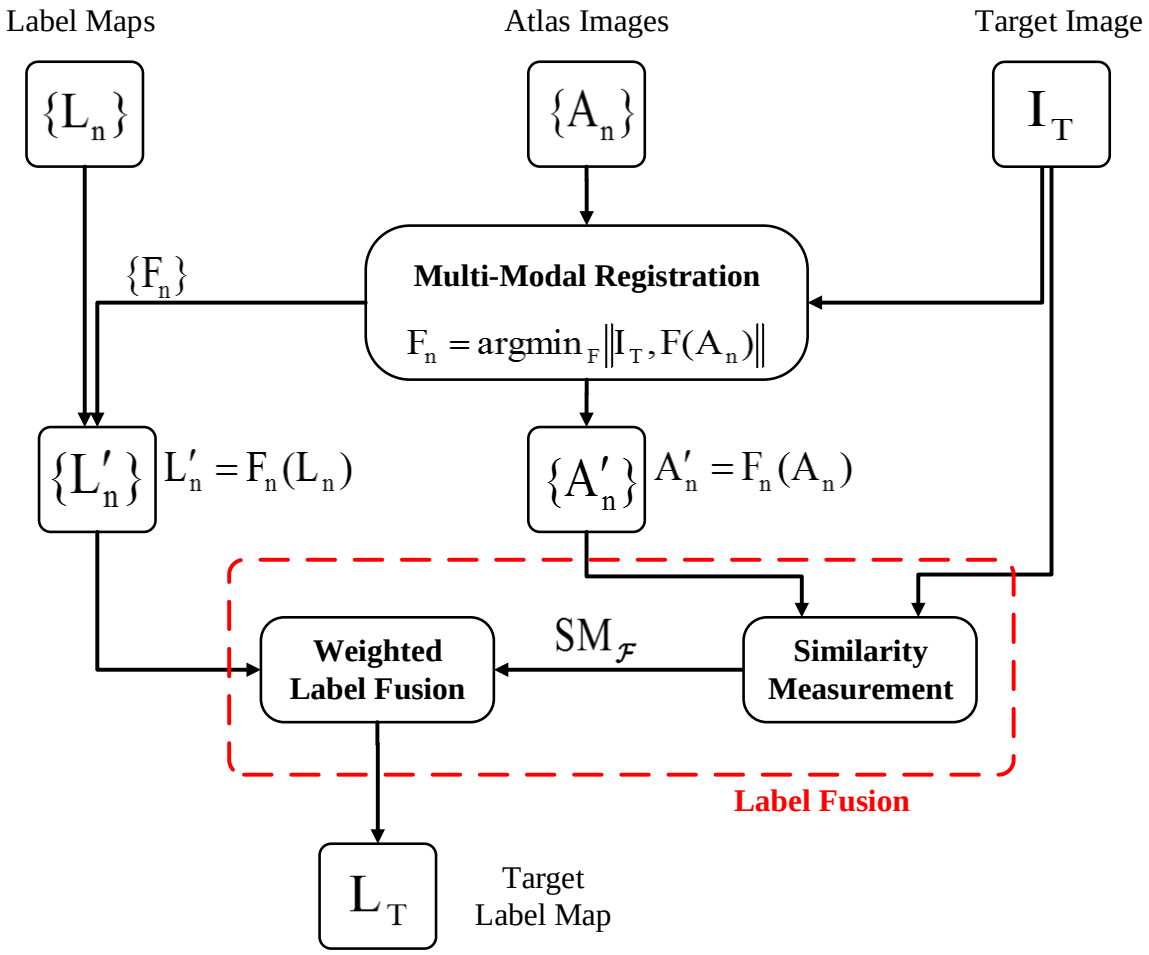


Figure 7.1: Block-diagram of the multi-atlas-based segmentation for multi-modal atlas database, which is shown in Fig. 3.1, with the focus on label fusion. The atlas-target similarity ($SM_{\mathcal{F}}$) is used to weight the atlas contributions to form the final segmentation result.

that are more similar to the target image:

$$\hat{L}_T(\mathbf{x}) = \underset{l\in\{1,\cdots,\mathcal{L}\}}{\operatorname{argmax}} \sum_{n=1}^{N_A} w_n(\mathbf{x}) L'_n(\mathbf{x}), \tag{7.3}$$

where w_n is the weight assigned to the nth atlas with

$$\sum_{n=1}^{N_A} w_n(\mathbf{x}) = 1. \tag{7.4}$$

If $w_n(\mathbf{x}) = w_n, \forall \mathbf{x}$, then the atlases would be ranked globally according to the atlas-target similarity. One way to estimate the set of weights $\{w_n\}$ is to locally measure the similarity of the target image and atlases after being registered, based on the assumption that similar regions are more likely to have the similar label maps. The local weighted voting is performed in a patch-based paradigm, in which the image likelihood $p\big(I_T(\mathbf{x})|A'_n\big)$ is defined on a neighbourhood $\mathcal{N}(\mathbf{x})$ centring at pixel $\mathbf{x}$ with patch size $(2r+1)^d$ for d dimensional images. To model the image likelihood, a Gaussian distribution is generally used as

$$p(I_T(\mathbf{x})|A'_n) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\big(I_T(\mathbf{x}) - A'_n(\mathbf{x})\big)^2\right), \tag{7.5}$$

with σ as the variance of the distribution [21, 107, 109]. However, this model relies on the intensity comparison of images and cannot model the intensity relationship in multi-modal cases.

