1-7-2014 12:00 AM

Image Fusion and Axial Labeling of the Spine

Brandon Miles
The University of Western Ontario

Supervisor
Aaron Fenster
The University of Western Ontario Joint Supervisor
Shuo Li
The University of Western Ontario

Graduate Program in Biomedical Engineering
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

IMAGE FUSION AND AXIAL LABELING OF THE SPINE
(Thesis format: Integrated Article)

by

Brandon Miles

Graduate Program in BioMedical Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

In order to improve radiological diagnosis of back pain and spine disease, two new algorithms have been developed to aid the 75% of Canadians who will suffer from back pain in a given year. With the associated medical imaging required for many of these patients, there is a potential for improvement in both patient care and healthcare economics by increasing the accuracy and efficiency of spine diagnosis. A real-time spine image fusion system and an automatic vertebra/disc labeling system have been developed to address this.

Both magnetic resonance (MR) images and computed tomography (CT) images are often acquired for patients. The MR image highlights soft tissue detail while the CT image highlights bone detail. It is desirable to present both modalities on a single fused image containing the clinically relevant detail. The fusion problem was encoded in an energy functional balancing three competing goals for the fused image: 1) similarity to the MR image, 2) similarity to the CT image and 3) smoothness (containing natural transitions). Graph-Cut and convex solutions have been developed. They have similar performance to each other and outperform other fusion methods from recent literature. The convex solution has real-time performance on modern graphics processing units, allowing for interactive control of the fused image. Clinical validation has been conducted on the convex solution based on 15 patient images. The fused images have been shown to increase confidence of diagnosis compared to unregistered MR and CT images, with no change in time for diagnosis based on readings from 5 radiologists.

Spinal vertebrae serve as a reference for the location of surrounding tissues, but vertebrae have a very similar appearance to each other, making it time consume for radiologist to keep track of their locations. To automate this, an axial MR labeling algorithm was developed that runs in near real-time. Probability product kernels and fast integral images combined with simple geometric rules were used to classify pixels, slices and vertebrae. Evaluation was conducted on 32 lumbar spine images and 24 cervical spine images. The algorithm demonstrated 99% and 79% accuracy on the lumbar and cervical spine respectively.

Keywords: Spine Diagnosis, Image Fusion, Automatic Labeling, Convex Optimization,

Graph-Cuts, Probability Product Kernel Classifiers, Integral Images

Co-Authorship Statement

Chapter 2 was published in IEEE Transactions on BioMedical Engineering as: Brandon Miles, Ismail Ben-Ayed, Max Law, Gregory J. Garvin, Aaron Fenster, Shuo Li, "'Spine Image Fusion via Graph-Cuts," IEEE Transactions on Biomedical Engineering (TBME), July 2013, Vol. 60 No. 5, pp. 1841-50. I formulated the energy functional and developed all the needed software for the image registration, fusion algorithm and the validation. I registered the images along with running all the software and performing all the validation. I also manually segmented the masks for validation. Dr. Ismail Ben-Ayed provided the initial idea to use graph-cuts for image fusion. Dr. Max Law and I developed the algorithm for registering the CT and MR images. Dr. Gregory J. Garvin provided clinical guidance for this work including verification of the validation masks and identification and description of the clinical case studies. The manuscript was prepared by myself, with extensive mentorship by Dr. Ismail Ben-Ayed with a particular emphasis on the formulation section and introduction. Dr. Aaron Fenster supervised this work providing extensive mentorship in preparing the introduction, methods, results and conclusion portions of the manuscript along with the validation procedure. The work in this paper was conducted under the supervision of Dr. Aaron Fenster and Dr. Shuo Li.

Chapter 3 is under major revisions with Medical Physics as: Brandon Miles, Jing Yuan, Ismail Ben-Ayed, Max W.K. Law, Vladislav Miropolsky, Rohit Joshi, Gregory J. Garvin, Aaron Fenster, Shuo Li, "'Real Time MR/CT Image Fusion for Lumbar Spine Diagnosis," Medical Physics. I developed most of the software, including all the software needed for image registration and the validation. The convex fusion algorithm and associated software was developed by Dr. Jing Yuan. I optimized his software for speed, and integrated it with my earlier software for graph-cut image fusion from chapter 3. The images were registered using the algorithm developed by Dr. Max Law and myself. I ran all the software including registering, fusing and validating the images. Dr. Gregory J. Garvin provided clinical guidance for this work including verification of the validation masks and feedback on the quality of the fused images.

Dr. Gregory J. Garvin, Dr. Vladislav Miropolsky and Dr. Rohit Joshi helped in identifying, acquiring and describing the clinical cases. Both Dr. Ismail Ben-Ayed and Dr. Aaron Fenster provided extensive mentorship of this work including revising the manuscript. The work in this paper was conducted under the supervision of Dr. Aaron Fenster and Dr. Shuo Li.

Chapter 4 is under preparation to be submitted to a clinical radiology journal. I wrote the software for the interactive fusion viewer. I also analyzed the data in consultation with Dr. Yves Bureau. The testing methodology and experimental protocol were developed by myself, Dr. Gregory J. Garvin, Dr. Ido Druckmann, Dr. Vladislav Miropolosky and Dr. Ismail Ben-Ayed, with input from Dr. Rohit Joshi. The images were acquired by Dr. Ido Druckmann. Initial protrusion measurements were made by Dr. Gregory J. Garvin. The rankings and measurements were made by Dr. Itai Gadiel, Dr. Olga Shmuilovich, Dr. Suha Ghoul, Dr. Giulio Muscedere and Dr. KengYeow Tay. Dr. Aaron Fenster provided mentorship included revisions of the final manuscript. The work in this paper was conducted under the supervision of Dr. Aaron Fenster and Dr. Shuo Li.

Chapter 5 is under submission to IEEE Transactions on BioMedical Engineering as: Brandon Miles, Seyed-Parsa Hojjat, Michael H. Wang, Gregory J. Garvin, Shuo Li, Aaron Fenster, Ismail Ben-Ayed, "'Spine Axial MR Classification via Integral Kernel Density Estimates," IEEE Transactions on Biomedical Engineering. The original algorithm was developed by myself and Dr. Ismail Ben-Ayed. I wrote all the software include the testing software and the interactive viewer. Dr. Parsa Hojjat manually labeled the lumbar spine data. He also validated this data for an early version of the paper. Michael Wang labeled the cervical spine data. I validate both data sets for the final paper. Dr. Gregory J. Garvin provided clinical guidance. The manuscript was prepared by myself, Dr. Parsa Hojjat, Dr. Ismail Ben-Ayed and Dr. Aaron Fenster. This work was supervised by Ismail Ben-Ayed and Aaron Fenster. I was under the supervision of Dr. Aaron Fenster and Dr. Shuo Li during the completion of this work.

Acknowlegements

I would like to acknowledge the help of the many people, without whom I would not have been able to complete this thesis.

I would like to start by thanking my supervisors: Shuo Li and Aaron Fenster. Dr. Li has provided me the opportunity to study at the University of Western Ontario allowing me to collaborate with so many excellent people. I would like to thank Dr. Fenster for all his supervision, mentorship and investment of time over the course of my Ph.D. Without his guidance and support this thesis would not have been possible.

I would also like to thank Dr. Ismail Ben-Ayed for serving as one of my supervisors in every way except by name, who with Dr. Fenster has served as my mentor through the latter half of my Ph.D.

Thank you also to Dr. Jing Yuan and Dr. Max Law for their many contributions. I would also like to thank Dr. Greg Garvin for all his clinical guidance along with other members of the radiology team including: Dr. Ido Druckmann, Dr. Rohit Joshi and Dr. Vladislav Miropolsky. Thank you also to Dr. Olga Veksler for her guidance throughout my degree and serving on my advisory committee.

Special thanks to my lab mates and colleagues: Abraam, Hamid, Kaymar, Mariam and Cyrus who is like a brother to me. I'd like to thank my parents, sisters, grandpa and other extended family for all their love and support. To Flora, my wonderful wife, you have supported me in so many ways. Thank you for all your love, encouragement and sacrifices through these years. You are my better half, I love you.

Contents

List of Figures

List of Tables

List of Abbreviations, Symbols, and Nomenclature

CABS	Convex Absolute Value
CLT	Contourlet
CSD	Convex Squared Difference
CT	Computed Tomography
CPU	Central Processing Unit
DSA	Digital Subtraction Angiography
DWT	Discrete Wavelet Transform
FLE	Fiducial Localization Error
GC	Graph-Cuts
GPU	Graphics Processing Unit
HU	Houndsfield Unit
ICC	Intra-class Correlation
KDE	Kernel Density Estimate
MI	Mutual Information
MR	Magnetic Resonance
PD	Proton Density
PET	Positron Emission Tomography
PPK	Probability Product Kernel
SPECT	Single Photon Emission Computed Tomography
SSD	Sum of Squared Difference
SSIM	Structural Similarity Metric
TP/FP/TN/FN	True Positive, False Positive, True Negative, False Negative
TRE	Target Registration Error

Chapter 1

Introduction

1.1 Back pain and spinal diseases

Back pain and spinal disease is a debilitating condition that is adversely affecting the lives of many Canadians and people all over the world. It is estimated that in Canada, musculo-skeletal disorders including the limbs and spine costs the economy over $20 billion dollars each year. Back pain and spinal disease is a significant part of this number [35]. It is estimated that 75% of people will experience back pain in their lives [43]. Additionally, in 1988 there were 101.8 million lost work days in the United States because of back pain [18].

A significant portion of this disease burden results in hospital care. In 2008 1.8% of all hospital visits were due to back pain and disease [43]. Additionally, there has been a large increase in medical imaging for the diagnosis and treatment of spine disorders. Between 1994 and 2006, there has been a 300% increase in lower spine magnetic resonance imaging (MR/MRI) scans [11].

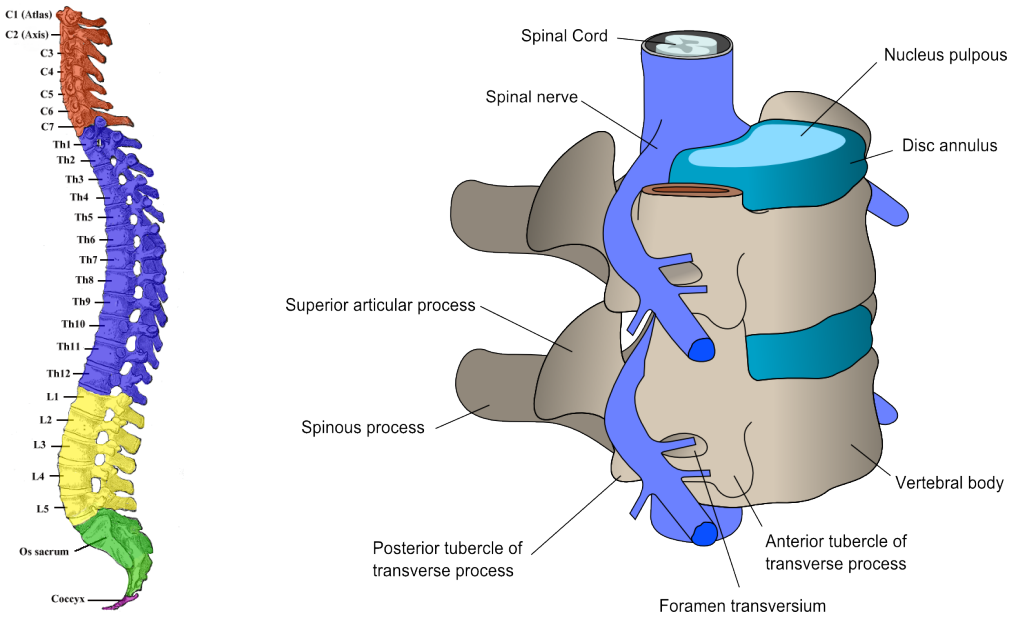
This prevalence of spine imaging motivates a need for effective analysis of these images. In particular, when both computed tomography (CT) and magnetic resonances images (MR/MRI) are available they provide complementary information. CT images highlight the bone details especially the cortical bone edges, while MR images highlight the soft tissue details such as

bone disc. In this work the combined visualization of CT and MR images is investigated for the diagnosis and treatment of back pain and spinal diseases. To further enhance diagnosis based on spine images the problem of accurately labeling the location of vertebrae on images of the spine is also considered.

1.2 Anatomy of the Human Spine

The human spine consists of a series of vertebrae, separated by intervertebral discs. These form a flexible column, which can be referred to as the *spinal column* since they appears in the form of a vertical column when a human is standing upright. The spinal cord runs along this column. There are five main sections. From top to the bottom these consists of the cervical spine, containing the seven cervical vertebrae, the thoracic spine, containing the twelve thoracic vertebrae, the lumbar spine containing five vertebrae, the sacral spine, containing five fused vertebrae and the coccyx. This column supports the trunk, connecting the head and limbs as well as protecting the spinal cord and nerves [38]. Intervertebral discs are located between adjacent vertebrae. They bear weight and allow for motion of the spine. The spinal cord runs through the vertebrae and behind the discs. This is shown in Fig. 1.1. Fig. 1.1 (a) shows five sections of the human spine and Fig. 1.1 (b) shows two vertebrae, the spinal cord, nerves and the intervertebral discs.

The individual vertebrae are bony structures (see Fig. 1.2). They consist of both cortical bone on the exterior of the vertebrae and trabecular bone found in the centre of the vertebral body. The vertebral body is found at the anterior (front side) of the vertebrae and the spinous process is found at the posterior (back) of the vertebrae. Transverse processes are found to the sides, with the facet joints located below the transverse processes. The facet joints allow for motion of vertebrae. The spinal cord runs behind the vertebral body in a gap referred to as the vertebral foramena. The intervertebral discs are located between adjacent vertebrae. On the exterior of the discs is the annulus fibrosus, which is a ring of fibrocartilage. This encases the



(a) (b)

Figure 1.1: Diagrams of spinal anatomy: a) The human vertebral column showing the cervical spine, thoracic spine, lumbar spine, sacrum and coccyx b) a typical vertebrae, showing the intervertebral discs, nerve tissue and vertebrae (used under Creative Commons attribution-share alike 3.0 unported, by Anuskafm).

nucleus pulposus, which at birth is mostly water, but stiffens with age. The purpose of these discs is for load bearing and to allow motion between the vertebrae [38].

1.3 MR and CT for Spine imaging

Magnetic Resonance Imaging (MRI/MR) and computed tomography (CT) are two common modalities for imaging the human spine. Both provide three dimensional (3D) images of the spine, highlighting different parts of the anatomy. An MR scanner works because the spins of certain nuclei will align in a strong magnetic field. One such nucleus is the proton, the nucleus of a hydrogen atom. Hydrogen atoms are typically found in the human body as parts of water and fat molecules, which make up most of our soft tissues. For the spine this is why the nerves, intervertebral discs, spinal cord and cerebral spinal fluid are visible in MR images.

To generate an image from the aligned protons, the magnetic field in the scanner is varied

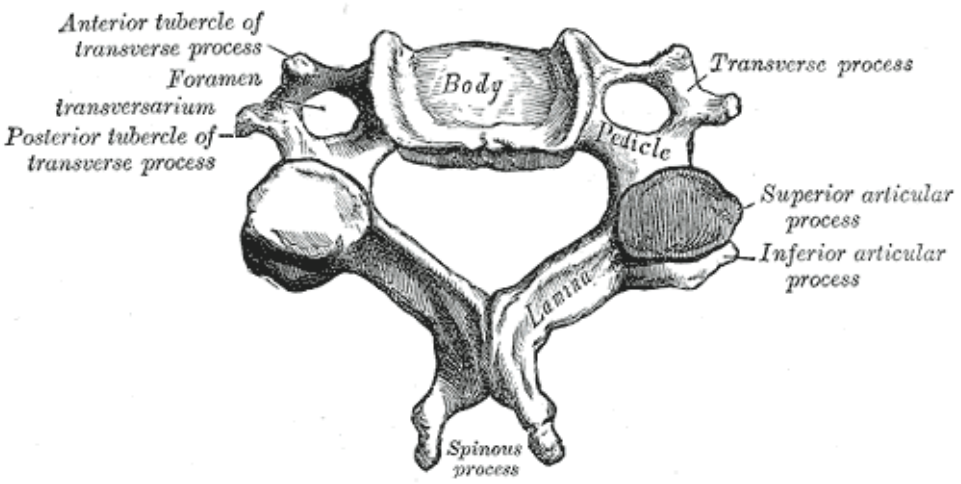


Figure 1.2: Diagram of a Cervical Vertebrae

over the volume of the patient. A pulse sequence of radio-frequency waves is transmitted to change the protons' alignment of the magnetic field. Over a period of time, the protons will relax back towards their original alignment, emitting radio waves as this occurs. It is the pattern of these emitted radio waves that allows the location and density of protons in the human body to be recorded as an image. Depending on the exact pulse sequence the image will be weighted: for fat contrast (T1), water contrast (T2) or the density of protons (PD). [45] For the human spine, T2 is the most commonly used, but T1 and PD scans can also provide valuable information.

Figure 1.3 shows a Seimens Avanto 1.5 Tesla MR scanner (Seimens AG, Erlangen, Germany). Under the blanket is the spine coil for acquiring radio frequency signals from the spine. The patient will be scanned lying on their back. For spine imaging one of the images typically acquired is generated from a T2 weighted 3D pulse sequence. This 3D image allows views of the spine in any direction, with images typically formatted as a series of two dimensional (2D) image slices in the sagittal plane (looking from the side), the axial plane (looking from head to toe), and coronal plane (looking from front to back).

Another common modality for spine imaging is CT. When X-rays pass through matter they are attenuated. Attenuation is tissue specific in that tissues with higher density and atoms with higher atomic numbers have an increased X-ray attenuation. Bone contains large amounts of

Figure 1.3: Seimens Avanto 1.5 Tesla MR scanner. A spine coil is located under the blanket.

calcium, which attenuates a high fraction of the incident X-rays. This is represented as a whiter region on an X-ray or CT image, and air appears black as it attenuates almost no X-rays. The Houndsfield scale is used to crepresent attenuation levels and is based on air having a value of -1000 Houndsfield Units (HU) and water having a value of 0 HU. Cortical bone is typically around 1000 HU [20].

To form a CT image, X-ray attenuation measurements are obtained from varying angles around the patient. These attenuation measurements are then used to reconstruct the 3D image either on a slice-by-slice 2D set of cross-sectional images, or as a 3D object [20]. In modern CT scanners, the image is acquired as a helical scan, where the patient is moved through the scanner as the scanner rotates around the patient. The reconstructed 3D image can then be used to generate sets of 2D slices in the sagittal, axial and coronal directions [17]. A modern CT

scanner is shown in Fig 1.4. Again the patient is typically scanned, while on their back.

Figure 1.4: A GE Healthcare Discovery CT750 HD CT scanner.

1.3.1 Images used in this Thesis

For the purpose of this thesis only images already acquired from patients were utilized. Furthermore the methods developed were only targeted towards better utilizing images that would have already been acquired without this research. This was to avoid the need for additional imaging with its associated healthcare costs and possible risks to the patients (such as for CT images).

Both the lumbar spine and cervical spine were investigated. For the lumbar spine T2 weighted 3D SPACE sequences were utilized, providing both sagittal and axial reformatted views. The T2 weighted images were used as they provided good contrast for the disc and

nerve structures in the lumbar spine. T1 axial images were also utilized for the purpose of labelling disc and vertebrae as they are commonly used for clinical diagnosis. For the cervical spine, T2 axial weighted images were utilized. CT helical images for the lumbar and cervical spine. Both sagittal and axial formatted images were used.

For both MR and CT the pixel and slice spacings varied for each patient. The ranges of spaces have been detailed for the data sets utilized (see Chapters 2,3 and 5). Additionally, unless explicitly stated no contrast was utilized in the images.

1.4 Diseases of the Spine

There are several diseases/pathologies that are common in the lower back and contribute to loss of mobility along with back pain. These include bone and joints disorders such as fractures and osteophyte formation. There are also nerve and tissue specific diseases including disc herniations, disc bulges, disc protrusion/extrusion, sciatica and spinal stenosis. Metastatic disease is also a concern.

Bone and Joint Diseases

Typical bone diseases/pathologies include fracture of vertebrae, joint disease and osteophyte formation. Examples of fractures include fractures caused by trauma to a cervical vertebral body and compression fractures where excessive force has been exerted axial along the spinal column, causing one or more vertebrae to fracture. Osteophytes are bony growths in response to areas of specific pressure on the spine. Depending on the location of the osteophyte, nerve roots or ligaments can be impacted, causing either localized back pain or local paralysis [38]. Various forms of arthritis can affect the spine and include osteoarthritis (usually a result of age), rheumatoid arthritis, which is an autoimmune disease and ankylosis, where the vertebrae become fused together.

Disc Disease

Degenerate disc disease encompasses a large number of disc abnormalities. These include but are not limited to bulging, herniations such as protrusions / extrusion, narrowing of the disc and osteophytes. A bulge is defined as the disc protruding beyond the end plates of the vertebrae over more than 50% of the circumference of the disc, i.e. bulging out into the regions beyond the vertebrae. A disc extending focally is called herniation and can be characterized as a protrusion/extrusion [14]. The biggest danger in these scenarios is that the disc material will impact the nerve tissue causing either pain or local paralysis. A diagram of a normal, herniated and protruding disc can be seen in Fig. 1.5.

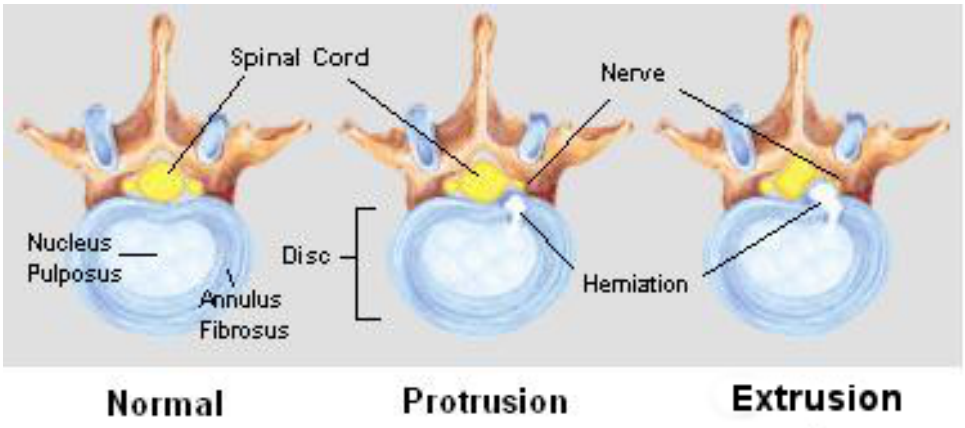


Figure 1.5: Illustrations of normal and protruding discs. In the right most image the nerve is being compressed due to the extrusion.

Nerve Diseases

Diseases and pathology that impact the nerves can be very debilitating as they can cause pain and/or local paralysis. An example of this is Sciatica, which is pain radiating downwards from the lower back into the thigh. This can be caused by compressed nerves at the L5 or S1 levels because of herniated discs or osteophytes. [38]. Spinal stenosis is another dangerous disorder, where there is a narrowing of the spinal cord, typically due to discs or osteophytes impacting the cord.

Metastatic Disease

The spine is a common location for the metastasis of tumors. This will often cause lesions in the vertebrae and is indicative of a primary cancer in the patient. Primary tumours can also form in the spine.

1.5 Overview of the Image Fusion Process

The goal of image fusion is to combine two or more images of the same object into a single image containing the salient details of the two or more input images. A simple example of this includes multifocal images [27], where two camera images with different focus settings are combined into a single image. This can also be applied to medical images, in particular image fusion of CT and MR images of the spine will be the subject of several of the following chapters.

In order to construct a fused image, several processing steps need to occur. These can be divded into image registration, pre-processing and fusion as show in Fig. 1.6. Image registration is the process of aligning multiple images and is essential for proper image fusion. If the images are not properly aligned prior to fusion, the output image would have only limited value to the observer. The second step is pre-processing, in this case the aligned images are processed to make them more suitable for fusion by filtering out any unwanted details, while enhancing relavent details. The final step is the pixel-wise image fusion of the aligned and pre-processed images. Image registration, and pre-processing will be discussed in this section. Pixel-wise fusion algorithms will be discussed in greater detail in section 1.6 along with chapters 2 and 3. Since pixel-wise image fusion was the primary focus of chapters 2 and 3, registration will only be briefly discussed and has only been considered in terms of it being a necessary step to achieved image fusion.

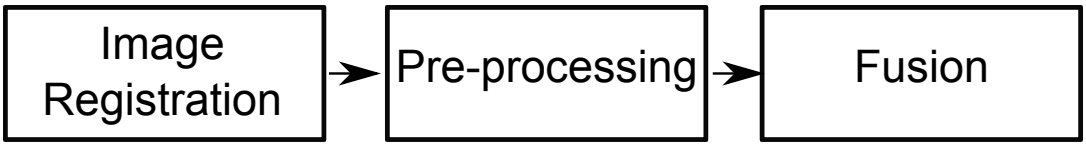


Figure 1.6: Typical fusion process, 1) Images are aligned (registration), 2) preprocessing occurs, 3) pixel-wise image fusion is carried out.

1.5.1 Distinction between Image Registration and Image Fusion

Image *registration* is the process of aligning multiple images together. When registered images are overlayed and displayed together they are often refered to as being *fused*, with *registration* and *fusion* being almost synonymous in some circles of the medical imaging community. This can be a source of confusion as pixel-wise image fusion refers only to the process of combining multiple preregistered images. In this thesis, *image fusion* is refered to as the process of combining multiple images into a single image and will *never* to be taken to mean image *registration*. The term image fusion will refer to either the entire fusion process of registration, pre-processing and pixel-wise fusion, or specifically to pixel-wise image fusion. It is hoped this will be self evident from the context. This distinction is detailed figure 1.7.

1.5.2 Image Registration

Typical algorithms for image registration combine three elements: a *transform*, an alignment *metric* and an *optimizer*. The first element, the geometric transform defines how a *study* image, can be aligned to match a *template*. Common types of transforms include rigid body transformation (rotation and translation), affine (rotation, translation and shearing) and non-rigid (warping the image). Once a suitable transformation has been selected, a measurement of how well the images are aligned needs to be determined. This is refered to as the metric. Common examples include the sum of squared differences (SSD) between overlapping pixels, or the mutual information (MI) between the images. The former method is useful when the images to be meausred are from the same modality as it only looks at pixel intensity differences. Similar images have low SSD values, while dissimilar images have high SSD values. The later: Mutual

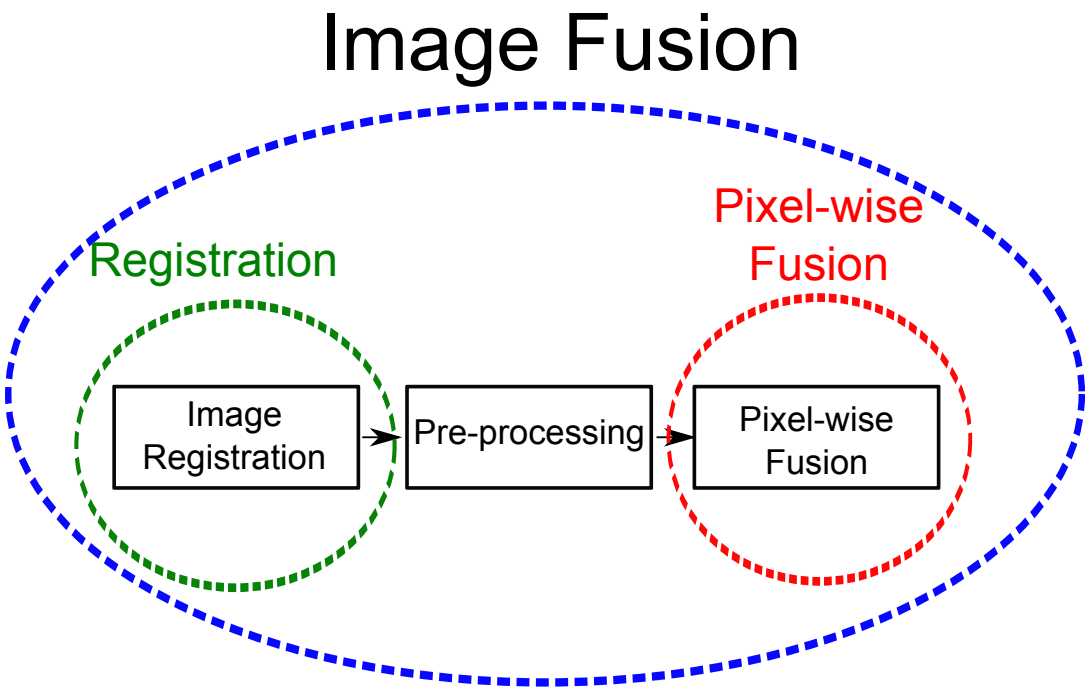


Figure 1.7: In this work, fusion should not be confused with registration. Image fusion can refer to both the entire process of registration, pre-processing and pixel-wise fusion, or pixel-wise fusion specifically.

information (MI), attempts to determine how well related the information in the two images. Since MI classifies information as opposed to pixel intensities, it is often used for multimodal image registration where information is similar between the images, but pixel intensities differ. A high value of MI indicates a good alignment, while a low MI value is indicative of poor alignment. The third element in a typical registration algorithm is the optimizer. The optimizer will adjust the current transform based on the results of the metric, searching for either a minimum or maximum value of the metric appropriately. By repeating this process iteratively the overall metric value can be optimized, which should correspond to a good alignment between the study and template images.

In this work spine MR and CT images were aligned. A rigid body transformation was used, although a piecewise rigid transformation (rigidly transforming the vertebrae) could also have been applied. Both methods preserve the structure of the spine, not allowing bones to deform in the registration process. For the lumbar spine it was found that rigid body deformation was

a suitable assumption (see chapter 2). This is likely because patients are scanned lying on their backs in similar positions for both CT and MR even when the images are acquired up to a year apart. For the rigid transformation a method using 3D versors (unit quarternions) was applied. For the metric, since MR and CT are from different modalities MI was used. The optimizer was designed for the optimization of versor transforms. Chapters 2,3 and 6 contain further details and discussion on the methods used in this work, but it should be noted that the approaches employed are fairly standard as image fusion as opposed to registration was the focus of this work.

1.5.3 Preprocessing

Preprocessing is used to enhance images prior to fusion. This is an optional step, that can be omitted for some applications. For spine image fusion of CT and MR images, this is a useful step as unwanted bone details can be removed from CT images. Additionally the relative intensities of MR and CT images can be adjusted to better highlight either bone or soft tissue detail when needed. A specific description of the preprocessing steps used is first provided in chapter 2, when CT/MR fusion is considered. CT soft tissue removal was accomplished by thresholding out detail below a given HU. This was followed by linearly adjusting the CT intensities so that the maximum CT pixel intensity approximately matched the maximum MR pixel intensity. These values are used again in chapter 3, but it is noted that their adjustment could be beneficial to clinical diagnosis. If the amount of CT soft tissue detail present in the fused image can be adjusted along with the relative intensity of the MR, it is possible to create pathology specific fusion settings. These values could then be adjusted interactively by the radiologists while they examine the spine images, potentially aiding diagnosis.

1.6 Background and prior art on Pixel-wise Image Fusion

In this section we present a discuss of applications of image fusion and then detail prior art using transform-based methods, followed by prior art using variational methods. Most of the methods that will be discussed were not introduced for the purpose of spine image fusion or even medical image fusion. It has been demonstrated that many of these methods are applicable to medical image fusion (e.g. [36]). This justifies a general discussion of these methods with potential medical applications being implied.

1.6.1 Applications of Image Fusion

As previously mentioned, one application of image fusion is for the combination of multifocal images [27]. In geographical applications, multi-spectral image fusion has proven to be very useful. Here images acquired from various wavelengths of light are combined into a single image, to create a composite image of a scene [51]. Arguably high dynamic range imaging, where multiple exposures of the same scene are combined, could also be classified as image fusion [31].

There have been many medical applications of fused images. Image fusion has been applied for brain imaging to combine CT and MR images, highlighting both the bone detail from the CT image and soft tissue detail from the MR image [32]. The fusion of MR and positron emission tomography (PET) images have been examined in the context of surgery planning for patients with epilepsy [50]. The use of fused images allowed for surgeries to be planned in a non-invasive manner, by highlighting the location of the epilepsy. MRI-SPECT (single photon emission computed tomography) fusion has also been studied for the purpose of detecting alzeihmers disease [9]. Here a preliminary study was undertaken to explore the potential of this technique. Applications of image fusion to digital subtraction angiography (DSA) has also been considered [55]. Multiple fluoroscopy images of arteries were acquired with and without a contrast agent. The non-contrasted images was subtracted from the contrasted images to create

DSA images at varying time points after contrast injection. Muiltiple DSA images were then fused to generate a fused contrast image over all time points. The fusion of CT images of the liver is another application that has been previously studied [34]. In this case the images were all from CT, but were taken at various points during an ablation of a liver tumor to highlight the progress of the procedure.

Basic methods of image fusion have been examined for the human spine. In this case MR and CT images are registered and then overlaid. These images have been used for assessing bone implants [23] and for surgery planning [21, 48]. The combination of these two modalities provides excellent bone details from the CT image and clear soft tissue details from the MR images.

1.6.2 Transform-based methods

The discrete wavelet transform (DWT) based image fusion method was proposed for fusing multifocal images [27]. In this method the two input images were first transformed into the wavelet image domain. Image fusion was accomplished in the transform domain by applying a set of simple rules, such as taking the maximum coefficient value, or the average coefficient value, depending on the detail band being fused. The output fused image could then be recovered by applying the inverse fusion transform. The wavelet transform works by decomposing an image into four subbands. These are the low detail band (a low resolution version of the original image) and three high detail bands in the vertical horizontal and diagonal directions. This is illustrated in Fig. 1.8.

One method that builds on the concept of the DWT is the contourlet transform (CLT) [12]. The 2D-DWT transform is not image oriented, as it only considers vertical, horizontal and diagonal edges. This may be sufficient at lower details scales, however when higher detail bands are considered, there are many edges that are not aligned parallel to any of these three directions. The contourlet transform seeks to address these drawbacks by considering an increasing number of directions at finer detail bands. This is accomplished in two steps. In first step, the

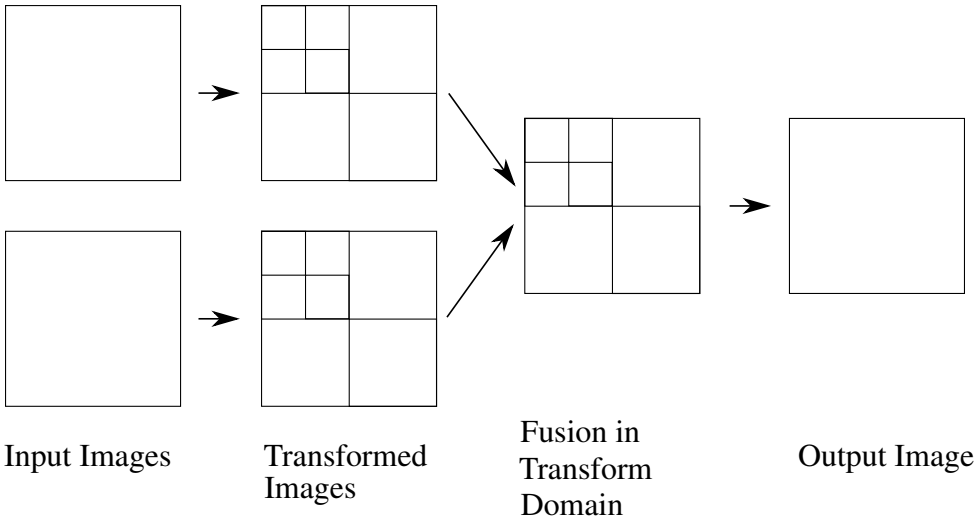


Figure 1.8: An illustration of the transform based fusion process, from left to right: the input images, the transformed input images showing various detailed bands, the image fused in the transform domain, the final output image.

image is decomposed into multiple resolutions. These levels are downscaled versions of the original image. In the second step a directional filter band is applied to all but the lowest scale. The highest detail version of the image, which is the original is decomposed into the most subdirections, with lower resolution versions being decomposed into fewer subdirections. This is demonstrated in Fig. 1.9. In this way the CLT considers a much greater amount of edge information than the DWT. For applications to image fusion, the input images are first transformed by the CLT. Fusion can then be carried out on the various detail and direction subbands of the CLT, creating a fused transformed image. The final fused image is then recovered by applying an inverse CLT transform [52].

Similar to the contourlet transform, the curvelet transform [39] also seeks to preserve edge structures in images. It has also been applied for image fusion. The advantage of the contourlet over the curvelet transform is that the contourlet transform has been directly formulated in the discrete domain, whereas the curvelet transform is formulated in the continous domain and needs to be sampled in order to be applicable to images. Other similar transformed based methods for image fusion include the additive wavelet decomposition [40], the complex wavelet transform [19, 26] and other wavelet approaches [42].

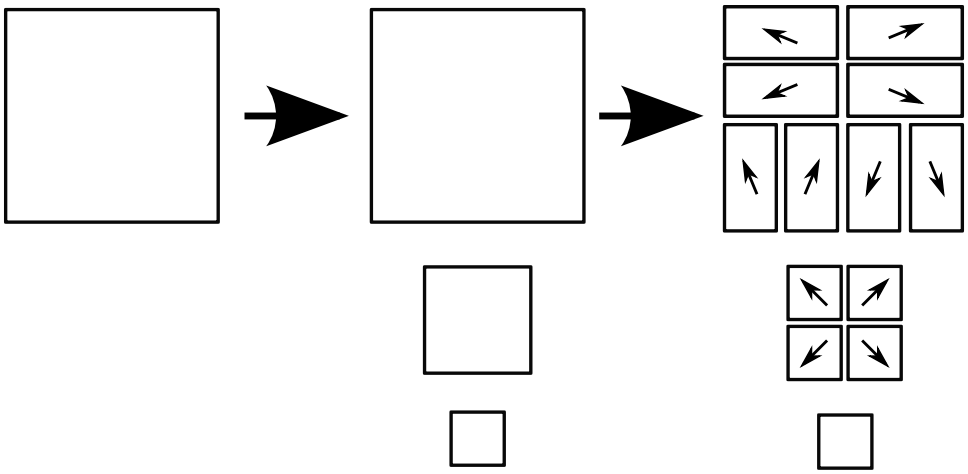


Figure 1.9: An illustration of the contourlet transform, showing the multiscale decomposition and the directional subbands. The arrows indicate approximate detail directions.

In all these methods decimations are included and there is a limit to the number of directions that can be encoded. This can introduce artifacts into the final image, especially when two dissimilar images are fused [36]. For medical purposes these artifacts have the potential to distort certain pathologies to the point where they may not be seen.

Another transform based fusion method includes the use of an independent component analysis basis [37]. This basis must be trained for the specific images in use and has the advantage of being problem specific; however its flexibility is reduced because of this. The polyharmonic local sine transform [28] has also been proposed; however, like all transform methods it is constrained by the quality of the transform.

1.6.3 Functional-optimization methods

Another set of approaches to solving the image fusion problem use functional optimization. Here the fusion problem is encoded as mathematical functional. Just like mathematical functions take a set of input variables and then output a variable, it is possible to design a set of functionals that take images as inputs and then output a value. We refer to this output value

as the energy of the functional. Typically the image that corresponds to the lowest possible energy of this functional is desired. Sometimes the maximum is desired. In these approaches the image fusion problem is stated as an energy functional consisting of a set of competing objectives that a fused image should minimize. Examples of possible objectives are similarity to the input images and smoothness of the final fused image. The energy functional is then optimized and the final solution is the fused image. These approaches have been considered in [2, 44, 49] offering the advantage of avoiding the artifacts that can be present due to the transform methods.

