Imperial College of Science, Technology and Medicine
Department of Computing

Patch-based Segmentation
with Spatial Context
for
Medical Image Analysis

Zehan Wang

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

Doctor of Philosophy

of

Imperial College London

September 2014

Abstract

Accurate segmentations in medical imaging form a crucial role in many applications from patient diagnosis to population studies. As the amount of data generated from medical images increases, the ability to perform this task without human intervention becomes ever more desirable. One approach, known broadly as atlas-based segmentation, is to propagate labels from images which have already been manually labelled by clinical experts. Methods using this approach have been shown to be effective in many applications, demonstrating great potential for automatic labelling of large datasets. However, these methods usually require the use of image registration and are dependent on the outcome of the registration. Any registrations errors that occur are also propagated to the segmentation process and are likely to have an adverse effect on segmentation accuracy. Recently, patch-based methods have been shown to allow a relaxation of the required image alignment, whilst achieving similar results. In general, these methods label each voxel of a target image by comparing the image patch centred on the voxel with neighbouring patches from an atlas library and assigning the most likely label according to the closest matches. The main contributions of this thesis focuses around this approach in providing accurate segmentation results whilst minimising the dependency on registration quality. In particular, this thesis proposes a novel kNN patch-based segmentation framework, which utilises both intensity and spatial information, and explore the use of spatial context in a diverse range of applications. The proposed methods extend the potential for patch-based segmentation to tolerate registration errors by redefining the "locality" for patch selection and comparison, whilst also allowing similar looking patches from different anatomical structures to be differentiated. The methods are evaluated on a wide variety of image datasets, ranging from the brain to the knees, demonstrating its potential with results which are competitive to state-of-the-art techniques.

3

Acknowledgements

I would like to thank my supervisor, Daniel Rueckert, for his advice and help during the course of my PhD. I'm grateful for the inspiration and motivation he provided, and for having faith and trust in my abilities.

In addition, I would also like to thank those around me, in the BioMedIA group as well as those who I initially shared an office with, for providing such a great atmosphere and environment during my PhD. Special thanks goes to Kanwal Bhatia, Anil Rao, Georgia Rajamanoharan (Sandbach) and Claire Donoghue for the guidance and support given to me, and for listening to my problems.

I would also like to acknowledge the department's computer support group (CSG) and Imperial ICT's high performance computing (HPC) for providing access to the computer resources and enabling me to run so many experiments.

Finally, I would like to thank my parents for being there and providing backup whenever I need it.

Dedication

I dedicate this to my family and friends, who have supported me, encouraged me and inspired me throughout my life.

"The man who moves a mountain begins by carrying away small stones."
Confucius

"Good work takes time."
Daniel Rueckert

Declaration of Originality

I hereby declare that the work described in this thesis is my own, except where specifically acknowledged

Zehan Wang

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

As the society becomes ever more data driven, the explosion in the amount of available data has yielded substantial advancement in machine learning and data analysis techniques as well as the scientific understanding of the world we live in. In medicine, the use of medical images has enabled better understanding of the diversity of biological processes and medical conditions, driving the advancement of medical knowledge. Further to this, computer aided diagnosis and clinical decision support systems also benefit from the growth in information. Image segmentation plays a crucial role in this, enabling quantitative analysis of the anatomical structures captured in the images. However, the amount of data generated from medical imaging can take a substantial amount of time for clinicians to manually segment, often becoming prohibitive as a regular task. Consequently, automatic methods for performing these tasks are becoming more important for image analysis and are an enabling factor for large population studies in addition to aiding patient-specific diagnosis. Nonetheless, obtaining accurate results is a key priority and still poses a challenge in many medical imaging applications.

In this thesis, the segmentation objective is broad and general, focusing on a specific methodology rather than a specific application. A diverse range of applications are used to test and evaluate the proposed methods, each of which providing different challenges to overcome. Although not exhaustive, these applications are topical and provide a general representation of the problems faced in medical imaging and enable the proposed solutions to be demonstrated

as being relevant in the field. The areas of applications in which the methods are applied and evaluated are as follows:

- **Hippocampus segmentation in brain magnetic resonance (MR) images.** The hippocampus is an important structure in the brain which provides cognitive functionality. Changes in the hippocampus provide a key biomarker for the onset of dementia and other cognitive or neurological problems. One approach in detecting these changes is by identifying strophy in the hippocampal volume over time, which can be calculated after obtaining the hippocampus segmentation [138], [89].

- **Bone and cartilage segmentation for the tibia and femur in knee MR images.** Knee injuries can occur from everyday life, and damage in the joints can build up over time. This commonly culminates in osteoarthritis of the knee, where the cartilage in the joints wear away, leading to painful conditions for patients [46]. Bone and cartilage segmentations can be used to provide biomarkers for disease severity and in diagnosis, treatment planning and prognosis of patients [87], [189].

- **Proximal pelvic limb muscles segmentation in canine leg MR images.** The study of these muscles provide advancement in the knowledge of muscular dystrophies, which are a group of genetic diseases that can occur worldwide and for all races. Currently there is no cure for this condition, but there are various treatment options and ongoing research. Similar genetic conditions also occur in dogs, and they are often used as test subjects for new treatments. Segmentation of the various muscles enable quantitative analysis in clinical research and trials [98], [181].

- **Left ventricle, myocadium and right ventricle segmentation in cardiac MR images.** Cardiovascular diseases are one of the major causes of death in the western world. Segmentations of the various components of the heart enable cardiac function analysis and provide estimations of measures such as the left ventricular mass, end-diastolic volume, ejection fraction, regional wall motion and regional wall thickening. These measures are import in routine clinical applications as well as in cardiovascular research [68], [37].

- **Liver, spleen, pancreas and kidney segmentation in abdominal computed tomography (CT) images.** Segmentation of the abdominal organs enable 3D navigable visualisations which can be used for surgical planning as well as providing quantitative analysis for the organs. A detailed mapping of the abdominal regional provides invaluable input for computer aided diagnosis and laparascopic surgery assistance [88]. In addition, they can be used for radiotherapy planning as well as cancer detection and staging [130], [154].

(a) Brain MRI   (b) Knee MRI   (c) Canine Leg MRI

(d) Cardiac MRI   (e) Abdominal CT
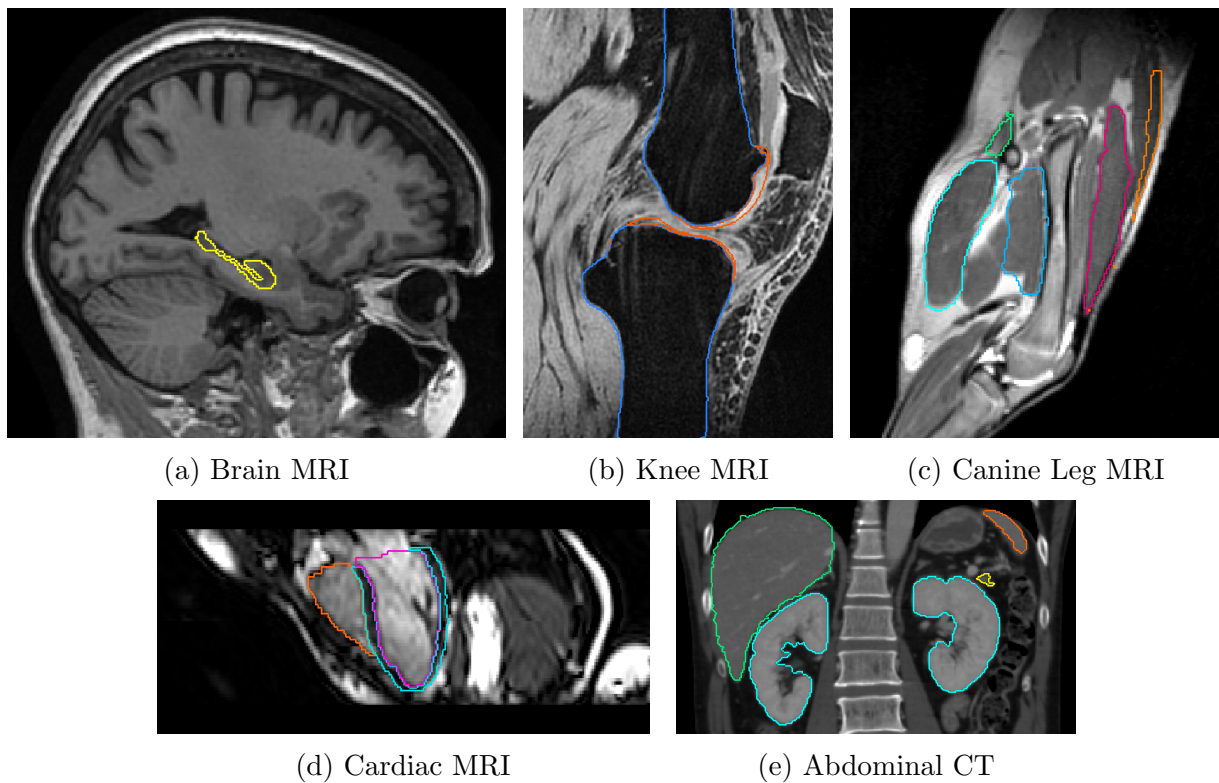
Figure 1.1: A snapshot of the different areas of application in this thesis, with anatomical structures of interest outlined in colour.

1.1   Image Modalities

There are many image modalities that can be employed in medical imaging, ranging from 2D to 4D scans and utilising very different technologies. This thesis focuses on the segmentation of 3D anatomical structures from Magnetic Resonance (MR) and Computed Tomography (CT) images, which are the most general and commonly used modalities for 3D structural analysis and
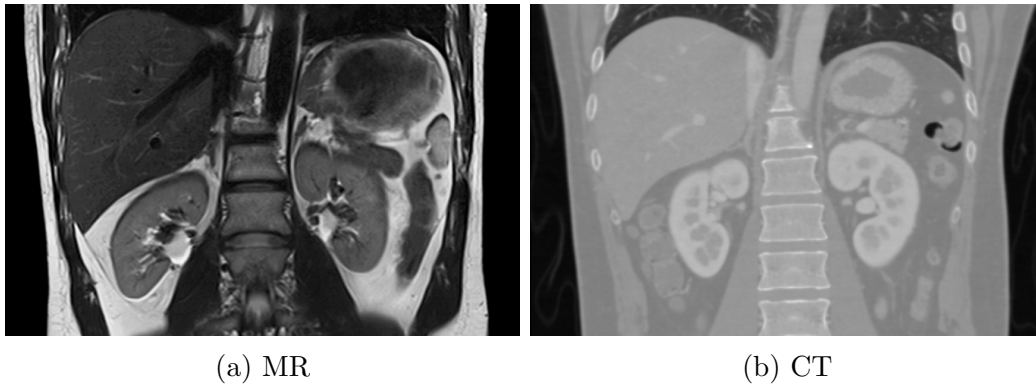
(a) MR   (b) CT

Figure 1.2: Comparing MR and CT images of the abdominal region from different subjects.

segmentation. MR and CT imaging both have their own advantages and limitations depending on the application, but are considered complementary imaging technologies in structural imaging [1].

1.1.1   Magnetic Resonance (MR)

Magnetic resonance imaging exploits a quantum mechanical property of protons, mostly in the form of hydrogen atoms[1], and uses strong magnetic fields and radio waves to excite these atoms whilst measuring the response. The principle behind this relies on detecting changes in the spin of hydrogen atoms, which can be localised, thus building an image of the presence of these atoms [75]. There is a high presence of hydrogen atoms within the body in the form of water and fat, with different quantities present in different tissues, which produce different signals.

Ordinarily, hydrogen atoms spin in random alignment, however a strong magnetic field can align the atoms in their spin axis. When aligned, the atoms can spin in two possible directions, and are at equilibrium with generally no net magnetisation. A radio frequency (RF) pulse can then be applied such that it causes the spin directions to change and align, affecting the spin equilibrium in a phenomenon known as *resonance*. The alignment in the spin of the atoms produces their own magnetic field, and when the RF pulse is turned off, the atoms lose their magnetisation and return to their original states in equilibrium in phenomenon known as

---

[1]Other atoms such as sodium have also been investigated.

relaxation. The relaxation releases energy which is detectable, and this is used to create the MR image. The energy can be released from the atoms in two ways, *spin-lattice* interactions and *spin-spin* interactions. In spin-lattice interactions, there is an energy exchange with the surroundings occurs in T1 time, whilst in spin-spin interactions, the atoms exchange energy between themselves in a more rapid process which occurs in T2 time. Different tissues have different T1 and T2 time constants, thus providing different signals and contrast in the acquired image. In addition, different images can be captured by altering the *echo time* (TE), the time between each pulse and signal measurement, as well as the *repetition time* (TR), the time between two RF pulses. The variation in TE and TR provides three main modalities:

- **T1 weighted** - if the TR and TE is short, tissues with shorter T1 times would have released more energy and will appear brighter than tissues with longer T1 times. Thus, the image intensities is dependent on the T1 time of the different tissues.

- **T2 weighted** - if the TR and TE is long, then all tissues would have time to release their energy, so then differences in signal strength would depend on the level of spin-spin interactions which occur.

- **Proton density weighted** - if a long TR and short TE is used, the strength of the signal depends on the strength of the acquired magnetisation, which is dependent on the density of the hydrogen atoms.

MR scans do not damage tissues or cause biological harm, however any ferromagnetic foreign bodies would cause problems, such as heating and imaging artifacts, due to the strong magnetic field. This means that certain medical devices and implants may not be safe as well as anyone with injuries from shell fragments or other metallic objects. Overall, it is highly suited to general medical imaging, but not emergency situations due to the potential presence of any ferromagnetic foreign bodies and also the time required for a scan, which differs depending on the scanner and application. It is also often used for clinical studies due to the non-ionising nature of the image acquisition process, which allows for repeat scans without causing subjects any harm.

1.1.2   Computed Tomography (CT)

Computed Tomography uses X-rays in a similar principle to standard 2D X-ray imaging, where projecting an X-ray through a body produces a shadow-like remnant when it is blocked by the various tissues within the body. X-rays are absorbed or attenuated differently by different tissues, where hard tissues like the bone attenuate X-rays much more than soft tissues like muscles. In CT scans, an X-ray source and detector[2], positioned on opposite sides of the scanner, acquire a series of cross-sectional scans of the target object at precise rotational steps around a predefined axis. The steps are usually defined at 0.25° to 1°, producing 1440 to 360 images in a full 360° rotation. These scans are then processed using tomographic reconstruction algorithms, such as Feldkamp's filtered back-projection algorithm [63] or more recent variations such as [121], [148] to generate 2D slices which can then be stacked up to create a 3D volume [84].

X-rays are an ionising radiation, and as such can be harmful to the body. However CT scans provide high resolution images which are much faster to acquire than MR images and are not limited by potential metallic objects within the patient, although metallic objects can produce artifacts. CT scans are more widely used in emergency situations than MRI but are also well suited for other applications such as bone injuries and cancer detection since CT scans provides good soft tissue differentiation as well as being able to outline bones inside the body very accurately.

1.1.3   Other Imaging Modalities

In addition to MR and CT, segmentation of images in other modalities are also often necessary, although they have not been used in the scope this thesis. There are a range of modalities which utilise other physical properties to generate 3D images and may have more application-specific uses. These vary from the use of high frequency sound waves such as in Ultrasound to using radioactive isotopes which emit gamma rays such as PET and SPECT.

---

[2]Some of the more recently developed scanners feature multiple pairs of x-ray sources and detectors

Ultrasound is often associated with imaging the fetus in pregnant women, but also has broader uses such as for abdominal organs, heart, breasts and other tissues in the body. The process is fast, non-ionising and relatively cheap compared to other modalities, however the images produced also tends to be much noisier and having lower quality compared to other modalities.

PET and SPECT, on the other hand, involves exposure to small doses of ionising radiation emitted by the radioactive isotopes and are much more expensive and time consuming for each acquisition. However, they provide images which are primarily used for functional analysis of organs such as the heart or the brain, in addition to cancer detection, rather than for structural analysis where MR and CT are better suited. In some cases it is possible to combine modalities, such as with PET-CT, PET-MRI or SPECT-CT images, which provides an image of both the functional and the structural aspects of the anatomy in a spatially coherent form. This can be performed by specialised hybrid machines [26], [54], [93], or they can fused from separate scans post-acquisition [188].

There are also additional modalities developed more recently which utilise other physical properties and effects to generate the image. These include photoacoustic (or optoacoustic) imaging [182] where the photoacoustic effect is used, combining ultrasound and electromagnet waves (particularly light), and tactile imaging, which measures mechanical pressure responses in soft tissues [60]. These are not yet widely used in clinical settings, but the development of new modalities for imaging are an active research area in the field of medical physics.

1.2 Challenges in Image Segmentation

The task of image segmentation is essentially a labelling task where each pixel or voxel is assigned a label according to its characteristics within the image. The label is usually predefined to have some semantic meaning such as to denote an object, and the process could be carried out completely manually by a human expert, semi-automatically with some human input, or automatically without any human input. The result of image segmentation is an image partitioned into regions where the pixels or voxels of each region have some common property
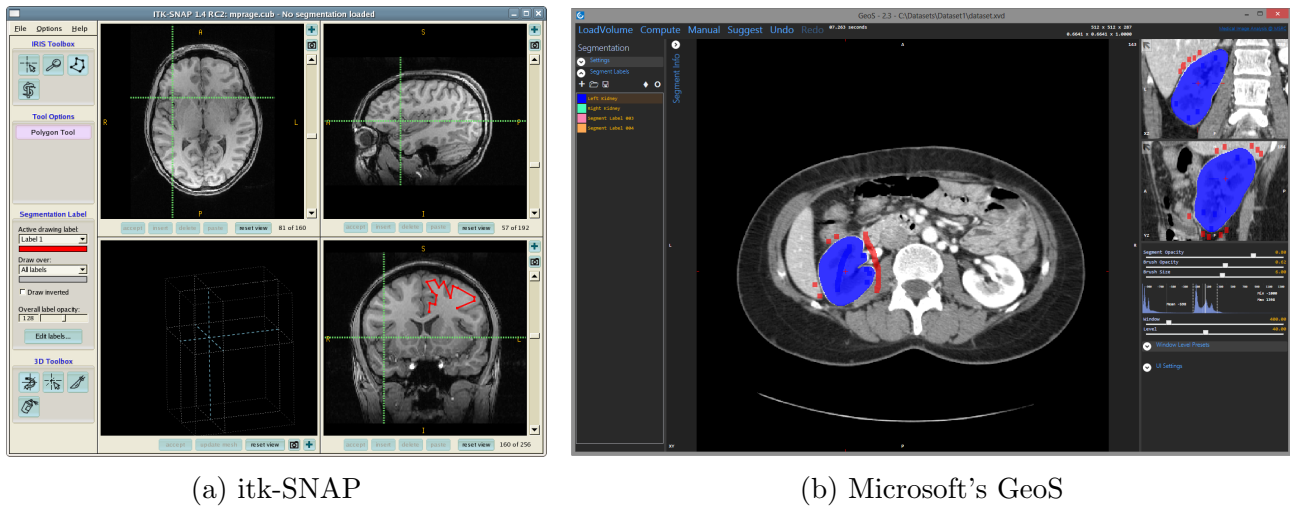
(a) itk-SNAP	(b) Microsoft's GeoS

Figure 1.3: Examples of GUI tools for performing manual segmentations. With itk-SNAP, the user places marks to create polygons to delineate relevant structures slice-by-slice throughout the volume. With GeoS, the user can place seeds in the image for different labels and then run an algorithm to resolve the non-seeded regions.

or meaning. In medical imaging, segmentation is often performed to impose structure on the data, identifying different tissue types and enabling quantitative analysis of what is captured in the image to be performed more easily.

Manual and semi-automatic segmentations generally involve the use of a graphical user interface where the user can either label individual voxels or place seeds within the image for each label and then running an algorithm, such as region growing, to resolve the non-seeded regions (see Figure 1.3). This can be time consuming for the user, and may require application specific knowledge. Furthermore, depending on the repetitive nature and difficulty of the task, each image may need to be segmented several times in order to minimise human errors. As such, this approach may not be scalable to large datasets, but is still useful in generating training data for automatic segmentation methods. Human error is often hard to completely avoid, but it can be taken into account for to determine the reliability of the manual segmentations. When a single rater is used to segment an image multiple times, the segmentation will often differ in some small way each time. This difference is known as *intra-observer error*. When multiple experts or raters are used to minimise the systematic error of an individual, the segmentation may also differ between the raters. This difference is known as *inter-observer error*.

The aim of automatic segmentation algorithms is to overcome the human resource requirement

of manual or semi-automatic segmentation tasks whilst also reducing potential human error and improving consistency. This would empower large-scale scientific studies as well as opening doors for other applications in computer vision such as object detection and recognition. The development of these methods is vital to many research projects, and would have a major impact if they can be transferred into clinical settings. Although a lot work has already been carried out in this area, resulting in significant advancement, there are still many limitations in the state-of-the art methods.

In its simplest form, image segmentation can be a binary labelling tasking, separating the background pixels from an object of interest (or the foreground), but the complexity of the task is very much dependent on the application. For example, a simple case could be when the background is highly dissimilar from the foreground, such as a dark object in front of a light background, however in practice it is rarely this straight forward. In many applications, there may not be much visual dissimilarity between the foreground and the background, and there might also be multiple objects of interest which all need to be correctly labelled as well. In addition, many factors can hinder the abilities of both human experts and computer algorithms alike, such as the presence of noise and artifacts in the image acquisition, which affects overall image quality. There are also further specific factors which can affect different modalities of imaging. For example, occlusion, shadows and reflections are present in visible spectrum images, whereas MR and CT images are generally free from these issues but poses other problems of its own which are often application dependent.

1.2.1   Challenges in Medical Imaging

One of the main challenges in medical image segmentation is due to the fact that similar tissues can take a range of intensities which can lead to ambiguity in tissue classification. One aspect of MRI which makes it particularly challenging for segmentation is that the range of intensities also vary from patient to patient and scanner to scanner, and there is often intensity inhomogeneity within the same tissues. As a result, there is rarely a linear mapping of intensities to tissues between patients or even within the same image. In CT scans, images outputs are standardised

using the Hounsfield scale, a quantitative measure for radiodensity, however there are often overlaps in the intensities of different tissues which can lead to ambiguity (which is also the case in MR images). Furthermore, the shapes of many anatomical structures are often complex, with large variations between subjects and may be hard to delineate or model.

For many medical imaging applications, there are some common artifacts that can arise from the acquisition process. These include motion artifacts, noise, streaks and partial volume effects, as well as other artifacts which are well documented in [135] and [159]. Both MR and CT images can suffer from artifacts but the extent of these artifacts occurring is somewhat dependent on the application as well as the scanner and modality used. For example, motion artifacts are more common in cardiac and fetal images than in brain images, and CT scans tend to have fewer motion artifacts due to the faster image acquisition rate. These artifacts affect the quality of the images, sometimes causing ambiguity when identifying different tissues and can have an adverse affect on segmentation results if they cannot be corrected or accounted for. Figure 1.4 presents an example of some of the artifacts and challenges in medical images.

Further to challenges within the image data, another challenge is in acquiring sufficient training data to effectively model or represent the full range of anatomical variations in addition to the variation in the image with regards to quality and how the anatomy is captured. Unlike with digital photos, which are more abundant, medical images are both more time consuming and expensive to acquire, and may be under legal and ethical agreements governing data use. Additionally, obtaining scans of "normal control subjects" is also difficult, as healthy patients normally don't have any reasons to visit the hospital for a scan. Furthermore, labelling 3D volumetric images can be a harder and more time consuming task than for 2D images, which adds to the limitations in training data.

There are also application-specific challenges which present particular problems which must be overcome, making it difficult for generic methods to work well in every case without some customisation and tuning. A major issue behind many applications is the sensitivity to the presence of different pathologies, and the capability to detect and segment the pathology for decision making. In many cases, grand challenges and challenge workshops have been created at
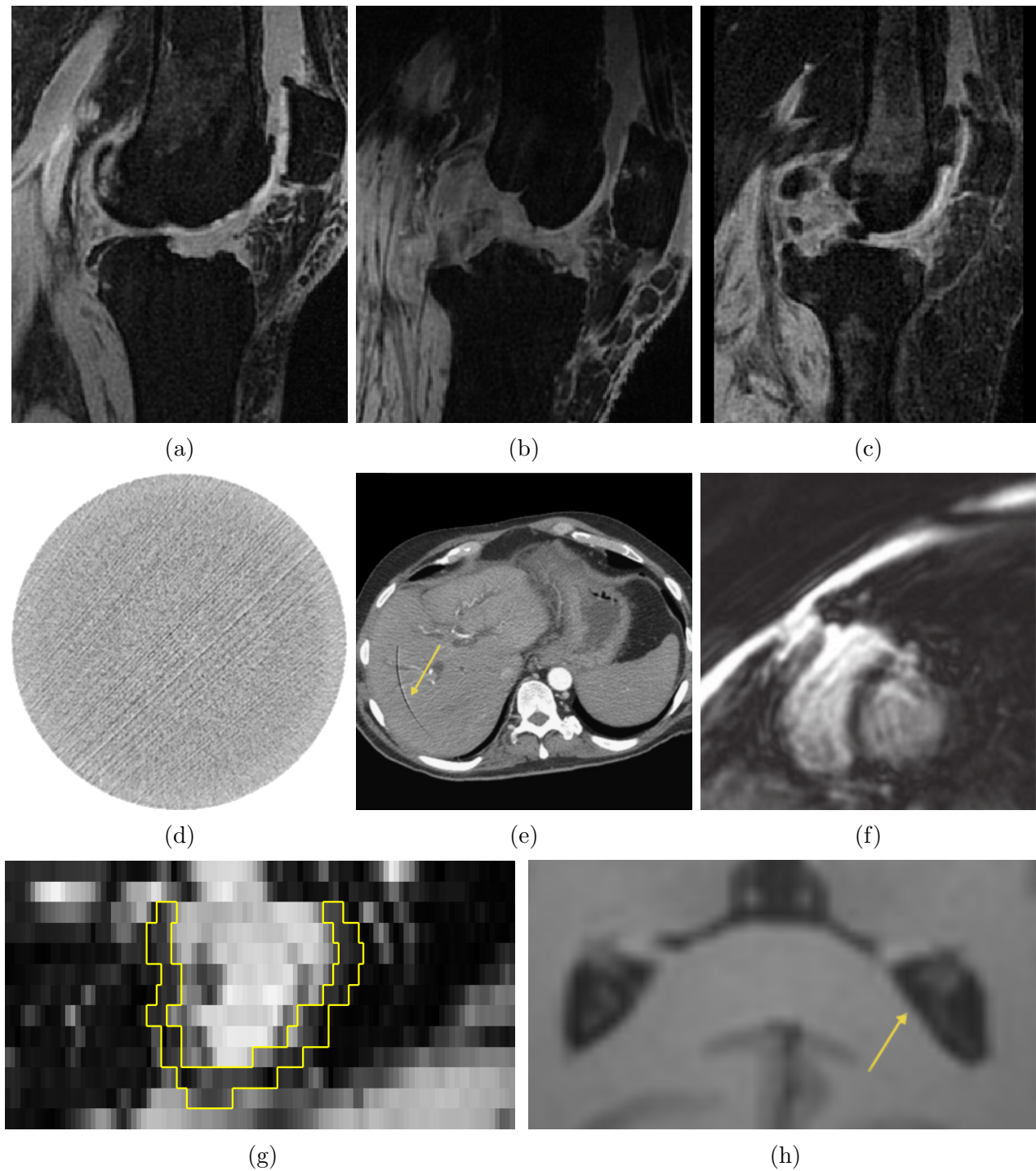
Figure 1.4: Examples of some of the artifacts and potential challenges. In (a), (b) and (c), tissue inhomogeneities can be seen, in addition to different levels of noise, different intensities for similar tissues and overlapping intensities between different tissues. (d) shows a streaks and (e) shows rings in CT images. Motion artifacts produce ghosting and blurring of the image in (f) and can also cause discontinuities between slices as seen in (g). Partial volume effects can lead to blurry boundaries such as in (h).

conferences to objectively compare and evaluate methods for specific applications. For example, these include challenges for brain image segmentation [119], knee cartilage segmentation [82], cardiac image segmentation [136]. In addition, there have been general segmentation workshops [102], [101] with specific applications such as the segmentation of brain substuctures, canine leg muscles and cardiac structures.

Challenges in Brain MRI

In many cases, automated tissue classification such as brain and non-brain, or grey matter, white matter and cerebrospinal fluid can be already be quite reliably performed [158], however the segmentation of anatomical structures within the brain can be much more challenging. This is largely due to many of the distinct structures exhibiting very similar intensities, and some structures are also composed of multiple tissue types. Anatomical variability exists in many of the structures, such that relying on image alignment and spatial properties alone may be insufficient for an accurate segmentation. For the hippocampus, there is variation in the shape and size between subjects and the exact definition of its boundaries can be ambiguous. Even between human experts, there are often variations in the delineation of the hippocampus for the same subject. In the past, there have been different protocols used for manual segmentations, and only recently has a standardised protocol been proposed [28].

Challenges in Knee MRI

The knee is made of many tissues and structures which exhibits variations in their volume and spatial distributions. Additionally, many of the structures have a globally cylindrical shape and there can be significant pose variation of the knee, which altogether provides challenging conditions for many tasks, not just segmentation. In recent experiments, it was discovered that 4-5% of all affine registrations fail for the knee [57]. For segmentation, the bones and the cartilage presents different challenges, although the cartilage is often considered the more challenging. There can be significant intensity variation and inhomogeneities in the bones, particularly for diseased patients. For the cartilage, the thickness and spatial distribution is

perhaps the most challenging aspect. In severe cases of osteoarthritis, the cartilage is thinner and much harder to identify than in healthy subjects, and there can be osteophytes (bony spurs or growths) on the bones which alters the shape and appearance of the knee joint.

Challenges in Canine Leg MRI

The application of segmentation methods to canine leg MRI is not as well studied as other other areas, presenting new challenges for existing methods. There is significant variation in the sizes of dogs and their legs, as well as their pose, which can be problematic for image alignment. Additionally, muscle tissues appear textured and do not have homogeneous intensities, whilst having very similar overall appearances for different muscle groups. Furthermore, the individual muscles can be interconnected and may not be clearly separable when they are in contact with each other. This can make it particularly difficult to correctly identify and label each muscle.

Challenges in Cardiac MRI

The acquisition of clear and detailed images of the heart is a challenging task for radiologists, largely due to the heart being in continuous motion. Additionally, patient motion and respiration further complicates the problem. As a result, the quality of the resulting images are often limited, particularly when compared to images of more static organs. Furthermore, there is also great variability between subjects in the shape, orientation and appearance of the heart and its substructures. In addition to appearance variations caused by the MR scan, pathological changes like infarction can cause local variations in contrast, and blood flow artifacts can create intensity inhomogeneities within some of the structures.

Challenges in Abdominal CT Images

In abdominal CT images, the contrast and quality are often good, however the complexity in the layout of the organs presents a significant challenge. There are a large number of organs in the abdominal region, which are deformable as well as being free to move individually. This makes

obtaining overall image alignment of the organs particularly challenging. Some organs, such as the liver, are large and easily identifiable, but others, such as the pancreas, are smaller and more difficult to locate. In addition, there is significant variation in the sizes of the patients and their organs, as well as variations in the spatial proximity of the organs between patients. The segmentation of the pancreas is a particularly difficult, due to its shape and spatial variation. Unlike other organs, such as the liver and the spleen, which are generally predictable in which side of the body they occupy, the location of the pancreas is much more variable and is not fixed to either the left or the right side of the body. Furthermore, it has a complex shape with many folds and bends along its surface, and can be interwoven between other organs in the abdomen.

1.2.2   Computational Challenges

The development of general methods or frameworks for image segmentation is an active and hotly contested area in research, however there is usually some level of customisation required for each separate application. In general, the ability to be adaptable and versatile, whilst maintaining a high level of accuracy is one of the biggest computational challenges for any automatic segmentation approach.

One aspect of using image data as a whole is sometimes known as the "curse of dimensionality"[3], and refers to the problems associated with using high dimensional data like images, but which do not exist in low dimensions. In most machine learning applications, an individual image is treated as a single data point in high dimensional space and at its most naive representation, an image with n pixels is a point that lives in n dimensional space. For example, an image with a resolution of $800{\times}600$ pixels is a point in $800{\times}600 = 480,000$ dimensional space and when looking at a 3D medical image that is just 128 voxels in each direction, the dimensional space becomes $128{\times}128{\times}128 = 2,097,152$. Analysis of high dimensional data poses many problems as the volume of the space grows exponentially when dimensionality is increased linearly, resulting in the available data becoming sparse. This then can mean that some of the

---

[3]a phrase coined by Richard E. Bellman [22]

| (a) | (b) | (c) |
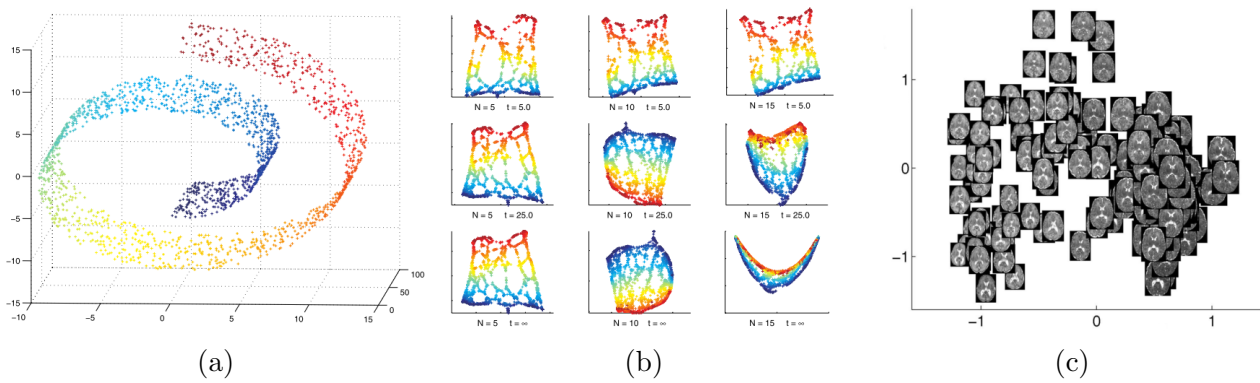
Figure 1.5: Data in high dimensional spaces can exist on a lower dimensional manifold. Dimensional reduction techniques aim to recover the data in a lower dimensional representation. As a synthetic 3D toy example, data may exist on a swiss roll, as shown in (a), which could be unrolled into a 2D flat surface. However, the results can differ depending on the choice of parameters, as seen in (b) with Laplacian Eigenmaps [21]. (c) shows the results of "unrolling" the images of brains from their high dimensional image representation after applying ISOMAP [166].

most common operations on this space such as search and data comparison are much more complicated to compute than in a lower dimensional space, and also prevents existing data organisation strategies from other areas of computer science from working efficiently.

The representation of images, as a whole and in terms of the objects captured within, is another open question and topic of research for many scientists. Objects of interest within images are often said to exist on a manifold where the intrinsic dimension of the manifold is much lower than the high dimensional space of the image, however extracting this manifold is non-trivial. Many dimensionality reduction techniques exists [109], but it is hard to verify what the true dimensional representation of any object looks like as it is generally not known. Instead, methods aim to approximate a manifold using the training data, and the outcomes are subject to the choice of objective or distance functions as well as the selected features of the images.

Overall, high dimensional data such as images presents a large number of complexities which can affect how well any algorithm can perform. These impact not just the computational speed and efficiency, but also what useful information can be extracted and how they can be used. Overcoming these challenges is a worthwhile endeavour and of interest for many applications, not just in medical imaging, but in extending human understanding of the natural world through computer vision and data analysis.

1.3   Contributions

The primary objective of this thesis is the formulation of new methods to obtain accurate automatic segmentation results in medical images. In particular, the focus is on a subcategory of label propagation methods called *patch-based segmentation* which propagates labels from existing labelled images to unlabelled images based on local image comparisons. This approach has been demonstrated to have potential in achieving highly accurate segmentation results [47], [142]. These methods utilise principles from machine learning, computer vision and image processing, and are currently of interest for both computer scientists and clinical researchers.

Traditionally, segmentation approaches which propagates labels from one image to another require accurate alignment between the images, however the process of fully registering images to the same coordinate space is complex and usually non-linear. Registration errors can often occur, although the amount and impact of the errors differ between applications, in some cases insignificant and other times more severe. In images where there is high anatomical variability, registration errors are more common and have a greater impact on the outcome of the segmentation process. Patch-based methods relax the dependency on registration outcome, often allowing the use of results after just a linear registration process, without requiring images to be fully aligned. In general, a patch-based approach labels each voxel of an image by looking at its local patch and comparing with similar local patches from an atlas library. In this thesis, the concept of patch-based segmentation is expanded upon with the inclusion of spatial context and a new framework for segmentation using a k nearest neighbour (KNN) perspective on the locality of patches. The key contributions are as follows:

- **A new patch-based segmentation framework which not only uses intensity information, but also spatial context.** The patch-based segmentation approach is reformulated to use the k nearest patches from a global perspective with regards to both intensity and spatial information. This removes the fixed window size limit on the locality of patches used for comparison in the standard patch-based methods, and enabling the dependency on registration outcome to be further reduced.

- **Development of different approaches to providing spatial context.** Three different approaches to spatial context are proposed and evaluated, from initially using spatial coordinates to using relative distances between structures, and then using geodesic distances within the image instead of Euclidean distances. The approaches allow the need for alignment between images to be decoupled in different ways, enabling the application of patch-based segmentation to be more general and versatile.

- **Application of proposed methods to a large variety of image datasets.** The proposed methods are evaluated on segmenting several different anatomical structures in different image types. These range from the hippocampus in brain MRI to tissues in cardiac MRI as well as major organs in abdominal CT scans. The methods are also applied to open challenge datasets such as the MICCAI SKI10 Grand Challenge, which involve bone and cartilage segmentations in knee MRI, and segmenting pelvic muscles in canine leg MRI as part of the SATA MICCAI challenge workshop.

1.4 Overview of Thesis

We start with a background chapter 2 detailing advances in automatic segmentation and the current state-of-the-art in medical imaging as well as some of the computational techniques and algorithms used. This is followed by three chapters where a novel kNN patch-based segmentation framework which uses spatial context is first proposed in Chapter 3 and then extended in Chapters 4 and 5 with a multi-resolution approach and different approaches to spatial context. In these chapters, the proposed methods are evaluated on a range of different applications, from the segmentation of the hippocampus in brain MRI to bone and cartilage segmentation in knee MRI as well as multi-organ segmentation in abdominal CT scans. Finally, we conclude with a summary of this thesis and potential future work in Chapter 6.

Chapter 2

Background

In this chapter, we provide an overview of some of the advances in image segmentation relevant to the work in this thesis. In addition, we also review the computational techniques and methods which are used in this thesis. Recent segmentation methods combine techniques from a range of computer science disciplines, from machine learning to signal and image processing. The task of image segmentation can be viewed from multiple viewpoints and often there are several routes to achieving the same goal.

We start with an overview of some of the pre-processing methods used in medical image analysis, followed by image registration, a vital process which is required for many segmentation methods as well other aspects of image analysis. Next, we examine the family of methods categorised under atlas-based segmentation and provide an overview of some other related approaches. Finally, we finish the chapter with some of the relevant computational techniques that are commonly used in image segmentation and data analysis.

2.1 Pre-processing Methods

Depending on the intended task, medical images often require some initial processing to reduce the effects of artifacts and to standardise image outputs. These are considered as pre-processing

tasks and treated as separate problems to any image analysis, but they can play a significant part in the final outcome. For all applications in this thesis, one or more of these pre-processing methods are used prior to segmentation.

2.1.1   Bias Correction

(a) Original image   (b) Bias corrected image   (c) Bias Field   (d) Difference
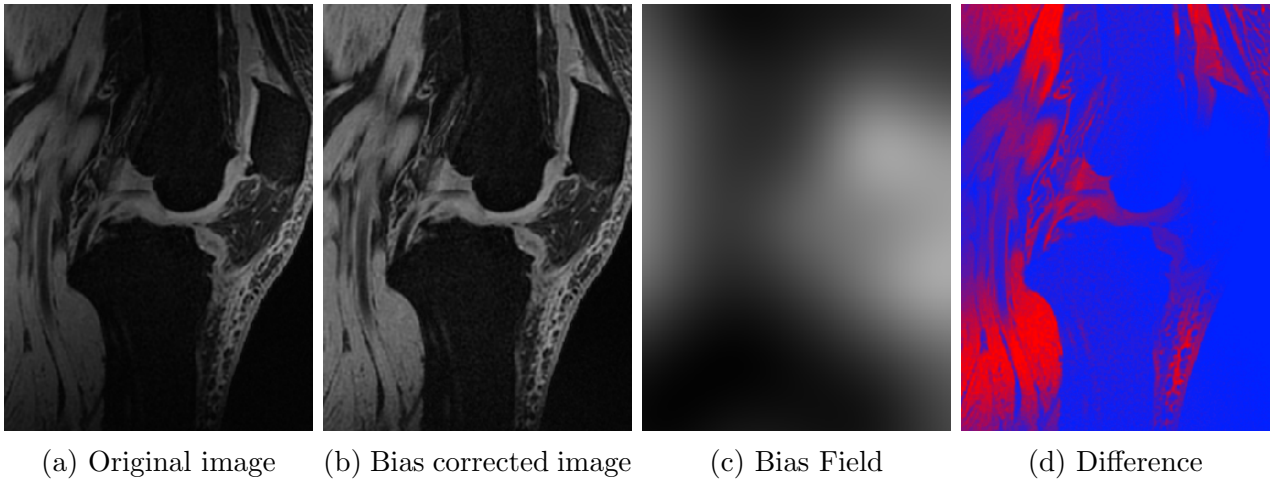
Figure 2.1: An example of the bias and intensity inhomogeneity present in MR images illustrated using an image of the knee. (a) shows the original uncorrected image, whilst (b) shows the result of bias correction using [172]. (c) shows the detected bias field and (d) is an image of the absolute differences, where bright red indicates large differences and blue indicates no difference.

A common artifact from MRI is the presence of a bias field, since the strength of the magnetic field within a scanned body will vary according to the surrounding tissues. Electromagnetic interaction with the scanned body, the positioning within the scanner and limitations in the equipment are primary factors in creating the bias field [156]. This causes intensity inhomogeneity, where the contrast and range of intensities in the image are uneven, with the effect varying by location. This often appears as a smooth gradient across the image, as shown in figure 2.1. It is not always noticeable at first glance, but the bias affects many automated image analysis tasks such as image registration and segmentation. In particular it alters the contrast and changes the intensity ranges for the same tissues within the image. The bias field can differ between subjects, thus it can also distort the outcome of similarity measures between different images.

One commonly used bias correction method is the nonparametric nonuniformity normalization

(N3) method presented in [157], which is fully automatic and applicable without any prior knowledge. The N3 method formulates the problem of bias correction in a fashion that is common to many other methods, modelling a given image v as:

$$v(\mathbf{x}) = u(\mathbf{x})b(\mathbf{x}) + n(\mathbf{x}) \tag{2.1}$$

for a pixel location $\mathbf{x}$ in the image, with u being the uncorrupted image, b the bias field and n the noise. The noise is usually assumed to be independent, which enables the bias correction algorithm to operate whilst ignoring the noise component. Assuming a noise-free scenario, this can then be reformulated by taking the logarithm (let $\hat{u} = \log u$, etc) as

$$\hat{v}(\mathbf{x}) = \hat{u}(\mathbf{x}) + \hat{b}(\mathbf{x}) \tag{2.2}$$

From a signal processing perspective, the bias field reduces the high frequency components in the intensity distribution of $\hat{u}$, thus the approach for bias field correction is to restore this content [157]. An iterative approach can then be used to estimate the bias field $\hat{b}$ and obtain an estimate of the uncorrupted image $\hat{u}$ in the ith iteration, such that the high frequency content of the distribution of intensities in $\hat{u}$ are maximised. The approach taken in [157] is:

$$\hat{u}^i = \hat{v} - \hat{b}^i \tag{2.3}$$

$$= \hat{v} - S\{\hat{v} - E[\hat{u}|\hat{u}^{i-1}]\} \tag{2.4}$$

where $S\{.\}$ is a smoothing operator, usually a B-spline approximator, and $E[\hat{u}|\hat{u}^{i-1}]$ is the expected value of the true image given the current estimate of the corrected image. The bias field is usually slowly and smoothly varying, hence the use of a smoothing operator $S\{.\}$ in modelling it. Initially, $\hat{u}^0 = \hat{v}$ and the initial bias estimate $\hat{b}^0$ is the typically set to 0.

In [172], a different B-spline approximator is used to that of [157] and instead of the total

bias field, the *residual bias field* is estimated at each iteration, which modifies the optimisation scheme for faster convergence. Although there are some recent variations of this approach such as [172], there have been relatively few improvements since. Reviews of this and other bias correction approaches are presented in [177] and [19].

2.1.2   Denoising

The presence of noise is a common and general problem with all modalities of digital imaging, from MRI to normal optical images. The term is derived from signal processing, where it is defined as an unwanted modification to a signal which causes it to degrade and suffer, and is also sometimes used to refer to a signal which carries no useful information. In images, noise can be caused by a number of factors, but primarily is an aspect of the detectors and sensors used and is related to their sensitivity and any signal amplifications used. It most visibly manifests as speckles or grains on the image, which can interrupt edges as well as create false edges and alter the homogeneity of tissue intensities.

Denoising, which is sometimes referred to as image restoration, is the process of removing the noisy signals whilst maintaining the relevant aspects of the image. However, it can be hard to determine exactly what is noise, and also the noise level in any image is generally unknown. Overestimating the level of noise may smooth the image and remove details such as the edges, whilst underestimating would not reduce enough of the noise. In addition, incorrect denoising may introduce additional artifacts and distortions in the image. In general, denoising is a hard challenge and there have been numerous approaches proposed; notable developments include the bilateral filter [169], anisotropic diffusion [132], Total Variation (TV) [145], nonlocal means (NLM) [33] and block-matching [51]. In addition, there have been recent optimisation for specific modalities such as MR [48] or CT [115]. A review of some recent methods are provided in [35].

In most methods, a similar model for image formulation as bias correction (see 2.1) is commonly used, however in contrast to bias correction, the focus is on estimating the noise and recovering

| (a) Original image | (b) Denoised using TV denoising | (c) Denoised using NLM |
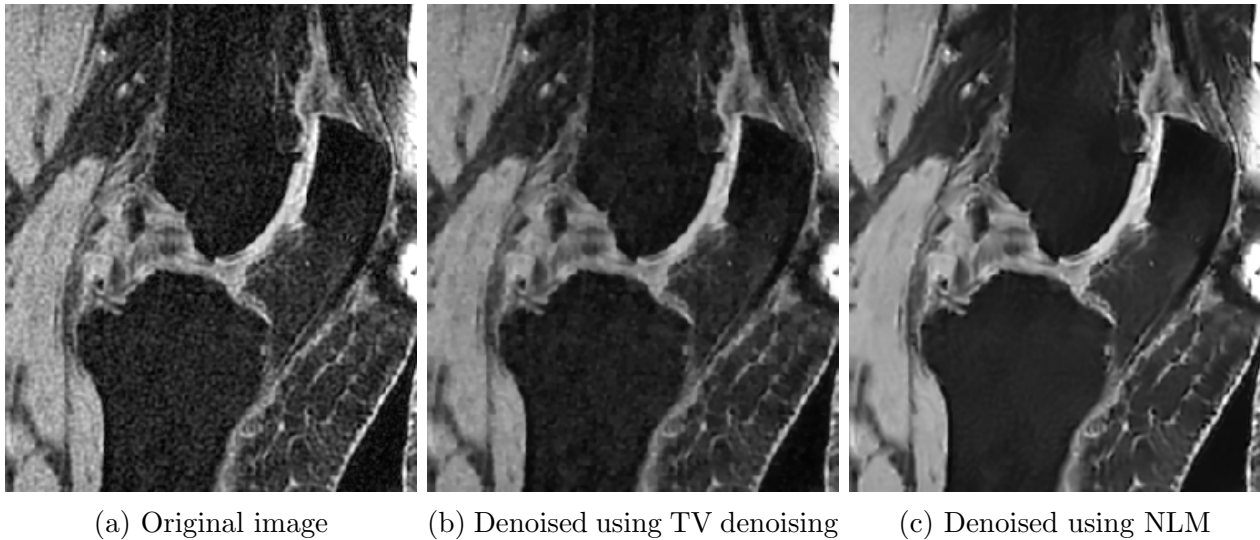
Figure 2.2: An example of noise present in knee MRI, shown in (a), with the result of applying TV denoising shown in (b) and nonlocal means in (c).

the image solely without noise and ignoring the bias field component. So for an image v, we have

$$v(\mathbf{x}) = u(\mathbf{x}) + n(\mathbf{x}) \tag{2.5}$$

where the aim is to recover the uncorrupted image u whilst estimating the noise n. Individual denoising methods will differ in the model chosen to describe the distribution of noise and the smoothing method used to recover the image u. An example of applying denoising to a MRI of the knee is shown in Figure 2.2.