7.3 Cross-Modality Label Fusion

Since images are obtained from different sensors, the intensity relationship between the images is complex and therefore the intensity-based image likelihood in Eq. 7.5 is not able to model atlas-target similarity. A label fusion method is proposed based on defining a structural similarity measure to approximate the similarity of the atlas and the target image, for which the block diagram is depicted in Fig. 7.2. As shown in the figure, multi-scale complex wavelet representation of the input images are constructed using an undecimated complex wavelet transform such as log-Gabor complex wavelet transform [110]. The multi-modal image representation based on complex wavelet coefficients is presented in Section 5.2. As

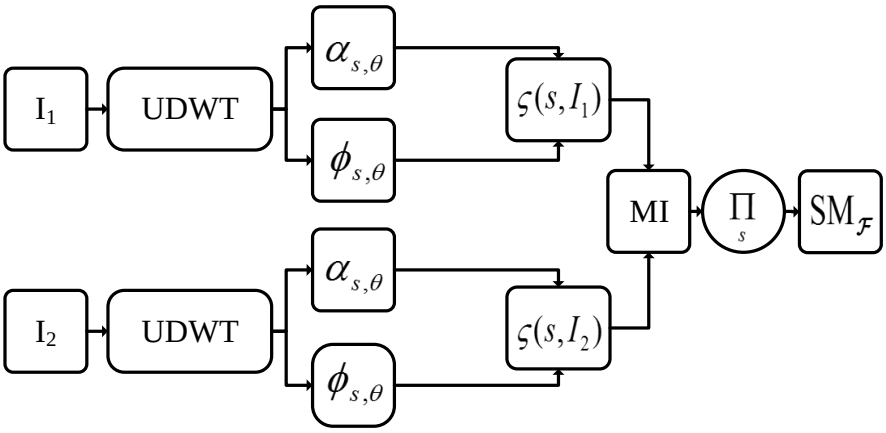


Figure 7.2: Similarity measure for multi-modal images based on structural features. Similarity measure is obtained by computing the mutual information of structural features captured by the UDWT.

in Eq. 5.2, the resulting wavelet coefficients for the scale s and orientation θ are noted as $\Upsilon_{s,\theta}(\mathbf{x})$ at location $\mathbf{x}$,

$$\Upsilon_{s,\theta}(\mathbf{x}) = \alpha_{s,\theta}(\mathbf{x}) \exp[j\phi_{s,\theta}(\mathbf{x})], \tag{7.6}$$

where $\alpha_{s,\theta}(\mathbf{x})$ and $\phi_{s,\theta}(\mathbf{x})$ are the amplitude and phase of the complex wavelet coefficients, respectively. The phase order, $\zeta(s, I(\mathbf{x}))$ at each scale can be defined as the normalised weighted summation of phase deviations from its mean value across all scales:

$$\zeta\big(s, I(\mathbf{x})\big) = \frac{\sum_\theta \alpha_{s,\theta}(\mathbf{x})\Lambda(\mathbf{x})}{\sum_{\theta,s} \alpha_{s,\theta}(\mathbf{x})}, \tag{7.7}$$

where

$$\Lambda(\mathbf{x}) = \cos(\phi_{s,\theta}(\mathbf{x}) - \bar{\phi}_\theta(\mathbf{x})). \tag{7.8}$$

Here, $\Lambda(\mathbf{x})$ is the phase deviation from the mean value of the complex phase $\phi_\theta(\mathbf{x})$. Fig. 7.3 shows the structural features of different modes of a brain MR slice from the BrainWeb simulated database [33]. As can be seen, the intensity information, which is the problematic part of the label fusion, is no longer present and instead the aspects which remain are the structural features that are almost the same in all modalities.

In order to measure the similarity between each atlas and the target image, the similarity is calculated across all scales based on the structural features represented by ρ_s.

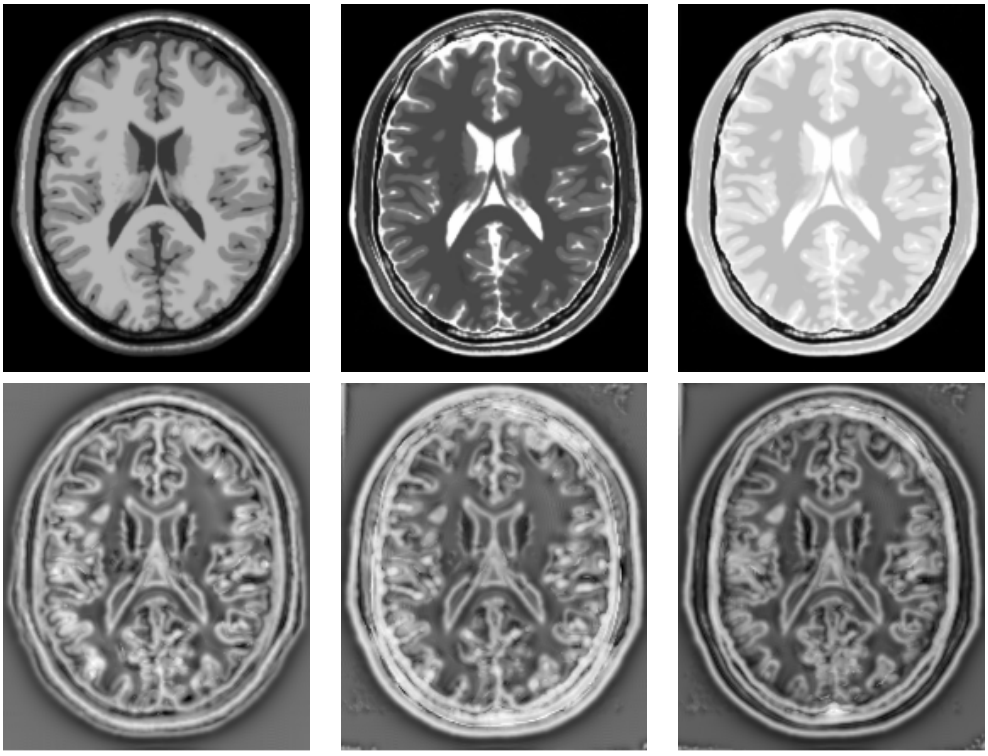


Figure 7.3: Structural features from different MR modes. The first row shows a slice of brain scans in T1, T2, and PD modes. Second row shows the structural features associated with the first row images extracted from the second scale of log-Gabor complex wavelet transform implemented in 4 scales and 6 orientations with wavelengths of 3, 9, 27, and 81 pixels.