The largest challenge in these methods is how to solve the desired functional. Typically, gradient descent is employed as in [2, 44, 49]. Given a starting solution the gradient descent method moves in the direction of steepest descent to find a lower energy solution. This continues until a minimum energy is found. The difficulty with this method is that it often gets trapped in locally minimum locations, as opposed to the globally minimum solution. It is desirable to find a solution that will give the globally minimum energy for a problem as opposed to local. Additionally, gradient descent procedures are typically computationally intensive, which may preclude their use in the three or higher dimensions commonly used for medical images.

One possible solution to the problem of gradient descent is to formulate the functional as a graph-cut optimization problem [6, 7, 25]. This can then guarantee a near global solution to the fusion problem. A second approach is the use of convex optimization [3, 4]. Here the energy is formulated as a convex functional, which can be quickly and globally solved.

1.7 Background and Prior Art on Spine Labeling

Adjacent vertebrae have very similar appearances so it can be time consuming for radiologists to determine the identity of a particular vertebra in an image. When making a diagnosis of the spine radiologists must scroll through multiple images slices keeping track of the vertebral level in the current image. This vertebrae level can then be used as a reference point for describing

nearby pathology. Radiologists will often cross reference the location of pathology in an axial slice to a sagittal slice, where it may be more easily located.

Several approaches have been considered in order to automate the labeling of vertebrae in spine images [1, 16, 22, 24, 29, 30, 41, 46, 54]; however automatic labeling remains a challenging problem due to a high variability in the image intensities, shape and orientation of various spine structures [1, 16, 54]. These variations can be seen both within individual patients and between different patients. Two major approaches have been considered by the previous algorithms.

(1) Most of the current algorithms address the labeling problem through intensive training from a manually-labeled data set [1, 15, 16, 41, 46, 54]. Such a training stage aims at learning the shapes, textures and appearances of different spinal structures. This knowledge is then used within a classification or regression algorithm (e.g., support vector machine [41], random forest regression [15, 46] or graphical models [16, 24]) to subsequently label different spinal structures in the test image. Such algorithms work very well on data sets that closely match the training data, but would require adjustment/retraining for different data sets or if the imaging modality and/or acquisition protocol are altered (e.g., an algorithm that is trained and built for CT images may not perform well on MR data [16, 24, 29, 30, 54]). This might impede the use of these algorithms in routine clinical practices, where a particular disorder might be analyzed radiologically using several different imaging modalities/protocols with widely variable imaging parameters (resulting in extremely high variation in image data).

(2) To the best of our knowledge, all of the current spine labeling algorithms focus on either the sagittal view [1, 15, 16, 22, 24, 41, 46, 54] or are restricted to CT images [29, 30]. However, the quantification and level-based reporting of common inter-vertebral disc displacements such as protrusion, extrusion and bulging require the radiologist to thoroughly inspect all individual axial MR slices [13], while visually cross-referencing such axial slices to their corresponding position in the sagittal view. It is, however, important to note that in some cases only the axial view is available for the patient while in other cases the two scans might be acquired

at different time points. Therefore, localizing the spinal structures in different views becomes a challenging task (even for an experienced radiologist), which motivates a stand alone axial spine detection/labeling algorithm. Such a system would facilitate the generation of radiologic reports.

1.8 Graph Cut Optimization

Functional optimization methods have been considered for image fusion purposes [2, 44, 49]. These methods have potential applications to medical images. However as noted in section 1.6.3, a large drawback to these methods is the reliance on gradient descent. One possible solution to the problem of gradient descent is to formulate the functional as a graph-cut optimization problem [6, 7, 25]. Indeed the possibility of utilizing graph-cuts for image fusion is examined in chapter 2. In this section a background on graph-cut methods is presented as a precursor to their application for image fusion.

Discrete optimization methods, which use graph cut algorithms have recently sparked a substantial research effort in computer vision, and led to very efficient algorithms in image segmentation [8, 47], stereo vision [6] and image restoration [5]. These algorithms seek to find an optimal pixel labeling for an image. This is based on minimizing an overall data cost of assigning a given label to each pixel in the image, and a smoothness cost, that seeks to enforce labeling similarity in local neighbourhoods around each pixel. We can express this as

$$\min_{\mathcal{L}} E(\mathcal{L}) = D(\mathcal{L}) + S(\mathcal{L}) \tag{1.1}$$

where $\mathcal{L}$ is a set of discrete integer labels, E is the energy we seek to minimize, D represents the data costs and S represents the smoothness costs.

1.8.1 Binary Labeling Problem

First we consider the binary labeling problem [7]. Exactly one label $l_b \in [0, 1]$ is given to each pixel in the image, with data and smoothness costs being given to the links in the graph. This is shown in Fig. 1.10. We define a graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, containing a set of nodes $\mathcal{V}$ and set of weighted edges $\mathcal{E}$. There is one node in $\mathcal{G}$ for every pixel in the image domain $\Omega \in \mathcal{R}^2$. There is an edge connecting each pixel to every other pixel in a local neighbourhood $\mathcal{N}$ (typically a 4, 8 or 16 connected region), which is denoted as an n-link. Additionally there is a sink node and a source node. There is an edge between the sink node and every pixel in the image, and an edge between the source node and every pixel in the image, these are denoted as t-links. The t-links encode data costs associated to a given label (either 0 or 1) and the n-links encode smoothness between labels.

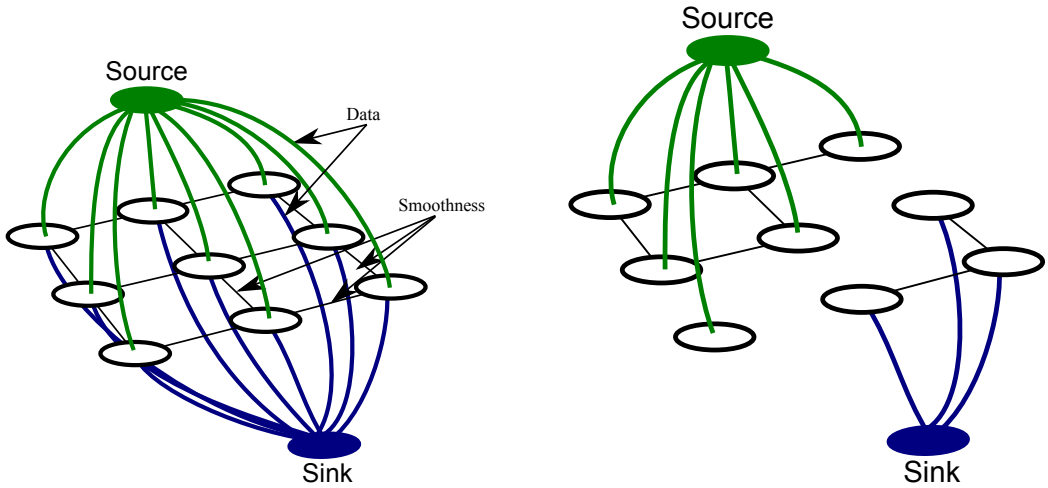


Figure 1.10: An illustration of the graph-cut problem: a) A binary graph showing the data cost of assigning a label to the sink/source and smoothness cost of assigning a labeling to adjacent pixel locations, b) the end result of the labeling of the graph.

A cut $C \subset \mathcal{E}$ is a set of edges that separates the sink node from the source node. The minimum-cut problem consists of finding the cut C with the lowest cost, denoted $|C|$. This cost is equal to the sum of all the edges in C and represent the minimum of $\mathcal{E}$. It can be solved globally and exactly using the algorithms of Boykov and Komologorov [7].

1.8.2 Multi-label Problem

An illustration of the multi-label graph cut problem is provided in Fig. 1.11. Exactly one label is given to each pixel in the image, with associated data and smoothness costs assigned to the links in the graph. Again there is a node for every pixel in the image, and edges connecting local neighbourhoods of pixels. Instead of a sink and source node, now there is a node for each label in the expanded set of labels $l_m \in [0, 1, \ldots, n]$. There are links between each label node and each pixel in the image. The edges between pixels are weighted based on the labels given to those pixels. In this case the minimum cut C is sought that creates a set of disjoint subgraphs with each pixel being assigned exactly one label.

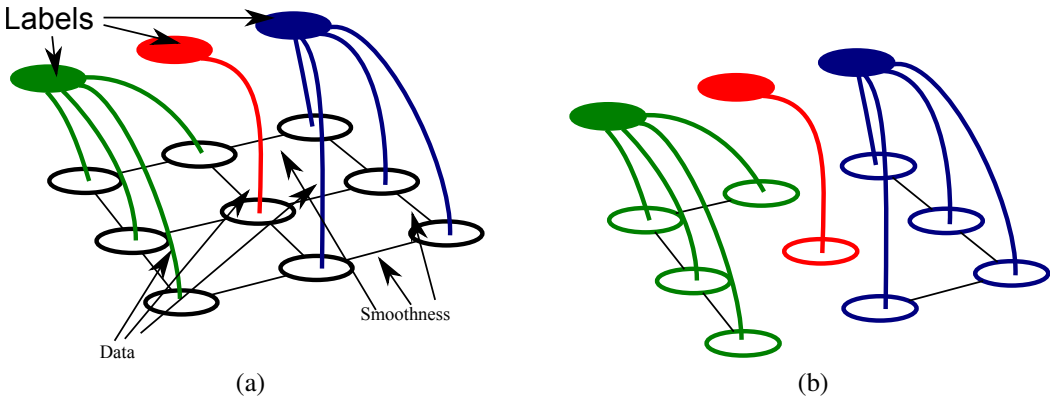


Figure 1.11: An illustration of the graph-cut problem: a) A graph with 3 possible labels showing the data cost of assigning a label to a node and smoothness cost of assigning a labeling to adjacent pixel locations, b) the end result of the labeling of the graph.

This problem can be solved in two ways: either by a series of alpha expansion moves or a series of swap moves [6]. In an alpha expansion moves, every pixel can either keep its original label or transfer to a new labeling l_α. This is iterated over every possible label, until the solution converges. The individual alpha expansion moves are solved globally using the binary graph cut algorithm. An alternative to this is the swap move. In this case, every pixel with a label l_α and l_β is considered. The pixels can either keep their original value, or swap values. Again this is iterated over the possible pixel combinations until convergence, and each individual swap move can be solved globally using the binary graph cut algorithm. Alpha expansions can be proven to give an answer within a given distance to the global minimum, while experimentally

swap moves have been shown to outperform alpha expansions [6].

1.9 Convex Optimization

The field of mathematical optimization deals with the problem of how to find the best possible solution to a given problem. Typically these problems are formulated as a form of energy functional $\min_{\mathbf{x}} f(\mathbf{x})$, where $\mathbf{x} = \mathbf{x_n},\ \mathbf{n} = \mathbf{1} \ldots \mathbf{N}$ is a vector valued variable and $f(\mathbf{x})$ is the functional to be minimized, which we call this the objective function. The goal of the optimization problem is to find the value $\mathbf{x}$ that corresponds to the smallest value of $f(\mathbf{x})$. Analogous to this is the energy maximization problem, where the goal is to *maximize* the objective function instead of *minimizing* it. If $\mathbf{x}$ is a continuous variable, then this is referred to as a continuous optimization problem and if $\mathbf{x}$ can only take on discrete values, we refer to the problem as a discrete optimization problem. Graph-cut optimization [7] is a special case of discrete optimization, since the pixel values can only take on discrete labels. We will focus on continuous optimization for the remainder of this section.

It is often valuable to be able to optimize $f(x_n)$ subject to some constraints on either f or x_n. These can take the form of inequality constraints e.g. $f(x_n) < 0$, $x_2 > 6$, $x_3 + x_4 < 5$ or equality constraints e.g. $f(x_n) = 0$, $x_2 = 7$, $x_3 + \sin x_4 = 2$. These problems are referred to as constrained optimization.

In general, optimizing $f(\mathbf{x})$ is a difficult problem, with or without constraints. There are often locally ”good” solutions that correspond to local minimas/maximas. These solutions may not be the best overall solution, which is more desirable and is referred to as the global optimum. One typical solution to these problems is the use of gradient descent. This is similar to walking down a mountain in fog. Given a starting point x^0, if you take a small step in the direction of steepest descent you should get a new point x^1, which represents a lower value of $f(\mathbf{x})$ or a lower *energy*. If you take enough steps, you should get to a point where you are no longer able to decrease the energy any further. The problem with this approach is that you

may be in a local optimum as opposed to a global optimum, and you have no way of knowing whether there is a better solution possible.

There are some problems that can be described by functions where one knows whether a solution is at a global optimum and that it can be reached easily. The global optimizability of these functions makes them advantageous to study. One common subset of globally optimizable functionals are linear functions. This consists of the set of all optimization problems where both the objective function and the constraints are linear in all variables. Efficient algorithms for solving these problems have been known since the 1950's [10]. However, this is not the only class of problems where a globally optimum solution can be guaranteed. Convex optimization problems guarantee a globally optimal solution and they are a superset of linear optimization problems.

Convex optimization deals with the optimization of functionals where the objective and inequality constraints are convex and any equality constraints must be linear. This takes the general form of:

$$\begin{aligned} \min_{x} \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0 \end{aligned} \tag{1.2}$$

where, $f_n(x)$ are convex functions, $i = 1, 2 \ldots n$.

A function is convex if and only if it satisfies Jensen's inequality [4] on every point on it's domain:

$$f(\theta x_1 + (1 - \theta x_2)) \leq \theta f(x_1) + (1 - \theta) f(x_2) \tag{1.3}$$

where $f(x), \mathcal{R} \rightarrow \text{dom}(f)$ is a convex function, $\theta \in [0, 1]$, and x_1, x_2 represent any two points on the domain of $f(x)$. Geometrically this means that any point x_s on a given chord between two points on the function is always greater than or equal to the function evaluated at that

point x_s as shown in Fig. 1.12. Simple examples include affine functions ($f(x) = ax + b$), the quadratic function ($f(x) = x^2$) and the absolute value function ($f(x) = |x|$). Convex functions of vector valued inputs can also be considered. Here the function is a vector valued function, and Jensen's inequality applies over every combination of two points on the function Convex functions play a significant role in numerical optimization because global optima can always be obtained and there are very efficient algorithms for finding those solutions [3, 4].

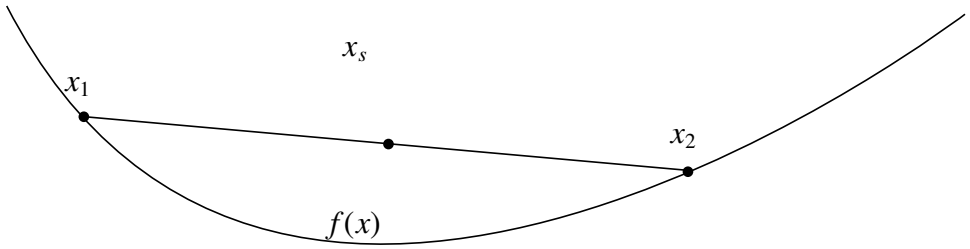


Figure 1.12: An illustration of Jensen's inequality. Any point on the chord between x_1 and x_2 is greater than the function at that point.

Image processing problems can be formulated as convex energy optimization problems. If we consider $\mathbf{x}$ to be a vector of all the image pixels, than $f(\mathbf{x})$ corresponds to optimizing over all the image pixels. This allows convex optimization techniques to be applied to images. In the context of image analysis, like graph cuts, convex optimization approaches have been the subject of recent research, [53]. Many of the same problems that can be solved via graph-cuts can also be optimized through an equivalent convex optimization, these include segmentation and stereo vision [53]. The most difficult aspect of their use is in finding an equivalent convex formulation to a given problem. When this can be found, very fast solutions can be obtained. Convex functions will be essential in the functional optimization methods used for image fusion presented in chapter 3.

1.10 Thesis objectives

The primary hypothesis of this work is that the use of fused images will enable more accurate and quicker diagnosis of spine images. Testing this will require the development of algorithms that are capable of real-time image fusion of MR and CT images of the spine. The methods should allow interactive adjustment of the fused images and be compatible with existing radiological work flows. Ideally these algorithms will be suitable for integration into existing radiology viewing software. To test the primary hypothesis the specific objectives are:

- Develop and validate an algorithm for off line registration of MR and CT images with minimal user input.

- Develop an image fusion method that accurately displays the clinically relevant details from both the CT and MR images in a single image.

- Develop a real-time image fusion method, which has similar accuracy to the above method, allowing for the interactive adjustment of fusion parameters on entire 3D volumes.

- Validate this system on clinical images, comparing the fused images to the current clinical practice. The images will be evaluated by radiologists based on suitability for clinical diagnosis.

An additional objective of this thesis is to develop an MR spine axial labeling algorithm for the correct annotation of MR axial images. To be clinical useful, this should have near real-time performance, while maintaining excellent accuracy. This should enable more accurate and quicker diagnosis of spine images.

1.11 Thesis Outline

1.11.1 Chapter 2 - Spine Image Fusion via Graph-Cuts, IEEE Transactions on BioMedical Engineering

This chapter details the design and validation of our image fusion method [36]. Here we present both a registration algorithm and a graph-cut based fusion algorithm. The registration method is a 3D rigid registration method, based on Mattes Mutual Information [33]. A 3D versor (unit quarternion) transform is used to perform this registration based on alignment of soft-tissue details in the MR and CT images. The accuracy is quantified and was found to be sub-voxel.

The fusion algorithm is based on the minimization of an energy functional via the use of graph cuts. The goal of this energy functional is to find a new image that is: 1) Similar to the input MR image, 2) Similar to the input CT image, and 3) contains natural transitions in the image. This approach offers the advantage of a near global solution avoiding the limits of transform based methods and gradient descent based energy optimizers.

The results of the fusion algorithm are validated based on the amounts of soft tissue and bone detail preserved in the fused image compared to the original image. These results are compared with 3 methods from the literature: 1) The discrete wavelet transform [27], 2). the contourlet transform [52] and 3). Piella's variational method [44]. In comparison with these works our methods are proven to have superior performance based on several quantitative evaluations/comparisons over 40 pairs of CT/MR images acquired from 20 patients. Additionally, four clinical cases studies are provided, highlighting the utility of our approach.

1.11.2 Chapter 3 - Multimodal Spine Fusion: A Convex Approach

In this chapter we detail a new fusion algorithm that provides significant advantages over the graph-cut (GC) based fusion algorithm of Chapter 2. The graph algorithm provides excellent image detail in the fused images; however, there are several drawbacks to this method. The GC method relies on approximations of the allowable gray level values of the image through the

introduction of multiple labels. Due to memory considerations it is also not extensible to 3D. Finally it is limited to off-line calculations due to the speed of the GC algorithm. Since GC is not a parallel problem, it is difficult to optimize it efficiently on modern computing hardware.

A convex formulation of the fusion energy function is proposed meeting the original criteria of: 1) being similar to the input MR image, 2) being similar to the input CT image and 3) containing natural transitions in the image. We examined two different criteria for preserving image similarity and present convex solutions to these two energy functionals.

We report comprehensive experiments and comparisons with images acquired from 30 subjects. We compared the results quantitatively to the original MR and CT images to determine the amount of clinically-relevant soft tissues and bone structures transferred to the fused images. Further comparisons with a recent graph cut based method show that the proposed methods yield competitive performances, while being applicable to 3D volumes. The results have less than a 10% average difference in pixel intensity from the target input images within the clinically relevant tissues. Our parallelized implementations on a GPU show that the proposed algorithms yield real-time solutions fusing 3D spine CT/MR volumes of typical sizes in 1-2s with an increase in speed in excess of 1000 times compared to the CPU. Case studies highlighting the utility of this approach are also included. These results indicate that image fusion may be a suitable technique for clinical use and are congruent with feedback we have received from radiologists based on the fused images; however, an observer performance study is now required to establish the utility of the technique.

1.11.3 Chapter 4 - Image Fusion for Spine Diagnosis

The real time fusion methods of chapter 3 allow for the creation of an interactive diagnosis tool. Being able to fuse entire volumes in less than a second allows for the adjustment of fusion parameters by radiologists. These parameters allow radiologists to have anatomy / pathology specific settings for viewing the fused images. This is similar, but complementary to window and level tools. Custom software has been designed for viewing fused images that has a similar

interface to existing radiology viewing software.

This software has been used to investigate the clinical performance our novel image fusion method, which was presented in chapter 3, for use in diagnosis of the lumbar spine. The fused images are designed to highlight bone detail from CT images and soft tissue detail from MR images. Institutional ethics review board approval was obtained for the use of data from 20 patients who had previously had imaging preformed, with the need for informed consent being waived. Images were obtained from consecutive patients who had an MR scan, and also had a previous CT scan within the past year. Five radiologists were employed to read the images. First unregistered MR and CT images were read. This was followed 4 weeks later by registered MR, CT and fused images. Comparisons were made based on confidence in diagnosis, observer agreement of protrusion measurements and time for completion of the previous two tasks.

Using the fused images radiologists had an overall higher confidence in diagnosis compared to CT, and the same or higher confidence compared to MR, except when assessing the exiting nerve root. There was not significant agreement between the readers for measurement of protrusions using either the fused or unregistered images. In general observers took the same amount of time to measure protrusions on fused and unregistered images. Overall, adding fused CT/MR images increases radiologist confidence when assessing CT or MR images of the lumbar spine, without an increase in time for diagnosis. As such, it would be desirable to include the use of fused images in clinical practice.

1.11.4 Chapter 5, Spine Axial MR Classification via Integral Kernel Density Estimates

Chapter 5 details a system for the automated annotation of MR axial spine images. Most anatomy in the spine is identified relative to its nearest vertebra. As a result knowing the correct labeling for the vertebra in the spine is essential for accurate diagnosis and treatment of spinal disorders. This can be a time consuming process and having an automated system would remove the need for crossing-referencing axial locations to sagittal images or mentally

tracking the current vertebral level.

The algorithm is based on a non-linear probability product kernel (PPK) classifier combined with a series of geometric constraints on the vertebrae/discs. Pixels are classified using the PPK classifier. From this, slice level features are generated, which are in turn used for single vertebrae classification. Multiple vertebrae are classified iteratively based on this method. Our classifier requires evaluations of computationally expensive integrals at each pixel of the image domain, and direct evaluations of such integrals would be prohibitively time consuming. We utilize an efficient computation of kernel density estimates and PPK evaluations for large images and arbitrary local window sizes via integral kernels, an extension of the well-known integral image method of Viola and Jones. Our method requires a single user click, runs in near real-time, and does not require an intensive external training. Results are presented using both T1-weighted and T2-weighted MR axial lumbar spine images from 32 patients and T2-weighted MR axial cervical spine images from 24 patients. A 99% overall classification accuracy is demonstrated for vertebrae in the lumbar spine, and 79% accuracy for the cervical spine.

1.11.5 Chapter 6 Summary, Conclusions and Future Work

The details of chapters 2 through 5 are briefly reviewed and conclusions as to the applicability of the fusion and labeling systems are presented. Additionally we explore future work related to image fusion, including: real-time approaches to image registration that utilize the output of the labeling algorithm, extensions of the fusion work and integration into a clinical product.

Bibliography

[1] Raja' S. Alomari, Jason J. Corso, and Vipin Chaudhary. Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model. *IEEE Trans. Med. Imaging*, 30(1):1–10, 2011.

[2] Coloma Ballester, Vicent Caselles, Laura Igual, and Joan Verdera. A Variational Model for P+XS Image Fusion. *International Journal of Computer Vision*, 69(1):4358, 2006.

[3] Dimitri P. Bertsekas. *Constrained optimiaztion and Lagrange multiplier methods*. Academic Press Inc, 1982.

[4] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2009.

[5] Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[6] Yuri Boykov and Vladimir Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 26(9):1124–1137, September 2004.

[7] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 23(11):1222–1239, November 2001.

[8] Yuri Boykov, Vladimir Kolmogorov, D Cremers, and Andrew Delong. An integral solution to surface evolution PDEs via geo-cut. In *Proceedings of European Conference on Computer Vision*, volume 3, pages 409–422, 2006.

[9] Arnaud Colin and Jean-Yves Boire. MRISPECT fusion for the synthesis of high resolution 3D functional brain images: a preliminary study. *Computer Methods and Programs in Biomedicine*, 60:107–116, 1999.

[10] George B. Dantzig, Alex Orden, and Philip Wolfe. The generalized simplex method for minimizing a linear form under linear inequality restraints. *Pacific Journal of Mathematics*, 5(2):183–195, 1955.

[11] Richard A. Deyo, Sohail K. Mirza, Judith A. Turner, and Brook I. Martin. Overtreating Chronic Back Pain: Time to Back Off? *Jour Am Board Fam Med*, 22(1):62–68, 2009.

[12] Minh N. Do and Martin Vetterli. The Contourlet Transform: An Efficient Directional Multiresolution Image Representation. *IEEE TRANSACTIONS ON IMAGE PROCESSING,*, 14(12):2091–2106, December 2005.

[13] D. F. Fardon and P. C. Milette. Nomenclature and classification of lumbar disc pathology: Recommendations of the combined task forces of the north american spine society, american society of spine radiology, and american society of neuroradiology. *Spine*, 26 (5):E93–E113, 2001.

[14] David F. Fardon and Pierre C. Milette. Nomenclature and classification of lumbar disc pathology recommendations of the combined task forces of the north american spine society, american society of spine radiology, and american society of neuroradiology. *Spine*, 26(5):E93–E113, 2001.

[15] B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, and A. Criminisi. Vertebrae localization in pathological spine ct via dense classi
cation from sparse annotations. In *MICCAI: LCNS Spinger (Heidelburg)*, 2013.

[16] Ben Glocker, Johannes Feulner, Antonio Criminisi, David R. Haynor, and Ender Konukoglu. Automatic localization and identification of vertebrae in arbitrary field-of-view ct scans. In *MICCAI (3)*, pages 590–598, 2012.

[17] Lee W Goldman. Principles of ct: Multislice ct. *Journal of Nuclear Medicine Technology*, 36(2):57–68, 2008.

[18] How-Ran Guo, Shiro Tanaka, William E. Halperin, and Lorraine L. Cameron. Back pain prevalence in us industry and estimates of lost workdays. *American Journal of Public Health*, 89:1029–1035, 1999.

[19] Paul Hill, Nishan Canagarajah, and David R Bull. Image Fusion Using Complex Wavelets. In *BMVC 2002*, 2002.

[20] Godfrey N. Hounsfield. Computed medical imaging. *Medical Physics*, 7(4):283–290, July 1980.

[21] Yangqiu Hua, Sohail K. Mirzac, Jeffrey G. Jarvikb, Patrick J. Heagerty, and David R. Haynor. MR and CT image fusion of the cervical spine: a noninvasive alternative to CT-Myelography. In *Proceedings of SPIE*, volume 5744, Bellingham, WA, 2005. SPIE.

[22] S.H. Huang, Y.H. Chu, S.H. Lai, and C.L. Novak. Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. *IEEE Trans. on Medical Imaging*, 28(10):1595–1605, 2009.

[23] C. A. Karlo, I. Steurer-Dober, M. Leonardi, C. W. A. Pfirrmann, M. Zanetti, and J. Hodler. MR/CT image fusion of the spine after spondylodesis: a feasibility study. *European Spine Journal*, 19:17711775, 2010.

[24] Tobias Klinder, Jrn Ostermann, Matthias Ehm, Astrid Franz, Reinhard Kneser, and Cristian Lorenz. Automated model-based vertebra detection, identification, and segmentation in ct images. *Medical Image Analysis*, 13(3):471–482, 2009.

[25] Vladimir Kolmogorov and Ramin Zabih. What Energy Functions Can Be Minimized via Graph Cuts? *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 26(2):147–159, February 2004.

[26] John J Lewis, Robert J. O'Callaghan, Stavri G. Nikolov, David R Bull, and NIshan Canagarajah. Pixel - and region-based image fusion with complex wavelets. *Information Fusion*, 8:119–130, 2007.

[27] H. Li, B. S. Manjunath, and S. K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, 1995.

[28] Shangzheng Liu, Jiuqiang Han, Bowen Liu, and Xinman Zhang. Novel image fusion algorithm based on human perception. *Optical Engineering*, 48(4):047002 1–6, April 2009.

[29] Jun Ma and Le Lu. Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. *Computer Vision and Image Understanding*, 117 (2013):10721083, 2013.

[30] Jun Ma, Le Lu, Yiqiang Zhan, Xiang Sean Zhou, Marcos Salganicoff, and Arun Krishnan. Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. In *MICCAI (1)*, pages 19–27, 2010.

[31] Steve Mann and Rosalind Picard. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T*, pages 442–448, 1995.

[32] G. K. Matsopoulos, S. Marshall, and J. N. H. Brunt. Multiresolution morphological fusion of MR and CT images of the human brain. *IEEE Proceedings on Vision Image Signal Processing*, 141(3):137–142, 1994.

[33] David Mattes, David R. Haynor, Hubert Vesselle, Thomas K. Lewellen, and William Eubank. PET-CT Image Registration in the Chest Using Free-form Deformations. *IEEE Transactions on Medical Imaging*, 22(1):120–128, January 2003.

[34] Pamela M. Mazurek, Anne M. Silas, and John M. Gemery. Summation of CT Scans During Radiofrequency Ablation for Assessing Target Lesion Coverage. *American Journal of Roentgenology*, 191:790:792, September 2008.

[35] Robin McGee, Stephen Bevan, and Tatiana Quadrello. Fit for work? musculoskeletal disorders and the canadian labour market. The Conference Board of Canada, July 2011.

[36] B. Miles, I. Ben Ayed, M. W. K. Law G. Garvin, A. Fenster, and S. Li. Spine image fusion via graph cuts. *IEEE Transactions On Biomedical Engineering*, 60(5):1841–50, July 2013. 2013 IEEE. Reprinted, with permission, from Brandon Miles, Ismail Ben-Ayed, Max W.K. Law, Gregory J. Garvin Aaron Fenster and Shuo Li, Spine Image Fusion via Graph-Cuts, IEEE Transactions on BioMedical Engineering, July 2013.

[37] Nikolaos Mitianoudis and Tania Stathaki. Pixel-based and region-based image fusion schemes using ICA bases. *Information Fusion*, 8(2):131142, 2007.

[38] Keith L. Moore and Anne M. R. Agur. *Essential Clinical Anatomy*. Lippincott Williams & Wilkins, third edition edition, 2007.

[39] Filippo Nencini, Andrea Garzelli, Stefano Baronti, and Luciano Alparone. Remote sensing image fusion using the curvelet trasnform. *Information Fusion*, 8:143–156, 2007.

[40] Jorge Núñez, Xavier Otazu, Octavi Fors, Albert Prades, Vicenç Palà, and Román Arbiol. Multiresolution-Based Image Fusion with Additive Wavelet Decomposition. *IEEE Transactions On Geoscience And Remote Sensing*, 37(3):1204–1211, May 1999.

[41] Ayse Betül Oktay and Yusuf Sinan Akgul. Localization of the lumbar discs using machine learning and exact probabilistic inference. In *MICCAI (3)*, pages 158–165, 2011.

[42] Xavier Otazu, Mara Gonzlez-Audcana, Octavi Fors, and Jorge Nez. Introduction of Sensor Spectral Response Into Image Fusion Methods. Application to Wavelet-Based Methods. *IEEE Transactions On Geoscience And Remote Sensing*, 43(10):2376–2385, October 2005.

[43] Pamela L. Owens, Maeve Woeltje, and Ryan Mutter. Emergency Department Visits and Inpatient Stays Related to Back Problems, 2008. URL `http://www.hcup-us.ahrq.gov/reports/statbriefs/sb105.pdf`.

[44] Gemma Piella. Image Fusion for Enhanced Visualization: A Variational Approach. *International Journal of Computer Vision*, 83(1):1–11, 2009.

[45] R. A. Pooley. Aapm/rsna physics tutorial for residents: fundamental physics of mr imaging. *Radiographics*, 25(4):1087–99, 2005.

[46] Martin G. Roberts, Timothy F. Cootes, and Judith E. Adams. Automatic location of vertebrae on dxa images using random forest regression. In *MICCAI (3)*, pages 361–368, 2012.

[47] Mohammed Ben Salah, Amar Mitiche, and Ismail Ben Ayed. Multiregion Image Segmentation by Parametric Kernel Graph Cuts. *IEEE Transactions on Image Processing*, 20(2):545–557, 2011.

[48] Moon-Jun Sohn, Dong-Joon Lee, Sang Won Yoon, Hye Ran Lee, and Yoon Joon Hwang. The effective application of segmental image fusion in spinal radiosurgery for improved targeting of spinal tumours. *Acta Neurochir 151:231238*, 151:231–238, 2009.

[49] Wei-Wei Wang, Peng-Lang Shui, and Xiang-Chu Feng. Variational Models for Fusion and Denoising of Multifocus Images. *IEEE Signal Processing Letters*, 15:65–68, 2008.

[50] Stephen T. C. Wong, Robert C. Knowlton, Randy A. Hawkins, and Kenneth D. Laxer. Multimodal Image Fusion for Noninvasive Epilepsy Surgery Planning,. *IEEE Transactions on Computer Graphics and Applications*, 16(1):30–38, 1996.

[51] Xiangnan Xu, Hua Li, and A Wang. The application of BEMD to multi-spectral image fusion. In *Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition*, 2007.

[52] L. Yang, B.L.Guo, and W.Ni. Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform. *Neurocomputing*, 72:203–211, 2008.

[53] Jing Yuan, Juan Shi, and Xue-Cheng Tai. A Convex AND Exact Approach to Discrete Constrained TV-L1 Image Approximation. Technical Report CAM-10-51, UCLA, CAM, UCLA, 2010.

[54] Yiqiang Zhan, Maneesh Dewan, Martin Harder, and Xiang Sean Zhou. Robust mr spine detection using hierarchical learning and local articulated model. In *MICCAI (1)*, pages 141–148, 2012.

[55] Guangming Zhang, Zhiming Cui, Fanzhang Li, and Jian Wu. DSA Image Fusion based on Dynamic Fuzzy Logic and Curvelet Entropy. *Journal of Multimedia*, 4(3):129–136, June 2009.

Chapter 2

Spine Image Fusion via Graph Cuts [20]

2.1 Introduction

For spine diseases and injuries, it is common for a patient to receive both an MR and a CT scan because of their individual benefits. MR images depict useful soft-tissue details including the spinal discs, nerves, cerebral spinal fluid and spinal cord. Therefore, it is the primary modality to diagnose protruding and degenerated discs. CT images clearly depict bony structures, especially the bone cortex, allowing the assessment of damaged joints or osteophyte growth (bony spurs at the margins of a joint).

Radiologists currently display MR and CT images side by side, when both images are available. This does provide them with all the available image information, but its accessibility is limited to visual correlation between the two images. It can be difficult to determine whether narrowing of a spinal canal is caused by tissue or bone from clinical MR images hence, both CT and MR can be employed [11]. Using both CT and MR images, as opposed to relying on a single modality can benefit diagnosis and treatment of osteophytes and degenerate discs that impact bone and nerve structures. In addition both modalities can aid post operative follow up after spinal surgery [12]. Here, both the CT and MR modalities provide complementary information. In order to properly visualize the related bone and soft tissue structures, the images

must be mentally aligned and fused together. Detecting changes on unregistered, uncombined images is an error prone task [26] . Therefore, it is highly desirable to fuse these two modalities into a single image showing the clinically significant CT and MR details as well as their relative locations on a single image. This will remove the need for mental juxtaposition when examining multiple views. Our goal is to provide and validate such a system.

Here we present a novel method for image fusion of the spine, which preserves the bone structures and soft tissue detail in a single image. Spine image fusion has the potential to enable more effective and efficient evaluations of spine disorders, more so as the number of spine scans increases very rapidly. For instance, in the United States, there has been a 300% increase in lower spine MR scans in the period between 1994 and 2006 [8], and the percentage of adults who have suffered from back pain is 75% [24].

Multi-modality image fusion has been studied in other fields, with applications varying from multifocal [16] to geographical images [23]. In medicine, image fusion has been used for brain imaging [17, 35], MRI-SPECT fusion [7], epilepsy treatment planning [34], liver ablation [19] and digital subtraction angiography [38]. For the spine, registered and overlaid CT and MR spine images have been used for surgery planning [11, 29] and evaluation of bone implants [12]. A wavelet based approach to image fusion has been proposed by Li et al. [16]. The two input images were fused in the wavelet domain, and an inverse transformation was applied to produce the result. Other variations of this technique include additive wavelet decomposition [23], the contourlet transform [35, 36], the curvelet transform [22] and the complex wavelet transform [10, 15]. The wavelet or transform based methods can suffer from pixelation artifacts when two dissimilar images are fused. This is a result of the decimations involved in the wavelet transform as well as the translation dependence of standard wavelets. This may result in small details being distorted, leading to a significant loss in image quality. Pathologies that should have been visible may no longer be seen.

Variational fusion methods have also been investigated [2, 25, 30, 31, 39]. These methods consist of finding the optimum of an energy functional, often via standard continuous optimiza-

tion techniques, e.g., gradient descent. For these variational approaches, the main difficulties come from the limitations of the optimizers. Gradient-descent procedures [2, 25, 30, 31, 39] yield sub-optimal solutions and have a very high computational load.

Discrete optimization methods, which use graph cut algorithms have recently sparked a substantial research effort in computer vision, and led to very efficient algorithms in image segmentation [6, 27], stereo vision [4] and image restoration [3]. For object recognition, graph cuts have been employed to segment planar surfaces from depth images [13]. Both depth and intensity values were combined as inputs to a graph cut based segmentation algorithm. Graph cuts have also been used to determine the optimal fusion rules for combining subbands of a beamlet transform [37]. Although this method utilizes a graph cut approach for combining subbands, it is still a transform based method and therefore is limited by the transformation performed. To the best of our knowledge, graph cut formulations have not been previously studied in the context of variational approaches to image fusion.

In this study, we state image fusion as a discrete multi-label optimization problem, which can be solved efficiently with graph cuts [4, 5, 14], via the well-known swap or alpha-expansion moves [4]. The proposed energy function [21] balances the contributions of three competing terms: (1) a squared error, which encourages the solution to be similar to the MR input, with preference to strong MR edges; (2) a squared error, which encourages the solution to be similar to the CT input, with preference to strong CT edges; and (3) a prior, favoring smooth solutions by encouraging neighboring pixels to have similar fused-image values. We further introduce a transparency-labeling formulation, which significantly reduces the computational load. The proposed graph-cut fusion guarantees nearly global solutions, while avoiding the pixelation artifacts that affect standard wavelet based methods. We report several quantitative evaluations/comparisons over 40 pairs of CT/MR images acquired from 20 patients. The results demonstrate very competitive performance in comparisons to existing variational and transform-based methods [25, 35, 16].

This work is a significant extension of a preliminary conference version [21]. The dataset

was increased from 9 to 20 patients, resulting in validation on 40 image slices. A comparison to Piella's variational method and validation results based on the structural similarity information measure (SSIM) have also been added. Two more clinical case studies have also been included. In section 2.2 we present our formulation of image fusion as a graph cut labeling problem. This is followed by a description of the dataset, its registration, preprocessing and fusion in section 2.3. We further discuss various case studies, and give a representative sample of the results in 2.4 with a discussion following in 2.5.

