MRI Image Analysis for Abdominal and Pelvic Endometriosis

Wenjun Chi

Magdalen College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2012

Acknowledgements

I would like to specially thank my supervisors Professor Sir Michael Brady and Dr. Julia Schnabel for their continued encouragement and firm backing throughout my D.Phil. Without their help and support I would not have got this far. I would also like to thank Enda McVeigh for his input on the clinical side, Niall Moore for providing patient data for the project, and Professor Stephen Kennedy for arranging funding for my D.Phil.

On a personal level, I must thank my parents for their relentless support both financially and emotionally, and my wife, Carrie, for being both critical and encouraging.

Finally, this project was funded by the Oxford Biomedical Research Centre, a partnership between the University of Oxford and Oxford University Hospitals, funded by the National Institute for Health Research (NIHR).

Abstract

Endometriosis is an oestrogen-dependent gynaecological condition defined as the presence of endometrial tissue outside the uterus cavity. The condition is predominantly found in women in their reproductive years, and associated with significant pelvic and abdominal chronic pain and infertility. The disease is believed to affect approximately 33% of women by a recent study [19]. Currently, surgical intervention, often laparoscopic surgery, is the gold standard for diagnosing the disease and it remains an effective and common treatment method for all stages of endometriosis. Magnetic resonance imaging (MRI) of the patient is performed before surgery in order to locate any endometriosis lesions and to determine whether a multidisciplinary surgical team meeting is required. In this dissertation, our goal is to use image processing techniques to aid surgical planning. Specifically, we aim to improve the quality of the existing images, and to automatically detect bladder endometriosis lesion in MR images as a form of bladder wall thickening.

One of the main problems posed by abdominal MRI is the sparse anisotropic frequency sampling process. As a consequence, the resulting images consist of thick slices and have gaps between those slices. We have devised a method to fuse multi-view MRI consisting of axial/transverse, sagittal and coronal scans, in an attempt to restore an isotropic densely sampled frequency plane of the fused image. In addition, the proposed fusion method is steerable and is able to fuse component images in any orientation. To achieve this, we apply the Riesz transform for image decomposition and reconstruction in the frequency domain, and we propose an adaptive fusion rule to fuse multiple Riesz-components of images in different orientations. The adaptive fusion is parameterised and switches between combining frequency components via the mean and maximum rule, which is effectively a trade-off between smoothing the intrinsically noisy images while retaining the sharp delineation of features. We first validate the method using simulated images, and compare it with another fusion scheme using the discrete wavelet transform. The results show that the proposed method is better in both accuracy and computational time. Improvements of fused clinical images against unfused raw images are also illustrated.

For the segmentation of the bladder wall, we investigate the level set approach. While the traditional gradient based feature detection is prone to intensity non-uniformity, we present a novel way to compute phase

congruency as a reliable feature representation. In order to avoid the phase wrapping problem with inverse trigonometric functions, we devise a mathematically elegant and efficient way to combine multi-scale image features via geometric algebra. As opposed to the original phase congruency, the proposed method is more robust against noise and hence more suitable for clinical data. To address the practical issues in segmenting the bladder wall, we suggest two coupled level set frameworks to utilise information in two different MRI sequences of the same patients - the T_2- and T_1-weighted image. The results demonstrate a dramatic decrease in the number of failed segmentations done using a single kind of image. The resulting automated segmentations are finally validated by comparing to manual segmentations done in 2-D.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The work of this dissertation aims to improve the non-invasive diagnosis of endometriosis via state-of-the-art image processing techniques. Although most of the work is technical and theoretical, the clinical problems we address are in fact very practical. In this chapter, we first broadly introduce the clinical aspects of the problem and the clinical end-point of this project; and then give a brief review of the main imaging modality we work on - magnetic resonance imaging, paving the way for in-depth analyses of the cutting-edge image processing methods that help us achieve our goals.

1.1 Introduction to Endometriosis and Clinical Motivation

Endometriosis is an oestrogen-dependent gynaecological condition defined as the presence of endometrial tissue outside the uterus cavity. Most frequently the ectopic endometrial spots affect the pelvic organs and the peritoneum. The condition is

predominantly found in women in their reproductive years, and disappears spontaneously after the menopause. The disease is associated with significant pelvic and abdominal pain during menstruation (dysmenorrhoea), painful intercourse (dyspareunia), painful urination (dysuria), spontaneous pain outside menstrual periods, and infertility. The psychological impact is often aggravated by an adverse effect of the disease on fertility. The exact prevalence of endometriosis is unclear, because laparoscopic examination of healthy women is rare. However, it is reasonable to assume the prevalence of the condition to be around 5 - 10%. According to a relatively old study conducted in 1991, the prevalence of endometriosis in asymptomatic women was 45.3% [40], which was higher than many previously believed at the time. In a more recent study in 2006, Guo and Wang investigated 11 previous published studies, adjusted for the difference in their year of publication, sample size, and symptom evaluation methods, and concluded that the prevalence among women with chronic pelvic pain is approximately 33% [19].

To date, the exact pathogenesis of endometriosis is unknown; a retrograde menstruation with implantation of viable endometrial cells seems to be a key component in the development of endometriosis. However, retrograde menstruation can be observed in 90% of women [20, 1], suggesting the involvement of additional factors in the implantation and growth of endometriotic lesions in women who go on to develop the disease [16]. Other theories include celomic metaplasia, embryonic cell rests, and lymphatic and vascular dissemination. However, no single theory can explain the location of ectopic endometrium in all cases of endometriosis. Other factors are also considered in a diagnosis of the disease [15]. These include genetic factors, menstrual

factors, environmental factors, and changes in hormones and/or immunity.

The biological factors affecting any individual's risk of developing endometriosis are poorly understood. Endometriosis is considered to be both an estrogen dependent and an inflammatory-immunological disease. The presence of ectopic endometrium in the peritoneum elicits a classical inflammatory response, which leads to the formation of adhesion between different locations of endometriosis lesions in the pelvis.

The principal objective in treating endometriosis is symptom-relief management. In addition to relieving pain, the goals of treatment for patients with endometriosis are to prevent or delay disease progression by reducing endometriosis implants through surgical treatment or medically induced atrophy of the implants. Surgery, alone or in combination with medical therapy, remains a common treatment method for all stages of endometriosis [41, 52].

Small deposits of endometrial cells scattered over the peritoneal, serosal, and ovarian surfaces are called peritoneal implants or lesions. They are typically described as superficial "powder-burn" or "gunshot" lesions due to their puckered bluish-black appearance. Atypical or "subtle" lesions can be found as red implants and serous or clear vesicles. Such lesions and implants can be easily removed during laparoscopic surgery.

When these implants are at least 5 mm beneath the peritoneal surface, the disease is known as deep infiltrative endometriosis, whose first description dates back to 1922 by Sampson [42]. When these tissues develop fibrosis, i.e. well-circumscribed nodular aggregates of hyperplastic smooth muscle and glandular elements, it is then called an endometriosis nodule. Ultimately, we are interested in deep endometriosis

implants with bladder and rectosigmoid involvement, two of the most severe forms of endometriosis. Identification of these two locations of deep endometriosis is clinically important, as management of urinary or rectal locations will lead to a multidisciplinary surgical approach. However, from an image processing point of view, the bladder and rectum are two structures that consist of very different salient features in images and require different image processing techniques. In this dissertation, we mainly focus on analysing the former type of endometriosis with bladder involvement, which usually leads to abnormal bladder wall thickening.

Clinically, the most common sites of deeply infiltrating endometriosis lesions in the pelvis are the uterosacral ligaments, vagina, bowel and bladder. The bowel was involved in 9.9% and the bladder was affected in 6.4% in a series of 241 patients with pathologically proven lesions of deep endometriosis (excluding adhesions) [25]. The diagnosis of endometriosis is based on the histological finding of ectopic endometrial tissue. In order to obtain this tissue for histology, surgical intervention is required. However a 'presumptive' diagnosis of endometriosis can be made based on the clinical history, a physical examination, and a radiological investigation. Obtaining the 'presumptive' diagnosis serves two purposes: 1. it can allow for symptomatic management of the condition without surgical intervention; and 2. the information obtained through the clinical history, physical examination, and radiological investigations can enable detailed planning of the surgical invention.

Bladder endometriosis is represented as localised bladder wall thickening, with occasional protrusion inside the bladder lumen. The lesion of the bladder detrusor is located anterior to the vesicouterine pouch and sometimes in contact with an ad-

enomyotic nodule of the anterior wall of the uterus [12]. Deep bowel endometriotic implants, on the other hand, are divided into four subtypes [26]:

- Type I (Figure 1.1 a.) - is characterised by a large pelvic area of typical and sometimes some subtle endometriotic lesions surrounded by white sclerotic tissue. Only during excision does it become obvious that the endometriotic lesion infiltrates deeper than 5 mm. Typically, the endometriotic area becomes progressively smaller as it grows deeper, and the lesion is thus cone shaped.

- Type II (Figure 1.1 b.) - lesions are characterised by retraction of the bowel. Clinically, they are recognised by the obvious bowel retraction around a small typical lesion. In some women, however, endometriosis cannot be seen through the laparoscope, and the bowel retraction is the only clinical sign. In some women, even the retraction is barely seen and the induration can hardly be felt. Only during excision does this type of endometriotic nodule become apparent, emphasising the need for a pre-operative diagnosis.

- Type III (Figure 1.1 c.) - lesions are spherical endometriotic nodules in the rectovaginal septum. This type of lesion is the most severe, and is often spread laterally up and around the uterine artery, sometimes causing sclerosis around the ureter.

- Type IV - sclerosing endometriosis, which is similar to the rectal endometriosis, invades the sigmoid, but is situated 10 cm above the rectovaginal septum.