Total Variation (TV) Denoising

Total Variation (TV) denoising was introduced in [145] based on the principle that noisy images contain excessive and spurious details, and have high total variation. From a signal processing perspective, the total variation is a measure of a signal defined by the integral of absolute gradient of the signal. Reducing the total variation whilst maintaining a close representation of the original signal, would then remove the unwanted detail (noise) from the signal. In TV

denoising, the denoised image u is obtained by solving a minimisation problem of the form:

$$\arg\min_u TV(u) + \lambda \sum |v(\mathbf{x}) - u(\mathbf{x})|^2 \qquad (2.6)$$

where v is the original image and $TV(u)$ is the total variation, i.e. $TV(u) = \sum |\nabla(u)|$. The parameter λ is related to the degree of filtering, with more regularity in the image as λ gets smaller. A corresponding Euler-Lagrange equation can be derived as

$$-\nabla \cdot \left(\frac{\nabla(u)}{|\nabla(u)|} \right) + 2\lambda(u(\mathbf{x}) - v(\mathbf{x})) = 0 \qquad (2.7)$$

which can then be solved using gradient descent [145]. The minimisation is also strictly convex, which enables a range of alternative techniques from convex optimisation to be used to solve it. More recent adaptions using alternative solutions include those of [39], [38] and [72].

Nonlocal Means (NLM)

The nonlocal means denoising method [33] does not assume any regularity in the images, which smoothing filters provide, and instead tries to exploit redundancies in the image at when comparing small windows of the image. The concept forms the basis for patch-based segmentation methods [47], [142] developed later. In the NLM approach, the value for each pixel $\mathbf{x}$ is estimated based on its Gaussian neighbourhood and comparing it with other pixels and their Gaussian neighbourhoods. This neighbourhood is essentially a local patch, and the weighted average value of pixels with the most similar patches are used to provide the estimate of the denoised image. Thus, we have for an image, v and pixels $\mathbf{x}$, $\mathbf{y}$ within the image v,

$$NLM(v)(\mathbf{x}) = \sum_{\mathbf{y}} w(\mathbf{x}, \mathbf{y})v(\mathbf{y}) \qquad (2.8)$$

where the weighting $w(\mathbf{x}, \mathbf{y})$ can be adapted for different noise distributions such as Rician or
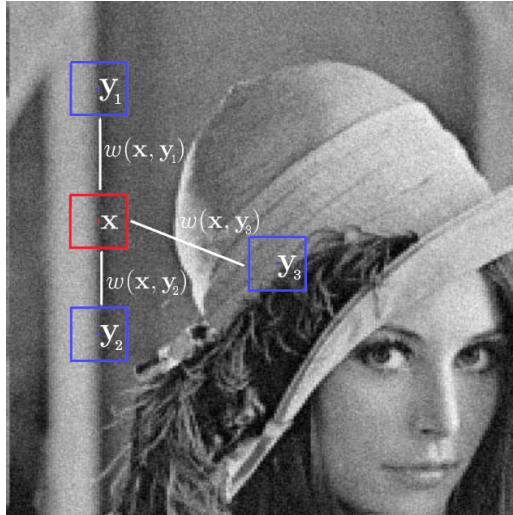
Figure 2.3: Nonlocal means: each pixel $\mathbf{x}$ is assigned a weighted average of values from other pixels $\mathbf{y}$ according to the similarity of their neighbourhoods. In this example, $\mathbf{y}_1$ and $\mathbf{y}_2$ would receive higher weights than $\mathbf{y}_3$ since their neighbourhoods more closely resembles that of $\mathbf{x}$.

Gaussian noise [48]. In the Gaussian case, it is defined as:

$$w(\mathbf{x}, \mathbf{y}) = \frac{1}{Z(\mathbf{x})} e^{\frac{-||\mathcal{N}_{\mathbf{x}} - \mathcal{N}_{\mathbf{y}}||_2^2}{h^2}} \tag{2.9}$$

where $\mathcal{N}_{\mathbf{x}}$ is the neighbourhood for $\mathbf{x}$ and $Z(\mathbf{x})$ is a normalisation factor $Z(\mathbf{x}) = \sum_{\mathbf{y}} e^{\frac{-||\mathcal{N}_{\mathbf{x}} - \mathcal{N}_{\mathbf{y}}||_2^2}{h^2}}$ and h is a parameter which controls the decay of the exponential function and therefore the decay of the weights.

2.1.3 Intensity Normalisation

In images, it is often desirable to have similar intensities for similar tissues so that analysis and comparisons of tissues can be performed more easily. Intensity normalisation is the process of standardising the range of intensity outputs between a collection of images in order to achieve this, and is particularly important for any image processes using intensity differences, such as any form of the L_p norm. This is often needed for MRI but not CT images, since the intensity outputs are already standardised as part of the acquisition protocol for CT scans but the complexity of MR acquisitions prohibits this. There are no standardised scales or interpretations for the intensities in MR images, and the intensities can differ for every scan.

Although there have been attempts at MR calibration such that intensities can be standardised during the acquisition [59], the calibration process is cumbersome and it is often more attractive to normalise the intensities as a post acquisition process.

There are several approaches for intensity normalisation; [123], [44], [110] provide a review of some popular methods. Naively, normalisation can be achieved by linearly scaling the intensities to match a given intensity range. Given an image u with minimum and maximum intensities u_{min} and u_{max} and a target intensity range $[\tau_{min}, \tau_{max}]$, the normalised image u' is obtained for each pixel $\mathbf{x}$ using the following:

$$u'(\mathbf{x}) = (u(\mathbf{x}) - u_{min})\frac{\tau_{max} - \tau_{min}}{u_{max} - u_{min}} + \tau_{min} \qquad (2.10)$$

Linear scaling is often insufficient, and instead most methods look at matching some aspect of the intensity distribution between images, not just the intensity range. In [123], normalisation is performed by matching to a target histogram profile and comparing landmarks on the histogram such as the quartiles or percentiles and rescaling the intensities (between landmarks) in the image accordingly to obtain the best match. The tails of the histograms are ignored so as not to be biased by outliers. Some region of interest (ROI) masks may be required to remove the influence of unwanted parts of the image. Thresholding and image registration would then be precursors to any intensity normalisation, however any errors from these tasks then affects the size of the ROI masks and reduces the effectiveness of the normalisation process. This is one of the challenges for intensity normalisation methods and remains an open problem.

2.2 Image Registration

Image registration is the process by which images are spatially aligned to the same space of reference or coordinate system. Registration is often required to allow comparison of data from multiple sources, enabling meaningful measurements to be made between different data. The goal of registration is to produce a mapping from one coordinate space (the source) to another

(the target or reference), which allows an image to be transformed from one space to the other. Image registration has been an key area of research in medical imaging for several decades, with numerous methods and frameworks. There are also a number of reviews and surveys, such as [32], [113], [76], [197] to mention just a few.

Formally, in 3D, registration produces a transformation $\mathbf{T} : (x_s, y_s, z_s) \rightarrow (x_t, y_t, z_t)$ for the source image I_s to the target image I_t such that it aligns I_s with I_t and maximises the similarity between the two images. To do this, all registration methods require

1. **A deformation or transformation model** - in order to describe how the source image can be mapped to the target image in a way that can be computed.

2. **A similarity measure** - this is used with an *objective function* which describes when the optimal alignment has been reached.

3. **An optimisation method** - a way to reach the best alignment as described by the objective function.

In general, registration methods can be categorised by the transformations that they produce, which can be loosely classed as being linear[1] or non-linear.

2.2.1 Linear Registration

Linear registration methods have linear transformations which are global and can be described as in a single function which is applied to the whole image. Strictly, a transformation T is linear if

$$T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}) \tag{2.11}$$

---

[1]Rigid and affine transformations are not strictly linear, but are often classified as such

and for any constant c,

$$cT(\mathbf{x}) = T(c\mathbf{x}) \tag{2.12}$$

Linear transformations preserve straight lines in the image and are a core component in rigid and affine registration. Rigid and affine transformations are linear in the sense that they preserve straight lines, but they also include translational components, which violates the strict definition of a linear mapping [85] in 3D Cartesian space. Many authors refer to rigid and affine registration as being linear and this is true if homogeneous coordinates are used, but it would be more correct to describe them as linear with translation.

This family of transformations can be sub-classed into the number of degrees of freedom the transformation provides: typically 6, 9 or 12 for 3D images. Usually, the first 6 degrees of freedom are for rigid transformations, whilst the next 3 add scaling and the final 3 add shearing. Broadly speaking, registrations which provide 9 or 12 degrees of freedom are categorised as affine registrations, which can be considered a superclass of rigid registrations.

Rigid Transformations

Rigid transformations preserve angles and distances in the image before and after the transformations. They have only 6 degrees of freedom: 3 for translation in the direction of each axis and 3 for rotation around each axis. Let $\mathbf{t} = (t_x, t_y, t_z)$ represent the translations in x, y, and z directions and let $\mathbf{R}$ be a orthogonal rotation matrix, then the transformation of a point $\mathbf{x} = (x, y, z)$ in 3D space can be expressed as a function:

$$T_{rigid}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{t} \tag{2.13}$$

Affine Transformations

Affine transformations extend the number of degrees of freedom from the 6 in rigid transformations to also include scaling and shearing, which each add another 3 degrees of freedom. This means that angles and distances are no longer preserved, but parallel lines still are. The affine transformation for a point $\mathbf{x} = (x, y, z)$ can be described by:

$$\mathrm{T}_{affine}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{t} \tag{2.14}$$

where $\mathbf{A}$ is a 9 parameter transformation matrix describing rotation, scaling and shearing, and $\mathbf{t}$ provides the translational components (t_x, t_y, t_z).

Computationally, both rigid and affine transformations can be performed in a single matrix multiplication:

$$\mathrm{T}_{affine}(\mathbf{x}) = \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix} = \begin{pmatrix} a_{00} & a_{01} & a_{02} & t_x \\ a_{10} & a_{11} & a_{12} & t_y \\ a_{20} & a_{21} & a_{22} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{2.15}$$

where $\{a_{ij} : i, j \in [0, 2]\}$ are parameters describing the rotation, shearing and scaling properties of the linear transformation matrix $\mathbf{A}$. For rigid transformations, $\{a_{ij}\}$ can be constrained to be provide solely rotation components of the orthogonal rotation matrix $\mathbf{R}$.

2.2.2 Non-linear Registration

In contrast to linear registration, non-linear methods, sometimes also referred to as non-rigid or deformable registration, can provide local transformations rather than just global ones. In general, this could enable better alignment than linear transformations, particular for soft non-rigid tissues, however the process is more complex and requires modelling local deformations which

are computationally more expensive. In many applications, the need for accurate alignment often outweighs the computational costs, and linear registration is often insufficient to handle the non-rigid nature of much of the human body. Additionally, in some segmentation approaches, such as most atlas-based methods (see section 2.3.2), obtaining an accurate alignment between images is a prerequisite.

Achieving such a feasible alignment is complex, difficult and sometimes application dependent, with many proposed approaches to modelling the deformations. These include elastic deformations [10], modelling as fluids [50], and using basis functions such as splines [30] and wavelets [193] to name just a few. A recent survey of deformable registration methods is detailed in [162], which covers the topic in much more detail than the scope of this thesis.

A general constraint of any non-linear registration method is that the deformations must be smooth and invertible, which ensures that the resulting transformations can be quantified, and are plausible and realistic. There are also other constraints that are desirable and sometimes applied [162]:

- **Inverse Consistency** - Ensuring the forward and backward transformations are inverse mappings of each other.

- **Symmetry** - Ensuring registering image I_s to image I_t has the same effect by symmetry, in terms of deformations, as registering image I_t to image I_s.

- **Topology Preserving** - Ensuring there is a one-to-one mapping from source to target image and that the mapping has a continuous inverse; also known as homeomorphism.

- **Diffeomorphism** - Ensuring the topology preserving constraint is satisfied and additionally that both the mapping function and its inverse are differentiable.

However, the necessity of these constraints are application dependent and may be difficult to satisfy for general purpose methods. Indeed, the majority of registration methods are asymmetric and may not satisfy any these constraints. As a consequence, the choice of the target image or domain is highly important to the registration outcome.

Free-Form Deformation

One commonly used deformation model is free-form deformation (FFD), a technique derived from computer graphics [149]. This method manipulates an underlying mesh of control points which can be displaced to enable local deformations. B-splines can be used as function to model the local deformation, allowing the deformation to be interpolated between control points and thus providing a mapping for all pixels in the image [146]. This method has been used to align images in several notable segmentation methods such as [140], [79] and [4]. Additionally, this approach has been used in [191], which we compare with in Chapter 5.

Given a $n_x \times n_y \times n_z$ mesh of regularly spaced control points $\phi_{i,j,k}$, the local transformation for voxel $\mathbf{x} = (x, y, z)$ can be modelled by:

$$\mathrm{T}_{local}(\mathbf{x}) = \mathbf{x} + \sum_l \sum_m \sum_n \mathcal{B}_{l,m,n} \phi_{i+l,j+m,k+n} \tag{2.16}$$

where $\mathcal{B}_{i,j,k}$ depends on the choice of blending function. One such function is the symmetric cubic B-spline [103], [104], in which case the 3D tensor product of 1D cubic B-splines can be used as:

$$\mathrm{T}_{local}(\mathbf{x}) = \mathbf{x} + \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} B_l(u) B_m(v) B_n(w) \phi_{i+l,j+m,k+n} \tag{2.17}$$

where $i = \lfloor \frac{x}{n_x} \rfloor - 1$, $j = \lfloor \frac{y}{n_y} \rfloor - 1$, $k = \lfloor \frac{z}{n_z} \rfloor - 1$, $u = \frac{x}{n_x} - \lfloor \frac{x}{n_x} \rfloor$, $v = \frac{y}{n_y} - \lfloor \frac{y}{n_y} \rfloor$, $w = \frac{z}{n_z} - \lfloor \frac{z}{n_z} \rfloor$, and B_l is the lth basis function of the B-spline [103], [104].

Usually in practice, a global alignment is first sought using rigid or affine registration before applying non-linear methods to align local areas. An overall transformation model [146] can then be described as

$$\mathrm{T}(\mathbf{x}) = \mathrm{T}_{global}(\mathbf{x}) + \mathrm{T}_{local}(\mathbf{x}) \tag{2.18}$$

where $T_{global}(\mathbf{x})$ is a global affine or rigid transformation and $T_{local}(\mathbf{x})$ is the local deformation as represented in (2.17). Changing the spacing between control points is one way to control the amount of allowable deformation, where larger spacing provides more global deformations whilst smaller control points enable more local and more detailed deformations. In order to allow detailed alignment whilst maintaining computational efficiency, a multi-resolution approach is often used for the optimisation process. This coarse-to-fine approach enables bigger, more global displacements to be made first before fine-tuning of local displacements are made.

As previously mentioned, all registration methods require a similarity or distance metric in order to compare the alignment between images. There are several metrics that can be used to this end, which we will discuss later in section 2.4.

2.3   Image Segmentation

Segmentation techniques vary in the type of image information they use as well as any prior knowledge or constraints about the problem. Consequently, the most suitable method is often dependent on both the modality of the image as well as the objects of interest within the image. [133] provides a general overview of early medical image segmentation methods whilst [12] and [108] provide more recent reviews. Additionally, [36] provides an overview of recent atlas-based approaches for brain MRI and [81] provides a review on using statistical shape models.

2.3.1   Intensity Modelling

An early approach to segmentation in medical images is to model the intensities of the various tissues using a finite mixture model, where it is assumed that the intensities are independent samples from a mixture of probability distributions. Construction of these models often examine the intensity histograms of the labelled data (see Figure 2.4), however determining the mixture of each distribution which best describes the data can be accomplished through a wide variety of approaches. One common approach is to use Gaussian distributions, and are

referred to specifically as Gaussian mixture models (GMMs) when applied. In these models, the expectation-maximisation (EM) algorithm [55] is usually used to fit the GMMs to the data. Other approaches for the finite mixture model include using parzen-windows [128] and Markov random fields [83] as well as a range of unsupervised methods [64].
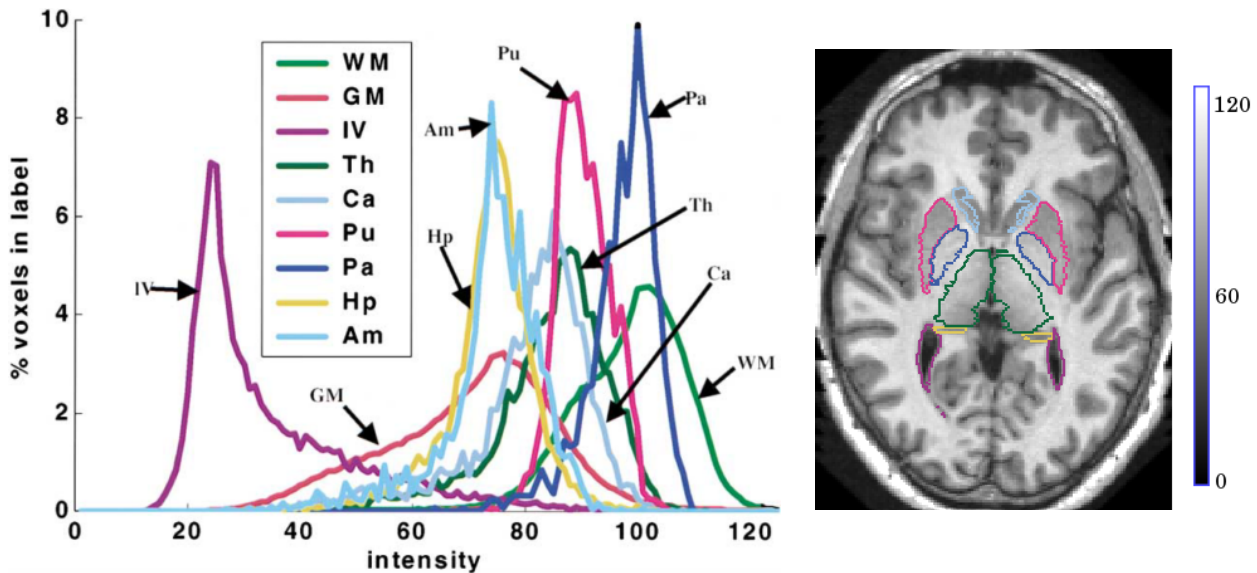
Figure 2.4: The intensity histograms for white matter (WM), grey matter (GM), lateral ventricle (lV), thalamus (Th), caudaute (Ca), putamen (Pu), pallidum (Pa), hippocampus (Hp) and amygdala (Am) in brain MRI from [66]. An example image with some of the structures outlined in their respective colours are shown on the right.

Originally, models were proposed to model intensities globally, however this approach can be unreliable and limited due to major overlaps in the intensity values and not taking into account any spatial information (see the histograms for hippocampus and amygdala in Figure 2.4). More recent methods address this limitation by combining intensity models with atlases.

2.3.2 Atlas-based Segmentation

One of the most popular concepts for segmentation is known under the umbrella term of atlas-based segmentation. An atlas in medical imaging is similar in definition to the term used in cartography, it is a term for an image(s) which is labelled and captures the properties and spatial relationships of the structures of interest. Atlases provide the ground truths and the examples

(a) An image   (b) A label map   (c) An atlas
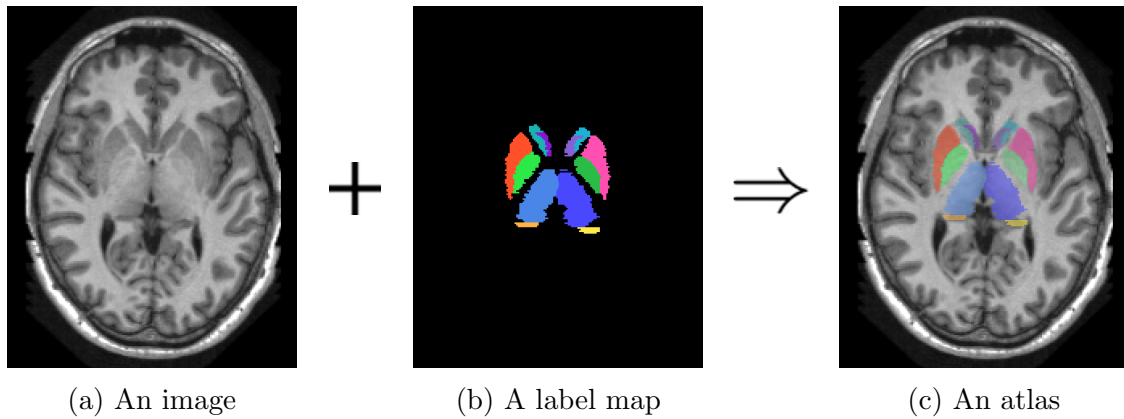
Figure 2.5: In this thesis, the term atlas is used to mean a pairing of an image and a label map. An atlas library or database is a collection of these atlases.

for which models can be built around. The formal definition of an atlas differs between authors, with some choosing it to mean a collection of images and label maps, whilst others use the term for individual images. At the most basic level, an atlas is a pair (I, M), where I is an image and M is the corresponding label map - this is the definition adopted in this thesis. Initially, atlases were constructed from a single subject which is chosen to represent the average, however this does not capture the anatomical variability present within the population. Modern atlases are composed from multiple subjects and aim to capture the full diversity of the population. Usually, atlases are created from individually segmented images, but they could also be created by integrating information from multiple images registered to a common template space.

One approach for segmentation is to transfer the labels, spatially, from an atlas to an unsegmented image by using the mapping derived from registration (see Figure 2.6). The idea here is to use registration to achieve an alignment of the two images, which also provides the an alignment for the structures of interest, thus a voxel-wise mapping can also be made from the labelled data to the unlabelled data. This is known as *label propagation* and is also sometimes known as *registration-based segmentation* [11], [45], [56]. This concept forms the basis for some recent developments such as probabilistic atlas, multi-atlas and patch-based methods.

The strategy for single atlas segmentation relies heavily on the registration process and in obtaining highly accurate alignments between images. However, registration errors are often present in most real-world applications. Whilst accurate segmentations are achievable for similar images, which is often the case for when they are from the same subject, the process is far

Atlas   Target   Segmentation
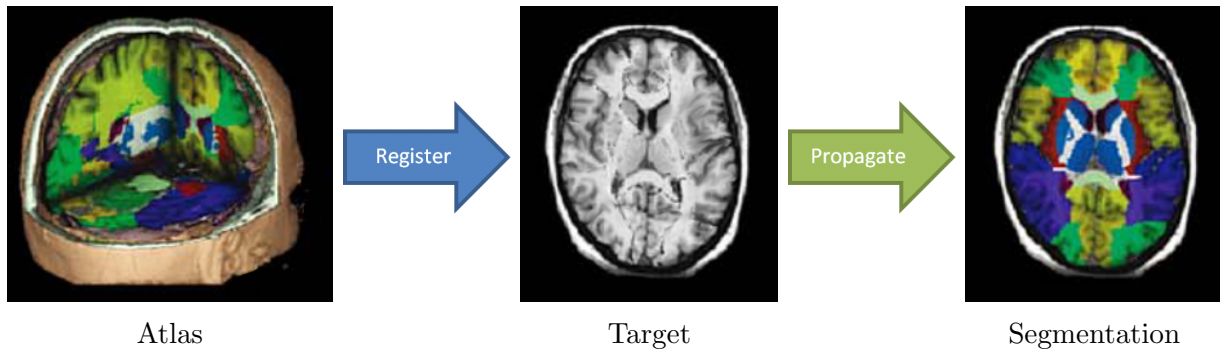
Figure 2.6: Atlas-based segmentation (label propagation): Register the atlas to the target image and transfer the label mapping from the atlas to obtain the segmentation.

less accurate when images are from different subjects, where there can be large differences. Misalignment after registration are far more substantial when images are dissimilar, and because labels are propagated directly from the resultant mapping, this then leads to incorrect labels being propagated from the atlas to the target image. Ultimately, this limits the applicability of this approach, particular for use in population studies where there can be significant differences between subjects. More recent methods use multiple atlases, probabilistic atlases or models to overcome this limitation.

2.3.3 Probabilistic Atlas-based Segmentation

Probabilistic atlases [62], [127] can be created from the scans and segmentations of multiple subjects once they have been registered to a common template space. For example, the MNI-ICBM152 template is often used in brain MRI applications. This provides spatially varying prior information, which can be propagated to images after registration and used in a number of statistical and variational frameworks such as in [66], [116], [78], [17], [153]. An example is shown for a cardiac probabilistic atlas in Figure 2.7.

The prior information from probabilistic atlases can be used in several ways for segmentation. A common approach is to combine the probabilities with intensity models, which often use Gaussian mixture models, where each tissue class is parameterised by a different Gaussian distribution. These parameters are then often optimised by using the *expectation-maximisation (EM)* algorithm [55] to explain the intensities for each target image. This was the approach

(a) Template   (b) Probabilities
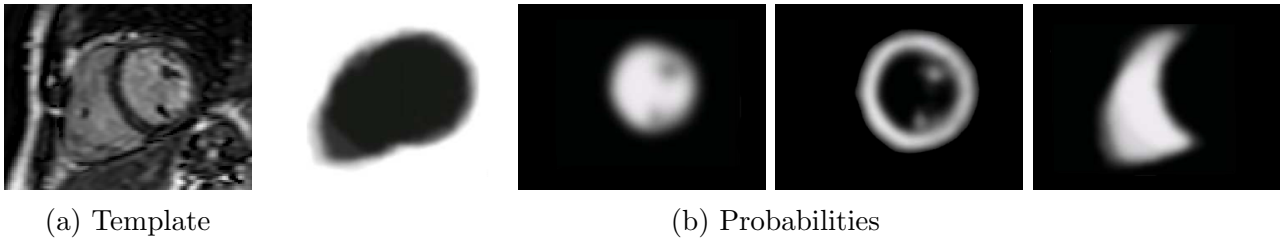
Figure 2.7: A probabilistic atlas provides a probability map for each of the labels in a template space. In this case, we have the probabilities for the background, left ventricle, myocardium and right ventricle in cardiac MRI. Probabilities are shown as proportional to intensity.

proposed in [175] to obtain model for white matter, grey matter and cerebrospinal fluid classification in brain MRI whilst also correcting for for intensity inhomogeneities from the MR acquisitions. In [175], this approach was further extended through the use of a Markov random field (MRF) to obtain a smooth segmentation for each of the tissues. Other related methods have since been proposed such as in [66], [106] and [134].

Usually, only a single registration is required, in order to align target images to the common template space. However, finding a common template space may be challenging depending on the anatomical variability between subjects, and registration accuracy would be a critical component for both the construction of probabilistic atlases and the accuracy of the prior information for segmentation.

2.3.4   Multi-atlas Label Propagation

Multi-atlas label propagation (MALP) approaches [140],[79],[4],[107] provide a more robust approach to segmentation compared to the single atlas predecessor. Using multiple atlases means that the segmentation results are not completely reliant on the outcome of a single registration. Instead, labels from multiple atlases are propagated to the target image after registration. This provides multiple segmentations, which are then combined using a label fusion process to determine the overall consensus segmentation (see Figure 2.8). Usually, only the most appropriate atlases are selected for label fusion by comparing the atlases with the target image with some predefined similarity measure. The use of multiple atlases can better account for the anatomical variability in images than a single atlas and minimise the effect

of registration errors through consensus voting. There are a number of strategies for MALP; variations are largely around the atlas selection scheme [4] and the label fusion process [7], which are often applied with other atlas-based segmentation approaches such as probabilistic and patch-based methods.

| Multiple Atlases | Target | Multiple Segmentations | Final Segmentation |
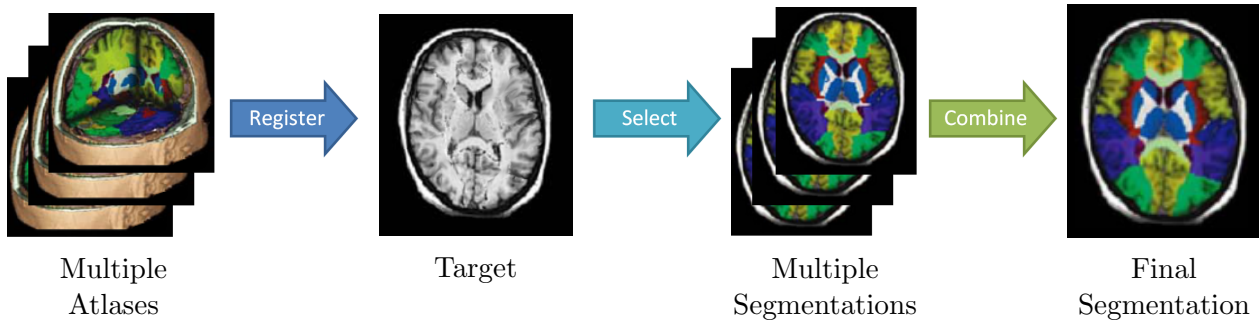
Figure 2.8: Multi-Atlas Label Propagation: Register each atlas to the target image, then select the most similar atlases and perform label fusion to combine their segmentations to obtain the final result.

Atlas Selection

Atlas selection strategies vary according to the distance or similarity measure and the number of atlases used. In general, it is desirable to use only similar atlases to propagate labels from, as registration errors tend to be reduced when subjects are similar, thus the labelling is more likely to be accurate. [140] and [4] provides a comparative study of some of these strategies, post-registration. Selecting atlases pre-registration is non-trivial and has not been well studied, however it is related to content based image retrieval (CBIR), which has many different approaches. [3] provides a review of some current techniques used in CBIR.

Similarity and distance metrics used for registration are often also used for atlas selection, with the most popular approaches using *sum of square differences* (SSD) and *normalised mutual information* (NMI) (see Chapter 2.4). These measures are applied after registration, and are calculated on a one-to-one pixel-to-pixel basis. This is performed between the target image and each of the atlases, so that the atlases can be ranked in order of similarity. This then allows the N most similar to be selected for label propagation. [140] and [4] demonstrated that using

multiple similar atlases is more effective than using the single most similar atlas, but only a limited number of atlases should be used rather than all available atlases. This is due to the effect of registration errors that can be minimised through consensus labelling from multiple atlases, whilst using all atlases would also propagate labels from dissimilar atlases where the labels are not likely to match the target image. The application of atlas selection was also shown to be more effective than selecting a random subset of atlases in [4], thus demonstrating that the atlas selection scheme is important for segmentation accuracy.

The choice of similarity metric is dependent on the application, as the use of SSD may not be applicable without adequate intensity normalisation, whilst NMI is not as sensitive to small local differences, but is well suited when there is widely differing levels of contrast and appearance.

Label Fusion

Label fusion, also sometimes known as decision fusion, is often considered analogous to traditional classification problems using multiple independent classifiers, since each registered atlas provides a classification [141]. These methods can be categorised into global methods, where weighting for each atlas is performed on an image-wide basis, and local methods, where weights vary spatially within each image. [7] and [36] provide comparative reviews of many of these approaches prior to the development of patch-based methods.

Generally, given a set of possible labels $\{L_i\}$, a consensus label $L(\mathbf{x})$ for each voxel $\mathbf{x}$ can be derived by the combining the outputs from each of the atlases $a_1(\mathbf{x}), ..., a_N(\mathbf{x})$, where N is the number of atlases, and by using some performance or evaluation model E:

$$L(x) = \arg\max_i p(\mathbf{x} \in L_i | a_j(\mathbf{x}), ..., a_j(\mathbf{x}), E) \tag{2.19}$$

The simplest approach is majority voting, which is also one of the first and most popular approaches, and often compared to as a baseline method in MALP. In this approach, each selected atlas casts a single vote for the label at each voxel and the label which obtains the

most votes is then chosen as the final labelling. This is equivalent to the *vote rule* for decision fusion, proposed for combining the results of multiple classifiers in [192]. This provides a global approach that is simple to implement, yet proven to be effective in several cases [79], [80]. Given N atlases with labels $a_i(\mathbf{x})$ for $i = 1, ..., N$ at each voxel $\mathbf{x}$, the final labelling L is given by:

$$L(\mathbf{x}) = \arg\max_i \sum_j^N \begin{cases} 1, & \text{if } i = a_j(\mathbf{x}) \\ 0, & \text{otherwise} \end{cases} \tag{2.20}$$

Since transformations are continuous, the voxel grids may not exactly align, thus the transformation of segmentations may result in contention for the same voxel by differing neighbouring labels. To resolve this, an interpolation scheme can be applied. In the simplest approach, nearest neighbour interpolation is used and each voxel is always has assigned a singular whole label, as is assumed in (2.20). In other schemes, such as *partial volume interpolation* [111], a weight for each label may be assigned at each voxel [141]. This then allows each atlas to provide a probabilistic or weighted vote for each voxel. A straight-forward extension of the *vote rule* to this case would be to sum up the probabilities from each atlas as weights, and then the class with the largest weight is chosen as the final label. This is known as the *sum rule*. A more in-depth review of the different decision rules that may be used, including the sum rule and vote rule, can be found in [94].

Another popular scheme is the STAPLE method [184], which uses the EM framework to iterate between an estimation of the "true" consensus segmentation and an estimation of the reliability parameters of each of the atlases. In each iteration, the current consensus segmentation can be used to measure the reliability of each of the atlases, and provide weights of the contribution of each atlas, based on its sensitivity and specificity. The derived weights are then used to generate the estimate of the consensus segmentation in the next iteration. This provides a *globally weighted* approach which is more sophisticated than majority voting and able to assign different weights to each atlas in order to minimise the impact of the dissimilar atlases. This can be particularly advantageous if there is a large variation in the selected atlases.

Locally weighted methods were developed more recently, with [7] proposing a generic approach where weights for each atlas vary spatially according to a local similarity measure, calculated using a distance metric within a fixed local neighbourhood. This can be expressed as

$$L(\mathbf{x}) = \arg\max_i \sum_j^N \begin{cases} [m(p,r)]^s, & \text{if } i = a_j(\mathbf{x}) \\ 0, & \text{otherwise} \end{cases} \tag{2.21}$$

where $m(p, r)$ is any local similarity measure that which compares a local neighbourhood of shape p, radius r and with s controlling the sensitivity. Potential measures include normalised mutual information, normalised cross correlation and mean square distance, which were evaluated in [7]. This is generally more computationally complex than globally weighted methods, but was found to perform more favourably. In general, locally weighted methods can better accommodate the range of anatomical differences between subjects (see Figure 2.9), which enables them to provide superior performance compared to globally weighted methods.

In other local approaches, spatially varying probabilities can be used to provide local weights. In [107], MALP was combined with local intensity models to refine the segmentation estimate as well as applying a graph cuts method based on [190]. Prior to this, a similar approach using local intensity appearance models was proposed in [174] which generated target-specific probabilistic atlases after registering individual atlases and then combining this with local intensity models as well as graph cuts refinement using [74].

More recently, the development of locally weighted methods has lead to approaches which perform weighted label fusion after local correspondence searches rather than direct voxel-to-voxel based weighting [47], [178], [142]. These methods can be classified under the category of *patch-based segmentation*, and are reviewed in more detail in the next section.

(a) Target   (b) Registered atlases   (c) Majority Voting
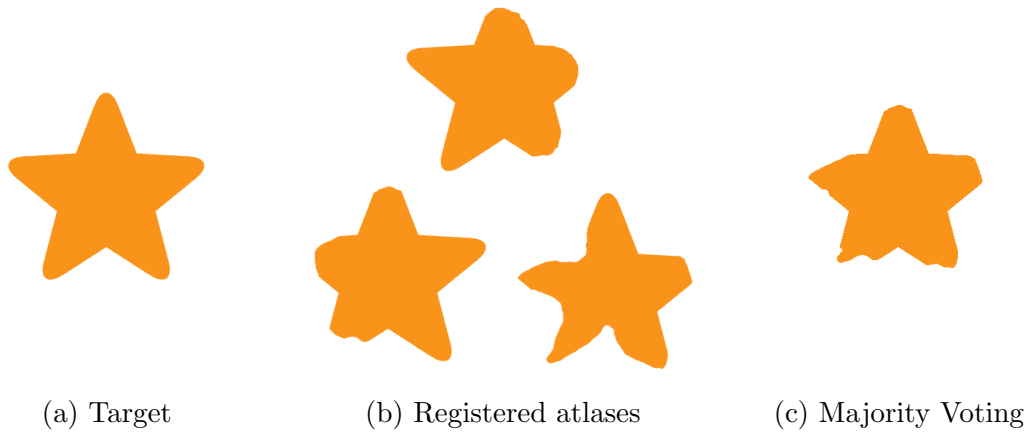
Figure 2.9: A simple example to illustrate the limitations of globally weighted local fusion. Individually, none of the atlases align perfectly to the target image. Locally, the atlases together would be able to accommodate for misaligned regions of the image, however if weights were assigned globally, this would not be possible. (c) provides an example of the result of a globally weighted label fusion - using majority voting.

2.3.5   Patch-based Segmentation

For most multi-atlas segmentation methods, the dependence on image registration can be problematic as inaccurate alignment adversely affects the performance of the segmentation. Additionally, finding suitable (fixed) registration parameters that yield accurate non-linear correspondences on different images can be a challenge on its own, particularly for anatomies that are highly variable. Furthermore, the range of atlases available may not always fully accommodate for the anatomical variability present between subjects, particularly as acquiring large datasets of atlases may be time consuming and expensive.

Patch-based methods for label propagation [47],[142] relax the dependence on registration accuracy and do not rely on explicit one-to-one correspondences between images. In general, these approaches label each voxel of a target image by comparing the image patch centred on the voxel with neighbouring patches from an atlas library and assigning the most likely label according to the closest matches (see Figure 2.10). Due to the relaxation of the required alignment between images, these methods are often able to use affine rather than non-rigid registration, yet still produce comparable results [142]. Patch-based methods for label fusion have also been shown to be effective for several applications in medical imaging [8], [179].

In contrast to previous MALP approaches, where atlas selection and label fusion is performed
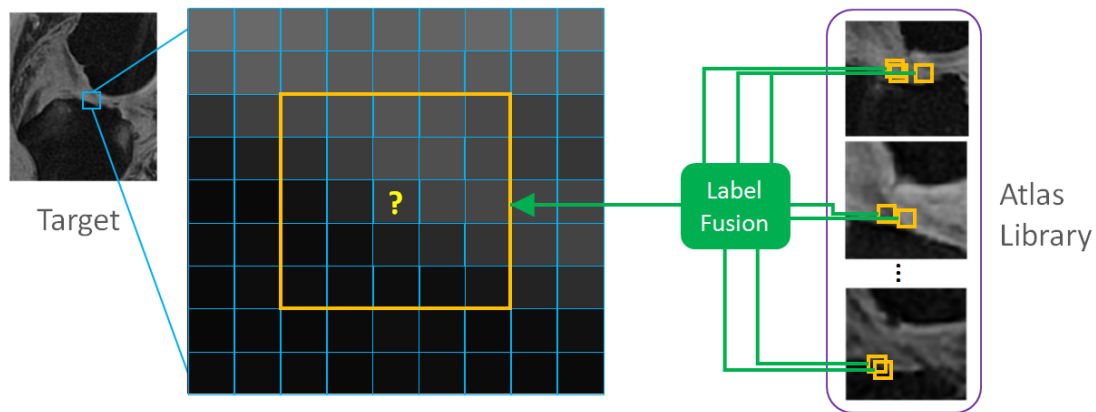
Figure 2.10: Patch-based Segmentation: For each voxel, labels are propagated from the most similar patches in the atlas library, rather than propagating one-to-one, voxel-to-voxel from the atlases to the image like in MALP.

on a global voxel-to-voxel basis, the patch-based approach to segmentation can be conceptually described as a similar process but within local neighbourhoods around each voxel. For each voxel within an image, a *patch* is essentially a raw feature descriptor of the local area which surrounds that voxel. Selecting similar patches from atlases for patch-level label fusion is analogous to selecting atlases for label fusion in MALP. The key advantages of this local approach are:

1. *Increased population size and reduced dimensionality.* There are many more patches to select from than whole atlases. The dimensionality of each patch is also much smaller than a whole image, decreasing the possible combinations of intensity values. This means the data space is much more dense, so increasing the chances that the selection results have high statistical significance.

2. *No assumption of a one-to-one mapping between images.* MALP transfers labels from atlases to images on a one-to-one voxel basis, but this relationship may not always be present between images, even after registration. A patch-based approach does not assume an explicit one-to-one mapping between images and label transfer is not limited to a one-to-one relationship between images. Voxel-wise, labels can be transferred to any number of voxels on the target image. This can overcome problems for multi-atlas segmentation where the anatomical variability cannot be fully accounted with the available atlases. This also relaxes the required accuracy of the registration algorithms, reducing the amount of manual input and bespoke customization needed.

Nonlocal Means Based Label Propagation

Many existing patch-based approaches [47, 142, 61, 191] apply a label fusion method based on the nonlocal means method [34] which was original proposed for patch-based denoising. When applied to label fusion, this approach derives a weighting for each label according to the intensity distances of the most similar patches. At each voxel location, $\mathbf{x}$, in the target image, let $P(\mathbf{x})$ be the patch extraction operator at $\mathbf{x}$ and let $\mathcal{N}_{\mathbf{x}}$ be a surrounding neighbourhood of $\mathbf{x}$ and $\mathbf{y}_{L,i} \in \mathcal{N}_{\mathbf{x}}$ represent voxels from the atlas library for label L which have similar patches to $\mathbf{x}$. A weighting for each label at voxel $\mathbf{x}$ is then determined as:

$$w_L(\mathbf{x}) = \frac{\sum_{\mathbf{y}_{L,i} \in \mathcal{N}_{\mathbf{x}}} w(\mathbf{x}, \mathbf{y}_{L,i})}{\sum_{L \in L_A} \sum_{\mathbf{y}_{L,i} \in \mathcal{N}_{\mathbf{x}}} w(\mathbf{x}, \mathbf{y}_{L,i})} \tag{2.22}$$

where $w(\mathbf{x}, \mathbf{y})$ is the weight of each patch and is determined by:

$$w(\mathbf{x}, \mathbf{y}) = e^{\frac{-||P(\mathbf{y}) - P(\mathbf{y})||_2^2}{h^2(\mathbf{x})}} \tag{2.23}$$

$h^2(\mathbf{x})$ is a decay parameter to control the level of influence of patches as the distance increases. In [47], an automatic estimation of this is calculated for each voxel based on the minimum distance between patch $P(\mathbf{x})$ and the relevant patches from the atlas library, $\{P(\mathbf{y}_i) : \mathbf{y}_i \in \mathcal{N}_{\mathbf{x}}\}$:

$$h^2(\mathbf{x}) = \min\{||P(\mathbf{x}) - P(\mathbf{y}_i)||_2^2\} \tag{2.24}$$

In approaches based on [47, 142], L is determined as the final label if $w_L(\mathbf{x})$ is greater than a predefined threshold t, otherwise the label defaults to the background label. For binary labelling, t is often simply defined as 0.5, which is equivalent to determining the final label by majority voting. Additionally in these methods, a sliding window approach is used to define $\mathcal{N}_{\mathbf{x}}$, and only patches from within this window are used for label fusion. In [47], a structural similarity measure [183] is used to preselect a subset of the patches within the search window

from for label fusion. This is defined as:

$$ss = \frac{2\mu_{\mathbf{x}}\mu_{\mathbf{y}}}{\mu_{\mathbf{x}}\mu_{\mathbf{y}}} \times \frac{2\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}} \qquad (2.25)$$

where μ and σ are the means and standard deviation of the intensities of the patches around voxel $\mathbf{x}$ in the target image and voxel $\mathbf{y}$ from the atlas. Only patches with structural similarities greater than a predefined threshold are used for label fusion. This was used to reduce the computational time, since heat maps of the means and standard deviations can be produced offline. In [47] and [61] only patches with $ss > 0.95$ were used.

Joint Label Fusion

In addition to nonlocal means derived methods, other alternative patch-based approaches have also been proposed, although not always explicitly described as such. One particular approach, called *joint label fusion* [179], was proposed to combine labels whilst evaluating the performance of the atlases together rather than independently as the other multi-atlas approaches have done so far. By doing so, it aims to take into account similar errors which can occur with different atlases and minimise the expected total labelling error. To do this, the pairwise dependency between atlases is modelled as a joint probability of two atlases making an error at each voxel. This is approximated by looking at the intensity similarities between each pair of atlases and the target image in a local neighbourhood, similarly to [7] and as shown in (2.21). Thereafter, the aim is to choose a set of weights $\mathbf{w}_{\mathbf{x}}^{*}$ for the atlases that minimise the error between the true segmentation and the consensus segmentation. This can be represented as:

$$\mathbf{w}_{\mathbf{x}}^{*} = \arg\min_{\mathbf{w}_{\mathbf{x}}} \mathbf{w}_{\mathbf{x}}^{\mathrm{T}}\mathbf{M}_{\mathbf{x}}\mathbf{w}_{\mathbf{x}} \qquad \text{subject to} \qquad \sum_{i=1}^{N}\mathbf{w}_{\mathbf{x}}(i) = 1 \;\text{ and }\; \mathbf{w}_{\mathbf{x}}(i) \geq 0 \qquad (2.26)$$

where $\mathbf{w}_{\mathbf{x}}$ is a vector of weights $[w_1(\mathbf{x}), ..., w_N(\mathbf{x})]$ for each atlas and $\mathbf{M}_{\mathbf{x}}$ is a pairwise dependency matrix with $\mathbf{M}_{\mathbf{x}}(i, j)$ being the probability atlas i and atlas j both produce the wrong segmentation label, given the local image similarity. Let I_t be the target image and I_i and I_j

be the images of atlases i and j, then the joint probability can be written as:

$$\mathbf{M_x}(i,j) \propto [||P_t(\mathbf{x}) - P_i(\mathbf{x'})|| P_t(\mathbf{x}) - P_j(\mathbf{x''})||]^s \tag{2.27}$$

with $P_t(.)$, $P_i(.)$ and $P_j(.)$ being patch extraction operators for image I_t, I_i and I_j respectively, and s being a parameter for sensitivity. The proportionality does not affect the choice of weights since multiplying by a positive constant does not change the solution w. $\mathbf{x'}$ is selected from a search window $\mathcal{N_x}$ around voxel $\mathbf{x}$ by:

$$\mathbf{x'} = \arg \min_{\mathbf{y} \in \mathcal{N_x}} ||P_t(\mathbf{x}) - P_i(\mathbf{y})||^2 \tag{2.28}$$

and similarly for $\mathbf{x''}$ with atlas j. This presents the biggest difference, ideologically, to approaches derived from [47] based on nonlocal means in that only a single patch is evaluated from each atlas at each voxel $\mathbf{x}$, whereas the nonlocal means compares multiple patches from each atlas. Prior to joint label fusion, the author also proposed a regression based label fusion approach using a similar principle [178].

For performance reasons and to provide context for patch comparisons, the window size for $\mathcal{N_x}$ is often limited in the most patch-based methods (typically less than $11^3 = 1331$ voxels). This means that images must be aligned within this margin of error in order for the patch selection and label fusion strategy to work well.

2.4 Image Similarity and Distance Metrics

It is often desirable to quantify the similarity, or conversely the distance, between two images in order to compare them. This is fundamental for many methods in machine learning as well as from a data analysis and retrieval perspective. For example, to rank search results, or to compare alignments for registration. There are many ways to compare images, however the meaningfulness of the comparison is very much dependent on the context. In general,

the range of methods can be broadly split into two main categories: distance metrics and similarity measures. They can thought of as being inverse to each other, the former is derived from mathematics and geometry, whilst the latter is largely rooted in signal processing and information retrieval.