In this way, features from fine and coarse scales of one mode are compared correspondingly to those extracted from the other mode and the results of scale-based comparison are combined to form a measure of similarity. Mutual information based on image intensity entropy is utilised to measure the similarity of structural features at each scale. MI for two images I_1 and I_2 is defined as

$$\mathrm{MI}(I_1, I_2) = H(I_1) + H(I_2) - H(I_1, I_2) \tag{7.9}$$

In this equation, $H(I_1)$ and $H(I_2)$ represent the entropy of the intensity in images I_1 and I_2 and $H(I_1, I_2)$ stands for the joint entropy of these two images. If the MI-based comparison

is performed over the whole image, the label fusion method would be a global weighting that ranks the contribution of warped atlases according to their global similarity to the target image. The MI-based comparison can be carried out in a patch-based paradigm to achieve higher segmentation accuracy by performing a local similarity measurement.

The proposed similarity measure is a function over all scales: the structural features at some scale from the two images are compared using mutual information applied to the phase order from (7.7):

$$\mathrm{SM}_{\mathcal{F}}(I_1, I_2) = \Xi\Big(\mathrm{MI}\big(\zeta(s, I_1), \zeta(s, I_2)\big), s\Big), \tag{7.10}$$

where Ξ denotes the fusion function that combines the MI-based comparison over the scale s. The function Ξ should return a high value when both fine and coarse scales have high similarities and low value when fine and coarse values have small mutual information. A simple example function could be a product of MI obtained from all scales:

$$\mathrm{SM}_{\mathcal{F}}(I_1, I_2) = \prod_{s} \mathrm{MI}\big(\zeta(s, I_1), \zeta(s, I_2)\big). \tag{7.11}$$

Finally, the resulting similarity measure is normalised and applied to Eq. 7.1, contributing to the label fusion paradigm by weighting labels from each atlas based on how similar each atlas image is to the target image:

$$\hat{L}_T(\mathbf{x}) = \underset{L_T}{\operatorname{argmax}} \sum_{n} p(L_T(\mathbf{x}) | L'_n) \mathrm{SM}_{\mathcal{F}}(I_T, A'_n). \tag{7.12}$$

7.4 Results and Discussion

7.4.1 Data

We have tested our method on the 3D brain MR scans from the BrainWeb simulated database [33], as described in Section 6.2, based on the T1, T2, and PD modalities with 3% noise and 20% intensity non-uniformity, and on the T1 images in the LONI real database [35]. The databases provide ground truth of tissue labels for white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF).

7.4.2 Experimental setup

To assess the proposed method, we compared our approach with conventional majority voting and mutual information [108] for segmenting real and simulated MR scans into WM, GM, and CSF tissues. The structural features are extracted using log-Gabor complex wavelet transform in 4 scales and 6 orientations, with wavelengths of 3, 9, 27, and 81 pixels to maintain bandwidths of two octaves. Mutual information is computed using Parzen windowing [102] in estimating the intensity histogram. 32 bits are used to quantise the intensity histogram. The experiments are performed on both simulated and real data.

Simulated Data: In the first test on simulated data, a set of training data was generated by an artificial deformation using thin-plate spline (TPS). Two different cases are examined: a single mode atlas database and a multi-modal atlas database with a target in a different mode from the atlas set. The registration utilised in this framework is undertaken using a non-rigid multi-modal image registration. The free-form deformation model with mutual information as the similarity measure implemented in the ITK, Segmentation & Registration Toolkit, is used. For these experiments, 25 different random deformation fields are generated and the whole process of segmentation is run ten times for each random deformation.

Real Data: To validate the method on real data, the second test was performed by using 40 real T1 atlases and a PD target image. A set of ten training scans out of 40 subjects is randomly selected to form the atlas database and this procedure is run ten times to obtain the segmentation results.

To quantitatively assess the accuracy of segmentation, the Dice similarity coefficient [111] is used, defined as

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \tag{7.13}$$

where A and B are the set of pixels in a segment in ground truth and the segmented image, respectively.

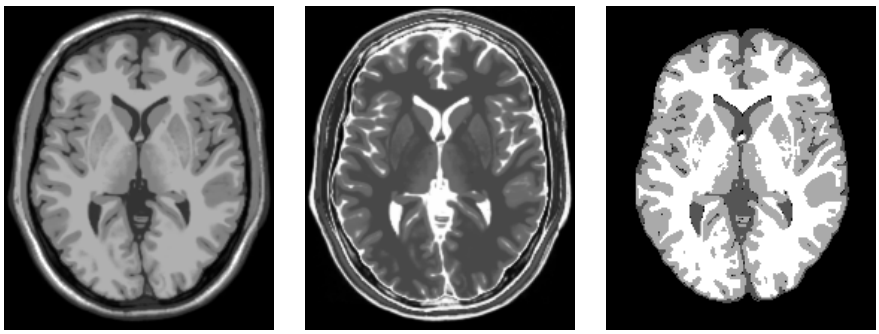
7.4.3 Results

Fig. 7.4 illustrates the advantage of using multi-modal atlases instead of single-mode ones. The effect of adding an atlas in a mode other than the target's mode on the segmentation accuracy is examined using simulated brain data. In this experiment, all atlases are in the same mode as the target image, and a slice of a T1 image is segmented using MV. The experiment is then repeated for the case that additional T2 training data is added. As is shown in this figure, the average Dice coefficient by MV method for the WM, GM, and CSF tissues is increase when using multi-modal training images. Comparatively, the proposed method shows an improvement over the MV method for the multi-modal case. The misclassification error in each of the segmentation results is shown in red color. One should note that, in the MV method, only label maps are used. However, the proposed method takes advantage of the structural features in the new mode as well as the label map to segment the target image.