Surgery for bladder and bowel endometriosis can be complex and challenging and

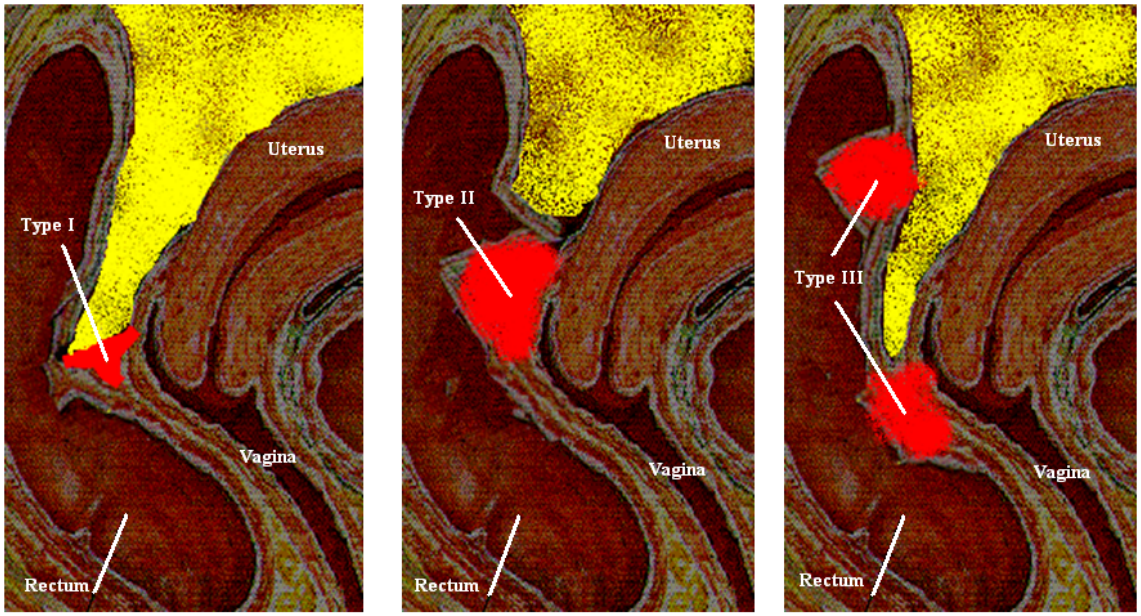

Figure 1.1: Schematic diagram of endometriosis lesions by type. a). Type I - typical and subtle lesions in cone shape. b). Type II - progressive lesions that cause bowel retraction. c). Type III - lesions that are in the form of spherical nodules and are often spread laterally up. Image sourced from E. McVeigh's presentation slides at the Colerectal Meeting at Verbier in 2007.

it often involves a multidisciplinary team, for example, in the case of type III and type IV lesions with bowel involvement. These types of lesions, usually along with bowel wall partial- or full-thickness defect, can lead to the need for enterotomy or colorectal resection, which is associated with increased morbidity and mortality to the patient. As the choice of surgery depends on the patient's wishes as well as on the surgical skill available, pre-surgical diagnosis of deep endometriosis, especially bladder and rectovaginal endometriosis, is essential in order to inform patients about the specific risks of surgery. From the doctor's perspective, the clinical usefulness for these investigations will be prediction of the difficulty of surgery, i.e. which endometriosis case can be operated by the gynaecologist and which case should be referred.

1.2 Clinical Diagnosis of Endometriosis and Clinical End Point

Ideally, the diagnosis of deep infiltrating endometriosis should precede surgery. A retrospective analysis has shown that by following routine clinical examination the percentage of correct diagnoses is only between 14.4% and 43.1%, varying significantly with the location of the implant [7]. Laparoscopy is currently the gold standard for the diagnosis of endometriosis when black or brown/blue nodules, resulting from the repeated haemorrhage and retention of haemosiderin, are seen on the peritoneal surfaces of structures in the pelvis. Other diagnostic methods include:

- Serum markers - there is growing evidence that carcinoembryonic antigen CA125 may help to evaluate selected populations at risk, as the concentrations of CA125 are increased in women with deep endometriosis and in women with endometriomas, a cystic form of endometriosis at the ovaries. However, this method still lacks the necessary sensitivity and specificity to serve as reliable screening tests for endometriosis and the ability to localise lesions.

- Ultrasound - this is the method of choice for the first-line examination of endometriosis due to its ease of use and immediate availability. The sensitivity and specificity of diagnosing ovarian endometriosis, or endometriomas using ultrasound are high because of the cystic nature of the lesion. However, it is less good at predicting the extent of deep infiltrating endometriosis, which often involves fibrotic nodules in muscles. Rectal endoscopic sonography has been reported to

have a sensitivity (the percentage of total true positives are correctly identified) and specificity (the percentage of total true negatives are correctly rejected) of 82% and 88% [3], respectively for the diagnosis of rectosigmoid endometriosis. However, ultrasound is limited by a small field of view and the difficulty in accurate anatomical mapping for surgical planning, which is one of the goals of our work.

- Computed Tomography (CT) - endometriomas may appear solid, cystic, or mixed solid and cystic, resulting in an overlap in the appearances with an abscess or ovarian cyst. However, owing to the poor specificity and high radiation dose, use of CT in the evaluation of pelvic endometriosis has been largely replaced by magnetic resonance imaging.

- Magnetic Resonance Imaging (MRI) - identification of endometriomas by MRI relies on the detection of pigmented haemorrhagic lesions. Sensitivities and specificities of 90% and 98%, respectively, have been achieved in diagnosing endometriomas [47]; and 90% and 91%, respectively, have been reported for deep endometriosis [2]. MRI will identify endometriosis in all locations with a high degree of accuracy, which seems the most promising MR imaging protocol for detecting deep implants as compared to the other modalities.

Typically, endometriosis lesions appear hyperintense on T_1 relaxation time weighted images and hypointense on T_2 relaxation time weighted images (see Section 1.3), owing to the presence of deoxyhaemoglobin and methaemoglobin. However, the signal characteristics may vary according to the age of the haemorrhage. Table 1 lists the

	T_1	T_2
Oxyhaemoglobin	Slightly dark	Very bright
Deoxyhaemoglobin	Dark	Very dark
Intracellular methaemoglobin	Very bright	Dark
Extracellular methaemoglobin	Extremely bright	Extremely bright
Haemosiderin	Extremely dark	Extremely dark
Bland fluid (urine)	Dark	Extremely bright
Protein containing fluid	Very bright	Very dark
Fat	Extremely bright	Very bright
Muscle	Dark	Extremely dark

Table 1.1: Intensity appearance of pelvic substances on T_1 and T_2 weighted MR images.

individual intensity appearances of common endometriosis substances on T_1 and T_2 weighted images. Acute haemorrhage occasionally appears hypointense on T_1 and T_2 sequences. Old haemorrhage occasionally appears hyperintense on T_1 and T_2 images. Multiple hyperintense lesions on T_1 sequences, irrespective of the intensity on T_2 sequences, are general characteristics of endometriosis.

Without the presence of a contrast agent, there are three substances that appear very bright on T_1 sequences – fat, blood (due to methaemoglobin) and protein fluid. T_1 fat suppressed sequences are designed to differentiate endometriosis from fat, normal flowing blood, and protein fluid.

- T_1 fat suppressed sequences - there are usually two substances that appear homogeneously bright – endometriosis and flowing blood in the arteries. Endometriomas are areas of homogeneous high intensities with clear capsulated boundaries. Endometriotic implants are bright lesions that usually cause distortion in surrounding structures. Blood vessels appear as homogeneous bright dots but also as tubular structures through sequential slices. Protein fluids are

bright but usually non-homogeneous inside the bowel structure.

- T_1 sequences - T_1 weighted images share a lot of similarities with T_1 fat suppressed sequences, except that fat appears extremely bright on T_1 images. Another way of differentiating blood vessels is that flowing blood is extremely dark on T_1 sequences, because it is difficult to magnetise flowing molecules.

- T_2 sequences - bland fluid such as urine is extremely bright on T_2 weighted images due to its low molecular density. Similarly, dense muscle tissues are extremely dark, which gives better muscle contrast than T_1 sequences. Usually, endometriosis appears dark on T_2 images due to deoxyhaemoglobin; but the intensities can vary considerably according to the age of the haemorrhage.

Note that the above observations only apply when one substance occupies an entire voxel. However, given the spatial resolution of MRI, especially the non-isotropic voxels in current MR protocols, they need to be tempered by the partial volume effect (PVE). That is, a large voxel may extend from one substance to another, which have mixed responses to MR signals. Hence, the intensity of that voxel is affected by the mix of the substances inside it. This is especially a problem for a thick slice at the boundary of an organ, where the long voxels are in the normal direction of the organ boundary. However, in clinical practice this problem does not attract much attention by radiologists.

Since no single MRI image can give all the information required for definitive clinical diagnosis, radiologists currently diagnose by visually comparing the intensities on different MRI pulse sequences. Structurally, it is difficult to examine a 3-D image.

As a consequence, radiologists visually compare MRI images in different orientations even with the same pulse sequence. Therefore, the fusion of composite information from different images into a single one is a fundamental component of this thesis work; and such fusion also naturally corrects for the intensity distortion due to the PVE.

Once we identify the areas of interest (extremely bright on T_1 and T_1 fat suppressed sequences, and very dark on T_2 sequences), we also need anatomical mapping of the signal, because we are most interested in the common regions with endometriosis involvement. The following list of sites was suggested by the collaborating radiologist:

- Uterus - is best seen on T_2 weighted sequences due to higher contrast of muscle tissues and endometrium tissues. It appears as a trilaminar structure with high signal intensity centre surrounded by a low signal intensity "band". In 3-D, the uterus is expected to be "pear" shaped.

- Ovaries - situated by the sides of the uterus. The ovary contains multiple locule structures corresponding to bland fluid signal intensities (extremely bright on T_2 sequences, very dark on T_1 images). A healthy ovary is expected to be of maximum size 4x3x2 cm.

- Rectum - is posterior to the uterus and anterior to the sacrum. The best visual identification of the rectum is that it is a tubular structure that extends across slices and is seen as a sigmoid shape in the sagittal view.

- Bladder - is anterior to the uterus with clear hypointense boundaries. It can be

easily identified since the content inside the bladder is bright on T_2 and dark on T_1 sequences.

This anatomical prior knowledge can possibly be deployed in order to enable the proposed algorithm to localise lesions automatically. The aim of this thesis is to define a new methodology for the interpretation of MRI images of the pelvic organs in women with endometriosis. Specifically, we report objectively on the degree of bladder or bowel wall infiltration by endometrial implants. Therefore, the clinical end point of this project is twofold:

1. to improve pre-operative diagnosis of endometriosis using the current MRI protocol,

2. to facilitate better pre-operative surgical planning which will include increased accountability to inform the patient of the surgical risks.