2.4.1 Distance Metrics

Commonly, when images are vectorised, distance metrics are applied to them using vector norms - for example the 2-*norm* which is also referred to as the **Euclidean distance**:

$$\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{2.29}$$

As a generalised form, this is also known as the *p-norm*, L_p-distance or **Minkowski distance** with order p:

$$\left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \tag{2.30}$$

When $p = 1$ this is also known as the **Manhattan distance** and is sometimes referred to as the **sum of absolute differences (SAD)**.

For $p \geq 1$, *p-norms* satisfy the mathematical definition of a distance metric. This is defined as a function, d, on a set $\mathcal{X}$ which returns a real number, where given $x, y, z \in \mathcal{X}$, the following constraints hold true:

1. *Non-negativity:*, $d(x, y) \geq 0$

2. *Identity:* $d(x, y) = 0 \iff x = y$

3. *Symmetry:* $d(x, y) = d(y, x)$

4. *Triangle Inequality:* $d(x, z) \leq d(x, y) + d(y, z)$

The triangle inequality is particularly important for several algorithms, such as metric trees for nearest neighbour search, which exploit it to avoid exhaustive search of the entire space. However, not all *distance functions* satisfy this constraint. Outside the pure mathematical context, distance functions and metrics are often used interchangeably without regard to the mathematical definition. A commonly used example is the **sum of squared differences (SSD)**. This is essentially the squared Euclidean distance and is defined as:

$$\sum_{i=1}^{n} (x_i - y_i)^2 \tag{2.31}$$

SSD does not satisfy the triangle inequality in a case where $x = (0,0)$, $y = (0,2)$, $z = (0,4)$: here, $d(x,z) = 16$, $d(x,y) = 4$ and $d(y,z) = 4$ but $16 \nleq 4 + 4$. However SSD is still often used for comparisons as it penalises large differences and is more sensitive than the SAD or Euclidean distance.

The choice of distance metrics are not particularly well studied for image comparison applications, although [2] provides a comparative review of Minkowski distances in high dimensional spaces, as well as suggesting the use of L_p-distances with $p < 1$. This is contrary to the majority of applications using distance metrics, where the Euclidean distance is chosen as a default.

Other relevant distances used include the *Hamming distance* [77] and the *Mahalanobis distance* [112], although they are generally not used directly for comparison in the image space. The Hamming distance is only valid between binary vectors, whilst the Mahalanobis distance is a statistical measure which requires the calculation of a covariance matrix first.

2.4.2 Similarity Metrics

Similarity can be considered the inverse of distances, although similarity measures in general can be defined more loosely than distance metrics. Recently, [40] presented a formal definition of a similarity metric as a counter to the established definition of a distance metric: for x, y, z in a set $\mathcal{X}$, a similarity metric is a function s which satisfies the following:

1. $s(x, x) \geq 0$

2. $s(x, x) \geq s(x, y)$

3. $s(x, y) = s(y, x)^2$

4. $s(x, y) + s(y, z) \leq s(x, y) + s(y, y)$

5. $s(x, x) = s(y, y) = s(x, y) \iff x = y$

Popular similarity measures, particularly for registration, includes cross-correlation, mutual information and normalised mutual information. [131] provides a review of these as well as some additional measures used in registration.

Cross-correlation (CC) measures similarities in intensities at corresponding points between images by their dot product, and can be considered analogous to the convolution (or interference) of two signals. For two images x and y, CC is defined as:

$$s_{cc} = \sum_{i=1}^{n} x_i \cdot y_i \tag{2.32}$$

Normalised cross-correlation (NCC) is a form of CC which reduces sensitivity to the differences in brightness and contrast between images x and y:

$$s_{ncc} = \sum_{i=1}^{n} \frac{(x_i - \mu_x) \cdot (y_i - \mu_y)}{\sqrt{(x_i - \mu_x)^2 \cdot (y_i - \mu_y)^2}} \tag{2.33}$$

where μ_x and μ_y are the mean intensities of x and y respectively.

In contrast to cross-correlation, mutual information and normalised mutual information provide entropy-based measures. These measures examine the joint probability distribution of the two images after binning the intensities at each voxel. **Entropy** [151] is defined by the probability

---

[2]This case is not always true - for example the correlation ratio which is sometimes used as a similarity metric.

p of (binned) intensity values i within an image x as:

$$H(x) = -\sum_i p(i) \log p(i) \tag{2.34}$$

and the **joint entropy**, which itself can be considered a similarity measure, for two images x and y with intensities i_x in x and i_y in y, and joint probability function p is given by:

$$H(x,y) = -\sum_{i_x,i_y} p(i_x,i_y) \log p(i_x,i_y) \tag{2.35}$$

Mutual information (MI) [176] is defined using entropy as:

$$s_{MI} = H(x) + H(y) - H(x,y) = \sum_{i_x,i_y} p(i_x,i_y) \log \frac{p(i_x,i_y)}{p(i_x) \cdot p(i_y)} \tag{2.36}$$

A normalised form of this also exist - **normalised mutual information (NMI)** [165] is defined as

$$s_{NMI} = \frac{H(x) + H(y)}{H(x,y)} \tag{2.37}$$

For image alignment, NMI is usually favoured since it is invariant to the amount of overlap of the background and low intensity regions between the images, whereas MI and joint entropy are both sensitive to this. NMI also allows comparison between images of different modalities, whereas intensity-based metrics like cross-correlation assumes that there is a linear relationship between intensities when comparing images.

Similarity measures are also used in evaluating segmentation accuracy. One commonly used measure is the **Dice coefficient**, which is defined for two segmentations, A and B as:

$$s_{Dice} = \frac{2|A \cap B|}{|A| + |B|} \tag{2.38}$$

where the intersection of two segmentations are calculated for each label c on a voxel-to-voxel basis, i.e. for each voxel $\mathbf{x}$ $A(\mathbf{x}) \cap B(\mathbf{x}) \iff A(\mathbf{x}) = B(\mathbf{x}) = c$

This defines a function in terms of overlap between the two segmentations, which is one way in which the similarity of two segmentations can be measured. A related function which is also used is the **Jaccard** index, which measures the similarity between segmentation A and B by:

$$s_{Jaccard} = \frac{|A \cap B|}{|A \cup B|} \tag{2.39}$$

where the union and intersection of the two segmentations are calculated as before, on a voxel-to-voxel basis.

2.4.3   Learned Similarities and Distances

In addition to predefined similarity measures, it is also possible to learn a task-specific similarity or dissimilarity measure. Approaches have been proposed for registration [86], segmentation [164], as well as general classification [186]. Recently, this line of thought has also led to an approximate nearest neighbour data structure using random forests [97]. However, the use of learned similarity or distance metrics remain an open research topic and is not yet widely adopted for practical applications.

2.5   Nearest Neighbour Search

Nearest neighbour search is a critical yet expensive step in a wide range of computer vision and image analysis methods, such as clustering [91], manifold learning [20], inpainting [58] and classification [29] to name just a few. k nearest neighbour (kNN) classifiers are also used in a number of segmentation methods, such as in [67], [187]. However due to the large dimensionality of images, fast kNN search is a challenging problem. Given a query in a collection of n images with d dimensions, a naive exhaustive linear search would cost time that is proportional to nd.

This would be acceptable when both n and d is expected to be low, but in many applications, such as looking at image patches in medical imaging, both n and d can be very large - for example a $125 \times 125 \times 125$ resolution image could be split into $(125 - 5)^3 = 1728000$ patches of $5 \times 5 \times 5 = 125$ dimensions, resulting in a significantly large value that makes nearest neighbour search a substantial bottle neck in the speed of information processing.

Both the k nearest neighbour and the related ϵ-radius near neighbour search problem has been tackled in many areas associated with machine learning and data mining and as a result, a wealth of methods exists that can be applied to imaging applications. In general, speed-ups come from either partitioning the data in some way so that exhaustive search is not necessary, requiring comparisons for only a small subset of the entire data, or using a lower dimensional representation or feature space, where the search process is less complex. Methods generally require constructing data structures prior to searches, in a classic computer science case of using a trade-off between speed and memory costs.

2.5.1 Trees

A tree is a data structure that partitions the data with each branch according to some decision metric and has leaf nodes at the bottom of the tree which each contain a significantly smaller subset of the total data. Trees are constructed recursively, partitioning each subset of the data at each branch until some termination criteria is met. The way the data is partitioned is generally what defines each type of tree. The most common type of trees are binary search trees - each node partitions the data into two subsets, however quadtrees [65] and trees with even more branches per node, such as hash-trees[120], have been previously developed.

Trees can also be grouped into two types according to how they categorise the data:

- **Projective trees** - these partition points based on their projection into some lower-dimensional space

- **Metric trees** - these use some metric on defined pairs of points in space, and generally take advantage of the triangle inequality for more efficient data access.

(a) Projective tree   (b) Projective tree   (c) Metric tree
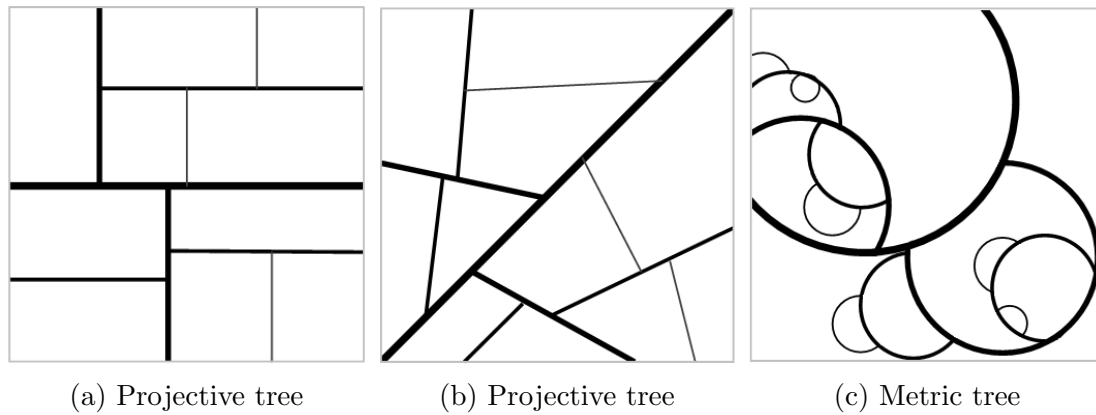
Figure 2.11: Example of how trees might partition a 2D space. (a) is an axis aligned projective tree, (b) is a non-axis aligned tree. (c) is a generalised metric tree

kd-Tree

Bentley's kd-tree [23] is a binary search tree and is one of the most commonly used algorithms - it is well known, simple to implement and there are also existing libraries for it in a variety of programming languages[3]. It splits the data using values from a single dimension at each branch until the size of the subset of data is below a certain threshold. For smaller trees[24], the dimension which with maximum variance is chosen and the split value is usually chosen to be the median value. This results in a fairly balanced partitioning of points into hyper-rectangles (see figure 2.11a) which are axis aligned, but this can lead to poor search performance if the data distribution does not favour this type of dimensional-partitioning. Furthermore, the number of neighbours for each leaf grows exponentially with dimensions and the data structure does not scale well with increases in dimensionality.

PCA Tree

Sproull [163] attempts to remedy the axis-alignment problem in kd-trees by applying Principle Component Ananylsis at each node to obtain the eigenvector corresponding to the maximum variance and then splitting the data along that direction. This still results in linear partitions

---

[3]scipy for Python, libkdtree++ for C++, Caltech Large Scale Image Search Toolbox for Matlab - to name a few

but they are no longer confined to the dimensional axes of the data (see figure 2.11b), but are partitioned by hyperplanes that can be of any orientation.

Ball Tree

The ball tree [124] is a metric tree which splits splits data according to which of two centres that it is closest to. It has similarities to the k-means algorithm, however the centres are chosen and fixed during the start of the construction of each node where construction is as follows:

1. Find the centroid of the dataset for the node.

2. Find the data point that has the greatest distance from the centroid and use that as the first centre.

3. Find the data point that has the greatest distance first centre and use that as the second centre.

4. Partition dataset into two subsets according to which centre the data is closest to.

There is no constraint on the number of data points assigned to each partition and the tree construction terminates when there is less than a certain number of data points at each leaf. The partitioning looks similar to the one for PCA tree, but each hyperplane bisects two centres and data points are partitioned according to distance to these centres rather than just absolute value. This method takes longer to construct than kd-tree and could be highly unbalanced, but could be significantly faster if it discovers the true distribution of data points in the data space.

vp-Tree

Vantage point (vp) trees [194] are similar to ball trees but rather than have multiple centres at each level of the tree, there is only a single centre at each level and data points are partitioned according to distance to it. The centre for each node can be chosen randomly or as the

centroid. This splits the data into "hypershells" around each centroid, and the thickness of the "hypershell" can be chosen in a variety of ways.

Other Trees

There are variety of other tree structures that can be used for nearest neighbour search such as cover trees[27], M-trees[42], R-trees[143] and their variants.

Cover tree are a type of metric tree which uses a refinement of the navigating nets data structure[99]. It tries to overcome the issue of space requirement by other tree structures such as kd-tree by making the space requirement linearly proportional to the dataset size, rather than exponentially proportional to the dimensionality. Cover trees are designed so that each level of the tree is a "cover" for the level beneath it, and the level is indexed by an integer scale which decreases as the tree is descended. Each node in the tree is associated to a data point, and each data point can be associated to multiple nodes in the tree, but appearing only once per level. There are 3 invariants that define the cover tree structure - let C_i denote the set of data points associated with the nodes at level i, d is the distance function:

- **Nesting** - $C_i \subset C_{i-1}$. This implies each data point occurs in every level below C_i once there is a node associated with it.

- **Covering tree** - for every $p \in C_{i-1}$, there is a $q \in C_i$ such that $d(p,q) < 2^i$ and the node in level i associated with q is a parent of the node in level $i-1$ associated with p

- **Separation** - for all distinct $p, q \in C_i$, $d(p,q) > 2^i$

M-trees are defined by an object at each node that identifies it, and a "ball" (within certain radius) around each node which defines the data points which belong to that node. It is similar to ball trees and vp-trees, but the "ball" around each node can overlap with that of another node. Each Leaf has a maximum population and a new node is created during insertion once the maximum population has been reached, splitting the population of the leaf into one of the two new leaves of the node.

R-trees are quite similar to *kd*-trees in that the data space is divided into hyper-rectangles, however in R-trees, each hyper-rectangles represent a "minimum bounding box" of the data at each level of the tree. There are several variants of R-trees such as R* tree [18], R+ tree[150], Hilbert R-tree [92] and X-tree [25].

2.5.2 Approximate Nearest Neighbour (ANN) Search

In many applications, it may be acceptable to trade accuracy for speed in the retrieval of the nearest neighbours, particularly if there is a large amount of high dimensional data. This is the idea behind approximate nearest neighbour (ANN) search, which often allows the user to set an approximation error bound to control the trade-off between speed and accuracy.

FLANN

One relatively popular choice for ANN search is the *fast library for approximate Nearest Neighbours (FLANN)* introduced by [122]. This library contains uses two principle data structures, randomised kd-trees and hierarchical k-means trees, which are used without backtracking and examining all candidate leaf nodes in order to achieve a speed up. A pre-defined level of precision, p can be set such that the correct nearest neighbour is returned for $p\%$ of the queries. FLANN has been used for a number of applications such as image classification [118], and as part of the point cloud library [147] in ROS[4].

Hashing

A data structure that is well known for its fast look-up times, which is on average a constant $O(1)$, is that of the hash table. There have been several hash table based data structures that have been applied to image search. These search methods are probabilistic, and are not guaranteed to find the exact nearest neighbour, however they can be employed for approximate nearest neighbour search. In an ordinary hash table, objects in the table are assigned a key,

---

[4]Robot Operating System `http://wiki.ros.org/`

based on a mathematical function of the object known as a "hash". There is a probability, depending on the size of the hash table, the hash function used, the length of the key and the dynamic nature of the table, that two object may have the same hash-key, or that a query of a non-existent item in the hash table will return an item, because the hash function returns the same value. When this occurs, it is known as a *collision*, and a lot of research exists on preventing and resolving collisions because it is undesirable when looking up a singular item. When hashing is applied to image search, collisions are actually desirable, since two similar images should return the same or similar hash-key.

Locality-sensitive hashing (LSH) [70], which can also be considered a form of dimensionality reduction, uses a family of hash functions in order to increase chances of collisions. Similar items are mapped to the same bucket with high probability via the family of hashes. The principle behind this method is that the probability of collision of two data points is closely related to the distance between then, so the larger the distance, the smaller the collision probability. This originally made use of the hamming distance but has since been extended for the Euclidean distance [6].

2.5.3   Other Approaches

Pyramids

If the dataset of images is maintained in pyramidal representations, this hierarchical representation can then be used to overcome the curse of dimensionality and has been proposed for use in image analysis [5]. Pyramid based data structures allow comparisons between images in the dataset at a variety of dimensionalities, so starting at the top of the pyramid, would be the lowest dimensional representation, and working down the pyramid, there is more and more of the data to compare. This allows the range of images to compare with to be narrowed as you go down the pyramid levels. Pyramids can be constructed in several ways, such as Gaussian pyramids or Laplacian pyramids or simply by sub sampling the images to different lower resolutions.

However, the application of pyramids in image search could be limited, particularly for image patches which may be too small to down sample. Furthermore, the loss of detail from sub-sampling may actually hinder the nearest neighbour search process between very similar images, or lead to incorrect matches at finer resolutions. Despite this, image pyramids provides a simple and effective hierarchical scheme for image processing, particularly if processing directly in the native resolution is computationally expensive.

Image Correspondence Search

A common application of nearest neighbour search is finding correspondences within images rather than comparing whole images. This is particularly important for tasks which try to analyse small parts of the images or different objects present within the images, such as applications of patch-based methods. Many approaches do not differentiate between searching within the image and searching a database of images, however the natural structure inherient within images could be exploited to speed up the search process for these tasks. This is what the recently proposed *PatchMatch* algorithm from [13] does for finding corresponding patches between images.

(a) Initialisation   (b) Propagation   (c) Search
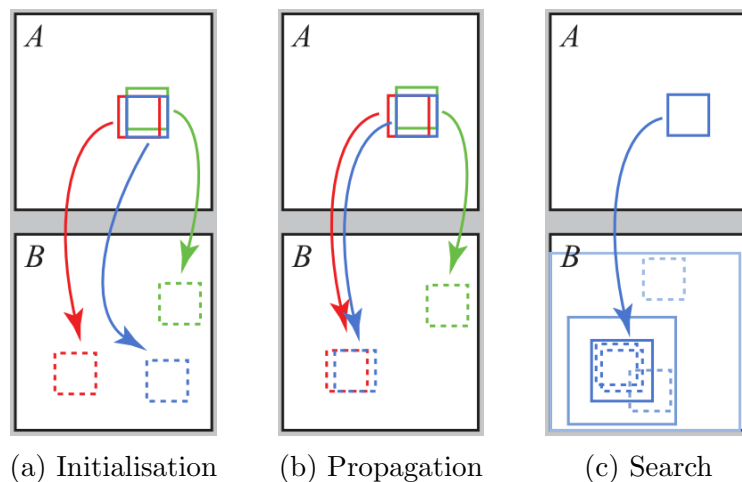
Figure 2.12: The PatchMatch random search algorithm: (a) patches are initially assigned random mappings; (b) the assignments are checked to find the best match, propagating if its good; (c) search randomly for improvements in concentric neighbourhoods.

The PatchMatch algorithm searches spatially in the 2D or 3D image coordinate space rather

than in the dimensional space of the patches to obtain a substantial speed benefit (see Figure 2.12). Random comparisons are made in the image and the position where the best match occurs is then refined by iteratively comparing adjacent patches. This approach has also been extended to a more generalised form for finding the k nearest patches rather than singular correspondences between images [14]. The PatchMatch algorithm is technically an ANN rather than exact kNN approach, since the random initiation does not guaranteed the exact nearest neighbours will be found, however it often works well enough for many common applications of nearest neighbour search such as denoising [14], inpainting [13], superresolution [152], and segmentation [53]. It is also used as part of a "advanced region fill" feature in Adobe Photoshop.

Feature-based Search

Search could also be carried by using a set of features about an image and comparing features rather than pixels. For example, Noah Snavely et al [160] makes use of SIFT features to search objects within a collection of photos. The dimensionality of the features would tend to be much less than the pixel dimensionality of an image, and if the feature is also a location descriptor such is the case of SIFT, it could work well to speed-up image comparison simply by reducing the number of values to compare. Additionally, using features could make search results more meaningful if the choice of features augment the variations of interest in a dataset.

Recently, Ender Konukoglu et al proposed *Neighbourhood Approximation Forests (NAFs)* [97] as a data structure which learns the local neighbourhoods of data according to a given distance labels. Unlike general purpose nearest neighbour data structures, this approach requires annotations or labels for each dataset, but it aims to improve the speed for task-specific nearest neighbour searches. This approach uses random forests which are trained to approximate "out-of-sample" data using the k nearest data items. The training process determines the best image features to use in order to predict the neighbourhood for new unseen data.

2.6   Summary and Conclusion

In this chapter, we reviewed some of the existing image segmentation approaches as well as the main computational methods which are relevant to the work in this thesis. The development of segmentation algorithms have gradually shifted over the years from global intensity-based methods to those which use registration and spatially varying information. The shift in atlas-based methods from globally weighted to locally weighted methods have provided better accuracy and robustness. However, despite improvements in registration methods, the dependency on registration can still pose a problem as registrations errors still commonly occur and can adversely affect segmentation results. The recent developments in using patch-based methods have provided a starting point in reducing dependency on registration and improving robustness to registration errors, and it is from this point where the main contributions of this thesis are introduced.

In the next chapter, a novel kNN patch-based segmentation framework is presented which uses several of the methods reviewed here, in particular kNN data structures and patch-based segmentation based on nonlocal means. The conceptual idea is to extend patch-based methods and further reduce the dependency on registration whilst improving the robustness and accuracy.

Chapter 3

Spatially Aware Patch-based Segmentation

This chapter is based on the following publication:

- Zehan Wang, Robin Wolz, Tong Tong, Daniel Rueckert. Spatially Aware Patch-based Segmentation (SAPS): An Alternative Patch-based Segmentation Framework. *Second International MICCAI Workshop on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging (MCV 2012)*. LNCS Volume 7766, pages 93-103. Springer Heidelberg 2013.

3.1   Introduction

Patch-based segmentation methods compare patches in a local neighbourhood in order to determine the label for each voxel. The underlying assumption is that patches with similar intensities and from similar local neighbourhoods are likely to be the part of the same anatomical structure. Traditionally, this locality is enforced by a sliding search window of a fixed size (typically less than 11^3 voxels). Label fusion then determines spatially-varying weights for each label according to the similarity of the corresponding patches within each voxel neighbourhood.

87

This neighbourhood, when defined as a fixed size search window, imposes a hard restriction on tolerance to any registration errors that occur. Increasing this search window increases the tolerance to registration errors but also increases the computational requirements and may yield patches with similar appearance but from different anatomical structures. Using hierarchical frameworks [61], [191] partly addresses these restrictions, however these approaches still use a fixed search window size for patch selection and comparison.

As an alternative approach, the patch selection process is reformulated so that it is no longer constrained by a fixed search window size. Instead patches are considered from all plausible regions of the image, without fixing or limiting the size of the search volume. To reduce the computational burden, kNN data structures are used so that an exhaustive search of all patches is not necessary. To differentiate between similar patches from different structures, the use of *spatial context* is employed to augment the intensity information for each patch. This means that for each voxel $\mathbf{x}$, a feature vector can be produced consisting of the intensity information for the patch centred on $\mathbf{x}$ in addition to the spatial information for $\mathbf{x}$ (see Figure 3.1). The local neighbourhood for patch comparison is then defined by the k nearest neighbours from each relevant label in the atlas library, in terms of both spatial distance and intensity distance, using the feature vector. This allows the search space of patches to be global whilst maintaining the sense of locality, thus removing the requirement for a fixed search window size to be set. Label fusion is performed by comparing the k nearest patches of each label to derive the final labelling.

For spatial context, the spatial coordinates of each voxel is used in this chapter, but other forms of spatial context will be discussed in later chapters.

3.1.1   Chapter Overview

This chapter introduces a novel patch-based segmentation framework which is used and extended with new approaches for spatial context in subsequent chapters. We start with a reformulation of an existing patch-based method [47] in a kNN framework which also includes spatial context as part of the label fusion process. This is followed by a proposal to use kNN
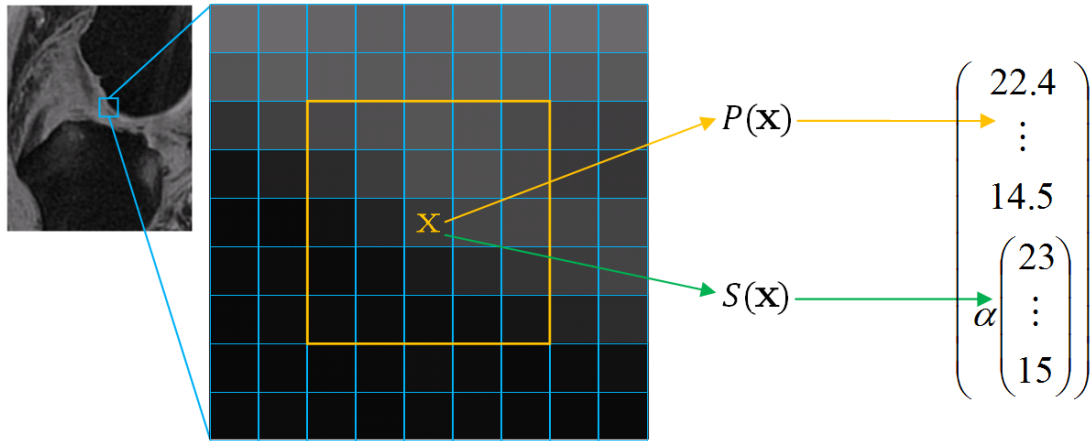
Figure 3.1: The proposed patch-based segmentation framework uses both the intensities of the patch, extracted using a patch extraction operator $P(\mathbf{x})$, and spatial information, provided by a spatial context function $S(\mathbf{x})$. These are concatenated to produce a feature vector, enabling kNN with standard data structures. The spatial weight α controls the contribution and balance of the two components. In this chapter, $S(\mathbf{x})$ is returns the image coordinates of voxel $\mathbf{x}$, but in subsequent chapters, other functions for the spatial context will be explored.

data structures to enable the patch search process to occur without constraining the search window size. Finally, we end with the experiments performed using the proposed framework for hippocampus segmentation in brain MRI and a discussion of the results.

3.2   *k*NN Spatially Aware Label Fusion

Let $P(.)$ be a patch extraction operator, so for each voxel $\mathbf{x}$, $P(\mathbf{x})$ provides the vector of intensities at the patch centered on $\mathbf{x}$ and let $S(.)$ be a spatial context function, such that $S(\mathbf{x})$ produces a vector of the spatial information at voxel $\mathbf{x}$. For this chapter, $S(\mathbf{x})$ is returns the image coordinates of voxel $\mathbf{x}$, but in subsequent chapters, other functions will be explored. For voxel $\mathbf{x}$, then let us denote for each label L, the k most similar patches from the atlas library as a set of voxels $\{\mathbf{y}_{L,i} : i \in 1, ..., k\}$ according the weighted distance of the intensities and the spatial context combined:

$$d_\alpha(\mathbf{x}, \mathbf{y}_{L,i}) = \sqrt{||P(\mathbf{x}) - P(\mathbf{y}_{L,i})||_2^2 + \alpha||S(\mathbf{x}) - S(\mathbf{y}_{L,i})||_2^2} \tag{3.1}$$

where α is the spatial weight which balances the relative importance between the spatial and

intensity components. In practice, the patch intensities and the spatial information for each voxel can be concatenated together, with α pre-applied, into a single vector rather than computing the two components separately. This enables use of existing kNN data structures without re-implementing their functionality, since d_α is then equivalent to the Euclidean distance of the overall vectors. This also enables other distance metrics to be used other than the Euclidean distance. However, on small scale images such as patches, the Euclidean distance is generally sufficient when intensities are normalised and the images are of the same modality.

A weighting w_L for each label L is then determined as follows:

$$w_L(\mathbf{x}) = \sum_{i=1}^{k} w(\mathbf{x}, \mathbf{y}_{L,i}) \tag{3.2}$$

where

$$w(\mathbf{x}, \mathbf{y}) = e^{\dfrac{-\{||P(\mathbf{x}) - P(\mathbf{y})||_2^2 + \alpha||S(\mathbf{x}) - S(\mathbf{y})||_2^2\}}{h^2(\mathbf{x})}} \tag{3.3}$$

and similarly to [47], $h^2(\mathbf{x})$ is determined by the minimum distance with regards to the set of the most similar patches for all labels $\{\mathbf{y}_i\} \in \bigcup_L \{\mathbf{y}_{L,i} : i \in 1, ..., k\}$:

$$h^2(\mathbf{x}) = \min\{||P(\mathbf{x}) - P(\mathbf{y}_i)||_2^2 + \alpha||S(\mathbf{x}) - S(\mathbf{y}_i)||_2^2\} + c \tag{3.4}$$

A small constant c is added to ensure $h^2(\mathbf{x}) \neq 0$. The final label $\hat{L}$ at voxel $\mathbf{x}$ is decided by majority vote, i.e.

$$\hat{L}(\mathbf{x}) = \arg\max_L w_L(\mathbf{x}) \tag{3.5}$$

Overall, the label fusion remains similar to the non-local means approach used in [47], [142] (see Chapter 2.3.5) and follows the same principles. However, the local neighbourhood $\mathcal{N}_\mathbf{x}$ around each voxel $\mathbf{x}$ is now defined by the k nearest patches for each label instead of the search

window size. This not only changes the neighbourhood definition but also means that there is no need to normalise the total weight for each label in order to compare each label's relative weighting. Since the final weighting is the decided by majority vote, using the same number of patches for each label means that the $\arg\max$ can be computed directly. Furthermore, the spatial information for each patch is also considered, and is used to determine the k nearest patches as well as being part of the label fusion process. The weight α is used to balance the contribution of the intensity and spatial information since the Euclidean distance is used and the two types of information are not directly comparable.

3.3 kNN Patch Selection

In general, an exhaustive search of any dataset for the k nearest neighbours would have a computational complexity that is linearly proportional to the size of the dataset and can be quite prohibitive in large datasets. In 3D volumetric images, the number of voxels is usually on the order of millions, and even using the masks to define regions of interest, the number of possible patch comparisons on a global basis would be a substantial computational bottleneck. This is another reason why existing patch-based methods use a small search window size, typically in the region of 11×11×11 = 1331 voxels, whilst patch sizes are typically 7×7×7 or 5×5×5 voxels. This limitation then reduces the tolerance to registration errors, and requires that collectively, any image misalignment must be within a few voxels between the atlases and the target image.

To increase the search volume size without a detrimental impact to the search speed, efficient kNN search data structures are required. In this thesis, a ball tree [124] (see Chapter 2.5.1) is used - these provide much better search performance than kd trees or brute force searches for high dimensional data [100]. Ball trees are metric trees which use a given distance metric to partition the data so that only a small part of the data need to be queried. The distance metric used must obey the triangle inequality for metric trees to work correctly. Since Euclidean distances are used in both patch based comparisons and atlas selection, and this obeys the

triangle inequality, ball trees can be used to provide the results to kNN queries.

3.3.1   Data Structure Construction

In principle all patches could be stored in a single tree, however, the memory requirements would grow prohibitively large as the number of atlases increases. In addition, search performances cease to obey the theoretical $\mathcal{O}(n \log n)$ computational complexity as the overhead for random memory access and tree traversal becomes more expensive as the memory usage increases. Furthermore, this also makes it difficult to use atlas selection to select the most appropriate atlases for label fusion. Instead, trees are constructed on an atlas-by-atlas basis and for each separate region of interest (ROI). This enables atlas selection to be performed on a region-by-region basis, providing flexibility in atlas usage whilst reducing memory requirements at the same time. This is particularly desirable as atlas selection has been demonstrated to be an important process in multi-atlas segmentation [140], [4] and also means that the exhaustive (and computationally expensive) search of the whole atlas library is not required. Additionally, patches are extracted and sorted by their label and a single tree is constructed for each label. This allows the k nearest patches for each label to be found in a straight forward way and organises the data structures in a modular and flexible fashion. Since the label fusion process uses the k nearest patches for each label to determine a weighting, it is easier to accomplish when the patches are sorted by their label. Furthermore, it enables patch search to be performed in parallel for each label and for each atlas.

The data structure construction can be considered analogous to the "training" process for many other machine learning methods and can be performed offline. However, the patch extraction and ball tree construction times are sufficiently fast that it can all be performed on-the-fly at run-time.
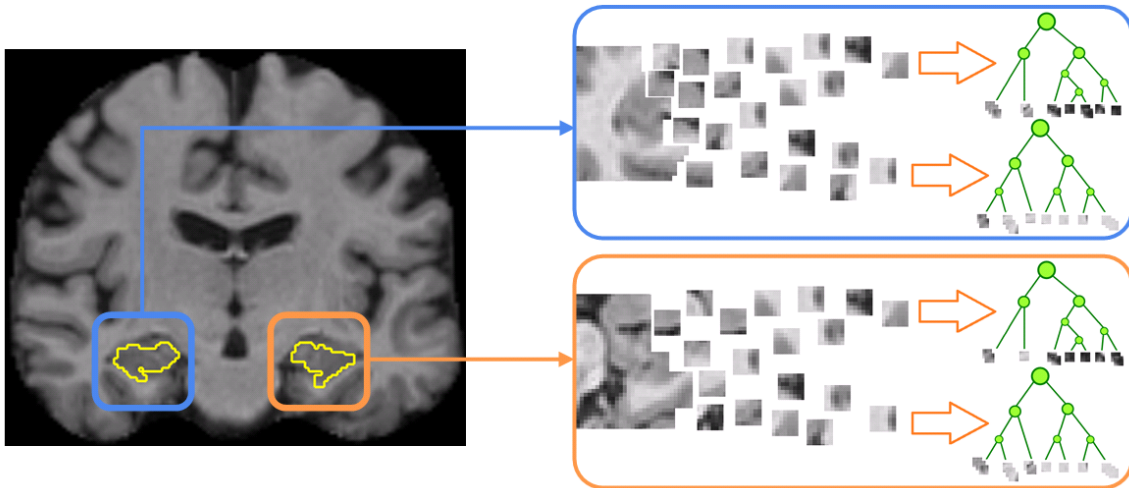
Figure 3.2: Example: ball tree construction using patches from the hippocampal regions. For each atlas, a tree is created for each label (including the background label) in each hippocampal region.

3.3.2 Atlas Selection and Patch Search

For each ROI that requires labelling, the nearest N atlases are found for each region by comparing them using a distance function $d(.)$. Performing atlas selection on a regional basis also allows for more granularity in the choice of atlases than doing so on a global image level. This relies on a globally defined mask for the regions of interest for all images, then any similarity or distance metric could be used for the region (See Chapter 2.4). For example, SSD could be used, then for each corresponding pixel $\mathbf{x}_i$ in the target image I_t, and atlas image I_a and a common ROI mask of size n, this distance is:

$$d(I_t, I_a) = \sum_{i=1}^{n} \left(I_t(\mathbf{x}_i) - I_a(\mathbf{x}_i)\right)^2 \tag{3.6}$$

This can be normalised by the number of pixels n to compare between different sized ROIs. However, when a common ROI is defined for all images, normalisation is not required.

In MALP, the choice of distance function for atlas selection often depends on the application. For the proposed framework, the distance metric used for atlas selection should be related to that used in the patch selection and label fusion. Since Euclidean distance is usually used for patch selection, then it follows that it or SSD would also be used for atlas selection, provided

(a) Atlas masks   (b) Union   (c) Dilation
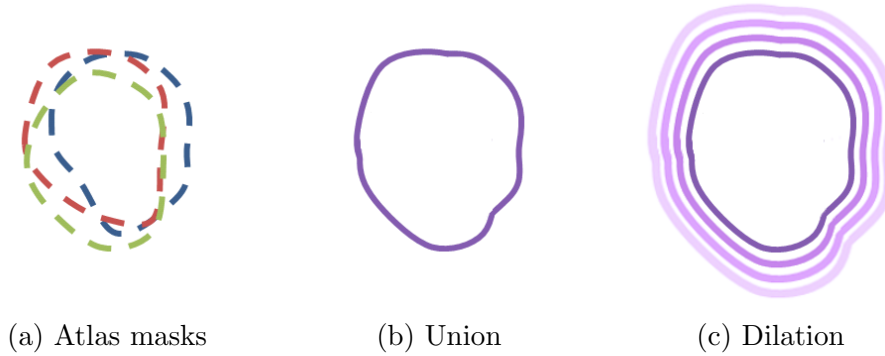
Figure 3.3: Creating a common ROI mask: take binary masks from each of the atlases and use the union of the masks as the base ROI mask. The base mask can then be dilated to enlarge it and allow for more variability.

that the image intensities are normalised. SSD provides higher sensitivity than the Euclidean distance to the image differences.

If images are all aligned, then the regional masks can be created by taking the unions of the labels from the atlases and dilating the results. This mask is used to narrow the search space and restrict search to valid areas where a label might appear. The mask needs to be large enough to allow for possible variations in anatomical variability, whilst being minimal in size so that the kNN search process can be more efficient. The mask can be dilated multiple times to enlarge it, if required.

Using a limited selection of the best subjects from the atlas library has been shown to provide more effective segmentation results [4] in MALP, but for the proposed framework, this is largely for easing the computational burden of the kNN search. This is due to the relationship for selecting the most appropriate "images" for label fusion being at patch level in this approach rather than the atlas level in traditional MALP. The k nearest patches are more likely to be chosen from similar atlases but it is also plausible that there are locally similar regions in atlases which would appear dissimilar with a global similarity measure. For this reason, multiple atlases will still be required, with more atlases likely to provide better results. For particularly large atlas libraries, kNN data structures could also be used for atlas selection, as well as patch selection, to speed up the search process.

After selecting the N nearest atlases, the corresponding kNN data structure for those atlas

regions are then used for patch-based segmentation. Patches are extracted from the target image using the same method as for the atlases, using the same patch size and spatial weight α. For each voxel, the k nearest patches with regards to both intensity and spatial information are retrieved for each label by querying the respective kNN data structures. These are then used for label fusion, as described above in section 3.2.

3.4 Framework Summary

In summary, the proposed framework can be split into two stages, the training or construction stage and the testing or run-time stage, as shown below in Figure 3.4. During the former stage, kNN data structures are constructed for extracted patches in each region of each atlas and for each label. In the latter stage, the k nearest patches are retrieved for each voxel and for each possible label, after performing atlas selection to select the N nearest atlases. Label fusion is used to determine the overall labelling for each voxel.

The label fusion process uses the k nearest patches with regards to both intensity information and spatial information, using a Gaussian function on their distance to derive weights for each label. The overall consensus label is then established by a weighted voting according to the derived weights. This is performed for each voxel in the image and can be computed independently and in parallel.

3.5 Application to ADNI Brain MRI Dataset

For evaluation, the proposed framework and methods are applied to hippocampus segmentation in brain MRI. The hippocampus is a key structure which provides cognitive functionality, and its volume can be used as an early biomarker for Alzheimer's disease [89]. In order to do so, accurate segmentations are required to correctly identify atrophy in the hippocampal volume [15]. Images from the Alzheimers Disease Neuroimaging Initiative (ADNI) database[1] were used

---

[1] www.loni.ucla.edu/ADNI

(a) Training/Construction   (b) Testing/Run-Time
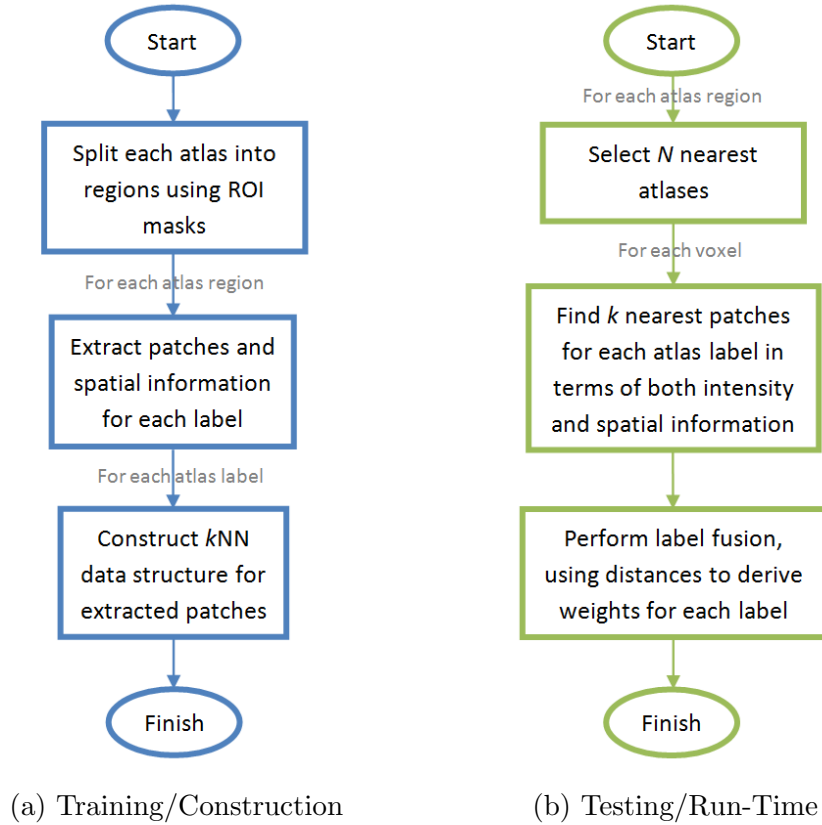
Figure 3.4: The two stages of the proposed framework. In the first stage in (a), patches and spatial information are extracted for each atlas region and kNN data structures are constructed for each label. In the second stage in (b), atlas selection and patch selection is performed, with resulting segmentation derived from the label fusion process.

for validation.

3.5.1   ADNI Dataset

The dataset used for evaluation are images from 202 randomly selected subjects from the ADNI database. These were acquired using different scanners from 68 normal control subjects, 93 subjects with mild cognitive impairment (MCI), and 41 subjects with Alzheimer's disease (AD). Images were prepared by the standard ADNI pipeline [90], where image correction methods were applied for gradient non-linearity and intensity non-uniformity on an individual case-by-case basis. Reference segmentations were obtained semi-automatically using a commercially available high dimensional brain mapping tool (Medtronic Surgical Navitgation Technologies, Louisville, CO) by propagating 60 manually labelled images.

3.5.2 Implementation

A leave-one-out validation strategy was applied where each image was segmented in turn using the remaining dataset as the atlas database. A patch size of $7\times7\times7$ was used whilst experiments were performed with the number of atlases used, N, the spatial weights, α, and the number of nearest neighbours for each patch, k. For spatial context, the spatial coordinates of each voxel were used. For atlas selection, sum of squared differences was used as the distance metric.

The main framework was implemented in Python and Cython using open source modules such as NumPy, SciPy and SciKit-learn. During testing, the computation time is around 10 minutes for each image using 8 cores clocked at 2.67GHz each and 8GB RAM when using 20 atlases and using the 100 nearest neighbours. The patch extraction and construction of trees for each atlas requires less than 5 seconds. The trees for each atlas require around 150MB if using 64-bit double precision for the data, but they do not all need to be loaded in memory at the same time. The main bottleneck for speed is the kNN lookup, however the choice of the best kNN data structure, still remains an open question in computer science.

Pre-processing

Atlases are all registered to the MNI-ICBM152 template space using affine registration and intensities are normalised using the method proposed by Nyúl and Udupa [123]. Two ROI masks for atlas selection and patch selection were created based around the left and the right hippocampus by taking the union of these labels from all atlases and dilating the result. The atlases are also denoised to improve robustness. Total Variation (TV) denoising was used as a quick and easy-to-apply method which can be effective in regularizing images without smoothing boundaries and edges [38]. Figure 3.5 illustrates the effect of using these methods.

Summary of Parameters

The parameters and specifications used are summarised below, in Table 3.1:

(a) Original   (b) Normalised   (c) Denoised
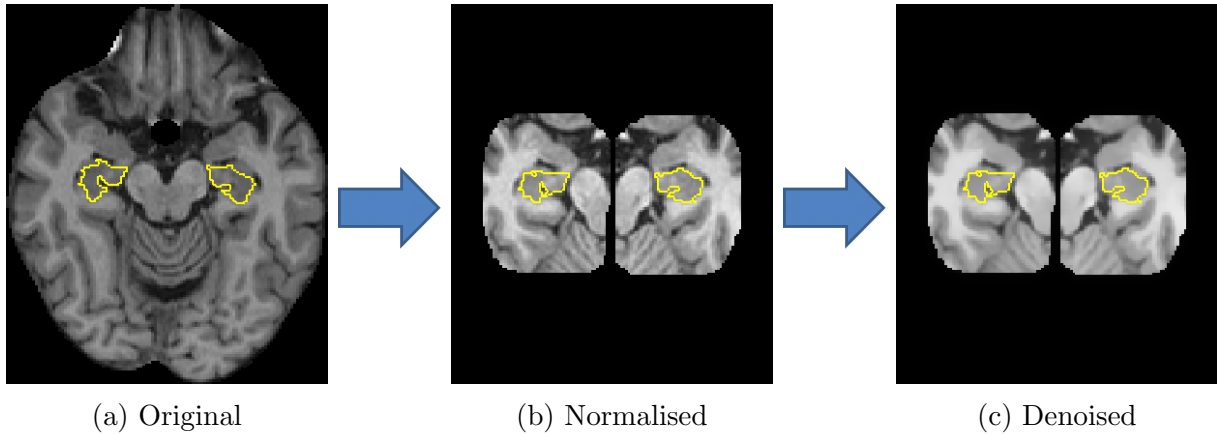
Figure 3.5: Examples from the pre-processing pipeline: (a) shows the original image, skull-stripped, in its native space. (b) shows the result of intensity normalisation after the ROI masks has been applied and having been registered to the MNI-ICBM152 template space. (c) shows the result of performing TV denoising. The hippocampus is outlined in yellow.

Table 3.1: Table of parameters for the ADNI dataset.