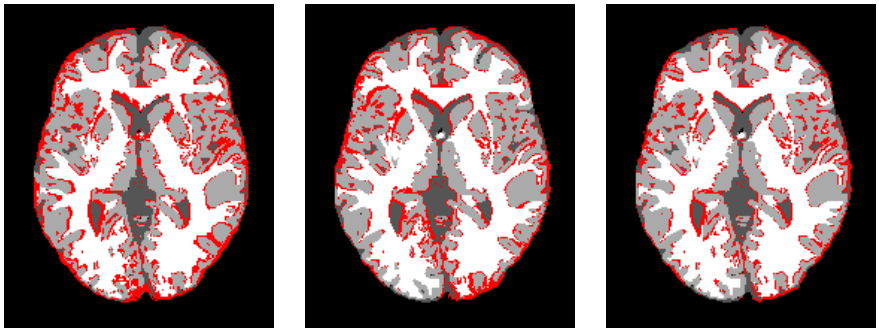
The first experiment on simulated data, which is illustrated in Fig. 7.5, considers the cross-modality segmentation with the single-mode atlas database. For this experiment, first, the target image is assumed to be in T2 mode while the atlas database is in T1. For the second case, the target is changed to PD mode. The atlas database is generated using artificial deformations applied on the simulated images from the BrainWeb database [33]. The segmentation results demonstrate improved performance of the proposed label fusion compared to the traditional MV and MI-based method.

A second experiment is performed to show how the method works for the complex cases with multi-modal atlases and the target image in a mode which does not have any representative in the atlas set. Table 7.1 reports the segmentation results when the database contains atlases of T1 and T2 mode scans and the target image is in PD mode. Results obtained from the proposed method significantly outperforms MV and shows considerable improvement over MI-based method. Also lower standard deviation for the accuracy measurement is achieved.

To evaluate on real data, the method is applied to segment a T2 target image given a set of T1 real normal images randomly selected from LONI database [35]. Table 7.2 shows the results for this experiment. Although the results of the proposed method does not show any improvement for segmenting the GM, it still does a promising job for delineation



T1 target image | T2 training image | Ground truth



MV-single mode 75.2% | MV-multi-mode 77.2% | Seg-multi-mode 80.1%

Figure 7.4: Multi-modal versus single-mode segmentation: the bottom row shows the results of MV and the proposed method, with the Dice coefficient D (7.13) given. The misclassification error of each case is shown in red color. The highest Dice performance is offered by the proposed approach.

Table 7.1: Segmentation results in terms of average Dice coefficient D and its standard deviation when the atlas database consists of T1 and T2 scans and the target scan is in PD mode: the performance of the proposed method (Proposed) is compared to the majority voting (MV) and MI-based weighting (MI).

Tissue	WM	GM	CSF
Proposed	**88.6±0.2**	**88.2±0.2**	**80.7±0.8**
MI	86.9±0.3	86.1±0.4	78.2±1.2
MV	85.6±0.4	85.4±0.5	77.6±1.3

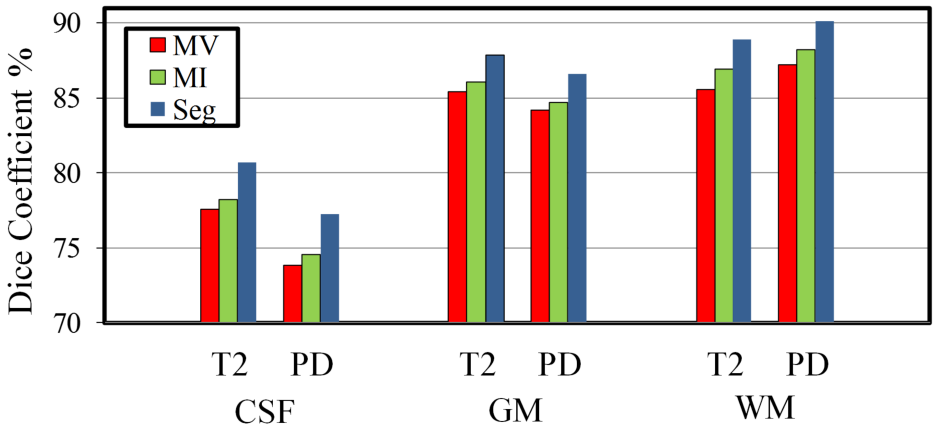


Figure 7.5: Single-mode multi-atlas segmentation results in terms of average Dice coefficient D for the proposed (Seg), majority voting (MV), and MI-based method (MI). The atlas set is in T1 while the target is in T2 and PD.

of the two other tissues. Furthermore, the method is shown to be robust over different atlas selections compared to other reported methods.

7.4.4 Discussion

Overall, the segmentation results demonstrate that the proposed weighted label fusion outperforms the classical MI-based weighted voting for cross-modality label fusion, specifically when the atlas database consists of atlases from different modes of MR images.

Table 7.2: Segmentation results in terms of average Dice coefficient D and its standard deviation when the atlas database consists of T1 scans and the target scan is in T2 mode: the performance of the proposed method (Proposed) is compared to the majority voting (MV) and MI-based weighting (MI).

Tissue	WM	GM	CSF
Proposed	**80.6±0.4**	75.0±**0.2**	**61.2±0.8**
MI	78.9±0.7	**75.2**±0.4	58.3±1.3
MV	77.6±0.8	72.4±0.4	55.1±1.7

In terms of computational complexity, the proposed method forces further loads of computations due to extracting structural features by complex wavelet transforms. However, if the whole label fusion procedure is designed in such a way that all the input atlases and target image are registered to a common space, then there will be no need to perform the whole procedure for every new target image. As a result, registration to the common space and also extracting structural features can be done offline. Estimating the similarity to the target's structural features over all scales and combining them to form the similarity measure in Eq. 7.10 are the steps that affect the computational time and complexity. For measuring the global similarity between each atlas and the target image after being aligned, it is required to compute the mutual information at each scale of structural representation. Since the structural representations are constructed by the over-complete wavelets, the size of the output at each scale will not vary from the input. Therefore, with s representing the number of scales, s MI-based similarity measurements are performed for each atlas. Comparing the proposed method to the classical MI-based weighted voting, we can deduce that the proposed method increases the amount of computations by a factor of s and the order of computations will remain the same.