1.3 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a non-invasive 3-D imaging modality capable of obtaining exquisite soft tissue contrast. It offers wide flexibility in imaging methods and voxel size with no radiation burden, unlike X-ray and CT imaging. A powerful magnetic field in the MRI machine causes electromagnetic dipoles, denoted β_H, of hydrogen atoms to align parallel or anti-parallel to the magnetic field lines. As predicted by Newtonian physics, the addition of the torque generated by β_H creates a gyroscopic precession, or "wobble" in these atoms. The frequency of precession in this

quantum wobble is proportional to the external magnetic field strength and called the Larmor frequency [30]:

$$\omega_0 = \gamma\beta_0 \tag{1.1}$$

where ω_0 is the angular frequency of precession, β_0 is the strength of the external magnetic field in Teslas, and γ is the ratio of a particle's magnetic dipole moment to its angular momentum (a.k.a. the gyromagnetic or the magnetogyric ratio) and is equal to 42.6 $^{\mathrm{MHz}}/_{\mathrm{T}}$ for hydrogen. In additional to its precession along the main magnetic field, an additional external torque is applied to the hydrogen atom by way of an external radio frequency (RF) pulse at the Larmor (resonant) frequency. As a result of the pulse, the individual dipoles may undergo a rotation, or flip, of the net magnetic field from the Z-axis to the XY plane. Removal of the external RF field causes the dipoles to recover their original orientation. The progress of each atom through this restoration generates a measurable electromagnetic signal in the RF coils, often those used originally to emit the RF field. The time elapsed between removal of the RF field and sampling of the signal at the coils is defined as the time to echo, T_E.

In order to generate an image, the RF field is repeatedly pulsed. The net Z-magnetism attains its maximum and the XY-magnetism its minimum prior to an RF field pulse and becomes its minima and maxima afterwards, respectively. The recovery of Z-magnetism and decay XY-magnetism can be conceptualised by a change in the local frame of reference rotating with the atoms, which reveals an exponen-

tial relationship. In hardware, this change of reference is accomplished through the implementation of an envelope detection algorithm. Signals are assumed to be in a moving frame of reference, or envelope detected, in the following discussion.

The time to recover net Z-magnetism after removal of the RF field is defined as the longitudinal relaxation time, T_1. This Z-magnetism recovery is,

$$M_z = M_0 \left(1 - e^{-\frac{T_E}{T_1}}\right) \tag{1.2}$$

where M_0 is the net magnetisation moment, and M_z is the magnetisation along the Z-axis. Conversely, the rate of decay of the magnetism in the XY-axis is expressed mathematically as,

$$M_{xy} = M_0 e^{-\frac{T_E}{T_2}} \tag{1.3}$$

where M_{xy} is the magnetic moment of the hydrogen atoms in the XY plane, and T_2 is the transverse relaxation time, i.e. the time in which the XY-magnetism decays after the RF field is removed. This definition only takes into account the decay of the XY magnetic field due to magnetic torque from β_0 and not the effect of adjacent atoms contributing to the degradation of M_{xy}. These additional effects are factored in as,

$$\frac{1}{T_2^*} = \frac{1}{T_{\text{additive}}} + \frac{1}{T_2} \tag{1.4}$$

where T_{additive} is the change in decay time from additional forces, and the T_2^* relaxation time is therefore the real XY-axis decay.

MRI machines are operated by trained radiographers who, by utilising a priori knowledge of the region or disorder that is under investigation, select an appropriate T_E to create an image which highlights differences in the T_1 or T_2 times. Such image types are referred to as T_1 or T_2 weighted images. Suppose contrast between two different types of tissues a and b is desirable, the choice of the optimum T_E is guided by,

$$T_E^{\text{opt}} = \ln\left(\frac{T_b}{T_a}\right)\frac{T_a T_b}{T_a - T_b} \tag{1.5}$$

where T_a is the relaxation time of tissue a, T_b is the relaxation time of tissue b, and T_E^{opt} is the optimal T_E. This can be used to find the optimal T_E for both T_1- and T_2-weighted images.

1.3.1 Spatial Encoding

All atoms inside the MRI machine are subjected to the scanner's magnetic and RF fields. To specify the location of excitation, MRI scanners use a combination of magnetic variants, gradients and phase delays to encode atoms. For simplicity, the derivation here, and the associated explanation, refers only to gradients used in the acquisition in axial imaging. The three gradients, denoted by G_x, G_y, and G_z are called the frequency, phase and slice select gradients, respectively. The slice select gradient G_z is used to isolate atoms across a specific thickness. Because MRI scanners have a uniform magnetic field along the Z-axis, the application of a magnetic gradient causes a change in the precession frequency. Recalling the relationship between

precession frequency and the magnetic field strength by equation 1.1, the encoded precession frequency in a particular slice becomes,

$$\omega(z) = \gamma(\beta_0 + zG_z) \tag{1.6}$$

where z is the distance from the start of the gradient (origin of the image frame). By using a steep G_z gradient and a narrow band RF field, only a thin slice of atoms have a precession frequency that is affected by the subsequent RF pulses. After a slice has been isolated, additional frequency and phase gradients are applied,

$$\omega_z(x, y) = \gamma(\beta_0 + xG_x + yG_y) \tag{1.7}$$

where G_x causes atoms in different columns of the thin slice (in the xy-plane) isolated by ω_z to precess at different frequencies, and similarly G_y gives a phase offset along the rows of the atoms in the xy-plane. Frequency data from one excitation is separated using power spectrum analysis. Frequency separated information is stored in a matrix referred to as K-space. K-space is a 2-dimensional matrix of frequency (x-axis) versus phase (y-axis). By varying the phase offset, but holding the frequency and phase gradients constant across multiple excitations, the recorded data fills the K-space matrix. By performing an inverse Fourier transformation on the K-space, the location and amplitude of the local excitation can be determined by,

$$f(t) = \int\int F(x, y)\, e^{j(xG_x + yG_y)} dxdy \tag{1.8}$$

where $F(x, y)$ is the signal intensity captured at x and y. The K-space is equivalent to the frequency space where low frequency components make up a majority of the information in the image and high frequencies contain the edge data, sharpness of an object contour and noise.

The resolution of the acquired image depends on the spatial encoding gradients G_x, G_y, and G_z. In fact, these gradients define the sampling grid in the frequency plane. A large gradient reflects a high frequency sampling rate. For example, a point in the K-space (x, y) is filled with the frequency component whose frequency is exactly (xG_x, yG_y). The higher the gradients G_x and G_y, the higher frequency the K-space is able to capture. For abdominal imaging, the slice selection gradient G_z is often set around 10 times smaller than the frequency and phase gradient. When the inverse Fourier transform is performed to get the spatial image, this discrepancy in spatial encoding gradients turns into a difference in spatial resolution of the image directly, as the spatial sampling rate has a reciprocal relationship with the frequency sampling rate. The fusion method proposed in this thesis attempts to reverse this process and to restore a high frequency sampling rate in all three directions in order to reconstruct a high resolution spatial image of the abdomen.

1.3.2 MRI Bias Field Effect

A smooth intensity variation is often observed in MR images. Such an intensity non-uniformity is called the MRI bias field effect. This bias field results from several factors [45], including:

- inhomogeneous radio-frequency (RF) excitation by the RF coils,
- non-uniform reception sensitivity,
- electrodynamic interactions with the object (which are often described as RF penetration and standing wave effects).

There are other less important effects that contribute to this non-uniformity, including:

- eddy currents driven by the switching of field gradients,
- poor tuning of the RF coil,
- bandwidth filtering of the data,
- geometric distortion.

The effects due to the latter category are usually considered negligible compared to those caused by the former. Usually the image distortion due to the less significant factors is at most a few millimetres and the corresponding changes in intensity are on the order of 1%. Moreover, modern MRI scanners are equipped with flat response RF coils and are often able to tune themselves for each patient individually, which reduces the effect of poor tuning of the RF coil and bandwidth filtering to the minimum. The influence of eddy currents is also small, but the only concern is that its effect increases with decreasing repetition time because shorter repetition time leads to faster switching of field gradients. However, eddy currents may be an issue for rapid image acquisition; and this is another factor that constrains the shortest repetition time that could be used, and hence limit the time needed for an MRI scan.

Among the first category of influencing factors, electrodynamic interactions with the object happen when the RF magnetic field is emitted and received during image acquisition. Spatial variations in both excitation field strength and reception sensitivity can produce significant variations in measured signal intensity. In addition, the relationship between the strength of the excitation field and the measured intensity is non-linear, while it is generally assumed to be linear in the following computations. These major factors generate significant distortions of the MR signals, and hence image intensities, and consequently attract the most attention.

A number of methods have been proposed to correct for the MR bias field, such as the non-uniform intensity normalisation (N3) approach by Sled et al. [45] and the more recent Nick's non-uniform intensity normalisation (N4) implementation by Tustison et al. [48]. However, most of this work focuses on brain MR images and there are few on MR bias field correction for abdominal images, which represent very different geometry and human tissue. In fact, from our initial evaluations, none of the off-the-shelf MR bias field correction algorithms seems to work well for our images. Therefore, we concluded not to go down this route; instead, we keep this problem in mind and devise a robust feature representation that is not significantly affected by the MR bias field.

1.4 Thesis Layout

This thesis consists of 7 chapters. The first chapter introduces the clinical problem and the objective of this project, then briefly describes the physics behind magnetic

resonance imaging. From Chapter 2 we start to explain in detail the theories in signal/image frequency analyses - from the simple one-dimensional Fourier transform, wavelet transform, to multi-dimensional isotropic filtering and the Riesz transform (also known as the generalised Hilbert transform). Isotropic filtering is the key to both efficient multi-dimensional feature detection and image decomposition for the subsequent fusion operations. Chapter 3 continues the background description to explain the idea of active contours via level sets, phase congruency and monogenic signal. Level-sets is an effective way to implement edge seeking active contours for automatic boundary segmentation. Phase congruency is a powerful local phase-based feature detector that (some evidence suggests) is similar to that of the human visual system. Together with monogenic signals as an isotropic way of computing local phases, Chapter 3 paves the background for our proposed segmentation method. Chapter 4 explains the use of the Riesz transform to decompose images, suggests several fusion rules for investigation, and validates the fusion with simulated data. Chapter 5 presents the problem of phase wrapping when computing local phases and proposes a method that avoids this problem via geometric algebra. In order to achieve effective filtering, Chapter 5 also investigates the properties of five well-known filters and suggests one for our application. Chapter 6 explains the practical issues that arise when the proposed methods are applied to real clinical images, and presents our solutions that address these issues, including a steerable fusion method, a solution for accurate mean phase calculation, and coupled level sets. The last chapter, Chapter 7 concludes this dissertation and provides suggestions for future research.