Method/Description	Parameter(s)
Registration Type	Affine, 12 parameters
Intensity Range	$[0, 68]$
TV Denoising[a]	weight=5
Patch Size	$7 \times 7 \times 7$
ROI mask size	50,015 voxels overall

---

[a]scikit-image 0.7 implementation

3.5.3   Experiments and Results

Effect of the Spatial Weight, α

Experiments using several values for spatial weights, α, showed that using spatial information to provide a soft-weighting has a significant impact on the segmentation accuracy (see Figures 3.6 and Table 3.2). Comparing results of using even a small weight, $\alpha=3$, with not using any spatial context, $\alpha=0$, yields p-values of 4.6×10^{-7}, 6.3×10^{-5}, and 5.8×10^{-7} for the left, right and overall hippocampus respectively with Welch's paired two sample T-test. Example segmentations, shown in Figure 3.7, highlights the problem with performing global patch comparison, where similar patches from different (background) structures yield greater weighting during label fusion for homogeneous regions of the hippocampus. The inclusion of spatial context with a suitable weighting provides a remedy for this problem.
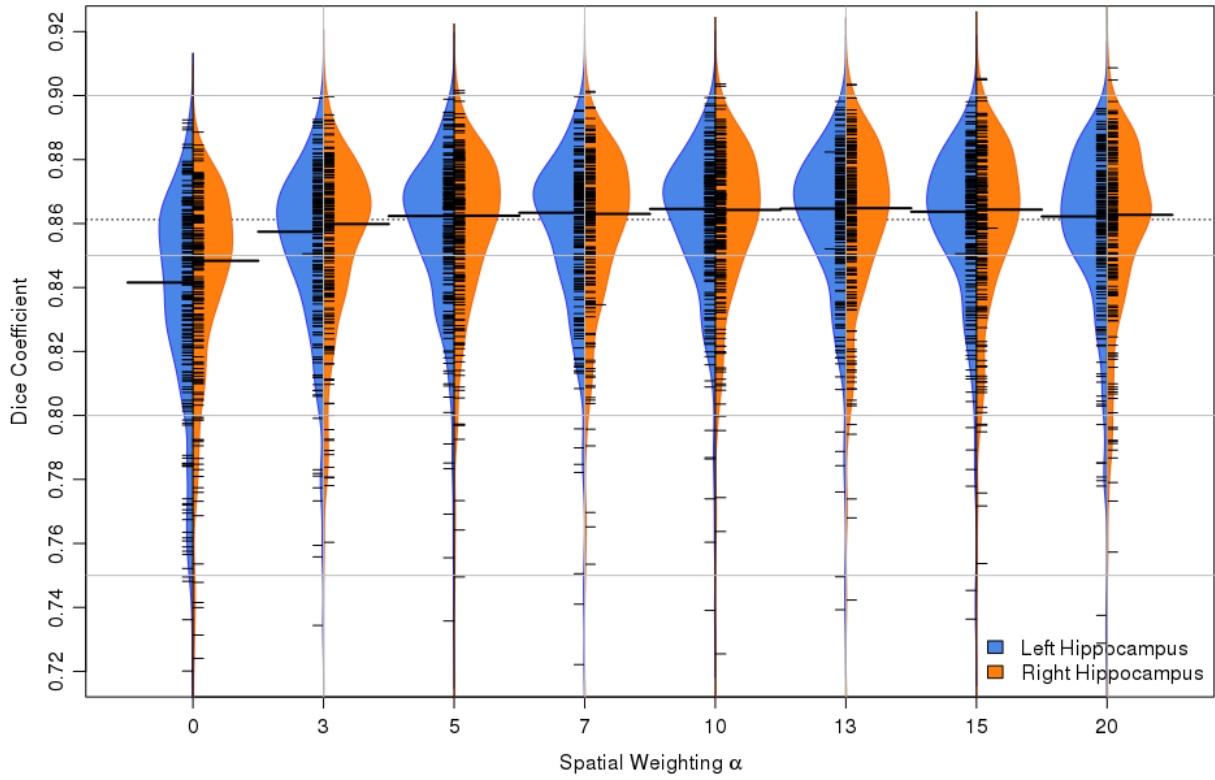
(a)
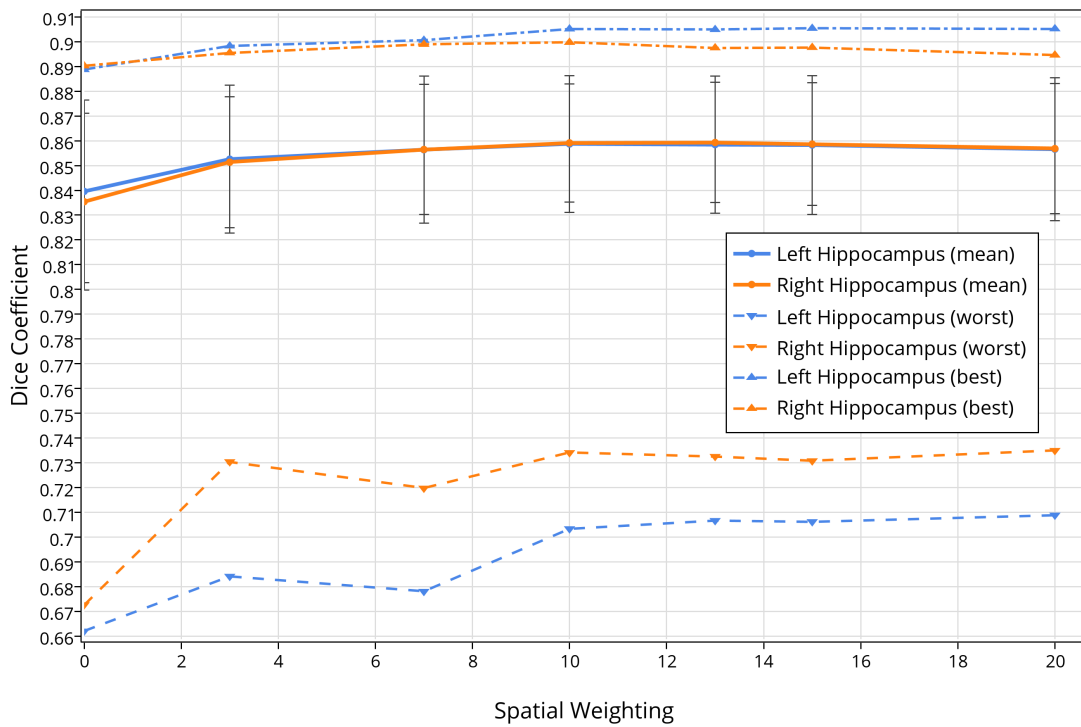
(b)

Figure 3.6: Comparison of Dice coefficients for a range of spatial weighting values α with N=20 and k=64. (a) is a beanplot where large thick lines indicate medians, dotted line indicates median across all α values. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean. (b) provides the line plot for the mean, worst case and best case results, whist error bars show the standard deviation.

Table 3.2: Dice coefficients, shown as mean (median) [worst, best], for the hippocampus (HC) when using different spatial weights α, with k=64 and number of atlases N=20. Best values are shown in red.

α	Left HC	Right HC	Overall
0	0.834 (0.842) $\pm$0.037	0.839 (0.848) $\pm$0.037	0.836 (0.844) $\pm$0.033
3	0.851 (0.857) $\pm$0.027	0.853 (0.860) $\pm$0.030	0.852 (0.857) $\pm$0.025
5	0.856 (0.862) $\pm$0.025	0.857 (0.862) $\pm$0.029	0.856 (0.862) $\pm$0.024
7	0.856 (0.863) $\pm$0.026	0.857 (0.863) $\pm$0.030	0.857 (0.862) $\pm$0.025
10	0.859 (0.865) $\pm$0.024	0.859 (0.864) $\pm$0.027	0.859 (0.863) $\pm$0.023
13	0.860 (0.865) $\pm$0.024	0.859 (0.865) $\pm$0.028	0.859 (0.864) $\pm$0.023
15	0.859 (0.864) $\pm$0.025	0.859 (0.864) $\pm$0.028	0.859 (0.863) $\pm$0.024
20	0.857 (0.862) $\pm$0.027	0.857 (0.863) $\pm$0.029	0.857 (0.862) $\pm$0.025
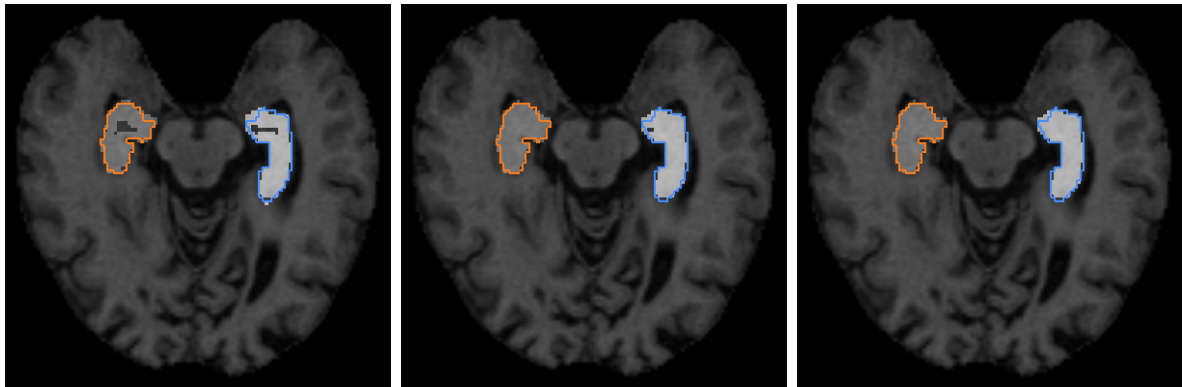
(a) α=0 [Dice: 0.863 0.868]   (b) α=5 [Dice: 0.893 0.902]   (c) α=13 [Dice: 0.894 0.903]

Figure 3.7: Example of results with different spatial context for the hippocampus, with k=64 and number of atlases N=20. Segmentations are filled in grey and the outlines of the ground truth shown in blue and orange for the left and right hippocampus respectively. In (a), the lack of spatial context leads to incorrect labelling, seen as "holes" in the hippocampus segmentation. As spatial context is introduced and the weights increased in (b) and (c), the holes disappear as similar patches from different structures can be better differentiated.

Additionally, the distribution of the results as seen in the beanplots in Figure 3.6 also suggests that both the accuracy and consistency of the results increases significantly when we use spatial information. The values attempted suggests that segmentation accuracy peaks at α=13. If the spatial weighting is too high, there is a detrimental effect on the segmentation accuracy as this soft-weighting becomes too restrictive when comparing patch intensities. As α increases, the behaviour becomes more like traditional multi-atlas segmentation approaches where one-to-one label fusion occurs.

Further analysis of the effect altering the spatial weighting α is presented in Figure 3.8 and Table 3.3. Here it can be seen that the most optimal values for α is between α=10 and α=13.

Table 3.3: Mean sensitivity, specificity, precision and F1-Score for different values of spatial weighting α, with k=64 and number of atlases N=20. Best values are shown in red.

α	Sensitivity	Specificity	Precision	F1-Score
0	0.8534	0.9774	0.8234	0.8366
3	0.8568	0.9797	0.8503	0.8523
5	0.8626	0.9795	0.8530	0.8566
7	0.8661	0.9789	0.8504	0.8569
10	0.8662	0.9794	0.8557	0.8597
13	0.8663	0.9792	0.8556	0.8596
15	0.8663	0.9790	0.8551	0.8593
20	0.8651	0.9785	0.8530	0.8576
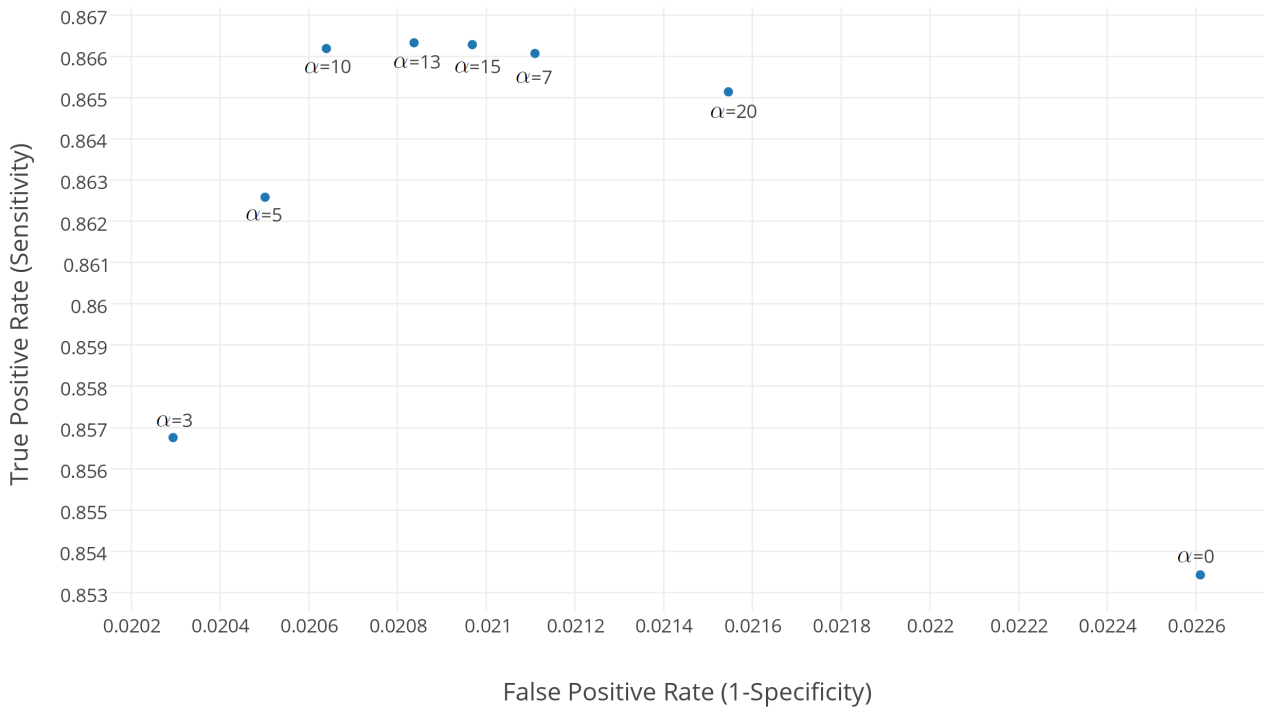
Figure 3.8: ROC analysis, comparing true positive rates (sensitivity) to false positive rates (1-specificity) for the different values of spatial weighting α.

Effect of the Number of Nearest Patches and Atlases Used

With the spatial weight fixed at $\alpha = 13$, experiments were conducted using a range of values for the number of patches, k, as well as the number of atlases, N. k is dependent on N as using more atlases would present a bigger selection of patches to choose from and, as seen in Figure 3.9, this means the optimal k value differs for the different N values.
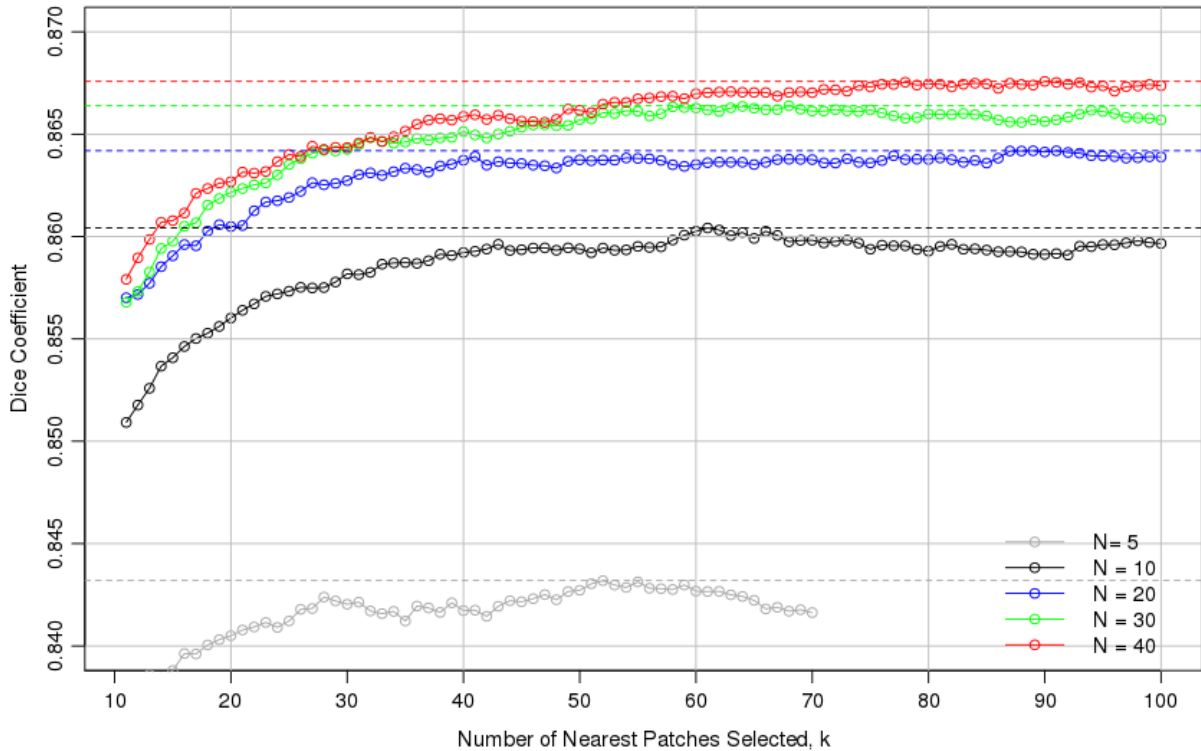
Figure 3.9: Median Dice coefficients for the whole hippocampus whilst using a range of k values with different N values.

Generally, the accuracy increases as k increases, but reaches a limit after $k > 60$. There is an increase in computational cost as k increases as more comparisons must be made in the kNN data structures, so it is most computationally optimal to select the lowest k value that provides the desired segmentation accuracy.

An increase in the number of atlases used generally increases segmentation accuracy, but the gain in accuracy after $N > 10$ becomes more marginal, particularly with changes in the distribution of the results. Given that the computational cost increases linearly with the number atlases used whilst the increase in segmentation appears to be logarithmic, the results suggest there is both a practical limit and a theoretical limit on the number of atlases to use to and

that there is a trade-off between the computational time spent and the accuracy gained.
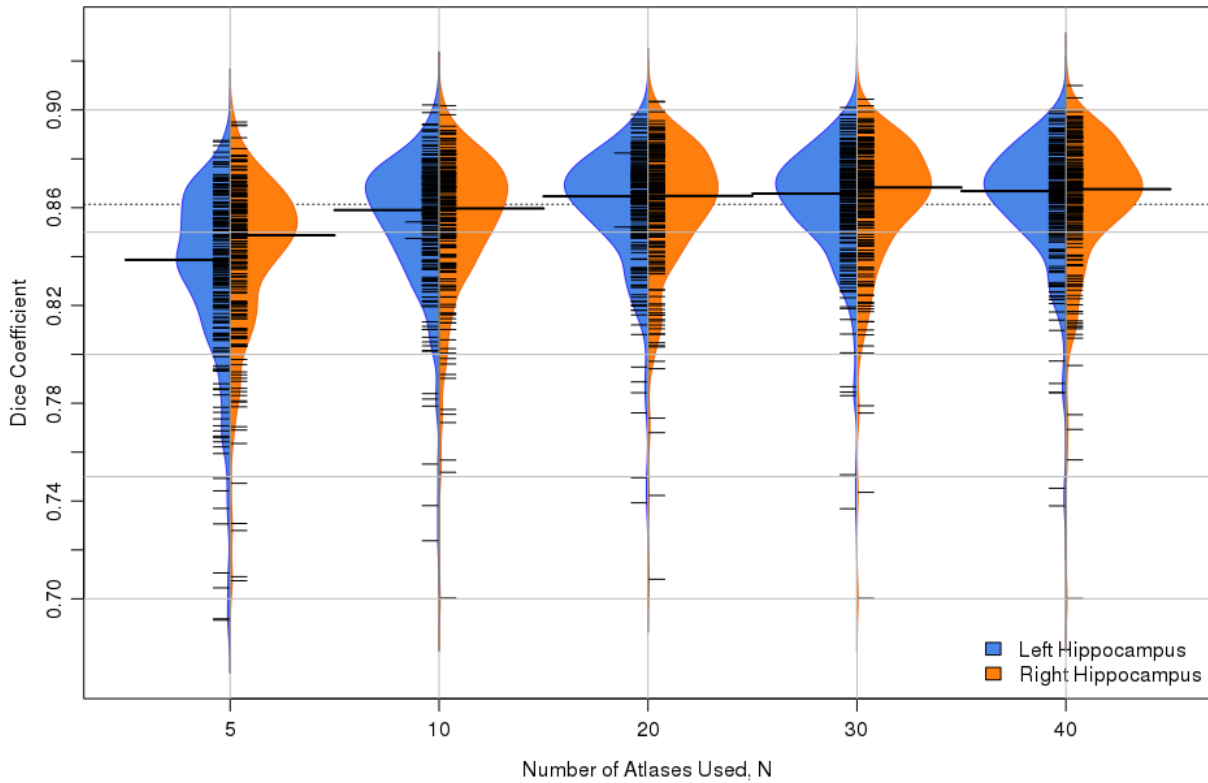
Figure 3.10: Beanplot showing overall Dice coefficients distributions for a range of N values with k=64. Large thick lines indicate medians, dotted line indicates median across all k values. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean.

Table 3.4: Dice coefficients, shown as mean (median) $\pm$ standard deviation, for the hippocampus (HC) when using different number of atlases, N, with k=64. Best values are show in red.

N	Left HC	Right HC	Overall
5	0.832 (0.839) $\pm$0.036	0.841 (0.849) $\pm$0.032	0.837 (0.842) $\pm$0.030
10	0.854 (0.859) $\pm$0.027	0.854 (0.860) $\pm$0.029	0.854 (0.860) $\pm$0.024
20	0.860 (0.865) $\pm$0.024	0.859 (0.865) $\pm$0.028	0.859 (0.864) $\pm$0.023
30	0.861 (0.866) $\pm$0.024	0.862 (0.868) $\pm$0.027	0.862 (0.866) $\pm$0.023
40	0.863 (0.867) $\pm$0.025	0.863 (0.868) $\pm$0.027	0.863 (0.867) $\pm$0.023

Effect of the Number of Total Available Atlases

To compare the effect of dataset size, experiments were conducted using random subsets of the total 202 available atlases, drawn independently for each case of 25, 50 and 100 total number of atlases used. Other parameters were fixed with N=20, k=64 and α=13. The results for each case can be seen in Figure 3.11 and Table 3.5.

Table 3.5: Dice coefficients, shown as mean (median) ± standard deviation, for the hippocampus (HC) when using different number of total atlases, with N=20, k=64 and α=13.

Total No. Atlases	Left HC	Right HC	Overall
25	0.843 (0.852) ±0.034	0.845 (0.850) ±0.034	0.844 (0.853) ±0.031
50	0.844 (0.851) ±0.036	0.851 (0.860) ±0.028	0.848 (0.853) ±0.029
100	0.855 (0.858) ±0.027	0.856 (0.860) ±0.025	0.856 (0.860) ±0.023
202	0.860 (0.865) ±0.024	0.859 (0.865) ±0.028	0.859 (0.864) ±0.023
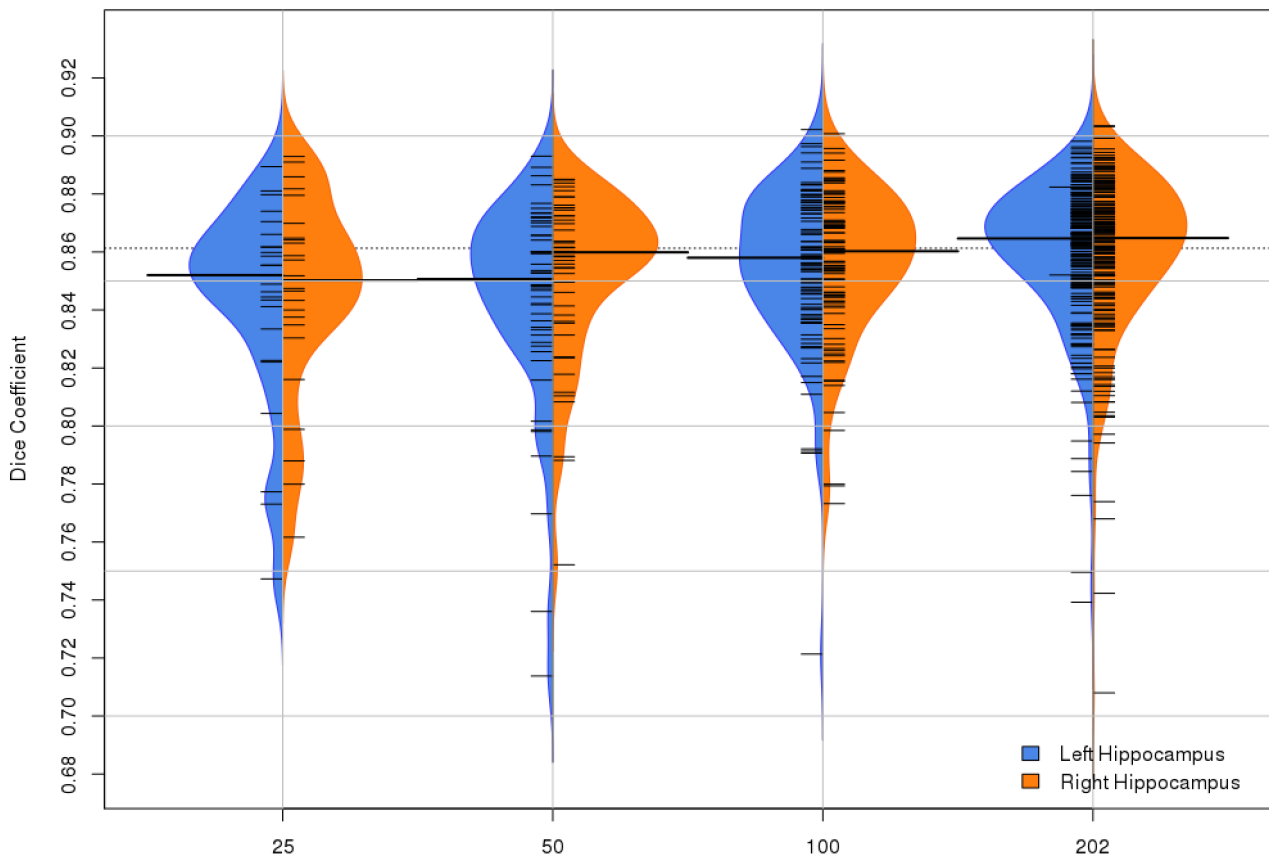
Figure 3.11: Dice coefficients distributions for results obtained for different sized atlas libraries with N=20, k=64 and α=13. Large thick lines indicate medians, dotted line indicates median across both datasets. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean.

When reducing the total number of atlases available, the overall performance and consistency suffers as can be expected given the nearest neighbour approach. With a reduced dataset, the similarity of the selected atlases is likely not to be as high and the number of appropriate patches for label fusion may not be sufficient.

Effect of Denoising

Comparing results from using non-denoised training data to those from using denoised training data, it can be seen that using denoised training data provides an improvement to the median segmentation accuracy (see Figure 3.12). Further to this, the range of the results is smaller with a more favourable distribution when using denoised training data and using Welch's paired two sample T-test provides p-values of 1.6×10^{-6}, 5.6×10^{-2} and 2.5×10^{-4} for the left, right and overall hippocampus respectively. This suggests that denoising generally improves accuracy and robustness of the framework, however given that the differences for the right hippocampus is not as large as for the left, it seems that denoising does not always provide a significantly large improvement and its use may be dependent on the application.
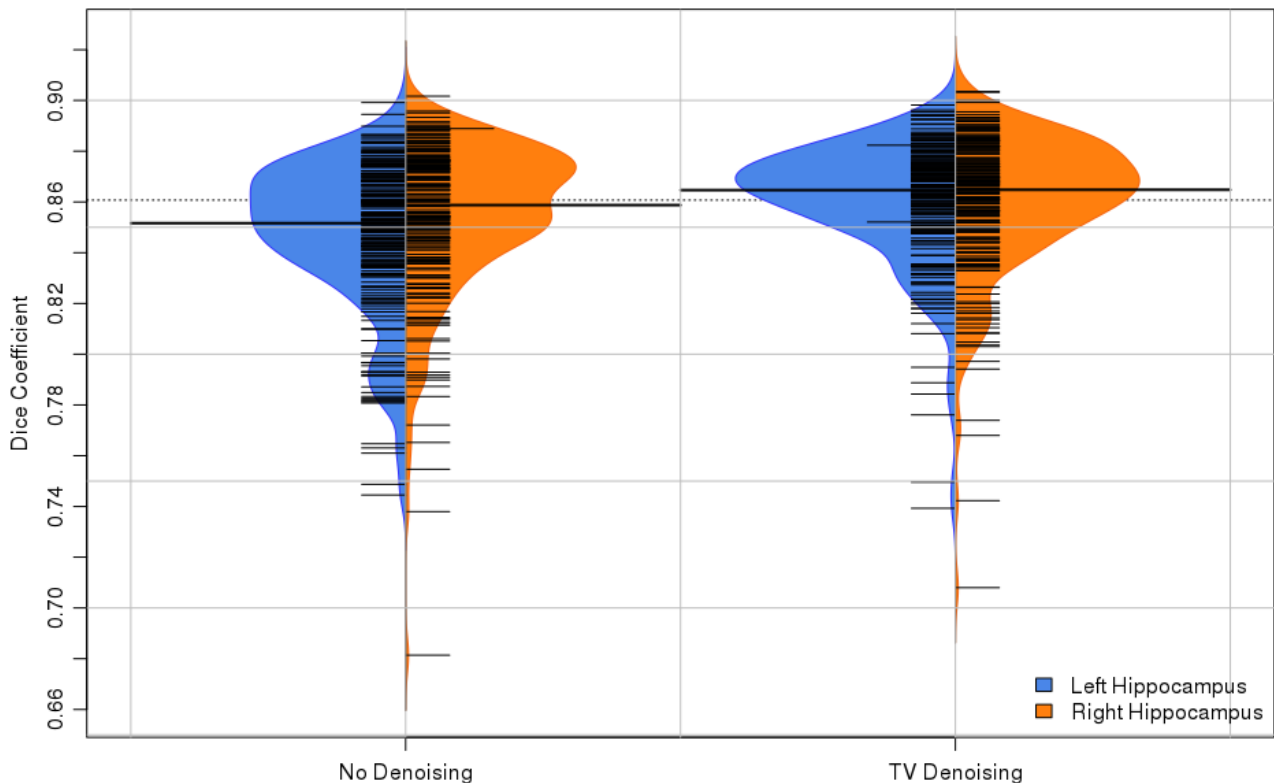
Figure 3.12: Dice coefficients distributions for results using denoised and non-denoised training data with N=20, k=64, α=13. Large thick lines indicate medians, dotted line indicates median across both datasets. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean.

Comparison of Results to an Existing Method

Finally, with the same dataset of ADNI images (and same ground truth segmentations), the results obtained by the proposed approach are compared to that using the current state-of-the-art patch-based method, as described by Coupé et al in [47], which does not use spatial context and has a fixed search window size. Parameters which were suggested as being most optimal for hippocampus segmentation were used, with patch sizes fixed at 7×7×7 and the search window size at 9×9×9. Results were compared for the case when the number of atlases, N, is set at 10.

It can be seen in Table 3.6 and Figure 3.13, that the proposed method generally outperforms the existing method and is more robust. When no spatial context is used with α=0, the label fusion of the two methods become very similar (see Table 3.2), and can be considered approximately equivalent with the approach from [47] using the whole ROI for patch comparison rather than a limited search window. This highlights the problem of using a search window size that is too large, and which then leads to comparisons with similar patches but from different structures, thus providing worse segmentation accuracy.

Table 3.6: Dice coefficients shown as mean (median) [worst, best] for the hippocampus (HC) comparing different methods with the number of atlases N=10 for both. Proposed method uses k=64, α=13 as its other parameters. Best values are highlighted in red. p-values of 1.5×10^{-5}, 4.6×10^{-3} and 1.1×10^{-4} were obtained for the left, right and overall hippocampus respectively with Welch's paired two sample t-test.

	Left HC	Right HC	Overall
[47]	0.842 (0.847) ±0.032	0.846 (0.851) ±0.034	0.844 (0.845) ±0.030
Proposed	0.854 (0.859) ±0.027	0.854 (0.860) ±0.029	0.854 (0.860) ±0.024

Applying Welch's paired two sample t-test on these results gave p-values of 1.5×10^{-5}, 4.6×10^{-3} and 1.1×10^{-4} for the left, right and overall hippocampus respectively. Additionally, there is also a 0.05 decrease in the standard deviation of the results compared to that of [47].

When comparing results across the different disease status of the subjects, we see in Table 3.7 and Figure 3.14 that the proposed approach improves upon that of [47] for all categories. Out
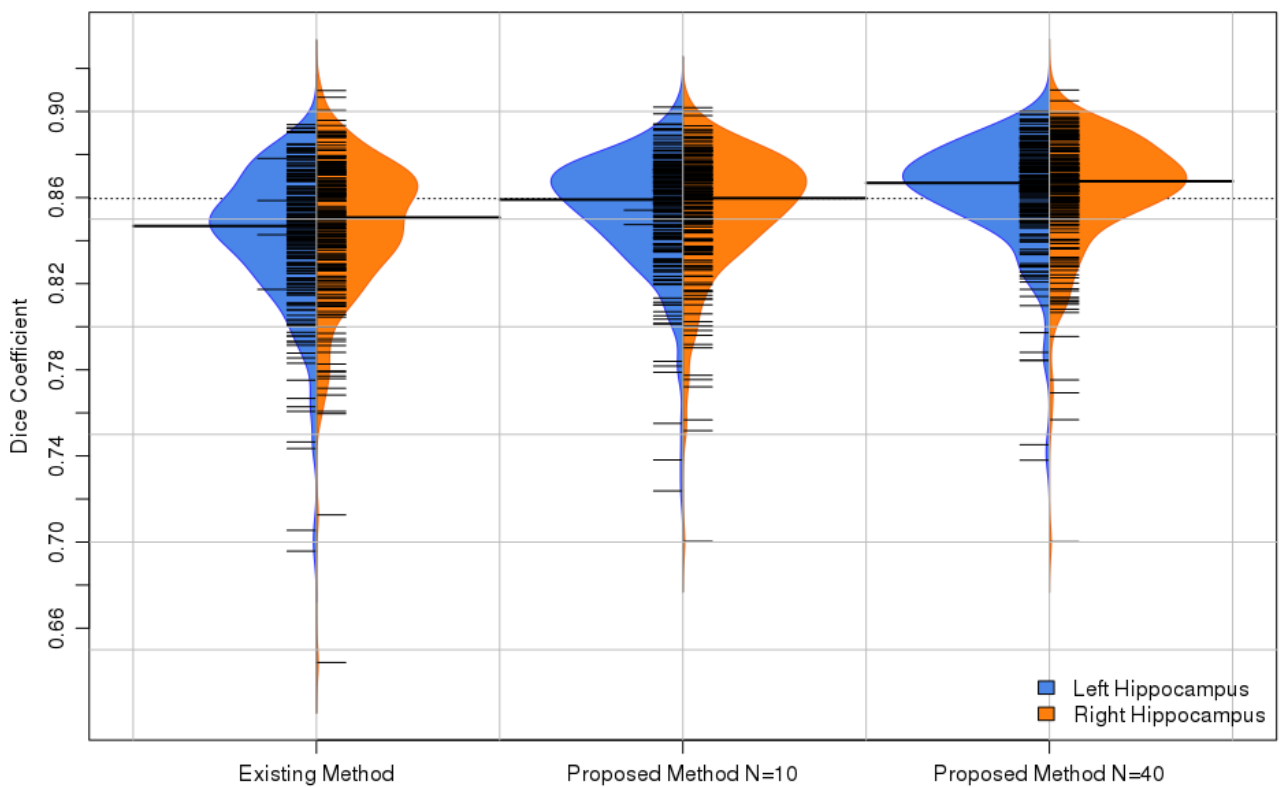
Figure 3.13: Comparison of Dice coefficients distributions for the proposed method versus [47] with k=64, α=13. Large thick lines indicate medians, dotted line indicates median across both datasets. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean. Results of the proposed method for $N = 40$ are shown for further comparison.

Table 3.7: Dice coefficients shown as mean (median) $\pm$ standard deviation for the hippocampus across disease status type. The proposed method uses $k=64$, $\alpha=13$ as its other parameters. Results of the proposed method for $N=40$ are shown for further comparison.

	Control	stable MCI
[47] ($N=10$)	0.855 (0.856) $\pm$0.022	0.850 (0.853) $\pm$0.023
Proposed ($N=10$)	0.861 (0.864) $\pm$0.020	0.859 (0.863) $\pm$0.019
Proposed ($N=40$)	0.868 (0.870) $\pm$0.019	0.866 (0.868) $\pm$0.018

	progressive MCI	AD
[47] ($N=10$)	0.835 (0.836) $\pm$0.033	0.828 (0.838) $\pm$0.034
Proposed ($N=10$)	0.845 (0.851) $\pm$0.030	0.847 (0.852) $\pm$0.025
Proposed ($N=40$)	0.854 (0.860) $\pm$0.029	0.859 (0.867) $\pm$0.024
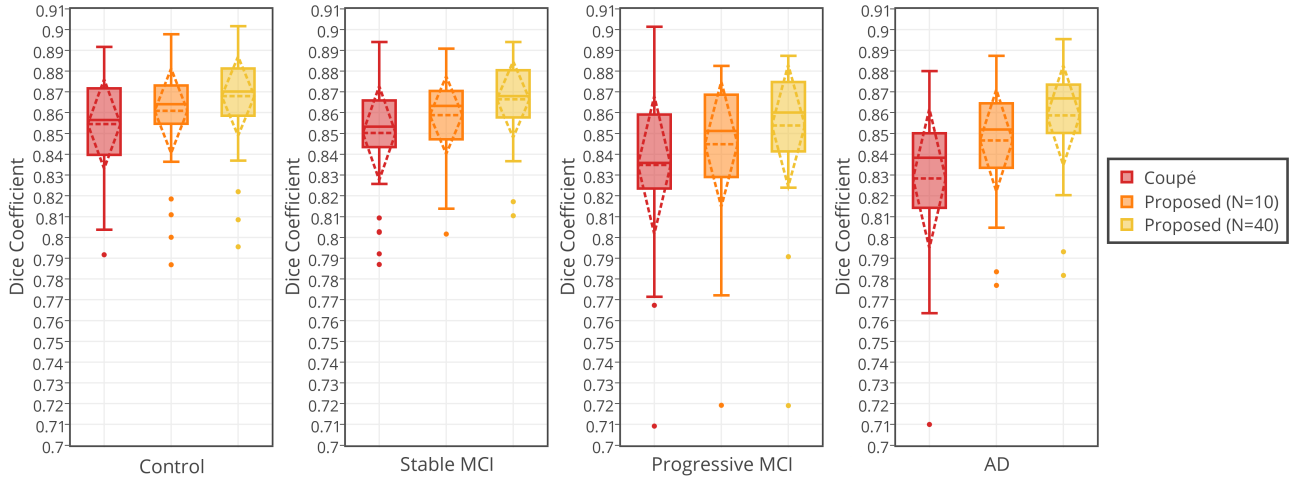
Figure 3.14: Comparison of segmentation accuracy for different disease status. Solid line indicates the median, the dashed line indicates the mean and standard deviation is shown by the dashed diamond.

of the 202 total subjects, 68 were control, 49 had stable MCI, 44 had progressive MCI and 41 had AD.

3.5.4   Discussion

The experiments indicate that the use of spatial context, when combined with using a global kNN approach without limiting the search volume size, improves segmentation results in terms of both accuracy and consistency. This not only provides validation for the proposed framework but also highlights that there is a problem with the reliance on registration accuracy for segmentation and that existing patch-based methods are unable to fully accommodate for the anatomical variability between subjects. Overall, the results demonstrate potential in the

Reference Segmentation   Best Subject - Dice: 0.910

Reference Segmentation   Median Subject - Dice: 0.867

Reference Segmentation   Worst Subject - Dice: 0.701
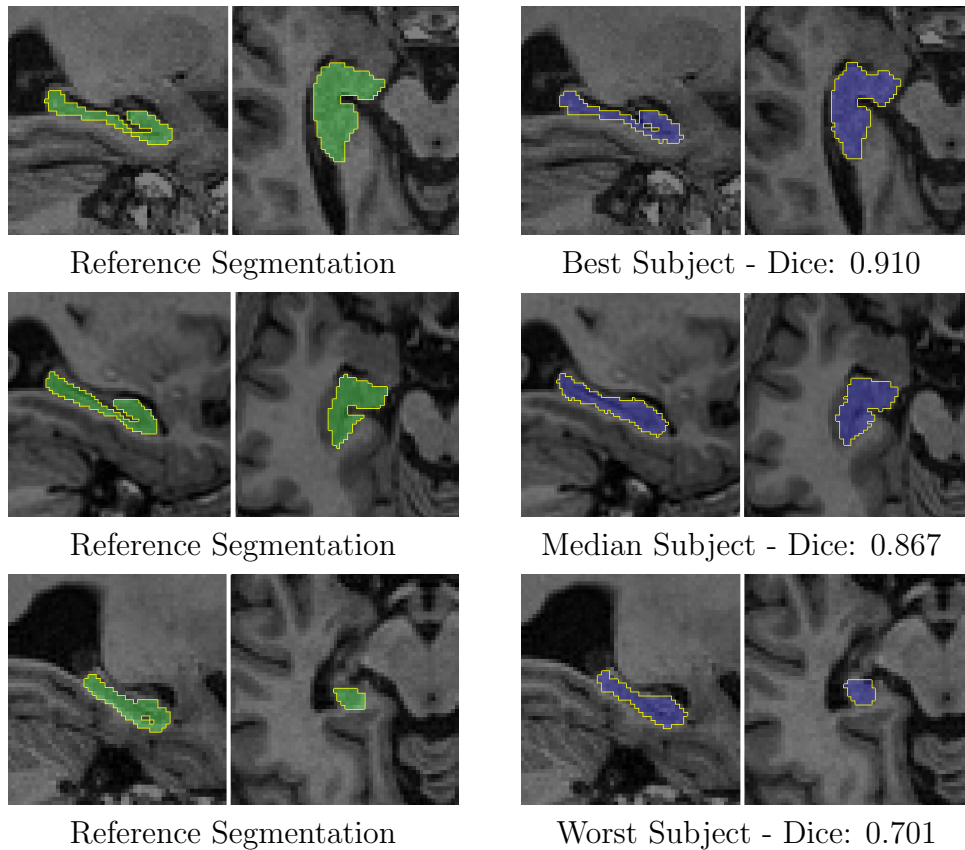
Figure 3.15: Example segmentations of the right hippocampus with parameters N=40, k=79, α=13

proposed approach to address to this problem, and provide accurate and robust segmentations.

One computational side effect of the proposed framework in extracting the patches from the images and storing within kNN data structures is that it allows the voxel labelling process to be easily exploited for parallel execution since voxels can be labelled independently. Additionally, the framework allows a trade-off between segmentation accuracy and speed through adjusting the number of atlases used. In general there is a logarithmic trend for improvements in segmentation accuracy with increases in the number of atlases, but a linear relationship for computational time with the number of atlases. Using patches from half as many atlases, would allow the segmentation to be completed in half as much time. At the lowest limit tested, using 5 atlases was still able to yield a median Dice coefficient of 0.842 for the whole hippocampus.

The findings on changing the number of atlases agree with those presented in [47] which also suggested that increasing the number of atlases N yields diminishing gains in accuracy as N increases following a logarithmic pattern. This is somewhat contrary to experiments with

traditional multi-atlas approaches [4], which suggested there is often an optimal number of atlases which is much lower than the total number within the atlas library and that increasing the number of atlases past this point would worsen the results. The reason for this is due to the non-local label fusion which does not simply use patches from all atlases to form a consensus labelling, but rather only using the most likely patches which are then weighted according to their similarities. Increasing the number of atlases for patch-based segmentation provides additional examples for which to select patches from, and because patches are weighted by similarity, it is not likely to worse the segmentation accuracy. This behaviour provides an additional advantage over traditional multi-atlas approaches in the ability to utilise the available data and to handle the anatomical variability between different images.

3.6   Summary and Conclusion

This chapter presented a new generalised framework for applying patch-based segmentation which uses spatial context to enable patch comparisons to be made on a global basis in order to better accommodate variations in image alignment. This approach no longer requires local search windows to be defined and extends the non-local label fusion approach of previous patch-based methods by incorporating spatial information and a spatial weighting to each patch. The proposed framework is validated against 202 ADNI images of patients at various stages of Alzheimer's disease and achieved an overall median dice coefficient of 0.867 using patches from the 40 most similar atlases. When compared with an existing and well established patch-based segmentation approach [47], the proposed approach is able to provide both higher segmentation accuracy on average and better consistency in the results.

In subsequent chapters, we will look at this approach can be extended to other applications as well as using different approaches for spatial context to further relax the dependency on image alignment.

Chapter 4

Patch-based Segmentation without Registration

This chapter is based on the following publication:

- Zehan Wang, Claire Donoghue, Daniel Rueckert. Patch-based Segmentation without Registration: Application to Knee MRI. *Fourth International Workshop on Machine Learning in Medical Imaging (MLMI 2013)*. LNCS Volume 8184, pages 98-105. Springer Heidelberg 2013.

- Zehan Wang, Anil Rao, Daniel Rueckert. Patch-based Segmentation without Registration: Application to Canine Leg MRI. *MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA)*. 2013.

4.1 Introduction

The previous chapter introduced the basis for a kNN patch-based segmentation framework which incorporates spatial context for improved segmentation performance. Using each voxel's coordinates as spatial context provided an improvement in segmentation accuracy compared to

previous patch-based methods which did not use spatial context. Now let us look at extending this approach to accommodate more challenging datasets in terms of registration.

One potential weakness of using voxel coordinates as spatial context is that this still relies on the alignment of the anatomical structures - and thus is still susceptible to registration errors. In some applications, it is not uncommon for registrations to fail altogether. Problems with affine registration of knee MRI have been previously quantified in [57] where the authors observed 4.08% of direct pairwise registrations fail without manual input. Here, failure is defined when registration results in the distance between manually identified landmarks within the image being greater than 10mm. There are many other applications where registration failures also occur with significant frequency without manual input. For example in cardiac images, manual landmarks are often required to initialise the registration [9], and in a recent segmentation challenge[1], around 20% of registrations failed in the canine leg dataset. This is something that receives little attention in many atlas-based segmentation methods, particularly if majority voting is used, since failed registrations can be ignored as long as they do not compromise the majority. However, this is not always guaranteed, and if weighted voting is used, the outliers are no longer ignored. This indicates that there is an issue with relying on the registration outcome, particularly if misalignment commonly occurs. The registration outcome has a direct effect on the segmentation accuracy as well as the scalability of the overall framework, particularly if manual input is required.

To avoid these issues, this chapter investigate the potential to perform segmentation without any kind of registration. Note that the goal here is to be able to handle the normal range of outputs from medical scanners and this is not the same as the ability to be invariant to arbitrary image orientations imposed after a normal scan. To overcome the dependency on registration, an approach is proposed to provide spatial context that is robust to the image alignment by using the concept of relative distances rather than using absolute coordinates within the image. Additionally, an atlas selection method is proposed which uses histograms with 3D oriented gradients as image descriptors to enable fast and generic similarity comparisons. Without registration, traditional atlas selection methods [140], [4] may not provide meaningful results

---

[1]MICCAI 2013 Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA)

as they use distance metrics that implicitly assume correspondences between image and atlas. Patch-based methods use local intensity information, so the atlas selection scheme is intended to reflect this in addition to being registration-free.

The proposed methods are applied within a multi-resolution framework and evaluated on two diverse and challenging datasets - knee MRI from the MICCAI SKI10 Grand Challenge and canine leg MRI from the SATA MICCAI challenge workshop. This framework can quickly obtain a coarse initial segmentation in the lowest resolution without relying on image correspondences and without requiring any information regarding alignment. This initial segmentation is then refined by propagating through subsequently higher resolutions until the required resolution has been reached.

4.1.1 Chapter Overview

In this chapter, we begin with a method for establishing spatial context without registration, using relative distances between established structures in the image. An approach for establishing these structures automatically in the initial segmentation is also proposed here. This is then followed by the introduction of a multi-resolution segmentation framework which uses the proposed spatial context approach in an iterative process, where both the spatial context and the segmentation is refined and updated with each iteration. Finally, we end with the atlas selection scheme before discussing the results of applying the proposed methods to knee and canine leg MRI datasets.

4.2 Providing Spatial Context without Registration

4.2.1 Adaptive Coordinate System

Without registration, using spatial coordinates as spatial context is not effective since the coordinates are not aligned. This is especially the case when dealing with potentially large
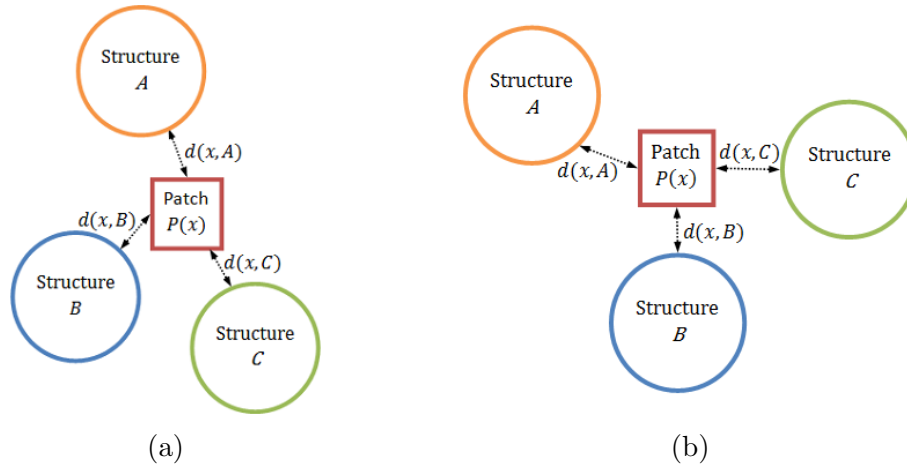
Figure 4.1: (a) and (b) represent the same anatomical structures but the two images are not aligned. Spatial context for patch $P(x)$ can be provided by the distances to these structures regardless of the how they are positioned within the image.

misalignments between images where spatial context based on explicit image coordinates can be unreliable. Spatial context should therefore be defined in a way that is robust to misregistrations. To this end, *relative distances* can be employed. If any reference points or an initial rough segmentation can be established, the distance to each of the labelled structures can be used to create a non-Cartesian, *patient-specific* coordinate system that is invariant to how anatomical structures are positioned within the image (Fig. 4.1).