7.5 Summary

This chapter presented a label fusion method for multi-modal images based on a structural similarity measure. Unlike most of previous label fusion methods that are working on single-mode multi-atlas segmentation, the proposed method is designed to deal with fusing labels across modalities or utilising single-mode atlas set to segment a target in different mode. For this purpose, a similarity measure is proposed based on structural features which can be extracted from undecimated wavelet coefficients. To validate our method, experiments for segmenting tissues in the simulated and real MR brain images were conducted.

Chapter 8

Conclusions

In this thesis research, the cross-modal multi-atlas segmentation framework is considered to segment brain images. In this framework, two major components, which are image registration and label fusion, are the focus of this research and undertaken independently. After highlighting the limitations of multi-atlas segmentation, specifically in multi-modal cases, in Chapter 3, methods have been proposed to deal with the multi-modal registration and cross-modality label fusion. A summary of the thesis contributions is given in the following section.

8.1 Thesis Contributions

The multi-modal image registration has been traditionally carried out using statistical similarity measures. To address the problem regarding the complex intensity relations in multi-modal images and also non-locality of the conventional similarity measures, the first approach is proposed based on comparing the self-similarity of images to be aligned. The relation of each pixel to other pixels in the image is considered and the most significant pixel-to-pixel relations are selected to transmit the required information for the comparison. The motivation and theory of this method is presented in detail in Chapter 4.

In an independent way of tackling the multi-modal registration problem, we focus on reducing the multi-modal problem to a mono-modal one by representing images in a new intensity mapping. Two separate representations are proposed in Chapter 5 to reduce

the registration problem, thus any intensity-based comparison can be utilised to measure the alignment accuracy. The use of undecimated complex wavelet transform along with gradient information is shown to be capable of extracting structural features from images in different MR modes. The alternative representation is take advantage of local entropy in a modified formulation to characterise the structural information in the image.

The similarity measure presented in Chapter 4 and structural representations in Chapter 5 are examined in registration frameworks separately in Chapter 6. The real and simulated brain scans in T1/T2/PD-weighted MRI and CT are utilised to evaluate the methods in both rigid and non-rigid registration paradigms. Experimental results show the superiority of the proposed approaches for multi-modal registration over classical and state-of-the-art methods.

The cross-modality label fusion proposed in Chapter 7 is an extension of the current weighted voting approaches in mono-modal label combination. The label combination method is proposed based on transforming the multi-modal images into a new space and comparing images in this new space. The space transformation is performed using an undecimated complex wavelet transform and the result is presented in different scales of resolution. The scale-based comparison between representations provides the atlas weights in a weighted voting paradigm. The experimental results using real and simulated brain MR images demonstrate the better performance of the proposed label fusion compared to the conventional method for the cross-modal label fusion.

As a summary, the contributions of the dissertation can be listed as:

- Introducing a similarity measure for multi-modal image registration based on comparing the self-similarity of images to be aligned,
- Reducing the multi-modal registration problem to a mono-modal and thus using a simple intensity-based similarity measure by
  - creating a structural representation not relying on the intensity mapping based on combining information extracted using undecimated complex wavelet transform and gradient magnitude of the image,
  - creating a structural representation based on measuring local image entropy in a modified formulation,

- Extending the label fusion to cross-modality label fusion by
  - extracting scale-based structural features using undecimated complex wavelet transform to represent images in a new representation,
  - defining a measure to make a cross-modality comparison between scale-based image representations that are not depending on the intensity of the original images.

8.2 Future Research

The work in this thesis results a number of general outcomes and directions of significance. The research presented in this dissertation provides a foundation for future research in cross-modal multi-atlas segmentation. Three potential research lines that can be pursued based on the research in this dissertation are presented in the following.

8.2.1 Performance Investigation Under Different Circumstances

This dissertation has resulted the methods in registration and segmentation of brain MR images in a multi-atlas segmentation framework. As the first line of research to pursue, we aim to investigate different conditions and circumstances in which the proposed approaches might behave in different way. There are a number of factors that matter when dealing with multi-modal medical images. As the first goal, we would like to investigate the effect of noise variations and changes in the bias field of the MR machine in the performance of the methods. Secondly, we are aiming to expand the application of this work to modalities other than MRI. As the third point, the segmentation in this research is evaluated by classifying three major tissues in the brain, however, in many cases correct labelling of the structures are highly of interest. We aim to expand the brain segmentation and evaluate the label fusion method to classification of different structures in the brain as well as tissues.

8.2.2 Unified Framework for Multi-Atlas-Based Segmentation

One major outcome of this dissertation is taking advantage of structural representations based on scale-based over-complete complex wavelet transform for multi-modal problems.

The label fusion approach and one of the methods proposed to present structural representation work based on the undecimated complex wavelet transform. The complex wavelet representation is shown promising in extracting the structural features in different modalities. Once the representation is made for the images, it is possible to use them for either registration step or label fusion. In a multi-atlas segmentation framework, we aim to yield a unified framework for solving the atlas-to-target registrations and label fusion steps simultaneously.

8.2.3 Joint Multi-modal Registration

With the availability of large databases, multi-atlas segmentation will becoming a more complex problem due to the increase in the number of atlases and anatomical variations in the database. Either of the proposed approaches in image registration, the proposed similarity measure and the structural representations, are designed for pair-wise registration of multi-modal images. A problem with doing pair-wise registrations is that the resulting alignment depends on which image is chosen as the template. The problem of template bias in pair-wise registration has been addressed by proposing groupwise registration in the literature. Congealing framework [112], which evaluates the entropy of a pixel stack, and ensemble registration, based on a maximum-likelihood clustering [104], are two examples. Since the structural representation aims to reduce the complexity of the multi-modal problem, it is possible to speed-up the matching procedure by employing an efficient optimiser based on using such representations. In this line of research, we aim to investigate an efficient objective function based on structural representations such that all images can be aligned in a simultaneous manner.