2.2 Formulation

2.2.1 Multi-label Formulation

We state image fusion as the following multi-label optimization problem:

$$\lambda^* = \min E(\lambda) \text{ with } E(\lambda) = \mathcal{D}(\lambda) + c_1\mathcal{R}(\lambda) \tag{2.1}$$

where:

- Variable λ is a labeling function that assigns each point in image domain Ω to a label l, which describes the intensity of the fused image at that point:

$$\lambda : p \in \Omega \rightarrow \lambda(p) \in \mathcal{L}, \tag{2.2}$$

with $\mathcal{L} \subset \mathcal{I}$ denoting a closed finite set of integers (the possible output intensities).

- Data term $\mathcal{D}$ is defined as:

$$\begin{aligned}\mathcal{D}(\lambda) &= \sum_{p\in\Omega} \mathcal{D}_p\left(\lambda(p)\right) \\ &= \sum_{l\in\mathcal{L}} \sum_{p\in R_l} \left[w_1\left(l - u_1(p)\right)^2 + w_2\left(l - u_2(p)\right)^2\right]\end{aligned} \tag{2.3}$$

where $u_1 : \Omega \to \mathbb{R}$ and $u_2 : \Omega \to \mathbb{R}$ denote the input images, and R_l is the l-label region defined by $\{p \in \Omega | \lambda(p) = l\}$. w_1 and w_2 are weights defined as follows:

$$\begin{aligned} s_1 &= |\nabla u_1| * K & s_2 &= |\nabla u_2| * K \\ w_1 &= \frac{s_1}{s_1 + s_2} & w_2 &= \frac{s_2}{s_1 + s_2} \end{aligned} \tag{2.4}$$

K is a kernel, for instance, a box filter. w_1 and w_2 bias the solution towards strong edges in u_1 and u_2, respectively.

The data term balances the contributions of two competing terms:

1. A squared error which, encourages the solution to be similar to the first input u_1, with preference to strong edges in u_1; and

2. A squared error which, encourages the solution to be similar to the second input u_2, with preference to strong edges in u_2.

- Smoothness term $\mathcal{R}$ favors smooth solutions by encouraging neighboring pixels to have similar fused-image values:

$$\mathcal{R}(\lambda) = \sum_{\{p,q\}\in\mathcal{N}} r\left(\lambda(p), \lambda(q)\right) \tag{2.5}$$

with $\mathcal{N}$ being a set containing all pairs of pixels p and q in a local neighborhood of p and

$r(\lambda(p), \lambda(q))$ is defined by the truncated absolute value:

$$r(\lambda(p), \lambda(q)) = \min\left(c_2, \left|l_p - l_q\right|\right) \tag{2.6}$$

with c_2 being a positive constant.

2.2.2 Alpha-blending Reformulation

The above formulation requires a one-to-one correspondence between the labels and pixel intensities. Therefore, the number of labels needed to express the output image is equal to the number of all possible pixel values. This may lead to a high computational load in the case of images with large dynamic ranges, as is common in medical imaging. To reduce the number of labels, we reformulate the data term as a transparency labeling. This is done by expressing the output image as a function of u_1 and u_2 via a transparency image α, with $\alpha(p) \in [0, 1] \; \forall \; p \in \Omega$:

$$u_\alpha = \alpha u_1 + (1 - \alpha) u_2 \tag{2.7}$$

where u_α denotes the output fused image.

Based on this formulation, we rewrite the data term in (2.3) as follows:

$$\begin{aligned} \mathcal{D}(\lambda) &= \sum_{p \in \Omega} D_p(\lambda(p)) \\ &= \sum_{l \in \mathcal{L}_\alpha} \sum_{p \in R_l} \Big[w_1 \left(u_\alpha(p, l) - u_1(p)\right)^2 + \\ &\qquad + w_2 \left(u_\alpha(p, l) - u_2(p)\right)^2 \Big] \end{aligned} \tag{2.8}$$

where

$$u_\alpha(p,l) = \frac{l}{N_l}u_1(p) + \left(1 - \frac{l}{N_l}\right)u_2(p) \qquad l \in \mathcal{L}_\alpha \tag{2.9}$$

with $\mathcal{L}_\alpha$ being a new (reduced) set of non-negative integer labels $\{0, 1, 2, ..., N_l\}$, parameterized by the user specified number of labels N_l.

2.2.3 Graph Cut Optimization

Our problem is amenable to efficient graph cut optimization [5, 4, 14]. An illustration of the multi-label graph cut problem is provided in Fig. 2.1. Exactly one label is given to each pixel in the image, with associated data and smoothness costs assigned to the links in the graph. To formulate this optimization let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a weighted graph, with $\mathcal{V}$ a set of nodes and $\mathcal{E}$ a set of weighted edges. $\mathcal{V}$ contains a node for each pixel in Ω and for each label in $\mathcal{L}_\alpha$. There is an edge $e_{\{p,q\}}$ between every pair of nodes p, q. A cut $C \subset \mathcal{E}$ is a set of edges that separates all the label nodes from each other, thereby creating a sub-graph for each label. The minimum-cut problem consists of finding a cut C with the lowest cost. The cost of this minimum cut, denoted $|C|$, equals the sum of the edge weights in C. By properly setting the weights of the graph, one can use a series of swap moves from combinatorial optimization [5] to efficiently compute the minimum-cost cuts corresponding to a minimum of functional E.

A swap move starts with a labeled graph and determines for a given pair of labels, p and q, whether each node having a value in p, q should (1) retain its current label or (2) be updated to the other label in the pair. Each swap is accomplished globally in an exact manner by finding the minimum cut on a binary graph consisting of only two labels. This can be extended to the multi-label case by iterating over the set of all possible pairs of labels. The minimum cut is selected at each stage, with the final labeling corresponding to a minimum of the energy functional. One can also use alpha-expansion moves [4] to optimize energy functions of the form E. It is well-known that alpha-expansion moves guarantee a solution that is within a

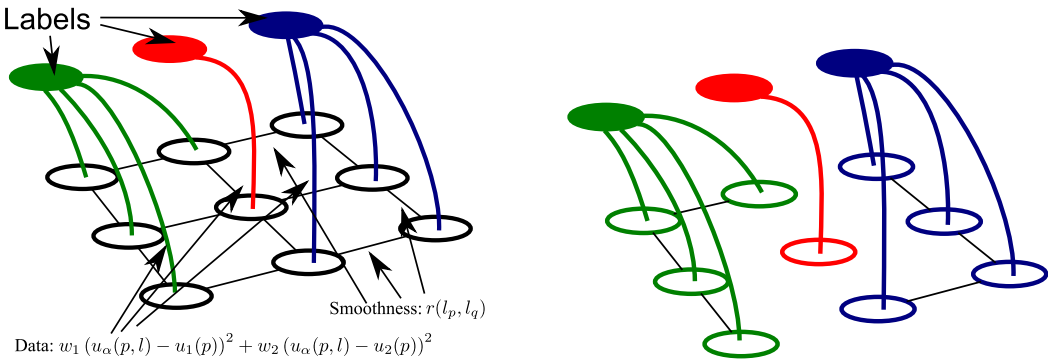


Figure 2.1: An illustration of the graph-cut problem: a) A graph with 3 possible labels showing the data cost of assigning a label to a node and smoothness cost of assigning a labeling to adjacent pixel locations, b) the end result of the labeling of the graph.

constant factor of the global optimum [4]. However, experimentally, it is well established that swap moves outperform alpha expansions [4]. Therefore, in this work, we used swap moves with the edge weights defined in Table 2.1; where $e_{\{l,p\}}$ denotes an edge between a label and a pixel, and $e_{\{p,q\}}$ an edge between two adjacent pixels.

Table 2.1: Weights assigned to the edges of the graph for minimizing the proposed fusion energy

edge	weight	for
$\{l, p\}$	$w_1 (u_\alpha(p,l) - u_1)^2 + w_2 (u_\alpha(p,l) - u_2)^2$	$p \in \Omega, l \in \mathcal{L}_\alpha$
$\{p, q\}$	$r(l_p, l_q)$	$p, q \in \Omega$

2.3 Methods

This retrospective study was approved by the Human Subjects Ethics Board of Western University, with the requirement for informed consent being waived. Twenty patient image sets were randomly selected with the criteria of patients having had both a lumbar MR and CT scan within a one-year time period. None of these patients had fractures, but other diseases such as degenerate / protruding discs, spinal stenosis and osteophytes were present. The images were acquired using either a Magnatom or Avanto Siemens 1.5T MR scanner (Seimens AG,

Erlangen, Germany), with varying CT scanners depending on the location the CT images were obtained. The MR scans were acquired using a 3D T2 weighted pulse sequence, and the CT scans were acquired from either helical or axial slice CT images. No contrast was used in either scan. The lumbar spine was assumed to be rigid between scans, because the patients were scanned in a feet first prone position, resulting in very similar postures. We evaluated the proposed method over 40 pairs of CT/MR images acquired from these 20 patients. Twenty pairs were from the center sagittal slice, and 20 were from the left side of the patient through the nerve root bundle. T2-weighted 3D MR images were used because they clearly present the discs, nerve root bundle and cerebral spinal fluid. The 3D MR/CT images were then registered and preprocessed. Finally the images were fused as sets of 2D images because radiologists typically view 3D volumes as stacks of 2D images.

Validation was completed based first on visual results of the fusion, studying clarity of the detail presented in the fused image, and second via a statistical comparison of the clinically significant bone and tissue transferred to the fused images. Four clinical case studies were then examined to illustrate the potential clinical value of this technique. Our method was compared to four methods: (A) an averaging of the two images, and three methods from recent literature: (B) the discrete wavelet transform (DWT) [16] (C) the contourlet transform (CLT) [35] and (D) Piella's variational method [25]. These methods were implemented using the parameters listed in their papers. Table 2.2 contains a summary of these features.

2.3.1 Registration and Pre-Processing

The input volumes were registered, using a rigid 3D versor based transform in ITK[1]. The optimizer used maximization of Mutual Information (MI) [18] to align soft tissue details present in both images (note the soft tissue details in the CT image are suitable for registration, but MR is better for diagnosis). For the purpose of aligning soft tissue each CT image was thresholded from -255 to 255 Houndsfield Units (HU) or -255 to 0 HU if needed. This kept many of the soft tissue details, but removed most of the bone detail. Both images were then scaled to an

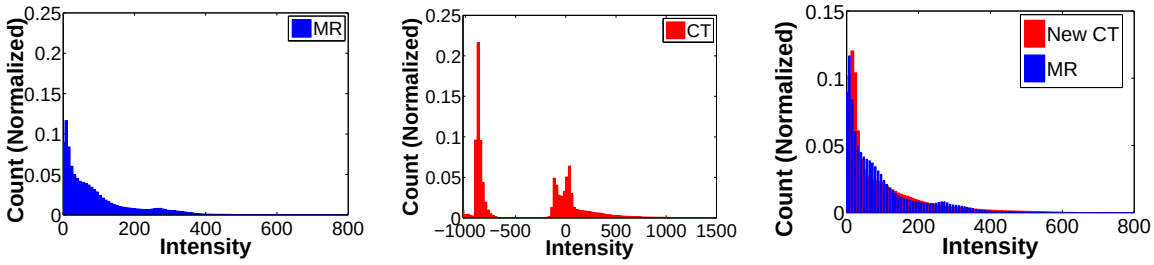
Table 2.2: Implementation details and parameters for the three methods from literature

Discreet Wavelet [16]	Contourlet Transform [35]	Piella Variational [25]
• **Subbands:** 3 subbands • **Wavelet:** Haar wavelet • **Lowpass Rule:** pixelwise averaging • **Highpass Rule:** pixelwise selection of the coefficient with the largest maximum value	• **Subbands:** 4 subbands, with 4,8,16 directional subbands in levels 2,3,4 (lowest detail to highest) • **Filters:** Lowpass 9-7 Filter, directional PKVA • **Lowpass Rule:** local energy in a 3x3 window • **Highpass Rule:** local contourlet contrast	• **Parameters** $\eta = 0.1$, $\beta = 0.5$, $\gamma = 0.3$, $\delta t = 0.15$ • **Kernel** w - Gaussian, $\sigma = 0.1$ • **Polynomial** J' $n = 7$, $\alpha = 10$, $k = 0.25$

intensity range of 0 to 255 to be in the same range. The transform was initialized using two corresponding user-selected points, one from the CT and the other from the MR image. After this, MI was calculated from the voxels in both images, and the versor transform was iteratively updated based on MI of the two images at each step. Using the obtained optimal transform, the original MR image (without intensity scaling) was transformed and resampled to the voxel spacing of the CT image.

Manual points were selected in the 3D images for the target registration error (TRE) and the fiducial localization error (FLE) evaluation. The TRE is the mean post registration Euclidean distance between corresponding pairs of fiducials from the input images. The FLE is the root mean-squared difference in locations when selecting the same fiducial multiple times in an image [9]. The TRE used 17 points from two image pairs. For the FLE, 5 distinct points were defined on the CT image. On 5 separate days, corresponding points in the MR image were identified producing a total of 25 point sets. These errors were used to validate the registration.

After registration, the original CT images were thresholded at 0 HU, setting any negative values to 0 HU and leaving other values unchanged. This removed most of the soft-tissue



(a) (b) (c)

Figure 2.2: Histograms of all the images, a) Histogram of all 40 MR images, b) Histogram of all 40 CT images, c) Histogram of all 40 CT images after Preprocessing (excluding pixels with an intensity of 0) and histogram of all 40 MR images.

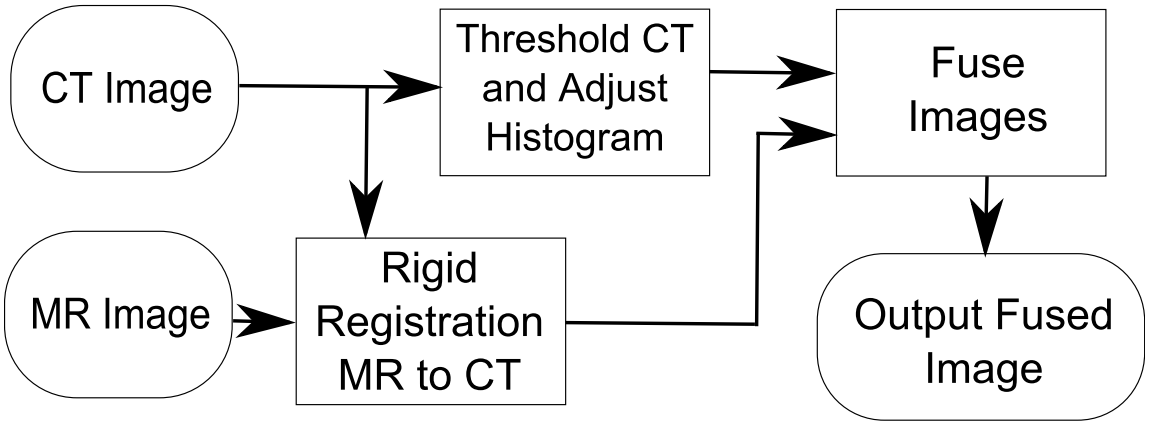


Figure 2.3: Flow chart of the image fusion process: The MR is registered to the CT image. The CT is thresholded and its histogram is adjusted to match the MR. Finally the images are fused.

details and was done because the MR presents the tissue detail with more clarity, so the CT tissue detail is undesirable for the fused image. For all 20 patients the MR images were found to have a maximum intensity of about 700, and the CT images were found to have a maximum intensity of about 1400. In view of this, the CT intensities were divided by two so that the MR and CT histograms would have similar intensity ranges prior to fusion. This was needed to eliminate bias in the fusion algorithm due to differing intensity ranges in the input images. MR and CT histograms are shown in Fig. 2.2 and Fig. 2.3 shows a flow chart describing the registration and pre-processing steps.

2.3.2 The Parameters of our Fusion Method

After preprocessing was completed, fusion could be carried out. For the purpose of these experiments, c_1 was set equal to $\frac{0.001(I_{\max})^2}{c_2}$, c_2 equal to $0.40\,(N_l)$ and N_l equal to 20, with $I_{\max}$ being the maximum intensity value in both inputs. N_l and c_1 were tuned manually to balance image quality and speed of computation. c_2, was set empirically for smoothness. Constant c_1 was set in relation to the maximum value in the data term and in relation to c_2.

2.3.3 Statistical Evaluation of the Fused Images

We compared each of the MR and CT images to the fused result within: (1) the regions of soft tissues, and (2) the regions of bone structures. The soft-tissue details consisted of the discs, nerves, and cerebral spinal fluid from the MR image, and the bone details were from the CT image, with a specific focus on the bone cortex. The trabecular bone does not contact soft tissue and so was omitted. We created image masks of the tissue and bone details for each patient. The tissue masks were created by manual segmentations of the MR images, and the bone masks were obtained by thresholding the CT images at a user selected HU for each image and then manually correcting any errors. Figure 2.4 shows sample masks of the tissue and bone detail.

We defined a fusion error as the mean absolute-value difference between the MR/CT images and the fused images in the tissue regions defined by the masks. For the MR images, we calculated the following two errors:

$$e_{\text{MR, Tissue}} = \frac{\sum_{M_{\text{Tissue}}} |I_{\text{MR}} - I_{\text{fused}}|}{\text{area of the tissue mask}} \tag{2.10}$$

$$e_{\text{MR, Bone}} = \frac{\sum_{M_{\text{Bone}}} |I_{\text{MR}} - I_{\text{fused}}|}{\text{area of the bone mask}} \tag{2.11}$$

where I_{MR} is the intensity of the MR image for a given pixel, I_{fused} is the intensity of the fused

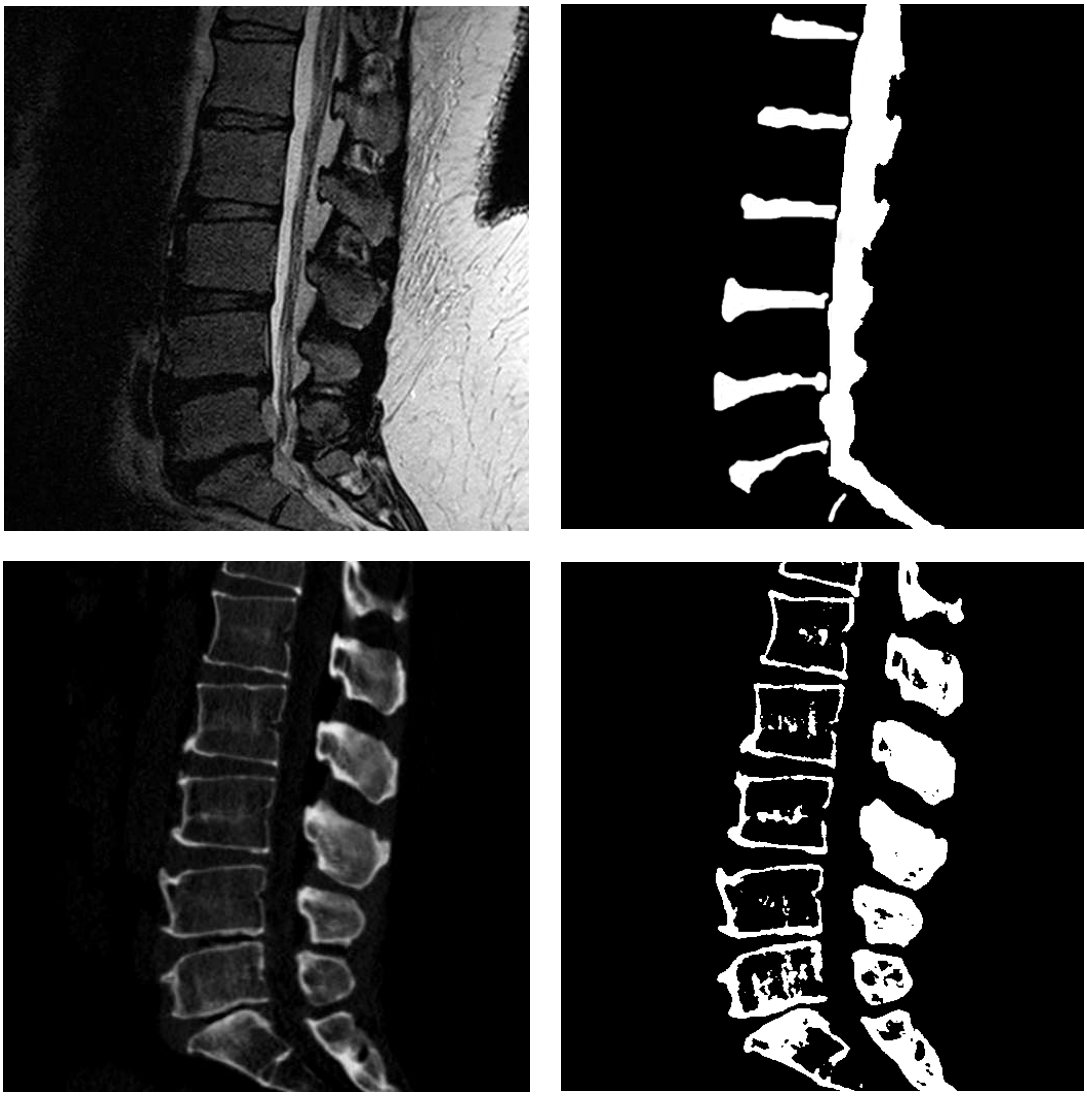


Figure 2.4: Sample masks for validating the quality of the fusion algorithm. Top left: sample MR image, Top right: soft tissue Mask, Bottom Left: sample CT image, Bottom right: bone detail mask

image at a given pixel and M_{Tissue}, M_{Bone} are the non-zero domains of the two masks. Similarly, two additional errors, $e_{\text{CT, Bone}}$ and $e_{\text{CT, Tissue}}$ were defined for the CT images. Ideally, there should be no tissue differences between the MR images and the fused images in the tissue regions ($e_{\text{MR, Tissue}} = 0$) and no bone difference between the CT images and the fused images in the bone regions ($e_{\text{CT, Bone}} = 0$). The hypothesis we tested was that the error obtained for the MR images is lower than the one obtained for the CT images within soft-tissue regions, i.e. $e_{\text{MR, Tissue}} < e_{\text{CT, Tissue}}$, and higher within bone regions $e_{\text{MR, Bone}} > e_{\text{CT, Bone}}$.

Each of the four errors were calculated for each patient. Some of the data was found to be non-normal using a Shapiro-Wilks test [28], thus a non-parametric Wilcoxon test [33] was used to compare sets of errors. The tissue errors: $e_{\text{MR, Tissue}}$ and $e_{\text{CT, Tissue}}$ were compared to each other and the bone errors: $e_{\text{CT, Bone}}$ and $e_{\text{MR, Bone}}$, where also compared, in order to determine if there was a statistical significance difference between them for the 40 patient image sets. These calculations were performed for each of the five fusion methods using version 20 of the SPSS statistical software (SPSS Inc., an IBM Company, Armonk, NY).

2.3.4 Additional Metrics for Evaluation

In addition to the above statistical tests, we have also examined the sensitivity and specificity or our algorithm along with the structural similarity in the masks [32]. For classification, we have defined true and false positives/negatives (TP,FP,TN,FN) per pixel as:

- $\text{TP}_{\text{tissue}}$ and TN_{bone} if ($e_{\text{MR, Tissue}} < e_{\text{CT, Tissue}}$)
- $\text{FP}_{\text{tissue}}$ and FN_{bone} if ($e_{\text{MR, Tissue}} \geq e_{\text{CT, Tissue}}$)
- $\text{TN}_{\text{tissue}}$ and TP_{bone} if ($e_{\text{MR, Bone}} > e_{\text{CT, Bone}}$)
- $\text{FN}_{\text{tissue}}$ and FP_{bone} if ($e_{\text{MR, Bone}} \leq e_{\text{CT, Bone}}$)

Sensitivity and specificity were calculated for each using the total number of TPs, FPs, TNs and FNs normalized over the image masks, which we denote by $\overline{n}$TP, $\overline{n}$FP, $\overline{n}$TN, and $\overline{n}$FN.

Sensitivity and specificity are defined in (2.12). Since the tissue sensitivity is equal to the bone specificity and the bone sensitivity is equal to the tissue specificity, only the two sensitivity values have been reported:

$$\begin{aligned} \text{Sensitivity} &= \frac{\overline{n}\text{TP}}{\overline{n}\text{TP} + \overline{n}\text{FN}} \\ \text{Specificity} &= \frac{\overline{n}\text{TN}}{\overline{n}\text{TN} + \overline{\text{n}}\text{FP}} \end{aligned} \tag{2.12}$$

The structural similarity metric [32] is defined as:

$$\text{SSIM}(x, y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{2.13}$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y, \sigma_{xy}$ represent the means in the x and y images, the variances in the x and y images and the covariance of the two images respectively. This metric has been applied over a local window for pixels within the given masks, comparing the MR images to the fused images in the tissue mask, and the CT images to the fused images in the bone masks. The window was defined as an 11x11 Gaussian kernel with $\sigma = 1.5$. $C_1 = 0.01$ and $C_2 = 0.03$ are positive constants.

2.4 Results

In the following, we describe a representative sample of the fusion results, report several statistical evaluations, and discuss four clinical case studies based on our fusion method.

2.4.1 Registration Validation

The TRE [9] was found to be 1.9 ± 0.6 mm with a CT voxel spacing of 3 x 0.3 x 0.3 mm for the tested images. The FLE was found to be 0.8 ± 0.4 mm. This demonstrates that the registration accuracy is sub-voxel, since the TRE is greater than the FLE, but less than the diagonal size of

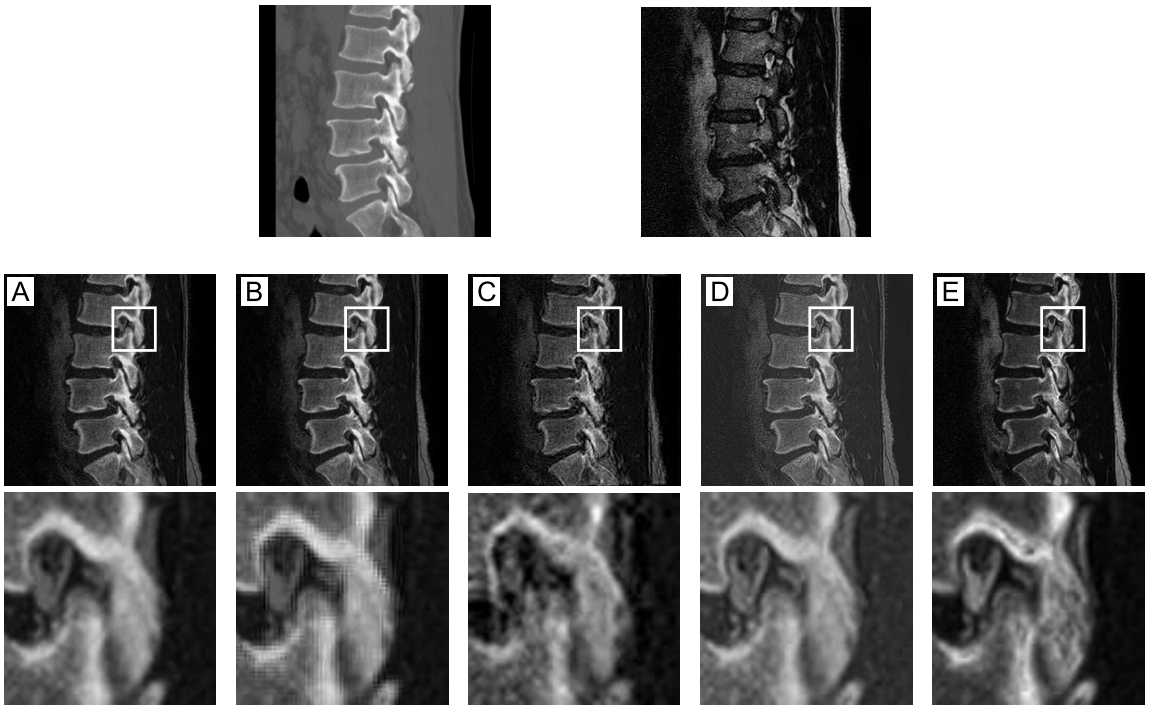


Figure 2.5: Sample fused Images: top left - input CT, top right input MR, middle fused images: A) averaging method, B) discrete wavelet transform, C) contourlet, D) Piella's variational method, E) our proposed graph cut method, bottom row: magnified images within the region of interest indicated in the images above.

the voxels.

2.4.2 Sample Images

Figure 2.5 shows sample input images, including the registered CT and MR inputs and the results of the five fusion methods. It shows that the algorithms perform very differently in preserving the CT/MR details. As expected, the averaging method (A) loses many details, whereas the wavelet method (B) introduces block-structure artifacts because it does not account for shift invariance. The contourlet method (C) significantly blurs the MR details and adds noise to the CT detail, making it difficult to identify the nerve structures and bones. The variational method (D) preserves the details, but significantly reduces the intensity range of the solution. The graph cut result (E) depicts sharp MR and CT details, has a much larger dynamic range than Piella's method (D), and is artifact-free.

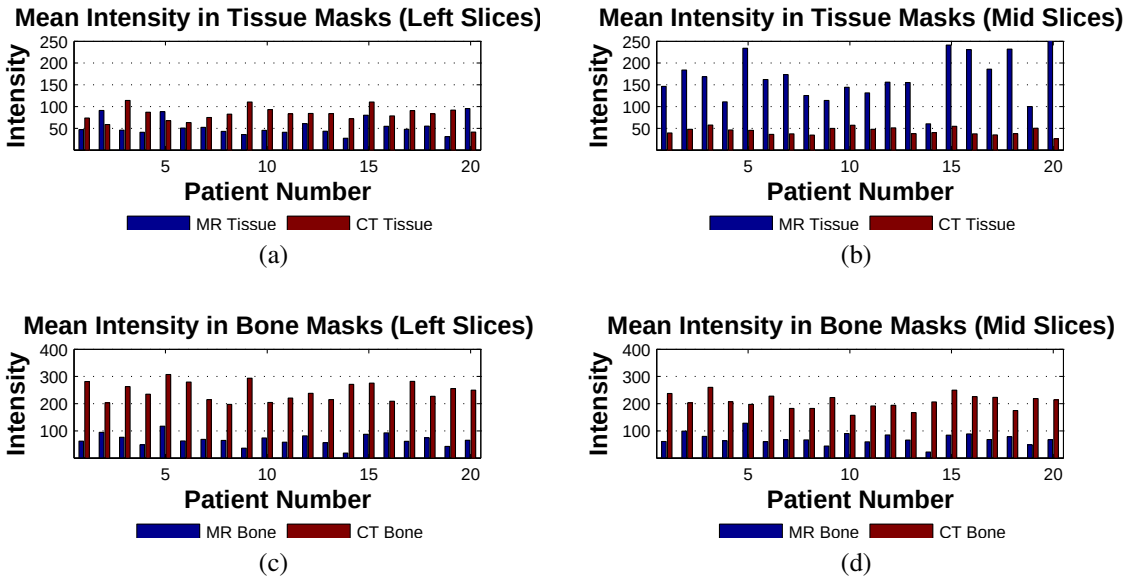


Figure 2.6: Masks Values before fusion, a) Tissue mask and left slices, b) Tissue mask and mid slices, c) Bone mask and left slices, d) Bone mask and mid slices

2.4.3 Evaluation of Masks

We have calculated the mean intensity in the tissue and bone masks, for all the MR and CT images. This provides a frame of reference for the fusion error calculations and to evaluate the effectiveness of the masks. These intensities are shown in Fig. 2.6, with separate graphs for the left sagittal slices through the nerve root bundle and mid sagittal slices, through the center of the subject. For the tissue mask, in the left slices the MR values are about 50, whereas the CT values are between 50 and 100, for the mid slices the CT values are about 50, whereas the MR values range between 100-250 for most patients. This demonstrates that the tissue masks perform well at discriminating between tissue and bone for the mid slices (high MR, low CT), but have less differentiation power for the left slices. For the bone masks, on both the left and mid slices the intensities prior to fusion are between 200 and 300, with the MR values around 100 or less. This shows clear differentiation between bony and tissue detail in the masks.

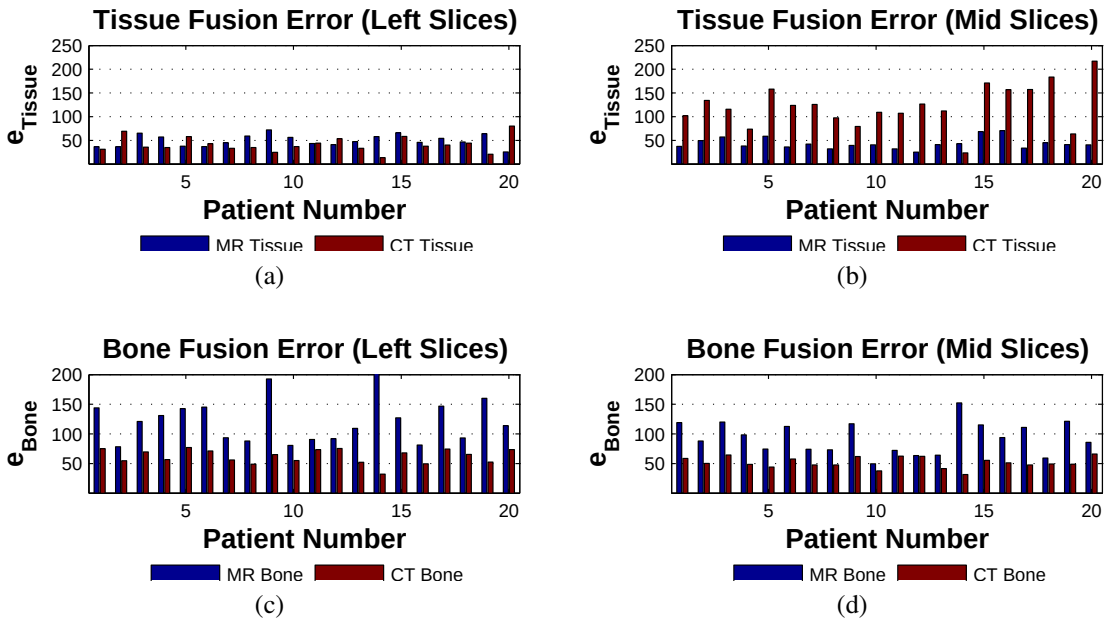


Figure 2.7: Fusion Error for our graph cut method: a) e_{Tissue} for the left slices, b) e_{Tissue} for the mid slices c) e_{Bone} for the left slices and d) e_{Bone} for the mid slices

2.4.4 Statistical Results

The results of the fusion errors for our graph-cut method are shown in Fig. 2.7. These show that $e_{\text{MR,Bone}}$ is greater than $e_{\text{CT,Bone}}$ error in all images as was hypothesized. For the tissue errors, $e_{\text{MR,Tissue}}$ is fairly constant at about 50 for all images, with $e_{\text{CT,Tissue}}$ being much lower on the left images, than the mid images, but higher than $e_{\text{MR,Tissue}}$ for the majority of subjects.

For all five methods, Table 2.3 reports the mean values of $e_{\text{MR,Tissue}}$, $e_{\text{CT,Tissue}}$, $e_{\text{CT,Bone}}$ and $e_{\text{MR,Bone}}$ over the 40 image sets. These were measured in pixel intensity. Table 2.3 also shows p-values for the pairwise and independent Wilcoxon tests comparing $e_{\text{MR,Tissue}}$ with $e_{\text{CT,Tissue}}$ and $e_{\text{CT,Bone}}$ with $e_{\text{MR,Bone}}$, again over the 40 sets of patient images.

As expected, $e_{\text{CT,Bone}} = e_{\text{MR,Bone}}$ and $e_{\text{MR,Tissue}} = e_{\text{CT,Tissue}}$ for the averaging method. For the wavelet/contourlet methods $e_{\text{MR,Tissue}}$ was slightly higher than $e_{\text{CT,Tissue}}$. This is the opposite of what is desired. On the contrary, for the proposed graph-cut method and Piella's method $e_{\text{MR,Tissue}}$ is lower than $e_{\text{CT,Tissue}}$. All the methods, except averaging, yielded $e_{\text{CT,Bone}} < e_{\text{MR,Bone}}$. Overall, the proposed graph cut method resulted in the lowest $e_{\text{CT,Bone}}$, and the lowest $e_{\text{MR,Tissue}}$,

Table 2.3: Mean fusion error values for the 20 patient data sets and statistical-significance test results for the fusion methods.

Method	$e_{\text{MR,Tissue}}$	$e_{\text{CT,Tissue}}$	P_{Pairwise}	$P_{\text{Independent}}$
Averaging	64.2 ± 26.0	64.2 ± 26.0	1	1
DWT [16]	64.9 ± 26.1	63.9 ± 25.9	< 0.001	0.707
Contourlet [35]	73.4 ± 28.6	64.2 ± 24.9	< 0.001	0.083
Piella [25]	152.8 ± 52.6	175.9 ± 54.0	0.002	0.006
Graph Cuts	46.6 ± 12.3	81.7 ± 52.6	0.006	0.020
Method	$e_{\text{CT,Bone}}$	$e_{\text{MR,Bone}}$	P_{Pairwise}	$P_{\text{Independent}}$
Averaging	82.5 ± 19.8	82.5 ± 19.8	1	1
DWT [16]	81.9 ± 19.7	83.7 ± 19.8	< 0.001	0.583
Contourlet [35]	82.6 ± 19.7	89.0 ± 18.6	< 0.001	0.121
Piella [25]	84.1 ± 40.6	188.7 ± 48.0	< 0.001	< 0.001
Graph Cuts	57.0 ± 11.9	108.0 ± 36.2	< 0.001	< 0.001

which corresponds well to our purpose. We obtained a mean $e_{\text{CT,Bone}}$ value of 57.0 based on the CT intensity dynamic range of 700. For the soft tissues $e_{\text{MR,Tissue}} = 46.6$. Note that Piella's method yielded the highest $e_{\text{MR,Tissue}}$. For bone regions, all the methods, except ours, yielded approximately the same $e_{\text{CT,Bone}}$.

With the exception of the averaging method $e_{\text{MR,Tissue}}$ was found to be pair-wise statistically different from $e_{\text{CT,Tissue}}$ and $e_{\text{CT,Bone}}$ was found to be pair-wise statistically different from $e_{\text{MR,Bone}}$. The independent Wilcoxon tests showed that only Piella's method and our graph cut method were statistically significantly different when analyzed as a group.

2.4.5 Additional Metrics

The images have also been analyzed for the sensitivity and specificity of the number of correctly fused pixels along with the structural similarity between the input images and the fused images within the mask regions (Table 2.4).

These results show that the averaging method has 0 sensitivity. The DWT and contourlet methods performed lower than the other two methods, with Piella's methods having the highest sensitivity for both CT and MR and graph cuts having the second highest for both. In regards to

Table 2.4: Mean Sensitivity for Tissue and Bone details along with Mean SSIM Index measures for all five methods.

Method	Sensitivity Tissue	Sensitivity Bone	SSIM Tissue	SSIM Bone
Averaging	0 ± 0	0 ± 0	0.59 ± 0.32	0.08 ± 0.04
DWT [16]	0.51 ± 0.01	0.76 ± 0.17	0.54 ± 0.32	0.11 ± 0.06
Contourlet [35]	0.50 ± 0.01	0.75 ± 0.17	0.22 ± 0.30	0.12 ± 0.08
Piella [25]	0.86 ± 0.11	0.87 ± 0.08	0.32 ± 0.31	0.15 ± 0.11
Graph Cuts	0.63 ± 0.12	0.84 ± 0.09	0.52 ± 0.33	0.21 ± 0.12

the SSIM index, the averaging method performed best on the MR data, followed by the DWT and Graph Cut methods. Piella's methods and the the contourlet transformed did much poorer. For bone details the graph cut method did the best, followed by Piella's method, the contourlet method, the DWT and finally averaging.