Let us define a *structure* as a landmark or labelled structure in the image, which can then be described spatially by the set of coordinates which it takes up in the image. This could be a single coordinate for a landmark or multiple coordinates to describe the location of an anatomical structure, essentially describing structures in a similar fashion as level sets. Then, for a voxel $\mathbf{x}$ and a set of n structures $\{R_i : i \in 1, ..., n\}$, the spatial context function $S_{\text{adaptive}}(.)$ for $\mathbf{x}$ can be defined according to the shortest distances between $\mathbf{x}$ and $\{R_i\}$ according to their voxel-wise coordinates. i.e.

$$S_{\text{adaptive}}(\mathbf{x}) = [d_{\min}(\mathbf{x}, R_1), d_{\min}(\mathbf{x}, R_2), ..., d_{\min}(\mathbf{x}, R_n)] \tag{4.1}$$

where $d_{\min} : \mathbb{R}^3 \times \{\mathbb{R}^3\} \to \mathbb{R}$ is a distance function which finds the minimum distance between $\mathbf{x}$ and R_i based on their coordinates and using some a defined distance metric. In this chapter,

the Euclidean distance is used as the metric, and the Euclidean distance transform (EDT) can be used to calculate $d_{\min}$ for all pixels. The EDT effectively provides a distance map (or image), where the value for each pixel is the minimum distance to the designated structure. It can be computed in linear time using the approach in [117] and enables us to calculate the spatial context on-the-fly during the segmentation.

Using a binary mask of each structure in turn, the EDT can be applied to provide a distance map for each voxel in the image, effectively providing a new coordinate system to use which is adaptive to the anatomical structures present in each image, and provides an efficient solution to establishing spatial context without registration. This can be calculated for each target image based on either an initial segmentation or established reference points and could be based on either an automatic approach or by manual interaction.

In principle, at least three structures are required to localise a point in 3D space, but this may not be necessary depending on how it is applied. For example, the EDT of two structures can localise a non-Cartesian line, whilst the EDT of a single structure can provide enough spatial context for a non-Cartesian plane or surface. This may provide enough additional information to distinguish between patches of similar intensities from different structures.

4.2.2 Establishing an Initial Segmentation

The approach for establishing an initial segmentation is influenced by the findings in [171], where, motivated by the remarkable tolerance of the human visual system to degradations in image resolution, the authors reported automated segmentation tasks can be performed on images with resolutions as small as in the range of 32×32 to 16×16 depending on the size of the object.

Initial experiments at low resolutions suggested that a patch-based approach could be used to determine an initial coarse segmentation both with and without spatial context. At this resolution, it is computationally feasible to make patch comparisons across the whole image without the need to define any regions of interest. Differences in global alignment of images

can be adequately accounted for on a patch level with atlases in similar orientations, regardless of their alignment. At low resolutions, the computational cost of searching through atlases for similar patches is also lower, enabling more atlases to be used, thereby potentially enabling the anatomical variability to be better accounted for at the patch level.

Spatial context based on relative distances between structures cannot be applied initially without having some prior knowledge of some of the structures locations, but weaker spatial context could be used such as scaled coordinates or the distance to the centre of the image. This sort of spatial context is relatively inaccurate without registration, but can still be beneficial when applied weakly. Results and examples of the low resolution segmentation are presented in sections 4.6.3 and 4.7.3. Once an initial segmentation has been established, it can then be refined with stronger spatial context by using the EDT-based approach, thus overcoming the misalignment between images and lack of direct image correspondences.

4.3 Segmentation Framework Overview
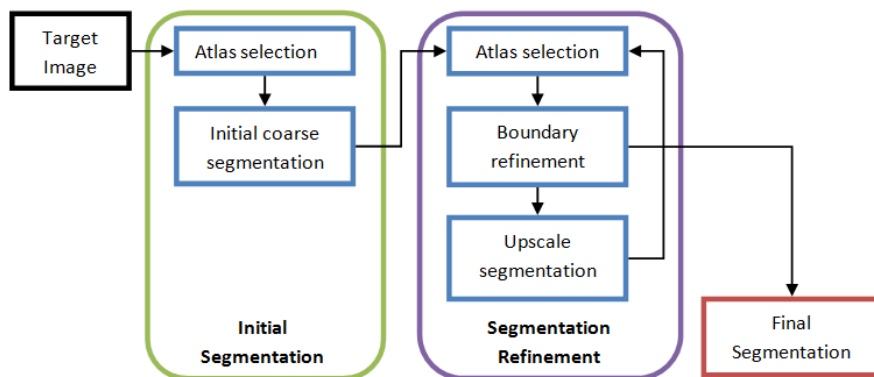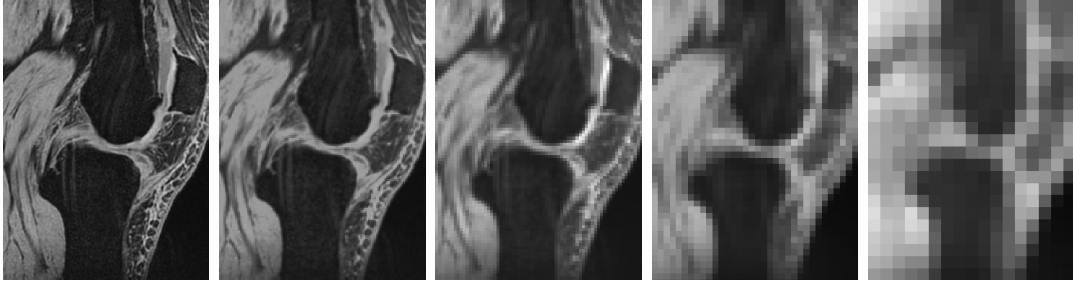
Figure 4.2: General overview of the proposed segmentation framework: once, the initial segmentation is established in the lowest resolution, the segmentation is then refined through subsequently higher resolutions until the required resolution is reached.

As an overall segmentation framework, a multi-resolution approach can be applied with the core methods described above and using the same segmentation principle as that introduced in 3. The segmentation labelling process uses the spatially weighted label fusion approach, previously introduced in 3.2, in a simple iterative manner. A high level overview of the proposed framework is illustrated in Figure 4.2.

This approach can quickly establish a coarse initial segmentation in the lowest resolution image and then refine it using subsequently higher resolutions with the EDT-based spatial context. With each subsequent refinement, not only is the segmentation being improved, but also the accuracy of the spatial context. The same method for atlas selection and patch-based label fusion (SAPS) is applied at all resolutions (see Table 4.1) to provide the initial, intermediate and final segmentations. The atlas selection approach is described in more detail in section 4.4.

4.3.1 Hierarchical Segmentation Strategy

Table 4.1: Example of a Gaussian image pyramid for a knee MR image from the SKI10 dataset with voxel size (V. Size) and image resolution (res.). Let level 0 denote the native resolution, then the levels are numbered upwards in ascending order, where level 1 is the highest sub-sampled isotropic resolution and each subsequent level is half the resolution of the previous level.

Level	0	1	2	3	4
V. Size	$0.39^2 \times 1$mm	0.78mm^3	1.56mm^3	3.12mm^3	6.24mm^3
Res.	$280 \times 400 \times 110$	$140 \times 200 \times 140$	$70 \times 100 \times 70$	$35 \times 50 \times 35$	$17 \times 25 \times 17$

Multiple resolutions of each image can be created by constructing a Gaussian image pyramid [5]. Given an initial segmentation from a low resolution image, only a boundary region, which can be calculated as a *morphological gradient* [139] defined by the difference between the *dilation* and *erosion* of each segmented structure, will need to be refined in a higher resolution (see Fig. 4.3). This is because a low resolution segmentation cannot represent the boundaries of each structure as well as a higher resolution. However, for internal voxels of each structure, the low resolution segmentation would be sufficient, so no further refinement would be required for these voxels.

The size and shape of the boundary region can be controlled by the structuring elements used for dilation and erosion as well as the number times the dilation and erosion are performed
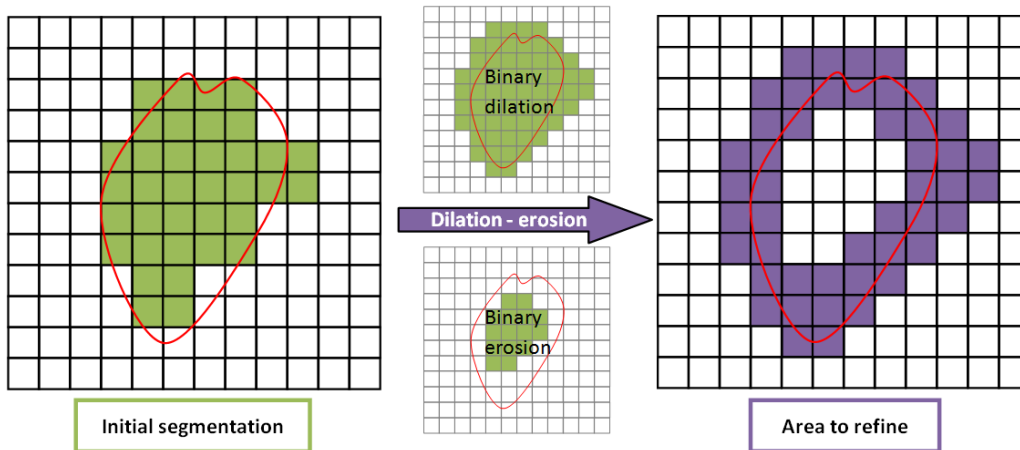
Figure 4.3: Determining the boundary region for refinement from an initial segmentation. The red outline represents the true segmentation boundary. This approach is also used to define the region of interest to use for atlas selection.

before taking their differences. For simplicity, this can be a '+ shaped' structuring element with a radius of 1 pixel in each axis for both dilation and the erosion, and the same number of dilation and erosion operations can be applied to produce the boundary region. This also allows the boundary size to be defined by the number of dilation (or erosion) operations used to create the morphological gradient. i.e. a boundary size of 3 means that it is created from the difference between 3 dilation operations and 3 erosion operations. This definition is used to allocate a ROI mask of equivocal boundary size in each image for the refinement process.

Multiple iterations of this boundary refinement can be carried out at each resolution to increase the accuracy. This forms a straight-forward and computationally-efficient strategy to process images through increasing resolutions, allowing a patch-based approach to be used in all resolutions. This approach to establishing a boundary region for refinement is similar to that used in [155]. However erosion is used in addition to dilation since the true boundary can lie within the segmentation from the lower resolution. This also allows some error correction to occur during the refinement process. The boundary size can be adjusted to account for potential errors from the initial segmentation, however it is desirable to minimise this size in order to limit the number of patch comparisons and thus the computational cost during each refinement.

4.4 Atlas Selection Using Histogram of Gradients

Previous works using patch-based approaches, such as in [47] or in the previous chapter, suggested that linear increases in the number of atlases yields logarithmic improvements in the segmentation. Therefore, as a trade-off between time and accuracy, only the nearest N atlases are used rather than the entire atlas library. Traditional atlas selection methods which typically use voxel-to-voxel based distance metrics cannot be used meaningfully when the images are not aligned, so a different type of similarity measure must be employed. Assuming images can be intensity normalised, one approach to compare images is by their intensity histograms. This can provide a measure of the high level differences in the distribution of intensities, however this discards local neighbourhood information which is relevant for patch comparisons. Instead, a better alternative would be to use a histogram of oriented 3D gradients based on [95] and [52], as this incorporates local neighbourhood information in providing gradients for each voxel.

Gradients are calculated using a 1D Sobel operator in each direction $(\vec{x}, \vec{y}, \vec{z})$ of the image to derive the gradient magnitudes for the three natural orthogonal axis of the image. The Sobel operator uses the convolution mask $[-1, 0, 1]$ to derive approximations for the gradient and presents a computationally simple and efficient manner to calculate the gradients for all voxels in the image. This then provides magnitudes in the three orthogonal axis directions of the image and must be binned in 3D to produce a 3D histogram. Ensuring the orientation bins are equidistant in 3D is not trivial, known otherwise as the Thomson problem [167], but using the centre positions of faces on regular polyhedrons such as the icosahedron is one solution. The icosahedron has 20 regular faces and was suggested as an appropriate choice for histogram binning in [95]. The 20 faces around a unit sphere centred on the original can be described by the following 20 centre points:

$$\{(\pm 1, \pm 1, \pm 1), (0, \pm 1/\phi, \pm\phi), (\pm 1/\phi, \pm\phi, 0), (\pm 1/\phi, 0, \pm\phi)\} \qquad (4.2)$$

where $\phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio. These can be used as vectors to describe the gradient orientation and to bin the gradients calculated by the Sobel operators. Let $\mathbf{P} = [\mathbf{p_1}, \mathbf{p_2}, ..., \mathbf{p_{20}}]^{\mathrm{T}}$
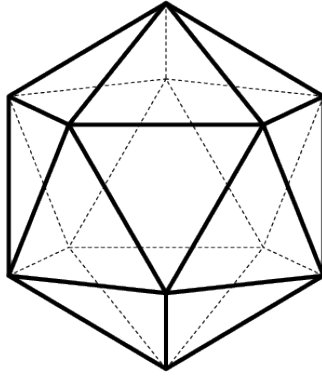
Figure 4.4: The icosahedron is a regular 20 sided shape. The centre of each of the faces are equidistant from each other, providing suitable orientations in which to bin gradients in 3D.

be the matrix of be the orientation vectors representing the centre of the faces of the icosahedron and let $\mathbf{q} \in \mathcal{R}^3$ be a vector of the gradients for a voxel in the image. Then $\mathbf{q}$ can be binned into one of 20 bins in the histogram, by first selecting the orientation bin h_i in the histogram as:

$$h_i = \arg \max_{\mathbf{p_i} \in \mathbf{P}} \mathbf{p_i} \cdot \mathbf{q} \tag{4.3}$$

After which, the total gradient magnitude, $||\mathbf{q}||_2$, is used as the value to bin, whilst there is no change to other bins. This is performed for all relevant voxels, as defined by any ROI masks, and summed up in order to build a histogram. Then for each image, a histogram of 3D gradients can be produced which can be used to compare and order images by similarity.

There are many distance measures that can be employed for histogram comparison such as the earth mover's distance [144], but for the applications in this chapter, the L_1 norm was found to be a simple and effective measure. For the initial segmentation, the atlas selection process compares histograms of the whole image, but once an initial segmentation has been established, regions of interest can be defined for the purpose of segmentation refinement using the boundary regions as shown in Figure 4.3. These regions can then be used to restrict the parts of the image where comparisons are made for atlas selection. To increase specificity, histograms are calculated separately for each structure's boundary region within each region and then concatenated to produce an overall histogram for that region.

4.5   Framework Summary

The methods proposed in this chapter, effectively extends the framework proposed previously in Chapter 3, using it in an iterative process to initialise and then refine a segmentation in multiple resolutions. A Gaussian image pyramid scheme is used to provide the multiple resolutions, and segmentation is carried in a *coarse-to-fine* fashion from the lowest resolution to the highest.

The new approach for spatial context, based on relative distances to structures rather than using the explicit coordinates of each patch, enables us to be less dependent on the alignment of the images and to provide spatial context without registration. The kNN patch selection and the label fusion process still uses the same approach which was presented previously in Chapter 3.2 and 3.3.

For atlas selection, in order to compare images without registration, histogram of 3D oriented gradients are computed from each image and then the N nearest atlases can be selected based their histograms similarities. Gradients for each voxel are binned regularly in 3D according to the 20 faces of an icosahedron to build up a histogram.

4.6   Application to MICCAI SKI10 Grand Challenge

For experimentation and evaluation, the proposed methods were applied to the MICCAI SKI10 Grand Challenge [82] dataset. The only assumption made is that the joint is the main focus for these images and that they are acquired in some approximately standardised orientation, i.e. none of the images are upside-down. However, no orientation information, nor any information regarding image alignment are explicitly used.

4.6.1   Dataset

The basis for the SKI10 dataset are a range of knee images originating from the surgical planning program of Biomet, Inc. This dataset contains an approximately equal number of

both left and right knee images from multiple patients, and were acquired at over 80 different centres in the USA using MRI machines from all major vendors, i.e. General Electric, Siemens, Philips, Toshiba and Hitachi. A range of MRI sequences were employed, with the majority using T1-weighting, but some also used T2-weighting. Many images used gradient echo or spoiled gradient echo sequences, and fat suppression techniques were common as well. Additionally, the field strength applied also varied, with 90% of cases using 1.5T, but the rest using 3T or 1T. All images were acquired in the sagittal plane with pixel spacing of $0.39 \times 0.39 \text{mm}^2$ and slice distance of 1mm without using contrast agents. The images were acquired for the purpose of surgical planning for partial or complete knee replacements, however information regarding pathologies were omitted. 100 images are provided for training purposes and 50 separate unseen images are used for testing. No reference segmentations are provided for the test images and results are validated independently by the challenge organisers.

4.6.2   Implementation

The framework from the previous chapter was extended to allow a multi-resolution approach as well as allowing a choice of spatial context to use. The implementation used open-source modules in Python and Cython when possible.

As a pre-processing step, all images are bias field corrected [172], re-sampled to isotropic voxel size and intensity normalised [123], with intensities scales to the range $[0, 100]$. All images are considered for each stage of the segmentation process without separation of MR field strengths, or into left or right knees.

Initial segmentations were obtained, as proposed in section 4.2.2, with voxel coordinates normalised to the range $[0, 100]$ used to provide spatial context and a weak spatial weighting α. Although images are not aligned, the use of normalised coordinates with a weak spatial weighting were found to provide more accurate initial segmentations than not using any spatial information at all. For segmentation refinement, the EDT from the tibia and the femur were used to provide relative distances as they constitute the main structures in these images which are subject to change in pose within the images. At the same time, the anatomy of the knee

naturally places the femur and tibia relative to each other spatially, hence making good candi-
dates to provide relative distances for spatial context. The choices for α at each resolution and
iteration were chosen empirically based on experiments with the training data.

Different patch sizes were found to provide the better results for each resolution, with a general
trend of increasing patch sizes as the resolution increased. For the initial segmentation, it was
found that larger patch sizes would have the most stable and consistent results whilst smaller
patch sizes could perform better on average but had greater variation in the range of results.
To combine the qualities of different patch sizes, an initial segmentation was established using
relatively larger patches but then refined in the same resolution using smaller patches before
propagation to higher resolutions. Since the native image resolution is non-isotropic, the choice
of patch size for the final resolution was also chosen to be non-isotropic so that the information
contained in each patch is more consistent in each direction.

The initial segmentation can have a great impact on the overall segmentation quality and it was
found that performing refinements with smaller patch sizes in the lowest resolution to obtain
a better coarse level segmentation would ultimately lead to better segmentation accuracy after
propagating through higher resolutions. Furthermore, the computational costs of performing
multiple refinements in the initial resolution is relatively cheap compared to doing so at a higher
resolution. The parameters used and computational times for each resolution are summarised
in Table 4.2.

4.6.3   Results and Discussion

The results for the SKI10 grand challenge[2] are presented in Tables 4.3, 4.4 and Figure 4.6,
showing promising potential for this framework and demonstrating the possibility of applying
patch-based segmentation in images without applying registration. It is interesting to note that
a segmentation of the knees, with an average surface distance to the bones of under 1.2mm,
can be achieved in under 2 minutes using the second lowest resolution level. The framework
does not make any prior assumptions about the different anatomical structures, yet achieves

---

[2]See `http://www.ski10.org/results.php` for a full list of results from other entries.

Table 4.2: Example segmentations, parameters and computation times for each resolution level. Segmentations from the proposed method are overlaid in green for the bone and yellow for the cartilage. The reference segmentations are outlined in red.
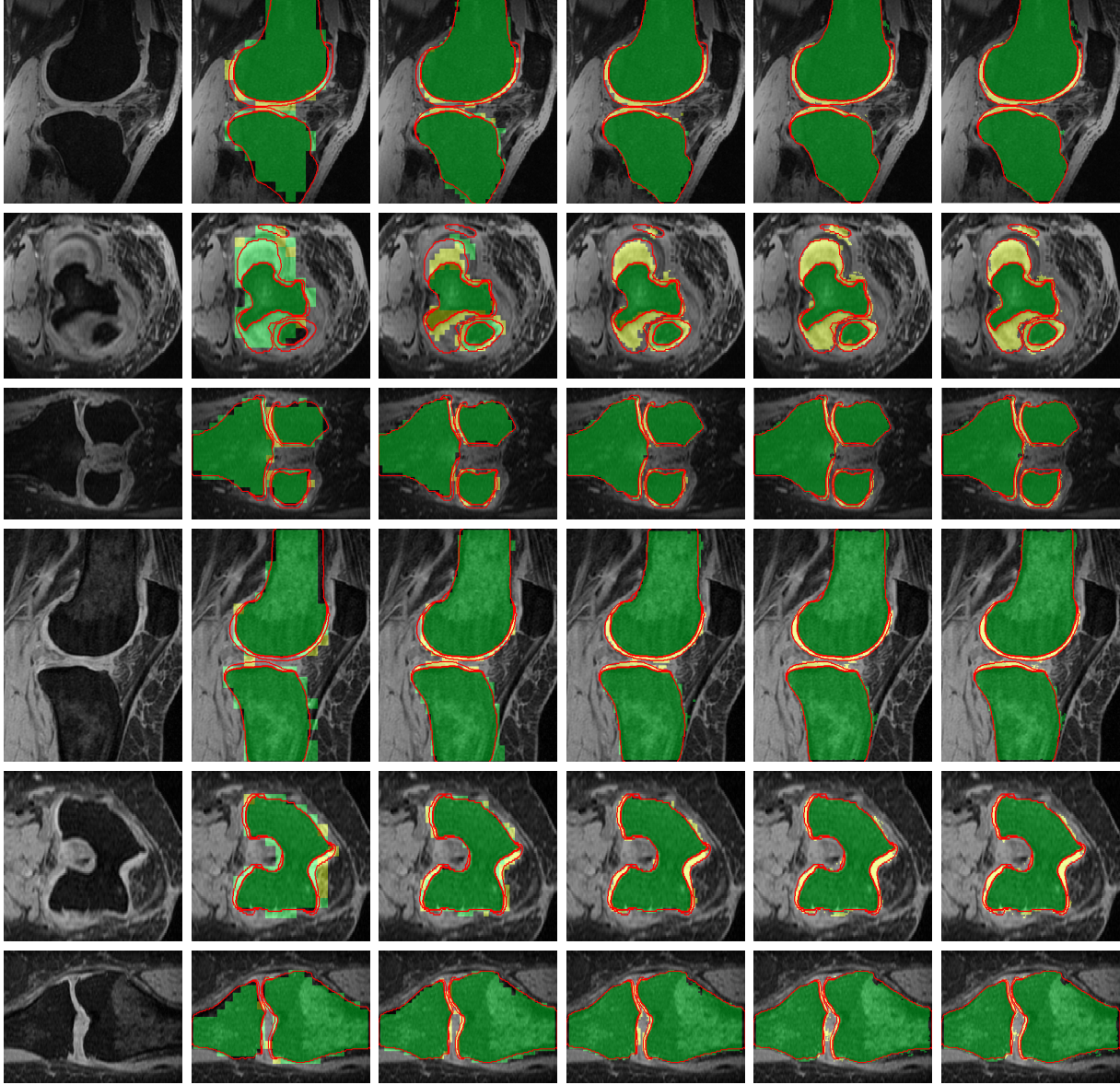
Level	4	3	2	1	0
Patch Size	$7^3, 5^3, 3^3$	3^3	5^3	7^3	$7 \times 7 \times 3$
α	5.5, 2.3, 1.5	2.4	13	35	60
k	15	15	25	40	40
Boundary Size	2	2	2	2	1
Time	~30 seconds	~40 seconds	~10 minutes	~2.5 hours	~1.5 hours
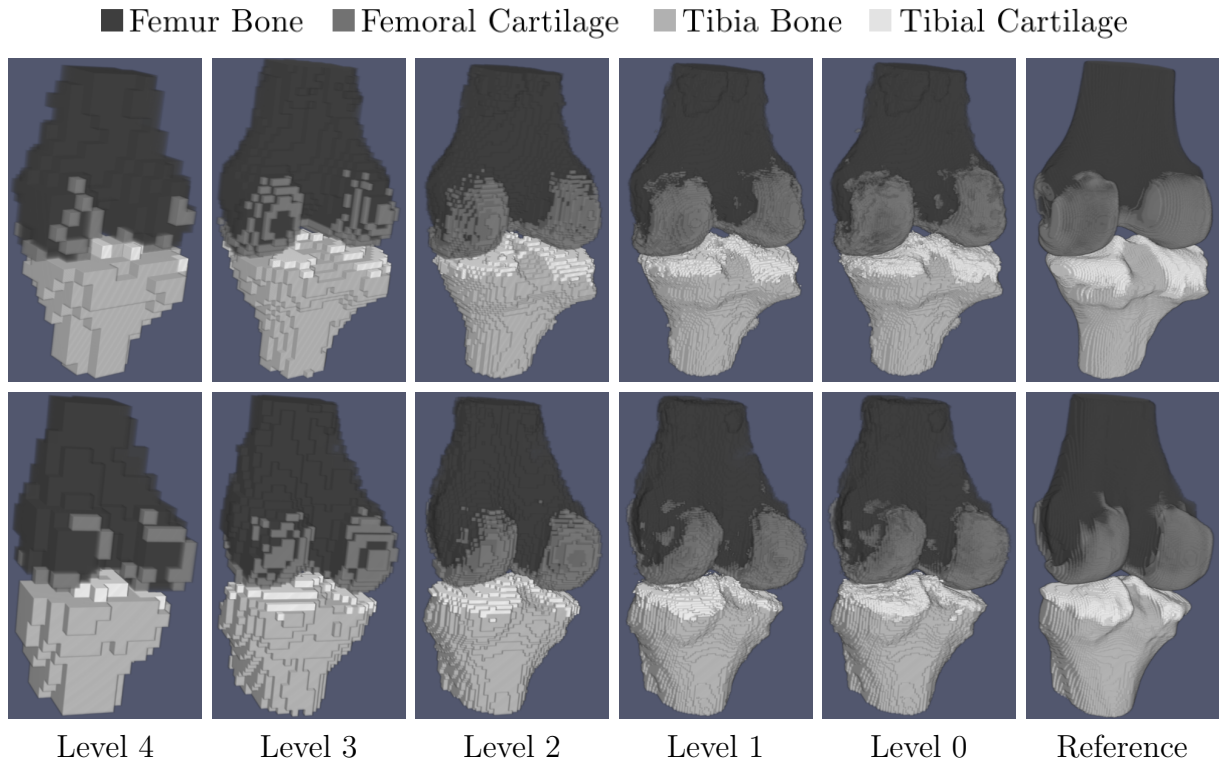
Figure 4.5: Example volume renderings of the segmentation results from each resolution level in comparison to the reference segmentation.

a good score for bone segmentation whilst the score for cartilage segmentation is comparable to the top scoring entries in the challenge. In the final resolution, the framework achieved an overall score of 58.3, with average scores of 64.9 for cartilage and 51.7 for bone. Overall within the challenge, these results rank 3rd for cartilage segmentation, 8th for bone segmentation, 7th overall. The SKI10 challenge operates on a rolling basis, and at the current time of writing, there are 14 total entries including this one.

In general, many of the top scoring methods use model-based approaches, which tend to have strong correlations between the bone and cartilage segmentation accuracy as cartilage segmentation relies on an accurate bone segmentation to define the bone-cartilage-interface. In contrast, the proposed method does not have such a tightly coupled relationship between the bone and cartilage segmentation results since bone and cartilage labelling are performed simultaneously. Although the relative bone positions influence the cartilage segmentation, it does not restrict the shape of the cartilage segmentation. As a result, this approach is able to work well on the cartilage, producing comparable scores with the top scoring methods, even with much
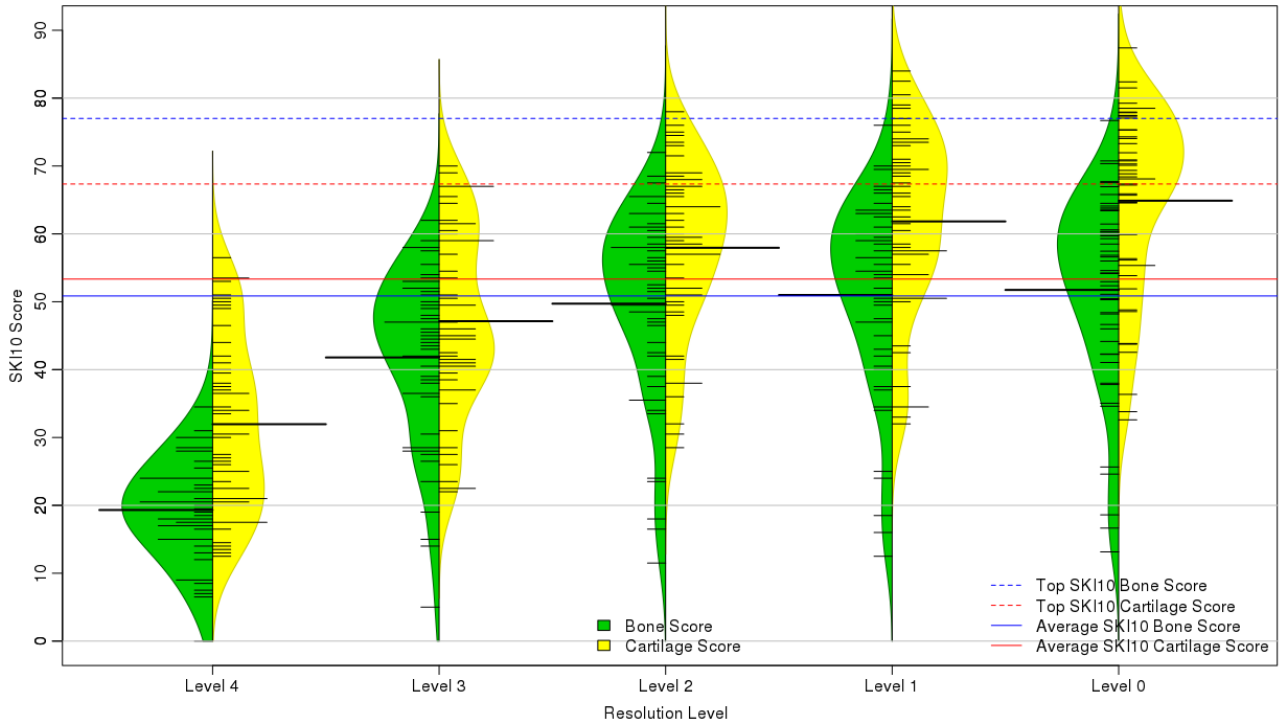
Figure 4.6: Beanplot of segmentation scores for each resolution level in comparison to the highest scoring method and the average score in the SKI10 Grand Challenge. The shape of the "bean" shows the distribution of the results and individual data points are shown as small lines on the bean. The overall score (calculated by the mean) for each resolution is indicated by the thick black line on the bean. Average SKI10 scores are calculated from the SKI10 results table, excluding the proposed method.

Table 4.3: Overall results from the SKI10 grand challenge showing average surface distance (AvgD), root mean squared surface distance (RMSD), volumetric overlap error (VOE), volumetric difference (VD), score (Scr) and their standard deviations for each resolution level.

Res. Level	Femur Bone			Tibia Bone			Femoral Cartilage			Tibial Cartilage		
	AvgD [mm]	RMSD [mm]	Scr	AvgD [mm]	RMSD [mm]	Scr	VOE [%]	VD [%]	Scr	VOE [%]	VD [%]	Scr
4	1.71	2.32	30.8	1.72	2.38	7.8	75.70	−6.10	37.5	76.42	−1.86	26.4
	±0.27	±0.50	±10.8	±0.60	±0.96	±6.2	±3.57	±16.00	±17.2	±9.21	±48.33	±17.6
3	1.18	1.88	49.1	1.15	1.91	34.4	56.17	−1.71	47.4	55.83	−1.05	46.9
	±0.39	±0.80	±15.0	±0.73	±1.30	±16.6	±4.27	±15.63	±17.9	±5.72	±24.56	±17.0
2	0.95	1.74	56.1	0.94	1.83	43.3	39.28	0.99	59.7	40.35	−1.20	56.2
	±0.42	±0.86	±15.9	±0.81	±1.44	±19.1	±5.54	±14.19	±17.3	±5.78	±18.35	±16.8
1	0.90	1.75	57.1	0.88	1.84	44.9	32.81	0.65	62.1	33.55	−2.36	61.6
	±0.44	±0.89	±16.3	±0.81	±1.46	±19.9	±6.69	±13.88	±17.5	±6.37	±15.96	±16.9
0	0.88	1.73	57.8	0.86	1.82	45.6	27.19	−0.894	65.4	27.96	−4.39	64.4
	±0.44	±0.89	±16.3	±0.80	±1.45	±20.1	±8.03	±13.16	±16.2	±7.05	±15.56	±18.1

Table 4.4: Overall scores and rankings from the SKI10 grand challenge. The rank is given as the current position in the SKI10 results table.

Res. Level	Bone Score	Rank	Cartilage Score	Rank	Overall Score	Rank
4	19.32 ± 7.20	12	31.96 ± 12.8	11	25.63 ± 7.5	11
3	41.79 ± 12.9	11	47.14 ± 13.6	11	44.44 ± 9.8	11
2	49.72 ± 12.1	8	57.95 ± 12.5	10	53.81 ± 9.2	10
1	50.99 ± 14.4	8	61.84 ± 13.6	4	56.42 ± 9.8	8
0	51.75 ± 14.5	8	64.88 ± 13.7	3	58.31 ± 9.9	7

lower bone segmentation accuracy. This can be particularly of an advantage in cases where it is challenging to accurately align correspondences on the bones. Additionally, the structural variability of the cartilage can be much greater than the bones so model-based approaches can be at a disadvantage to nearest-neighbour approaches if their models do not fully explain the complexities. However, the model-based approaches tend to perform better for bone segmentation. This is in part due to the limited shape variation for individual bones, so a model-based approach is able to capture this during training fairly well. Patch-based methods do not incorporate shape information and can be misled by the intensity information if there are not enough sufficiently locally-similar examples within the atlas library. Given that the bones are relatively large and that variations in intensities can differ from region to region, even within a single subject, it is likely that the atlases in the training set were not sufficiently diverse enough for the proposed approach to fully account for it. The intensity normalisation process also greatly affects the intensity variability between images, and therefore, the segmentation results. The fact that the SKI10 dataset were produced from different scanners with differing image acquisition protocols certainly poses a challenge for intensity normalisation across the images, especially when they are not aligned. This a weakness of nearest-neighbour approaches based on absolute intensity information, and partly explains why the model-based approaches perform better in this case. This could potentially be addressed in changing the distance metrics used for patch comparison, such as by using a NMI-like metric instead of an Euclidean one. This is plausible in the proposed framework as long as it is a true metric which obeys the triangle inequality, which the ball tree relies upon. Otherwise, an alternative kNN data structure will be required.

The use of k-NN data structures such as the ball tree removed the requirement for using fixed-size search windows and in turn, enabled the proposed framework to allow segmentation without any kind of registration between images. However, the speed is still dependent on the number of voxels within the image. It can be seen in Table 4.2 that the increase in computation times grows exponentially as resolution increases. Even though the use of a binary search tree such as the ball tree theoretically reduces search time to $\mathcal{O}(\log_2 n)$, the number of voxels grows by $\mathcal{O}(n^3)$ with each increase in resolution. This puts a constraint on the size of the images that a voxel-by-voxel patch-based approaches can be suitably applied to. One possible solution to overcome this problem is to exploit parallelism, since voxels are labelled independently from each other, so the segmentation process could be potentially run on GPUs or CPU clusters.

Overall, the proposed framework streamlines and simplifies the image segmentation pipeline, requiring fewer separate processes than model-based approaches where correct image alignment and extensive model training are required. The results for this dataset demonstrate potential in an alternative approach that is not usually used for this application, although more can be done to improve the overall performance. One possibility is to combine model-based approaches with the proposed framework, using the patch-based approach to refine and further adapt trained models to each target image. This could be performed after model fitting, where an initial segmentation can be estimated, thereby providing structures for relative distances and enabling the proposed approach for spatial context. Patch-based refinement can then carried out to improve the initially fitted model.

4.7   Application to Canine Leg MRI

The proposed framework was also applied to the canine leg MRI dataset from the 2013 MICCAI challenge workshop on Segmentation: Algorithms, Theory and Applications (SATA). This operated in a similar fashion to the SKI10 challenge, where the challenge organisers independently evaluate the performance of each method on an unseen test set.
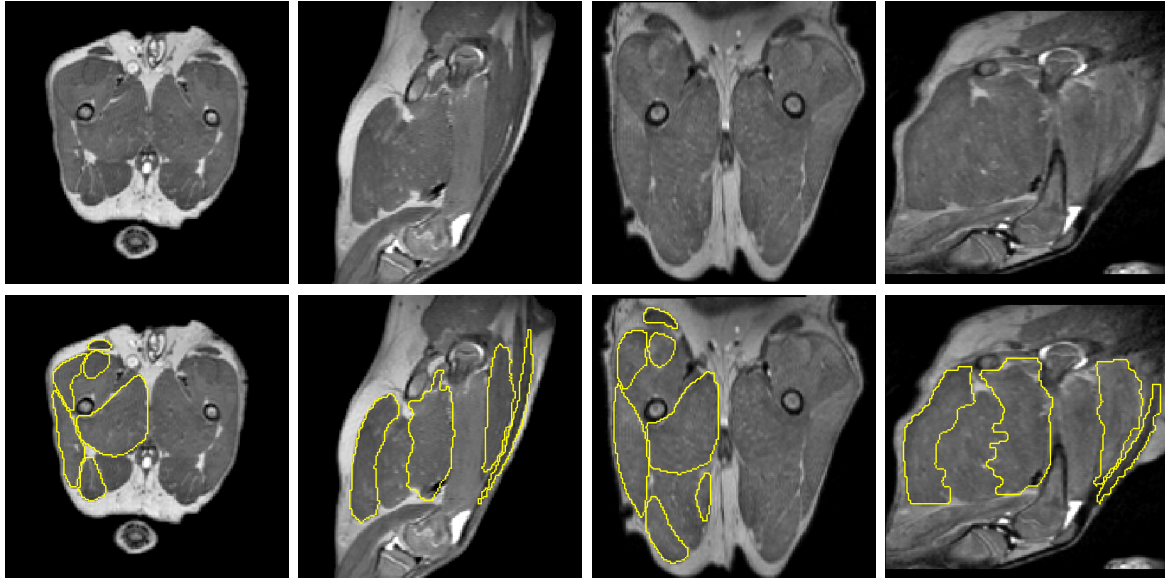
Figure 4.7: Example of images and reference segmentations in the SATA canine leg MRI dataset. Reference Segmentations are outlined in yellow for each of the muscle groups visible.

4.7.1 Dataset

The dataset contains MR images of the proximal pelvic limbs of 8 normal dogs and 10 golden retrievers with muscular dystrophy at approximately 3 and 6 months of age. Additional scans were performed on 4 of the normal dogs and 5 of the golden retrievers at 9 months of age. The images were acquired with Siemens 3T MRI scanners using T2-weighted protocols and have resolutions of $256 \times 256 \times 180$ voxels with each voxel having a resolution of $1 \times 1 \times 1 \text{mm}^3$. The labels for each atlas consisted of 7 proximal pelvic limbs muscles manually segmented by 3 experts by delineating the outline of each muscle for every 5th image slice and using linear interpolation to connect up the segmentations in 3D. The muscles segmented were the cranial sartorius, rectus femoris, semitendinosus, biceps femoris, gracilis, vastus lateralis and adductor magnus. In total, 22 images were provided for training and 23 additional unseen images were used for testing. Segmentation results were evaluated independently by the challenge organisers.

4.7.2 Implementation

A similar approach was applied as for the SKI10 dataset, initialising using spatial coordinates, normalised to the range $[0, 100]$, and then using the EDT to each of the muscle groups as

spatial context. Pre-processing was also similar, however Otsu thresholding [125] was applied to remove the background prior to intensity normalisation. Multiple boundary refinements were performed for each resolution using different patch sizes (see Table 4.5 for parameters). The small size of the atlas library meant it was computationally feasible to use all atlases to provide patches for segmentation, so atlas selection was not required or used for this dataset.

Background Removal using Otsu Thresholding

Otsu's method [125] is an automatic approach to find a threshold value, which can be used for background removal. This can be particularly useful in the situations where the background can be clearly removed, but the threshold value varies between images. The method exhaustively searches for a threshold that minimises the intra-class variance for two classes, the foreground and the background, defined as a sum of variances of the two classes:

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_1(t)\sigma_1^2(t) \tag{4.4}$$

where weights w_i are the probabilities of the two classes with variances σ_i^2, and t is the threshold value that separates them. This can be equivalent to maximising the inter-class variance which is easier to compute:

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = w_1(t)w_2(t)[\mu_1(t) - \mu_2(t)]^2 \tag{4.5}$$

where μ_i are the class means. The class probability $w_i(t)$ is computed from the intensity histogram with t as a limit, so $w_1(t) = \sum_{i=0}^{t} p(i)$ and $w_2(t) = \sum_{i=t}^{n} p(i)$ for the n bins of the histogram. The means are computed in a similar fashion using the threshold t, so $\mu_1(t) = [\sum_{i=0}^{t} p(i)x(i)]/w_i$ and $\mu_2(t) = [\sum_{i=t}^{n} p(i)x(i)]/w_i$, where $x(i)$ is the central value for each bin of the histogram. The values are then computed iteratively until convergence, selecting the best threshold t each time. The final t value can then be used to threshold the background from the foreground.

Table 4.5: Parameters used for each resolution. Resolution levels are numbered such that level 1 denotes the highest resolution segmented and each subsequent level is half the resolution of the previous level. Parameters are chosen by experimentation with the training set.

Resolution Level (voxel size)	3 - Initial ($8mm^3$)	3 ($8mm^3$)	2 ($4mm^3$)	1 ($2mm^3$)
Patch Size	7^3	$5^3, 3^3$	$7^3, 5^3$	7^3
α	10	3.5, 0.9	3.5, 2	5
k	15	15	15	25
Boundary Size	-	3	3	3

4.7.3 Results and Discussion

The results from the SATA segmentation challenge[3] are presented in table 4.6 and the final results for each muscle group presented in Figure 4.9. A comparison of the results with some of the other methods used in the challenge is presented in Table 4.7.

(a) Level 3, $8mm^3$ (b) Level 2, $4mm^3$ (c) Level 1, $2mm^3$
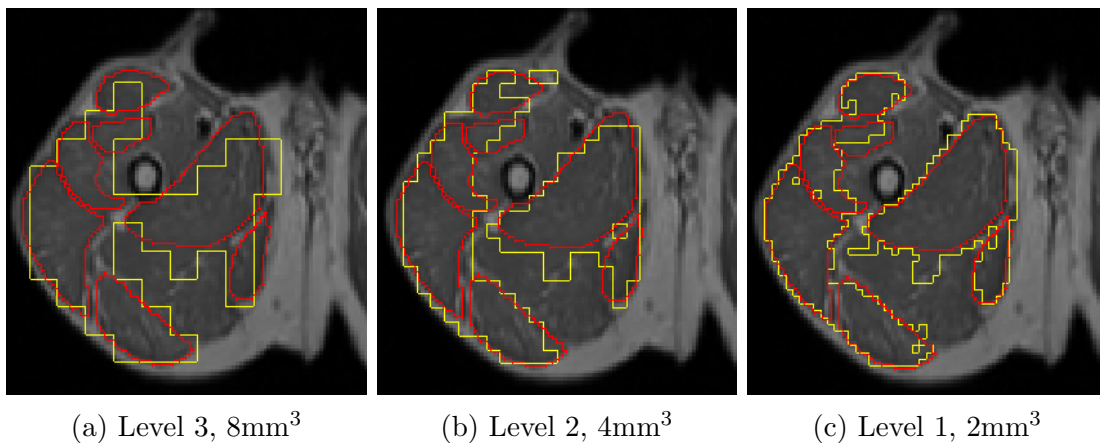
Figure 4.8: Example segmentations are outlined in yellow with ground truth outlined in red, shown in respective ordering to resolution levels. Resolution levels are numbered such that level 1 denotes the highest resolution segmented and each subsequent level is half the resolution of the previous level. Parameters are chosen by experimentation with the training set.

Table 4.6: Overall results for each resolution level. Time taken is an average estimate per image for the resolution on an 8 core 2.8GHz CPU. Infinities (Inf) and undefined (NaN) are caused by failed segmentations where a structure is missing in some of the results.

Res. Level (Voxel Size)	Dice Metric Mean (median) $\pm$ Std.	Hausdorff Dist. (mm) Mean (median) $\pm$ Std.	Time Taken
3 - Initial ($8mm^3$)	0.441 (0.436) $\pm$ 0.078	Inf (42.931) $\pm$ NaN	~5 secs
3 ($8mm^3$)	0.464 (0.460) $\pm$ 0.060	Inf (37.692) $\pm$ NaN	~5 secs
2 ($4mm^3$)	0.568 (0.571) $\pm$ 0.074	32.101 (32.580) $\pm$ 6.671	~2 mins
1 ($2mm^3$)	0.597 (0.587) $\pm$ 0.089	33.383 (33.097) $\pm$ 6.881	~20 mins

---

[3]See `http://masi.vuse.vanderbilt.edu/submission/leaderboard.html` for all other submitted results as well as results for other challenges
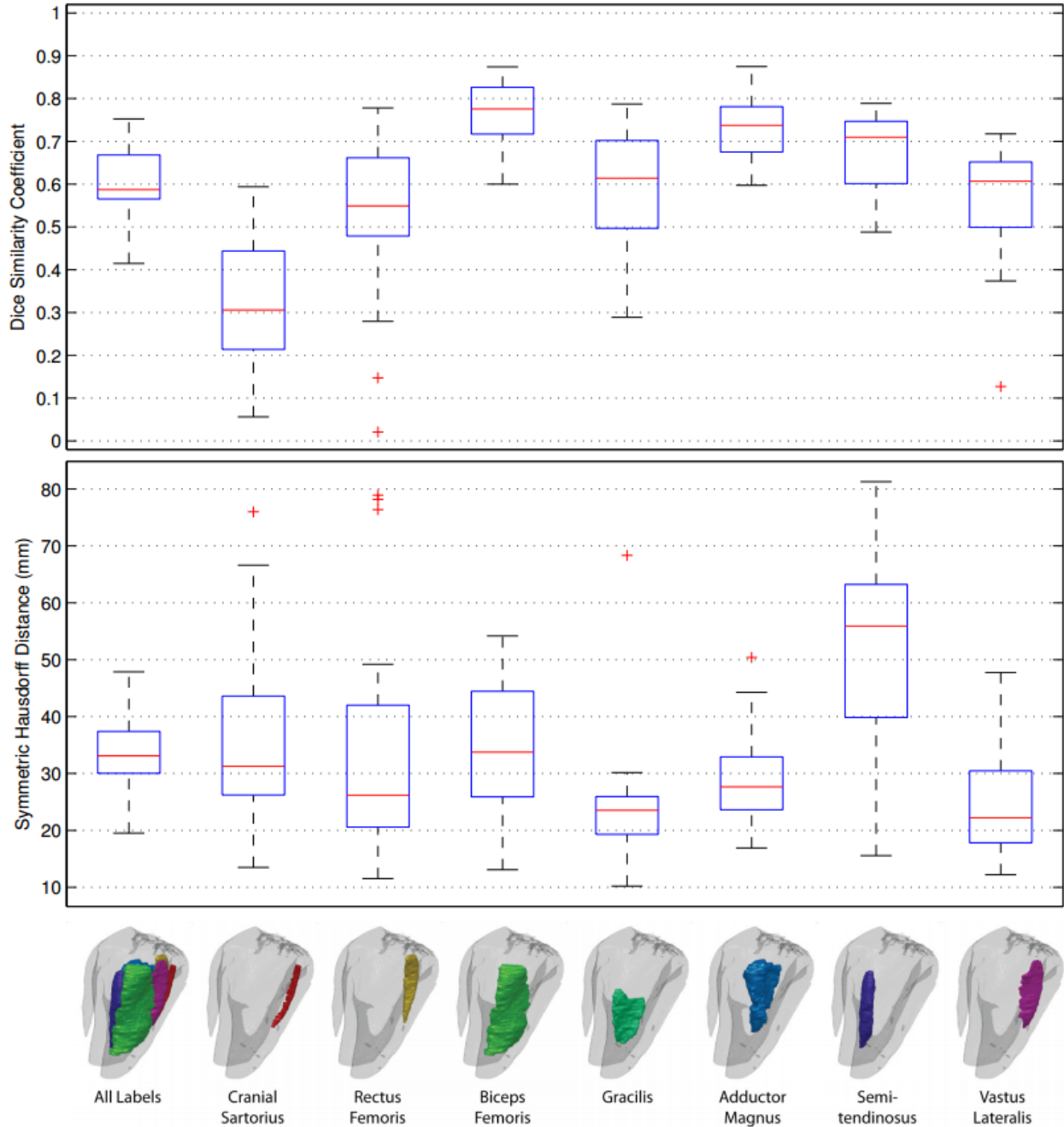
Figure 4.9: Segmentation results for the final resolution level (2mm³) of the proposed method for each of the muscle groups.