References

[1] D. D. Blatter, E. D. Bigler, S. D. Gale, S. C. Johnson, C. V. Anderson, B. M. Burnett, N. Parker, S. Kurth, and S. D. Horn. Quantitative volumetric analysis of brain MR: Normative database spanning 5 decades of life. *American Journal of Neuroradiology*, 16(2):241–251, 1995.

[2] Y. Ge, R. I. Grossman, J. S. Babb, M. L. Rabin, L. J. Mannon, and D. L. Kolson. Age-related total gray matter and white matter changes in normal adult brain. Part I: Volumetric MR imaging analysis. *American Journal of Neuroradiology*, 23(8):1327–1333, 2002.

[3] E. Courchesne, H. J. Chisum, J. Townsen, A. Cowles, J. Covington, B. Egaas, M. Harwood, S. Hinds, and G. A. Press. Normal brain development and aging: Quantitative analysis at in vivo MR imaging in healthy volunteers 1. *Radiology*, 216(3):672–682, 2000.

[4] A. F. Fotenos, A. Z. Snyder, L. E. Girton, J. C. Morris, and R. L. Buckner. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*, 64(6):1032–1039, 2005.

[5] M. Sonka and J. M. Fitzpatrick. *Handbook of Medical Imaging*, pages 422–430. 2000.

[6] D. Pham, C. Xu, and J. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2:315–337, 2000.

[7] A. Wee-Chung Liew and H. Yan. Current methods in the automatic tissue segmentation of 3D magnetic resonance brain images. *Medical Imaging Reviews*, 2006.

[8] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158–e177, 2011.

[9] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans. Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*, 3(3):190–208, 1995.

[10] T. Rohlfing, R. Brandt, C. R. Maurer Jr, and R. Menzel. Bee brains, B-splines and computational democracy: Generating an average shape atlas. In *Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis–MMBIA*, pages 187–194, 2001.

[11] J. E. Iglesias and M. R. Sabuncu. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1):205–219, 2015.

[12] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer Jr. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4):1428–1442, 2004.

[13] 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*, 33:115–126, 2006.

[14] J. M. P. Lötjönen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage*, 49(3):2352–2365, 2010.

[15] T. Rohlfing, R. Brandt, R. Menzel, D. B. Russakoff, and J. C. R. Maurer. Quo vadis, atlas-based segmentation? In *Handbook of Biomedical Image Analysis*, pages 435–486. Springer, 2005.

[16] J. E. Iglesias, M. R. Sabuncu, and K. Van Leemput. A generative model for multi-atlas segmentation accross modalities. In *Proceedings of the IEEE International Symposium on Biomedical Imaging–ISBI*, pages 888–891, 2012.

[17] J. E. Iglesias, M. R. Sabuncu, and K. Van Leemput. A unified framework for cross-modality multi-atlas segmentation of brain MRI. *Medical Image Analysis*, 17(8):1181–1191, 2013.

[18] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.

[19] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[20] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[21] M. R. Sabuncu, B. T. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland. A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging*, 29:1714–1729, 2010.

[22] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz de Solorzano. Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Transactions on Medical Imaging*, 28:1266–1277, 2009.

[23] T. R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, M. A. Viergever, M. V. Vulpen, and J. P. W. Pluim. Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). *IEEE Transactions on Medical Imaging*, pages 2000–2008, 2010.

[24] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 3 edition, August 2007.

[25] A. Elnakib, G. Gimelfarb, J. S. Suri, and A. El-baz. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies: Medical Image Segmentation: A Brief Survey*. Springer ScienceBusiness Media, 2011.

[26] Y. F. Shih. *Image Processing and Pattern Recognition: Fundamentals and Techniques*. Wiley-IEEE Press, 2010.

[27] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.

[28] D. Rueckert and A. F. Frangi. Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014–1025, August 2003.

[29] P. L. Bazin and D. L. Pham. Statistical and topological atlas-based brain image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention–MICCAI*, volume I, pages 94–101. Springer, 2007.

[30] M. Kuklisova-Murgasova, P. Aljabar, L. Srinivasan, S. J. Counsell, V. Doria, A. Serag, I. S. Gousias, J. P. Boardman, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert. A dynamic 4D probabilistic atlas of the developing brain. *NeuroImage*, 54:2750–2763, 2011.

[31] J. Talairach and P. Tournoux. Co-planar stereotaxic atlas of the human brain. 1988.

[32] A. C. Evans, A. L. Janke, D. L. Collins, and S. Baillet. Brain templates and atlases. *NeuroImage*, 62:911–922, 2012.

[33] McConnell brain imaging center. BrainWeb: simulated brain database. `http://www.bic.mni.mcgill.ca/brainweb/`, February 2016.

[34] Carnegie Mellon University's CCBI. ICBM: International consortium for brain mapping. `http://www.loni.ucla.edu/ICBM/`, February 2016.

[35] LONI. The UCLA laboratory of Neuro Imaging. `http://www.loni.ucla.edu/`, February 2016.

[36] D. Rueckert and J. A. Schnabel. *Biomedical Image Processing*, chapter 5, pages 131–149. Springerverlag Berlin Heidelberg, 2011.

[37] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46:R1–R45, 2001.

[38] J. A. Richards. *Remote Sensing Digital Image Analysis: An Introduction*. Springer, 5th edition, 2012.

[39] A. Goshtasby. *2-D and 3-D Rmage Registration for Medical, Remote Sensing, and Industrial Applications*. Wiley, 2005.

[40] F. Khalifa and G. M. Beache. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, chapter 9, pages 235–264. Oxford : Wiley-Blackwell, 2 edition, 2011.