2.4.6 Case Studies

Visual Inspection - Lumbar Spine, Joint and Disc Disease We present the first clinical case study for our fusion technique in Fig. 2.8. In the first case, the patient had a protruding spinal disc and damaged facet joint. The disc can be seen in the MR image as a hypointense region, whereas the facet joint is visible in the CT image. There is significant osteoarthritis in the joint. The fused image clearly shows both of these pathologies in a single image, allowing for a better diagnosis.

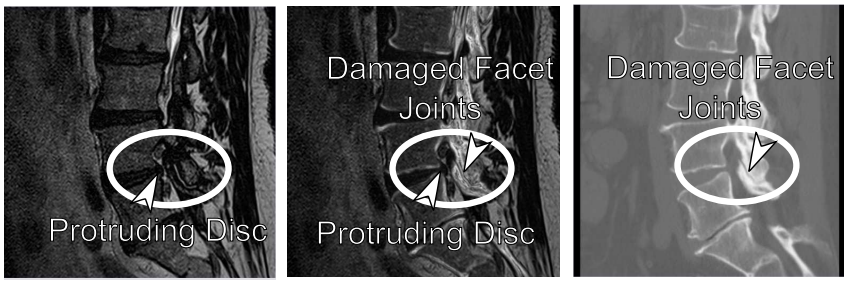


Figure 2.8: Images of damaged spine: left - MR showing a protruding disc (hypointense region); middle - Fused image showing the disc, the spinal cord and the damaged facet joints; right - CT image showing damaged facet joints.

Visual Inspection - Osteophyte Growth The second case study shows osteophyte growth (see Fig. 2.9), which is the formation of bony spurs at the margins of a joint. On the MR image alone, it is difficult to see the location of the osteophyte. The CT shows the osteophyte, but none of the surrounding soft tissue. The fused image shows both the formation of the bony spurs and the surrounding soft tissue on a single image.

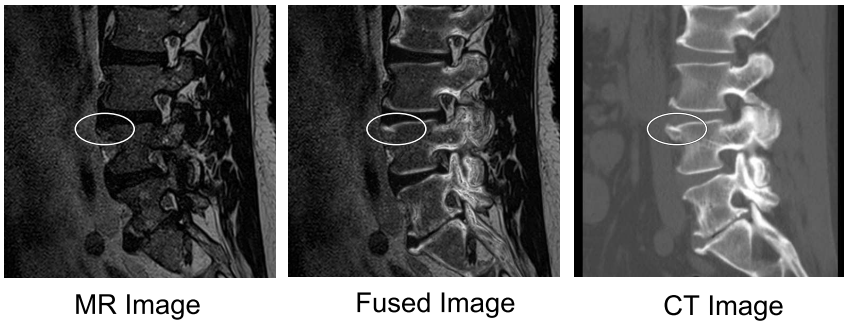


Figure 2.9: Images of bony spur formation: left - in the MR, the osteophyte is not identifiable; middle - in the fused image, the osteophyte is clearly visible along with the surrounding soft tissue; right - the CT image shows the osteophyte, but not the soft tissue.

Visual Inspection - Abnormal Vertebrae and Cord damage The third case study shows spinal cord damage (Fig. 2.10) and an abnormal vertebrae on a single fused image. The cord damage is not visible on the CT image, while the abnormal vertebrae is difficult to see on the MR image. The fused image presents both.

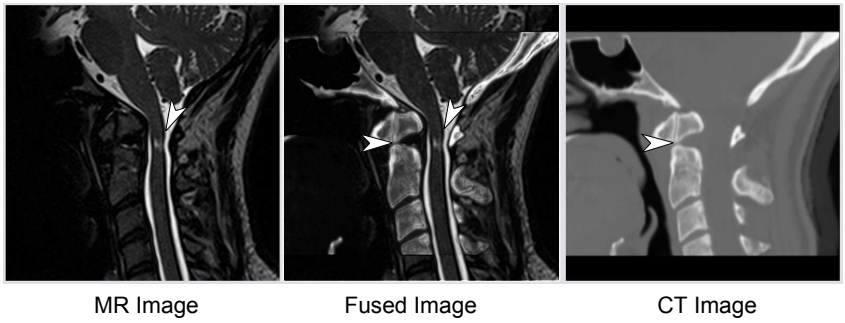


Figure 2.10: Image of an abnormal vertebrae and cord damage: left - in the MR the cord damage is easily visible; center - the fused image: the cord damage and the abnormal vertebral body can be seen clearly; right - the CT image shows the abnormal vertebrae.

Visual Inspection - Osseous Erosion Secondary to Pannus In this patient with rheumatoid arthritis (Fig. 2.11), pannus is eroding the posterior aspect of the dens. On the MR, the chronic

pannus is dark and cannot be distinguished from the underlying bony cortex. On the CT, the margins of the bone are well seen, but soft-tissue contrast is poor. The relationship of the pannus to the underlying bone is best seen on the fused image.

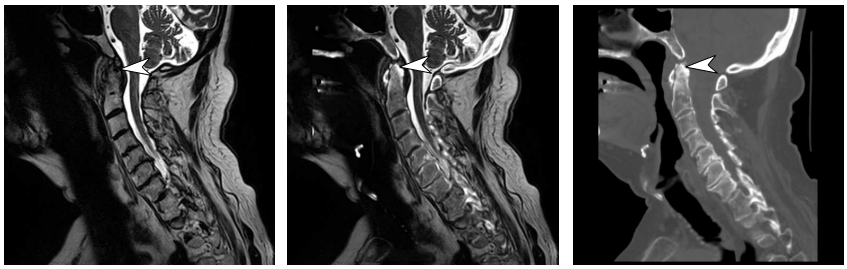


Figure 2.11: The arrows show the pannus eroding the posterior aspect of the tip of the dens. The relationship between the pannus and the surface of the bone is best seen on the fused image.

2.5 Discussion

We have investigated a novel CT/MR spine image fusion algorithm based on graph cuts. We have successfully fused MR and CT images to create a single fused image, providing a new and effective combined modality for diagnosis. Images were registered, pre-processed and then fused. This has been tested on 40 sets of clinical images from 20 patients. The graph cut results show better performance than the averaging method and the three state-of-the-art methods from the literature. Our method successfully transfers bone detail and soft tissue detail to the resulting fused image, with only a 57.0 difference in intensity values for the bone details and 46.6 intensity different for the soft-tissue details, in a dynamic range of 700. Visual inspection confirms these results, with graph cuts showing the sharpest detail for both the bone and soft tissue details.

The statistical tests showed pairwise significance for the CT vs MR error in every method except averaging, however, only Piella's method and our graph cut method showed groupwise statistical significance between errors, which is a stronger test. This indicates that these two methods perform better than the others, in transferring bone detail and soft tissue detail to the

fused image. With regard to the additional methods, Piella's method performed better than graph cuts in regards to the sensitivity test. However, the graph cut method outperformed Piella's method in terms of structural similarity, a test where Piella's method is expected to perform well. Piella's method may have a higher sensitivity than graph-cuts because sensitivity is strongly influenced by the number of false negatives. Since the MR bone error for Piella's method is very high compared to graph-cuts, this would lead to a reduction in false negatives even though the CT bone error is higher for Piella's method compared to graph-cuts. Sample images are shown in Fig. 2.5. In view of all the numeric and visual results, the graph cut method can be concluded to outperform the existing state-of-the-art methods.

The subvoxel accuracy of the registration ensures that the fusion errors are a result of the fusion techniques and not misregistration for our 20 patients. In the proposed method, the rigid image registration assumes minimal structure deformations and patient posture variations. If significant deformation or patient posture difference was present, the non-rigid registration, as a pre-processing step, could be replaced with local-affine or non-rigid image registration, without affecting the quality of the fused images.

One note of interest is that $e_{\text{MR,Tissue}}$ and $e_{\text{CT,Tissue}}$ for the DWT are similar or lower than the averaging, contourlet and Piella's methods, even though the DWT is visually worse than Piella's method. The DWT also has a high SSIM value. There was pairwise statistical significance of the fusion errors for the DWT, but no groupwise statistical significance. This indicates its poor ability to discriminate between tissue and bone detail, which is essential for clinical use.

In this study, the evaluation focuses on fusion of sagittal non-contrast CT and T2 MR Data. Nonetheless, the proposed graph cut fusion method is general for fusing different image modalities and images formatted in different reconstruction planes. It is expected to share similar success if the model parameters are properly adjusted. This is one of the future directions of this research.

We have also shown the benefit of our fusion system on four clinical cases, where the fused

image clearly shows both the bone and soft tissue detail on a single image. This highlights the pathology on a single image. Our method can successfully combine CT and MR images of the lumbar spine, while retaining the significant clinical detail. This eliminates the need for radiologists to mentally align and fuse two separate datasets, along with the associated potential for errors. Although we do not intend to have fused images replace CT and MR scans for clinical use, we do see this as a strong tool to add to the current practice and aid radiologists in completing more accurate and quicker diagnosis.

2.6 Acknowledgments

The authors would like to thank the clinical collaborators who have discussed these ideas with us. This research was supported by funding from the Computer Assisted Medical Intervention (CAMI) training program and the BioMedical Engineering program at Western University, Canada.

Bibliography

[1] Insight ToolKit. URL `http://www.itk.org/`.

[2] Coloma Ballester, Vicent Caselles, Laura Igual, and Joan Verdera. A Variational Model for P+XS Image Fusion. *International Journal of Computer Vision*, 69(1):4358, 2006.

[3] Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[4] Yuri Boykov and Vladimir Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 26(9):1124–1137, September 2004.

[5] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 23(11):1222–1239, November 2001.

[6] Yuri Boykov, Vladimir Kolmogorov, D Cremers, and Andrew Delong. An integral solution to surface evolution PDEs via geo-cut. In *Proceedings of European Conference on Computer Vision*, volume 3, pages 409–422, 2006.

[7] Arnaud Colin and Jean-Yves Boire. MRI-SPECT fusion for the synthesis of high resolution 3D functional brain images: a preliminary study. *Computer Methods and Programs in Biomedicine*, 60:107–116, 1999.

[8] Richard A. Deyo, Sohail K. Mirza, Judith A. Turner, and Brook I. Martin. Overtreating Chronic Back Pain: Time to Back Off? *Jour Am Board Fam Med*, 22(1):62–68, 2009.

[9] J. Michael Fitzpatrick, Jay B. West, and Calvin R. Maurer. Predicting Error in Rigid-Body Point-Based Registration. *IEEE Transactions on Medical Imaging*, 17(5):694–702, 1998.

[10] Paul Hill, Nishan Canagarajah, and David R Bull. Image Fusion Using Complex Wavelets. In *BMVC 2002*, 2002.

[11] Yangqiu Hu, Sohail K. Mirzac, Jeffrey G. Jarvikb, Patrick J. Heagerty, and David R. Haynor. MR and CT image fusion of the cervical spine: a noninvasive alternative to CT-Myelography. In *Proceedings of SPIE*, volume 5744, Bellingham, WA, 2005. SPIE.

[12] C. A. Karlo, I. Steurer-Dober, M. Leonardi, C. W. A. Pfirrmann, M. Zanetti, and J. Hodler. MR/CT image fusion of the spine after spondylodesis: a feasibility study. *European Spine Journal*, 19:1771–1775, 2010.

[13] Olaf Khler, Erik Rodner, and Joachim Denzler. On fusion of range and intensity informa-

tion using graph-cut for planar patch segmentation. *Int. J. Intelligent Systems Technologies and Applications*, 5(3/4):365–373, 2008.

[14] Vladimir Kolmogorov and Ramin Zabih. What Energy Functions Can Be Minimized via Graph Cuts? *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 26(2):147–159, February 2004.

[15] John J Lewis, Robert J. O'Callaghan, Stavri G. Nikolov, David R Bull, and NIshan Canagarajah. Pixel - and region-based image fusion with complex wavelets. *Information Fusion*, 8:119–130, 2007.

[16] H. Li, B. S. Manjunath, and S. K. Mitra. Multisensor Image Fusion using the Wavelet Transform. *Graphical Models and Image Processing*, 57(3):235–245, 1995.

[17] G. K. Matsopoulos, S. Marshall, and J. N. H. Brunt. Multiresolution morphological fusion of MR and CT images of the human brain. *IEEE Proceedings on Vision Image Signal Processing*, 141(3):137–142, 1994.

[18] David Mattes, David R. Haynor, Hubert Vesselle, Thomas K. Lewellen, and William Eubank. PET-CT Image Registration in the Chest Using Free-form Deformations. *IEEE Transactions on Medical Imaging*, 22(1):120–128, January 2003.

[19] Pamela M. Mazurek, Anne M. Silas, and John M. Gemery. Summation of CT Scans During Radiofrequency Ablation for Assessing Target Lesion Coverage. *American Journal of Roentgenology*, 191:790:792, September 2008.

[20] B. Miles, I. Ben Ayed, M. W. K. Law G. Garvin, A. Fenster, and S. Li. Spine image fusion via graph cuts. *IEEE Transactions On Biomedical Engineering*, 60(5):1841–50, July 2013. ©2013 IEEE. Reprinted, with permission, from Brandon Miles, Ismail Ben-Ayed, Max W. K. Law, Gregory J. Garvin, Aaron Fenster and Shuo Li, Spine Image Fusion via Graph-Cuts, IEEE Transactions on BioMedical Engineering, July 2013.

[21] Brandon Miles, Max W. K. Law, Ismail Ben-Ayed, Greg Garvin, Aaron Fenster, and Shuo Li. Pixel Level Image Fusion for Medical Imaging - an Energy Minimizing Approach. In *Proceedings of SPIE Medical Imaging*, 2012.

[22] Filippo Nencini, Andrea Garzelli, Stefano Baronti, and Luciano Alparone. Remote sensing image fusion using the curvelet transform. *Information Fusion*, 8:143–156, 2007.

[23] Jorge Núñez, Xavier Otazu, Octavi Fors, Albert Prades, Vicenç Palà, and Román Arbiol. Multiresolution-Based Image Fusion with Additive Wavelet Decomposition. *IEEE Transactions On Geoscience And Remote Sensing*, 37(3):1204–1211, May 1999.

[24] Pamela L. Owens, Maeve Woeltje, and Ryan Mutter. Emergency Department Visits and Inpatient Stays Related to Back Problems. 2008. URL `http://www.hcup-us.ahrq.gov/reports/statbriefs/sb105.pdf`.

[25] Gemma Piella. Image Fusion for Enhanced Visualization: A Variational Approach. *International Journal of Computer Vision*, 83(1):1–11, 2009.

[26] Ron Rensink. Change blindness. *Annual Review of Psychology*, 53:245–277, 2002.

[27] Mohammed Ben Salah, Amar Mitiche, and Ismail Ben Ayed. Multiregion Image Segmentation by Parametric Kernel Graph Cuts. *IEEE Transactions on Image Processing*, 20(2):545–557, 2011.

[28] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591611, 1965.

[29] Moon-Jun Sohn, Dong-Joon Lee, Sang Won Yoon, Hye Ran Lee, and Yoon Joon Hwang. The effective application of segmental image fusion in spinal radiosurgery for improved targeting of spinal tumours. *ACTA Neurochir*, 151:231–238, 2009.

[30] Chao Wang and Zhong-Fu Ye. Perceptual Contrast-Based Image Fusion: A Variational Approach. *ACTA Automica Sinica*, 33(2):132–137, February 2007.

[31] Wei-Wei Wang, Peng-Lang Shui, and Xiang-Chu Feng. Variational Models for Fusion and Denoising of Multifocus Images. *IEEE Signal Processing Letters*, 15:65–68, 2008.

[32] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[33] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6): 80–83, December 1945.

[34] Stephen T. C. Wong, Robert C. Knowlton, Randy A. Hawkins, and Kenneth D. Laxer. Multimodal Image Fusion for Noninvasive Epilepsy Surgery Planning,. *IEEE Transactions on Computer Graphics and Applications*, 16(1):30–38, 1996.

[35] L. Yang, B. L. Guo, and W. Ni. Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform. *Neurocomputing*, 72:203–211, 2008.

[36] Zheng Youzhi, Qin Zheng, and Yang Jingyu. Image Fusion Using a Hybrid Representation of Empirical Mode Decomposition and Contourlet Transform. In *Proceedings of Information Science and Technology (ICIST)*, 2011.

[37] Guang Ming Zhang and Zhi Ming Cui. A novel image fusion method using beamlet transform and graph cuts. *Key Engineering Materials*, 467 - 469:1092–1096, 2011.

[38] Guangming Zhang, Zhiming Cui, Fanzhang Li, and Jian Wu. DSA Image Fusion based on Dynamic Fuzzy Logic and Curvelet Entropy. *Journal of Multimedia*, 4(3):129–136, June 2009.

[39] Wei Zhang and Le Yu. SAR and Landsat ETM+ image fusion using variational model. In *2010 International Conference on Computer and Communication Technologies in Agriculture Engineering*, 2010.

Chapter 3

Real time MR/CT Image Fusion for Lumbar Spine Diagnosis

3.1 Introduction

3.1.1 Background and clinical interest

The current availability of various imaging modalities presents clinicians with various views, each of which highlights distinct anatomical and functional details. Integrating information from multiple modalities into a single view, often referred to as image fusion, allows depicting complementary and clinically important details [8, 20, 34]. An application of high clinical interest is the fusion of MR and CT data of the spine [22]. MR and CT images are two main modalities for the diagnosis and treatment of most spine diseases. These two modalities provide complementary information: an MR scan, especially a T2-weighted image, depicts useful soft-tissue details of the intervertebral discs, - including diagnosis of protruding and degenerated discs, information about the spinal cord and its associated nerves as well as the cerebral spinal fluid space. A CT scan depicts the high contrast bony structures, especially the bone cortex, which allows for the assessment of damaged bone and joints. In routine clinical use, radiologists often examine the two modalities side-by-side, which requires mental reposition-

ing, alignment and fusion of the respective images in order to determine the relative locations of details in the two images. This is an error-prone and time-consuming process. Even when the MR and CT images are closely aligned the position of anatomy in one image relative to anatomy in the other image may not be clear, such as the exact edges of a protrusion relative to surrounding bone tissue.

As a complement to MR and CT images currently used, the addition of a fused image of the spine containing both the soft tissue and bone details from the MR and CT images promises to facilitate more accurate diagnoses in less time. This is particularly important given the exponential increase in spine scans during the last decade. For instance, the study in [9] reports a 300% increase in the number of lower spine MR scans in the period between 1994 and 2006. Also, the study in [9] reports that 75% of adults will suffer from back pain, which was the cause of 1.8% of all hospital visits in the United States in 2008.

Image fusion has seen multiple medical applications. These include: image fusion used for brain imaging [20, 34], MRI-SPECT fusion [8], non-invasive epilepsy surgery planning [32], and digital subtraction angiography [38]. For the treatment and diagnosis of the human spine, registered and overlaid images have been utilized in the evaluation of bone implants [15] and for surgery planning [14, 29]. More recently, the fusion of MR and CT spine images was studied in [22], which was demonstrated to be the state-of-the-art for spine image fusion. The largest drawback to this method is its computational time, making it unsuitable to real-time image fusion.

Clinically, a large drawback to long computational time is the inability to interactively adjust the fused image. Currently when viewing medical images, radiologists can adjust the window and level, to optimally highlight a specific tissue range. These settings are often changed multiple times throughout a reading. For fused images, similar controls are also desirable. Allowing clinicians control over the fused images will allow radiologists to explore different tissue / disease settings. An example of this capability is the need to adjust the relative balance between the CT and MR images. A real-time image fusion system is the easiest way to provide

this capability.

3.1.2 Technical Overview of Prior Art on Image Fusion

In addition to the medical applications of [8, 14, 15, 20, 22, 29, 32, 34, 38] there has been extensive research into non-medical image fusion applications, which vary from multifocal images [18] to multi-spectral imaging (often with geographical applications) [33]. Both medical and non-medical applications are implemented using similar techniques and in general these formulations of image fusion algorithms fall within two main categories: (1) Transform-based methods [13, 17, 18, 24, 25, 34, 35] and (2) Functional-optimization methods [2, 22, 27, 30].

Transform-based methods

Transform-based methods fuse the input images in a transformation domain, not the image domain. For instance, a wavelet-based approach has been proposed in [18], where two images of different focal lengths were fused in the wavelet domain, and an inverse transformation was applied to recover the final result. Additionally wavelet transform variations include: additive wavelet decomposition [25], the contourlet transform [34, 35], the curvelet transform [24], the complex wavelet transform [13, 17] and other wavelet approaches [26]. Other methods include using an independent component analysis basis [23] and the polyharmonic local sine transform [19]. Due to the limitations of the transforms used, transform-based methods often distort fine details, which may affect important clinical information. For example, wavelet-based methods suffer from pixelation artifacts when two dissimilar images are fused. This is a result of the decimations needed for the forward and inverse transforms, which are compounded because standard wavelets are not image oriented. Such artifacts may lead to a significant loss of image quality, making it difficult to visualize certain pathologies to the point where they may not be seen. Contourlets [34, 35] and complex wavelets [13, 17] were also studied to address this problem. Unfortunately, such descriptions are limited in the number of directions that can be represented in their structures. Use of an independent component analysis basis [23] is another

alternative. However, the basis must be trained to generate an application specific transform. This has the advantage of being problem specific, but also leads to a high dependency on the trained basis, reducing its flexibility.

Functional-optimization methods

Functional-optimization statements of image fusion have recently attracted several research efforts [2, 22, 27, 30], and can address some of the limitations of transform based methods. Unfortunately, the main difficulties in these methods come from the limitations of the optimizers. For instance, the methods in [2, 27, 30] rely on standard gradient-descent procedures. Such optimizers yield sub-optimal solutions and a very large computational load, which precludes their use in 3D (or higher) dimension, as is common in medical imaging. The recent study in [22] attempted to address spine image fusion with a multi-label graph cut formulation, where the fused image is described by a finite set of labels. Using established graph-cut optimizers from combinatorial optimization (e.g., swap expansion moves) [6], the authors of [22] showed promising performances in the 2D case, outperforming the methods of [18, 27, 34]. It is well known that graph cuts can yield an excellent performance in the case of 2D grids [5], with a relatively small number of labels and neighborhood systems. However, the extension of graph-cut optimizers to applications where the number of labels is large, as is the case of image fusion problems, or when the image dimension is high (3D or higher) as is the case of medical data, may result in a very large computation and memory load [16]. The complexity of multi-label graph cuts increases super-linearly with the number of labels because pairwise combinations of the labels are considered. Furthermore, graph cuts cannot be easily parallelized because of the pairwise potentials [36]. Ideally a real-time image fusion system could be utilized to allow interactive adjustment of the fused images.

3.1.3 Contributions of this study

This study investigates multi-modal spine image fusion algorithms based on convex optimization. We are able to solve a similar model to [22], while allowing for interactive adjustment of the fused images via real-time computation. We optimize an objective functional that contains three terms: (1) an edge weighted data term, seeking a solution that is similar to the MR image, (2) an edge weighted data term, seeking a solution that is similar to the CT image, and (3) a total variational term that encourages edge smoothness in the output image. The data terms are constrained by their L1 and L2 norms (absolute value and squared difference) and weighted by gradients in the images. Unlike the graph cut approaches [22] we are able to ensure a globally optimal solution, without the need for multiple labels in the image, via convex optimization. We demonstrate these results on 30 patient volumes.

The major contributions of this work include four new convex approaches to spine image fusion. These are shown to have similar or better performance than the state-of-the-art graph cut approach [22], which has only been applied to 2D image slices. The four novel methods are applicable to both 2D and 3D images, and achieve real-time results on a GPU for 3D volumes. This allows interactive adjustment of the fused images. These methods include convex absolute value difference methods (CABS) and convex squared difference methods (CSD), which are both implemented in 2D and 3D.

The remainder of this paper is organized as follows. In section 3.2, we state image fusion as a functional optimization and discuss two solutions, namely: convex absolute value difference methods and convex squared difference methods, based on a previous conference paper by one of the authors [37]. In section 3.3, the experimental methods are described, including data selection and preprocessing, along with registration methods and our validation scheme. Several experiments with their results are presented in section 3.4, including the registration results, visual and numerical fusion results and results on the computational runtimes of the algorithms. This is followed by a brief discussion in section 3.5 with proofs from section 3.2 contained in the appendix.

3.2 Convex Optimization Approaches to Image Fusion

Here we present an overview of the fusion problem, including its formulation with a focus on the details impacting the clinical results of this work. The mathematical details including a rigorous definition of equations, along with further derivations, proofs and implementation details are provided in the appendix.

We seek to minimize an energy functional that balances three competing terms in order to create a single fused image that preserves the important clinical details in the input MR and CT images. These terms and their goals are designed as follows:

1. An edge weighted data term, biasing the solution towards a fused image that is similar to the input MR image. This aims to preserve soft tissue detail from the discs, nerves and spinal cord. In the T2-weighted MR image, the nerve roots, spinal cord and cerebral spinal fluid present with strong edges and high pixel intensities. The discs in the T2-MR image also present with a high intensity value.

2. An edge weighted data term, biasing the solution towards a fused image that is similar to the input CT image. This aims to preserve cortical bone detail, which has both strong edges and high pixel intensity corresponding to the high Houndsfield Units (HU).

3. An edge preserving smoothness term, biasing the solution to maintain natural transitions in the image.

We can express the above criteria as the minimum of a functional of the following general form:

$$\min_{u} \; E(u) \; := \; \mathcal{D}(u, f_1, f_2) + \lambda \mathcal{S}(u)\,. \tag{3.1}$$

where, u, represents our output image, f_1 represents the first input image (the MR scan), f_2 represents the second input image (the CT scan), and λ is a positive constant. Additionally $\mathcal{D}$

represents a data term, dependent on the input images and the output images, and $\mathcal{S}$ represents a smoothness term, which is only dependent on the output image.

Specifically we define the data term $\mathcal{D}$ as:

$$\mathcal{D}(u) = \int_\Omega w_1 A(f_1 - u)\, dx + \int_\Omega w_2 A(f_2 - u)\, dx, \tag{3.2}$$

Where A is some positive convex function (e.g. $(f-u)^2$ or $|f-u|$), w_1, w_2 are weights that depend on the edges of the images and Ω is the domain of the image. This encodes objectives 1) and 2) resulting in $\mathcal{D}$ measuring an edge-weighted conformity of a solution u.

We define the smoothness term $\mathcal{S}(u)$ to be the total variation of u, which is designed to create an output image containing natural transitions, encoding objective 3). The use of the total variational smoothness term preserves strong edges in the image, while preserving smoothness.

We choose the L1-norm (absolute value) and the L2-norm (squared difference) for A as these are two of the most commonly used norms. This results in the following two data terms:

$$\textit{L1-norm problem: } \mathcal{D}_{\mathrm{ABS}}(u) = \int_\Omega w_1 \,|f_1 - u|\, dx + \int_\Omega w_2 \,|f_2 - u|\, dx \tag{3.3}$$

or

$$\textit{L2-norm problem: } \mathcal{D}_{\mathrm{SD}}(u) = \int_\Omega w_1 \,(u - f_1)^2\, dx + \int_\Omega w_2 \,(u - f_2)^2\, dx\,. \tag{3.4}$$

These data terms result in the following optimization problems, which we denote as E_{ABS} for the absolute value norm and E_{SD} for the squared difference norm:

$$\min_{u\in\Omega}\left\{E_{\mathrm{ABS}} = \int_\Omega w_1\,|f_1 - u|\,dx + \int_\Omega w_2\,|f_2 - u|\,dx + \lambda\int_\Omega |\nabla u|\,dx\right\} \tag{3.5}$$

or

$$\min_{u:\Omega} \left\{ E_{\mathrm{SD}} = \frac{1}{2} \int_{\Omega} w_1 (u - f_1)^2 dx + \frac{1}{2} \int_{\Omega} w_2 (u - f_2)^2 dx + \lambda \int_{\Omega} |\nabla u| \, dx \right\} \tag{3.6}$$

We can solve these two energies, using two different algorithms. We refer to these algorithms as CABS (convex absolute value) and CSD (convex squared difference) for E_{ABS} and E_{SD} respectively. We have left further details of these two algorithms for the appendix. It is important to note that these solutions are convex, which means that we can always find the image that corresponds to the optimal energy and we can do so very efficiently. Additionally the algorithms used to find the optimal fused image are highly parallel and amenable to implementation on modern graphics hardware. This addresses the biggest drawback of the state-of-the-art algorithm [22] allowing for an interactive image fusion system.

3.3 Data Selection and Evaluation Methods

This retrospective study was approved by the Human Subjects Ethics Board of the University of Western Ontario, with the requirement for informed consent waived. A total of 30 patients were randomly selected, who had received lumbar MR and CT scans at most one year apart. The patients had a range of diseases including osteophytes, spinal stenosis and disc protrusions. The MR scanners used were a Magnetom and Avanto Siemens 1.5T MR scanners (Seimens AG, Erlangen, Germany). The CT scanner used varied based on the location where the images were acquired. Pairs of CT and T2 weighted 3D MR images were selected for each patient. These CT and MR images were sagittal reconstructions from axial CT slices and 3D T2-weighted MR scans respectively. The CT pixel spacing was 0.9mm by 0.9mm, with a slice thickness varying between 1.5-3mm, depending on the patient. The MR pixel spacing varied from 0.41 mm by 0.41mm to 0.94 by 0.94 mm, with the slice thickness varying between 0.27 mm and 0.41mm, depending on the patient. The image pairs were then registered, preprocessed and fused into a single image. For each patient a pair of corresponding sagittal slices, from the

registered images, were selected at the middle of the patient. For 2D fusion algorithms, the image pairs were selected prior to fusion. These were then fused and validation was performed on these slices. For 3D fusion algorithms, slice selection was completed after the fusion of the 3D images and validation was performed on these slices.

The techniques were validated based first on visual results for the clarity and detail present in the fused images. Statistical comparisons were then made, based on the clinically relevant tissues transferred from the input images to the fused images. Time for computation was also analyzed, including runtimes for the fusion algorithms and a comparison of runtimes with varying image sizes. Finally, clinical case studies are presented to illustrate the clinical potential of this technique. A total of six methods were compared: 1) a simple averaging of the two input images, 2) a multi-label graph cut method (GC) [22], 3) our absolute value convex method in 2D (CABS-2D), 4) our absolute value convex method in 3D (CABS-3D), 5) our squared difference convex method in 2D (CSD-2D), and 6) our squared difference convex method in 3D (CSD-3D).

3.3.1 Registration

Each pair of 3D MR and CT images were rigidly registered to each other. This was accomplished using rigid registration based on the Mattes mutual information metric [21] and was implemented using the Insight toolkit [1]. The 3D versor class and its associated optimizer were used for the transformation. For each image pair, registration was initialized based on a user identified corresponding point located on each of the two images. The MR images were registered to the CT images based on soft tissue correspondence and then resampled to the CT voxel size. The CT soft tissue is suitable for registration purposes, but MR is the preferred modality for diagnosis.

3.3.2 Validation of Registration

A Target Registration Error (TRE) and Fiducial Localization Error (FLE) [11] were calculated to validate the registration of the images. A total of 48 fiducials, based on 12 fiducials each from four patients, were used for calculating the TRE. These were identified manually in corresponding CT and MR images based on four points located on the vertebral edges on each of three separate vertebrae per patient. For calculating the FLE, four fiducials were identified in a sample CT image and corresponding fiducials were located in the MR image. The MR fiducial selection was repeated 5 more times resulting in a total of 24 samples being used to calculate the FLE. To account for memory bias, the minimum gap between repetitions was 8 hours.

3.3.3 Preparation of the Dataset

The registered images were preprocessed prior to image fusion. In order to remove unwanted soft tissue detail from the CT image, the intensities were cut-off below a user specified minimum cut-off $I_{\min}$. To ensure that the CT bone structures would not eclipse the soft tissue structures in terms of relative brightness, the images were scaled to be within similar ranges, again by a user adjustable parameter d. This preprocessing is detailed in (3.7),

$$I = \begin{cases} 0 & I_o \leq I_{\min} \\ (I_o - I_{\min})/d & I_o > I_{\min} \end{cases} \tag{3.7}$$

where I_o is the initial CT intensity (in HU), and I is the output intensity.

To maintain a fair comparison, the algorithms were tested with the same parameters as [22]: $d = 0$ and $I_{\min} = 2$. In clinical use these could be easily varied. Histograms of the 30 mid-sagittal slices are shown in Fig. 3.1 before and after fusion using these parameters. A flow chart of the registration, preprocessing and fusion processing is shown in Fig. 3.2.

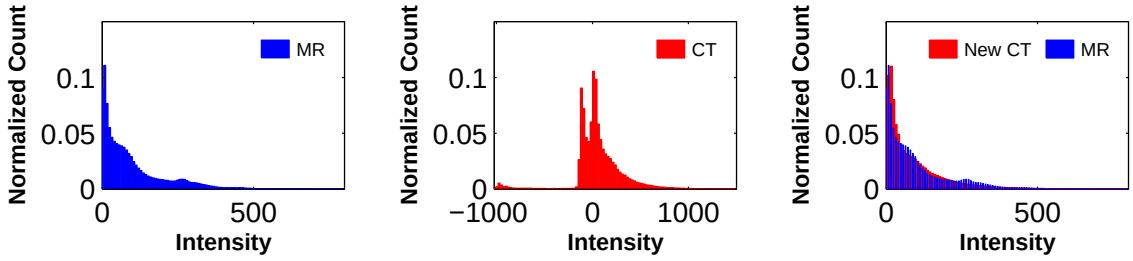


Figure 3.1: Histograms of the 30 mid-sagittal images, a) Histogram of the 30 MR images, b) Histogram of the 30 CT images, c) Histogram of the 30 CT images after Preprocessing (excluding pixels set to 0) and histogram of the 30 MR images

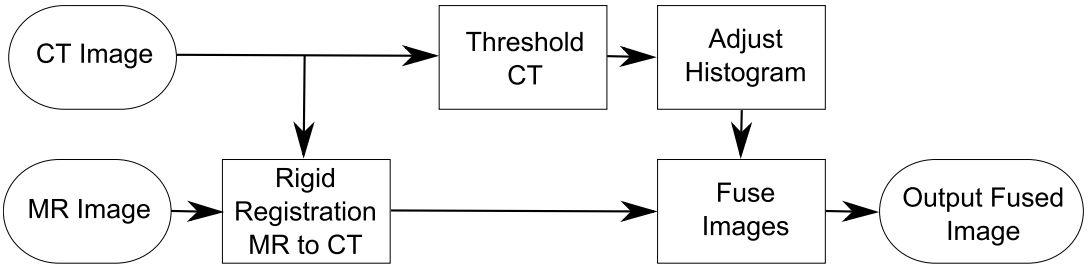


Figure 3.2: A flow chart of the registration, preprocessing and fusion of the data.

3.3.4 Parameters for our Image Fusion Methods

For our fusion methods, the kernel K, used in (3.11) was a 3x3, or 3x3x3 box filter, for the 2D and 3D methods respectively. These kernels have been normalized so the elements sum to 1. The individually tuned parameters are given in Table 3.1. These were divided based on λ from (3.5) and (3.6), which was tuned for visual quality, and c, τ, from (3.27), (3.34) and (3.44) which were tuned for rate of convergence.

Table 3.1: Parameters used for the convex fusion methods. Visual quality was affected by the weight of smoothing λ, while the rate of convergence for the algorithm was determined by step sizes c and τ.

Fusion Method	Visual Quality	Speed
CABS-2D	$\lambda = 0.2$	$c = 25, \tau = 0.15$
CSD-2D	$\lambda = 1.0$	$\tau = 0.23$
CABS-3D	$\lambda = 0.2$	$c = 28, \tau = 0.11$
CSD-3D	$\lambda = 1.0$	$\tau = 0.10$

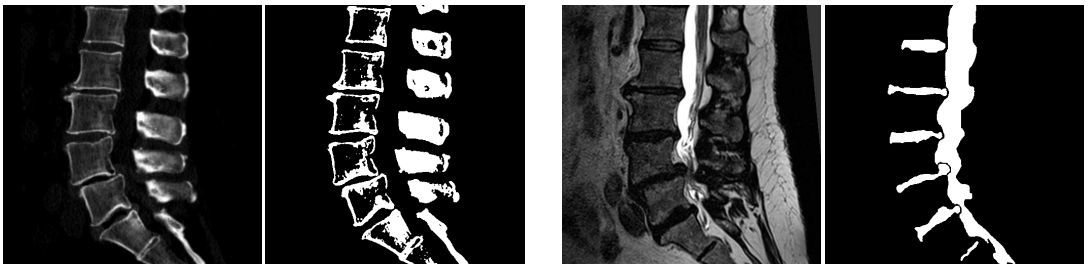


Figure 3.3: Sample masks for validating the quality of the fusion algorithm. From the left to right: - CT input image, bone detail mask, MR input image, soft tissue detail mask

3.3.5 Fusion Validation

The goal of the fusion is to create a fused image containing the soft tissue detail from the MR images, and the bony detail from the CT images. To test this, first the areas of soft tissue and the regions of bone were defined for each of the patients. The areas of soft tissue consist of the spinal cord, the cerebral spinal fluid, the discs and the nerve tissue. Tissue masks were created by manually segmenting these soft tissue details from the mid-sagittal slices taken from each of the registered 3D MR images. The bony details consist of the bone, with an emphasis on the bone cortex. Bone masks, for each patient, were created from the CT images using the mid-sagittal slices. The slices have been thresholded to include any CT value corresponding to bone, with a user-selected threshold in HU for each slice. These were then manually refined, to correct any errors. The masks were checked for correctness by a radiologist (GJG). Sample masks and images are shown in Fig. 3.3.

Based on these masks, we define four fusion errors. For the CT image, the two errors are $e_{\text{CT,Bone}}$, which is the error between the fused image and the CT image in the bone regions and $e_{\text{CT,Tissue}}$, which is the error between the fused image and the CT image in the tissue regions. These are formulated as:

$$e_{\text{CT,Bone}} = \frac{\sum_{M_{\text{Bone}}} |I_{\text{CT}} - I_{\text{Fused}}|}{\text{area of the bone mask}} \qquad e_{\text{CT,Tissue}} = \frac{\sum_{M_{\text{Tissue}}} |I_{\text{CT}} - I_{\text{Fused}}|}{\text{area of the tissue mask}}$$

where I_{CT} denotes the intensity of the CT image at a given pixel, I_{Fused} denotes the intensity of the fused image at a given pixel, and M_{Bone}, M_{Tissue} are the non-zero domains of the bone and

tissue masks. In a similar way two more errors: $e_{\text{MR,Bone}}$ and $e_{\text{MR,Tissue}}$ were defined for the MR images. Ideally the fused image will be exactly the same as the MR image in the soft tissue regions and the fused image will be exactly the same as the CT image in the bone regions. This corresponds to $e_{\text{MR,Tissue}} = 0$ and $e_{\text{CT,Bone}} = 0$, with $e_{\text{CT,Tissue}}$ and $e_{\text{MR,Bone}}$ being much higher. The hypothesis we tested was that the CT error should be lower than the MR error in the bone regions i.e., $e_{\text{CT,Bone}} < e_{\text{MR,Bone}}$ and higher in the tissue regions, i.e., $e_{\text{CT,Tissue}} > e_{\text{MR,Tissue}}$.

To compare the six fusion methods, the four errors were calculated for each patient using each fusion method. The values could then be evaluated for statistical significance. For each method, the values of $e_{\text{MR,Tissue}}$ were compared to $e_{\text{CT,Tissue}}$, and the values of $e_{\text{CT,Bone}}$ were compared to $e_{\text{MR,Bone}}$. Some of the data was found to be non-parametric by a Shapiro-Wilks test [28], thus a non-Parametric Wilcoxon test [31] was used. These calculations were performed using version 20 of the SPSS statistical software (SPSS Inc., an IBM Company, Armonk, NY).