Chapter 2

Background for Image Fusion

Our aim is to reconstruct an approximately isotropic, fine resolution 3-D image from thick slice MRI scans that are acquired in routine clinical practice. These thick slices consist of very long, thin voxels as a result of anisotropic frequency sampling. Fortunately, these voxels are usually available in multiple directions, which often are orthogonal. This chapter establishes the groundwork for reversing the anisotropic frequency sampling and decomposition, which allows us to re-combine the appropriate frequency components and reconstruct a close approximation to the original signal. In the next section we revisit the frequency domain via the Fourier transform, where acquisition of MR images begins, then look at the ideas of the windowed Fourier transform and the wavelet transform for frequency filtering. At the end of this chapter we introduce the isotropic generalisation of an analytic signal via the Riesz transform.

2.1 The Fourier Transform

The acquisition of MR images is, in essence, a frequency sampling process. The journey of frequency sampling starts with the Fourier transform, which converts a signal in the time/spatial domain into the frequency domain. MRI signals are collected in k-space, which is a discrete frequency space of the digitised MRI image. The spatial image is obtained by performing the inverse Fourier transform of the k-space. It is important for us to recall some key properties of the Fourier transform, as this forms the basis of frequency sampling in MRI.

2.1.1 One-dimensional Fourier Transform

The Fourier transform of a one-dimensional function $f(x)$ is defined as:

$$F(u) = \mathcal{F}(f(x)) = \int_{-\infty}^{+\infty} f(x) \exp(-jux)\, dx \tag{2.1}$$

where $F(u)$ belongs to the complex frequency domain (1-D k-space). To avoid convergence issues of the integration, the Fourier transform is first defined over the space $\mathbf{L}^1(\mathbb{R})$ of integrable functions. It is then extended to the space $\mathbf{L}^2(\mathbb{R})$ of finite energy functions. If $f(x) \in \mathbf{L}^1(\mathbb{R})$, the Fourier integral can be shown to converge and

$$|F(u)| \leq \int_{-\infty}^{+\infty} |f(x)|\, dx < +\infty \tag{2.2}$$

Thus the Fourier transform is bounded, and $F(u)$ is a continuous function of u. If $F(u)$ is also integrable, it can be shown that the inverse Fourier transform is

$$f(x) = \mathcal{F}^{-1}(F(u)) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} F(u)\exp(jux)\,du \tag{2.3}$$

The most important property of the Fourier transform for signal and image processing is the convolution theorem, which is another way to express the fact that sinusoidal waves e^{jux} are eigenvalues of convolution operators. Let $f(x) \in \mathbf{L}^1(\mathbb{R})$ and $h(x) \in \mathbf{L}^1(\mathbb{R})$. The function $g(x) = h(x) * f(x)$ is in $\mathbf{L}^1(\mathbb{R})$ which in the Fourier domain can be expressed as:

$$G(u) = H(u)F(u) \tag{2.4}$$

In effect, each frequency component e^{jux} of amplitude $F(u)$ is amplified or attenuated by $H(u)$. Thus a convolution can be expressed as frequency filtering, and $H(u)$ is the transfer function of the filter.

2.1.2 Conservation of Energy

Extending the Fourier transform to the space $\mathbf{L}^2(\mathbb{R})$, also known as the Hilbert space, of functions $f(x)$ with finite energy, requires the definition of inner product. The inner product of $f(x) \in \mathbf{L}^2(\mathbb{R})$ and $h(x) \in \mathbf{L}^2(\mathbb{R})$ is

$$\langle f, h\rangle = \int_{-\infty}^{+\infty} f(x)\,h^*(x)\,dx \tag{2.5}$$

where $h^*(x)$ is the complex conjugate of $h(x)$. The resulting norm of the function $f(x)$ in $\mathbf{L}^2(\mathbb{R})$ is

$$\|f\|^2 = \langle f, f \rangle = \int_{-\infty}^{+\infty} |f(x)|^2 \, dx \tag{2.6}$$

If f and h are in $\mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$, then Parseval's theorem shows that

$$\langle f, h \rangle = \int_{-\infty}^{+\infty} f(x) \, h^*(x) \, dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(u) \, H^*(u) \, du = \frac{1}{2\pi} \langle F, H \rangle \tag{2.7}$$

This means that if f and h are continuous and finite in energy, then the Fourier transform conserves their continuity and energy up to a factor of 2π. Since in medical image processing applications signals are always continuous and finite, the condition $f \in \mathbf{L}^1(\mathbb{R}) \cap \mathbf{L}^2(\mathbb{R})$ generally holds.

2.1.3 Multi-dimensional Fourier Transform

The Fourier transform in $\mathbb{R}^n$ is a dot-product extension of the one-dimensional Fourier transform. For an n-dimensional integrable function $f(\mathbf{x}) \in \mathbf{L}^1(\mathbb{R}^n)$, its Fourier transform is:

$$F(\mathbf{u}) = \mathcal{F}(f(\mathbf{x})) = \int_{-\infty}^{+\infty} f(\mathbf{x}) \exp(-j\mathbf{u} \cdot \mathbf{x}) \, d\mathbf{x} = \left\langle f, \mathrm{e}^{j\mathbf{u}\cdot\mathbf{x}} \right\rangle \tag{2.8}$$

where $F(\mathbf{u}) \in \mathbf{L}^1(\mathbb{R}^n)$. In the case of a two-dimensional function, this is

$$F(u_1, u_2) = \int \int_{-\infty}^{+\infty} f(x_1, x_2) \exp(-j(u_1 x_1 + u_2 x_2)) \, dx_1 dx_2 \tag{2.9}$$

where $\mathbf{x} = (x_1, x_2)$, $\mathbf{u} = (u_1, u_2)$ and $\int f(\mathbf{x})\, d\mathbf{x} = \int\int f(x_1, x_2)\, dx_1 dx_2$. In polar coordinates, the exponent $\exp(-j\mathbf{u}\cdot\mathbf{x})$ can be written as

$$\exp(-j\mathbf{u}\cdot\mathbf{x}) = \exp(-j(u_1x_1 + u_2x_2)) = \exp(-j|\mathbf{u}|(x_1\cos\theta + x_2\sin\theta)) \quad (2.10)$$

with $|\mathbf{u}| = \sqrt{u_1^2 + u_2^2}$ and $\theta = \arg(\mathbf{u}) = \arctan({}^{u_2}/_{u_1})$. This is a plane wave that propagates in the direction of θ and oscillates at frequency $|\mathbf{u}|$. The value of $F(\mathbf{u})$ is the Fourier coefficient of the sinusoidal decomposition in the direction of $\arg(\mathbf{u})$ at frequency $|\mathbf{u}|$. The frequency isolines in the frequency domain are contours of the $\mathbf{L}^2$ norm of $\mathbf{u}$. This means the n-dimensional Fourier space is isotropic. A real-valued unit circle on the two-dimensional Fourier plane represents a point-symmetric cosine wave of 1 Hz in the spatial domain. This explains why frequency filtering in multi-dimensions should be isotropic as well.

2.1.4 Properties of Fourier Transform

The properties of an n-dimensional Fourier transform are essentially the same as in one dimension. The important results are summarised below.

- If $f(\mathbf{x}) \in \mathbf{L}^1(\mathbb{R}^n)$ and $F(\mathbf{u}) \in \mathbf{L}^1(\mathbb{R}^n)$, then

$$f(\mathbf{x}) = \mathcal{F}^{-1}(F(\mathbf{u})) = \frac{1}{(2\pi)^n}\int_{-\infty}^{+\infty} F(\mathbf{u})\exp(j\mathbf{u}\cdot\mathbf{x})\, d\mathbf{u} = \frac{1}{(2\pi)^n}\left\langle F, \mathrm{e}^{j\mathbf{u}\cdot\mathbf{x}}\right\rangle \quad (2.11)$$

  This describes an inverse Fourier transform in multi-dimensions.

- If $f(\mathbf{x}) \in \mathbf{L}^1(\mathbb{R}^n)$ and $h(\mathbf{x}) \in \mathbf{L}^1(\mathbb{R}^n)$, then the convolution

$$g(\mathbf{x}) = f * h(\mathbf{x}) = \int\int f(\tau)\, h(\mathbf{x} - \tau)\, d\tau \tag{2.12}$$

has a Fourier transform

$$G(\mathbf{u}) = H(\mathbf{u})\, F(\mathbf{u}) \tag{2.13}$$

This result is the same as in 1-D except that now the terms are multi-dimensional functions and the convolution operation is a n-D convolution as well.

- Parseval's theorem proves that

$$\int_{-\infty}^{+\infty} f(\mathbf{x})\, h^*(\mathbf{x})\, d\mathbf{x} = \frac{1}{(2\pi)^n} \int_{-\infty}^{+\infty} F(\mathbf{u})\, H^*(\mathbf{u})\, d\mathbf{u} \tag{2.14}$$

- If $f(\mathbf{x}) \in \mathbf{L}^2(\mathbb{R}^n)$ is separable, which means that

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n) = f_1(x_1)\, f_2(x_2) \ldots f_n(x_n), \tag{2.15}$$

then its Fourier transform is also separable:

$$F(\mathbf{u}) = F(u_1, u_2, \ldots, u_n) = F_1(u_1)\, F_2(u_2) \ldots F_n(u_n) \tag{2.16}$$

where F_i are the one-dimensional Fourier transforms of f_i. In fact, the separability of the Fourier transform underlies the ability to spatially encode MR signals in the x-, y- and z-directions separately.

- If $f_{\mathbf{R}}(\mathbf{x})$ is a rotated version of $f(\mathbf{x})$ by a rotation matrix $\mathbf{R}$ as

$$f_{\mathbf{R}}(\mathbf{x}) = f(\mathbf{R}\mathbf{x}) \tag{2.17}$$

then its Fourier transform is also rotated by $\mathbf{R}$

$$F_{\mathbf{R}}(\mathbf{u}) = F(\mathbf{R}\mathbf{u}) \tag{2.18}$$

This is the key to steerable filtering, which will be crucial for image processing when multiple images are not in perpendicular orientations to each other.