[41] P. C. Lebby. *Brain Imaging: A Guide for Clinicians*. Oxford University Press, 2013.

[42] L. Hallpike and D. J. Hawkes. Medical image registration: An overview. *Imaging (British Institute of Radiology)*, 14:455–463, 2002.

[43] M. Khader and A. B. Hamza. An entropy-based technique for nonrigid medical image alignment. In *Proceedings of the International Workshop of Combinatorial Image Analysis–IWCIA*, pages 444–455. May 2011.

[44] 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 Transactions on Medical Imaging*, 22(6):730–741, 2003.

[45] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations:
textscApplication to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[46] N. M. Grosland, R. Bafna, and V. A. Magnotta. Automated hexahedral meshing of anatomic structures using deformable registration. *Computing Method in Biomechanical and Biomedical Engineering*, 12(1):35–43, 2009.

[47] S. Gefen, O. Tretiak, and J. Nissanov. Elastic 3D alignment of rat brain histological images. *IEEE Transactions on Medical Imaging*, 22(11):1480–1489, 2003.

[48] F. P. M. Oliveira and J. M. R. S. Tavares. Medical image registration: A review. *Computer Methods in Biomechanics and Biomedical Engineering*, pages 1–21, 2012.

[49] B. Zitova and J. Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:997–1000, 2003.

[50] C. E. Shannon. A mathematical theory of communication. *Bell System Technology*, 27:379–423/623–656, 1948.

[51] E. M. van Rikxoort, I. Isgum, Y. Arzhaeva, M. Staring, S. Klein, M. A. Viergever, J. P. W. Pluim, and B. van Ginneken. Adaptive local multi-atlas segmentation: Application to the heart and the caudate nucleus. *Medical Image Analysis*, 14(1):39–49, 2010.

[52] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

[53] W. R. Crum, T. Hartkens, and D. L. G. Hill. Non-rigid image registration: Theory and practice. *British Journal of Radiology*, (2):S140–53, 2004.

[54] C. Studholme and D. L. G. Hill andD. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern recognition*, 32(1):71–86, 1999.

[55] Y. Keller and A. Averbuch. Multisensor image registration via implicit similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):794–801, 2006.

[56] 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. *Medical Physics*, 35:1407, 2008.

[57] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens. Nonrigid image registration using conditional mutual information. *IEEE Transactions on Medical Imaging*, 29(1):19–29, 2010.

[58] H. Rivaz, Z. Karimaghaloo, V. S. Fonov, and D. L. Collins. Nonrigid registration of ultrasound and MRI using contextual conditioned mutual information. *IEEE Transactions on Medical Imaging*, 33(3):708–725, 2014.

[59] H. Rivaz, Z. Karimaghaloo, and D. L. Collins. Self-similarity weighted mutual information: A new nonrigid image registration metric. *Medical Image Analysis*, 18(2):343–358, 2014.

[60] M. Ghantous, S. Ghosh, and M. Bayoumi. A multi-modal automatic image registration technique based on complex wavelets. In *Proceedings of the IEEE International Conference on Image Processing–ICIP*, pages 173–176, 2009.

[61] Y. S. Kim, J. H. Lee, and J. B. Ra. Multi-sensor image registration based on intensity and edge orientation information. *Pattern Recognition*, 41(11):3356–3365, 2008.

[62] A. Wong, D. A. Clausi, and P. Fieguth. CPOL: Complex phase order likelihood as a similarity measure for MR–CT registration. *Medical Image Analysis*, 14(1):50–57, 2010.

[63] C. Wachinger and N. Navab. Structural image representation for image registration. In *Proceedings of the Computer Vision and Pattern Recognition Workshops–CVPRW*, pages 23–30, 2010.

[64] E. Haber and J. Modersitzki. Intensity gradient based registration and fusion of multi-modal images. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI*, pages 726–733. 2006.

[65] I. Isgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. Ginneken. Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans. *IEEE Transactions on Medical Imaging*, 28:1000–1010, 2009.

[66] A. R. Khan, N. Cherbuin, W. Wen, K. J. Anstey, P. Sachdev, and M. F. Beg. Optimal weights for local multi-atlas fusion using supervised learning and dynamic information (SuperDyn): Validation on hippocampus segmentation. *NeuroImage*, 56:126–139, 2011.

[67] M. Wu, C. Rosano, P. Lopez-Garcia, C. S. Carter, and H. J. Aizenstein. Optimum template selection for atlas-based segmentation. *NeuroImage*, 34(4):1612–1618, 2007.

[68] K. Kasiri, D. A. Clausi, and P. Fieguth. Multi-modal image registration using structural features. In *Proceedings of the International Conference of Engineering in Medicine and Biology Society–EMBC*, pages 5550–5553, 2014.

[69] K. Kasiri, P. Fieguth, and D. A. Clausi. Structural representations for multi-modal image registration based on modified entropy. In *Proceedings of the International Conference on Image Analysis and Recognition–ICIAR*, pages 82–89. Springer, 2015.

[70] K. Kasiri, P. Fieguth, and D. A. Clausi. Cross modality label fusion in multi-atlas segmentation. In *Proceedings of the IEEE International Conference on Image Processing–ICIP*, pages 16–20, 2014.

[71] K. Kasiri, P. Fieguth, and D. A. Clausi. Self-similarity measure for multi-modal image registration. *Accepted for Publication in Proceedings of the IEEE International Conference on Image Processing–ICIP*, 2016.

[72] K. Kasiri, P. Fieguth, and D. A. Clausi. Sorted self-similarity for multi-modal image registration. *Accepted for Publication in Proceedings of the International Conference of Engineering in Medicine and Biology Society–EMBC*, 2016.

[73] 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 Transactions on Medical Imaging*, 25(5):626–639, 2006.