3.3.6 Runtime Validation

The mean times for computation and the number of iterations to convergence were calculated. For the 2D case the mean runtime was calculated using the mid-sagittal slices, for the 30 patients with the error rate for convergence set at $\epsilon < 10^{-4}$. For the 3D cases, the convergence was set at $\epsilon < 10^{-3}$. The computations were performed using a quad core 2.0 GHz Xeon processor (Intel, Santa Clara, CA, USA), 2GB of RAM and an NVidia GEFORCE 680 GTX (Nvidia, Santa Clara, CA, USA) with 1536 CUDA cores and 2GB of video RAM. The CPU implementations were written in C++ and the GPU implementations were written in CUDA. Matlab version 7.8.0 (The Mathworks Inc, Natick, MA, USA) was used to load the images and call the CPU / CUDA algorithms.

Table 3.2: Target Registration Error based on four subjects and Fiducial Localization Error based on one subject. All measurements are in mm.

	Axial	Sagittal	Coronal	Total
TRE	0.97 ± 0.40	0.89 ± 0.26	1.03 ± 0.30	1.90 ± 0.51
FLE	0.51 ± 0.23	0.83 ± 0.64	0.84 ± 0.21	0.62 ± 0.44

3.4 Experiment Results

3.4.1 Registration Results and Validation

For the registration, the TRE was found to be 1.9 ± 0.5 mm, based on four subjects with voxel sizes in the range of: 0.30-0.48 x 0.30-0.48 x 2.5-3 mm and the FLE of was found to be 0.6±0.4 mm based on one subject with a voxel spacing of 0.33 x 0.33 x 2.0 mm. Table 3.2 contains a breakdown of these values in the axial, sagittal and coronal directions. The TRE voxel sizes are based on the CT images from the four subjects used to evaluate the registration, while the FLE voxel size is based on the CT image from the subject used. For all 30 subjects, the registered MR images have been resampled to match the corresponding CT image spacing for each CT-MR image pair. It should be noted that the TRE values for the sagittal and coronal planes are greater than the in-plane pixel spacing; however, these measurements are limited by the FLE, which is also greater than the in-plane pixel spacing. The overall TRE is less than the voxel size for the CT and resampled MR images so it can be concluded that the registration error is subvoxel.

3.4.2 Visual Results

A visual comparison of the methods can be seen in Fig. 3.4 and Fig. 3.5. Sagittal slices through the center of the spine are shown in the first image, and slices through the dorsal root ganglion are shown in the second. The six fusion methods: 1) averaging, 2) graph cuts, 3) CABS-2D (the 2D version of our convex absolute value method), 4) CABS-3D (the 3D counterpart to CABS-2D), 5) CSD-2D (the 2D version of our convex squared differences method) and 6)

CSD-3D (the 3D counterpart to CSD-2D) were compared for clarity of bone and soft tissue detail present in the fused images. Additionally, cross-sectional views from the 3D methods CABS-3D and CSD-3D are shown in Fig. 3.6.

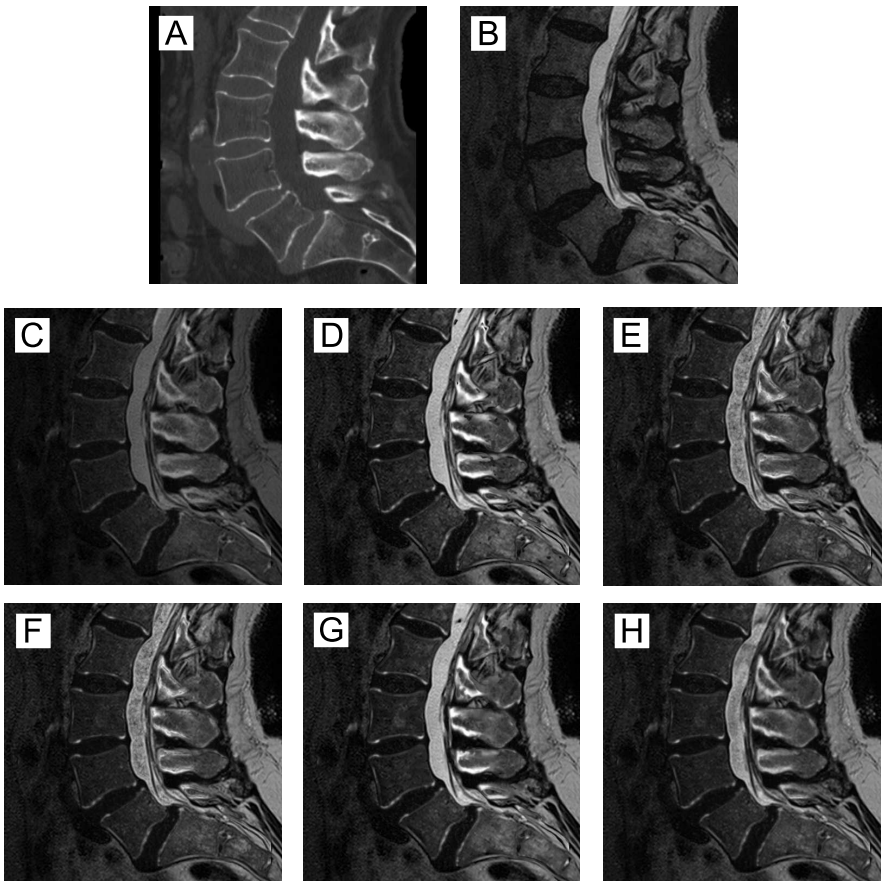


Figure 3.4: Fused sagittal midslices showing the cerebral spinal fluid and vertebrae of the patient. A: Input CT image, B: Input MR image, C: Fused averaging method, D: Fused CABS-2D, E: Fused CSD-2D, F: Fused graph cut method, G: Fused CABS-3D, H: Fused CABS-2D

Based on the results in Fig. 3.4 and Fig. 3.5 some comments can be made on the visual advantages of the various methods. The averaging results, shown in image C in both figures, do fuse the images, but the results are blurry and lose some details. The CABS methods, which are shown for 2D in image D and 3D in image G (for both figures), produce some noticeable artifacts in the images. These two images are clearer than the averaged image, but these artifacts affect the quality of the results. In Fig. 3.4, they are significantly less noticeable than in Fig. 3.5. The 3D version appears to have more smoothing than the 2D version. The

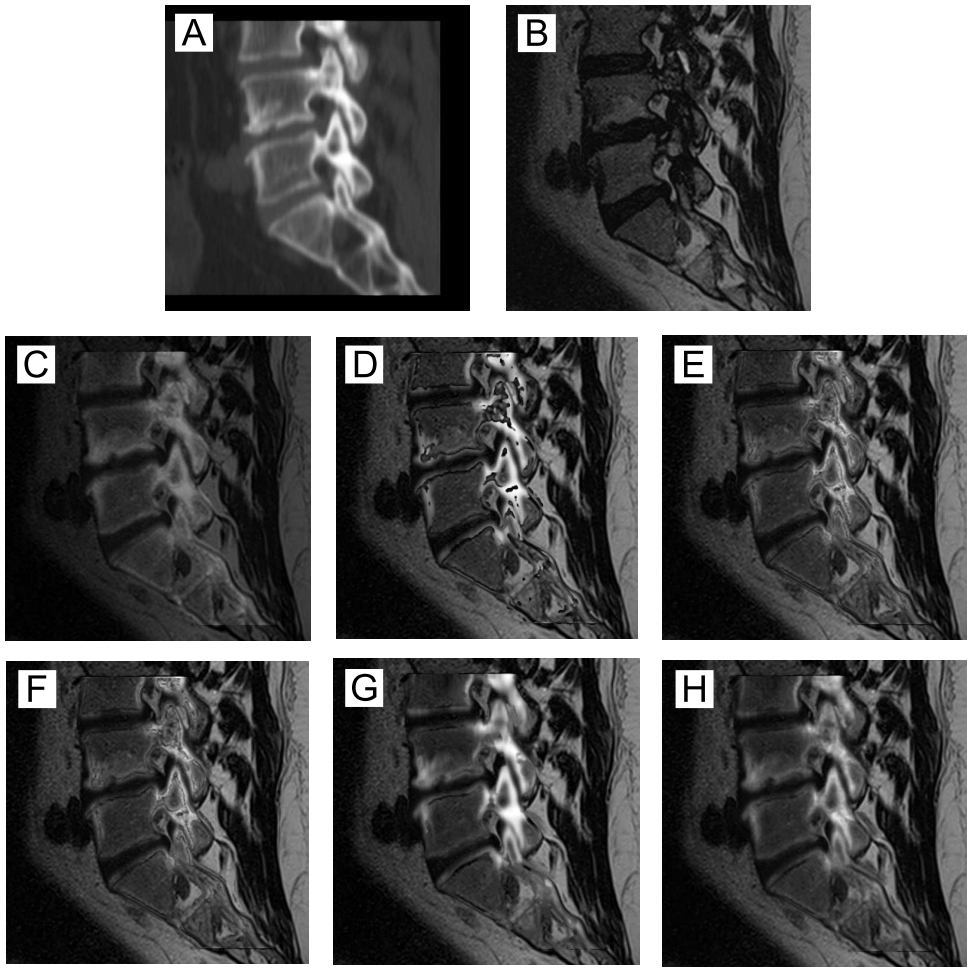


Figure 3.5: Fused sagittal slices showing the nerve roots of the patient, A: Input CT image, B: Input MR image, C: Fused averaging method, D: Fused CABS-2D, E: Fused CSD-2D, F: Fused graph cut method, G: Fused CABS-3D, H: Fused CABS-2D

CSD results, found in image E for 2D and image H for 3D (for both figures), are much clearer than the averaging method, and delineate the edge details clearly. Again the 3D version appears to have more smoothing than the 2D version. Finally, the graph cut results also show crisp edge details. Visually the CSD-2D and the graph cut 2D cases appear very similar.

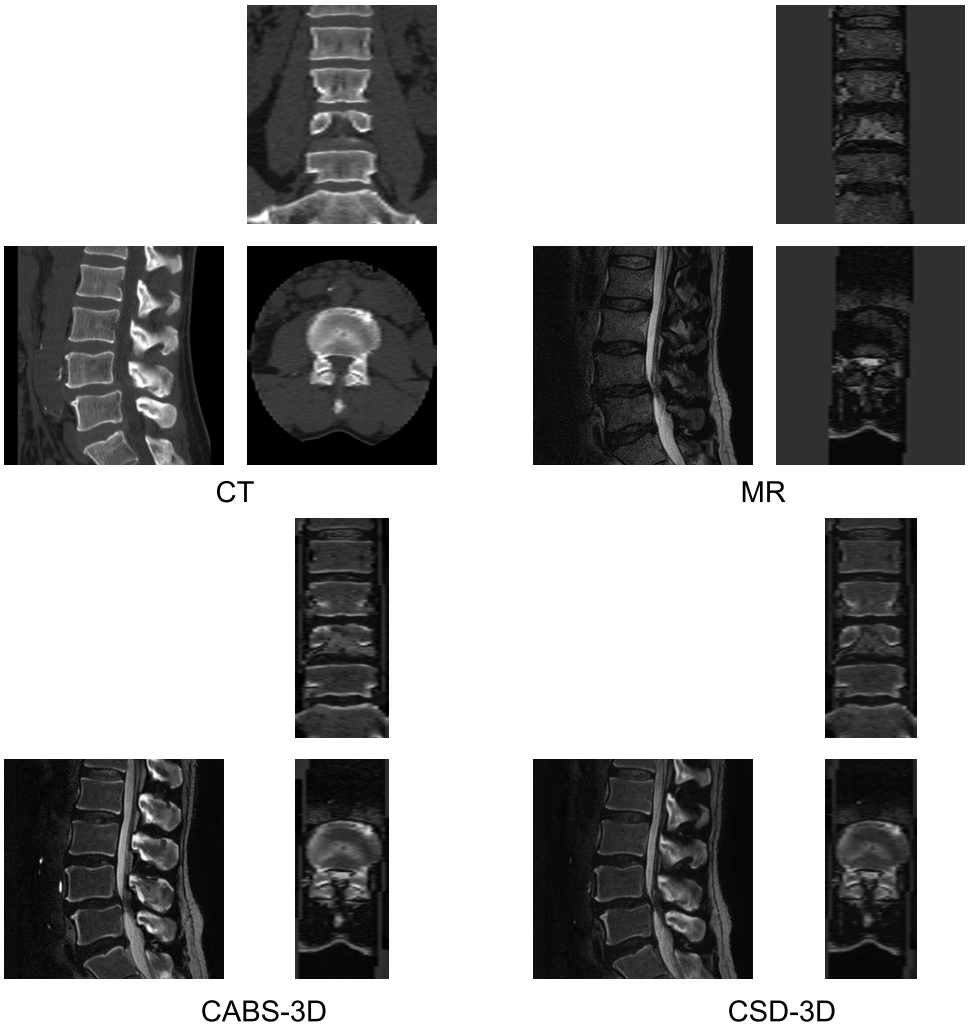


Figure 3.6: 3D fused images. The sets of groups are as labeled, with the CT and MR on the top and the CABS-3D and CSD-3D results below. Within each group of 3 images, the coronal image is on the top, the saggital view is on the bottom left and the axial is on the bottom right. This demonstrates the ability to fuse an image completely in 3D

For the 3D views shown in Fig. 3.6, the details in both sets of images are very clear, although there are more smoothing artifacts in the CABS-3D images than the CSD-3D images. This is most noticeable in the sagittal plane likely because of the pixel spacing differences

between the sagittal and other directions.

3.4.3 Validation of the Fusion Masks

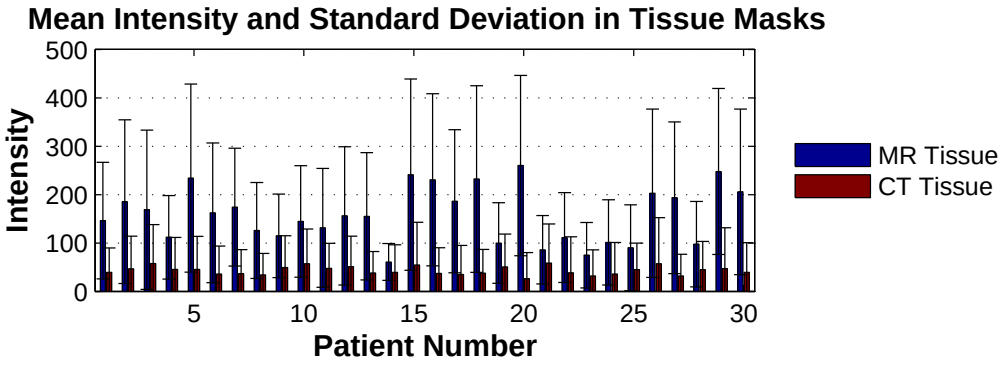
Bone and tissue masks were generated for each patient. For each patient, the mean intensity values in the MR and the preprocessed CT images for the two masks are shown with error bars for standard deviation in Fig. 3.7. In the tissue masks, the mean MR intensity values vary between 100 and 250, depending on the image. This is consistent with soft tissue detail. For the CT images the mean intensity is about 50 for all images. Again, this is consistent with minimal soft tissue detail present in the CT image and demonstrates that the tissue mask can differentiate between soft tissue detail in the MR and CT images. In the bone masks, the CT images have mean intensity values between 175 and 275, which is consistent with bone detail. In the same mask the MR values are in the range of 30-130. Additionally, for all images, the standard deviation was approximately proportional to the mean signal. These results demonstrate that the bone mask can differentiate between bone detail present in the CT image and a lack of bone detail in the MR image.

3.4.4 Statistical Evaluation

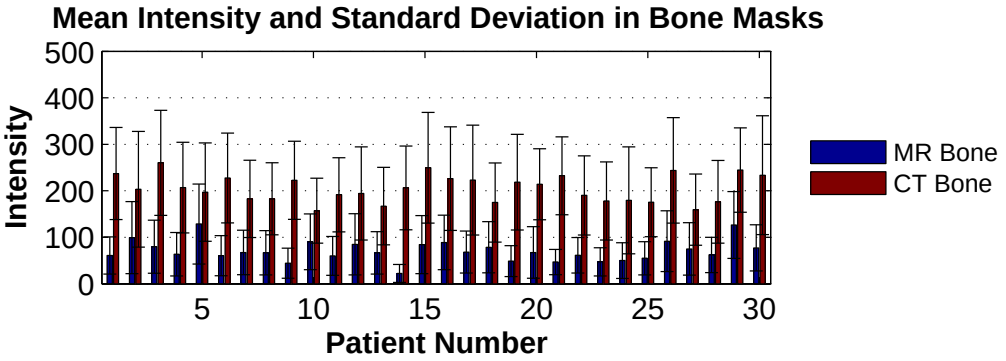
Based on the bone and tissue masks, the fusion errors: $e_{\text{MR,Tissue}}$, $e_{\text{CT,Tissue}}$, $e_{\text{CT,Bone}}$ and $e_{\text{MR,Bone}}$ were calculated for each patient and each of the six fusion methods. For our CABS and CSD methods, these errors are shown for the 2D algorithms in Fig. 3.8 and for the 3D algorithms in Fig. 3.9. These show the general trend that $e_{\text{MR,Tissue}} < e_{\text{CT,Tissue}}$ and $e_{\text{CT,Bone}} < e_{\text{MR,Bone}}$ for our methods.

The data for all six fusion methods was tested for normality. The results for the statistical significance of the data being normally distributed based on a Shapiro Wilks test [28] are shown in Table 3.3. Since some of the data was found to be non-normal ($p < 0.05$), further comparisons were made using a non-parametric Wilcoxon [31] test.

Using a pairwise Wilcoxon test [31] $e_{\text{MR,Tissue}}$ was compared to $e_{\text{CT,Tissue}}$ and $e_{\text{CT,Bone}}$ was



(a)

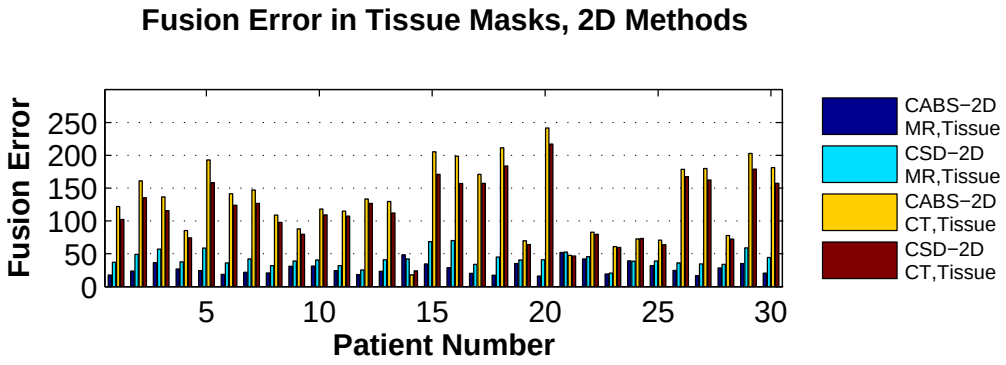


(b)

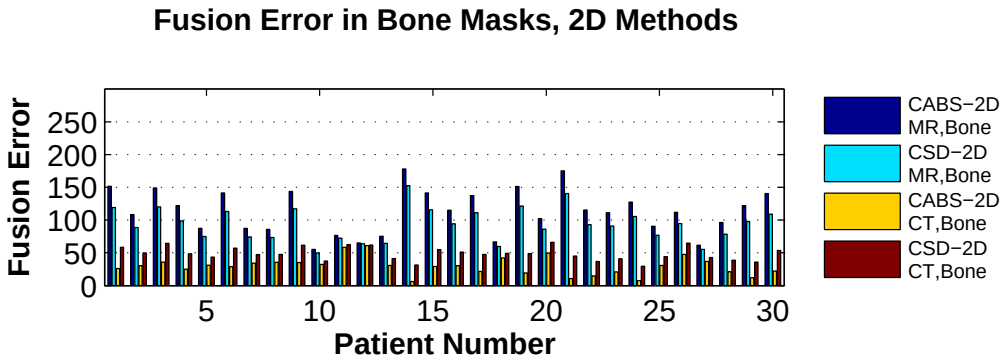
Figure 3.7: Mean image intensity values in the MR and preprocessed CT images prior to fusion, with one standard deviation given by the error bars, a) mean pixel intensities with standard deviations in the tissue mask, b) mean pixel intensities with standard deviations in the bone masks.

Table 3.3: Statistical significance of the normality tests of the fusion errors, (p-values for normality tests)

	P-$e_{\text{MR,Tissue}}$	P-$e_{\text{CT,Tissue}}$	P-$e_{\text{CT,Bone}}$	P-$e_{\text{MR,Bone}}$
Average	.455	.455	.324	.324
GC	.057	.701	.456	.759
CABS-2D	.022	.660	.362	.532
CSD-2D	.051	.696	.474	.763
CABS-3D	.732	.721	.369	.594
CSD-3D	.020	.851	.475	.836

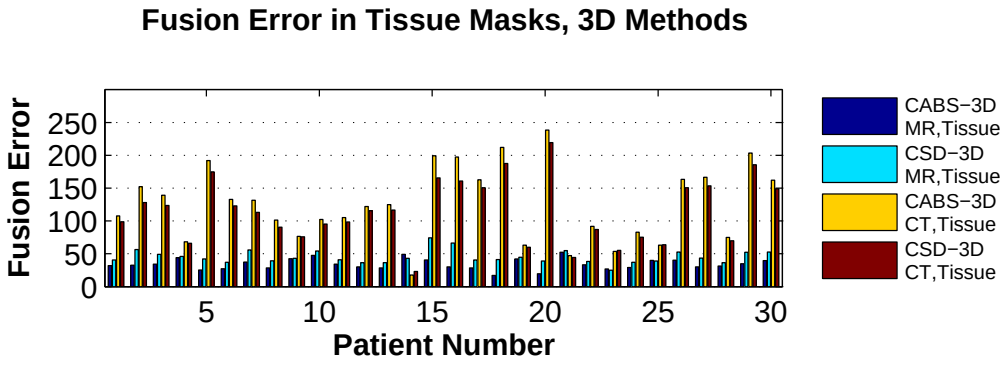


(a)

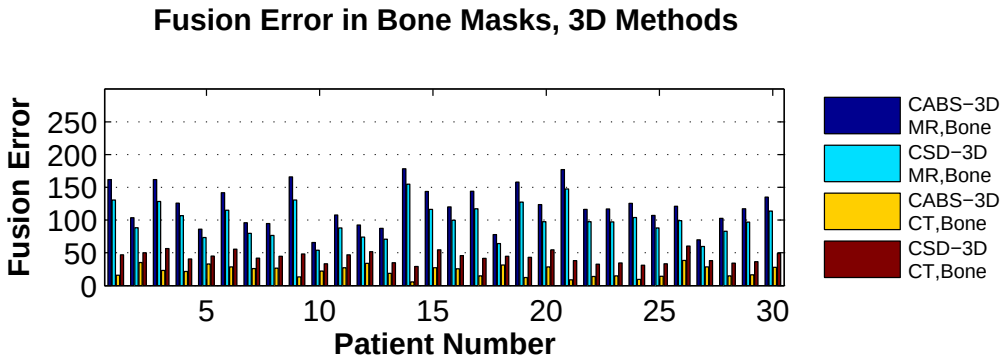


(b)

Figure 3.8: Fusion Error values for our 2D Fusion Methods, a) Tissue errors for MR and CT b) Bone errors for MR and CT.



(a)



(b)

Figure 3.9: Fusion Error values for our 3D Fusion Methods, a) Tissue errors for MR and CT b) Bone errors for MR and CT.

Table 3.4: Mean fusion errors and statistical significance of Wilcoxon test results for the six fusion methods.

	$e_{MR,Tissue}$	$e_{CT,Tissue}$	P - Value
Average	79.58 ± 26.08	79.58 ± 26.08	1
GC	42.68 ± 11.44	116.76 ± 47.32	< 0.001
CABS-2D	27.67 ± 9.47	131.69 ± 56.74	< 0.001
CSD-2D	42.41 ± 11.41	116.82 ± 47.25	< 0.001
CABS-3D	34.29 ± 8.28	125.19 ± 56.32	< 0.001
CSD-3D	45.26 ± 10.04	114.01 ± 48.19	< 0.001
	$e_{CT,Bone}$	$e_{MR,Bone}$	P - Value
Average	71.24 ± 13.83	71.24 ± 13.84	1
GC	48.95 ± 10.13	93.52 ± 25.23	< 0.001
CABS-2D	29.54 ± 13.45	113.08 ± 33.61	< 0.001
CSD-2D	48.73 ± 10.16	93.75 ± 25.21	< 0.001
CABS-3D	21.87 ± 8.70	120.82 ± 31.04	< 0.001
CSD-3D	43.21 ± 8.45	99.28 ± 25.40	< 0.001

compared to $e_{MR,Bone}$. Table 3.4 shows the mean values for these errors for the patient data sets, along with the p-values for the Wilcoxon test. In all cases except for the averaging method, the results were statistically significantly different. As expected for the averaging method, for a given mask, the errors were identical for the two input images. For the other five methods, the results show that the soft tissue detail is coming primarily from the MR images with the bone detail primarily coming from the CT images. In terms of relative comparisons, the CABS results are the best numerically with the lowest error in soft tissue and bone detail in both the 2D and 3D cases. As expected the averaging results are the worst. The graph cuts and CSD results are similar. The graph cuts have less error for the soft tissue and more error for the bony detail compared to the CSD results. Based on this analysis, the CABS results were optimal. Aside from the averaging method, the other results performed close to one another with less than a 10% difference between the ideal (the dynamic range was found to be about 700 pixel intensity levels for all images). These percentages are presented in Table 3.5.

Table 3.5: Fusion errors as a percentage of the image's dynamic range (700)

	$e_{\text{MR,Tissue}}$	$e_{\text{CT,Tissue}}$	$e_{\text{CT,Bone}}$	$e_{\text{MR,Bone}}$
Average	11%	11%	10%	10%
GC	6.1%	17%	7.0%	13%
CABS-2D	4.0%	19%	4.2%	16%
CSD-2D	6.1%	17%	7.0%	13%
CABS-3D	5.0%	18%	3.1%	17%
CSD-3D	6.5%	16%	6.2%	14%

Table 3.6: Mean computational times and iterations to convergence, of the 2D fusion methods, for a single sagittal slice. Convergence was set at $\epsilon < 10^{-4}$

	Time per slice(s)	Iterations to Convergence
GC - CPU	36	N/A
CABS-2D - CPU	207	329
CABS-2D - GPU	0.77	328
CSD-2D - CPU	8.2	21
CSD-2D - GPU	0.052	21

3.4.5 Runtime Analysis

For the 2D algorithms, the mean algorithm runtime per slice and the mean number of iterations for convergence are summarized in Table 3.6. These results were based on the mid-sagittal slices for each of the 30 patients. For the 3D algorithms, the mean total runtime, the mean algorithm runtime, the mean algorithm runtime per slice and the mean number of iterations to convergence are summarized in Table 3.7. These results were based on 3D volumes.

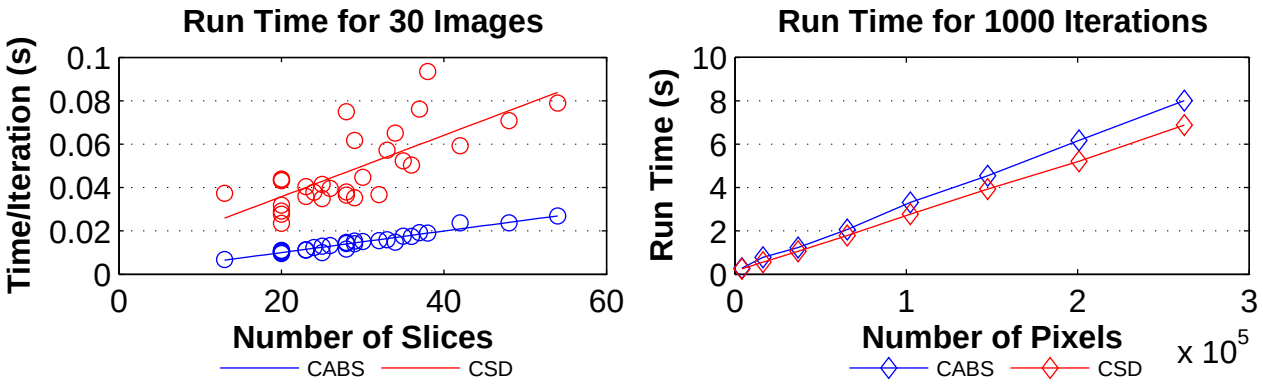
Based on the evaluation results shown in tables 3.6 and 3.7, the CSD method clearly has the shortest runtime in both the 2D and 3D cases. The graph cut method execution time is shorter on the CPU than the CABS method, but longer than the CSD method. The methods

Table 3.7: Mean computational times and iterations to convergence for a volume. Convergence was set at $\epsilon < 10^{-3}$

	Total Time (s)	Algorithm Time (s)	Computation Time per slice (s)	Iterations to Convergence
CABS-3D - CPU	~ 2100	~ 2100	~ 80	154
CABS-3D - GPU	2.1	1.7	0.074	154
CSD-3D - CPU	327	~ 327	10	10
CSD-3D - GPU	0.46	0.14	0.016	10

implemented on the GPU required the shortest time, with the CSD method being shorter than the CABS method, and both methods having results in real-time making them suitable for clinical applications.

Two different experiments were performed to demonstrate that the convex algorithms' runtimes are linear with respect to the number of pixels in the image. In the first experiment, the mean time to calculate one iteration of the algorithm was plotted vs the number of slices in the image, using a fixed slice size of 512x512 pixels. The images used were from the 30 patient datasets. As can be seen in Fig. 3.10 a) these results show a linear relationship between runtime and number of slices, for both the CABS-3D and the CSD-3D algorithm. For the second experiment, runtime was compared with image size variation in the x and y directions. Eight different images where used, having x and y sizes from $\{64 \times 64, 128 \times 28, 172 \times 712 \ldots 512 \times 512\}$ with 20 slices in the z direction, These were plotted versus the algorithm run time for 1000 iterations. Again, in both cases, a linear relationship was found see Fig. 3.10 b). Combined, the results in Fig. 3.10 show that the algorithm run time is linear with the number of pixels in the image.



(a) (b)

Figure 3.10: a) Runtime results for the CABS-3D and CSD-3D method, compared with slice number b) Runtime results for various image sizes for both the CABS-3D and CSD-3D methods.

Figure 3.11: Presence of osteophytes. Left: MR image, the osteophyte is not easily visible. Middle: fused image (CSD-3D), the osteophyte and the soft tissue details are clearly depicted. Right: CT image, the osteophyte is visible.

3.4.6 Case Studies

Four case studies of fused images are shown detailing a patient with osteophytes (formation of bony spurs at the margins of a joint) in the lumbar spine in Fig. 3.11, a lumbar spine patient with pathology near the L4 and L5 vertebrae in Fig. 3.12 and a patient with a fracture in their L1 vertebrae Fig. 3.13. The final case presents a patient with metastatics disease visible in the spine shown with multiple preprocessing settings to better highlight the pathology (see Fig. 3.14).

On the MR image shown in Fig. 3.11, the osteophyte is poorly seen and it is difficult to distinguish the osteophyte from the adjacent disc. On the CT image, the osteophyte is well seen but the disc is not. Both are well seen on the fused image and it is easy to distinguish the disc from the osteophyte.

In cases of lumbar spine disease (Fig. 3.12), the fused image can accurately show facet damage and a protruding disc on a single image. The T2 - weighted MR sequence clearly highlights a protruding disc between L4 and L5, and the nerve roots can also be seen above and below this disc. In the CT image, the facet joints between L4 and L5 can be clearly visualized. The fused image clearly presents all of these details and their relative locations as shown in Fig. 3.12.

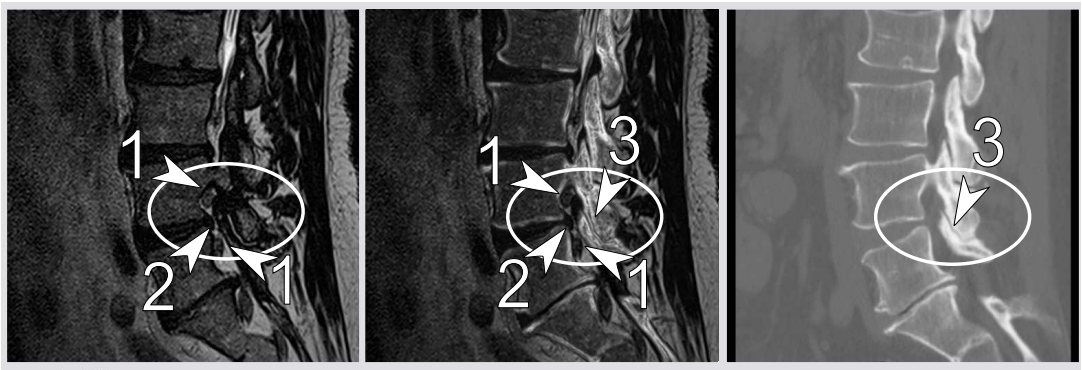


Figure 3.12: Patient with lumbar degenerative disc disease. Left: the MR image clearly depicts the bulging disc. Middle: the fused image (CSD-3D) depicting the bulging disc and facet joint arthrosis. Right: the CT image presents the facet arthrosis. 1. nerve roots, 2. bulging disc, 3. facet joint arthrosis.

In the third case (Fig. 3.13), vertebroplasty cement post fracture is visible in the CT image and appears only as a black mass in the MR image. The fused image (CSD-3D) shows a combined view of both the input images.

The final case study shows a patient with metastatic disease in the lumbar spine. Four different views are presented in Fig. 3.14 highlighting the utility of adjustable preprocessing. For each of the four views, the CT image is shown on the left, the fused image in the centre and the MR image is shown on the right. The metastatic disease is visible as bright white regions on the C6 - T2 vertebrae in the MR images on Figs. 3.14a - 3.14c. In Fig. 3.14a; the CT highlights the bone edges, while the MR image shows the disc, spinal cord and vertebrae detail. The fused image used the standard settings (see eq: (3.7)). Here the metastases are not visible, hence the fused image is of limited use. Fig. 3.14b has utilized an alternate preprocessing scheme for the fused image allowing the contrast changes due to the metastatases to be visible along with the cortical bone, spinal cord, nerve tissue and intervertebral disc. Fig. 3.14c utilizes fusion settings highlighting the cortical bone edges and metastases in particular. Finally, Fig. 3.14d shows a view of the patient through the nerve root bundle, with both the bone edges and nerve roots visible in the fused image. It is important to note that since the fusion occurs in real-time, the preprocessing settings can be interactively adjusted in a patient specific manner, as was

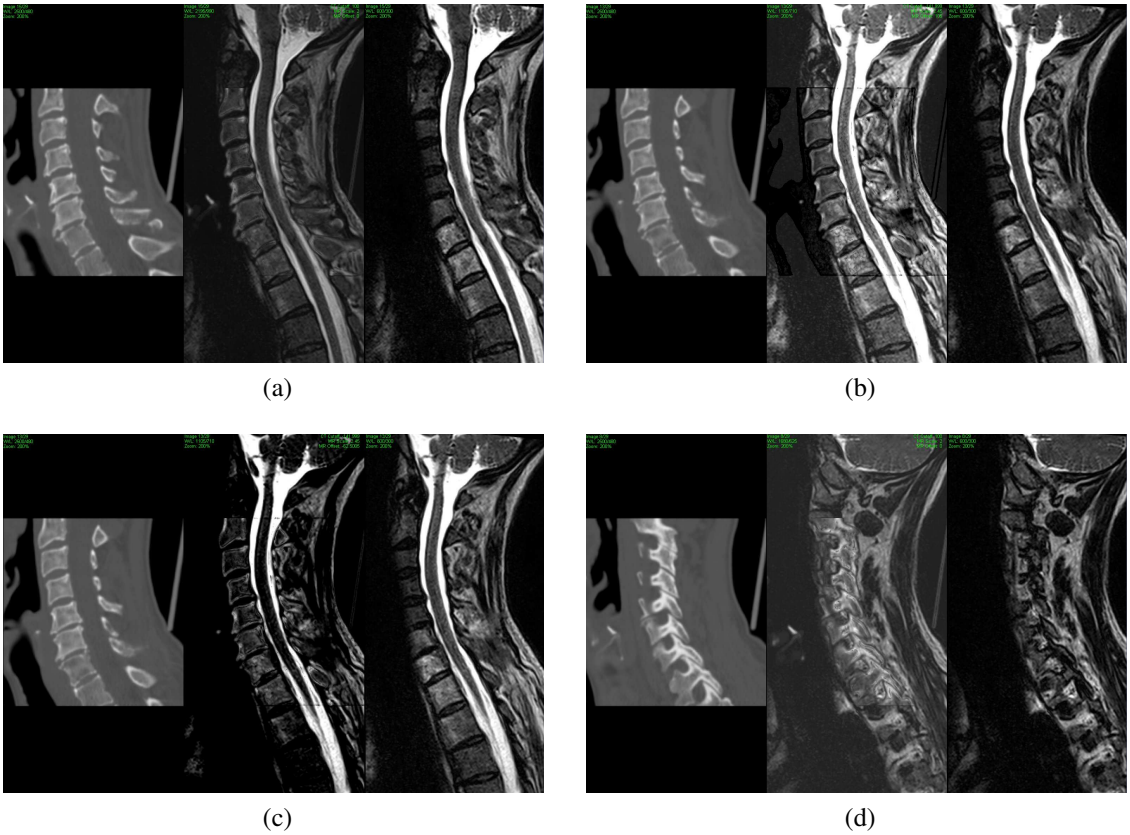
Figure 3.13: Fracture case: L1 compression fracture post vertebroplasty. On MR, vertebroplasty cement shows low signal which cannot be differentiated from disc or compressed bone. On CT, cement can be easily differentiated from disc and compressed bone, but the disc margins are not well delineated. The fused image differentiates and delineates all of these structures well.

done in this case.

These clinical examples illustrate the added benefit to diagnosis that these image fusion techniques offer to complement the original MR and CT images.

3.5 Conclusions and Discussions

We have presented four methods of image fusion based on convex energy minimization methods in 2D and 3D. We have achieved results with real time execution for the 2D and 3D images, while maintaining similar results to the state-of-the-art graph cut method based on a validation using images from 30 patients. The TRE was found to be 1.9 ± 0.6 mm, which is less than the voxel size. The in-plane TRE is about 1mm in the sagittal and coronal directions, but these values are limited by the FLE measurements which are also about 1mm. The registration algorithm provides sub-voxel accuracy on our data set, which is necessary for pixel-wise image fusion, otherwise misregistration would play a significant factor in the fusion. It should be noted that the subvoxel registration accuracy shows that for these cases the lumbar spine can be assumed to be a rigid body, and deformable registration is not needed. If a more flexible registration technique was utilized for images with significant deformations our fusion method



(a) (b)

(c) (d)

Figure 3.14: Four different views of a patient with metastatic disease. a) Fused image with normal settings where the metastases are not visible in the fused image, b) alternate settings highlighting the metastases in the fused image, c) additional settings highlighting both cortical bone and the contrast due to metastases in the fused image, d) view through the nerve root bundle, with the fused image showing both the nerve roots and the cortical bone.

could be applied to these registered images.

The fusion results show there is less than 10% difference in pixel intensity in the areas of interest (soft tissue including the spinal cord and the bone structures). The error for the CABS-2D method is the least, with similarly low errors for the CABS-3D method. In view of the visual results however, the CSD results and graph cuts appear better in this case.