2.2 Wavelet Transform

While the Fourier transform connects the time/spatial and frequency domain, the wavelet transform introduces the important concept of scale, which allows us to decompose a signal systematically in another dimension. The idea of wavelets is described thoroughly and consolidated by Mallat [33]. This section describes "local" frequency analysis and the generalisation of it. It can be shown that multiscale analysis is done via filtering, which is important for feature detection and noise suppression.

2.2.1 Time/space-frequency Atoms

The wavelet transform is an important class of local time/space-frequency decomposition. A linear time/space-frequency transform correlates the signal with a dictionary of waveforms that are concentrated both in time/space and in frequency. Let us

consider a general dictionary of such waveforms $\mathcal{D} = \{\phi_\gamma\}_{\gamma \in \Gamma}$, where γ is an index parameter. We assume that $\phi_\gamma(\mathbf{x}) \in \mathbf{L}^2(\mathbb{R}^n)$ and that $\|\phi_\gamma\| = 1$. The corresponding linear time/space-frequency transform of $f(\mathbf{x}) \in \mathbf{L}^2(\mathbb{R}^n)$ is defined by

$$\Phi_\gamma(f(\mathbf{x})) = \int_{-\infty}^{+\infty} f(\mathbf{x}) \phi_\gamma^*(\mathbf{x}) d\mathbf{x} = \langle f, \phi_\gamma \rangle \tag{2.19}$$

In each individual dimension $\phi_\gamma(x) \in \mathbf{L}^2(\mathbb{R}^1)$ if $\phi_\gamma(x)$ is concentrated in space at a centre location τ_γ, then $\langle f, \phi_\gamma \rangle$ depends only on the values of f in the neighbourhood of τ_γ. Similarly, if $\Phi_\gamma(\mathbf{u})$ is concentrated in the frequency domain at a centre frequency ξ_γ, then $\langle f, \phi_\gamma \rangle$ reveals the properties of F in the neighbourhood of ξ_γ. The time/space-frequency resolution of ϕ_γ is represented in the time/space-frequency plane (x, u), also known as the Heisenberg plane, by a rectangle centred at $(\tau_\gamma, \xi_\gamma)$. The size of the rectangle, or Heisenberg box, depends on the time/space-frequency spread of ϕ_γ, (σ_x, σ_u). This is illustrated in Figure 2.1.

The Heisenberg uncertainty principle states that the area of the Heisenberg box is at least one-half:

$$\sigma_x \sigma_u \geq \frac{1}{2} \tag{2.20}$$

This limits the joint resolution of ϕ_γ in time/space and frequency. Since there is no function that is concentrated perfectly at a point of time/location and a frequency, the rectangles must be manipulated carefully in order to cover the whole Heisenberg plane.

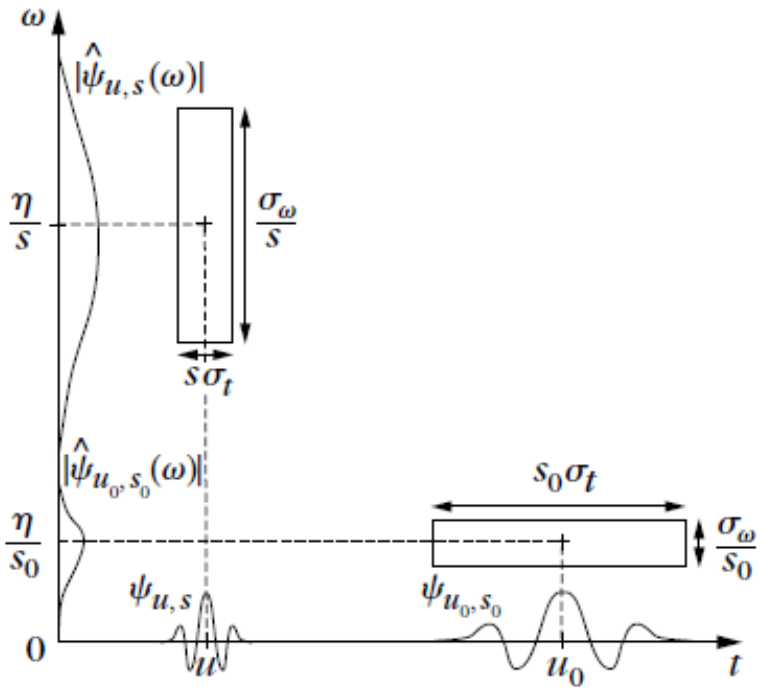

Figure 2.1: Time-frequency atom at $(\tau_\gamma, \xi_\gamma)$ can be illustrated as tile on the time-frequency plane. An increase in size of a time-frequency atom in time must accompany a decrease in size in the frequency domain, and vice versa. Image taken from [33] Fig. 4.9 on page 110.

2.2.2 Windowed Fourier Transform

The windowed Fourier transform introduced by Gabor [14] is one way to construct a time/space-frequency atom. A windowed Fourier atom is defined by a real and symmetric window $g(x) = g(-x)$ that is translated by τ and modulated by a sinusoidal with the frequency ξ:

$$g_{\tau,\xi}(x) = A\mathrm{e}^{j\xi x} g(x-\tau) \tag{2.21}$$

where A is a constant which ensures that $g_{\tau,\xi}(x)$ is normalised so that $\|g_{\tau,\xi}\| = 1$ for any (τ, ξ). The resulting windowed Fourier transform of $f \in \mathbf{L}^2(\mathbb{R}^1)$ is

$$F_{\mathrm{w}}(\tau,\xi) = \mathcal{F}_{\mathrm{w}}(f(\tau,\xi)) = \langle f, g_{\tau,\xi}\rangle = \int_{-\infty}^{+\infty} f(x)\, g(x-\tau) \exp(-j\xi x)\, dx \tag{2.22}$$

The original function f can be reconstructed by the inverse windowed Fourier transform.

$$\begin{array}{rcl} f(x) & = & \mathcal{F}_{\mathrm{w}}^{-1}(F_{\mathrm{w}}(\tau,\xi)) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \left\langle F_{\mathrm{w}}, g^*_{\tau,\xi}\right\rangle d\tau \\ & = & \frac{1}{2\pi}\int\int_{-\infty}^{+\infty} F_{\mathrm{w}}(\tau,\xi)\, g(x-\tau) \exp(j\xi x)\, d\xi d\tau \end{array} \tag{2.23}$$

A major problem with the windowed Fourier transform is that the functions $\{g_{\tau,\xi}\}_{\tau,\xi\in\mathbb{R}^2}$ are highly redundant in $\mathbf{L}^2(\mathbb{R})$, because there are many overlaps over the integration $\int_{-\infty}^{+\infty} \left\langle F_{\mathrm{w}}, g^*_{\tau,\xi}\right\rangle d\tau$. It is plausible to introduce a window scaling, which imposes a constraint between τ and ξ when modifying the time/space-frequency localisation of $g_{\tau,\xi}$.

Suppose that g has a Heisenberg time/space and frequency width, respectively, equal to σ_x and σ_u. Let $g_s(x) = s^{-1/2} g(x/s)$ be its dilation by s. Note the coefficient

$s^{-1/2}$ keeps a unit norm such that $\|g_s\| = 1$. From the definition of variance it is shown that the Heisenberg time/space and frequency width of g_s are, respectively, $s\sigma_x$ and σ_u/s. It follows that the area of the Heisenberg box is not modified, but it is dilated by s in time and compressed by s in frequency. The choice of a particular scale s depends on the desired resolution trade-off between time and frequency.

2.2.3 Wavelet Transform

Wavelets are also time/space-frequency atoms with different time/space supports that file the whole of the Heisenberg plane. The wavelet transform decomposes signals over dilated and translated wavelets. A "mother" wavelet is defined as a function $\psi \in \mathbf{L}^2(\mathbb{R})$ with a zero average:

$$\int_{-\infty}^{+\infty} \psi(x)\, dx = 0 \tag{2.24}$$

It is normalised $\|\psi\| = 1$ and centred in the neighbourhood of $x = 0$. A dictionary of time/space-frequency atoms is obtained by scaling ψ by s and translating it by τ:

$$\mathcal{D} = \left\{ \psi_{\tau,s}(x) = \frac{1}{\sqrt{s}} \psi\left(\frac{x-\tau}{s}\right) \right\}_{\tau \in \mathbb{R}, s \in \mathbb{R}^+} \tag{2.25}$$

These atoms remain normalised so that $\|\psi_{\tau,s}\| = 1$. The wavelet transform of $f \in \mathbf{L}^2(\mathbb{R})$ at time/location τ and scale s is

$$W_f(\tau, s) = \mathcal{W}(f(\tau, s)) = \langle f, \psi_{\tau,s} \rangle = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{s}} \psi^*\left(\frac{x-\tau}{s}\right) dx \tag{2.26}$$

where ψ^* is the complex conjugate of ψ. The wavelet transform integral can be rewritten in the form of a convolution product. In fact, this is the cross-correlation of f and ψ. That means the wavelet transform of scale s is equivalent to a similarity measurement of the signal and the wavelet at that scale.

$$W_f(\tau, s) = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{s}} \psi^* \left(\frac{x-\tau}{s} \right) dx = f(\tau) * \frac{1}{\sqrt{s}} \psi^* \left(\frac{-\tau}{s} \right) \tag{2.27}$$

The inverse wavelet transform was first proved by Calderon [4] under a weak admissibility condition. Grossman and Morlet [18] also proved the same formula independently. For $\psi \in \mathbf{L}^2(\mathbb{R})$, the inverse wavelet transform is

$$f(x) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_f(\tau, s) \frac{1}{\sqrt{s}} \psi \left(\frac{x-\tau}{s} \right) d\tau \frac{ds}{s^2} \tag{2.28}$$

where $C_\psi = \int_0^{+\infty} \frac{|\Psi(u)|^2}{u} du < +\infty$. In fact, the reconstruction of $f(x)$ is easier to understand in the Fourier domain. The wavelet transform is essentially a frequency band-decomposition of a signal. Since a wavelet at a particular scale s is effectively a bandpass filter $\Psi_s(u)$, the inverse wavelet transform ought to restore the spectrum of the original signal $F(u)$. With careful design of the wavelets, the wavelet transform can be a useful tool for multiscale signal/image analysis. For example, let $\mathcal{D} = \left\{ \Psi_s \mid \int_0^{+\infty} \Psi_s ds = 1 \right\}_{s \in \mathbb{R}^+}$, then we have

$$f(x) = \mathcal{F}^{-1}(F(u)) = \mathcal{F}^{-1} \left(\int_0^{+\infty} F(u) \Psi_s(u) ds \right) = \mathcal{F}^{-1} \left(\int_0^{+\infty} W_f^s(u) ds \right) \tag{2.29}$$