Table 4.7: Comparison of results with the top performing method and with baseline methods performed by the challenge organisers using a standardised registration method. Infinities (Inf) and undefined (NaN) are caused by failed segmentations where a structure is missing in some of the results.

Method	Dice Metric Mean (median) ± Std.	Hausdorff Dist. (mm) Mean (median) ± Std.
Proposed (Level 1, 2mm^3)	0.597 (0.587) ± 0.089	33.383 (33.097) ± 6.881
[179]	0.762 (0.797) ± 0.098	27.129 (27.257) ± 8.073
Majority Vote	0.418 (0.424) ± 0.108	Inf (39.198) ± NaN
Spatial STAPLE	0.559 (0.540) ± 0.125	38.6044 (38.141) ± 8.017

The proposed approach was the only one in the challenge which attempted to perform segmentation without any registration. The fact that it was able to outperform both the baseline majority vote and spatial STAPLE methods, which used a standard affine and non-rigid deformable registrations, shows promise and highlights the issues with dependence on accurate registrations for segmentation. However, ultimately the results fell short of the top performing method which did use registration, although their label fusion used a patch-based approach which also allows some relaxation in the dependence on voxel-wise correspondences between images.

The segmentation of muscles, especially ones which neighbour each other proved to be very difficult for the proposed approach as well as many other approaches. One of the issues is in part due to the large variation in sizes of the muscles and the variation in the sizes of the dogs in the dataset. Using two breeds of dogs lead to a large variation in the anatomical sizes and given the relatively small dataset, it meant that the available atlases could not fully account for the total variation that could be exhibited. Another issue was the quality of the MRI scans which varied greatly and how it affected the intensity normalisation outcome. One particular aspect of the proposed framework which may have not worked well was the multi-resolution approach and the initial resolution used. Here, it often became difficult to separate neighbouring muscle groups as downsampling to such a low resolution removed the visible boundaries between them. Given that the muscle groups were generally quite similar in intensities, it is often difficult to differentiate between the different muscle groups, particularly for the initial segmentation when strong spatial context could not be provided.

The approach for spatial context using relative Euclidean distances between muscle groups allowed the provision of spatial context once an initial segmentation was established, however it was easily mislead by incorrectly labelled structures. Additionally, it did not allow much flexibility in accounting for the two different types of dogs which had large variation in the shape and sizes of their legs since. This potentially could have been solved by a registration approach which could enable the dogs to be realigned to a similar scale. However, judging by the results from applying baseline segmentation approaches using a current and standardised registration method, achieving good alignment between these images is a challenging task and remains open to significant improvements.

4.8   Summary and Conclusion

In this chapter, we looked at the possibility of segmentation without registration using a new multi-resolution framework for applying patch-based segmentation and an approach to provide spatial context which is not dependent on voxel-wise correspondences between images. In addition, an atlas selection method was also proposed which uses histograms of 3D gradients for image comparison to enable atlas selection without registration. The overall framework was evaluated without any post-processing methods by applying it to two challenging and publicly available datasets from the MICCAI SKI10 grand challenge and the canine leg dataset from the SATA MICCAI challenge workshop.

This is the first time a purely patch-based method has been applied to segmenting knee images, producing results which are comparable to many of the other methods used in the challenge. The scores indicate a promising first application of the proposed framework but potential improvements could be achieved in the intensity normalisation process and the adaptability to inter-subject appearance variation. Additionally, the use of more training data could may have also have provided a performance increase, as well as post-processing refinements.

Using the EDT from different structures as relative distances to provide an adaptive coordinates allowed the use of spatial context for images which are not aligned. However the Euclidean

distances are subject to variations in scale between subjects, particularly as images are not registered. Additionally, use of the EDT may not fully compensate for incorrectly labelled regions from the initial segmentation which do not lie within the boundary region for refinement. Although it is plausible to perform segmentation without registration, the proposed approach is ultimately not able to outperform the most successful methods in the segmentation challenge. However, the proposed framework could be coupled with model-based approaches, which have more commonly been used in knee segmentations, in order to overcome some of the problems that have been mentioned.

In the next chapter we will explore the use of another approach for spatial context which uses the geodesic distances along the image intensities as an alternative to using the Euclidean distance.

Chapter 5

Geodesic Patch-based Segmentation

This chapter is based on the following publication:

- Zehan Wang, Kanwal K. Bhatia, Ben Glocker, Antonio de Marvao, Tim Dawes, Kazunari Misawa, Kensaku Mori, Daniel Rueckert. Geodesic Patch-based Segmentation *Medical Image Computing and Computer-Assisted Intervention – MICCAI*. LNCS Volume 8673, pages 666-673. Springer International 2014.

5.1   Introduction

So far, we have looked at the use of spatial context within a patch-based segmentation framework to regularise patch selection, enabling similar looking patches from different structures to be distinguished and removing the requirement for the use of search windows. The previous chapter introduced a new approach to provide spatial context using relative distances between anatomical structures to provide an adaptive coordinate system which is independent of voxel-wise image correspondences. This then lead to an investigation for applying it as a plausible approach for segmentation without registration, which could also be used to cope with registration errors.

Previously, the EDT was applied from labelled structures to provide relative distances for use as spatial context, but the Euclidean distance is sensitive to anatomical variability as well as the quality of initial segmentation. As spatial features to regularise patch comparisons, it would be desirable to have relative distances between structures which are comparable in the presence of shape and size variability between different subjects, particularly for applications where accurate registration can prove challenging. In these circumstances, using the Euclidean distance may be insufficient for this purpose, particularly since it provides a linear measure which ignores any information within the image such as boundaries between structures or the presence of flat homogeneous regions.

As an alternative, this chapter looks at using the geodesic distances within the image, which are able to contribute information on the locality of structure boundaries, to provide spatial context and an *anatomically-adaptive* coordinate system. The use of geodesic distances has been shown to be effective in interactive segmentation [49] and this is adopted within the spatially aware patch-based framework proposed in previous chapters. In addition to the proposed application of geodesic distances as spatial features, it has also recently been used with random forest classifiers in [96].

5.1.1   Chapter Overview

This chapter begins by presenting a new approach for spatial context, based on using the geodesic distance within the image, and showing how it can be calculated efficiently using a distance transform. This approach is used within the multi-resolution segmentation frame-work presented in the previous chapter, and evaluated with multi-structure segmentation in 20 cardiac MR images as well as multi-organ segmentation in 150 abdominal CT images.

5.2 Anatomically-Adaptive Coordinate System

To overcome the shortfalls of the Euclidean distance in providing adaptive spatial context, geodesic distances within the image could be used instead. This takes into account image gradients and describes distance between two points using the shortest path along the image intensities rather than just through physical (empty) space.

| (a) Euclidean distance | (b) Geodesic distance |
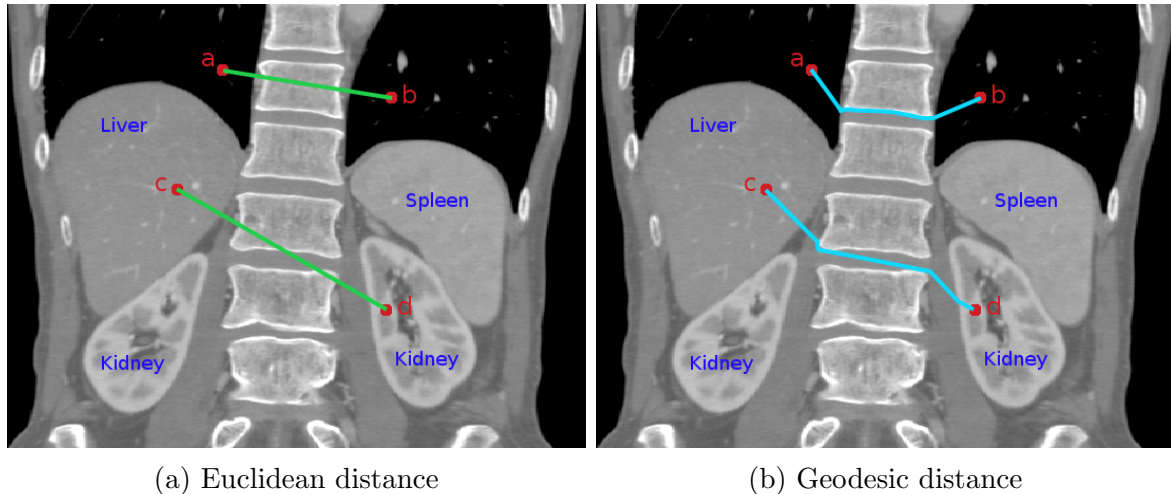
Figure 5.1: A comparison of Euclidean distances in (a) and geodesic distances in (b) within an image, shown in green and cyan respectively.

In general, the geodesic distance between two points $\mathbf{x}$, $\mathbf{y}$ within an image I is defined as follows:

$$d(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{\Gamma} \in \mathbf{P}_{\mathbf{x},\mathbf{y}}} \int_0^{l(\mathbf{\Gamma})} \sqrt{1 + \gamma^2 (\nabla I(\theta) \cdot \mathbf{\Gamma}'(\theta))^2} \, \mathrm{d}\theta \qquad (5.1)$$

where $\mathbf{\Gamma}$ is a path in the set of all paths, $\mathbf{P}_{\mathbf{x},\mathbf{y}}$ between $\mathbf{x}$ and $\mathbf{y}$ and is parametrised by its arclength $\theta \in [0, l(\mathbf{\Gamma})]$. The Euclidean distance can be considered a special case of the geodesic distance, since these are equivalent when γ is set to 0, and in practice it is possible to tune for some balance between a purely Euclidean and purely gradient-derived distance measure, depending on the application.

The same principle in providing spatial context is use as that described previously in Chapter 4.2.1, using relative distances to structures, but using geodesic distances instead of Euclidean distances. This provides a coordinate system that is then not only adaptive to the anatomical

structures used to provide relative distances, but also adaptive to anatomical features within the image between these structures. This substantially alters the relative distances between structures as the distances are now affected by visible boundaries and takes into account the image gradients between structures (see Figure 5.2). Using geodesic distances from multiple structures then allows patches to be localised in a way that is more comparable between different subjects than using Euclidean distances, particularly when there are large differences in the physical sizes of the subjects' anatomies.

(a)   (b)
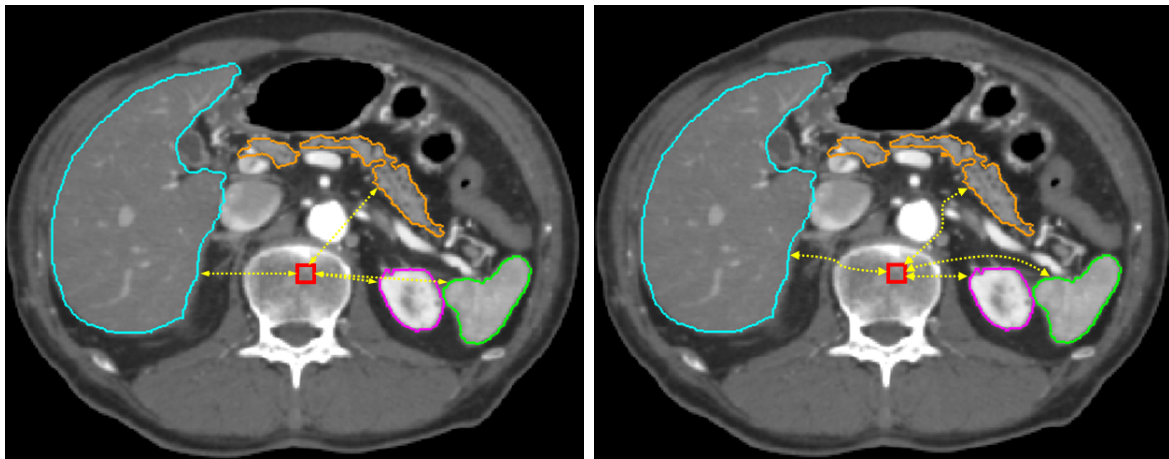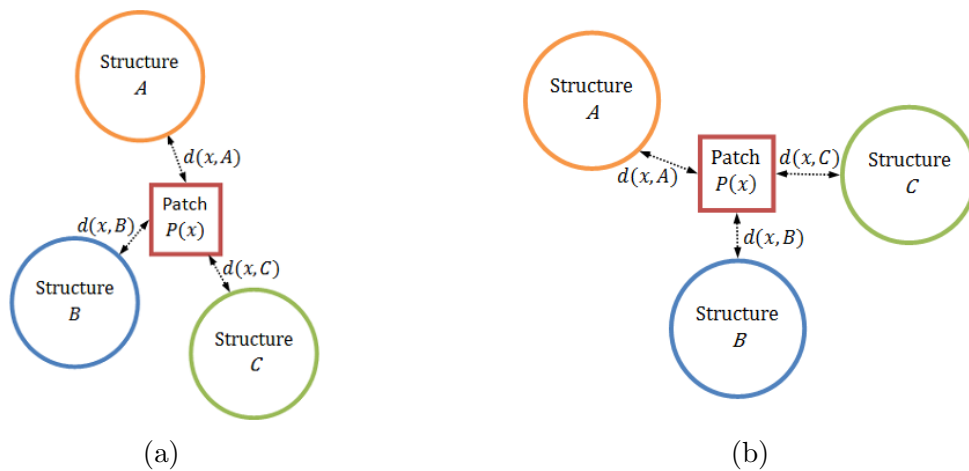
(c) Euclidean distances   (d) Geodesic distances

Figure 5.2: (a) and (b) represent the same anatomical structures but the two images are not aligned. Spatial context for patch $P(x)$ can be provided by the distances to these structures regardless of the how they are positioned within the image. (c) and (d) provide an example where this type of spatial context can be used and how the distances will be different using the geodesic distances compared to Euclidean distances.

To efficiently calculate geodesic distances for all pixels in an image, a distance transform can be applied in a similar fashion to the EDT which was used in the previous Chapter. This

calculates the minimum distances to a reference point or labelled structure within the image and provides a distance map, describing the distances for each of the pixels to the structure. This can be performed for multiple structures to obtain the relative distances to each, thus providing spatial context in the form $S_{\text{adaptive}}(\mathbf{x}) = [d_{\min}(\mathbf{x}, R_1), d_{\min}(\mathbf{x}, R_2), ..., d_{\min}(\mathbf{x}, R_n)]$ for each pixel $\mathbf{x}$ and structures R_i as previously defined in Chapter 4.2.1.

5.2.1 Geodesic Distance Transform (GDT)

There are several geodesic distance transform algorithms [185]; this chapter uses the approach from [168], which was also used in [49] and demonstrated to have good performance with linear computational requirements. This approach uses a local wavefront propagation algorithm with a fast sweeping method [195], which iteratively sweeps from one corner of the image to the opposite corner in alternating directions, updating the distances of each path in a local kernel in favour of the shortest.

This requires discretising the image domain and the geodesic distance formulation in (5.1), then traversing the image grid using a local neighbourhood structure. Given two pixels $\mathbf{x}$ and $\mathbf{y}$, paths between them can be constructed using chains of neighbouring pixels $(\mathbf{s}_0 = \mathbf{x}, \mathbf{s}_1, ..., \mathbf{s}_n = \mathbf{y})$. Along such a chain, the integral in (5.1) can be approximated as

$$\sum_{i=1}^{n} [|\mathbf{s}_i - \mathbf{s}_{i-1}|^2 + \gamma^2 |I(\mathbf{s}_i) - I(\mathbf{s}_{i-1})|^2]^{\frac{1}{2}} \tag{5.2}$$

Based on this, the GDT can then be calculated by finding the minimum paths. This can be performed using a wave-front propagation approach, where the distances for the current pixel $\mathbf{x}$ is updated according to the minimum:

$$D(\mathbf{x}) = \min\{D(\mathbf{x} + \mathbf{a}_i) + [|\mathbf{a}_i|^2 + \gamma^2 |I(\mathbf{x}) - I(\mathbf{x} + \mathbf{a}_i)|^2]^{\frac{1}{2}}\} \tag{5.3}$$

D is current depth map and eventual output of the GDT, and $\mathbf{a}_i$ defines the local neighbourhood

which represents the local wave-front. The fast sweep method updates the distances by first sweeping from one corner of the image to the opposite corner and then sweeping in the opposite direction, e.g. the upper-top-left corner to the lower-bottom-right corner and vice versa. A local window of size 3×3×3 is usually sufficient in providing the local neighbourhood for the GDT, although larger ones could be used which produce better approximations with each sweep but with a significant reduction in speed [49]. It takes a few iterations of this to converge to the distance of the shortest path from a seeded region to all other points in the image, and thus provide the correct distance map. Results from [168] suggests it requires 3-10 iterations, depending on the nature and size of the image. The algorithm has an optimal computational complexity of $\mathcal{O}(N)$ where N is the number pixels, and accesses the image data in memory in contiguous blocks allowing it run quite fast in practice.

(a) GDT map from kidneys   (b) GDT map from pancreas

(c) GDT map from liver   (d) Original image   (e) EDT map from liver
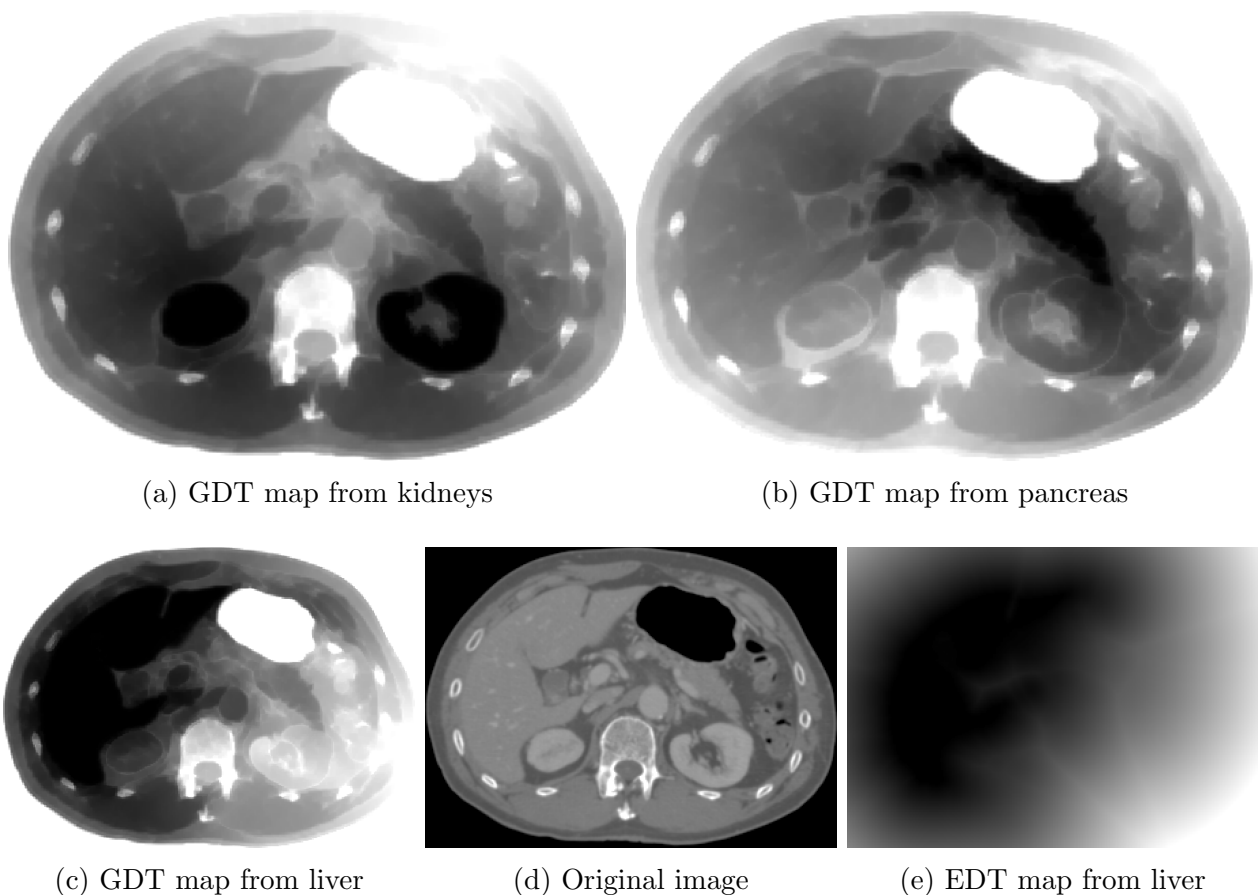
Figure 5.3: Comparing the EDT to the GDT as distance maps created using eroded (×2) versions the anatomical structures in abdominal CT images. Distances are shown as being proportional to the intensity.

Examples of distance maps using geodesic distances in comparison to the Euclidean distance

| (a) Original image | (b) EDT map | (c) GDT map |
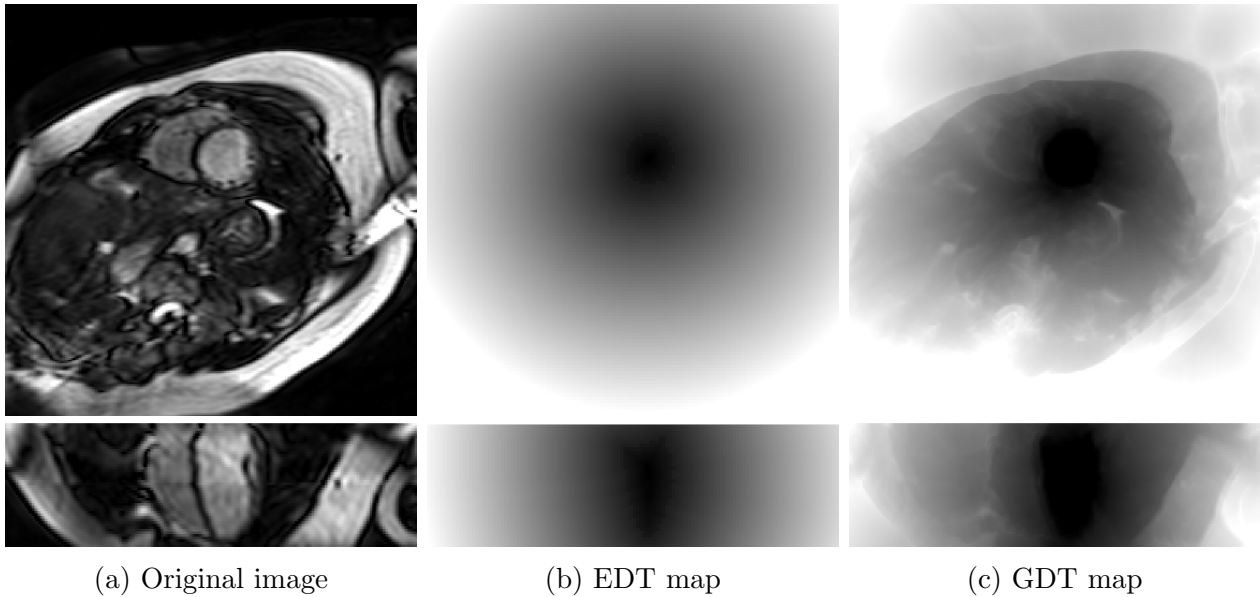
Figure 5.4: Comparing the EDT to the GDT as distance maps created using eroded ($\times 2$) versions the left ventricle in cardiac MR images, shown from two perpendicular view points. Distances are shown as being proportional to the intensity.

is shown in Figures 5.3 and 5.4. Here it can be seen, that contrary to the EDT distance map, structure boundaries are clearly visible and distances within the same anatomical structures are more similar than that from other structures in the GDT distance maps.

5.3 Framework Overview

As an overall segmentation framework, the multi-resolution approach previously introduced in Chapter 4.3 can be adopted with the same principles for patch-based segmentation as in Chapter 3. Although it is plausible to perform segmentation without registration using the proposed approach for spatial context, it is difficult to obtain an initial segmentation with sufficient robustness for all applications without registration. This was seen with the results of the canine leg MRI challenge in the previous chapter, where errors in the initial segmentation could not be fully corrected by the refinement process, which also adversely affected the relative distances used for spatial context. This could potentially be addressed with a coarse registration so that the organs are in the same orientation and share a similar frame of view, enabling the initial segmentation to benefit from the use of coordinates as spatial context. This can be achieved

using linear registration without requiring non-rigid deformable registration. Registration failures can still occur depending on the application, but only a subset of atlases are required for the initial segmentation.

5.3.1   Initial Segmentation

Given that images are coarsely aligned, there are several options for an initial segmentation to enable the use of the GDT for spatial context. One possibility is to use the approach used previously in Chapter 4.3, where coordinates are used as spatial context with a weak spatial weight in very low resolution. One issue with this approach is that the resolution must be low enough such that it is computationally plausible to perform image-wide kNN patch search without relying upon well defined and concise ROI masks. This may cause problems if the structures of interest are not distinguishable at this resolution, resulting in incorrectly labelled initial segmentations.

Another possibility is to establish globally defined reference structures based on the binary intersection of each label from the atlases. This can then be used to provide relative distances to patches for all images. This option has the benefit of being fast and simple to obtain but relies upon the registration outcome to be sufficiently accurate such that the the intersection of the atlases does not yield an empty set. This is unlikely to be the case in images where there is substantial variability in the location of the anatomical structures.

Finally, the initial segmentation could be obtained by another independent segmentation approach. The framework used in this chapter enables a boundary region to be set for any segmentation and to refine it using the proposed patch-based approach with spatial context derived from the eroded versions of the structures. The region for refinement can be established by taking the difference between the dilation and the erosion of the structures as shown in Figure 4.3 in the previous Chapter.

Ultimately, the choice in how to obtain the initial segmentation, as well as the resolution to obtain it at, may be dependent on the application. For example, in the cardiac dataset, the

intersection of the atlases can be used, but in the abdominal dataset, the intersection of the atlases results in an empty set. Further details are presented on the different approaches with regards to specific applications in the next section.

5.3.2 Segmentation Refinement

As part of the multi-resolution framework, once an initial segmentation has been established, only the boundary region for each label requires refinement after each iteration. In the same manner as presented in Chapter 4.3.1, the boundary region is defined by the morphology gradient (dilation minus erosion) of each structure. The boundary region size is controlled by the number of dilations and erosions performed before taking their differences.

5.3.3 Atlas Selection

Another consequence of using registration is that it enables the use of established atlas selection schemes in MALP. This allows the approach used in Chapter 3.3.2 to be adopted, where a pixel-to-pixel distance measure such as SSD or Euclidean distance is used. In this chapter, the SSD is used when atlas selection is required, in the same manner as was previously described in Chapter 3.

5.3.4 Label Fusion

The label fusion uses the same spatially aware label fusion approach that was introduced in Chapter 3.2, using relative geodesic distances, calculated with the GDT, as the spatial component. The GDT is always calculated in the native resolution (and downsampled if required) so that the same spatial weighting α can be used at all resolutions.

5.4   Application to Cardiac MRI Dataset

5.4.1   Dataset

Experiments were performed using end-diastolic frames of cardiac MR images, captured from 20 subjects using a 1.5T Philips Achieva system, evaluating the segmentation of the left ventricle, myocardium and right ventricle. The images were acquired in a single breath-hold using cine imaging sequences with 2mm gaps in the left ventricular short axis direction and have a native resolution of $256{\times}256{\times}64$ voxels with voxel sizes of $1.25{\times}1.25{\times}2\text{mm}^3$. The ground truth segmentations were provided by two experts manually labelling each voxel using freely available software (ITK-SNAP). In addition, they also provided six pre-defined landmarks on each image, so that the different orientation of the heart could be accounted for when aligning the images. The landmarks were mostly based around of the left ventricle and myocardium, with a single landmark on the right ventricle (see Figure 5.5).

The experiments also compared using the GDT with using the EDT in providing relative distances for an adaptive coordinate system as well as using explicit image coordinates as spatial context. In addition, these methods using spatial context were also compared to the standard patch-based approach from Coupé [47]. Each method was evaluated using leave-one-out cross validation with all available atlases (19) to segment each test image. Affine registration using only the six pre-defined landmarks were applied to align the atlases to each test image in turn.

5.4.2   Implementation

Atlas selection was not used due to the low number of available atlases (19 atlases for each test image). For the methods using relative distances (using the GDT or the EDT), the initial segmentation is defined by the intersection of the atlases, and distances to the left ventricle, right ventricle and background labels were used to provide spatial context. These methods used a multi-resolution framework to refine the quality of the segmentations that provide the relative distances as spatial context. In total, 3 resolution levels were used, with the lowest
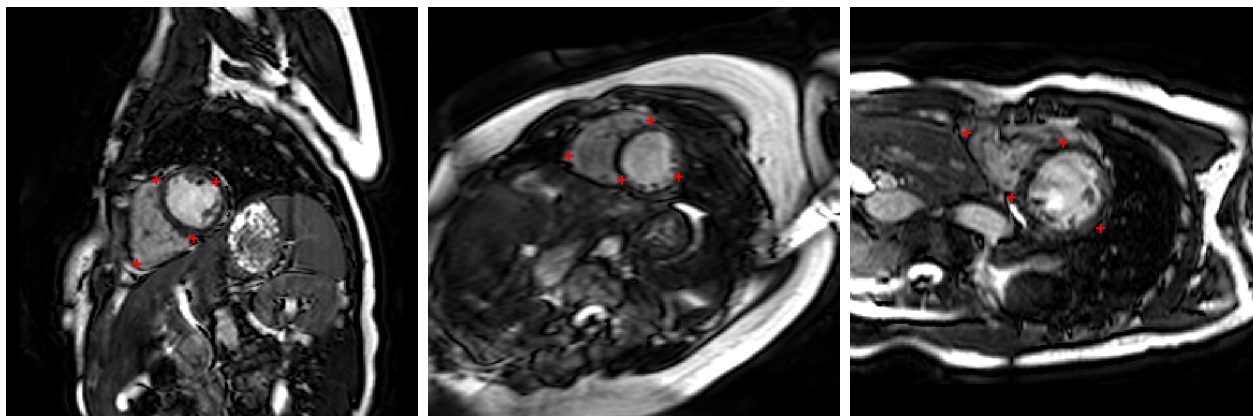
(a) Landmarks
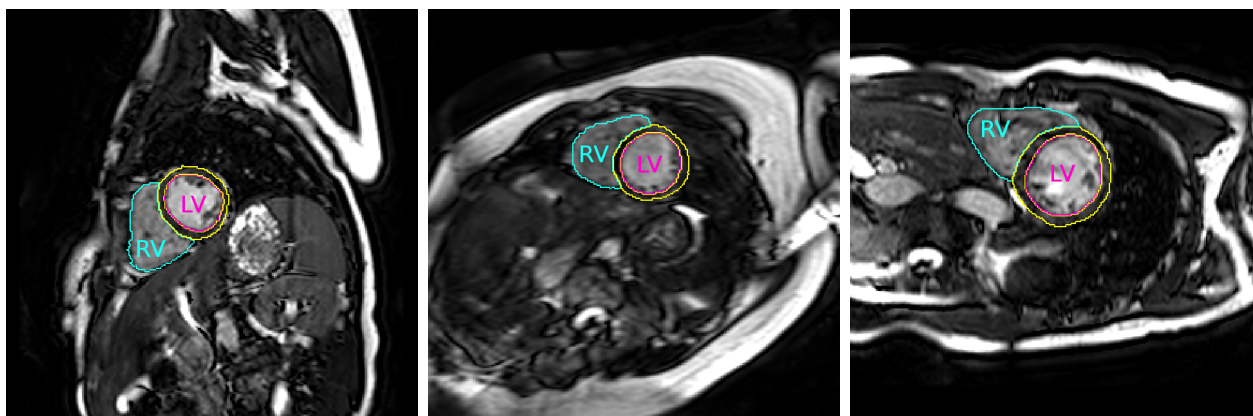
(b) Segmentations

Figure 5.5: Examples from the cardiac dataset in their native space, shown all at the same scale. (a) shows four of the landmark positions as red crosses. These four landmarks are positioned in the middle slice along the long axis of the left ventricle, whilst the other two landmarks (not shown) are positioned at the top and the bottom of the centre of the left ventricle. Reference Segmentations are shown in (b) with left ventricle (LV), myocardium, right ventricle (RV) outlined in pink, yellow and cyan respectively. The myocardium is the wall around the left ventricle.

resolution at 5×5×5mm³ voxel sizes, the intermediate level at 2.5×2.5×2.5mm³ voxel sizes and the highest at the native resolution with 1.25×1.25×2mm³ voxel sizes. For the standard patch-based approach and the approach using explicit image coordinates, the segmentation outcome using multi-resolution and single resolution approaches would be the same since they are not affected by the intermediate results as they are not adaptive to any changes in the segmentation between iterations.

A patch size of 5×5×5 voxels was used for all methods and all resolutions and k was fixed at 40 for the kNN methods. For the different approaches to spatial context, α was selected in the lowest resolution and then applied for all subsequent resolutions whilst γ was set at 100 for GDT. For spatial context using image coordinates, the weights are adjusted for each resolution to reflect the same overall effect in real-world distance. The parameters and specifications used are summarised in Table 5.1. Total computational time was about 4 hours per image running on a 16-core machine clocked at 2.8Ghz.

Table 5.1: Table of parameters for the cardiac MRI dataset.

Method/Description	Parameter(s)
Registration Type	Affine (landmark-based) - 9 parameters (no shearing)
Number of Landmarks	6
Intensity Range	$[0, 100]$
Patch Size	5×5×5
k	40
Boundary Size	3
Number of Resolutions	3

5.4.3   Experiments and Results

Effect of the Spatial Weight α

The results, presented in Figure 5.6, suggest that for each method, there is a different optimal spatial weighting α and this is also somewhat dependent on the application. The results also suggests that, even in the lowest resolution, methods using adaptive spatial context are able to provide more accurate segmentations than the other methods. Using these results, α=7, α=13 and α=5 provide suitable parameters to use, respectively, for coordinates, EDT and GDT as

(a) Left Ventricle

(b) Myocardium

(c) Right Ventricle

(d) Key
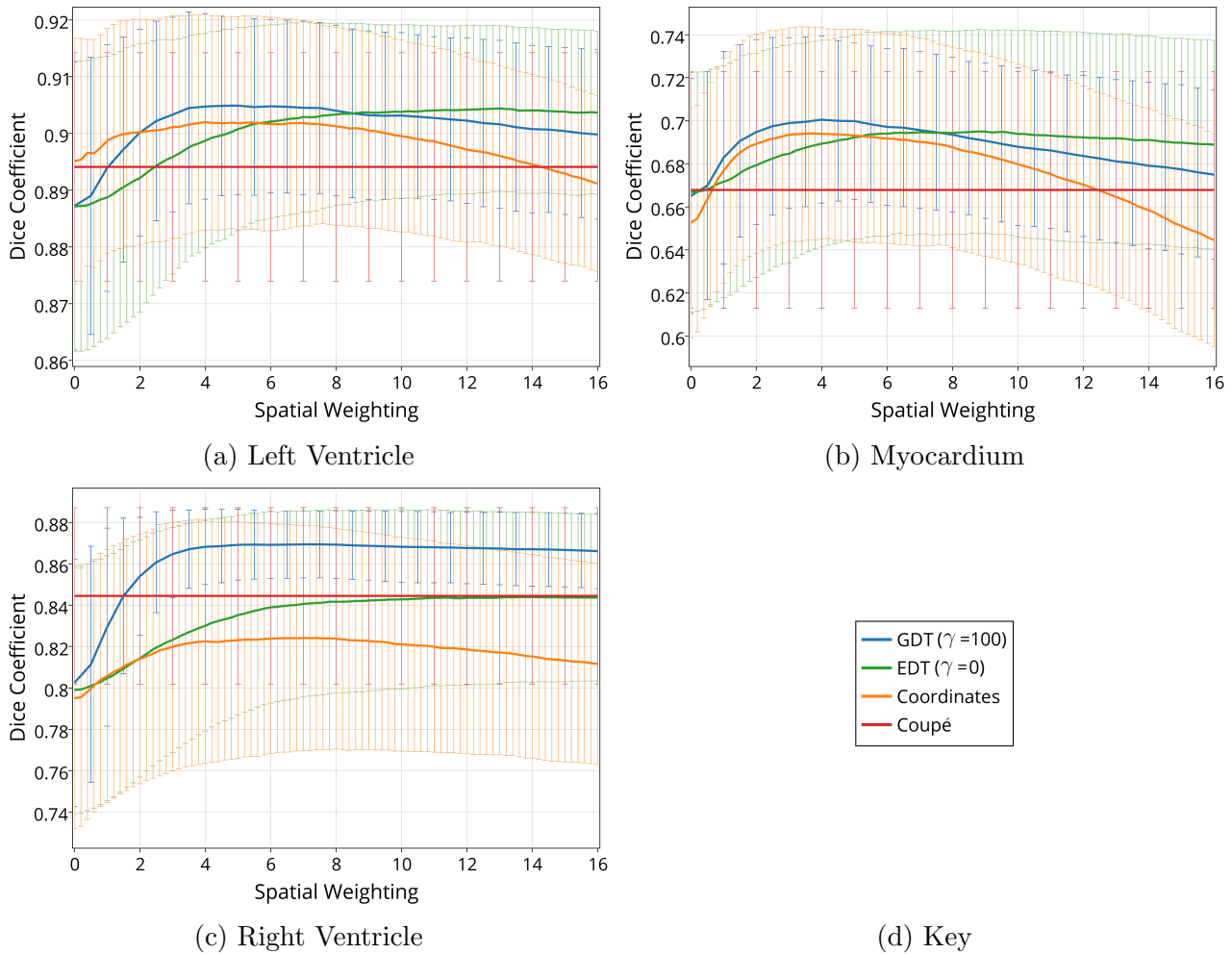
Figure 5.6: Comparison of mean segmentation accuracy in the initial resolution level with regards to spatial weighting and different spatial context. Error bars represent the standard deviation. Results using a standard patch-based approach [47] with a window size of 3 voxels is shown for comparison.

spatial context in subsequent resolutions. In general, there is a minimum value for which α should be to have a positive effect, but the effect of changing α is fairly stable and predictable; a rough ballpark number can provide reasonable results and precise tuning of this parameter is not required.

Final Segmentation

Figure 5.7 and Table 5.2 summarises the final segmentation accuracy for all approaches to providing spatial context as well as using a standard patch-based approach with different search window sizes (WS) for patch comparison. Examples of segmented images from all methods are
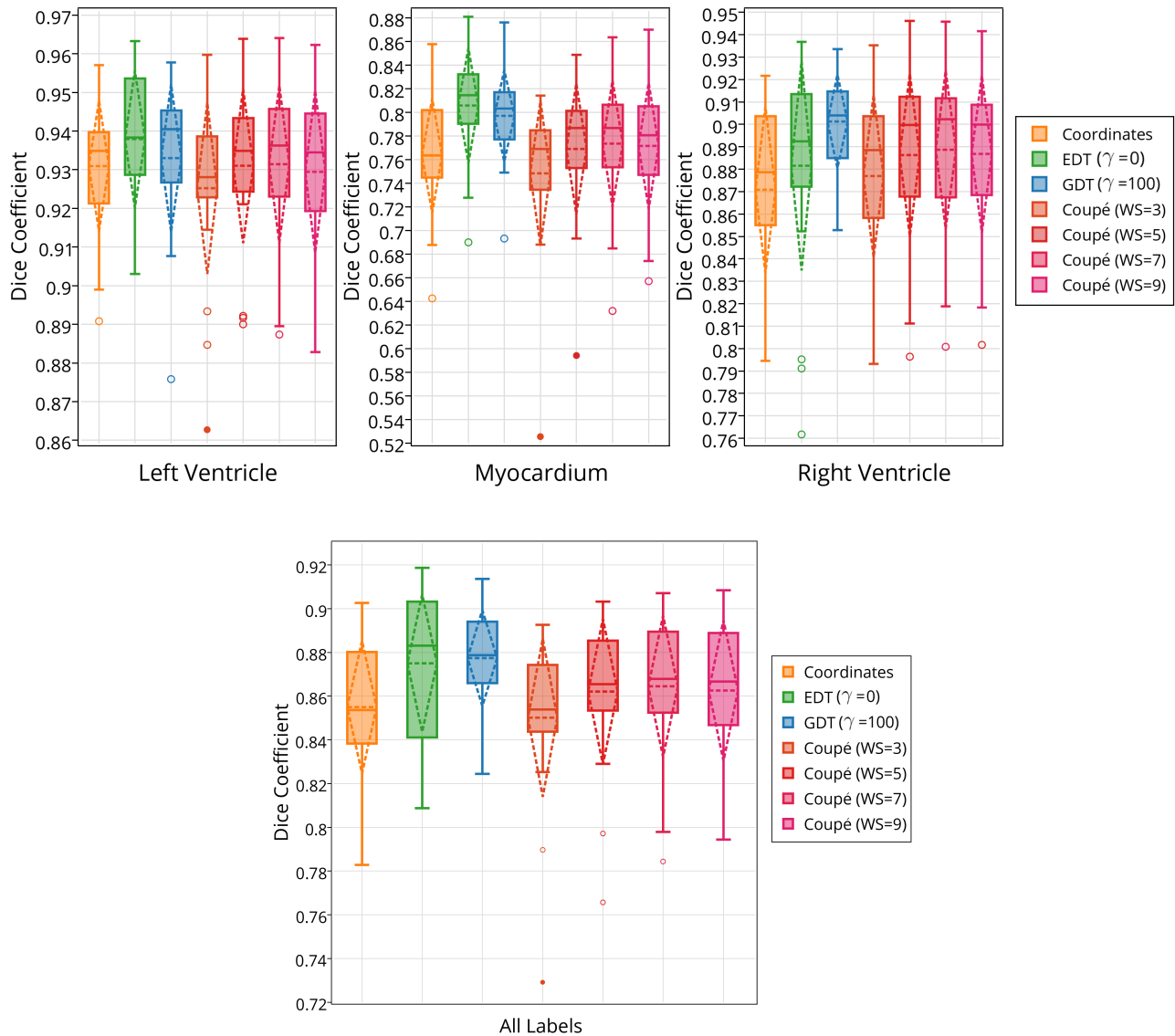
Figure 5.7: Comparison of segmentation accuracy for each label with regards to different spatial context and different search window sizes. Solid line indicates the median, the dashed line indicates the mean and standard deviation is shown by the dashed diamond.

presented in Figure 5.8.

Refining an Existing Segmentation

As another possibility for the initial segmentation, a segmentation provided by an independent segmentation process could be used as the initial segmentation. The boundary regions could be refined using the proposed framework, much in the same manner as if was the result was from a previous iteration. To demonstrate this, experiments were conducted using the results from applying Coupé's patch-based method [47] with a window size of 7^3 voxels (see Table 5.2) as

(a) Coordinates (Dice: 0.938, 0.797, 0.905) (b) Coordinates (Dice: 0.944, 0.801, 0.912)

(c) EDT (Dice: 0.936, 0.828, 0.919) (d) EDT (Dice: 0.961, 0.859, 0.937)

(e) GDT (Dice: 0.929, 0.801, 0.905) (f) GDT (Dice: 0.953, 0.838, 0.934)

(g) Coupé [47] (Dice: 0.942, 0.801, 0.922) (h) Coupé [47] (Dice: 0.947, 0.823, 0.925)

(i) Coupé [47]+GDT (Dice: 0.938, 0.808, 0.930) (j) Coupé [47]+GDT (Dice: 0.947, 0.831, 0.930)

(k) Reference Segmentation (l) Reference Segmentation
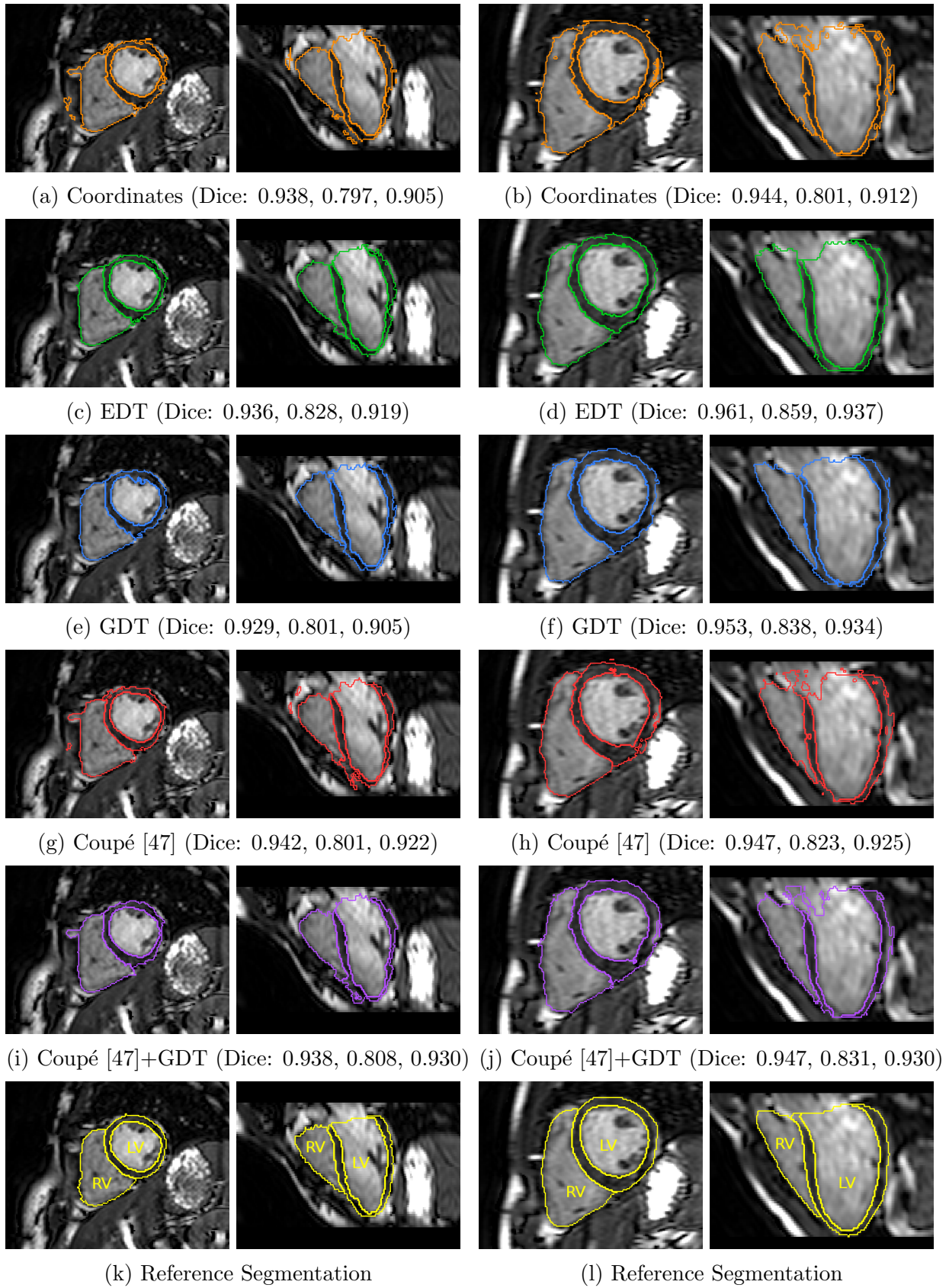
Figure 5.8: Examples segmentations from all methods for two subjects, shown from two orthogonal directions. Dice coefficients are shown for the left ventricle, myocardium and right ventricle respectively. "[47]+GDT" indicates results which used [47] (WS=7) to provide the initial segmentation, followed by refinement with γ=100.