For the mean tissue and bone values in the masks, shown in Fig. 3.7, there is a variation in soft tissue image intensity, especially for subject 14. Further inspection of the images, showed that there are no obvious abnormalities with any of the images, but rather for subject 14 there is lower overall signal, with a corresponding lower standard deviation.

The CSD-2D method is better than graph cut for the soft tissue detail and worse for the bony detail. Visually the graph cut results are similar to the CSD-2D results, being the clearest and showing the best edge detail. In terms of runtime, the CSD-3D requires the shortest time to converge of the 3D methods, with a mean runtime of 0.461 seconds on the GPU, while the CABS-3D method had a mean runtime of 2.06 s on the GPU. Based on execution time, the lack of smoothing artifacts and statistically good fusion results (although not the best), we have determined the CSD-3D method to be the most desirable, for further clinical testing.

The utility of real-time fusion has been demonstrated in the fourth case study (Fig. 3.14). In this case the standard fusion settings did not highlight the pathology on the fused image however, using alternate settings the pathology is clearly highlighted. It should be noted that all the views in the images in Fig. 3.14 were live screen captures from our interactive fusion viewer and the settings were determined interactively in real-time.

With this technique, physicians can see diseases such as, osteophyte growth, protruding discs and damaged facet joints on a fused image, as a complement to the original CT and MR images. This should allow for easier and more accurate diagnosis than would be available without the fused images. These results indicate that image fusion may be a suitable technique for clinical use and are congruent with feedback we have received from radiologists based on the fused images; however, an observer performance study is now required to establish the

utility of the technique.

3.6 Acknowledgements

B. Miles would like to acknowledge the financial support of this research from the Graduate Program in Biomedical Engineering at the University of Western Ontario and the support of the Computer Assisted Medical Intervention (CAMI) training program, which is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. A. Fenster holds a Canada Research Chair in Biomedical Engineering, and acknowledges the support of the Canada Research Chair Program. Both I. Ben-Ayed and S. Li are employees of GE Healthcare.

3.7 Appendix

Here we present a rigorous treatment of the fusion problem. Based on the three objectives in Section 3.2 We can express the fusion problem as the minimum of a functional of the following general form:

$$\min_{u} \; E(u) \; := \; \mathcal{D}(u, f_1, f_2) + \lambda \mathcal{S}(u)\,. \tag{3.8}$$

where $u : \Omega \to \mathbb{R}$ is an image function encoding the fusion solution, $f_1, f_2 : \Omega^2 \to \mathbb{R}^2$ are the input scans (MR and CT in this work), $\Omega \subset \mathbb{R}^n$ is the image domain, typically $\mathbb{R}^2$ or $\mathbb{R}^3$, and λ a positive constant.

Functional (3.8) includes two different types of terms:

- A data term $\mathcal{D}(u, f_1, f_2)$ (denoted as $\mathcal{D}(u)$ for convenience), encoding objectives 1) and 2), which measures an edge-weighted conformity of a solution u to input images f_1 or f_2

(MR and CT):

$$\mathcal{D}(u) \;=\; \int_\Omega w_1 A(f_1 - u)\, dx \;+\; \int_\Omega w_2 A(f_2 - u)\, dx\,, \tag{3.9}$$

with $A(\cdot)$ a positive convex function, (we will study different choices of $A(\cdot)$ later). $w_1 : \Omega \to \mathbb{R}$ and $w_2 : \Omega \to \mathbb{R}$ are two non-negative edge-weight functions verifying

$$w_1(x) \;+\; w_2(x) \;=\; 1\,, \quad \forall x \in \Omega\,, \tag{3.10}$$

and are defined as follows:

$$w_1(x) \;=\; \frac{g_1(x)}{g_1(x) + g_2(x)}\,, \quad w_2(x) \;=\; \frac{g_2(x)}{g_1(x) + g_2(x)}\,, \quad \forall x \in \Omega \tag{3.11}$$

where $g_i(x) = |\nabla(K * f_i)|$, $i = 1, 2$, and K is a kernel, for instance a Gaussian filter. w_1 and w_2 bias the solution towards strong edges in f_1 and f_2, respectively.

- A smoothness (total-variation) functional, $\mathcal{S}(u)$ encoding objective 3):

$$\mathcal{S}(u) \;=\; \int_\Omega |\nabla u|\, dx \tag{3.12}$$

This penalizes transitions in solution u, while preserving edge detail, encouraging neighboring pixels/voxels to have similar fused-image values, while allowing for transitions at edges.

In the following, we will study the corresponding optimization problems when the convex penalty function $A(u)$ is the L1-norm (absolute value) or L2-norm (squared difference), i.e., the corresponding data term is one of the following:

$$\textit{L1-norm problem: } \mathcal{D}_{\text{ABS}}(u) \;=\; \int_\Omega w_1\, |f_1 - u|\; dx \;+\; \int_\Omega w_2\, |f_2 - u|\; dx \tag{3.13}$$

or

$$\textit{L2-norm problem: } \mathcal{D}_{\mathrm{SD}}(u) = \int_\Omega w_1 (u - f_1)^2 \, dx + \int_\Omega w_2 (u - f_2)^2 \, dx . \tag{3.14}$$

Given the convexity of both functions $\mathcal{D}(u)$ and $\mathcal{S}(u)$ in (3.1), the proposed optimization problem for image fusion reduces to a convex optimization problem for which the global optimum exists, avoiding the gradient descent approaches of [30, 27] and graph cut approach of [22]. In the following we will show that: (1) The L1-norm solution can be obtained efficiently by solving an equivalent constrained problem via the augmented Lagrange multiplier method; and (2) The L2-norm solution can be obtained by a projection onto a convex set.

3.7.1 Equivalent Convex Formulation (Absolute Value)

We consider functional (3.1), with data term $\mathcal{D}(u)$ defined by (3.3). This functional shall be denoted E_{ABS} because of the absolute value data term it contains. Hence our problem can be written as:

$$\min_{u \in \Omega} \left\{ E_{\mathrm{ABS}} = \int_\Omega w_1 \, |f_1 - u| \, dx + \int_\Omega w_2 \, |f_2 - u| \, dx + \lambda \int_\Omega |\nabla u| \, dx \right\} \tag{3.15}$$

Proposition 1: *The unconstrained minimization problem (3.5) is equivalent to the following constrained optimization problem:*

$$\begin{aligned} &\max_{q_1, q_2} \max_{p \in C_\lambda} \int_\Omega q_1 f_1 dx + \int_\Omega q_2 f_2 dx \\ &\quad s.t. \;\; q_1(x) \in [-w_1(x), w_1(x)], \qquad q_2(x) \in [-w_2(x), w_2(x)]. \\ &\quad\quad q_1 + q_2 = div \; p \end{aligned} \tag{3.16}$$

where C_λ the set of continuously differentiable vector functions with compact support contained

in $\Omega \subset \mathbb{R}^n$, and having an L1-norm less than λ:

$$C_\lambda := \left\{ p | p \in C_c^1(\Omega, \mathbb{R}^n), |p(x)| \leq \lambda, \forall x \in \Omega \right\} \tag{3.17}$$

Note that variable function u does not appear in the constrained problem in (3.16). As we shall see shortly, u will act as a multiplier to equality constraint $q_1 + q_2 = div\ p$..

Proof of Proposition 1:

Proof The conjugate function A^* of $A = |x|$ can be computed by:

$$A^*(y) = \sup_{x \in \mathbf{dom} A} \{yx - |x|\} \tag{3.18}$$

where **dom** A is the domain of A, which is $\mathbb{R}$ in this case. Notice that $A^*(y) = 0\ \forall y \in [-1, 1]$ and is unbounded everywhere else. Refer to [4], section 3.3, for further details on conjugate functions.

Now, we define functions A_1 and A_2 as follows:

$$A_1(v) = A_1^{**}(v) = \max_{s_1} \{v s_1 - A_1^*(s_1)\}, \qquad A_2(v) = A_2^{**}(v) = \max_{s_2} \{v s_2 - A_2^*(s_2)\} \tag{3.19}$$

This results in a data term of the following form:

$$\begin{aligned} \mathcal{D}_{\text{ABS}} &= \max_{s_1, s_2 \in [-1,1]} \int_\Omega w_1 \left((f_1 - u) s_1 - A_1^*(s_1) \right) dx + \int_\Omega w_2 \left((f_2 - u) s_2 - A_2^*(s_2) \right) dx \\ &= \max_{s_1, s_2 \in [-1,1]} \int_\Omega w_1 f_1 s_1 dx + \int_\Omega w_2 f_2 s_2 dx - \int_\Omega u \left(w_1 s_1 + w_2 s_2 \right) dx \end{aligned} \tag{3.20}$$

The smoothness term $\mathcal{S}(u)$ can be expressed as follows [12]:

$$\mathcal{S}(u) = \lambda \int_\Omega |\nabla u| \, dx = \max_{p \in C_\lambda} \int_\Omega u \ \text{div}\ p \, dx \tag{3.21}$$

Substituting (3.20) and (3.21) into (3.5), we obtain:

$$\min_{u} \max_{s_1,s_2} \max_{p\in C_\lambda} \int_\Omega w_1 s_1 f_1 dx + \int_\Omega w_2 s_2 f_2 dx + \int_\Omega u \text{ div } p \, dx - \int_\Omega u\,(w_1 s_1 + w_2 s_2)\, dx$$
$$\text{s.t. } \; s_1(x) \in [-1, 1], \qquad s_2(x) \in [-1, 1]. \tag{3.22}$$

The original problem is now expressed as a minimization of maximums. For (3.22), the conditions of the minimax theorem hold (see [10], Chapter 6, Proposition 2.4). The constraints on s_1 and s_2 are convex, and the functional is linear in u, p, s_1 and s_2. This implies that the functional is convex l.s.c (lower semi-continuous) for fixed u and concave u.s.c (upper semi-continuous) for fixed p, s_1 and s_2. It follows that the order of the min and max operators can be interchanged [10]. Now, using variable changes $q_1 = w_1 s_1$ and $q_2 = w_2 s_2$, we further express (3.22) as the following constrained problem:

$$\max_{q_1,q_2} \max_{p\in C_\lambda} \min_{u} \int_\Omega q_1 f_1 dx + \int_\Omega q_1 f_2 dx + \int_\Omega u \text{ div } p \, dx - \int_\Omega u\,(q_1 + q_2)\, dx$$
$$\text{s.t. } \; q_1(x) \in [-w_1(x), w_1(x)], \qquad q_2(x) \in [-w_2(x), w_2(x)], \tag{3.23}$$

Now, by minimizing over u, we obtain:

$$\text{div } p - (q_1 + q_2) = 0 \tag{3.24}$$

Therefore, (3.23) can be formulated as the following constrained optimization problem:

$$\max_{q_1,q_2} \max_{p\in C_\lambda} \int_\Omega q_1 f_1 dx + \int_\Omega q_2 f_2 dx$$
$$\text{s.t. } \; q_1(x) \in [-w_1(x), w_1(x)], \qquad q_2(x) \in [-w_2(x), w_2(x)].$$
$$q_1 + q_2 = \text{div } p \tag{3.25}$$

Note that u can act as a multiplier to the equality constraint in (3.16). In view of this the

Lagrangian function of (3.16) is:

$$\max_{q_1,q_2}\max_{p\in C_\lambda}\min_{u}\int_\Omega q_1 f_1 dx+\int_\Omega q_1 f_2 dx+\int_\Omega u\ \text{div}\ p\ dx-\int_\Omega u\,(q_1+q_2)\,dx$$
$$\text{s.t.}\ \ q_1(x)\in[-w_1(x),w_1(x)],\qquad q_2(x)\in[-w_2(x),w_2(x)], \tag{3.26}$$

Therefore, defining the corresponding augmented Lagrangian function (c is a scalar):

$$L_c(q_1,q_2,p,u)=\int_\Omega q_1 f_1 dx+\int_\Omega q_2 f_2 dx+\int_\Omega u\,(\text{div}\ p-(q_1+q_2))\,dx-\frac{c}{2}\|\text{div}p-(q_1+q_2)\|^2 \tag{3.27}$$

we can solve (3.5) using the augmented Lagrangian algorithm [3], as shown in algorithm 1. Variables q_1, q_2 and p_k are first optimized pointwise by a projection onto their associated convex sets. We define **Proj** as constraining a value to be within a given convex set. For scalar variable q_1, $\mathbf{Proj}_{|q_1|\le w_1}$ results in simply restricting the final value to be within $[-w_1,w_1]$. i.e.

$$\mathbf{Proj}_{|q_1|\le w_1}(q_1)=\begin{cases}-w_1 & \text{if } q_1<-w_1\\ q_1 & \text{if } -w_1\le q_1\le w_1\\ w_1 & \text{if } w_1<q_1\end{cases} \tag{3.28}$$

For vector variable p, function $\mathbf{Proj}_{C_\lambda}$ results in constraining the vector to lie within a ball or radius λ:

$$\mathbf{Proj}_{C_\lambda}(p)=\begin{cases}p & \text{if } \|p\|\le\lambda\\ \lambda\frac{p}{\|p\|} & \text{if } \|p\|>\lambda\end{cases} \tag{3.29}$$

With the use of a splitting method, we proceed by first optimizing the inner variables, q_1, q_2, p. Variables q_1 and q_2 can be updated pointwise at each iteration, while p can be updated

with Chambolle's projection [7], which is solved iteratively by:

$$p^{i+1} = \mathbf{Proj}_{C_\lambda}\left(p^i + \tau\nabla\left\{\mathrm{div}p^i - \left(\left(q_1^{i+1} + q_2^{i+1}\right) + u^i/c\right)\right\}\right) \tag{3.30}$$

where τ is the step size of each iteration. Finally u is updated, repeating until convergence. This is shown in algorithm 1. The solution will converge with only one iteration of (3.30) within each main loop of the iterations. Note that this optimization can be completed in parallel for each point in the image. This algorithm is easily implemented in both 2D and 3D. It can also be easily parallelized and executed on a graphics card.

3.7.2 Equivalent Convex Formulation (Squared Differences)

We now consider the data term in (3.4), which gives the following minimization problem:

$$\min_{u:\Omega}\left\{E_{\mathrm{SD}} = \frac{1}{2}\int_\Omega w_1(u-f_1)^2dx + \frac{1}{2}\int_\Omega w_2(u-f_2)^2dx + \lambda\int_\Omega |\nabla u|\,dx\right\} \tag{3.36}$$

To minimize this functional, first let us consider the following proposition:

Proposition 2: The problem in (3.6) is equivalent to the following optimization problem:

$$\min_{p\in C_\lambda}\int_\Omega (w_1f_1 + w_2f_2 - \mathrm{div}p)^2\,dx \tag{3.37}$$

Proof of Proposition 2:

Proof Combining (3.6) and (3.21) leads to the formulation

$$\min_u\max_{p\in C_\lambda}\frac{1}{2}\int_\Omega w_1(u-f_1)^2dx + \frac{1}{2}\int_\Omega w_2(u-f_2)^2dx + \int_\Omega u\ \mathrm{div}\ p\,dx \tag{3.38}$$

Since this functional satisfies the conditions of the minimax theorem [10], we can change the

Algorithm 1: L1-norm Optimization (E_{ABS})

For each k-iteration,

1. Optimize q_1^{k+1} by fixing q_2^k, p^k and u^k, giving:

$$\begin{aligned} q_1^{k+1} &= \underset{|q_1(x)|\leq w_1(x)}{\arg\max}\ L_c(q_1^k, q_2^k, p^k, u^k) \\ &= \mathbf{Proj}_{|q_1(x)|\leq w_1(x)}\left(f_1/c + \left(\text{div } p^k - q_2^k(x) - u^k/c\right)\right) \end{aligned} \tag{3.31}$$

   which can be solved pointwise.

2. Optimize q_2^{k+1} by fixing q_1^{k+1}, p^k and u^k, giving:

$$\begin{aligned} q_2^{k+1} &= \underset{|q_2(x)|\leq w_2(x)}{\arg\max}\ L_c(q_1^{k+1}, q_2^k, p^k, u^k) \\ &= \mathbf{Proj}_{|q_2(x)|\leq w_2(x)}\left(f_2/c + \left(\text{div } p^k - q_1^{k+1}(x) - u^k/c\right)\right) \end{aligned} \tag{3.32}$$

   which can be solved pointwise.

3. Optimize p^{k+1} by fixing q_1^{k+1}, q_2^{k+1} and u^k, giving:

$$p^{k+1} = \underset{p\in C_\alpha}{\arg\min} \left\| \text{div } p - \left(q_1^{k+1} + q_2^{k+1} + u^k/c\right)\right\|^2 \tag{3.33}$$

   which can be iteratively solved, by projecting $\left(q_1^{k+1} + q_2^{k+1} + u^k/c\right)$ onto C_λ using:

$$p^{i+1} = \mathbf{Proj}_{C_\lambda}\left(p^i + \tau\nabla\left\{\text{div} p^i - \left(\left(q_1^{i+1} + q_2^{i+1}\right) + u^i/c\right)\right\}\right) \tag{3.34}$$

4. Update u^{k+1} by

$$u^{k+1} = u^k + c\left(q_1^{k+1} + q_2^{k+1} - \text{div } p^{k+1}\right) \tag{3.35}$$

   which can be solved pointwise.

5. let $k = k + 1$

6. repeat until the solution converges

order of optimization. Rearranging, and using $w_1 + w_2 = 1$, (3.38) becomes:

$$\max_{p \in C_\lambda} \min_{u} \int_\Omega \left\{ \frac{u^2}{2} - u\,(w_1 f_1 + w_2 f_2) + \frac{w_1 f_1^2 + w_2 f_2^2}{2} + u \text{ div } p \right\} dx \tag{3.39}$$

Defining $f = w_1 f_1 + w_2 f_2$, and remembering that $w_1 f_1^2 + w_2 f_2^2$ is a constant relative to u (and to p), we obtain:

$$\max_{p \in C_\lambda} \min_{u} \int_\Omega \left\{ \frac{u^2}{2} - uf + u \text{ div } p + \text{const} \right\} dx \tag{3.40}$$

From setting the derivative of the integrand in (3.40) equal to zero and noting the convexity of (3.40), we obtain the optimal value of u, denoted by u^*:

$$\begin{aligned} u^* - f + \text{div} p &= 0 \\ u^* &= f - \text{div} p \end{aligned} \tag{3.41}$$

Rearranging (3.40) gives:

$$\max_{p \in C_\lambda} \min_{u} \int_\Omega \left\{ u \left(\frac{u}{2} - f + \text{ div } p \right) + c \right\} dx \tag{3.42}$$

Substituting u^* from (3.41), results in:

$$\begin{aligned} &\max_{p \in C_\lambda} \int_\Omega \left\{ (f - \text{ div } p) \left(-\frac{f - \text{ div } p}{2} \right) + c \right\} dx \\ &= \max_{p \in C_\lambda} \int_\Omega \left\{ -\frac{1}{2} (f - \text{ div } p)^2 + c \right\} dx \\ &= \min_{p \in C_\lambda} \int_\Omega (w_1 f_1 + w_2 f_2 - \text{div } p)^2 \, dx \end{aligned} \tag{3.43}$$

Using the result in proposition 2, our problem reduces to a projection of the weighted image onto the convex set C_λ. This can be solved efficiently via Chambolle's projection algorithm [7],

by iterating:

$$p^{i+1} = \mathbf{Proj}_{C_\lambda}\left(p^i + \tau\nabla\left\{(w_1 f_1 + w_2 f_2) - \mathrm{div}p^i\right\}\right) \tag{3.44}$$

This algorithm is easily implemented in both 2D and 3D. It can also be easily parallelized and executed on a graphics card, resulting in a very efficient solution.

3.7.3 Extension to n-image fusion

The formulations of section 3.7.2 can be extended to n images. An application of such an extension would be the fusion of T1-weighted MR, T2-weighted MR and CT scans of the spine. In this case, the multi-image fusion problem becomes:

$$\min_{u:\Omega\to\mathbb{R}} \frac{1}{2}\sum_1^n \int_\Omega w_n(u - f_n)^2 dx + \alpha\int_\Omega |\nabla u|\, dx \tag{3.45}$$

where f_n are the input images, and the weights w_n are extended as follows:

$$\begin{aligned} g_n &= |\nabla f_n| * K \\ w_n &= \frac{g_n}{\sum_1^n g_n} \end{aligned} \tag{3.46}$$

Proposition 3: The problem in (3.45) is equivalent to the following optimization problem:

$$\min_{p\in C_\lambda} \int_\Omega \left(\sum_n w_n f_n - \mathrm{div}p\right)^2 \tag{3.47}$$

The proof follows the same steps as the proof of Proposition 2 (see appendix). Similarly to section 3.7.2, our problem reduces to a projection of the weighted image onto the convex set C_λ. This can be solved efficiently via the Chambolle's projection algorithm [7], by iterating:

$$p^{i+1} = \mathbf{Proj}_{C_\lambda}\left(p^i + \tau\nabla\left(\sum_n w_n f_n - \mathrm{div}p^i\right)\right) \tag{3.48}$$

where τ is the step size of each iteration.

Bibliography

[1] Insight ToolKit. URL `http://www.itk.org/`.

[2] Coloma Ballester, Vicent Caselles, Laura Igual, and Joan Verdera. A Variational Model for P+XS Image Fusion. *International Journal of Computer Vision*, 69(1):4358, 2006.

[3] Dimitri P. Bertsekas. *Constrained optimiaztion and Lagrange multiplier methods*. Academic Press Inc, 1982.

[4] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2009.

[5] Yuri Boykov and Vladimir Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 26(9):1124–1137, September 2004.

[6] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 23(11):1222–1239, November 2001.

[7] Antonin Chambolle. An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):8997, 2004.

[8] Arnaud Colin and Jean-Yves Boire. MRISPECT fusion for the synthesis of high resolution 3D functional brain images: a preliminary study. *Computer Methods and Programs in Biomedicine*, 60:107–116, 1999.

[9] Richard A. Deyo, Sohail K. Mirza, Judith A. Turner, and Brook I. Martin. Overtreating Chronic Back Pain: Time to Back Off? *Jour Am Board Fam Med*, 22(1):62–68, 2009.

[10] Ivar Ekeland and Roger Teman. *Convex analysis and variational problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999.

[11] J. Michael Fitzpatrick, Jay B. West, and Calvin R. Maurer. Predicting Error in Rigid-Body Point-Based Registration. *IEEE Transactions on Medical Imaging*, 17(5):694–702, 1998.

[12] Enrico. Giusti. *Minimal surfaces and functions of bounded variation*. Australian National University, Canberra, 1977.

[13] Paul Hill, Nishan Canagarajah, and David R Bull. Image Fusion Using Complex Wavelets. In *BMVC 2002*, 2002.

[14] Yangqiu Hua, Sohail K. Mirzac, Jeffrey G. Jarvikb, Patrick J. Heagerty, and David R. Haynor. MR and CT image fusion of the cervical spine: a noninvasive alternative to CT-Myelography. In *Proceedings of SPIE*, volume 5744, Bellingham, WA, 2005. SPIE.

[15] C. A. Karlo, I. Steurer-Dober, M. Leonardi, C. W. A. Pfirrmann, M. Zanetti, and J. Hodler. MR/CT image fusion of the spine after spondylodesis: a feasibility study. *European Spine Journal*, 19:17711775, 2010.

[16] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *ECCV*, volume 1, pages 332–345, 2008.

[17] John J Lewis, Robert J. O'Callaghan, Stavri G. Nikolov, David R Bull, and NIshan Canagarajah. Pixel - and region-based image fusion with complex wavelets. *Information Fusion*, 8:119–130, 2007.

[18] H. Li, B. S. Manjunath, and S. K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, 1995.

[19] Shangzheng Liu, Jiuqiang Han, Bowen Liu, and Xinman Zhang. Novel image fusion algorithm based on human perception. *Optical Engineering*, 48(4):047002 1–6, April 2009.

[20] G. K. Matsopoulos, S. Marshall, and J. N. H. Brunt. Multiresolution morphological fusion of MR and CT images of the human brain. *IEE Proceedings on Vision Image Signal Processing*, 141(3):137–142, 1994.

[21] David Mattes, David R. Haynor, Hubert Vesselle, Thomas K. Lewellen, and William Eubank. PET-CT Image Registration in the Chest Using Free-form Deformations. *IEEE Transactions on Medical Imaging*, 22(1):120–128, January 2003.

[22] B. Miles, I. Ben Ayed amd M. W. K. Law G. Garvin, A. Fenster, and S. Li. Spine image fusion via graph cuts. *IEEE Transactions On Biomedical Engineering*, 60(5):1841–50, July 2013.

[23] Nikolaos Mitianoudis and Tania Stathaki. Pixel-based and region-based image fusion schemes using ICA bases. *Information Fusion*, 8(2):131142, 2007.

[24] Filippo Nencini, Andrea Garzelli, Stefano Baronti, and Luciano Alparone. Remote sensing image fusion using the curvelet trasnform. *Information Fusion*, 8:143–156, 2007.

[25] Jorge Núñez, Xavier Otazu, Octavi Fors, Albert Prades, Vicenç Palà, and Román Arbiol. Multiresolution-Based Image Fusion with Additive Wavelet Decomposition. *IEEE Transactions On Geoscience And Remote Sensing*, 37(3):1204–1211, May 1999.

[26] Xavier Otazu, Mara Gonzlez-Audcana, Octavi Fors, and Jorge Nez. Introduction of Sensor Spectral Response Into Image Fusion Methods. Application to Wavelet-Based Methods. *IEEE Transactions On Geoscience And Remote Sensing*, 43(10):2376–2385, October 2005.

[27] Gemma Piella. Image Fusion for Enhanced Visualization: A Variational Approach. *International Journal of Computer Vision*, 83(1):1–11, 2009.

[28] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591611, 1965.

[29] Moon-Jun Sohn, Dong-Joon Lee, Sang Won Yoon, Hye Ran Lee, and Yoon Joon Hwang. The effective application of segmental image fusion in spinal radiosurgery for improved targeting of spinal tumours. *Acta Neurochir 151:231238*, 151:231–238, 2009.

[30] Wei-Wei Wang, Peng-Lang Shui, and Xiang-Chu Feng. Variational Models for Fusion and Denoising of Multifocus Images. *IEEE Signal Processing Letters*, 15:65–68, 2008.

[31] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6): 80–83, December 1945.

[32] Stephen T. C. Wong, Robert C. Knowlton, Randy A. Hawkins, and Kenneth D. Laxer. Multimodal Image Fusion for Noninvasive Epilepsy Surgery Planning,. *IEEE Transactions on Computer Graphics and Applications*, 16(1):30–38, 1996.

[33] Xiangnan Xu, Hua Li, and A Wang. The application of BEMD to multi-spectral image fusion. In *Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition*, 2007.

[34] L. Yang, B.L.Guo, and W.Ni. Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform. *Neurocomputing*, 72:203–211, 2008.

[35] Zheng Youzhi, Qin Zheng, and Yang Jingyu. Image Fusion Using a Hybrid Representation of Empirical Mode Decomposition and Contourlet Transform. In *Proceedings of Information Science and Technology (ICIST)*, 2011.

[36] Jing Yuan, Juan Shi, and Xue-Cheng Tai. A Convex AND Exact Approach to Discrete

Constrained TV-L1 Image Approximation. Technical Report CAM-10-51, UCLA, CAM, UCLA, 2010.

[37] Jing Yuan, Juan Shi, Xue-Cheng Tai, and Yuri Boykov. A Study on Convex Optimization Approaches to Image Fusion. In *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*, 2011.

[38] Guangming Zhang, Zhiming Cui, Fanzhang Li, and Jian Wu. DSA Image Fusion based on Dynamic Fuzzy Logic and Curvelet Entropy. *Journal of Multimedia*, 4(3):129–136, June 2009.

Chapter 4

A Clinical Evaluation of Fused MR/CT images for Spine Diagnosis

4.1 Introduction

The information on magnetic resonance (MR) and computed tomography (CT) of the lumbar spine is complementary [4]. By fusing these two modalities it is possible to create a single image that combines the bony detail highlighted in the CT scan and the soft tissue structures that are more clearly visualized in the MR images [4]. This enhanced view will allow the interpreting clinician to localize the pathologies on a single image allowing for better characterisation of the interactions between bone and soft tissues, particularly at bony margins. The intent is to create an additional view that adds value for the interpreting physician, rather than a replacement for both source views.

Image fusion techniques have been applied to liver ablation [3] and to MRI SPECT images [1], where techniques such as the addition of images will suffice. However the spine presents a difficult technical challenge because of the complexity of the two image sets being fused. We have created a system for real-time image fusion designed and proven [4] to preserve soft tissue detail from the MR image and cortical bone detail from the CT images based on novel

algorithms. Interactive controls allow the clinician to adjust the fused image via, 1) CT soft tissue removal, based on removing soft tissues below a given Houndsfield unit, and 2) changing the relative intensity of the MR image compared to the CT image. An example of a fused image is shown in Fig. 4.1.

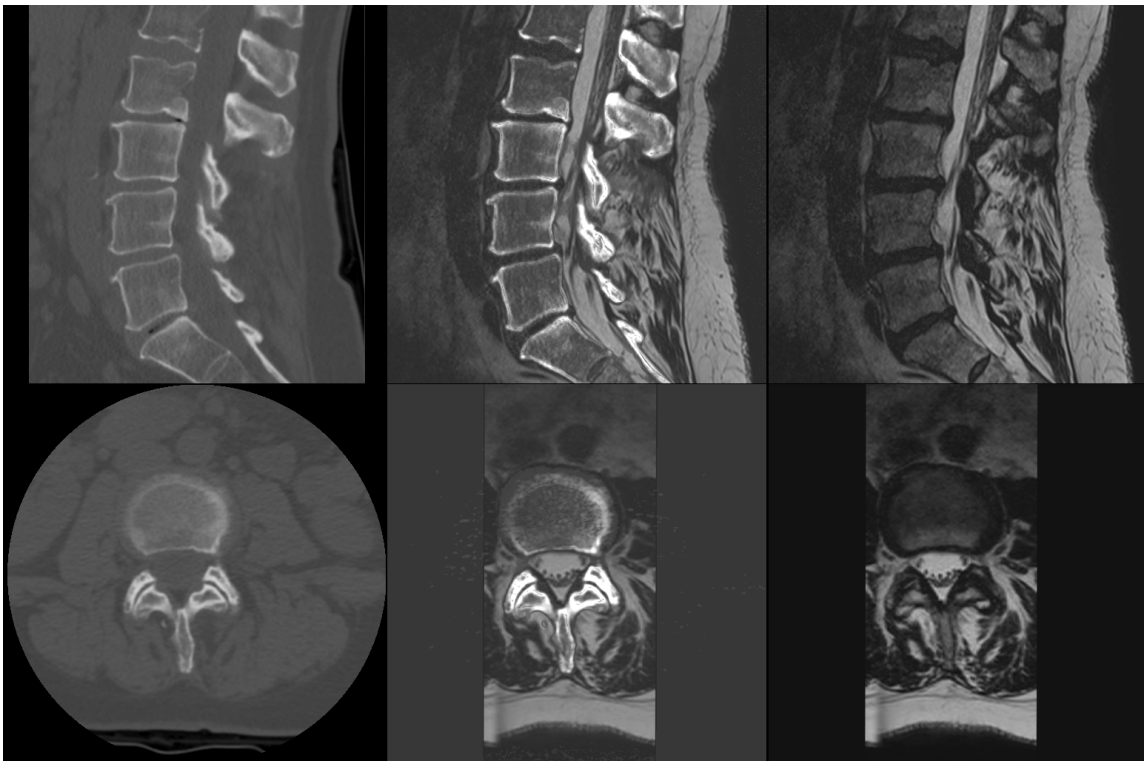


Figure 4.1: Comparison of CT, MR and Fused images. Top: sagittal lumbar spine images, Bottom, axial lumbar spine slices. Left: CT, Center: new fused images, Right: T2-weighted MR image.

It was hypothesized that employing these fused images for spine diagnosis compared to unregistered MR and CT images, radiologists would have a higher confidence in their diagnosis, have a lower inter-observer variability in measuring protrusions and spend less time on diagnosis than without using fused images.

4.2 Materials and Methods

4.2.1 Data Selection

Institutional ethics review board approval was obtained for the use of data from 20 patients who had previously had imaging preformed. Individual consent was not needed as per the approval of the ethics review board. To obtain a routine set of clinical images, MR lumbar spine images were collected from the last 20 sequential patients at St. Joseph's Health Care London who also had a lumbar CT scan within a year of their lumbar MR scan. The corresponding CT images were also collected. Thoracic and cervical spine images were not included in the 20 subjects as their lower prevalence would not provide enough statistical power to provide meaningful results. The MR images were acquired using a SIEMENS 1.5T Avanto (Seimens AG, Erlangen, Germany). These were 3D T2 weighted images acquired using the SPACE pulse sequence with variable slice thickness and number of slices. The CT images came from a variety of scanners at multiple sites with variable slice thickness and number of slices.

Of the 20 images, 15 were then registered and fused using the methods of Miles et. al [4] and Yuan et. al. [5]. The other five images could not be registered due to non-rigid deformations or poor image quality such as motion artifacts and were omitted from the final analysis.

4.2.2 Reader Selection

The images were read by five radiologists, with varying levels of experience. Reader #1 had 2 years of experience since completion of residency including a one year fellowship in musculoskeletal imaging and 1 year of breast imaging, Radiologist #2 had 5 years of post-residency experience with fellowship training in Breast Imaging and Cross Sectional Imaging. Radiologist #3 had 5 years of post-residency experience including a 1 year Women's and abdominal radiology fellowship and a 2 year breast and cross sectional imaging fellowship. Radiologist #4 had 15 years of post-residency experience and completed a fellowship in cross-sectional

and women's imaging. Radiologist #5 has 9 years of experience including a one year Neuroradiology subspecialty training followed by an 18 month Neuroradiology fellowship.

4.2.3 Evaluation of Images

To evaluate the suitability of fused images for use in clinical practice, a comparison was made between current practice using unregistered CT and MR images and the proposed use of registered CT, MR and fused images. Evaluation occurred in two phases. In the first phase, unregistered CT and MR images were evaluated. In the second phase, registered CT, MR and fused images were evaluated. In order to eliminate memory bias, there was atleast a 4 week interval between the first and second phase for each reader. Additionally, to blind the participants, they were not told about the purpose of the study until the second phase of reading.

The images were evaluated based on two tasks: 1) ranking of the images and 2) measurement of protrusions. For the rankings, the readers were asked their confidence in assessing: 1) disc contours, 2) disc protrusions, 3) spinal stenosis, 4) foraminal stenosis, 5) exiting nerve roots, 6) the facet joints (based on osteoarthritis, synovial cysts, edema, erosions, and ankylosis), 7) ligamentum flavum, 8) differentiating between disc and osteophyte, 9) making the overall correct diagnosis and 10) whether in aggregate the images were aesthetically pleasing. The ranking was done using a 1-5 scale with one being no confidence and five being extremely confident. These rankings were made for each subject and each modality provided (i.e., MR, CT for the first phase of testing and MR, CT and fused for the second phase of testing).

The readers were also asked "to determine the greatest sagittal and transverse dimension in mm of the largest extrusion/protrusion, not including osteophytes, providing the best measurement that would be suitable in a radiological report." They were provided with the definition of protrusions/extrusions given in [2]. They were also provided with the level of the protrusion, which had been previously identified by a senior radiologist with 19.5 years of experience including a one-year fellowship in musculoskeletal radiology and 6 months fellowship in interventional radiology, who was not included in the 5 reviewing radiologists. The readers were

given the choice of using either the CT or MR for the measurement in phase 1 and were instructed to use the fused image for phase 2.

4.2.4 Software Viewer

Custom software designed to provide similar functionality to GE PACS radiology software (GE Healthcare, Milwaukee, Il) was used for the testing. All the readings and measurements were timed by the software to determine how long the participants spent on the given cases. Both sagittal and axial views were presented to the radiologists as shown in 4.1.

4.2.5 Statistical Analysis Methods

All data was analysed using SPSS Statistics 21 (IBM Corp, Armonk, New York) statistical significance was set at $P < 0.05$. Comparisons of the ranking results for each questions was conducted first using a Friedman test to determine if there was a significant difference in ranks, followed by non-parametric Wilcoxon tests between modalities, with statistical significance set at $P < 0.01$ to account for multiple comparisons.

The measurement results were imputed to account for missing data. This was determined to be valid as less than 15% of the data was missing and the changes in results of further statistical tests were small between the imputed and non-imputed data. Consistency of measurements was determined using a two-way mixed single measures intra-class correlation based on absolute agreement. This was computed for all reviewers together with agreement defined to be an ICC value greater than .8. Additionally, Levene's test was conducted to analyze whether the variance was different between the two phases of the study for individual reviews and all reviewers.

Time results were also imputed. Again this was determined to be valid as less than 15% of the data was missing and the changes in effect size were small between the imputed and non-imputed data. The imputed results were analysed using a two-way repeated measures ANOVA to test for differences between readers and between phases of the study.

4.3 Results

4.3.1 Ranking Results

The average rankings with standard error for the five readers in a given phase of the study and modality are presented in Table 4.1. A comparison was made between the fused images and the other modalities in both phases of the study to determine whether there were significant differences between rankings. Based on Friedman tests there was a significant difference in ranks between modalities $P < .001$ for all 10 questions. Individual modalities were compared and the P values are contained in Table 4.2, with any significant results marked with a *.

Table 4.1: Average confidence ratings over the 5 readers for the given anatomy/question and for the given modality and phase of the study.

Question	Phase1		Phase 2		
	CT	MR	CT	MR	Fused
Disc Contours	3.30 ± .32	4.37 ± .18	3.43 ± .28	4.70 ± .09	4.37 ± .23
Disc Protrusions	3.20 ± .44	4.40 ± .19	3.30 ± .29	4.57 ± .13	4.33 ± .18
Spinal Stenosis	3.93 ± .26	4.73 ± .08	3.93 ± .29	4.77 ± .06	4.63 ± .17
Foraminal Stenosis	3.87 ± .20	4.57 ± .16	3.60 ± .19	4.63 ± .12	4.43 ± .20
Exiting Nerve Root	3.47 ± .20	4.43 ± .12	3.07 ± .29	4.53 ± .16	4.13 ± .23
Facet Joints	4.33 ± .04	4.37 ± .06	4.07 ± .11	4.27 ± .11	4.27 ± .11
Ligamentum Flavum	3.57 ± .27	4.67 ± .08	3.32 ± .25	4.67 ± .08	4.53 ± .18
Disc vs Osteophytes	4.47 ± .08	4.03 ± .17	4.40 ± .12	3.70 ± .07	4.67 ± .15
Overall Confidence	3.53 ± .32	4.47 ± .15	3.67 ± .22	4.37 ± .17	4.50 ± .18
Aesthetically Pleasing	4.13 ± .18	4.43 ± .11	3.97 ± .34	4.57 ± .13	4.17 ± .27

Based on these values, it can be seen that the fused image is better for differentiating disc and osteophytes than the registered MR and provides an overall higher confidence in diagnosis. The registered MR provided a higher confidence than the fused image for assessing the exiting nerve root. Readers had a higher confidence using the fused images compared to the CT images from phase 2 for all 10 questions.

It should be noted that the CT images were exactly the same in both phases of testing as the MR images were registered to the CT images. There are fewer statistically significant differences compared to the fused image for phase 1 CT than phase 2 CT. Users likely adjusted

Table 4.2: Statistical Power (P Value) for differences between the given modalities and phases of testing. Note significance was set at $P < 0.01$ to account for multiple comparisons. Any statistically significant values have been marked with a *.