2.2.4 Analytic Wavelet Functions

One of the most important applications of the wavelet transform is to measure the time/space evolution of frequency transients. This requires using a complex analytic wavelet, which can separate the amplitude and phase components. Any real wavelet $\psi_{\mathrm{r}} \in \mathbf{L}^2(\mathbb{R})$ can be written in an analytic form by applying the Hilbert transform:

$$\psi_{\mathrm{a}}(x) = \psi_{\mathrm{r}}(x) + j\mathcal{H}(\psi_{\mathrm{r}}(x)) = \psi_{\mathrm{r}}(x) + jh(x) * \psi_{\mathrm{r}}(x) \tag{2.30}$$

where j is the unit imaginary defined as $j = \sqrt{-1}$ and $h(x) = \frac{1}{\pi x}$. Since $\mathcal{H}(\psi_{\mathrm{r}}(x))$ and $\psi_{\mathrm{r}}(x)$ make up a quadrature pair, if one of them is an even function then the other one is odd. By a convenient convention, we restrict ourselves to only consider analytic wavelets whose real components are even functions $\Re(\psi_{\mathrm{a}})(-x) = \Re(\psi_{\mathrm{a}})(x)$ and whose imaginary components are odd functions $\Im(\psi_{\mathrm{a}})(-x) = -\Im(\psi_{\mathrm{a}})(x)$. Then we have $\psi_{\mathrm{a}}^*(-x) = \psi_{\mathrm{a}}(x)$. Under this convention we can drop the subscript in ψ_{a} and the wavelet convolution product becomes:

$$W_f(\tau, s) = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{s}} \psi\left(\frac{\tau - x}{s}\right) dx = f(\tau) * \frac{1}{\sqrt{s}} \psi\left(\frac{\tau}{s}\right) = f(\tau) * \psi_s(\tau) \tag{2.31}$$

This shows that the wavelet transform of $f \in \mathbf{L}^2(\mathbb{R})$ at scale s is equal to the cross-correlation, indeed equal to the convolution, of f and the diluted mother wavelet ψ scaled by a factor s. The Fourier transform of $\psi_s(x)$ is

$$\Psi_s(u) = \mathcal{F}(\psi_s(x)) = \mathcal{F}\left(\frac{1}{\sqrt{s}} \psi\left(\frac{x}{s}\right)\right) = \sqrt{s}\Psi(su) \tag{2.32}$$

Since $\Psi(0) = \int_{-\infty}^{+\infty} \psi(x)\, dx = 0$ and $\Psi(u)$ has finite support in the frequency domain, it appears that Ψ is the transfer function of a band-pass filter. The convolution computes the wavelet transform with dilated band-pass filters:

$$W_f^s(x) = f(x) * \psi_s(x) = f(x) * \Re(\psi_s(x)) + jf(x) * \Im(\psi_s(x)) \tag{2.33}$$

From the Fourier convolution theorem, we can express the wavelet transform at scale s in the frequency domain as

$$\mathcal{F}\left(W_f^s(x)\right) = W_f^s(u) = F(u)\,\Psi_s(u) = F(u) * \Re(\Psi_s(u)) + jF(u) * \Im(\Psi_s(u)) \tag{2.34}$$

2.2.5 Scaling Functions

When $W_f^s(x)$ is known only for $s < s_0$, to recover f we need complementary information that corresponds to $W_f^s(x)$ for $s > s_0$. This is obtained by defining the scaling function ϕ which is the aggregation of wavelets at scales larger than s_0. Inversely, the wavelet function ψ can be obtained by the segregation of scaling functions at two different scales. Therefore, ϕ is defined by a given ψ and vice versa. The scaling function can be interpreted as the impulse response of a lowpass filter.

$$\phi_{\tau,s}(x) = \frac{1}{\sqrt{s}}\phi\left(\frac{x-\tau}{s}\right) \tag{2.35}$$

The low-frequency approximation of f at scale s is

$$L_f(\tau, s) = \langle f, \phi_{\tau,s} \rangle = f(\tau) * \phi_s(\tau) \tag{2.36}$$

Since the scaling function is defined to be the aggregation of wavelets at scales larger than s_0, or more precisely $\Phi_{s_0}(u) = \int_{s_0}^{+\infty} \Psi_s(u)\, ds$ in the Fourier domain, it can be shown that ϕ is complementary to ψ:

$$\int_0^{s_0} \Psi_s(u)\, ds + \Phi_{s_0}(u) = \int_0^{s_0} \Psi_s(u)\, ds + \int_{s_0}^{+\infty} \Psi_s(u)\, ds = 1 \tag{2.37}$$

2.2.6 Dyadic Frequency Band Selection

Like the windowed Fourier transform, the wavelet transform is a redundant representation of the signal f. In order to simplify computer implementations, the scale s is often chosen in a dyadic fashion, i.e. as powers of 2. The dictionary of wavelets becomes:

$$\mathcal{D} = \left\{ \psi_{\tau,2^k}(x) = \frac{1}{\sqrt{2^k}} \psi\left(\frac{x-\tau}{2^k}\right) \right\}_{\tau \in \mathbb{R}, k \in \mathbb{Z}} \tag{2.38}$$

The resulting dyadic wavelet transform of $f \in \mathbf{L}^2(\mathbb{R})$ is defined by

$$W_f\left(\tau, 2^k\right) = \left\langle f, \psi_{\tau,2^k} \right\rangle = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{2^k}} \psi^*\left(\frac{x-\tau}{2^k}\right) dx = f(\tau) * \psi_{2^k}(\tau) \tag{2.39}$$

In the frequency domain, this corresponds to dyadic frequency band selection. When the frequency axis is completely covered by dilated dyadic wavelets, as shown in Figure 2.2, a dyadic wavelet transform defines a complete and stable representation. The reconstruction is then given by:

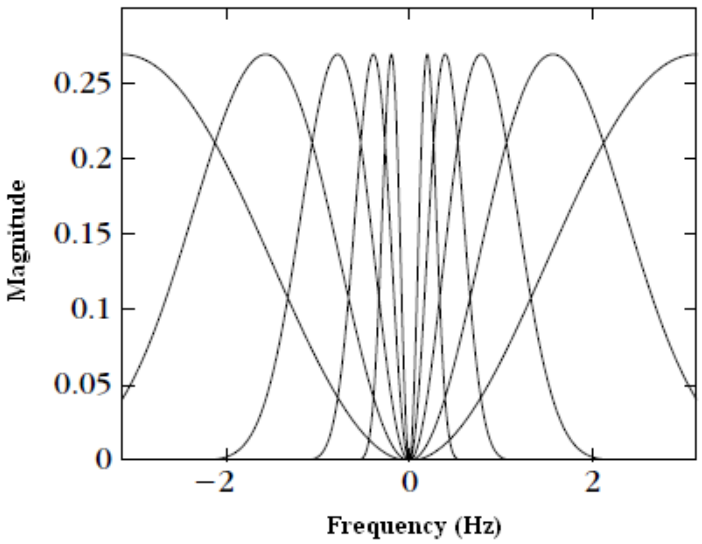

Figure 2.2: Dyadic frequency band selection. An illustration of a set of dyadic bandpass filters, where the mean frequency of the $(k-1)$-th bandpass filter is half of the k-th filter's.

$$f(x) = \sum_{k=-\infty}^{+\infty} \frac{1}{2^k} W_f\left(\tau, 2^k\right) * \psi_{2^k}(\tau) \tag{2.40}$$

This forms the basis for multiresolution approximation, which is entirely characterised by the scaling function $\Phi(u)$. It can be shown that any scaling function is specified by a discrete filter called a conjugate mirror filter [33]. The relationship between the scaling function at the k-th level and $(k-1)$-th level is

$$\Phi_k(u) = \Phi_{k-1}(2u) \tag{2.41}$$

Also, the wavelet function at the k-th scale level can be interpreted as the difference

between the k-th level and $(k-1)$-th level scaling function:

$$\Psi_k(u) = \Phi_{k-1}(u) - \Phi_k(u) = \Phi_{k-1}(u) - \Phi_{k-1}(2u) \tag{2.42}$$

Such wavelets are called orthogonal wavelets.

2.3 Early Implementations of the Wavelet Transform

Two common implementations of the wavelet transform are presented and we focus on the implementation aspects of the techniques. The discrete wavelet transform (DWT) is a discretised real valued wavelet transform, which was studied by Mallat and Daubechies in the 1980s [32, 11]. The dual-tree complex wavelet transform (DT-CWT) is a popular implementation of the complex wavelet transform (CWT) originally proposed by Kingsbury in 1998 [24, 43]. In this section, we assess their strengths and weaknesses in detail.

2.3.1 One-dimensional Discrete Wavelet Transform

The DWT was one of the earliest numerical implementations of the wavelet transform for discretely sampled signals. In addition to a wide range of families of wavelets designed for different applications, several developments have made the DWT particularly popular. The most prominent change is the recursive application of the filter bank. Since dyadic frequency band selection is used, the wavelet kernel of the k-th level ψ_k is a dilated version of ψ_{k-1} by a factor of 2. Instead of scaling the kernel

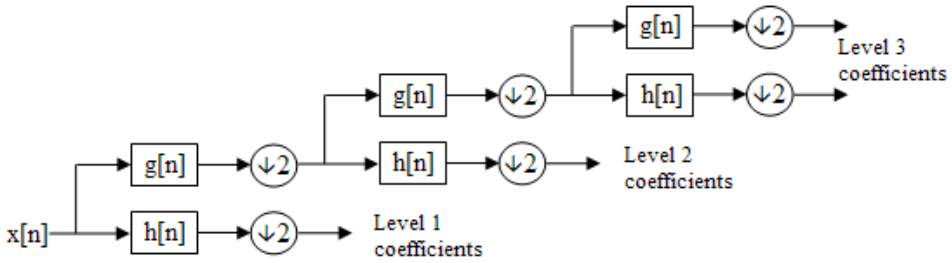

Figure 2.3: Wavelet filter bank of 3 levels. A signal function x[n] is split into a high-passed component h[n] and a low-passed component g[n], which are both subsequently down-sized by 2. The low-passed component at level 1 is then further split into a high- and low-passed component in level 2, and so on. The original signal function x[n] is decomposed in a cascade fashion. Image sourced from Wikipedia page about the "Discrete Wavelet Transform". A similar construction was originally shown in [32].

for each level, it can be achieved by downsampling the signal by a factor of 2 after filtering in each level. As a result, the filter bank need not change and can be applied recursively (Figure 2.3).