[74] X. Zhuang, S. Arridge, D. J. Hawkes, and S. Ourselin. A nonrigid registration framework using spatially encoded mutual information and free-form deformations. *IEEE Transactions on Medical Imaging*, 30(10):1819–1828, 2011.

[75] C. Wachinger and N. Navab. Entropy and Laplacian images: Structural representations for multi-modal registration. *Medical Image Analysis*, 16(1):1–17, 2012.

[76] D. Rueckert, M. J. Clarkson, D. L. G. Hill, and D. J. Hawkes. Non-rigid registration using higher-order mutual information. In *Medical Imaging*, pages 438–447. International Society for Optics and Photonics, 2000.

[77] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition–CVPR*, volume 2, pages 60–65, 2005.

[78] P. M. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, J. M. Brady, and J. A. Schnabel. Non-local shape descriptor: A new similarity metric for deformable multi-modal registration. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI*, pages 541–548. Springer, 2011.

[79] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel. MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration. *Medical Image Analysis*, 16(7):1423–1435, 2012.

[80] P. Coupé, P. Yger, and C. Barillot. Fast non local means denoising for 3D MR images. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI*, pages 33–40. Springer, 2006.

[81] L. Liu, P. Fieguth, D. A. Clausi, and G. Kuang. Sorted random projections for robust rotation-invariant texture classification. *Pattern Recognition*, 45(6):2405–2418, 2012.

[82] D. De Nigris, D. L. Collins, and T. Arbel. Multi-modal image registration based on gradient orientations of minimal uncertainty. *IEEE Transactions on Medical Imaging*, 31(12):2343–2354, 2012.

[83] Y. Li and R. Verma. Multichannel image registration by feature-based information fusion. *IEEE Transactions on Medical Imaging*, 30(3):707–720, 2011.

[84] L. G. Nyúl, J. K. Udupa, and P. K. Saha. Incorporating a measure of local scale in voxel-based 3-D image registration. *IEEE Transactions on Medical Imaging*, 22(2):228–237, 2003.

[85] P. K. Saha. Tensor scale: A local morphometric parameter with applications to computer vision and image processing. *Computer Vision and Image Understanding*, 99(3):384–413, 2005.

[86] L. Li, M. Rusu, S. Viswanath, G. Penzias, S. Pahwa, J. Gollamudi, and A. Madabhushi. Multi-modality registration via multi-scale textural and spectral embedding

representations. In *Proceedings of the SPIE Medical Imaging*, pages 978446–978446. International Society for Optics and Photonics, 2016.

[87] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (7):710–732, 1992.

[88] A. P. Bradley. Shift-invariance in the discrete wavelet transform. In *Proceedings of VIIth Digital Image Computing: Techniques and Applications*, 2003.

[89] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7):1160–1169, 1985.

[90] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12):2379–2394, 1987.

[91] A. K. Jain, N. K. Ratha, and S. Lakshmanan. Object detection using Gabor filters. *Pattern Recognition*, 30(2):295–309, 1997.

[92] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, pages 14–19, 1990.

[93] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos. DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Medical Image Analysis*, 15(4):622–639, 2011.

[94] J. Liu, B. C. Vemuri, and J. L. Marroquin. Local frequency representations for robust multimodal image registration. *IEEE Transactions on Medical Imaging*, 21(5):462–469, 2002.

[95] D. A. Clausi and M. E. Jernigan. Designing Gabor filters for optimal texture separability. *Pattern Recognition*, 33(11):1835–1849, 2000.

[96] P. Kovesi. Image features from phase congruency. *Videre: Journal of Comput. Vision Research*, 1(3):1–26, 1999.

[97] P. Kovesi. Phase congruency detects corners and edges. In *Proceedings of the Australian Pattern Recognition Society Conference–DICTA*, 2003.

[98] M. C. Morrone and D. C. Burr. Feature detection in human vision: A phase-dependent energy model. volume 235, pages 221–245. The Royal Society, 1988.

[99] S. Venkatesh and R. Owens. An energy feature detection scheme. In *Proceedings of the IEEE International Conference on Image Processing–ICIP*, 1989.

[100] RIRE. Retrospective Image Registration Evaluation. `http://www.insight-journal.org/rire/`, February 2016.

[101] J.M. Fitzpatrick, J. B. West, and C. T. Maurer Jr. Predicting error in rigid-body point-based registration. *IEEE Transactions on Medical Imaging*, 17(5):694–702, 1998.

[102] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[103] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):567–585, 1989.

[104] J. Orchard and R. Mann. Registering a multisensor ensemble of images. *IEEE Transactions on Image Processing*, 19(5):1236–1247, 2010.

[105] C. Sjöberg and A. Ahnesjö. Multi-atlas based segmentation using probabilistic label fusion with adaptive weighting of image similarity measures. *Computer Methods and Programs in Biomedicine*, 110(3):308–319, 2013.

[106] X. Artaechevarria, A. Muñoz-Barrutia, and C. Ortiz de Solorzano. Efficient classifier generation and weighted voting for atlas-based segmentation: Two small steps faster and closer to the combination oracle. In *Proceedings of the Medical Imaging*, pages 69141W–69141W. International Society for Optics and Photonics, 2008.

[107] H. Wang, J. W. Suh, S. Das, J. Pluta, M. Altinay, and P. Yushkevich. Regression-based label fusion for multi-atlas segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition–CVPR*, pages 1113–1120, 2011.

[108] P. Aljabar, R. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert. Classifier selection strategies for label fusion using large atlas databases. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI*, pages 523–531. 2007.

[109] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich. Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):611–623, 2013.

[110] S. Fischer, F. Šroubek, L. Perrinet, R. Redondo, and G. Cristóbal. Self-invertible 2D log-Gabor wavelets. *International Journal of Computer Vision*, 75(2):231–246, 2007.

[111] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[112] L. Zöllei, E. Learned-Miller, E. Grimson, and W. Wells. Efficient population registration of 3D data. In *Computer Vision for Biomedical Image Applications*, pages 291–301. Springer, 2005.