Question	Phase 1 CT vs Fused	Phase 2 CT vs Fused	Phase 1 MR vs Fused	Phase 2 MR vs Fused	Phase 1 CT vs Phase 2 CT	Phase 1 MR vs Phase 2 MR
Disc Contours	*.001	*.001	.369	.031	.513	*.002
Disc Protrusions	*.001	*.001	.322	.201	.374	.017
Spinal Stenosis	*.001	*.001	.833	.865	.422	.861
Foraminal Stenosis	*.001	*.001	.076	.076	*.006	.610
Exiting Nerve Root	*.001	*.001	.025	*.005	.032	.117
Facet Joints	.583	*.001	.134	.039	*.002	*.010
Ligamentum Flavum	*.001	*.001	.382	.678	.319	.316
Disc vs Osteophytes	*.006	*.001	*.001	*.001	.156	.022
Overall Confidence	*.001	*.001	.529	*.003	.179	*.008
Aesthetically Pleasing	.164	*.003	.251	.065	.742	.179

their rankings slightly when the fused image was presented, making a relative comparison between the CT and fused image. A similar effect occurred between testing phases with the MR images. We have reported comparisons in rankings between the CT/MR images from phase 2 vs the fused image as a result of this.

4.3.2 Measurement Results

The average protrusion size found by the five readers for phase 1 and phase 2 of the study in both the axial and sagittal planes is shown in Table 4.3. Intra-class correlations (ICC) showing the degree of agreement between readers are also show in Table 4.3. All values were less than .8 demonstrating lack of agreement between readers.

Table 4.3: Average protrusion size (mm) and intra-class correlations for all 5 readers. The Intra-class correlation is two-way mixed model absolute measure, with agreement defined as ICC > 0.8.

	Unregistered (Phase 1)		Fused (Phase 2)	
	Axial	Sagittal	Axial	Sagittal
Average Protrusion Size (mm)	17.2 ± 5.4	11.9 ± 2.6	20.3 ± 11.6	9.3 ± 2.7
Intra-class Correlation	.367	.202	.417	.484

Levene's tests were conducted to determine whether there was an inhomogeneity of variance between testing phases for individual readers and for all readers. These results are shown in Table 4.4. Reader 3 showed an inhomogeneity of variance between testing phases. There was no other significant inhomogeneity. This indicates a possible learning effect for reader 3.

Table 4.4: Significance results (P-values) of Levene's test for homogeneity of variance between testing phases.

	Axial	Sagittal
Reader 1	.520	.936
Reader 2	.869	.358
Reader 3	.001	.168
Reader 4	.432	.111
Reader 5	.553	.690
All Readers	.646	.871

4.3.3 Time Results

The average times for completing the rankings and measurement tasks per patient are presented in Table 4.5.

Table 4.5: Average time per subject in seconds for completing the rankings and measurement tasks in phase 1 (non-registered images) and phase 2 (registered and fused images) of the study.

		Reader 1	Reader 2	Reader 3	Reader 4	Reader 5	All Readers
Phase 1	Ranking	597 ± 260	661 ± 242	496 ± 212	227 ± 48	115 ± 71	419 ± 238
	Measurement	581 ± 196	555 ± 259	560 ± 204	191 ± 48	426 ± 287	463 ± 164
Phase 2	Ranking	94 ± 42	42 ± 22	494 ± 396	31 ± 9	138 ± 73	160 ± 192
	Measurement	126 ± 114	48 ± 30	230 ± 138	38 ± 10	94 ± 95	107 ± 77

An analysis of variance was conducted for both the rankings and measurements based on the readers (radiologists) and phases of the study (unregistered vs registered). The results are contained in Table 4.6. Based on the results of the ANOVA test, there was no significant difference in times for ranking $P = .085$ or measurement $P = .179$ between the two phases of the study. There was a significant difference between radiologists for rankings $F(3, 44) = 38.47$,$P < 0.001$ and measurements $F(1, 13) = 22.28$,$P < 0.001$ and for a combination of

radiologist and phase for the rankings $F(3,41) = 6.17$,$P = 0.002$, however these differences were not examined further as the primary goal of the study was a comparison between phases.

Table 4.6: ANOVA results for times to perform rankings and times to make measurements, with statistically significant result denoted with a *.

		Degrees of Freedom	P-value	Effect Size η^2	Power
Ranking	Radiologist	$F(3,44)$	*.000	.733	1.00
	Phase	$F(1,14)$	.085	.197	.407
	Radiologist*Phase	$F(3,41)$	*.002	.306	.940
Measurement	Radiologist	$F(1,13)$	*.000	.712	.998
	Phase	$F(1,9)$	.179	.191	.256
	Radiologist*Phase	$F(1,11)$	.102	.257	.387

4.4 Discussion

A previously presented technique [4] was tested for clinical utility. A fused image was created that successfully combined the bone cortex detail from the CT and the soft tissues visible in the MR modality.

Based on the results from five readers, it has been shown that in general the readers are more confident making diagnosis using the fused images than the CT. They have similar confidences for the majority of questions when making diagnosis using the fused images and MR images. The fused images were better for distinguishing between disc and osteophyte than the MR images and readers had an overall higher confidence in diagnosis using the fused images compared to the MR images. Readers were more confident using the MR images than the fused for examining the exiting nerve root.

When measuring protrusion sizes the intra-class correlation coefficients did not show agreement between readers for either phase of the study and Levene's tests only show inhomogeneity for reader three on the axial images. This is likely a learning effect. In general this indicates the source of variation in measurements is primarily due to variation between radiologists as opposed to the different phases of the study.

There was no statistical difference in the times taken for rankings and measurements between the unregistered images and the fused images. It should be noted that the unregistered images were not linked so the readers scrolled independently through these images, whereas the fused images were registered to the MR and CT images and could be linked if the user elected to do so. This may have made completing the tasks for the fused images more efficient. There were three series of images compared for phase 2, compared to two for phase 1, but the registered MR images were resampled so contained fewer slices, which may also have affected the time for completion.

A limitation of this study is that the users were shown the unregistered images first, followed by the fused images. This may have introduced a bias. A further limitation of the study is in determining whether the measured improvements are primarily due to the linking and registering of the images as opposed to fusing the images. To be more certain a future third phase of the study could be conducted comparing registered MR and CT without the fused images. With that said, the results of the current study indicate that the improvement was due to fusion as opposed to linking and registering because the readers had a higher confidence in making a diagnosis using the fused images compared to the registered CT images and in specific areas compared to the registered MR images.

In general the performance of the fused images was similar to the unregistered images in terms of time to review images. Fused images were associated with overall higher or the same confidence in assessment compared to MR and CT, except when assessing the exiting nerve root. The variability in measuring protrusions is primarily from differences in radiologists as opposed to fused vs non-registered MR/CT images. In view of these advantages and considering that using the fused images did not increase the average time needed to evaluate each subject it would be a useful additional modality to complement CT and MR images when both are available. It is suggested that MR images be used for diagnosis exiting nerve roots as opposed to the fused images.

A limitation in this technique is the difficulty in accurate registration of CT and MR spine

images. The registration technique employed for this study assumes a rigid transformation between the CT and MR images, demonstrating subvoxel registration accuracy when this assumption is valid [4]. For patients with non-rigid deformations between the MR and CT scans, this technique is not applicable as demonstrated by the 5 patient images that were not able to be registered. These non-rigid deformations include positional changes and potential for anatomical changes between scans. Additionally due to running time the registration process currently takes place off-line.

Future work could involve the investigation of different possible user interfaces for the fused images, including as reader 5 suggested, having the fused image able to be toggle on or off over the MR image. This would effectively toggle the bone detail on and off over the MR image. It would also be useful to evaluate the fusion of other MR modalities with CT images, such as T1 images, along with more flexible and efficient registration methods

In conclusion, adding a series of fused CT/ MR images improves radiologist's confidence in assessing CT and MR images of the lumbar spine without adding time to the interpretation. The magnitude and number of structures showing improvement is greater for CT. The improvement for MR is mainly in assessment of disc vs osteophyte, though improvement for the study as a whole was also shown. These results suggest that commercial development of CT/ MR fusion software for the lumbar spine would be beneficial.

4.5 Acknowledgements

The authors would like to acknowledge the contributions of Yves Bureau who is affiliated with Lawson Imaging for his role as a statistical consultant for this study. Brandon Miles would like to acknowledge the support of the BioMedical Engineering Graduate Program at the University of Western Ontario.

Bibliography

[1] Arnaud Colin and Jean-Yves Boire. MRISPECT fusion for the synthesis of high resolution 3D functional brain images: a preliminary study. *Computer Methods and Programs in Biomedicine*, 60:107–116, 1999.

[2] D. F. Fardon and P. C. Milette. Nomenclature and classification of lumbar disc pathology: Recommendations of the combined task forces of the north american spine society, american society of spine radiology, and american society of neuroradiology. *Spine*, 26(5): E93–E113, 2001.

[3] Pamela M. Mazurek, Anne M. Silas, and John M. Gemery. Summation of ct scans during radiofrequency ablation for assessing target lesion coverage ,. *American Journal of Roentgenology*, pages 790–792, September 2008.

[4] B. Miles, I. Ben Ayed, M. W. K. Law G. Garvin, A. Fenster, and S. Li. Spine image fusion via graph cuts. *IEEE Transactions On Biomedical Engineering*, 60(5):1841–50, July 2013.

[5] Jing Yuan, Brandon Miles, Juan Shi, Greg Garvin, Xue-Cheng Tai, and Aaron Fenster. Efficient convex optimization approaches to image fusion. *Numerical Mathematics: Theory, Methods and Applications*, 2014. To Appear.

Chapter 5

Spine Axial MR Classification via Integral Kernel Density Estimates

5.1 Introduction

Radiologic assessment is an essential step in managing patients with spinal diseases or disorders. This is particularly true given the prevalence of spinal diseases and the increase in medical imaging of the spine. Up to 75% of people will experience back pain in their lives with 1.8% of all hospital visits in the USA being a result of back pain and related disorders [19]. There has also been a 300% increase in the number of MR scans in the USA from the period between 1996 and 2002 [4]. In the diagnosis of spinal diseases, accurate detection and labeling of different spinal structures is vital, as many interventions require precise anatomic information [1, 8, 11, 13, 16, 18, 21, 26]. Correct annotation of the vertebral level of an image slice is particularly important as these slices serve as a reference for describing locations of additional pathology, such as the intervertebral discs [5]. Incorrectly identifying vertebrae may lead to wrong level surgery [10, 17] or misdiagnosis.

It can be time consuming for radiologists to determine the identity of a particular vertebra in an image, since adjacent vertebras have very similar appearance. Even when a vertebra has

been identified, a radiologist must remember its identity and location as they make a diagnosis from multiple slices in an image. Axial locations are cross-referenced from sagittal images. However, this is still dependent on the radiologist locating the cross-referenced vertebra in the sagittal image. Correctly identifying vertebra is especially difficult in cases where the sagittal image is not available, or the cross-referencing cannot be relied upon, i.e., when there are deformations between sagittal and axial images due to patient repositioning. Even when the sagittal images are available for cross referencing, the topmost and bottommost axial slice belonging to a given vertebrae need to be identified for annotation of all axial image slices. This is a non-trivial task based on sagittal images alone and is more suitable to be performed on axial images. Accurate automatic labeling of axial slices would alleviate the need for radiologists to mentally label slices and would also provide consistency for all practitioners referencing a given patient image. Furthermore, vertebrae detection and identification is required for image processing tasks such as registration [22], automated diagnosis [25], 3D spine reconstruction [14] and segmentation [15, 16].

5.1.1 Previous Labeling Work

Generating these labels in a manual fashion is tedious, subjective, and time-consuming. Therefore, automating the process is desired and has recently sparked an impressive research effort [1, 8, 11, 13, 15, 16, 18, 21, 26]. Automated labeling of such images is, however, a challenging problem as the field-of-view (i.e., the number of visible vertebral levels), the distributions of image intensities, and the sizes, shapes, and orientations of different spinal structures are highly variable among different patients [1, 8, 26]. Algorithms for the lumbar spine are more common than the cervical spine, with the lumbar spine being studied by [1, 7, 8, 11, 15, 16, 18, 21, 26] and the cervical spine being studied by [7, 8, 11, 13, 18, 21, 26]

There are two major limitations with current automated spine labeling algorithms:

(1) Most of the current algorithms address the labeling problem through intensive training from a manually-labeled data set [1, 7, 8, 18, 21, 26]. Such a training stage aims at learning

the shapes, textures and appearances of different spinal structures. This knowledge is then used within a classification or regression algorithm (e.g., support vector machine [18], random forest regression [7, 21] or graphical models [8, 13]) to subsequently label different spinal structures in the test image. Such algorithms work very well on data sets that closely match the training data, but would require adjustment/retraining for different data sets or if the imaging modality and/or acquisition protocol are altered (e.g., an algorithm that is trained and built for CT images may not perform well on MR data [8, 13, 15, 16]). This might impede the use of these algorithms in routine clinical practices, where a particular disorder might be analyzed radiologically using several different imaging modalities/protocols with widely variable imaging parameters (resulting in extremely high variation in image data).

(2) To the best of our knowledge, most of the current spine labeling algorithms focus on either the sagittal view [1, 7, 8, 11, 13, 18, 21] or are restricted to CT/X-Ray images [7, 8, 15, 16, 13, 21]. The work of [26] utilizes 3D MRI volumes directly. Axial labeling is beneficial as the quantification and level-based reporting of common inter-vertebral disc displacements such as protrusion, extrusion and bulging require the radiologist to thoroughly inspect all individual axial MR slices [5], while visually cross-referencing such axial slices to their corresponding position in the sagittal view (Fig. 5.5). It is, however, important to note that in some cases only the axial view is available for the patient while in other cases the two scans might be acquired at different time points. Therefore, localizing the spinal structures in different views becomes a challenging task (even for an experienced radiologist), which motivates a stand-alone axial spine detection/labeling algorithm. Such a system would facilitate the generation of radiologic reports. Again, even the when there is a sagittal view to cross-reference from, the lowermost and uppermost slices still need to be determined to label all sagittal slices. Additionally none of the MR based methods offer near real-time performance that we are aware of.

5.1.2 Integral Images

Integral images are well known in computer vision literature and allow for the fast computation of the sum of pixels of any rectangular regions of an image. Introduced for the purpose of human face detection [23], integral images are very efficient, demonstrating a speed up of several orders of magnitude compared to direct computation. Their computational complexity depends only on the size of the image rather than the size of the rectangular region, reducing the computational complexity to $O(1)$ for each sub region [3]. Extensions of this work include the calculation of histograms via integral images [20, 24] and calculating geometric distributions [9] for the purposes of template matching in photographs. Extensions involving probability product kernels have also been considered for template matching in photographs [3].

5.1.3 Contributions of this work

In this work, we present a robust, near real-time axial MR labeling algorithm. Individual pixels are classified through a non-linear probability product kernel (PPK) based classification. Based on these classified pixels, per slice features are generated, which are used to classify individual vertebra. Multiple vertebra are then iteratively identified based on the location of previous vertebrae. Our classifier requires evaluations of computationally expensive integrals at each pixel of the image domain. However, direct evaluations of such integrals would be prohibitively time consuming. We utilize an efficient computation of kernel density estimates and PPK evaluation for large volumes and arbitrary local window sizes via *integral kernels*. This algorithm is $O(nz)$ with respect to n and z, where n is the number of pixels in the image and z is the number of histogram bins. Results can be achieve in near real-time with a GPU implementation. Furthermore, the proposed method does not require intensive external training. The performance of the algorithm was quantified on 32 sets of T1-weighted 3D MR images of the lumbar spine and 24 sets of T2-weighted 3D MR images from the cervical spine. Each 3D image contained multiple 2D slices.

The paper is organized as follows: in section 5.2, the labeling algorithm is introduced,

where pixels are classified using a PPK classifier, with slices and vertebrae then classified based on this information. After this, the experiments are outlined in section 5.3 followed by the results in section 5.4. Finally, conclusions are presented.

5.2 Formulation

To properly find the correct axial annotations, our algorithm works on a hierarchy of feature levels, with the features from the current level used as inputs to the next level. A single user-selected point is the only user input needed. Based on this point, pixels can be classified, followed by slices, followed in turn by single vertebra and finally, followed by multiple vertebras. At each stage we aim to minimize the number of tuned parameters, while utilizing robust geometric information in order to provide a flexible model for classification. A diagram of this hierarchy is shown in Fig. 5.1. Each of these steps will be detailed in subsequent sections.

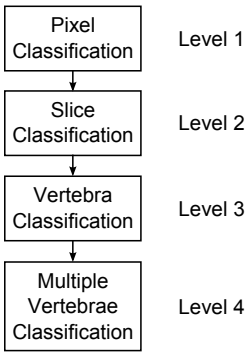


Figure 5.1: The hierarchy of feature levels for classifying discs and vertebrae.

5.2.1 Efficient Pixel Classification via Integral Kernel Images

Pixelwise Probability Kernel Matching

We propose a non-linear classifier, which determines whether the neighbourhood of each pixel p matches a target distribution. Such neighbourhood distributions contain contextual information, thereby providing much richer inputs to the classifier than individual pixel intensities. Let

$\mathcal{D}_j : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, $j \in [1 \ldots N]$, be a set of input images, which correspond to the axial slices of a given spine series. Ω is the image domain and N is the number of slices in the series. For each $\mathcal{D} \in \{\mathcal{D}_j, j = 1 \ldots N\}$ and each pixel $p : (x, y) \in \Omega$, we seek to create a non-linear kernel based classifier of the form:

$$\text{sign}\left(\phi\left(P_{p,\mathcal{W},\mathcal{D}}, P^L\right) - \rho\right) \tag{5.1}$$

where P^L is an *a priori* learned distribution, ρ is a constant and $P_{p,\mathcal{W},\mathcal{D}}$ is the kernel density estimate (KDE) of the distribution of image data $\mathcal{D}$ within a window $\mathcal{W}$ centered at pixel $p : (x, y) \in \Omega$:

$$P_{p,\mathcal{W},\mathcal{D}} = \frac{\sum_{p \in \mathcal{W}} k_z^{\mathcal{D}}(p)}{|\mathcal{W}|} \quad \forall z \in \mathcal{Z} \tag{5.2}$$

where $|\mathcal{W}|$ is the number of pixels within $\mathcal{W}$ and $\mathcal{Z}$ is a finite set of bins encoding the space of image variables. $k_z^{\mathcal{D}}$ is a kernel function defined by the following:

$$k_z^{\mathcal{D}}(p) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{\frac{(z-\mathcal{D}(p))^2}{2\sigma^2}} \\ \qquad or \\ \delta\left(z - \mathcal{D}(p)\right) \end{cases} \tag{5.3}$$

where σ is the width of the kernel and $\delta(t)$ is the dirac delta function, which is equal to 1 if $t = 0$, and 0 elsewhere. The first case corresponds to the Gaussian kernel, and the second case to bin counting, i.e., normalized histogram.

ϕ is a *probability product kernel* [12], which measures the degree of similarity between two distributions:

$$\phi\left(P_{p,\mathcal{W},\mathcal{D}}, P^L\right) = \sum_{z \in \mathcal{Z}} \left[P_{p,\mathcal{W},\mathcal{D}}(z) P^L(z)\right]^\gamma, \quad \gamma \in [0, 1] \tag{5.4}$$

Based on these pixelwise classifications, we further add regional geometrical constraints to

produce a classification for each slice $\mathcal{D} \in \{\mathcal{D}_j, j = 1 \dots N\}$.

Efficient Computation of PPK via Integral Kernels

To embed rich contextual information about the vertebrae/discs, we need to use large-size windows in our classifiers. For large windows, the computation of (5.4) for each pixel in $\mathcal{D}$ is very expensive computationally if performed by direct evaluation. In the following, we describe an efficient computation of kernel density estimates and PPK evaluations for large images and arbitrary window sizes via integral kernels. Such integral-kernel method can be viewed as an extension of the integral-image method of Viola and Jones [23], which is well-known in computer vision. First, recall the integral-image method.

Integral images: Given an input image $\mathcal{D}$, the corresponding integral image $\mathcal{I}^{\mathcal{D}}$ is defined as the sum of all pixel intensities to the left and above the current pixel:

$$\mathcal{I}^{\mathcal{D}}(x, y) = \sum_{u \le x} \sum_{v \le y} \mathcal{D}(u, v) \tag{5.5}$$

The sum of intensities of all pixels within an arbitrary rectangular window defined on the domain of $\mathcal{D}$ can be computed from the integral image using only the corners of the rectangle:

$$\begin{aligned} \sum_{u=x_1}^{x_2} \sum_{v=y_1}^{y_2} \mathcal{D}(u, v) = \mathcal{I}^{\mathcal{D}}(x_1, y_1) + \mathcal{I}^{\mathcal{D}}(x_2, y_2) \\ - \mathcal{I}^{\mathcal{D}}(x_1, y_2) - \mathcal{I}^{\mathcal{D}}(x_2, y_1) \end{aligned} \tag{5.6}$$

where (x_1, y_1) are the coordinates of the upper left corner of the rectangle and (x_2, y_2) are those of the lower right corner. Since (5.5) can be computed efficiently for the entire image and (5.6) can be computed very efficiently for a given rectangle, this method is very efficient when multiple windows need to be computed from the same image.

Integral kernels: To extend the idea of integral images to integral kernels and efficiently compute the PPK in (5.4), we build for each slice $\mathcal{D}$ a set of separate kernel images defined

over Ω: $k_1^{\mathcal{D}}, k_2^{\mathcal{D}}, \ldots k_z^{\mathcal{D}}, z \in \mathcal{Z}$, as shown in Fig. 5.2 a. We can then proceed to compute an integral kernel image based on each k_z. In a similar way to (5.5), and for each $p = (x, y) \in \Omega$, we define the integral kernel image as follows:

$$\mathcal{I}_z^{\mathcal{D}}(x, y) = \sum_{u \leq x} \sum_{v \leq y} k_z^{\mathcal{D}}(u, v) \tag{5.7}$$

Now we can easily show that the kernel density estimates $P_{p,\mathcal{W},\mathcal{D}}$ can be calculated from the integral kernel images using five simple operations for each $p(x, y) \in \Omega$:

$$P_{p,\mathcal{W},\mathcal{D}}(z) = \frac{\mathcal{I}_z^{\mathcal{D}}(x_1, y_1) + \mathcal{I}_z^{\mathcal{D}}(x_2, y_2) - \mathcal{I}_z^{\mathcal{D}}(x_1, y_2) - \mathcal{I}_z^{\mathcal{D}}(x_2, y_1)}{(x_2 - x_1 + 1)(y_2 - y_1 + 1)} \tag{5.8}$$

where $x_1 = x - \frac{w}{2}$, $x_2 = x + \frac{w}{2}$, $y_1 = y - \frac{h}{2}$ and $y_2 = y + \frac{h}{2}$, with w and h being the width and height of $\mathcal{W}$ (an illustration is given in Fig. 5.2 b).

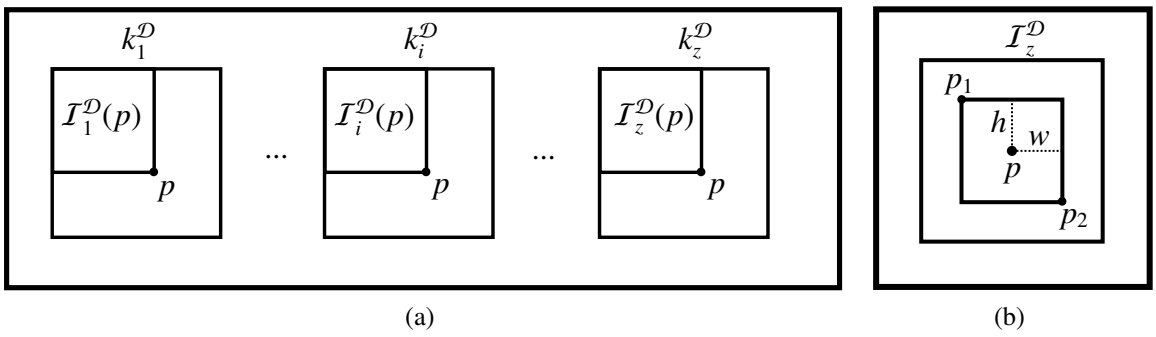
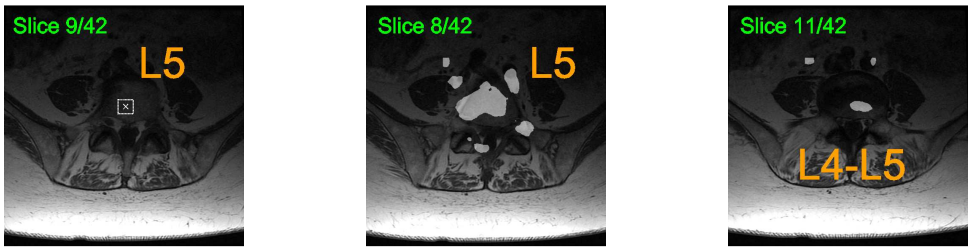


Figure 5.2: Diagram of the Integral Kernels: (a) Formation of integral kernel images $1 \ldots z$, using point $p = (x, y)$, (b) A diagram of the window centered at pixel $p = (x, y)$ defined by $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$.

This leads to the following efficient evaluation of classifier (5.1) for every $(x, y) \in \Omega$:



(a) (b) (c)

Figure 5.3: Diagram of pixel classifications from images: a) a simple user input used for all slices, b) pixel classifications for a vertebra, c) pixel classifications for an intervertebral disc.

$$\operatorname{sign}\left(\sum_{z\in\mathcal{Z}}\left[\frac{P^L(z)\left(\mathcal{I}_z^{\mathcal{D}}(x_1,y_1)+\mathcal{I}_z^{\mathcal{D}}(x_2,y_2)\right.}{(x_2-x_1+1)}\right.\right.$$
$$\left.\left.\frac{\left.-\mathcal{I}_z^{\mathcal{D}}(x_1,y_2)-\mathcal{I}_z^{\mathcal{D}}(x_2,y_1)\right)}{(y_2-y_1+1)}\right]^{\gamma}-\rho\right) \tag{5.9}$$

The above function has a computational complexity that is linearly proportional to the number of pixels in Ω and the cardinality of $\mathcal{Z}$, and is independent of the window's size. This method is also highly suited to modern graphics cards because it is amenable to parallel implementations.

Pixel-level Classifications via Integral Kernel Images

To provide the initial training, a single user point $p_o = (x_o, y_o) \in \Omega$ is selected on a single slice in the series. Then, the prior distribution P^L is learned from a window of size $w \times h$ centered on p_o. Examples of the obtained pixel-level classifications are shown in Fig. 5.3, where the results are depicted for both vertebrae and disc slices, based on the simple training input from a different slice. To avoid processing the entire slice, only pixels within a region-of-interest R_s around the input point are considered in the pixel-level classification process. The pixel-level classifications we obtained from (5.1) will be used to further generate slice-level classifications.

5.2.2 Slice-level Classifications

The second level of features is the area of pixels classified as vertebrae in a given slice. We group vertebra pixels into sets of 4-connected regions: S^i, $i = 1, 2, \ldots$. These connected regions are then filtered, discarding any region with an area less than a specified area $A_{\min}$ or whose centroid $c_i \in \Omega$ is further than $d_{\max}$ from p_o using the standard Euclidean distance. The area of the largest remaining S^i is then used as a feature for slice-level classifications. If all the connected regions are removed, we assign value 0 to this feature. The following describes the criteria for including or discarding a region:

$$\text{Criteria} = \begin{cases} \text{Include} & \left\{S^i | \text{area}(S^i) > A_{\min}, \right. \\ & \left. \|c_i, p_0\| < d_{\max}\right\} \neq \emptyset \\ \text{Discard} & \text{otherwise} \end{cases} \tag{5.10}$$

5.2.3 Vertebra Classification

The next level of classification is identifying individual vertebrae. We start with an input set of adjacent slices $\mathcal{D}_k, k \subset [1, N]$ in the neighbourhood of a vertebra. These slices are all the slices with a given search height H_s either centered on the initial point or starting at a previously identified vertebra boundary. We need to classify these slices as either vertebrae or not. As an input we use the areas A_k from the largest S_i for each slice. This becomes a one dimensional data analysis problem. We start by applying a small smoothing filter to the data $A_k^s = A_k * K$, were A_k^s is the smoothed data and K is a one dimensional convolution kernel. After this the slices are assigned a binary classification as either vertebrae: 1 or not: 0 based on

$$\text{sign}\left(A_k^s - t_{\text{area}}\right) \tag{5.11}$$

where t_{area} is a threshold given by $t_{\text{area}} = c_a \mu_{\text{area}}$, with c_a being a user defined factor and μ_{area} being the average of areas A_k^s. The largest number of adjacent slices classified as vertebrae are

then defined to be the vertebra. This provides the location of the vertebra including an uppermost and lower most slice defining the vertical boundaries of the vertebra. We also impose a minimum vertebrae height $H_{\min}$ criteria. If no set of adjacent slices meets the minimum height criteria then no vertebra is classified.

5.2.4 Multiple Vertebra Classification

In order to improve the classification accuracy of the algorithm, iterative model updating can be employed. By using the location of the most recently found vertebra, a new learned probability P^L can be determined at the center of that vertebra, and a search region for finding the next vertebrae can be identified. Classification then proceeds up and down the spine. For the first vertebrae, the initial search height H_s^0, is defined to be twice the height of a vertebrae, centered at the input point. This is because the input point could be selected at either the top or the bottom of the vertebrae. For finding subsequent vertebrae, the search range H_s begins at the boundary of the previous vertebrae and extends for the height of a vertebra plus two intervertebral disc spaces, which should include an entire vertebra.

5.2.5 Algorithm Summary:

An overview the algorithm can be found in Algorithm 2. Step 3 is the most computationally intensive. However, with the use of integral images, it can be reduced to order nz, where n is the number of pixels in all image slices and z is the number of features.

5.3 Methods

This retrospective study was approved by the Human Subjects Ethics Board of Western University, with the requirement for informed consent being waived. A total of 56 suhjects were included in this study. A set of axial T1-weighted Magnetic Resonance (MR) lumbar spine 3D images, from 32 patients, was used for the lumbar spine (LSpine) and an additional set

Algorithm 2: Vertebrae Classification Algorithm

- Given an initial input $p = p_0$ and Vertebrae $V_n = V_0 \in [V_{\min}, V_{\max}]$

  1. Learn the target probability distribution P^L.
  2. Set the search height $H_s = H_s^0$
  3. For each slice $\mathcal{D}_j$ in H_s:
     (a) Use sign $\left(\phi\left(P_p, P^L\right) - \rho\right)$ to classify each pixel p via integral kernel images.
     (b) Identify the largest connected region $S_j^{\mathcal{D}}$, satisfying $A_{\max}$ and $d_{\min}$.
     (c) Calculate area A_k from $S_j^{\mathcal{D}}$
  4. Using sign $\left(A_k^s - t_{\text{area}}\right)$, find the uppermost and lowermost slices for each vertebrae.
  5. Update the vertical search region H_s and target distribution P_L based on the V_n.

- **If** $V_n \leq V_{\max}$:

  6. While $V_n < V_{\max}$,
     (a) Let $n = n + 1$
     (b) repeat steps 3-5

- **Else:**

  7. Set $V_n = V_0$. Update the vertical search H_s and target distribution P_L based on V_0.
  8. While $V_n \geq V_{\min}$
     (a) Let $n = n - 1$
     (b) Repeat steps 3-5.

of axial formatted T2-weighted Magnetic Resonance (MR) cervical spine 3D images, from 24 patients, was used for the cervical spine (CSpine). The T2 volumes were acquired using a 3D, T2-weighted MR sequence. The MR model and manufacturer varied based on the location the images were acquired. For the T1-images, the slice thickness ranged from 4 to 5 mm and the in-plane voxel spacing ranged from 4.4 to 10 mm. For the T2 3D images, the slice thickness ranged from 1.9 to 3.0mm with an in-plane voxel spacing ranging from 0.39 to 0.47 mm. For the C-spine data, all seven cervical vertebrae (C1-C7) and corresponding inter-vertebral discs were visible in every image. For the lumbar spine, the number of visible vertebrae varied based on the subject. These results are shown in Table 5.1.

Table 5.1: The number of vertebrae visible at each level for the lumbar spine data sets acquired from 32 subjects.

Vertebral Level	L5	L4	L3	L2	L1	T12
Number of Visible Vertebrae	32	32	21	11	5	1

The initial user click was placed on a single axial slice, on the C7 vertebra for the cervical spine labeling and on the L5 vertebra for the lumbar spine labeling. The algorithm labeled the axial slices as either vertebra or intervertebral disc (Fig. 5.5 and 5.6 depict typical examples for lumbar and cervical spines respectively).

5.3.1 Choices of the Parameters and Input Selection

The algorithm parameters were selected based either on tuning the proposed classifier or on the geometric properties of the human spine. The classifier-related parameters were tuned experimentally. The geometric properties were also set experimentally, but corresponded to well documented measurements from the literature [6, 2]. For both the cervical spine and lumbar spine images, we used the Dirac delta function to compute the kernel density estimates. Variable γ was set equal to 0.5, which corresponds to the Bhattacharya distance between distributions. Pixel-level classification threshold ρ was set equal to 0.75, a value determined experimentally. The number of bins was experimentally set to be 100. The 1D convolution parameter K was

set to [0.3 1 0.3], while the area threshold factor c_a was fixed equal to 0.75, both determined experimentally.

The remaining parameters are geometric in nature and have been defined in terms of millimeters, to be invariant to the voxel spacing of the images. These are the search window $\mathcal{W}$, the minimum classification area $A_{\min}$, the maximum distance $d_{\max}$, the initial search height H_s^0, the subsequent search height H_s, the minimum vertebrae height $H_{\min}$ and the region of interest R_s. Table 5.2 shows vertebra and disc measurements (height and width) reported in the literature [6]. Furthermore, based on data for the lumbar spine [2], the major axis length of the vertebrae was found to be about 1.5 times the minor axis length (width of the vertebrae). Based on these values, the cross-sectional area of a vertebra can be overestimated by a rectangle of $1.5w^2_{\text{vertebrae}}$ and underestimated by an oval of size $0.375\pi w^2_{\text{vertebrae}}$, with $w_{\text{vertebrae}}$ denoting vertebra width. Our experimental heights, search ranges and area measurements correspond to these measurements, which can be found in Table 5.2. The search window size was set experimentally.

Table 5.2: Vertebra and Disc Measurements as well as Overestimates/Underestimates for Cervical and Lumbar Vertebrae.

Measurement	L-Spine	C-Spine
Vertebra Height (*mm*)	27.3 ± 1.2	14.6 ± 0.8
Disc Height (*mm*)	8.8 ± 0.9	4.2 ± 0.5
Vertebrae Width (*mm*)	34.3 ± 1.8	15.5 ± 0.9
Vertebrae Area Overestimate (mm^2)	360	1765
Vertebrae Area Underestimate (mm^2)	224	1110

For all the experimentally determined parameters, the tuning was performed on a single lumbar and c-spine image and verified on two other lumbar and five other c-spine images before testing the entire dataset. Evaluation was completed using all images, as the tuning set was so small.

Table 5.3: Parameter selection for the C-Spine and L-Spine labeling algorithms.

Parameter	Symbol	L-Spine	C-Spine
Pixel Threshold	ρ	0.75	0.75
Number of Bins	$\mathcal{Z}$	100	100
Search Window (*mm*x*mm*)	$\mathcal{W}$	12 x 12	8 x 8
Region of Interest (*mm*x*mm*)	R_s	80 x 80	40 x 40
Minimum Area (mm^2)	$A_{\min}$	400	100
Max Distance (*mm*)	$d_{\max}$	40	20
Minimum Vertebrae Height (*mm*)	$H_{\min}$	12.5	5
Area Threshold Factor	c_a	0.75	0.75
Initial Search Height (*mm*)	H_s^0	50	25
Subsequent Search Height (*mm*)	H_s	45	16.2

5.3.2 Validation of Labeling

The performance of the algorithm was validated based on the correct classification of vertebrae, the classification of individual slices and the distance of the vertebral uppermost and lowermost slices from the ground truth. The ground truths were manually generated from the axial images. Each slice was classified as either vertebra or disc based on the percentage of the vertebral column cross sectional area containing vertebra/disc in that slice. If more than 50% of the vertebral column consisted of a single vertebra, then that slice was labeled as belonging to that vertebra; otherwise, it was labeled as disc. This resulted in an uppermost and lowermost slice for each vertebra (e.g., L3 could be manually labeled to extend from axial slice 24 to slice 30). Since the C1 and C2 vertebrae are special cervical vertebrae with anatomically overlapping bone structures, their classification required precise boundary definitions. The C1 vertebrae were defined from the top of C1's anterior arch to the bottom of C1's inferior articular surface of lateral mass. C2 vertebrae were defined from the top of C2's dens to the bottom of C2's vertebral body. The overlapping structure can be seen in Fig. 5.4. Due to the distinctive shape of the C1 vertebra it is easily identifiable manually and was omitted from the automatic classification. The C2 vertebrae have been labeled excluding the dens.

To validate the correct classification of vertebrae, a vertebra was considered to be *correctly labeled* if: 1) there was at least one correctly labeled vertebral slice for that vertebra and 2) no

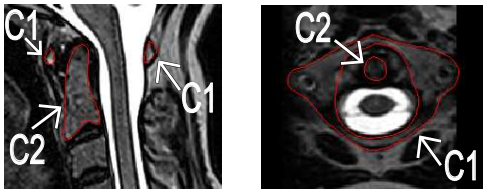


Figure 5.4: Sagittal (left) and Axial (right) views of the C1 and C2 vertebrae, with the vertebrae outlined in red. In the axial view the dens of C2 is located inside of C1, resulting in both vertebrae being visible in some axial slices.

slices were incorrectly labeled as another vertebra. If only condition 2) was met, the vertebra was considered to be *unlabeled*, since it was not given any label. The vertebra was considered to be *incorrectly labeled* if any of the vertebra's slices were incorrectly labeled as a different vertebra. Ideally, all vertebrae will be correctly labeled. An *incorrectly labeled* vertebra is a concern, since this can lead to incorrect diagnosis or treatment, whereas an *unlabeled* vertebra is merely inconvenient.

To validate the correct classification of individual slices, slices were considered to be correctly labeled if they matched the ground truth and incorrectly labeled otherwise. Slices beyond the extent of the vertebral column, such as in the sacrum, were excluded from this analysis.

To validate the vertebral uppermost and lowermost slice boundaries, a comparison was made with the manually identified ground truths. For each vertebra, the distance, in number of slices between the labeled uppermost slice of the vertebra and the ground truth uppermost slice was calculated. The same principle was applied for the lowermost slices. Comparisons of these distances were then made for each vertebra over the set of subjects by calculating both the mean distances and the maximum distances in number of slices.

5.3.3 Run Time Validation

The run time for the algorithm was tested. The CPU code was written in Matlab (the Mathworks Nattick MA, USA) and the GPU code was written in CUDA. The experiments were run on a computer with a Xeon quad core processor (Intel, Santa Clara, CA, USA) 2Gb of ram and an NVidia GeForce GTX 680 graphics card (Nvidia, Santa Clara CA, USA).

5.3.4 Sensitivity Analysis of Parameters

In order to determine how robust the algorithm was with regard to changes in parameters, a sensitivity analysis was performed. For a given parameter, using a sample subject, the classification accuracy for vertebrae and slices was calculated for a range of possible algorithm parameters, surrounding the given parameter. The range was specific to each parameter extending over a wide choice of possible values and was centered at the actual value utilized in the algorithm (see Table 5.3). The parameters that were analyzed for sensitivity included: the Bhattacharya threshold, the x and y location of the input point, the search window size, the number of bins and the area threshold.