A benefit of arranging the filtering in such a cascade fashion is that it preserves the memory it requires to store the decomposed signal. Since the decomposed components are downsampled by half at each level, the total number of pixels of all the decomposed components is always the same as the number of pixels of the original image. As a result, it does not require extra memory to store the decomposition of the image despite the number of levels of decomposition. However, downsampling of the image, or decimating, also generates other problems such as shift variance and aliasing, which will be discussed further in section 2.3.3.

2.3.2 Multi-dimensional Discrete Wavelet Transform

The extension of the 1-D discrete wavelet transform to higher dimensions was studied by Mallat [32], who proposed to use the tensor product. To construct a filter bank for an n-D signal, the tensor product operation is applied to n sets of 1-D filters. As a result, the 2-D DWT consists of a filter bank of four bandpass filters, each of which is a separable scaling function ϕ and wavelet function ψ (equation 2.43). Since the scaling function is a lowpass filter and the wavelet function is highpass, the four filters for the 2-D DWT are also called the low-low (LL), low-high (LH), high-low (HL) and high-high (HH) filters, respectively.

$$W_s^{\mathrm{2D}}(x,y) = f(x,y) * \left(\begin{bmatrix} \phi^{\mathrm{1D}} \\ \psi^{\mathrm{1D}} \end{bmatrix}_x \otimes \begin{bmatrix} \phi^{\mathrm{1D}} \\ \psi^{\mathrm{1D}} \end{bmatrix}_y \right) = f(x,y) * \begin{bmatrix} \phi_x^{\mathrm{1D}}\phi_y^{\mathrm{1D}} \\ \phi_x^{\mathrm{1D}}\psi_y^{\mathrm{1D}} \\ \psi_x^{\mathrm{1D}}\phi_y^{\mathrm{1D}} \\ \psi_x^{\mathrm{1D}}\psi_y^{\mathrm{1D}} \end{bmatrix} = \begin{bmatrix} W_s^{\mathrm{LL}} \\ W_s^{\mathrm{LH}} \\ W_s^{\mathrm{HL}} \\ W_s^{\mathrm{HH}} \end{bmatrix} \tag{2.43}$$

where $\otimes$ is a tensor product, and ϕ^{1D} and ψ^{1D} is the 1-D scaling and wavelet function in the x- or y-direction, respectively. The complexity of such a tensor product construction increases dramatically as the number of dimensions increases. More precisely, the number of decomposed components of an n-dimensional DWT W_s^{nD} is equal to 2^n.

More importantly, the tensor product construction creates a chequerboard pattern in the Fourier domain for frequency selection, while the frequency isolines are

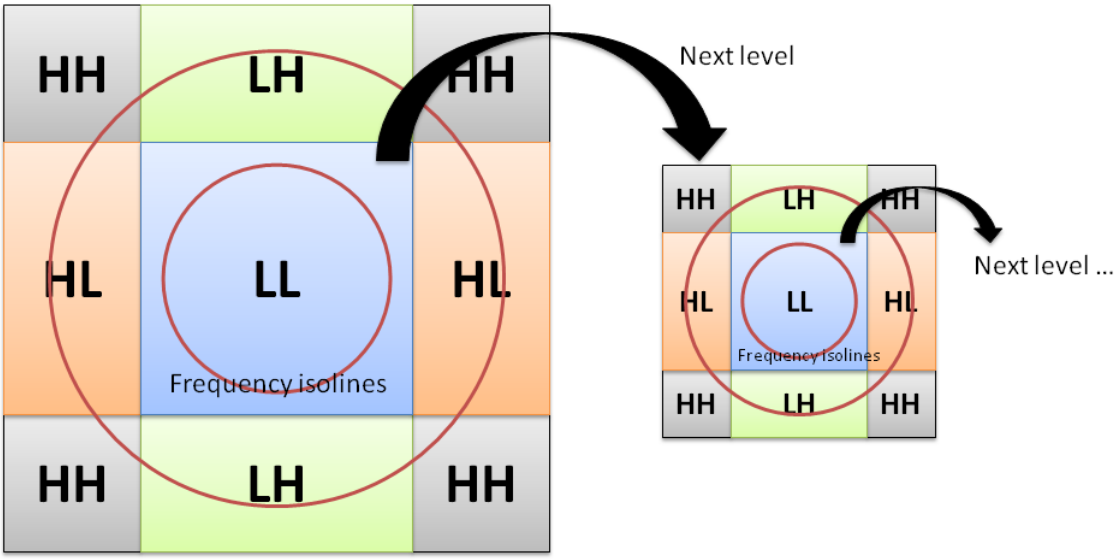

Figure 2.4: Discrete wavelet transform (DWT) frequency filtering. Left: at scale 0, the DWT decomposes an 2-D image spectrum into 4 partitions (low-low, low-high, high-low, and high-high). Right: at the next scale 1, the DWT decomposes the low-low component at scale 0 recursively into another 4 partitions (again, low-low, low-high, high-low, and high-high). The isotropic frequency isolines are shown in red circles.

concentric circles in 2-D (Figure 2.4).

Although this chequerboard cascading covers the frequency space completely, the frequency filtering is not isotropic. As a result, in a particular cascading level of the DWT the scale of the chosen wavelet function varies with orientation. This filtering pattern inevitably biases frequency selection along the x- and y-directions and amplifies noise in the diagonal direction. Moreover, it lacks the directionality to tune the filtering to be along an arbitrary orientation.

2.3.3 Dual-Tree Complex Wavelet Transform

In order to measure frequency transients, analytic complex wavelets are required, and the complex wavelet transform (CWT) is a natural extension of the traditional real

valued wavelet transform via the application of the Hilbert transform. The dual-tree complex wavelet transform (DT-CWT) proposed by Kingsbury [24, 43] is the most popular implementation of the CWT. During his development of this method, Kingsbury started with an attempt to solve the four well-known problems of the traditional DWT. They are listed in the following:

- Problem 1: Oscillations. Since wavelets are bandpass functions, the real valued wavelet coefficients tend to oscillate positively and negatively around singularities. Depending on its local phase, sometimes the wavelet coefficient can be very small or even zero at the location of a singularity. This considerably complicates wavelet-based processing, making singularity extraction and feature detection very challenging.

- Problem 2: Shift Variance. As a consequence of decimating the signal, a small shift of the signal can substantially perturb the wavelet coefficient pattern around singularities. This will particularly be a problem when a consistent representation of a feature is desired, because shift variance changes the representation of the same feature when it is shifted slightly.

- Problem 3: Aliasing. This is another consequence of the iterated downsampling operations with non-ideal lowpass and highpass filters. The wide spacing of the wavelet coefficient samples after each downsampling results in substantial aliasing. This aliasing is cancelled by the inverse DWT, but only if the wavelet and scaling coefficients are not changed. Any wavelet coefficient processing (thresholding, filtering, and quantisation) upsets the balance between the for-

ward and inverse transforms, leading to artefacts in the reconstructed signal.

- Problem 4: Lack of Directionality. Because of the chequerboard pattern produced by standard tensor product construction, the wavelets are simultaneously oriented along the x- and y-directions. This lack of directional selectivity complicates modelling and processing geometric features in an arbitrary direction.

The DT-CWT was proposed as a solution to all four of these problems. As regards oscillations, a pair of analytic complex wavelets constructed by the Hilbert transform is used. In theory, any exactly analytic wavelet must have infinite support and slow decay. Thus, there is a trade-off between analyticity and finite support. In practice, we must accept that wavelets are only approximately analytic. When a complex wavelet pair is constructed, the real and imaginary wavelet is applied to the image separately as two traditional DWTs. This leads to two parallel cascading processes. Hence it forms a so-called dual-tree (Figure 2.5). In fact, the oscillations are a problem of real valued bandpass filtering. The application of analytic filters should solve this problem. Once a complex pair of filters is used, the amplitude of the complex wavelet coefficients, or local energy, can be then computed as a feature representation. Oscillations around singularities will be avoided, and the DT-CWT is a relatively consistent representation of features.

In order to address the problem of shift variance and aliasing, the DT-CWT uses half of the pixels (odd pixels) for one tree of filtering and another half of pixels (even pixels) for the other tree of filtering, so that the two trees contain similar and complementary signal information. In addition, the DT-CWT also involves several

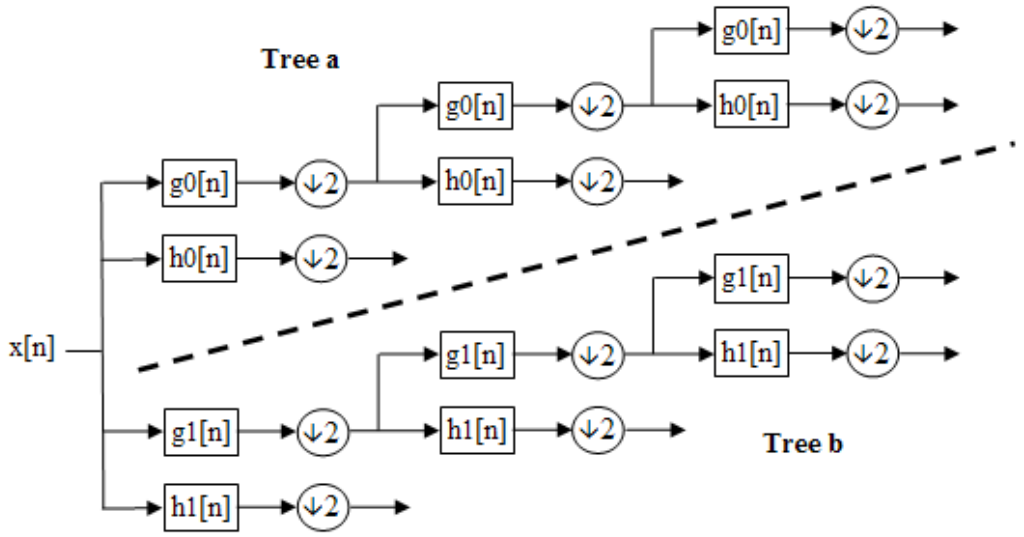

Figure 2.5: Parallel cascading processes for dual-tree complex wavelet transform (DT-CWT). A signal function x[n] is split into high- and low-passed components by the real wavelet transform as in "Tree a". In addition, a complex wavelet bank (generated by the Hilbert transform) is applied to x[n] to generate "Tree b". Image sourced from Wikipedia page on "Complex Wavelet Transform". Similar instructive images are shown in [24, 43].

implementation adjustments such as satisfying the half-sample delay condition and alternating real/imaginary filters between the two trees. Even so, the DT-CWT only achieves being approximately amplitude/phase shift invariant, approximately analytic, and free from aliasing with limitations. Since both problems of shift variance and aliasing are really the consequences of decimating the image, a simple alternative solution to these is not to downsample. Indeed, this will require more computer memory to store the decomposed image. However, for applications where accuracy outweighs computing cost this may be a price worth paying for.