Table 5.2: Final segmentation accuracy from all methods for each label. "[47]+GDT" indicates results which used [47] (WS=7) to provide the initial segmentation. Dice coefficients are shown as mean (median) ±standard deviation. Best values are show in red.

Method/Description	Left Ventricle	Myocardium	Right Ventricle
Coordinates (α=7)	0.931 (0.934) ±0.016	0.763 (0.763) ±0.049	0.871 (0.879) ±0.037
EDT (α=13, γ=0)	0.938 (0.938) ±0.017	0.806 (0.814) ±0.049	0.882 (0.893) ±0.047
GDT (α=5, γ=100)	0.934 (0.941) ±0.019	0.797 (0.803) ±0.039	0.901 (0.904) ±0.021
Coupé [47] (WS=7^3)	0.931 (0.936) ±0.020	0.773 (0.787) ±0.053	0.889 (0.902) ±0.035
[47]+EDT (α=10, γ=0)	0.940 (0.942) ±0.017	0.803 (0.814) ±0.046	0.896 (0.906) ±0.033
[47]+GDT (α=10, γ=0.1)	0.938 (0.940) ±0.018	0.802 (0.816) ±0.047	0.897 (0.910) ±0.033
[47]+GDT (α=10, γ=1)	0.937 (0.941) ±0.017	0.789 (0.793) ±0.048	0.900 (0.911) ±0.031
[47]+GDT (α=10, γ=10)	0.935 (0.939) ±0.016	0.783 (0.786) ±0.044	0.897 (0.908) ±0.031
[47]+GDT (α=10, γ=100)	0.935 (0.938) ±0.016	0.784 (0.782) ±0.041	0.894 (0.908) ±0.031

Method/Description	All Labels
Coordinates (α=7)	0.855 (0.854) ±0.030
EDT (α=13)	0.875 (0.883) ±0.031
GDT (α=5)	0.877 (0.879) ±0.022
Coupé [47] (WS=7^3)	0.865 (0.868) ±0.032
[47]+EDT (α=10, γ=0)	0.879 (0.884) ±0.026
[47]+GDT (α=10, γ=0.1)	0.879 (0.882) ±0.028
[47]+GDT (α=10, γ=1)	0.875 (0.880) ±0.027
[47]+GDT (α=10, γ=10)	0.872 (0.875) ±0.025
[47]+GDT (α=10, γ=100)	0.871 (0.874) ±0.026
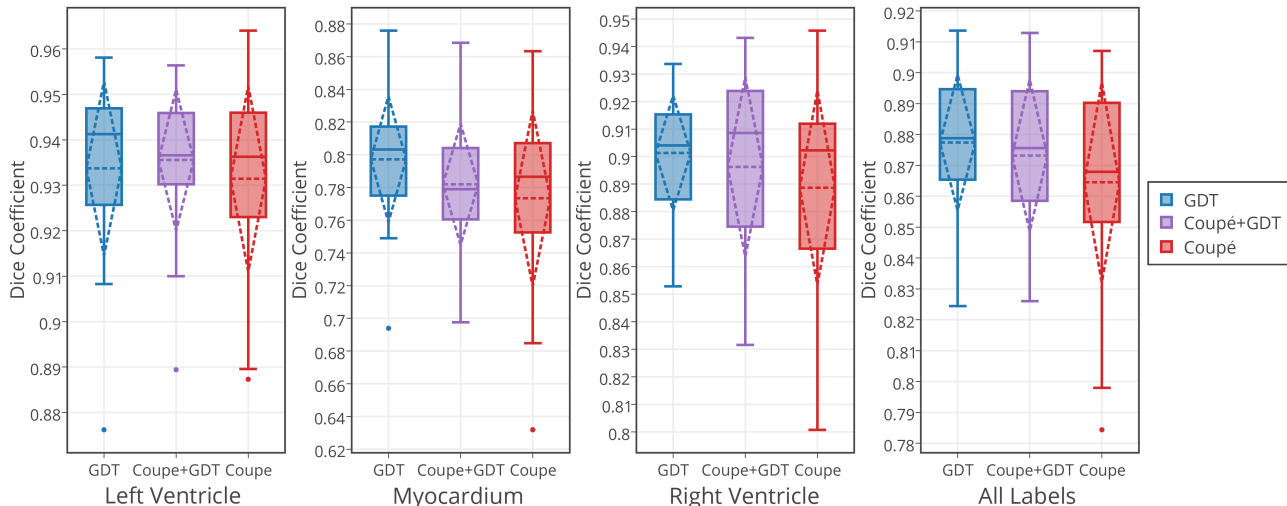
Figure 5.9: Comparison of segmentation accuracy for each label with regards to refining an existing segmentation. "Coupé+GDT" indicates the results which used [47] (WS=7) to provide the initial segmentation, followed by refinement with γ=100. Solid line indicates the median, the dashed line indicates the mean and standard deviation is shown by the dashed diamond.

the initial segmentation. This was refined in the native resolution, using the GDT from eroded versions of the background, left ventricle, myocardium and right ventricle to provide spatial context. Further to this, several different γ values were also applied under this scenario.

The results are presented in Figure 5.9 and Table 5.2. Comparing the refined results (γ=100) and the original unrefined results using Welch's paired two sample T-test yields p-values of 0.11, 0.023, 0.00061 and 0.0076 for the left ventricle, myocardium, right ventricle and overall across all labels. This represents the extreme case, and there is further improvement as γ is reduced, thus p-values would be also be smaller for smaller values of γ as the differences are more significant. The best values for γ seem to be different for the different structures, with $\gamma = 1$ providing the best results for the right ventricle, whilst $\gamma = 0$ provides better results for the left ventricle and myocardium.

5.4.4 Discussion

The results indicate that using relative geodesic distances to provide spatial context is competitive with other methods and is often able to provide more accurate and consistent results on average. The geodesic parameter γ was not tuned for the experiments, but it could be optimised for a performance that is between the results given by using the EDT (γ=0) and the GDT(γ=100) presented here. The results from using several different γ values for the GDT suggest that the best value is likely to be dependent on the application and may differ from structure to structure.

One potential reason for the GDT not outperforming the EDT in segmenting the left ventricle and myocardium is due to the lack of visible boundaries in certain parts of the image, leading to an overflow in the segmentation, especially as local patches are also similar in these areas. This is most evident at the top of the left ventricle, as can be seen in Figure 5.10. In the outlying case, part of an unknown image artifact in the left ventricle was incorrectly identified as myocardium. This is likely caused by the local appearance similarities to the myocardium and the initial segmentation not establishing it as part of the left ventricle, which then leads to a higher geodesic distance in that area compared to the rest of the left ventricle. Using

(a) Segmentation overflow at the top of the left ventricle

(b) Image artifact in the left ventricle incorrectly labelled as myocardium
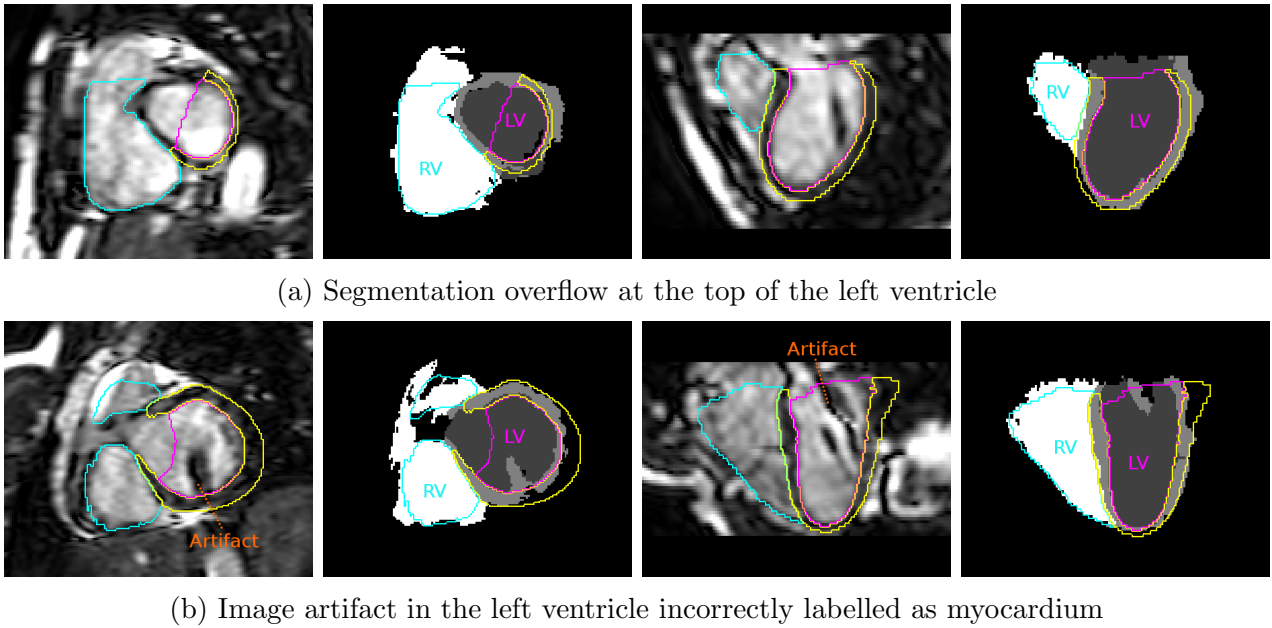
Figure 5.10: Example showing where using the GDT performs relatively poorly. Reference segmentations for the left ventricle (LV), myocardium and right ventricle (RV) are outlined in pink, yellow and cyan respectively, whilst the segmentation from using the GDT is shown in grey and white.

the Euclidean distance ignores image artifacts and limits the overflow in the segmentation more than the geodesic distance, thus providing a more accurate segmentation. This could be addressed by lowering the weight γ for the gradient component of the geodesic distance, which would increase the weight of the Euclidean component of the distance measure and improve the segmentations. This would also lower the impact of noise in images, which would increase the geodesic distance in otherwise homogeneous regions and make distances less comparable between images with differing noise-to-signal ratios.

One region where using the GDT notably outperformed the other methods is with the right ventricle. A clear example is shown in Figure 5.11. Here using the EDT yielded many more outliers, which lowered the average performance. The cause for this is likely caused by the choice of landmarks for registration which were focused around the left ventricle. This would most likely not have been enough to fully capture all the anatomical differences, resulting in size differences for the right ventricle after registration. This would also explain why using explicit coordinates and the standard patch-based approach also gave a much bigger range in results. Using the GDT to provide spatial context provided more comparable spatial features

(a) EDT (0.762)   (b) GDT (0.899)   (c) Reference

(d) Coordinates (0.793)   (e) Coupé (0.819)   (f) Coupé+GDT (0.832)
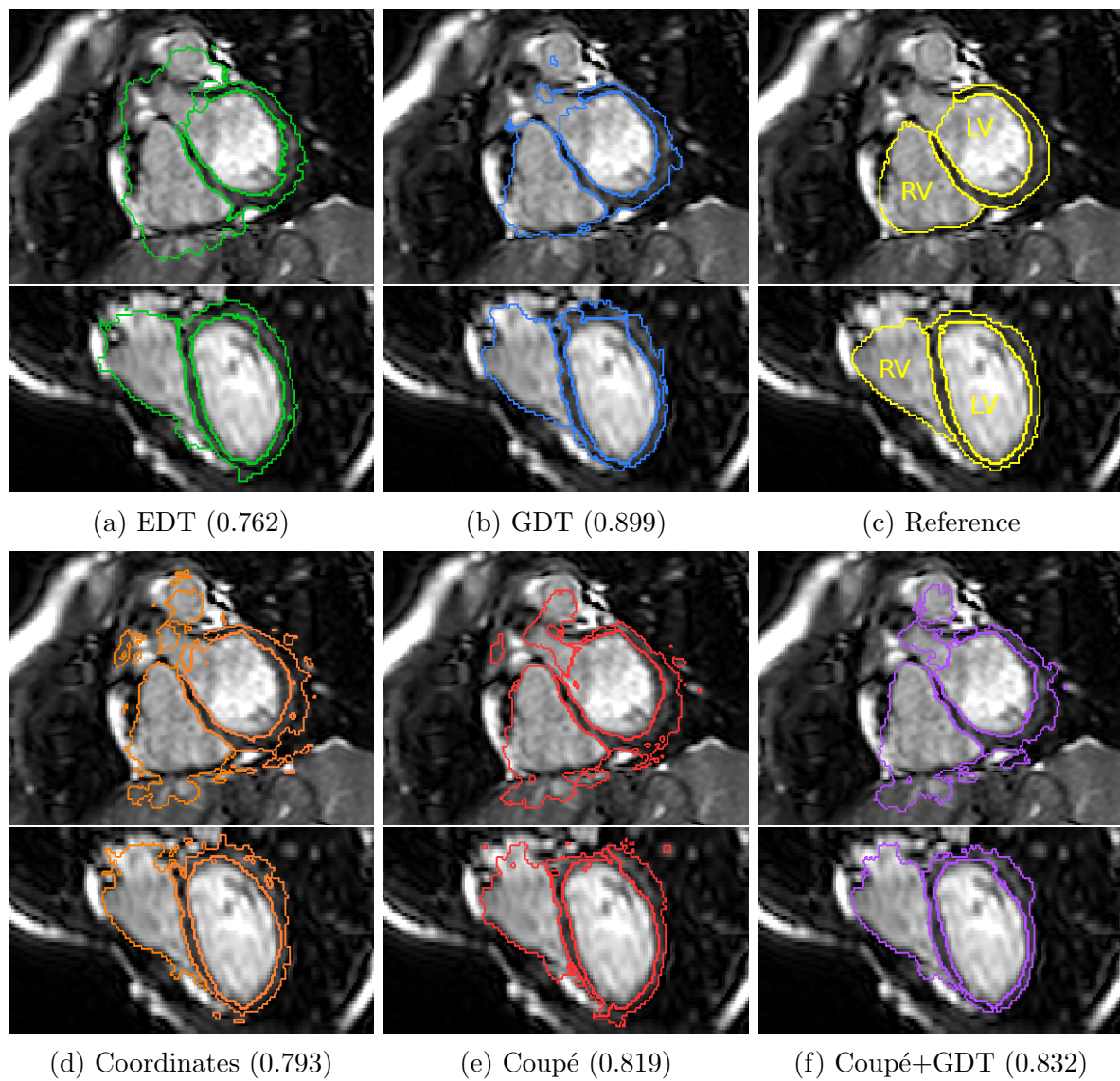
Figure 5.11: Example, shown from two orthogonal directions, where using the GDT clearly outperforms the other approaches for the right ventricle. Dice coefficients are shown in brackets for the right ventricle only.

across different subjects and accommodated for the anatomical variability. In contrast, for the left ventricle, there were generally less variance in the results for all methods since the majority of the landmarks were defined around this region.

The ability to refine an existing segmentation from an established segmentation method also demonstrates the validity and versatility of the refinement approach in the proposed framework. For all labels, the refinement of [47] resulted in improved results on average and significantly so for the right ventricle. When using a high γ value, the left ventricle and myocardium did not show as big a difference, largely due to the segmentation overflow and image artifacts as just discussed. However, for smaller values of γ, the improvements become much more significant.

Overall, the results show that using the GDT to provide spatial context works well and is adaptive to different registration outcomes in this dataset. Although, it did not always outperform the EDT in providing spatial context when there is low contrast or in the presence of image artifacts, the results still demonstrate great potential in this approach.

5.5   Application to Abdominal CT Dataset

5.5.1   Abdominal CT Dataset

In addition to the cardiac dataset, the proposed framework was also evaluated using leave-one-out cross validation on abdominal CT scans of 150 subjects, both male and female, ageing from 26 to 84 years with an average age of 62.8±12.0. The images have an in plane resolution of 512×512 voxels with voxel sizes ranging from 0.55 to 0.82mm and contain between 263 to 538 slices with spacing ranging from 0.4 to 0.8mm depending on the field of view and the slice thickness. All images were acquired from the Aichi Cancer Center in Nogaya, Japan for the purpose of laparoscopic resection of the stomach and gallbladder glands or colon. Out of the 150 subjects, 141 subjects had early or advanced gastric cancer, one subject had cholecystitis cancer and eight subjects had colorectal cancer.

For each scan, manual segmentations of the liver, spleen, pancreas and the kidneys were pro-
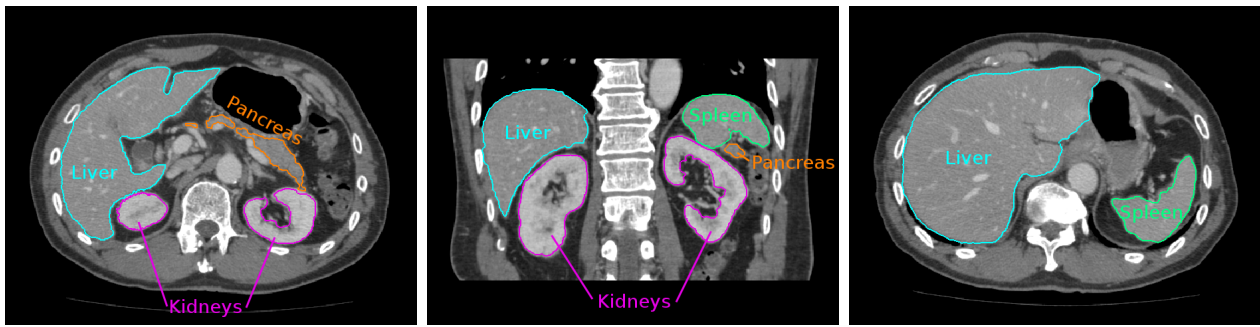
(a) Subject 1
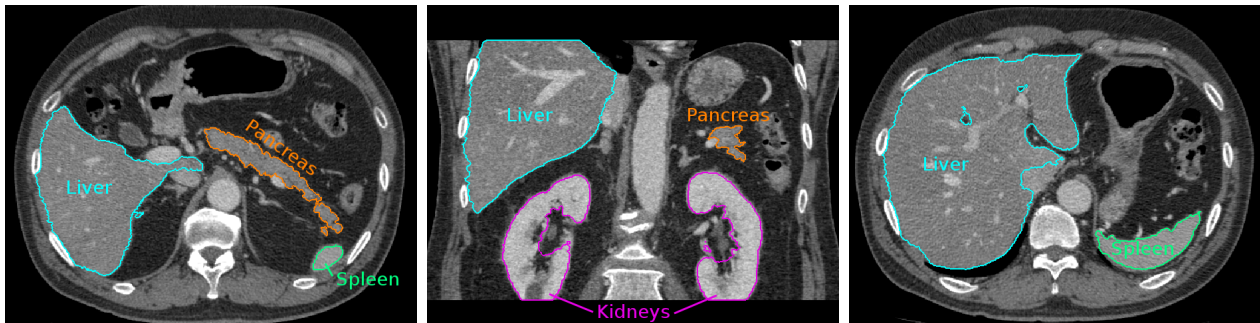
(b) Subject 2

(c) Subject 3
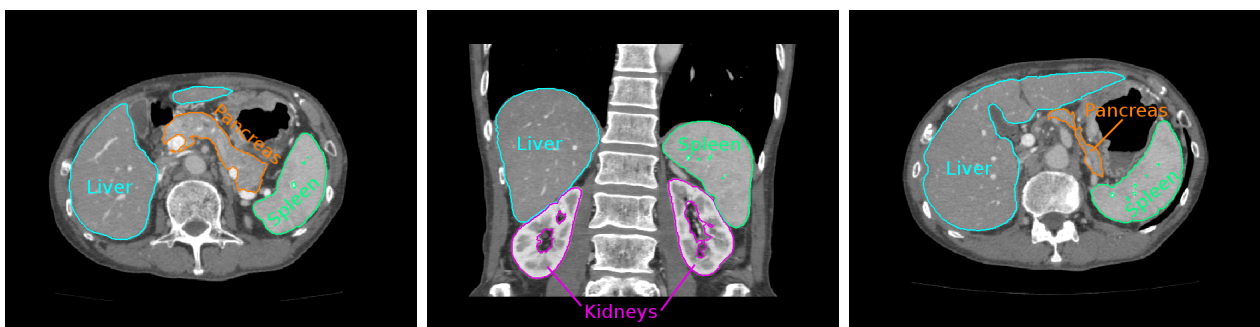
(d) Subject 4

Figure 5.12: Examples of 4 random subjects from the abdominal dataset. Images differ in their quality and have high variability in size, shape and location of organs between subjects. In some images, some of the organs have been clipped during the scan - as can be seen in subject 2. Images have been intensity thresholded to the range [1800, 2400] and are presented at the same physical scale in their native space.

duced by one of three trained raters (examples shown in Figure 5.12). The large variability in the positions of each organ presents challenging conditions for registration and also may be ill-suited for using the EDT to provide relative distances for spatial context.

5.5.2   Implementation

For this application, three resolutions levels were used, at 4mm^3, 2mm^3 and 1mm^3 voxel sizes. For each test image, affine registration was used to align the atlases and the 50 nearest atlases were selected using sum of squared differences as the distance measure. Initial coarse segmentations were established by using coordinates normalised to the range $[0, 100]$ as spatial context with $\alpha{=}0.6$, whilst subsequent refinements using GDT as spatial context used $\alpha{=}7$. Similarly to both [191] and [41], graph cuts [31] was applied as post processing to obtain the final segmentation, using the same adaptations and parameters as proposed in [191].
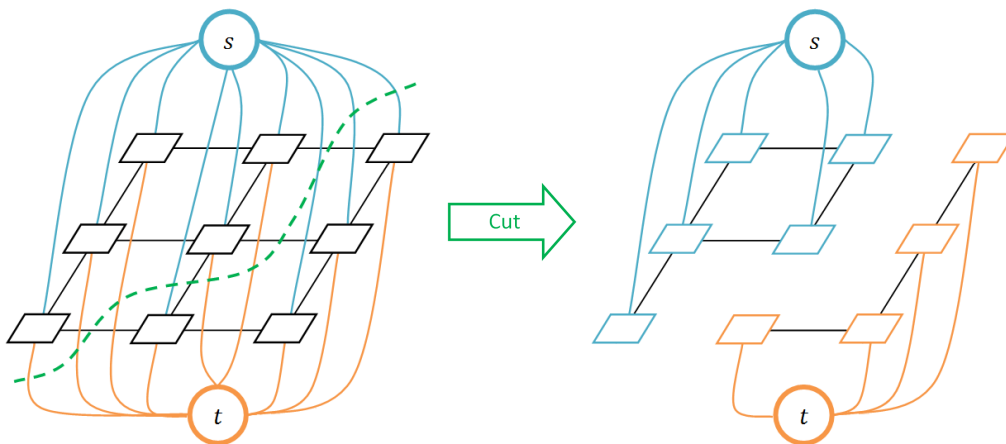
Figure 5.13: Graph cuts uses the max-flow min-cut principle: Treating an image as a graph, find a cut between terminal nodes s and t which partitions the graph into two sets with s and t in different sets such that only the minimum capacity edges are removed. The cut thus provides a foreground/background segmentation.

Graph cuts is a maximum-flow minimum-cut graph based optimisation which has been widely used for image segmentation [161], [174], [105]. It automatically estimates an intensity model for a structure of interest based on spatial prior in the image before finding the optimal cut in a graph representation, which segments the structure (foreground) from the background. An image can be considered a graph $\Gamma = (V, E)$ with vertices $v \in V$ representing pixels and

edges $e \in E$ representing the flow of between vertices and two terminal nodes, s and t. A segmentation is obtained by determining an s-t cut on Γ, which splits s and t into two different sets (see Figure 5.13). In graph cuts, the flow is maximised to solve an MRF-based energy function:

$$\mathcal{E}(l) = \lambda \sum_{\mathbf{x} \in I} D_{\mathbf{x}}(l_{\mathbf{x}}) + \sum_{\{\mathbf{p},\mathbf{q}\} \in \mathcal{N}_{\mathbf{x}}} \mathcal{V}_{\mathbf{p},\mathbf{q}}(l_{\mathbf{p}}, l_{\mathbf{q}}) \tag{5.4}$$

where $\mathcal{N}_{\mathbf{x}}$ is a neighbourhood around voxel $\mathbf{x}$ and $l_{\mathbf{x}}$ is the labelling of $\mathbf{x}$ in the unseen image in I. $D_{\mathbf{x}}(l_{\mathbf{x}})$ is a data term which measures the disagreement between a prior probabilistic model and the observed data, whilst $\mathcal{V}_{i,j}$ is a smoothness term, which penalises discontinuities in the grey value appearance. $D_{\mathbf{x}}(l_{\mathbf{x}})$ is defined as

$$D_{\mathbf{x}}(l_{\mathbf{x}}) = -\beta \ln P_A(\mathbf{y}, l_{\mathbf{y}}) - (1 - \beta) \ln P_i(\mathbf{x}, l_{\mathbf{x}}) \tag{5.5}$$

where $P_i(p, l_{\mathbf{x}}), i \in s, t$ are Gaussian intensity models estimated for foreground and background of structure l_i and $P_A(q, l_k)$ is spatial prior. In this implementation, the spatial prior P_A is obtained from the existing segmentation by setting a Gaussian neighbourhood around the each segmented structure with a fixed size G_σ. This can be performed simply by Gaussian blurring each segmented structure in turn with σ as the width of the Gaussian. The foreground intensity model P_x is estimated from all voxels which have at least 95% confidence for the associated structure in the spatial prior P_A based on the intensity distribution model, i.e. $P_s(p, l_i) = P(I(x)|l_i)$. The background distribution model P_t is described as a Gaussian distribution of background voxels estimated from a dilated region of voxels that have a foreground probability of more than 0 but with a maximum of 5%. The intensity distribution for this region then provides the probabilities, in a similar fashion as for P_s.

Table 5.3: Table of parameters for the abdominal CT dataset.

Method/Description	Parameter(s)
Registration Type	Affine - 12 parameters
Intensity Range	$[1800, 2400]$
Patch Size	$5{\times}5{\times}5$
k	40
α	0.6 (initialisation), 7
Boundary Size	3
Number of Resolutions	3
Number of Atlases	50, 30 in final resolution
Graph cuts - λ	1
Graph cuts - β	0.1
Graph cuts - c	0.5

Following [161] and [191], $\mathcal{V}_{i,j}$ is defined as:

$$\mathcal{V}_{p,q}(l_p, l_q) = c\left(1 + \ln\left(1 + \frac{1}{2}\left(\frac{|I(p) - I(q)|}{\sigma}\right)^2\right)\right)^{-1} + (1 - c)\left(1 - \max_{\mathbf{x} \in M_{p,q}} B(\mathbf{x})\right) \quad (5.6)$$

where $M_{p,q}$ is a line joining p and q, B is the intervening contour probabilistic map derived from the gradient image [114] and σ is the robust scale of the image I [161]. The intensity information and the contour probability information are balanced by the weight c.

The parameters used are summarised in Table 5.3.

5.5.3   Results

The results presented in Figure 5.15 and Table 5.4, demonstrate competitive accuracy compared to other state-of-the-art methods, also presented in Table 5.4. The results from [191] and [41] both required non-rigid registration, with the former performing separate registrations for each organ and the latter using multiple cubic divisions of each image and registering them in a hierarchical framework. In contrast, the proposed approach only used global affine registration. The dataset used in this experiment were also used for [191] and [41], although [41] only experimented with 100 of the 150 total images available.
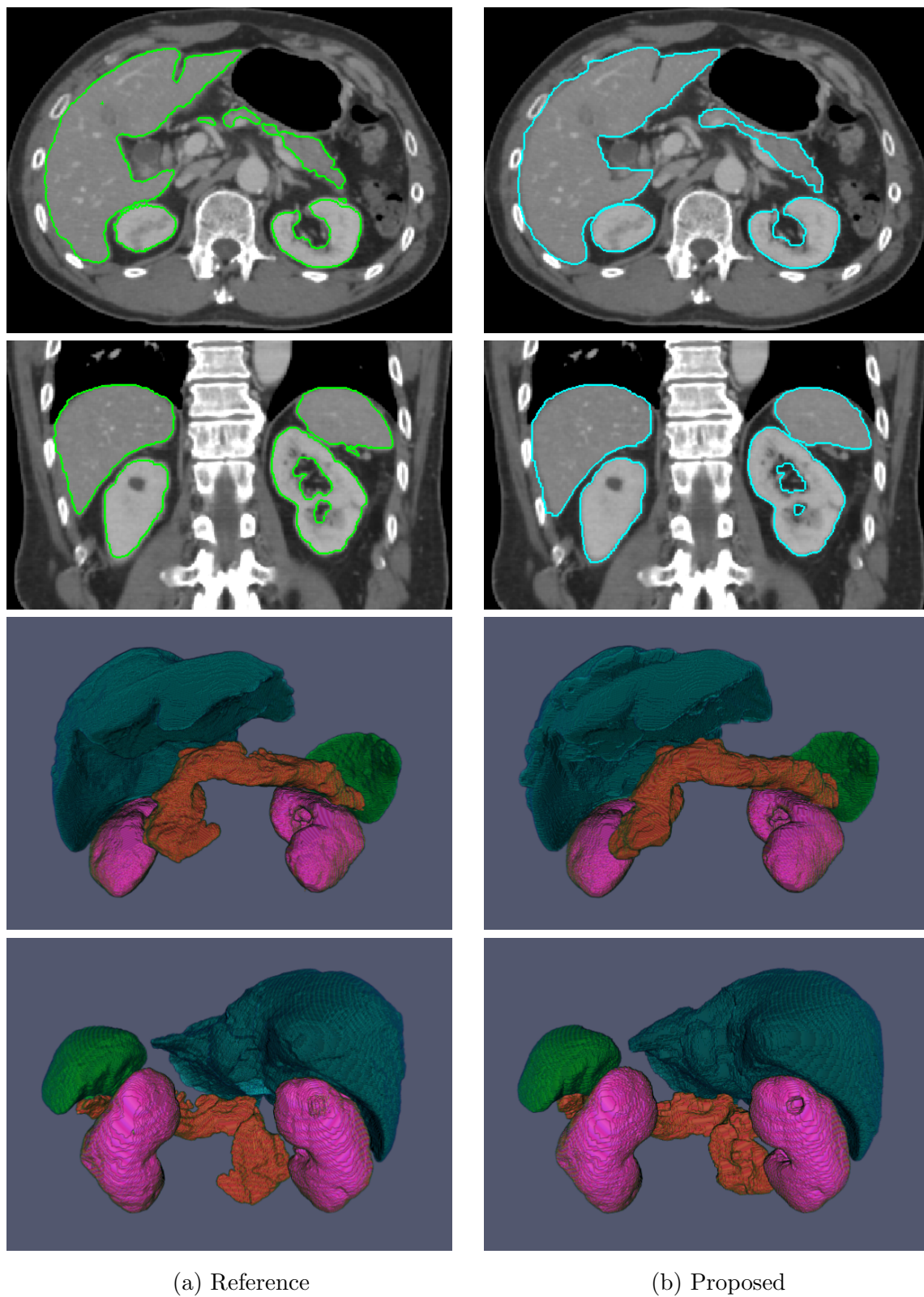
(a) Reference   (b) Proposed

Figure 5.14: Examples segmentations for the Liver, Spleen, Pancreas, Kidneys with Dice of 0.947, 0.955, 0.826, 0.924 respectively. Reference segmentations are outlined on the left. Volume renderings shown on bottom from two different perspectives.

Table 5.4: Overall Dice coefficients in comparison to other state-of-the-art methods shown as mean±standard dev. with (median) and [worst, best] where available. Best values are shown in red.

	Proposed (150 images)	[191] (150 images)	[41] (100 images)
Liver	0.947 (0.957) ±0.036 [0.654, 0.979]	0.940 ±0.028 [0.814, 0.974]	0.951 ±0.010
Spleen	0.934 (0.958) ±0.077 [0.425, 0.983]	0.920 ±0.092 [0.264, 0.982]	0.914 ±0.057
Pancreas	0.673 (0.731) ±0.189 [0.031, 0.907]	0.696 ±0.167 [0.069, 0.909]	0.691 ±0.153
Kidneys	0.933 (0.956) ±0.083 [0.344, 0.980]	0.925 ±0.072 [0.515, 0.982]	0.901 ±0.050
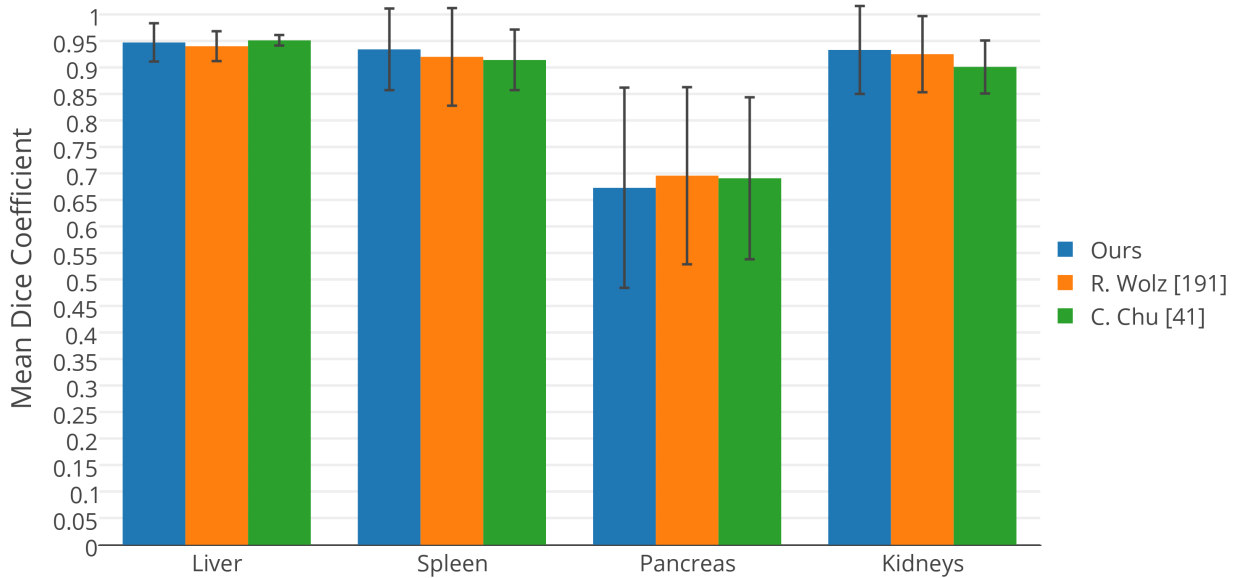
Figure 5.15: Comparison of mean Dice Coefficients with other state-of-the-art methods.

Effect of Graph Cuts Post-Processing

A comparison of the results before and after performing graph cuts [31] as a post-process refinement are outlined below in Table 5.5 and Figure 5.16. Use of graph cuts provides an improvement in segmentation accuracy for almost every subject across all organs, although the level of improvement differs from organ-to-organ. When Welch's paired two sample t-test is applied, the p-values[1] obtained are $<< 0.0001$ for all organs.

Table 5.5: Comparison of segmentation accuracy before and after graph cuts post processing. Results showing Dice coefficients mean±standard dev. with (median) and [worst, best].

	Before Graph Cuts	After Graph Cuts
Liver	0.940 (0.951) ±0.038 [0.627, 0.976]	0.947 (0.957) ±0.036 [0.654, 0.979]
Spleen	0.918 (0.946) ±0.080 [0.425, 0.971]	0.934 (0.958) ±0.077 [0.425, 0.983]
Pancreas	0.651 (0.700) ±0.180 [0.030, 0.875]	0.673 (0.731) ±0.189 [0.031, 0.907]
Kidneys	0.922 (0.946) ±0.083 [0.351, 0.968]	0.933 (0.956) ±0.083 [0.344, 0.980]

---

[1]Exact p-values: 7.2×10^{-31}, 1.0×10^{-29}, 8.1×10^{-32}, 1.4×10^{-41} for the liver, spleen, pancreas and kidneys
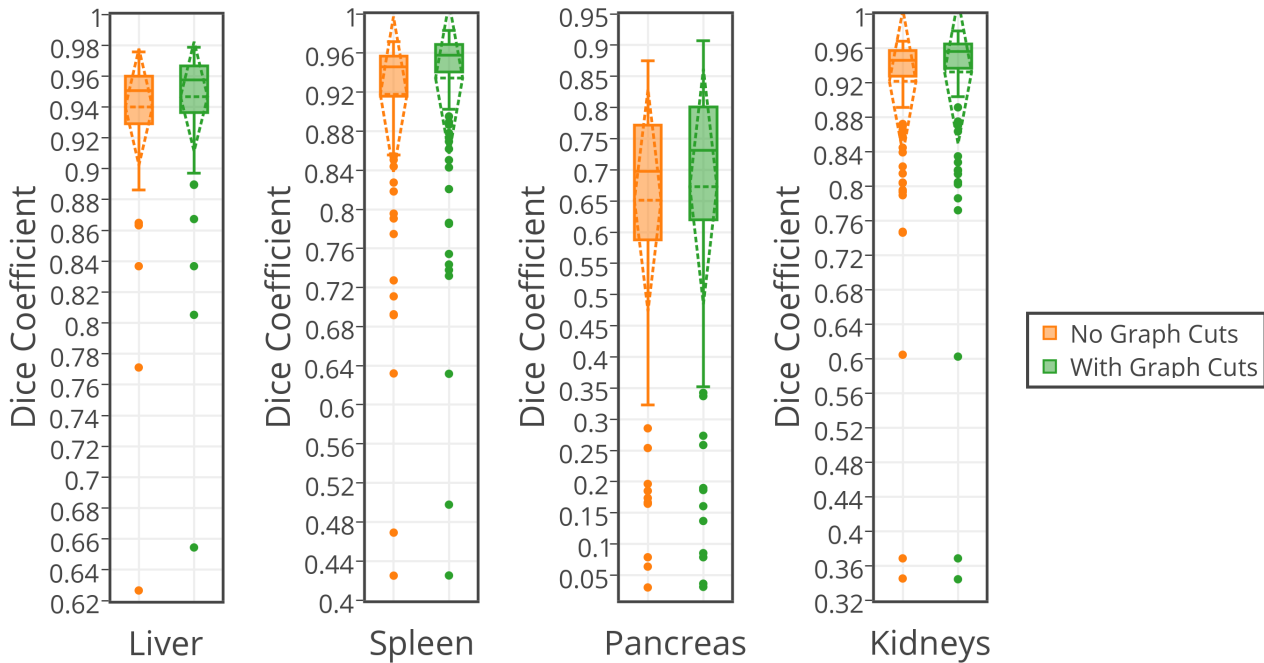
Figure 5.16: Segmentation accuracy comparison across all labels, with and without graph cuts post processing. Solid line indicates the median, the dashed line indicates the mean and standard deviation is shown by the dashed diamond.

5.5.4   Discussion

The proposed approach was able to achieve results competitive with the current state-of-the-art whilst only relying on a global affine registration to align images, contrary to the approaches used in [191] and [41], which also used non-rigid deformable registrations at various partitions of the image space as well. This demonstrates the robustness of using geodesic distances for spatial context and the ability to be less reliant on the registration outcome.

The pancreas remains one of the most challenging organs for automated segmentation, as there is a particularly large variance in both the shape and position of this organ between subjects. Furthermore, the organ's size is relatively small in comparison to the other organs and can be harder to locate in the large volume of the abdominal region. This can lead to incorrect initialisations which do not get corrected during the refinement.

For all organs, there are several outlying cases where the proposed approach did not perform so well (see Figure 5.16), thus reducing the mean accuracy significantly below that of the median. The predominant cause for this may be due to the different diseases and disease stages of the

subjects, which often affects the intensity outputs in the CT images. Whilst the majority, 141 out of 150, had gastric cancer, the remainder had different types of cancer. Furthermore, three of the subjects had only one kidney, which differed in the location within the body and had a significant impact on the spatial configuration of the organs. The outlying cases were not well represented within the atlas library, thus finding similar examples for patch-based label propagation often lead to incorrect labelling in these cases. This highlights one of the weaknesses of using a kNN approach for segmentation.

Overall, the results still demonstrates great potential in the proposed approach. Despite, the shortcomings in a small minority of cases, the results on average provide improved accuracy for some organs such as the spleen and kidneys whilst maintaining a comparable level for the liver and pancreas.

5.6   Summary and Conclusion

This chapter further explored options for patch-based segmentation with spatial context and proposed the use of the geodesic distance transform to measure relative distances between structures in providing an anatomically-adaptive coordinate system. This approach was evaluated on cardiac MRI and abdominal CT scans which are two very different datasets, yet demonstrating good results with both. The approach compares favourably with other approaches that have been proposed for spatial context and the GDT can be fine tuned to provide behaviour that is similar to the EDT if required. In both applications, the proposed approach only uses global affine registration, whilst many existing methods require the use of non-rigid deformable registration, sometimes on an multiple hierarchical levels as well as requiring initial global alignment. The results demonstrate robustness in the proposed approach and the ability to provide accurate results when images are not well aligned. Overall this approach shows much potential for automated image segmentation, particularly in more challenging datasets where achieving accurate registration is difficult.

Chapter 6

Conclusions

6.1 Summary of Achievements

During the course of this thesis, a general segmentation framework was developed which uses patch-based segmentation with spatial context. The framework reformulates the patch-based segmentation approach in [47] and [142] as a global kNN approach which incorporates spatial context and removes the requirement for a limited search window for patch comparison. In addition, three methods of providing spatial context for patches within the proposed framework were developed and evaluated, which are listed below in descending order of dependence on registration accuracy:

1. Voxel coordinates within the image space.

2. Euclidean distances between structures or reference points.

3. Geodesic distances within the image, between structures or reference points.

These approaches have different suitability for different types of applications, each with their own strengths and weaknesses. A more detailed comparison of these approaches are given in later section 6.2.

6.1.1   Applications

The proposed framework and methods have been applied to a large variety of image datasets, which are summarised below:

- **Brain MRI - hippocampus.** 202 images from the ADNI dataset were used to evaluate the initially proposed kNN patch-based framework using a leave-one-out cross-validation strategy. The image coordinates of each patch were used to provide spatial context, demonstrating superior performance to an established patch-based method [47]. Mean Dice scores of 0.863 were achieved for the hippocampus using only affine registration to align the images to the MNI-152 template space.

- **Knee MRI - tibia and femur (bone and cartilage.** The SKI10 Grand Challenge provided a challenging but relevant dataset to evaluate the newly proposed multi-scale patch-based framework. Here, 100 images were provided for training and 50 additional unseen images were used for testing. An adaptive coordinate system using relative distances between structures was used as to provide spatial context. The proposed framework was applied without any registration, demonstrating plausibility in such an approach. The results ranked 3rd for cartilage, 8th for bone and 7th overall, out of a total of 16 entries.

- **Canine Leg MRI - 7 proximal pelvic limb muscles.** The SATA segmentation challenge provided another application to continue the investigation on segmentation without registration. Here, the canine leg dataset was used, which provided 22 images for training and 23 unseen images for testing. The same method was used as for the SKI10 knee MRI dataset, and obtained results which outperformed standard majority voting and spatial STAPLE methods which used non-rigid deformable registration.

- **Cardiac MRI - myocardium, left and right ventricles.** 20 cardiac MR images were used to evaluate a new spatial context approach based on geodesic distances. A leave-one-out cross-validation strategy was used, and all previously proposed spatial context methods were also evaluated. The results demonstrates superior performance to [47] when only affine registration is used. Mean Dice scores of 0.934, 0.797, 0.901 were achieved for

the left ventricle, myocardium and right ventricle when using geodesic distances to provide spatial context.

- **Abdominal CT - liver, spleen, pancreas and kidneys.** Further to the cardiac dataset, the proposed approach was also applied to 150 abdominal CT images and evaluated using a leave-one-out cross-validation strategy. It obtained results which are competitive to current state-of-the-art methods such as [191] and [41], achieving mean Dice scores of 0.947, 0.934, 0.673, 0.933 for the liver, spleen, pancreas and kidneys. Furthermore, the proposed approach only used a global affine registration whilst the other methods also apply additional non-rigid deformable registration at various different hierarchical levels.

6.2 Comparison of Spatial Context Approaches

The proposed approaches for spatial context each have different suitability for different applications. Table 6.1 below summarises their strengths and weaknesses relative to each other:

Table 6.1: Summary comparison of proposed spatial context approaches.

	Voxel Coordinates	EDT-based	GDT-based
Registration dependence	High	Low	Low
Requires initial segmentation or reference structures	No	Yes	Yes
Suitability to single structure segmentation	High	Low	Low
Suitability to multi-structure segmentation	Low	Medium	High
Suitability for high anatomical variability	Low	Medium	High
Aware of image boundaries	No	No	Yes
Affected by image noise	No	No	Yes

Use of voxel coordinates as spatial context works best in segmenting single structures and where the anatomical variability is low. It is not as effective in multi-structure segmentation as there is generally more anatomical variability in those cases. However, it can be used to provide a coarse initial segmentation for the other approaches for spatial context.

Use of relative distances require an initial segmentation or reference in order to work. They're

most effective when segmenting multiple structures where more spatial information can be provided. Using Euclidean distances or geodesic distances with a low value for γ works best in cases where there is more image noise and where image contrast is low - such as certain MR applications. However, using the Euclidean distance is not as effective where there are large variances in the relative positions of the anatomical structures as well as the scale of the structures. The use of geodesic distances, with a moderate to high value for γ, is more effective in this situation. However geodesic distances do not work as well in noisy images or where image artifacts commonly occur.

6.3   Limitations and Future Work

There are several key areas which limit the proposed framework, in both the general applicability and the adoption by research scientists and clinicians or as part of a clinical decision support system. These areas can be interesting avenues to explore in extending or adopting the methods presented in this thesis for future work.

6.3.1   Speed

One particularly unfavourable aspect of the proposed framework is the speed and computational requirement for each segmentation. This is primarily due to performing global kNN search for the patches in each atlas and is particularly a problem for larger images since each voxel is labelled individually. This imposes a restriction on both the number of atlases and the dimensional size of the images that can be used. The two are related, and it may be for larger images, fewer atlases can be used, as the computational time requirement could be halved by simply halving the number of atlases used. However, using fewer atlases reduces the anatomical variability that can be expressed through the atlases and thus the robustness of the segmentation performance.

Implementation

Without altering the algorithm of the proposed methods, one area that would yield improvements in speed is in the implementation, particularly in exploiting parallelism. Although the independent labelling for each voxel enables parallelism to be exploited, this has so far been with CPU-based architecture, which is limited and more costly for massively parallel tasks. Using a GPU-based implementation for kNN search could provide a significant boost in performance, and recent works such as [69] and [16] demonstrate the possibility for such an approach. However, the memory limitation of the GPU can still present a significant engineering challenge to overcome.

Alternative Data Structures

In the software implementations, the kNN datastructure used was a ball tree from scikit-learn, an open-source module for Python. This is a generalised kNN data structure but may not be the most optimal, in terms of speed and memory usage, for patch search and comparison. The proposed segmentation framework is suitably modular such that the ball tree could be replaced with alternatives, and a wide range of these were reviewed in chapter 2. Data structures that provide approximate nearest neighbours could also be used instead of exact kNN. Both the PatchMatch algorithm and the neighbourhood approximation forests (NAFs) are potential candidates that could be integrated into the proposed framework, although the training times for NAFs would be prohibitive for on-the-fly spatial context extraction and would limit the flexibility of the framework.

One aspect of performing kNN search that is not often taken into account in many data structures is the similarity of the queries. In general, applications of kNN search rarely only perform a singular kNN query, but performs them more like a batch process which are repeated many times. This has been termed as *general N-body problems* by Alexander Gray et al [73]. It would be plausible to exploit the fact that similar queries should return similar results, and this is proposed by works such as [73] and [137] which propose the use of dual-trees, where a tree

is constructed for the query items as well the data library. This is related to the all nearest neighbour problem which has been studied in computational geometry [43], [173].

Alternative Classification Frameworks

A more drastic change to improve the speed would be to alter the classification framework from relying on using the kNN at run time, to one which does not use such a dense and potentially redundant representation of the training data. The approach of using random forests in [196] provides a potential candidate for which the spatial context methods developed in this thesis could be easily adopted. Another alternative could be using random ferns [126] and the extremely random variant [71], which could allow similar on-the-fly functionality for incorporating spatial context as the proposed framework. Additionally, dictionary learning could be applied to learn the range of patches and how they can be expressed in order to reduce the redundancy between them. A recent approach [170] uses such a principle for segmentation of the hippocampus, and could provide a template for future work.