5.4 Results

5.4.1 Classification Accuracy

For the lumbar spine, a total of 102 vertebrae were classified. A representative sample labeling for these images can be seen in Fig. 5.5 for the lumbar spine and in Fig. 5.6 for the cervical spine. Of the 102 vertebrae, 101 were correctly identified and only 1 was incorrectly identified for a 99% classification accuracy. The per slice classification accuracy was found to be $88\% \pm 7\%$, and the error in identifying the uppermost and lowermost vertebrae slice boundaries was found to be 0.83 ± 0.46 slices, with the average maximum distance from the classified uppermost and lowermost slice boundaries to the ground truth boundaries (over the 32 patients) being 1.44 ± 0.91 slices. These results are summarized in Table 5.4. It should be noted that for the one vertebra that was defined as wrong, only one slice was incorrectly classified, with the rest of the vertebra being correctly classified. This error could be easily identified by a clinician. Additionally, for the majority of vertebra the boundaries are within one slice of the manually identified boundaries.

For the cervical spine, a total of 144 vertebrae were classified, with 114 being correctly

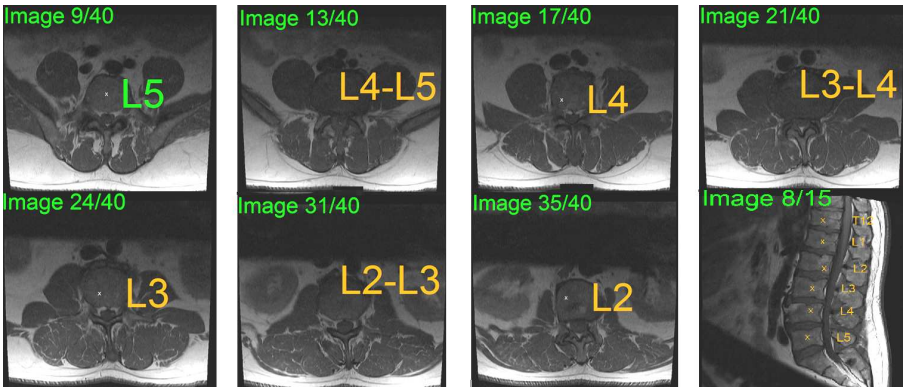


Figure 5.5: Representative output of the lumbar spine detection algorithm displaying axial slices from each analyzed level with the initial user input chosen at L5, with a labeled sagittal view provided for illustrative purposes.

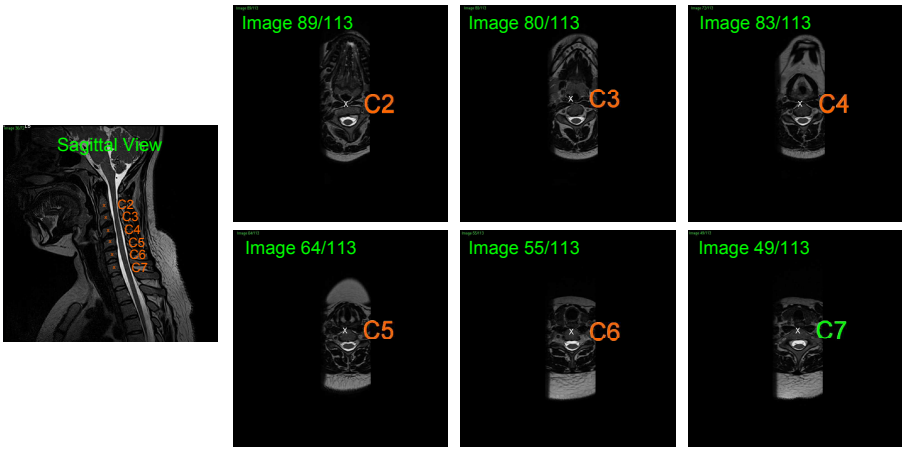


Figure 5.6: Representative output of the cervical detection algorithm displaying axial slices from each analyzed level with the initial user input chosen at C7, with a labeled sagittal view provided for illustrative purposes.

identified, 12 missed and 18 incorrectly labeled for a 79% classification accuracy. The slice accuracy was 71% ± 21%, the average boundary distance was 2.3 ± 2.5 slices and the average maximum boundary distance per patient was 5.5 ± 5.1. Out of the 24 patients, it was found that 4 had significant curvature in the lumbar spine that the algorithm was not set up to handle, hence there were a high number of mislabeled vertebrae on these images. An example of this can be seen in Fig. 5.7. New results were obtained with these four patients removed. Out of 120 vertebrae remaining, 107 were correctly labeled, 3 were missed and only 10 were incorrectly labeled for an 89% classification accuracy. The new slice accuracy was 77%±17%, the average boundary distance was 2.1±2.4 slices and the average maximum boundary distance per patient was 5.3 ± 5.0. Both sets of cervical spine results are summarized in Table 5.4. It should be noted that the cervical results are worse than [18] which had a 97% identification accuracy on sagittal images, but similar to [26] which had a performance of 79% again on sagittal images.

(a) (b)

Figure 5.7: Examples of cervical spines: a) a straight spine, which the algorithm succeeds on; b) a spine with significant curvature that the algorithm fails on.

5.4.2 Sensitivity to Input Parameters

The results of varying the input parameters are shown in Fig. 5.8 for the lumbar spine and in Fig. 5.9 for the cervical spine. For the lumbar spine, varying the Bhattacharya threshold (Fig. 5.8a) produced very similar results for values in the range of 0.5-0.8, with a large drop

Table 5.4: Detection accuracy for lumbar and cervical spine.

	No. Vertebrae	No. Correct	No. Missed	No. Wrong
LSpine	102	101	0	1
CSpine	144	114	12	18
CSpine (No Deformed Cases)	120	107	3	10
	Vertebrae Accuracy (Percent)	Slice Accuracy (Percent)	Boundary Distance (Slices)	Max Boundary Distance (Slices)
LSpine	99%	88% ± 7%	0.83 ± 0.46	1.44 ± 0.91
CSpine	79%	71% ± 21%	2.3 ± 2.5	5.5 ± 5.1
CSpine (No Deformed Cases)	89%	77% ± 17%	2.1 ± 2.4	5.3 ± 5.0

in performance for values above 0.8. This is expected since, as the threshold gets higher, the classified pixels must match the target distribution more closely. This makes the algorithm less robust to variations away from the target distribution, excluding many pixels that are actually part of the vertebrae. For the X and Y inputs, Fig. 5.8b and Fig. 5.8c show constant performance until about 20 pixels from the origin, giving a wide range of areas for selecting the initial point. The method was also very robust to window sizes (Fig. 5.8d) with any window size greater than 9mm x 9mm producing excellent results. This showed that the larger the input area for the classification, the better the method performed. Surprisingly, the method was not sensitive at all to the number of bins (Fig. 5.8e) or the area threshold (Fig. 5.8f). This demonstrated that these were not important parameters in the algorithm and a possible speed up could be realized by reducing the number of bins.

For the cervical spine, varying the Bhattacharya threshold (Fig. 5.9a) produced good results in the range of 0.7-0.87, with a large drop in performance for most values outside this range. This is not surprising but, unlike the lumbar spine, there was both a lower and upper cut-off. For the X and Y inputs, Fig. 5.9b and Fig. 5.9c show constant performance until about 8 pixels from the origin, giving a much narrower range for selecting the initial point than for the lumbar spine. This makes sense given the relative size of the structures. The method was also very robust to window sizes (Fig. 5.9d). Any window size in the range of 8x8-13x13mm^2

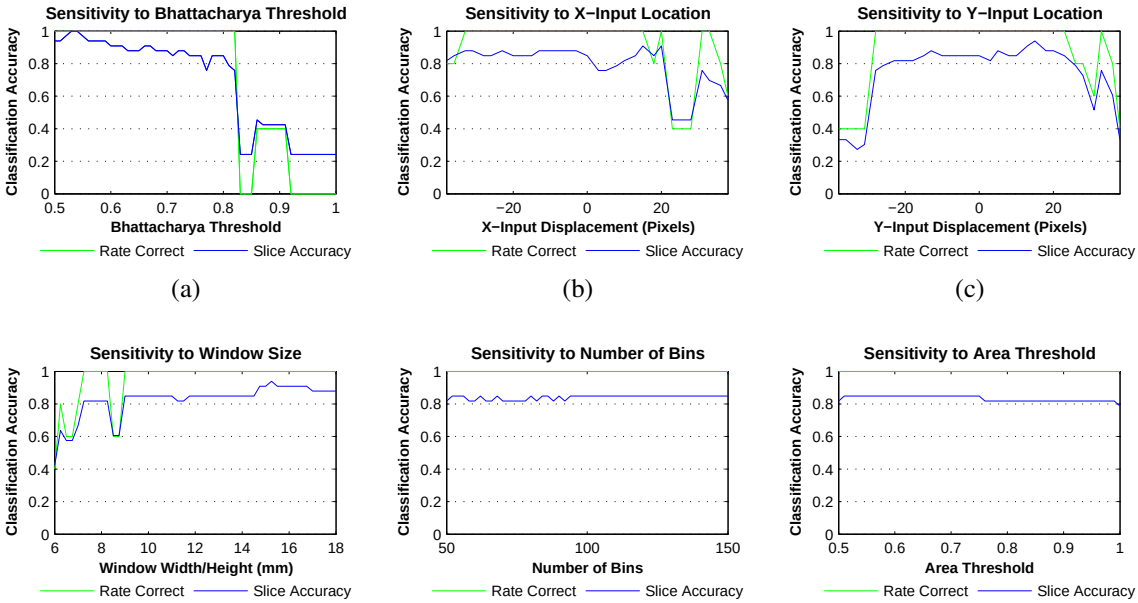


Figure 5.8: Lumbar spine analysis of the sensitivity of the algorithm to changes in various parameters:, with original values given in parenthesis a) Bhattacharya threshold (0.75), b) X-input location (0), c) Y-input location (0), d) Window Size (12), e) Number of histogram bins (100), f) Area threshold (0.75).

showed excellent performance, with a cut-off above and below these values. The upper cut-off is likely linked to the size of the structures in the cervical spine, whereas the lower cut-off is likely due to small input size decreasing the classification accuracy. Again, surprisingly the method was not sensitive to the number of bins (Fig. 5.9e) or the area threshold (Fig. 5.9f), demonstrating that these are not important parameters in the algorithm and a possible speed up could be realized by reducing the number of bins.

5.4.3 Run-time results

Table 5.5 shows the runtime along with the computational complexity of our algorithm using a conventional calculation, the integral image method and a GPU implementation of the integral images. The GPU based method required 2.95 second to complete compared to the CPU method, which required 49.2 seconds. This was based on a T1 lumbar spine 3D image with 42 axial slices.

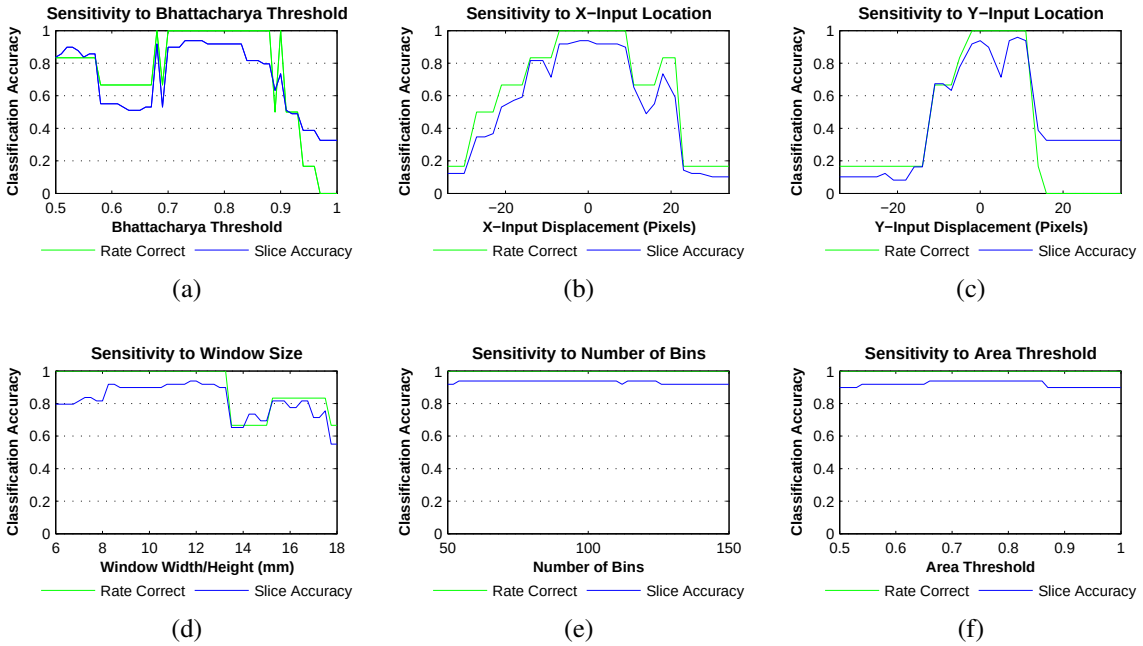


Figure 5.9: Cervical spine analysis of the sensitivity of the algorithm to changes in various parameters, with original values given in parenthesis: a) Bhattacharya threshold (0.75), b) X-input location (0), c) Y-input location (0), d) Window Size (8), e) Number of histogram bins (100), f) Area threshold (0.75).

Table 5.5: Runtime for 42 axial slices. Local window (w x h) is 50x50 pixels, n is the number of pixels in the image, and z is the number of kernel features. The computational order of the integral image method is independent of window size.

	CPU-conventional	CPU-Integral Images	GPU-Integral Images
Runtime (s)	N/A	49.2	2.95
Order	$w \times h \times n \times z$	$n \times z$	$n \times z$

5.5 Conclusion

We proposed an efficient algorithm for the classifications of T1-weighted, axial MR images. We have used an extension of integral images for the efficient computation of image intensity distributions around local image windows. A probability product kernel method was used with the intensity distributions to classify pixels in the image. We can classify vertebra and disc structure using these pixel classifications and simple geometric rules.

The classification accuracy for lumbar spine vertebrae was found to be 99% and for cervical spine vertebrae to be 79%, with a substantial increase to 89% when only 4 cases with large deformations were removed. The algorithm also executes in near real-time, taking only a few seconds to label all the 2D slices in a 3D image. In the single lumbar-spine case where the algorithm incorrectly labeled a vertebra, only one slice was incorrect, with the remainder of the slice being given the correct labeling. This should be easily identified by a human operator and not lead to medical errors.

For the cervical spine, the performance is good, but not as good as for the lumbar spine. There are some vertebrae that were incorrectly labeled, requiring clinicians to check the results carefully and manually correct any errors. A sample workflow would be to label the image and have the clinician check the results, indicating which vertebrae are correct. A new input point would then be given for a previously missed or incorrectly labeled vertebra with this being used to attempt to classify any remaining vertebrae. This decrease in accuracy is likely because of the significantly smaller structures in the cervical spine compared to the lumbar spine. The maximum errors between the detected uppermost and lowermost vertebrae boundaries and the ground truth boundaries are high for the cervical spine. This may be in part due to the C2 vertebrae having a larger height than the other cervical vertebrae. This is not currently accounted for in the algorithm, but could be the subject of future improvements.

The lower accuracy of the cervical spine algorithm does preclude it from being used as an unsupervised input to other algorithms; however, its quick run-time makes interactive use possible. The lumbar spine algorithm is very suitable to being used as an unsupervised input.

Limitations to this approach include the need for accurate first slice detection. The user is required to present at least one valid input. Errors in the labeling of the initial point will be propagated by the algorithm. This work is also limited by the lack of pathological test cases, so its performance in those circumstances is uncertain. It should be noted that even with these limitations, if this method works for most clinical cases then it is likely to be used by clinicians.

In this work a machine learning based approach was taken as opposed to segmentation for reasons of speed and accuracy. Although we don't doubt that it may have been possible to accurately segment the spine, in our experience, segmentation of 3D volumes can be computationally expensive. There is also the question of accurate segmentation of the vertebrae and how the errors would propagate to labeling accuracy. Furthermore, since labeling is a classification problem, a classification as opposed to segmentation approach was taken.

In the future, we will test this method on other MR modalities, such as proton density, as well as CT images. We expect excellent results, as recognition is based solely on a learnt intensity distribution, which is not limited to any specific modality.

5.6 Acknowledgements

Brandon Miles would like to acknowledge the support of the Graduate Program in BioMedical Engineering and the Computer Assisted Medical Intervention (CAMI) training program, which is funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Bibliography

[1] Raja' S. Alomari, Jason J. Corso, and Vipin Chaudhary. Labeling of lumbar discs using both pixel- and object-level features with a two-level probabilistic model. *IEEE Trans. Med. Imaging*, 30(1):1–10, 2011.

[2] James L. Berry, James M. Moran, William S. Berg, and Arthur D. Steffee. A morpho-

metric study of human lumbar and selected thoracic vertebrae. *Spine*, 12(4):362–367, 1987.

[3] Huang-Wei Chang and Hwann-Tzong Chen. A square-root sampling approach to fast histogram-based search. In *CVPR 2010*, 2010.

[4] R. A. Deyo, S. K. Mirza, J. A. Turner, and B. I. Martin. Overtreating chronic back pain: time to back off? *J. Am. Board. Fam. Med.*, 22:62–68, 2009.

[5] D. F. Fardon and P. C. Milette. Nomenclature and classification of lumbar disc pathology: Recommendations of the combined task forces of the north american spine society, american society of spine radiology, and american society of neuroradiology. *Spine*, 26 (5):E93–E113, 2001.

[6] Issachar Gilad and Moshe Nissan. A study of vertebra and disc geometric relations of the human cervical and lumbar spine. *Spine*, 11(2):154–157, 1986.

[7] B. Glocker, D. Zikic, E. Konukoglu, D. R. Haynor, and A. Criminisi. Vertebrae localization in pathological spine ct via dense classi
cation from sparse annotations. In *MICCAI: LCNS Spinger (Heidelburg)*, 2013.

[8] Ben Glocker, Johannes Feulner, Antonio Criminisi, David R. Haynor, and Ender Konukoglu. Automatic localization and identification of vertebrae in arbitrary field-of-view ct scans. In *MICCAI (3)*, pages 590–598, 2012.

[9] J.W. Bell H. Schweitzer and F. Wu. Very fast template matching. In *ECCV*, 2002.

[10] John Hsiang. Wrong-level surgery: A unique problem in spine surgery. *Surgical Neurology International*, 2:47, 2011.

[11] S.H. Huang, Y.H. Chu, S.H. Lai, and C.L. Novak. Learning-based vertebra detection and iterative normalized-cut segmentation for spinal mri. *IEEE Trans. on Medical Imaging*, 28(10):1595–1605, 2009.

[12] Tony Jebara, Risi Imre Kondor, and Andrew Howard. Probability product kernels. *Journal of Machine Learning Research*, 5:819–844, 2004.

[13] Tobias Klinder, Jrn Ostermann, Matthias Ehm, Astrid Franz, Reinhard Kneser, and Cristian Lorenz. Automated model-based vertebra detection, identification, and segmentation in ct images. *Medical Image Analysis*, 13(3):471–482, 2009.

[14] Fabian Lecron, Jonathan Boisvert, Said Mahmoudi, Hubert Labelle, and Mohammed Benjelloun. Fast 3d spine reconstruction of postoperative patients using a multilevel statistical model. In *MICCAI. LNCS, Springer, Heidelberg*, pages 446–453, 2012.

[15] Jun Ma and Le Lu. Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. *Computer Vision and Image Understanding*, 117 (2013):10721083, 2013.

[16] Jun Ma, Le Lu, Yiqiang Zhan, Xiang Sean Zhou, Marcos Salganicoff, and Arun Krishnan. Hierarchical segmentation and identification of thoracic vertebra using learning-based edge detection and coarse-to-fine deformable model. In *MICCAI (1)*, pages 19–27, 2010.

[17] Milan G. Mody, Ali Nourbakhsh, Daniel L. Stahl, Mark Gibbs, Mohammad Alfawareh, and Kim J. Garges. The prevalence of wrong level surgery among spine surgeons. *Spine*, 33(2):194–198, 2008.

[18] Ayse Betül Oktay and Yusuf Sinan Akgul. Localization of the lumbar discs using machine learning and exact probabilistic inference. In *MICCAI (3)*, pages 158–165, 2011.

[19] Pamela L. Owens, Maeve Woeltje, and Ryan Mutter. Emergency Department Visits and Inpatient Stays Related to Back Problems. 2008. URL `http://www.hcup-us.ahrq.gov/reports/statbriefs/sb105.pdf`.

[20] Fatih Porikli. Integral histogram: A fast way to extract histograms in cartesian spaces. In *CVPR*, 2005.

[21] Martin G. Roberts, Timothy F. Cootes, and Judith E. Adams. Automatic location of vertebrae on dxa images using random forest regression. In *MICCAI (3)*, pages 361–368, 2012.

[22] Sebastian Steger and Stefan Wesarg. Automated skeleton based multi-modal deformable registration of head & neck datasets. In *MICCAI. LNCS, Springer, Heidelberg*, pages 66–73, (2012).

[23] Paul A. Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[24] Yichen Wei and Litian Tao. Efficient histogram-based sliding window. In *CVPR*, 2010.

[25] Jianhua Yao, Joseph E. Burns, Hector Munoz, and Ronald M. Summer. Detection of vertebral body fractures based on cortical shell unwrapping. In *MICCAI. LNCS, Springer, Heidelberg*, pages 509–516, (2012).

[26] Yiqiang Zhan, Maneesh Dewan, Martin Harder, and Xiang Sean Zhou. Robust mr spine detection using hierarchical learning and local articulated model. In *MICCAI (1)*, pages 141–148, 2012.

Chapter 6

Summary, Conclusions and Future Work

6.1 Summary of Work

The goal of this work is the creation of tools for increasing the accuracy and efficiency of radiological diagnosis of the lumbar spine. Two problems in particular have been considered. The first problem is CT/MR spine image fusion and the second is automatic labeling of MR spine images. To solve these problems, the development of novel algorithms along with software prototypes simulating their clinical use was completed. Due to the nature of image fusion, a human observer study was used to validate its clinical performance, whereas the axial labeling algorithm results can be directly compared to labels identified manually. The following is a summary of these two projects.

6.1.1 Graph Cut Image Fusion

In chapter 2 we have developed a method of spine image fusion utilizing an energy minimization approach solved via graph-cuts. The goal of this technique was to combine the cortical bone detail, present in CT images with the soft tissue of the spine present in T2-weighted MR images. Our energy minimization approach avoids the introduction of artifacts present in transform-based approaches of image fusion [5, 10]. Compared to other energy minimiza-

tion approaches, we can guarantee a near global optimum energy value and avoid the intensive computations that are common with gradient descent based methods.

We formulated the energy minimization problem as an optimization of three competing terms. The goal of the first term is to create an image that is similar to the input MR image, weighted for strong edges. The goal of the second term is to create an image that is similar to the input CT image, again weighted for strong edges. The final term aims to create an output image that contains natural transitions between neighbouring pixels. To reduce the computational complexity of the problem, it has been reformulated as the problem of finding an optimal alpha labeling, which represents the percentage of each pixel that comes from the MR image and the percentage of each pixel that comes from the CT image. The reformulated problem can be efficiently solved using graph-cuts via swap moves from combinatorial optimization [2].

To test this approach a total of 20 MR/CT image pairs were registered based on soft tissue correspondence. These were then preprocessed and fused. The resulting images were compared to existing methods both visually and numerical. The images fused using graph-cuts did not contain noticeable artifacts unlike wavelet [5] and Contourlet [10] approaches. It also preserved the dynamic range of the output images unlike Piella's variational approach [7].

To numerically compare the images, manually segmented bone and tissue masks were generated. Fusion errors were calculated based on the pixel-wise difference between the fused images and MR images in the areas of soft tissue and the fused images and CT images in the areas of bone detail. The errors for the graph-cut fusion method were less than all the other methods, which is congruent with a better fused image. Our method also outperformed the other methods when compared using a sensitivity analysis and structural similarity index measure.

This fusion system provides an enhanced visualization of the lumbar spine and outperforms existing methods from the literature in terms of visual appearance and numerical accuracy. It is hoped that more accurate and more efficient radiological diagnoses can be realized through the use of these fused images.

6.1.2 Real Time Image Fusion via Convex Approaches

In order to address some of the deficiencies of the graph-cut approach from chapter 2, we explored convex optimization approaches to image fusion. The largest drawback to the graph-cut methods is that they are computationally intensive to calculate [4]. By being non-parallel in nature they are not amenable to efficient implementation on a GPU and do not achieve real-time performance. Multi-label graph-cuts are also not practical to implement for 3D images, due to increased memory and computational load. The largest drawback this imposes is that for clinical use the fused images would need to be calculated offline, precluding user interaction with the fused images.

In chapter 3 we considered convex approaches to solving similar energy functionals to the one proposed in chapter 2. Here convex methods were employed as opposed to graph-cut approaches. We presented two new families of fusion algorithms based on optimizing a squared difference (SD) between the fused image and input MR/CT images, and an absolute value difference (ABS) between the fused image and input MR/CT images. We presented efficient convex algorithms (CSD) and (CABS) for solving the SD and ABS energy formulation.

Based on 30 MR/CT images pairs, the CSD and CABS are comparable to the graph methods when validated visually and numerically. CABS performs slightly better numerically than CSD and graph-cuts, however, the CSD method performs better visually than CABS and similarly to graph-cuts both visually and numerically. When run on the CPU, the CABS method takes longer than the graph-cut methods, while the CSD method has a similar running time (10-15 minutes). When implemented on a GPU, the CSD method can calculate an entire fused volume in under a second, providing real-time performance. The CABS takes about 2 second on the GPU. Additionally these runtime results scale linearly with the number of pixels in the image.

Through the use of case studies the clinical importance of these images was presented. In particular, the importance of being able to adjust the settings of the fused images interactively was demonstration. For a patient with multiple metastases in their spine, the standard fusion

algorithm settings (such as those used in the graph-cut approach) did not show the metastases in the fused images. When alternate settings were determined interactively in real-time, the metastases could be clear highlighted. Additional settings clearly highlighted the cortical bone detail and the metastases on a single fused image.

The results of this study paved the way for a clinical validation of the interactive real-time image fusion system.

6.1.3 Clinical Analysis of Image Fusion

In chapter 5 we studied the clinical use of the real-time interactive fusion system developed in chapter 4. To evaluate its clinical potential for lumbar spine diagnosis, the fusion system was compared to the use of unregistered MR and CT images, which is the current clinical standard.

Based on the readings of 5 radiologists, the unregistered MR and CT images wee compared to registered CT, MR and fused images. The images were presented using custom image fusion viewing software. The readers were asked to rank each presented modality based on confidence of diagnosis in eight areas of anatomy, along with their overall confidence in diagnosis and whether the images were aesthetically pleasing. They were also asked to measure the diameter of the largest protrusion in both the axial and sagittal views. These tasks were timed in order to determine whether they were completed more efficiently using the fused images.

It was found that the readers had similar confidence or higher confidence using the fused images compared to the MR images except for diagnosis the exiting nerve but more confidence compared to the CT images. For the measurement task there was a similar variability when measuring protrusions on the fused image versus the unregistered MR/CT images. In regard to time for completing the rankings and measurements, we did not find an overall difference between using the fused images and the unregistered images.

Based on these results, it can be concluded that the fused images do provide a benefit to the existing MR and CT images as they generally increase the reader's overall confidence in diagnosis. They also did not increase the time required to complete the individual tasks even

though they added additional options when reading the images. This image fusion is targeted for inclusion in future versions of the GE PACS radiology software.

6.1.4 Spine Axial Annotations

In chapter 6 we considered the problem of properly annotating the level of vertebrae and discs in axial MR spine images. Like image fusion, this project is motivated by increasing the accuracy and efficiency of diagnosis by radiologists. The precise level of a given vertebrae serves as a reference for surrounding tissue in the spine and incorrectly identified vertebrae can lead to wrong level surgery [3, 6] or incorrect diagnosis. Identifying this level is also a time consuming task for radiologists.

We have developed a one click method to label axial slices in both lumbar spine and cervical spine MR images. This is based on the user identifying a single point inside a single vertebra and providing the correct annotation for that vertebrae. This avoids the need to constantly cross reference location based on the sagittal slices, or mentally record the vertebral level.

Using integral images [9] and geometric constraints, classifications can be made for pixels, then slices, then whole vertebra followed by multiple vertebrae in an image. When implemented on the GPU results are achieved in near-real time. The method has been tested on T1 lumbar spine images, and T2 cervical spine images. We have achieved 99% classification accuracy on the lumbar spine and 79% accuracy on the cervical spine.

This algorithm should aid in diagnosis of the spine. It can also serve as an input to future algorithms for spine segmentation and registration. Ideally both this and the image fusion work will be implemented in future versions of the GE PACS radiology software.

6.2 Additional Findings

6.2.1 A note on Colour Imaging

Colour imaging has the potential to allow humans to visualize multiple data sources on a signal image. Intuitively this could be a powerful method for enhanced visualization of medical images. It could eliminate ambiguity between detail that presents with similar intensity levels in both CT and MR images.

To this end, the ability to visualize the cortical bone detail from the CT image on a fused image was incorporated into the software Fig. 6.1. At least 8 radiologists were given the opportunity to test the fusion software and were instructed in the availability of the colour enhanced image. Without fail every radiologist ended up turning off the colour enhancement, preferring grey scale images. This may be because of familiarity, since radiologist's typical workflow involves grey scale images for spine diagnosis, but that is only a potential hypothesis. It does present a possible direction for future research.

Figure 6.1: Custom fused image viewer showing the cortical bone detail from the CT highlighted in orange.

6.3 Limitations of the Thesis

It is important to not overstate the contributions of this work. Here we present a discussion of the major limitations of this thesis, in order to provide motivation for future research directions.

6.3.1 Graph Cut Image Fusion

In chapter 2 a first approach to image fusion is presented. An assumption of rigid deformation between CT and MR images, correctable by rigid registration, has been made. This is a strong assumption. As discussed in that chapter it has been verified to be accurate for the given dataset, however there is possibility of anatomical repositioning between scans, especially if these techniques are to be extended to the cervical spine. Piece-wise rigid registration techniques will be discussed as a future area of research. Additional sources of registration error could include: anatomical changes (e.g. disc deformations, bone healing, tumour growth, inflammation) between scans. The limitations of registration accuracy are something that the clinicians utilizing these techniques would need to consider if fusion is incorporated into radiological practice.

The techniques presented in this chapter also only involve CT and T2 image fusion. An extensive evaluation of fused CT/T1 and CT/PD images has not been performed. Extensions to C-spine have also not been thoroughly evaluated, although a small number of cases have been examined. These are areas of future investigation.

In regards to the algorithms utilized in this work, the largest drawbacks are 1) speed of computation, 2) lack of extensibility to 3D, and 3) utilization of a fixed number of labels quantizing the output results. These are considered in chapter 3.

6.3.2 Real Time Image Fusion via Convex Approaches

As these results are an extension of those presented in chapter 2, the limitations of registration and lack of verification on the C-spine and CT/T1, CT/PD images are present.

This work does address the 3 main algorithmic deficiencies of chapter 2, however a limita-

tion of this work is the assumption of isotropic voxel spacing that was made when extending the method to 3D. Some investigations were made into this. In particular it is possible to update the smoothing of the edge weights to account for a non-isotropic voxel spacing.

6.3.3 Clinical Analysis of Image Fusion

As the testing this chapter involved human observers it was the hardest to control experimentally. As has been discussed in chapter 4, one of the largest limitations is that the unregistered images were examined first, followed by the registered and fused images. There is potential bias due to the order the images were presented.

Since the registered and fused images were presented together, there is potential that the effects seen were more a result of the registration and linking of the images rather that the fusion of the images. There are indications from the data, that this is not the case, however these are indirect. A future study could include a separate testing phase for registration in addition to fusion.

6.3.4 Spine Axial Annotations

In this work the largest limitation is the lack of evaluation on highly pathological cases. It would be beneficial to understand the performance of this algorithm for severely diseased spines. It should be noted that everyone's spine is undergoing degeneration and as such every patient has some degree of degeneration. It would be beneficial to better understand the limits of the labeling algorithm. The performance on the cervical spine is another limitation of this work. This is further discussed in the section on future work.

6.4 Future Work

This project has succeeded in many respects. Image fusion has been shown to be feasible and desirable in routine radiological diagnosis of the spine. In spite of this success, there are

several drawbacks to the current implementation of the registration, which remain areas for future investigation. Addressing these areas should further improve the system and result in a more desirable end product. In parallel with these improvements, the system is ready for the investigation of new clinical applications, both for the spine and other areas of anatomy such as the hand and upper limb.

6.4.1 Real Time Spine Registration

Currently registration stands as a bottleneck for the speed of the fusion. It currently takes about 5-15 minutes and is completed offline, as only the fusion needs the constant input of the radiologist. This makes the system slightly more difficult to integrate into an end product. It would be ideal to maintain the accuracy of the current registration, with a real-time or near real computational speed.

Near real-time registration speeds of 5-10 second per volume would allow radiologists to use the fused images without offline registration, while run times of less than 1 second per volume would be ideal. This would increase the flexibility of image fusion as a tool in general. Further increases in speed could be considered for applications to image fusion for surgery. This would likely involve the use of ultrasound or fluoroscopy instead.

Two possible directions for the increase in registration speed are algorithm optimization and parallelizing the code. Currently, the code runs on the CPU, a GPU implementation could potentially speed up the code. This would require a parallel implementation of the metric and the resampler. This may be enough to achieve real-time performance, assuming a one hundred times speed up by switching to the GPU. Since this would only be a reimplementation of existing techniques, its research potential may be limited.

An alternative to this method is to first segment the images and then utilize surface matching approaches to quickly align these structures. Max-flow based segmentation algorithms could be employed for this [1, 11].

6.4.2 Piecewise Deformable Registration of the Spine

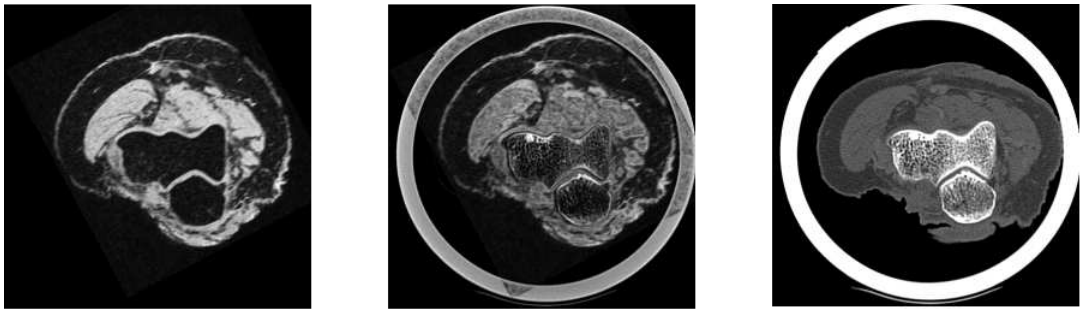
One of the greatest difficulties with our current registration methods is that they are rigid. It was found that assuming the lumbar spine is rigid was valid for most of the subjects examined. However, the cervical spine is likely to demonstrate more deformation between scans. Creating an algorithm that is capable of rigidly registering these structures would be advantageous.

Piecewise registration of the spine has been proposed in recent literature [8]. The idea behind this approaches is to treat individual vertebrae as rigid structures since they consist of bone and align these with the MR image. Soft tissue deformations could then be interpolated from these rigid deformations and/or a combination of image information.

Having the axial slice boundaries, would be an excellent starting point for this algorithm. These boundaries, which are already determined by the axial labeling algorithm, could provide initial localization. From there, rigid registration methods could be employed locally on each vertebra, using either the existing mutual information based method, or another rigid registration method. Methods involving segmentation of the images and then applying surface based registration might be of higher value as long as they can easily be extended to other MR images such as T1 and PD. Ideally these methods would also work in near real-time (5-10 second per volume).

6.4.3 Extensions of Image Fusion to T1, PD and other MR modalities

This work has primarily focused on the fusion of CT and T2-weighted lumbar spine images. Fusion of CT images with T1 images, PD images or other MR modalities may be advantageous. Ideally the current fusion methods would only need minor modifications to work in these cases. Registration would need to be examined. Directly registering the CT to T1/PD/Other images could be investigated. If these methods fail, then registering to the T2 images, which would in turn be registered to the CT images, may be possible. The new fused results could then be compared to the CT-T2 fusion results in a clinical study.



MR Fused CT

Figure 6.2: Fused images of an elbow, showing views of the CT image, fused image and MR image.

6.4.4 Other Clinical Applications

There are other potential clinical targets for image fusion. Preliminary work with elbow images has demonstrated that it is possible to fuse these images, showing both the bone and soft tissue on a single view. This is shown in Fig. 6.2. Other targets could include, the shoulder, knee, hand and foot. They are all musculo-skeletal targets containing soft tissue and bone structures, which would be desirable to see on a single image. Additional registration methods may be needed for these images; however, the current lumbar spine methods can be used as a starting point for this investigation.

6.5 Conclusion

We have developed and validated two systems designed for increasing the accuracy and efficiency of radiological diagnosis. This has included the generation of prototype viewers incorporating these tools. Both tools show promising results and are now ready for inclusion in clinical products. Future work could involve examining new clinical targets of this work, along with speed and flexibility improvements to the registration algorithm employed for image fusion.

Bibliography

[1] I. Ben Ayed, H.-M. Chen, K. Punithakumar, I. Ross, and S. Li. Max-flow segmentation of the left ventricle by recovering subject-specific distributions via a bound of the bhattacharyya measure. *Medical Image Analysis*, 16:87–100, 2012.

[2] Yuri Boykov and Vladimir Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions On Pattern Analysis and Machine Intelligence (PAMI)*, 26(9):1124–1137, September 2004.

[3] John Hsiang. Wrong-level surgery: A unique problem in spine surgery. *Surgical Neurology International*, 2:47, 2011.

[4] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *ECCV*, volume 1, pages 332–345, 2008.

[5] H. Li, B. S. Manjunath, and S. K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, 1995.

[6] Milan G. Mody, Ali Nourbakhsh, Daniel L. Stahl, Mark Gibbs, Mohammad Alfawareh, and Kim J. Garges. The prevalence of wrong level surgery among spine surgeons. *Spine*, 33(2):194–198, 2008.

[7] Gemma Piella. Image Fusion for Enhanced Visualization: A Variational Approach. *International Journal of Computer Vision*, 83(1):1–11, 2009.

[8] Sebastian Steger and Stefan Wesarg. Automated skeleton based multi-modal deformable registration of head & neck datasets. In *MICCAI. LNCS, Springer, Heidelberg*, pages 66–73, (2012).

[9] Paul A. Viola and Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[10] L. Yang, B.L.Guo, and W.Ni. Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform. *Neurocomputing*, 72:203–211, 2008.

[11] Jing Yuan, Egil Bae, Xue-Cheng Tai, and Yuri Boykov. A study on continuous max-flow and min-cut approaches part i: Binary labeling. Technical report, (UCLA CAM 10-61), 2010.

Curriculum Vitae

Name:	Brandon Miles
Post-Secondary Education and Degrees:	Queen's University Kingston, ON 2007-2009 M.Sc.(Eng) University of Western Ontario London, ON 2009 - 2013 Ph.D.
.	
Related Work Experience:	Teaching Assistant Queen's Univeristy 2007 - 2009 Teaching Assistant University of Western Ontario 2009 - 2013

Publications:

B. Miles, I. Ben Ayed, M. W. K. Law G. Garvin, A. Fenster, and S. Li. Spine Image Fusion via Graph Cuts. IEEE Transactions On Biomedical Engineering, 60(5):184150, July 2013.