6.3.2   Atlas Selection

One aspect of multi-atlas segmentation approaches which could benefit from further investigation is the atlas selection process, particularly for large datasets and where it is desirable to select atlases without registration first. This was briefly visited in chapter 4, where an approach using histogram of 3D oriented gradients was proposed, but was not explored to substantial depth. Methods from content-based image retrieval could be used in alternative atlas selection schemes, but this has received relatively little attention in the medical imaging field.

Additionally, traditional atlas selection strategies of selecting the most similar atlases for label propagation should be reconsidered. Although selecting the most similar atlases would reduce the errors, it does not guarantee the atlases can account for the anatomical differences between them. This may have been the most optimal approach for multi-atlas approaches which are reliant on registration accuracy, but it may not hold for patch-based frameworks where there is

less dependency on voxel-wise image correspondence. Perhaps an alternative strategy should be to select a set of nearby atlases which are diverse enough to account for a wide range of local differences instead.

Furthermore, there is a trade-off in speed and accuracy with regards to the number of atlases to use. An increase in the number of atlases used yields logarithmic returns in the improvement to segmentation performance, as both the experimental results in chapter 3 and the results from [47] suggest. An atlas selection strategy could also be optimised to produce the maximum local diversity with the minimum number of atlases, which would be highly beneficial to the computational performance.

6.3.3   Patch and Spatial Comparison

One general limitation of patch-based segmentation methods, particularly for MRI application, is how patches are compared. The use of intensity based metrics requires that images are normalised in order for their differences to be meaningful. Intensity normalisation can sometimes be challenging with MR images, especially if they are acquired from different scanners. Most normalisation methods use approaches based on histogram analysis, but this is then dependent on the anatomy captured within the images. For datasets with large anatomical variability, this can limit the effectiveness of intensity normalisation, particularly if there is a subtle difference in diseased and healthy tissues of the same organ. Better intensity normalisation methods could address this weakness, but this area is currently lacking further development. Alternative patch comparison metrics could be used, such as mutual information or cross correlation, but this may not be as efficient for kNN search and the current label fusion process based on non-local means would have to be altered.

The use of spatial context augments the intensity information and has been shown to beneficial, particularly when there is ambiguity when comparing patches, however balancing the two components requires adjustment of the spatial weighting α. This is due to the use of distances for the non-local means approach to label fusion which evaluates the two components together. The parameter α is currently chosen manually and requires some level of trial and error in

order to select a reasonable value. Ideally, this could be either selected automatically, allowing for α to vary spatially, or the label fusion could be altered to take into account the intensity and spatial information differently. This presents a possible extension to the current framework which could utilise other recent developments in machine learning.

6.3.4   Segmentation Initialisation

A major limitation of using the approach for spatial context proposed in chapters 4 and 5, based on relative distances to known structures or reference points, is correctly identifying them at initialisation. So far, the proposed frameworks have used coarse low resolutions obtained using the approach from chapter 3 with weak spatial weighting α. This coarse initial segmentation is then eroded as false negatives in the labelling would be detrimental to the accuracy of the relative distances. However, this process is limited by the size of the organ in the initial resolution. Performing this initialisation in higher resolutions is more computationally costly and more dependent on registration accuracy.

To overcome this limitation, a classifier could be trained to locate each organ within the image in the original resolution. A bounding box for these organs could be enough to provide relative distances and would also reduce the search space for patch selection and comparison. Random forest based classifiers, which have recently been proposed for organ location [129], could be adopted as a potential solution.

6.3.5   Summary

The work presented in this thesis demonstrates the potential for patch-based segmentation methods with spatial context to be a robust and versatile approach for image segmentation. The various approaches to spatial context reduces dependence on the registration outcome, enabling multi-atlas label propagation in applications where obtaining accurate image alignment is challenging. The methods proposed here could also be adapted to segmentation problems we have not explored, such as other organs as well as those in different image modalities.

It is also plausible to apply the methods for tumour segmentation, where registration is also challenging [180]. Although there are some limitations to the proposed segmentation framework and methods, they present interesting areas for future development.

Chapter 7

List of Publications

- Zehan Wang, Robin Wolz, Tong Tong, Daniel Rueckert. Spatially Aware Patch-based Segmentation (SAPS): An Alternative Patch-based Segmentation Framework. *Second International MICCAI Workshop on Medical Computer Vision: Recognition Techniques and Applications in Medical Imaging (MCV 2012)*. LNCS Volume 7766, pages 93-103. Springer Heidelberg 2013.

- Zehan Wang, Claire Donoghue, Daniel Rueckert. Patch-based Segmentation without Registration: Application to Knee MRI. *Fourth International Workshop on Machine Learning in Medical Imaging (MLMI 2013)*. LNCS Volume 8184, pages 98-105. Springer Heidelberg 2013.

- Zehan Wang, Anil Rao, Daniel Rueckert. Patch-based Segmentation without Registration: Application to Canine Leg MRI. *MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA)*. 2013.

- Zehan Wang, Kanwal K. Bhatia, Ben Glocker, Antonio de Marvao, Tim Dawes, Kazunari Misawa, Kensaku Mori, Daniel Rueckert. Geodesic Patch-based Segmentation *Medical Image Computing and Computer-Assisted Intervention – MICCAI*. LNCS Volume 8673, pages 666-673. Springer International 2014.

Bibliography

[1] ACHARYA, R., WASSERMAN, R., STEVENS, J., AND HINOJOSA, C.: Biomedical imaging modalities: a tutorial. *Computerized Medical Imaging and Graphics* 19(1), 3–25, Jan. 1995.

[2] AGGARWAL, C., HINNEBURG, A., AND KEIM, D.: On the Surprising Behavior of Distance Metrics in High Dimensional Space. In *Database Theory ICDT 2001*, J. Van den Bussche and V. Vianu, Eds., vol. 1973 of *LNCS*, pp. 420–434. Springer Berlin Heidelberg, 2001.

[3] AKGÜL, C. B., RUBIN, D. L., NAPEL, S., BEAULIEU, C. F., GREENSPAN, H., AND ACAR, B.: Content-Based Image Retrieval in Radiology: Current Status and Future Directions. *Journal of Digital Imaging* 24(2), 208–22, Apr. 2011.

[4] ALJABAR, P., HECKEMANN, R. A., HAMMERS, A., HAJNAL, J. V., AND RUECKERT, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46(3), 726–738, July 2009.

[5] ANDERSON, C. H., BERGEN, J. R., BURT, P. J., AND OGDEN, J. M.: Pyramid Methods in Image Processing, 1984.

[6] ANDONI, A., AND INDYK, P.: Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions. In *IEEE Symposium on Foundations of Computer Science*. 459–468. Oct. 2006.

[7] ARTAECHEVARRIA, X., MUNOZ-BARRUTIA, A., AND ORTIZ-DE SOLORZANO, C.:
Combination strategies in multi-atlas image segmentation: application to brain MR data.
IEEE Transactions on Medical Imaging 28(8), 1266–77, Aug. 2009.

[8] ASMAN, A., AND LANDMAN, B.: Non-local STAPLE: An Intensity-Driven Multi-atlas
Rater Model. In *Medical Image Computing and Computer-Assisted Intervention – MIC-
CAI*, pp. 426–434. Springer Berlin Heidelberg, 2012.

[9] BAI, W., SHI, W., O'REGAN, D., TONG, T., WANG, H., JAMIL-COPLEY, S., PE-
TERS, N., AND RUECKERT, D.: A Probabilistic Patch-Based Label Fusion Model for
Multi-Atlas Segmentation with Registration Refinement: Application to Cardiac MR
Images. *IEEE Transactions on Medical Imaging* 32(7), 1302–1315, July 2013.

[10] BAJCSY, R., AND KOVAČIČ, S.: Multiresolution elastic matching. *Computer Vision,
Graphics, and Image Processing* 46(1), 1–21, Apr. 1989.

[11] BAJCSY, R., LIEBERSON, R., AND REIVICH, M.: A computerized system for the elastic
matching of deformed radiographic images to idealized atlas images. *Journal of computer
assisted tomography* 7(4), 618–625, 1983.

[12] BALAFAR, M. A., RAMLI, A. R., SARIPAN, M. I., AND MASHOHOR, S.: Review of
brain MRI image segmentation methods. *Artificial Intelligence Review* 33(3), 261–274,
Jan. 2010.

[13] BARNES, C., SHECHTMAN, E., FINKELSTEIN, A., AND GOLDMAN, D. B.: Patch-
Match: a randomized correspondence algorithm for structural image editing. *ACM Trans-
actions on Graphics* 28(3), 2, July 2009.

[14] BARNES, C., SHECHTMAN, E., GOLDMAN, D., AND FINKELSTEIN, A.: The General-
ized PatchMatch Correspondence Algorithm. In *Computer Vision ECCV 2010*, K. Dani-
ilidis, P. Maragos, and N. Paragios, Eds., vol. 6313 of *LNCS*, pp. 29–43. Springer Berlin
Heidelberg, 2010.

[15] BARNES, J., FOSTER, J., BOYES, R. G., PEPPLE, T., MOORE, E. K., SCHOTT,
J. M., FROST, C., SCAHILL, R. I., AND FOX, N. C.: A comparison of methods for the

automated calculation of volumes and atrophy rates in the hippocampus. *NeuroImage* 40(4), 1655–71, May 2008.

[16] BARRIENTOS, R., GÓMEZ, J., AND TENLLADO, C.: kNN Query Processing in Metric Spaces Using GPUs. In *Euro-Par 2011 Parallel Processing*, pp. 380–392. Springer Berlin Heidelberg, 2011.

[17] BAZIN, P.-L., AND PHAM, D. L.: Homeomorphic brain image segmentation with topological and statistical atlases. *Medical Image Analysis* 12(5), 616–25, Oct. 2008.

[18] BECKMANN, N., KRIEGEL, H.-P., SCHNEIDER, R., AND SEEGER, B.: The R*-tree: an efficient and robust access method for points and rectangles. In *ACM SIGMOD*. 322–331. ACM Press, New York, New York, USA, 1990.

[19] BELAROUSSI, B., MILLES, J., CARME, S., ZHU, Y. M., AND BENOIT-CATTIN, H.: Intensity non-uniformity correction in MRI: existing methods and their validation. *Medical Image Analysis* 10(2), 234–46, Apr. 2006.

[20] BELKIN, M., AND NIYOGI, P.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. *Science* 1, 585–591, 2002.

[21] BELKIN, M., AND NIYOGI, P.: Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation* 15(6), 1373–1396, June 2003.

[22] BELLMAN, R. E., AND CORPORATION, R.: *Dynamic programming.* Princeton University Press, 1957.

[23] BENTLEY, J. L.: Multidimensional binary search trees used for associative searching. *Communications of the ACM* 18(9), 509–516, 1975.

[24] BENTLEY, J. L.: K -d trees for semidynamic point sets. *Proceedings of the sixth annual symposium on Computational geometry SCG 90* 187–197, 1990.

[25] BERCHTOLD, S., KEIM, D., AND KRIEGEL, H.: The X-tree: An index structure for high-dimensional data. In *Readings in Multimedia Computing and Networking.* Morgan Kaufmann, 2001.

[26] BEYER, T., TOWNSEND, D. W., BRUN, T., KINAHAN, P. E., CHARRON, M., RODDY, R., JERIN, J., YOUNG, J., BYARS, L., AND NUTT, R.: A Combined PET/CT Scanner for Clinical Oncology. *Journal of Nuclear Medicine* 41(8), 1369–79, Aug. 2000.

[27] BEYGELZIMER, A., KAKADE, S., AND LANGFORD, J.: Cover trees for nearest neighbor. *Proceedings of the 23rd international conference on Machine learning ICML 06* 1, 97–104, 2006.

[28] BOCCARDI, M., GANZOLA, R., BOCCHETTA, M., PIEVANI, M., REDOLFI, A., BARTZOKIS, G., CAMICIOLI, R., CSERNANSKY, J. G., DE LEON, M. J., DETOLEDO-MORRELL, L., KILLIANY, R. J., LEHÉRICY, S., PANTEL, J., PRUESSNER, J. C., SOININEN, H., WATSON, C., DUCHESNE, S., JACK, C. R., AND FRISONI, G. B.: Survey of protocols for the manual segmentation of the hippocampus: preparatory steps towards a joint EADC-ADNI harmonized protocol. *Journal of Alzheimer's disease : JAD* 26 Suppl 3(0 3), 61–75, Jan. 2011.

[29] BOIMAN, O., SHECHTMAN, E., AND IRANI, M.: In defense of Nearest-Neighbor based image classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–8. June 2008.

[30] BOOKSTEIN, F.: Thin-plate splines and the atlas problem for biomedical images. *Information Processing in Medical Imaging* 511, 326–342, 1991.

[31] BOYKOV, Y., VEKSLER, O., AND ZABIH, R.: Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(11), 1222–1239, 2001.

[32] BROWN, L. G.: A survey of image registration techniques. *ACM Computing Surveys* 24(4), 325–376, Dec. 1992.

[33] BUADES, A., COLL, B., AND MOREL, J.-M.: A non-local algorithm for image denoising. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 60–65. 2005.

[34] BUADES, a., COLL, B., AND MOREL, J. M.: A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation* 4(2), 490–530, Jan. 2005.

[35] BUADES, A., COLL, B., AND MOREL, J. M.: Image Denoising Methods. A New Non-local Principle. *SIAM Review* 52(1), 113–147, Jan. 2010.

[36] CABEZAS, M., OLIVER, A., LLADÓ, X., FREIXENET, J., AND CUADRA, M. B.: A review of atlas-based segmentation for magnetic resonance brain images. *Computer methods and programs in biomedicine* 104(3), e158–77, Dec. 2011.

[37] CAUDRON, J., FARES, J., BAUER, F., AND DACHER, J.-N.: Evaluation of left ventricular diastolic function with cardiac MR imaging. *Radiographics : a review publication of the Radiological Society of North America, Inc* 31(1), 239–59, 2011.

[38] CHAMBOLLE, A.: An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision* 20(1), 89–97, 2004.

[39] CHAN, T. F., OSHER, S., AND SHEN, J.: The digital TV filter and nonlinear denoising. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* 10(2), 231–41, Jan. 2001.

[40] CHEN, S., MA, B., AND ZHANG, K.: On the similarity metric and the distance metric. *Theoretical Computer Science* 410(24-25), 2365–2376, May 2009.

[41] CHU, C., ODA, M., KITASAKA, T., MISAWA, K., FUJIWARA, M., HAYASHI, Y., NIMURA, Y., RUECKERT, D., AND MORI, K.: Multi-organ Segmentation Based on Spatially-Divided Probabilistic Atlas from 3D Abdominal CT Images. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., vol. 8150 of *LNCS*, pp. 165–172. Springer Berlin Heidelberg, 2013.

[42] CIACCIA, P., PATELLA, M., AND ZEZULA, P.: M-tree: An Efficient Access Method for Similarity Search in Metric Spaces. *Compute* 25, 426–435, 1997.

[43] CLARKSON, K. L.: Fast Algorithms for the All Nearest Neighbors Problem. In *24th Annual Symposium on Foundations of Computer Science (sfcs 1983)*. 226–232. Nov. 1983.

[44] COLLEWET, G., STRZELECKI, M., AND MARIETTE, F.: Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magnetic resonance imaging* 22(1), 81–91, Jan. 2004.

[45] COLLINS, D. L., AND EVANS, A. C.: Animal: validation and applications of nonlinear registration-based segmentation. *International Journal of Pattern Recognition and Artificial Intelligence* 11(08), 1271–1294, 1997.

[46] CONAGHAN, P. G., PORCHERET, M., KINGSBURY, S. R., GAMMON, A., SONI, A., HURLEY, M., RAYMAN, M. P., BARLOW, J., HULL, R. G., CUMMING, J., LLEWE-LYN, K., MOSCOGIURI, F., LYONS, J., AND BIRRELL, F.: Impact and therapy of osteoarthritis: the Arthritis Care OA Nation 2012 survey. *Clinical rheumatology* June 2014.

[47] COUPÉ, P., MANJÓN, J. V., FONOV, V., PRUESSNER, J., ROBLES, M., AND COLLINS, D. L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *NeuroImage* 54(2), 940–954, 2011.

[48] COUPE, P., YGER, P., PRIMA, S., HELLIER, P., KERVRANN, C., AND BARILLOT, C.: An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. *IEEE Transactions on Medical Imaging* 27(4), 425–41, Apr. 2008.

[49] CRIMINISI, A., SHARP, T., AND BLAKE, A.: GeoS : Geodesic Image Segmentation. In *ECCV*. vol. 5302, 99–112. 2008.

[50] CRUM, W. R., TANNER, C., AND HAWKES, D. J.: Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. *Physics in medicine and biology* 50(21), 5153–74, Nov. 2005.

[51] DABOV, K., FOI, A., AND KATKOVNIK, V.: Image denoising with block-matching and 3D filtering. In *SPIE*. vol. 6064, 1–12. 2006.

[52] DALAL, N., TRIGGS, B., AND TRIGGS, W.: Histograms of Oriented Gradients for Human Detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. vol. 1, 886–893. June 2005.

[53] DE MARVAO, A., DAWES, T. J. W., SHI, W., MINAS, C., KEENAN, N. G., DIAMOND, T., DURIGHEL, G., MONTANA, G., RUECKERT, D., COOK, S. A., AND O'REGAN, D. P.: Population-based studies of myocardial hypertrophy: high resolution cardiovascular magnetic resonance atlases improve statistical power. *Journal of cardiovascular magnetic resonance* 16(1), 16, Jan. 2014.

[54] DELSO, G., FÜRST, S., JAKOBY, B., LADEBECK, R., GANTER, C., NEKOLLA, S. G., SCHWAIGER, M., AND ZIEGLER, S. I.: Performance Measurements of the Siemens mMR Integrated Whole-Body PET/MR Scanner. *Journal of Nuclear Medicine* 52(12), 1914–22, Dec. 2011.

[55] DEMPSTER, A., LAIRD, N., AND RUBIN, D.: Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal statistical . . .* 39(1), 1–38, 1977.

[56] DHAESE, P.-F., DUAY, V., MERCHANT, T. E., MACQ, B., AND DAWANT, B. M.: Atlas-Based Segmentation of the Brain for 3-Dimensional Treatment Planning in Children with Infratentorial Ependymoma. *Medical Image Computing and Computer-Assisted Intervention – MICCAI* 2879, 627–634, 2003.

[57] DONOGHUE, C. R., RAO, A., PIZARRO, L., BULL, A. M. J., AND RUECKERT, D.: Fast and Accurate Global Geodesic Registrations using Knee MRI from the Osteoarthritis Initiative. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 50–57. June 2012.

[58] DRORI, I., COHEN-OR, D., AND YESHURUN, H.: Fragment-based image completion. *ACM Transactions on Graphics* 22(3), 303, July 2003.

[59] EDELSTEIN, W. A.: A signal-to-noise calibration procedure for NMR imaging systems. *Medical Physics* 11(2), 180, 1984.

[60] EGOROV, V., AND SARVAZYAN, A. P.: Mechanical imaging of the breast. *IEEE Transactions on Medical Imaging* 27(9), 1275–87, Sept. 2008.

[61] ESKILDSEN, S. F., COUPÉ, P., FONOV, V., MANJÓN, J. V., LEUNG, K. K., GUIZARD, N., WASSEF, S. N., Ø STERGAARD, L. R., AND COLLINS, D. L.: BEaST:

brain extraction based on nonlocal segmentation technique. *NeuroImage* 59(3), 2362–73, Feb. 2012.

[62] Evans, A. C., Kamber, M., Collins, D. L., and MacDonald, D.: An MRI-Based Probabilistic Atlas of Neuroanatomy. In *Magnetic Resonance Scanning and Epilepsy*, S. D. Shorvon, D. R. Fish, F. Andermann, G. M. Bydder, and H. Stefan, Eds., vol. 264 of *NATO ASI Series*, pp. 263–274. Springer US, 1994.

[63] Feldkamp, L. A., Davis, L. C., and Kress, J. W.: Practical cone-beam algorithm. *Journal of the Optical Society of America A* 1(6), 612, June 1984.

[64] Figueiredo, M., and Jain, A.: Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(3), 381–396, Mar. 2002.

[65] Finkel, R. a., and Bentley, J. L.: Quad trees a data structure for retrieval on composite keys. *Acta Informatica* 4(1), 1–9, 1974.

[66] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., and Dale, A. M.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33(3), 341–55, Jan. 2002.

[67] Folkesson, J., Dam, E., Olsen, O. F., Pettersen, P., and Christiansen, C.: Automatic segmentation of the articular cartilage in knee MRI using a hierarchical multi-class classification scheme. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, J. Duncan and G. Gerig, Eds., vol. 3749 of *LNCS*, pp. 327–34. Springer Berlin Heidelberg, Jan. 2005.

[68] Frangi, a. F., Niessen, W. J., and Viergever, M. a.: Three-dimensional modeling for functional analysis of cardiac images: a review. *IEEE Transactions on Medical Imaging* 20(1), 2–25, Jan. 2001.

[69] GARCIA, V., DEBREUVE, E., AND BARLAUD, M.: Fast k nearest neighbor search using GPU. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–6. June 2008.

[70] GIONIS, A., INDYK, P., AND MOTWANI, R.: Similarity search in high dimensions via hashing. In *Proceedings of the International Conference on Very Large Data Bases*. 518–529. 1999.

[71] GLOCKER, B., IZADI, S., SHOTTON, J., AND CRIMINISI, A.: Real-time RGB-D camera relocalization. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 173–179. Oct. 2013.

[72] GOLDSTEIN, T., AND OSHER, S.: The Split Bregman Method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences* 2(2), 323–343, Jan. 2009.

[73] GRAY, A., AND MOORE, A.: 'N-Body' Problems in Statistical Learning. In *Advances in Neural Information Processing Systems (NIPS)*. 2001.

[74] GREIG, D., PORTEOUS, B., AND SEHEULT, A.: Exact Maximum A Posteriori Estimation for Binary Images. *Journal of the Royal Statistical Society* 51(2), 271–279, 1989.

[75] HAACKE, E. M., BROWN, R. W., THOMPSON, M. R., AND VENKATESAN, R.: *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley-Blackwell, 1999.

[76] HAJNAL, J. V., HILL, D. L. G., AND HAWKES, D. J.: *Medical Image Registration*. Taylor & Francis, 2001.

[77] HAMMING, R.: Error Detecting and Error Correcting Codes. *Bell System technical journal* 1950.

[78] HAN, X., AND FISCHL, B.: Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE Transactions on Medical Imaging* 26(4), 479–86, Apr. 2007.

[79] HECKEMANN, R. A., HAJNAL, J. V., ALJABAR, P., RUECKERT, D., AND HAMMERS, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33(1), 115–26, Oct. 2006.

[80] HECKEMANN, R. a., HAJNAL, J. V., ALJABAR, P., RUECKERT, D., AND HAMMERS, A.: Multiclassifier Fusion in Human Brain MR Segmentation: Modelling Convergence. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, R. Larsen, M. Nielsen, and J. Sporring, Eds., vol. 4191 of *LNCS*, pp. 815–822. Springer Berlin Heidelberg, Jan. 2006.

[81] HEIMANN, T., AND MEINZER, H.-P.: Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis* 13(4), 543–63, Aug. 2009.

[82] HEIMANN, T., AND MORRISON, B.: Segmentation of Knee Images: A Grand Challenge. In *MICCAI Workshop on Medical Image Analysis for the Clinic*. 207–214. 2010.

[83] HELD, K., ROTA KOPS, E., KRAUSE, B. J., WELLS, W. M., KIKINIS, R., AND MÜLLER-GÄRTNER, H. W.: Markov random field segmentation of brain MR images. *IEEE Transactions on Medical Imaging* 16(6), 878–86, Dec. 1997.

[84] HERMAN, G. T.: *Fundamentals of Computerized Tomography*. Springer London, London, 2009.

[85] HILL, D. L., BATCHELOR, P. G., HOLDEN, M., AND HAWKES, D. J.: Medical image registration. *Physics in medicine and biology* 46(3), R1–45, Mar. 2001.

[86] HOFMANN, M., STEINKE, F., ALTUN, Y., CAHILL, N., AND SCHOLKOPF, B.: Learning similarity measure for multi-modal 3D image registration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 186–193. June 2009.

[87] HOHE, J., ATESHIAN, G., REISER, M., ENGLMEIER, K.-H., AND ECKSTEIN, F.: Surface size, curvature analysis, and assessment of knee joint incongruity with MRI in vivo. *Magnetic Resonance in Medicine* 47(3), 554–561, 2002.

[88] HOWE, R. D., AND MATSUOKA, Y.: Robotics for surgery. *Annual review of biomedical engineering* 1, 211–40, Jan. 1999.

[89] JACK, C., PETERSEN, R. C., XU, Y. C., OBRIEN, P. C., SMITH, G. E., IVNIK, R. J., BOEVE, B. F., WARING, S. C., TANGALOS, E. G., AND KOKMEN, E.: Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology* 52(7), 1397–1403, 1999.

[90] JACK, C. R., BERNSTEIN, M. A., FOX, N. C., THOMPSON, P., ALEXANDER, G., HARVEY, D., BOROWSKI, B., BRITSON, P. J., L. WHITWELL, J., WARD, C., DALE, A. M., FELMLEE, J. P., GUNTER, J. L., HILL, D. L. G., KILLIANY, R., SCHUFF, N., FOX-BOSETTI, S., LIN, C., STUDHOLME, C., DECARLI, C. S., KRUEGER, G., WARD, H. A., METZGER, G. J., SCOTT, K. T., MALLOZZI, R., BLEZEK, D., LEVY, J., DEBBINS, J. P., FLEISHER, A. S., ALBERT, M., GREEN, R., BARTZOKIS, G., GLOVER, G., MUGLER, J., AND WEINER, M. W.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging* 27(4), 685–691, Apr. 2008.

[91] JAIN, A. K.: Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31(8), 651–666, June 2010.

[92] KAMEL, I., AND FALOUTSOS, C.: Hilbert R-tree: An improved R-tree using fractals. In *VLDB*. vol. 8958546. 1994.

[93] KEIDAR, Z., ISRAEL, O., AND KRAUSZ, Y.: SPECT/CT in Tumor Imaging: Technical Aspects and Clinical Applications. *Seminars in nuclear medicine* 33(3), 205–18, July 2003.

[94] KITTLER, J., HATEF, M., DUIN, R., AND MATAS, J.: On Combining Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(3), 226–239, Mar. 1998.

[95] KLAESER, A., MARSZALEK, M., AND SCHMID, C.: A Spatio-Temporal Descriptor Based on 3D-Gradients. *Procedings of BMVC* 99.1–99.10, 2008.

[96] Kontschieder, P., Kohli, P., Shotton, J., and Criminisi, A.: GeoF: Geodesic Forests for Learning Coupled Predictors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 65–72. Ieee, June 2013.

[97] Konukoglu, E., Glocker, B., Zikic, D., and Criminisi, A.: Neighbourhood approximation using randomized forests. *Medical Image Analysis* 17(7), 790–804, Oct. 2013.

[98] Kornegay, J. N., Bogan, J. R., Bogan, D. J., Childers, M. K., Li, J., Nghiem, P., Detwiler, D. a., Larsen, C. A., Grange, R. W., Bhavaraju-Sanka, R. K., Tou, S., Keene, B. P., Howard, J. F., Wang, J., Fan, Z., Schatzberg, S. J., Styner, M. a., Flanigan, K. M., Xiao, X., and Hoffman, E. P.: Canine models of Duchenne muscular dystrophy and their use in therapeutic strategies. *Mammalian Genome* 23(1-2), 85–108, Feb. 2012.

[99] Krauthgamer, R., and Lee, J. R.: Navigating nets : Simple algorithms for proximity search. *Proceedings of the fifteenth annual ACMSIAM symposium on Discrete algorithms* 798–807, 2004.

[100] Kumar, N., Zhang, L., and Nayar, S.: What is a Good Nearest Neighbors Algorithm for Finding Similar Patches in Images? In *ECCV*. vol. 5303 of *LNCS*, 364–378. Springer, Springer, 2008.

[101] Landman, B., and Warfield, S.: MICCAI Challenge Workshop on Segmentation: Algorithms, Theory and Applications (SATA), 2013.

[102] Landman, B. A., Warfield, S. K., Hammers, A., Akhondi-asl, A., Asman, A. J., Ribbens, A., Lucas, B., Avants, B., Ledig, C., Ma, D., Rueckert, D., Vandermeulen, D., Maes, F., Holmes, H., Wang, H., Wang, J., Doshi, J., Kornegay, J., Hajnal, J., Gray, K., Collins, L., Cardoso, M. J., Lythgoe, M., Styner, M., Armand, M., Miller, M., Aljabar, P., Suetens, P., Yushke-vich, P., Coupe, P., Wolz, R., and Heckemann, R.: MICCAI 2012 Workshop on Multi-Atlas Labeling, 2012.

[103] Lee, S., Wolberg, G., Chwa, K.-Y., and Shin, S. Y.: Image metamorphosis with scattered feature constraints. *IEEE Transactions on Visualization and Computer Graphics* 2(4), 337–354, 1996.

[104] Lee, S., Wolberg, G., and Shin, S.: Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics* 3(3), 228–244, 1997.

[105] Linguraru, M. G., Pura, J. a., Chowdhury, A. S., and Summers, R. M.: Multi-organ Segmentation from Multi-phase Abdominal CT via 4D Graphs Using Enhancement, Shape and Location Optimization. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, T. Jiang, N. Navab, J. Pluim, and M. Viergever, Eds., vol. 6363 of *LNCS*, pp. 89–96. Springer Berlin Heidelberg, Jan. 2010.

[106] Lorenzo-Valdés, M., Sanchez-Ortiz, G. I., Elkington, A. G., Mohiaddin, R. H., and Rueckert, D.: Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. *Medical Image Analysis* 8(3), 255–65, Sept. 2004.

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

[108] Ma, Z., Tavares, J. a. M. R. S., Jorge, R. N., and Mascarenhas, T.: A Review of Algorithms for Medical Image Segmentation and their Applications to the Female Pelvic Cavity. *Computer Methods in Biomechanics and Biomedical Engineering* 13(2), 235–46, Jan. 2010.

[109] Maaten, L. J. P. V. D., Postma, E. O., Herik, H. J. V. D., van den Herik, J., van der Maaten, L., and van den Herik, J.: Dimensionality Reduction: A Comparative Review. *Journal of Machine Learning Research* 10(February), 1–35, 2007.

[110] Madabhushi, A., and Udupa, J. K.: New methods of MR image intensity standardization via generalized scale. *Medical Physics* 33(9), 3426, 2006.

[111] MAES, F., COLLIGNON, A., VANDERMEULEN, D., MARCHAL, G., AND SUETENS, P.: Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging* 16(2), 187–98, Apr. 1997.

[112] MAHALANOBIS, P. C.: On the generalised distance in statistics. *Proceedings National Institute of Science India* 2(1), 49–55, 1936.

[113] MAINTZ, J., AND VIERGEVER, M. A.: A survey of medical image registration. *Medical Image Analysis* 2(1), 1–36, Mar. 1998.

[114] MALIK, J., BELONGIE, S., LEUNG, T., AND SHI, J.: Contour and Texture Analysis for Image Segmentation. *International Journal of Computer Vision* 43(1), 7–27, 2001.

[115] MANDUCA, A., YU, L., TRZASKO, J. D., KHAYLOVA, N., KOFLER, J. M., MC-COLLOUGH, C. M., AND FLETCHER, J. G.: Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Medical Physics* 36(11), 4911, 2009.

[116] MARROQUIN, J. L., VEMURI, B. C., BOTELLO, S., CALDERON, F., AND FERNANDEZ-BOUZAS, A.: An accurate and efficient bayesian method for automatic segmentation of brain MRI. *IEEE Transactions on Medical Imaging* 21(8), 934–45, Aug. 2002.

[117] MAURER, C. R., RENSHENG, Q., AND RAGHAVAN, V.: A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(2), 265–270, 2003.

[118] MCCANN, S., AND LOWE, D. G.: Local Naive Bayes Nearest Neighbor for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 3650–3656. June 2012.

[119] MENDRIK, A., BIESSELS, G. J., KUIJF, H., VINCKEN, K., AND VIERGEVER, M.: MRBrainS: MICCAI Grand Challenge on MR Brain Image Segmentation, 2013.

[120] MERKLE, R. C.: A Digital Signature Based on a Conventional Encryption Function. In *Advances in Cryptology Crypto 87*, C. Pomerance, Ed. vol. 293 of *CRYPTO '87*, 369–378. International Association for Cryptologic Research, Springer-Verlag, 1987.

[121] MUELLER, K.: Accelerating popular tomographic reconstruction algorithms on commodity PC graphics hardware. *IEEE Transactions on Nuclear Science* 52(3), 654–663, June 2005.

[122] MUJA, M., AND LOWE, D.: Fast approximate nearest neighbors with automatic algorithm configuration. In *International Conference on Computer Vision Theory and Applications*. 2009.

[123] NYÚL, L. G., UDUPA, J. K., AND ZHANG, X.: New Variants of a Method of MRI Scale Standardization. *IEEE Transactions on Medical Imaging* 19(2), 143–50, Feb. 2000.

[124] OMOHUNDRO, S. M.: Five Balltree Construction Algorithms. Tech. Rep. 063, International Computer Science Institute, Berkeley, 1989.

[125] OTSU, N.: A threshold selection method from gray-level histograms. *Automatica* 20(1), 62–66, 1975.

[126] OZUYSAL, M., FUA, P., AND LEPETIT, V.: Fast Keypoint Recognition in Ten Lines of Code. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1–8. Ieee, June 2007.

[127] PARK, H., BLAND, P. H., AND MEYER, C. R.: Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Transactions on Medical Imaging* 22(4), 483–92, Apr. 2003.

[128] PARZEN, E.: On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics* 33(3), 1065–1076, Sept. 1962.

[129] PAULY, O., GLOCKER, B., CRIMINISI, A., MATEUS, D., MÖLLER, A. M., NEKOLLA, S., AND NAVAB, N.: Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In *Medical Image Computing and Computer-Assisted Intervention*

– *MICCAI*, G. Fichtinger, A. Martel, and T. Peters, Eds., vol. 6893 of *LNCS*, pp. 239–247. Springer Berlin Heidelberg, 2011.

[130] PEKAR, V., McNUTT, T. R., AND KAUS, M. R.: Automated model-based organ delineation for radiotherapy planning in prostatic region. *International Journal of Radiation Oncology, Biology, Physics* 60(3), 973–80, Nov. 2004.

[131] PENNEY, G. P., WEESE, J., LITTLE, J. A., DESMEDT, P., HILL, D. L., AND HAWKES, D. J.: A Comparison of Similarity Measures for Use in 2-D3-D Medical Image Registration. *IEEE Transactions on Medical Imaging* 17(4), 586–95, Aug. 1998.

[132] PERONA, P., AND MALIK, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(7), 629–639, July 1990.

[133] PHAM, D. L., XU, C., AND PRINCE, J. L.: Current methods in medical image segmentation. *Annual review of biomedical engineering* 2, 315–37, Jan. 2000.

[134] POHL, K. M., BOUIX, S., NAKAMURA, M., ROHLFING, T., McCARLEY, R. W., KIKINIS, R., GRIMSON, W. E. L., SHENTON, M. E., AND WELLS, W. M.: A Hierarchical Algorithm for MR Brain Image Parcellation. *IEEE Transactions on Medical Imaging* 26(9), 1201–12, Sept. 2007.

[135] POPILOCK, R., SANDRASAGAREN, K., HARRIS, L., AND KASER, K. A.: CT artifact recognition for the nuclear technologist. *Journal of Nuclear Medicine Technology* 36(2), 79–81, June 2008.

[136] RADAU, P.: Cardiac MR Left Ventricle Segmentation Challenge. In *MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge.* 2009.

[137] RAM, P., LEE, D., MARCH, W., AND GRAY, A.: Linear-time Algorithms for Pairwise Statistical Problems. *NIPS* 1–10, 2009.

[138] REIMAN, E. M., UECKER, A., CASELLI, R. J., LEWIS, S., BANDY, D., DE LEON, M. J., DE SANTI, S., CONVIT, A., OSBORNE, D., WEAVER, A., AND THIBODEAU,

S. N.: Hippocampal volumes in cognitively normal persons at genetic risk for Alzheimer's disease. *Annals of neurology* 44(2), 288–91, Aug. 1998.

[139] RIVEST, J.-F., SOILLE, P., AND BEUCHER, S.: Morphological gradients. *Journal of Electronic Imaging* 2(4), 326, Oct. 1993.

[140] ROHLFING, T., BRANDT, R., MENZEL, R., AND MAURER, C. R.: Evaluation of Atlas Selection Strategies for Atlas-Based Image Segmentation with Application to Confocal Microscopy Images of Bee Brains. *NeuroImage* 21(4), 1428–42, Apr. 2004.

[141] ROHLFING, T., AND MAURER, C. R.: Multi-classifier framework for atlas-based image segmentation. *Pattern Recognition Letters* 26(13), 2070–2079, Oct. 2005.

[142] ROUSSEAU, F., HABAS, P., AND STUDHOLME, C.: A supervised patch-based approach for human brain labeling. *IEEE transactions on medical imaging* 30(10), 1852–1862, May 2011.

[143] ROUSSOPOULOS, N., KELLEY, S., AND VINCENT, F.: Nearest neighbor queries. *ACM SIGMOD Record* 24(2), 71–79, 1995.

[144] RUBNER, Y., TOMASI, C., AND GUIBAS, L.: The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision* 40(2), 99–121, 2000.

[145] RUDIN, L. I., OSHER, S., AND FATEMI, E.: Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60(1-4), 259–268, Nov. 1992.

[146] RUECKERT, D., SONODA, L. I., HAYES, C., HILL, D. L., LEACH, M. O., AND HAWKES, D. J.: Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* 18(8), 712–21, Aug. 1999.

[147] RUSU, R. B., AND COUSINS, S.: 3D is here: Point Cloud Library (PCL). In *IEEE International Conference on Robotics and Automation*. 1–4. May 2011.

[148] SCHÄFER, D., BORGERT, J., RASCHE, V., AND GRASS, M.: Motion-Compensated and Gated Cone Beam Filtered Back-Projection for 3-D Rotational X-ray Angiography. *IEEE Transactions on Medical Imaging* 25(7), 898–906, July 2006.

[149] SEDERBERG, T. W., AND PARRY, S. R.: Free-form deformation of solid geometric models. *ACM SIGGRAPH* 20(4), 151–160, Aug. 1986.

[150] SELLIS, T., ROUSSOPOULOS, N., AND FALOUTSOS, C.: The R+-Tree: A Dynamic Index For Multi-Dimensional Objects. In *International Conference on Very Large Data Bases*. 1987.

[151] SHANNON, C. E.: A Mathematical Theory of Communication. *The Bell System Technical Journal* 26(3), 1948.

[152] SHI, W., CABALLERO, J., LEDIG, C., ZHUANG, X., BAI, W., BHATIA, K., DE MARVAO, A., DAWES, T., OREGAN, D., AND RUECKERT, D.: Cardiac Image Super-Resolution with Global Correspondence Using Multi-Atlas PatchMatch. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., vol. 8151 of *LNCS*, pp. 9–16. Springer Berlin Heidelberg, 2013.

[153] SHIEE, N., BAZIN, P.-L., OZTURK, A., REICH, D. S., CALABRESI, P. A., AND PHAM, D. L.: A topology-preserving approach to the segmentation of brain images with multiple sclerosis lesions. *NeuroImage* 49(2), 1524–35, Jan. 2010.

[154] SHIMIZU, A., KIMOTO, T., KOBATAKE, H., NAWANO, S., AND SHINOZAKI, K.: Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. *International Journal of Computer Assisted Radiology and Surgery* 5(1), 85–98, Jan. 2010.

[155] SINOP, A. K., AND GRADY, L.: Accurate Banded Graph Cut Segmentation of Thin Structures Using Laplacian Pyramids. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, R. Larsen, M. Nielsen, and J. Sporring, Eds., vol. 4191 of *LNCS*, pp. 896–903. Springer Berlin Heidelberg, 2006.

[156] SLED, J., AND PIKE, G.: Understanding intensity non-uniformity in MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, vol. 1496, pp. 614–622. Springer Berlin Heidelberg, 1998.

[157] SLED, J. G., ZIJDENBOS, A. P., AND EVANS, A. C.: A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data. *IEEE Transactions on Medical Imaging* 17(1), 87–97, Feb. 1998.

[158] SMITH, S. M., JENKINSON, M., WOOLRICH, M. W., BECKMANN, C. F., BEHRENS, T. E. J., JOHANSEN-BERG, H., BANNISTER, P. R., DE LUCA, M., DROBNJAK, I., FLITNEY, D. E., NIAZY, R. K., SAUNDERS, J., VICKERS, J., ZHANG, Y., DE STEFANO, N., BRADY, J. M., AND MATTHEWS, P. M.: Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23 Suppl 1, S208–19, Jan. 2004.

[159] SMITH, T. B., AND NAYAK, K. S.: MRI artifacts and correction strategies. *Imaging in Medicine* 2(4), 445–457, Aug. 2010.

[160] SNAVELY, N., SEITZ, S. M., AND SZELISKI, R.: Photo tourism: exploring photo collections in 3D. In *ACM SIGGRAPH*. 835–846. New York, NY, USA, 2006.

[161] SONG, Z., TUSTISON, N., AVANTS, B., AND GEE, J. C.: Integrated Graph Cuts for Brain MRI Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, R. Larsen, M. Nielsen, and J. Sporring, Eds., vol. 4191 of *LNCS*, pp. 831–838. Springer Berlin Heidelberg, Jan. 2006.

[162] SOTIRAS, A., DAVATZIKOS, C., AND PARAGIOS, N.: Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging* 32(7), 1153–90, July 2013.

[163] SPROULL, R. F.: Refinements to nearest-neighbor searching in k-dimensional trees. *Algorithmica* 6(1-6), 579–589, 1991.

[164] STAUFFER, C.: Learning a Probabilistic Similarity Function for Segmentation. 50–50. June 2004.

[165] STUDHOLME, C., HILL, D., AND HAWKES, D.: An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition* 32(1), 71–86, Jan. 1999.

[166] Tenenbaum, J. B., Silva, V. D., and Langford, J. C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290, 2000.

[167] Thomson, J.: On the Structure of the Atom: an Investigation of the Stability and Periods of Oscillation of a number of Corpuscles arranged at equal intervals around the Circumference of a Circle; with Application of the Results to the Theory of Atomic Structure. *Philosophical Magazine* 7(39), 237–265, 1904.

[168] Toivanen, P. J.: New geodosic distance transforms for gray-scale images. *Pattern Recognition Letters* 17(5), 437–450, May 1996.

[169] Tomasi, C., and Manduchi, R.: Bilateral filtering for gray and color images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 839–846. 1998.

[170] Tong, T., Wolz, R., Coupé, P., Hajnal, J. V., and Rueckert, D.: Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. *NeuroImage* 76, 11–23, Aug. 2013.

[171] Torralba, A., Fergus, R., and Freeman, W. T.: 80 Million Tiny Images: a Large Data Set for Nonparametric Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(11), 1958–70, Nov. 2008.

[172] Tustison, N. J., Avants, B. B., Cook, P. a., Zheng, Y., Egan, A., Yushkevich, P. a., and Gee, J. C.: N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging* 29(6), 1310–20, June 2010.

[173] Vaidya, P. M.: An O(n log n) Algorithm for the All-Nearest-Neighbors Problem. *Discrete and Computational Geometry* 4(2), 101–115, 1989.

[174] van der Lijn, F., den Heijer, T., Breteler, M. M. B., and Niessen, W. J.: Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage* 43(4), 708–20, Dec. 2008.

[175] VAN LEEMPUT, K., MAES, F., VANDERMEULEN, D., AND SUETENS, P.: Automated model-based bias field correction of MR images of the brain. *IEEE Transactions on Medical Imaging* 18(10), 885–96, Oct. 1999.

[176] VIOLA, P., AND III, W. W.: Alignment by Maximization of Mutual Information. *International Journal of Computer Vision* 24(2), 137–154, 1995.

[177] VOVK, U., PERNUS, F., AND LIKAR, B.: A Review of Methods for Correction of Intensity Inhomogeneity in MRI. *IEEE Transactions on Medical Imaging* 26(3), 405–21, Mar. 2007.

[178] WANG, H., SUH, J. W., DAS, S., PLUTA, J., ALTINAY, M., AND YUSHKEVICH, P.: Regression-Based Label Fusion for Multi-Atlas Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1113–1120. June 2011.

[179] WANG, H., SUH, J. W., DAS, S. R., PLUTA, J., CRAIGE, C., AND YUSHKEVICH, P. A.: Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(3), 611–623, June 2012.

[180] WANG, H., AND YUSHKEVICH, P.: Multi-atlas Segmentation without Registration: A Supervoxel-Based Approach. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, vol. 8151, pp. 535–542. 2013.

[181] WANG, J., FAN, Z., VANDENBORNE, K., WALTER, G., SHILOH-MALAWSKY, Y., AN, H., KORNEGAY, J. N., AND STYNER, M. A.: A computerized MRI biomarker quantification scheme for a canine model of Duchenne muscular dystrophy. *International Journal of Computer Assisted Radiology and Surgery* 8(5), 763–74, Sept. 2013.

[182] WANG, L. V.: Multiscale photoacoustic microscopy and computed tomography. *Nature photonics* 3(9), 503–509, Aug. 2009.

[183] WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P.: Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing* 13(4), 600–12, Apr. 2004.

[184] WARFIELD, S. K., ZOU, K. H., AND WELLS, W. M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging* 23(7), 903–21, July 2004.

[185] WEBER, O., DEVIR, Y. S., BRONSTEIN, A. M., BRONSTEIN, M. M., AND KIMMEL, R.: Parallel algorithms for approximation of distance maps on parametric surfaces. *ACM Transactions on Graphics* 27(4), 1–16, Oct. 2008.

[186] WEINBERGER, K. Q., AND SAUL, L. K.: Distance Metric Learning for Large Margin Nearest Neighbor Classification. *Journal of Machine Learning Research* 10, 207–244, 2009.

[187] WEISENFELD, N. I., AND WARFIELD, S. K.: Automatic segmentation of newborn brain MRI. *NeuroImage* 47(2), 564–72, Aug. 2009.

[188] WELLS, W. M., VIOLA, P., ATSUMI, H., NAKAJIMA, S., AND KIKINIS, R.: Multimodal volume registration by maximization of mutual information. *Medical Image Analysis* 1(1), 35–51, 1996.

[189] WLUKA, A. E.: How does tibial cartilage volume relate to symptoms in subjects with knee osteoarthritis? *Annals of the Rheumatic Diseases* 63(3), 264–268, Mar. 2004.

[190] WOLZ, R., ALJABAR, P., RUECKERT, D., HECKEMANN, R. A., AND HAMMERS, A.: Segmentation of subcortical structures and the hippocampus in brain MRI using graph-cuts and subject-specific a-priori information. In *IEEE International Symposium on Biomedical Imaging (ISBI)*. 470–473. June 2009.

[191] WOLZ, R., CHU, C., MISAWA, K., FUJIWARA, M., MORI, K., AND RUECKERT, D.: Automated Abdominal Multi-Organ Segmentation With Subject-Specific Atlas Generation. *IEEE Transactions on Medical Imaging* 32(9), 1723–30, Sept. 2013.

[192] XU, L., KRZYZAK, A., AND SUEN, C.: Methods of Combining Multiple Classifiers and Their Applications to Handwriting Recognition. *IEEE Transactions on Systems, Man, and Cybernetics* 22(3), 418–435, 1992.

[193] Xue, Z., Shen, D., and Davatzikos, C.: Determining Correspondence in 3-D MR Brain Images Using Attribute Vectors as Morphological Signatures of Voxels. *IEEE Transactions on Medical Imaging* 23(10), 1276–91, Oct. 2004.

[194] Yianilos, P. N.: Data Structures and Algorithms for Nearest Neighbor Search in General Metric Spaces. In *ACM-SIAM Symposium on Discrete Algorithms.* 311–321. 1993.

[195] Zhao, H.: A fast sweeping method for Eikonal equations. *Mathematics of Computation* 74(250), 603–628, May 2004.

[196] Zikic, D., Glocker, B., and Criminisi, A.: Atlas Encoding by Randomized Forests for Efficient Label Propagation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, Eds., vol. 8151 of *LNCS*, pp. 66–73. Springer Berlin Heidelberg, 2013.

[197] Zitová, B., and Flusser, J.: Image registration methods: a survey. *Image and Vision Computing* 21(11), 977–1000, Oct. 2003.