For higher dimensional processing, the DT-CWT maintains the tensor product based construction. In addition, it separates the real and imaginary part of the band-pass components such as $\Re\left(W_s^{\mathrm{HH}}\right)$ and $\Im\left(W_s^{\mathrm{HH}}\right)$ in order to differentiate between

the $+45°$ and $-45°$ directions. The directional selectivity in 2-D has increased from only 3 directions for the DWT to 6 directions for the DT-CWT. However, this still does not provide the flexibility to decompose a signal in an arbitrary orientation. Moreover, as it is for the DWT, the computational complexity of DT-CWT increases exponentially as the number of dimensions increases. For an n-dimensional input signal, the number of decomposition components from the DT-CWT is 2^{n+1}.

To summarise, the DT-CWT is a fast implementation of the CWT that preserves the amount of memory required to process a 1-D signal. It indeed provides partial solutions to the four crucial problems of the DWT. However, the problem of oscillation can be solved by any complex filtering methods; shift variance and aliasing can be avoided by non-decimation with the price of requiring more processing memory; and the chequerboard filtering problem and lack of directionality is not really solved at all. Therefore, the DT-CWT is not an ideal CWT for multi-dimensional signal processing especially when accuracy is paramount.

2.4 Isotropic Filtering and Generalised Hilbert Transform

Steerability is conceptually easy to understand as a valid one-dimensional filtering is applied in a multi-dimensional space along every direction at each local point, but it is computationally very demanding and impractical to implement in this way with a great degree of accuracy. In fact, steerability is equivalent to applying the one-

dimensional filtering to the Radon transform of the multi-dimensional signal due to the Fourier slice theorem. When this 1-D filtering is orientation invariant, it is called isotropic filtering. The generalised Hilbert transform is an isotropic multi-dimensional extension of the 1-D Hilbert transform.

2.4.1 Wavelet Steerability

The DWT and CWT work well in one-dimension. However, the standard tensor product construction of multi-dimensional wavelets produces a chequerboard pattern that is simultaneously oriented along several directions and generates filtering and orientation selectivity problems. One of the earliest attempts to address this problem was by making use of the steerability of one-dimensional waves. In order to achieve this, two additional parameters are introduced: the orientation θ and the radial window spread α. Figure 2.6 illustrates the steerable filtering in the frequency plane. On the left, it shows a frequency plane tiling in which the radial window spread α is the same for any orientation θ; the wavelet of such frequency filtering is sometimes called a curvelet. On the right is the discrete version of the left due to the fact that most images acquired in practice, and hence their frequency planes, are rectangular. Each discrete curvelet in a rectangular frequency plane is defined as a trapezoidal wedges. Ideally we want to cover all orientations and let α become infinitesimally small, but this is computationally demanding and impractical. In practice, a finite number of orientations are used and the radial spread depends on the scale.

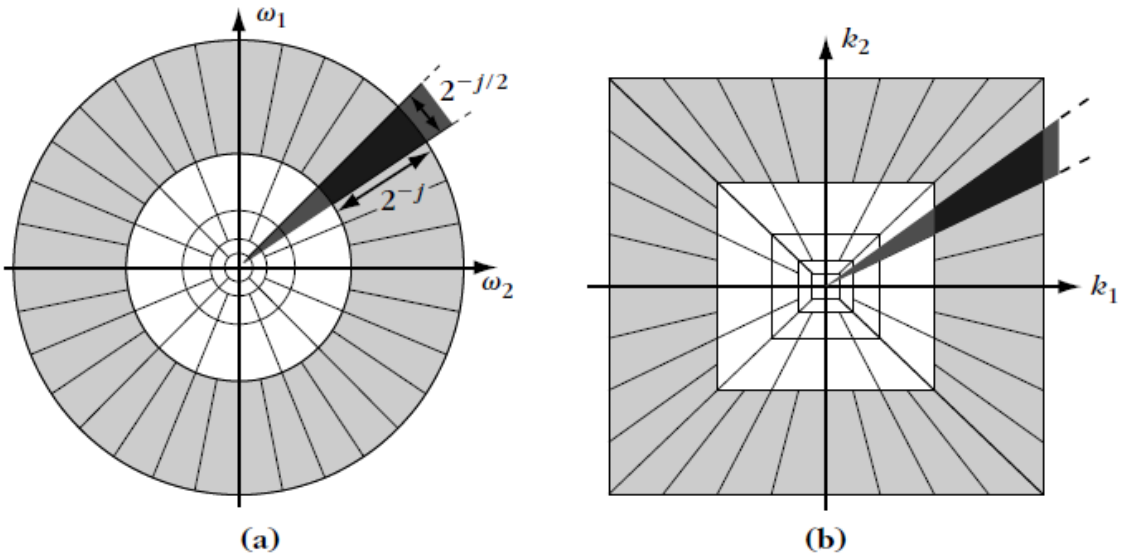

Figure 2.6: Steerability illustrated as frequency plane tiling. Left: curvelet frequency plane tiling in which the radial window spread is the same for different angles. Right: discrete curvelet frequency plane tiling in which radial and angular windows are defined in trapezoidals.

2.4.2 Isotropic Filtering

This is related to the Fourier slice theorem via the Radon transform - isotropic n-dimensional filtering in one particular direction in the frequency domain is equal to the corresponding 1-dimensional filtering of the Radon transform of the original function in that particular direction. In signal theory, the Radon transform $\mathcal{R}$ maps a 2-D signal onto an orientation parameterised family of 1-D signals by integrating the 2-D signal on the line given by the orientation parameter:

$$\mathcal{R}\left\{f\right\}(t,\theta) = \int_{\mathbb{R}^2} f(\mathbf{x})\,\delta_0\left(\langle \mathbf{x}, \mathbf{n}_\theta\rangle - t\right) d\mathbf{x} \tag{2.44}$$

where δ_0 is the Dirac delta that masks out the signal, and $\mathbf{n}_\theta$ is the normal vector pointing in the direction given by $\theta \in [0, \pi)$. The subscript 0 of the Dirac delta func-

tion δ_0 emphasizes that it masks out the signal only at the point where its inputs are equal to zero, i.e. the line along the $\mathbf{n}_\theta$ direction and shifted by t. The integration operation over the 2-D space $\int_{\mathbb{R}^2}$ on a masked signal is an alternative way to define a 1-D integration ove the line. Geometrically, the Radon transform projects (orthogonally) the 2-D signal onto a line with orientation θ. The Fourier slice theorem states that the 1-D Fourier transform of a projection of the signal is equal to the slice of the signal's 2-D Fourier transform corresponding to the projection direction.

$$\mathcal{F}_{1\mathrm{D}}\left\{\mathcal{R}\left\{f\right\}\right\}(u,\theta) = F\left(u\mathbf{n}_\theta\right) \tag{2.45}$$

As for the windowed Fourier transform, a window function $g(x)$ can be applied to the Fourier bases e^{-jux}, which can then be considered as a linear wavelet. The idea of the Radon transform relates the wavelet transform of the 1-D signal to its steerability extension for multi-dimensions. As a consequence of the slice theorem, this is equal to a slice passing through zero in the Fourier domain along the direction θ. As the steerability extension of the 1-D wavelet transform covers all possible directions in the time/spatial domain, the frequency domain is filled by radial slices. This means that a 1-D filtering along all directions in the time/spatial domain is a radial filtering, or isotropic filtering, in the frequency domain.

2.4.3 Generalised Hilbert Transform

The Hilbert transform shifts the signal by 90 degrees in the complex frequency plane and creates a quadrature representation of the signal. This is a necessary step to

generate the analytic signal representation, from which three additional quantities can be drawn - the local energy, local phase and local orientation. Each of these is discussed further in Section 3.3.3. To start with the classical 1-D Hilbert transform, it is defined as

$$\mathcal{H}\{f\}(x) = f_{\mathrm{H}}(x) = h(x) * f(x) \tag{2.46}$$

where $h(x) = \frac{1}{\pi x}$. In the frequency domain, it is

$$\mathcal{H}\{F\}(u) = F_{\mathrm{H}}(u) = H(u) \cdot F(u) \tag{2.47}$$

where $H(u) = j\frac{u}{|u|}$. Analysis of multi-dimensional signals will be done by the generalised Hilbert transform which will be abbreviated by "Hilbert transform". The Hilbert transform is also known as Riesz transform. The 2-D Riesz filter is defined in the frequency domain as

$$\mathbf{H}(\mathbf{u}) = (i,j)\frac{\mathbf{u}}{|\mathbf{u}|} = (i,j)\begin{pmatrix} H_1 \\ \\ H_2 \end{pmatrix} = i\frac{u_1}{|\mathbf{u}|} + j\frac{u_2}{|\mathbf{u}|} \tag{2.48}$$

where (i,j) are quaternion units, and $i^2 = j^2 = -1$. The 2-D Riesz transform of a real-valued signal generates two complex components, all of which are mutually in quadrature pairs.

$$\mathcal{H}\{F\}(\mathbf{u}) = \mathbf{H}(\mathbf{u})\,F(\mathbf{u}) = i\frac{u_1}{|\mathbf{u}|}F(\mathbf{u}) + j\frac{u_2}{|\mathbf{u}|}F(\mathbf{u}) \tag{2.49}$$

The Riesz transform is an isotropic extension of the 1-D Hilbert transform and

they are related by the Radon transform. Using the linearity of the Riesz, Hilbert, and Radon transforms, it can be shown that

$$\mathcal{R}\left\{f_{\mathrm{H}}\right\}(x,\theta) = (i,j)\,\mathbf{n}_\theta h_{1\mathrm{D}}(x) * \mathcal{R}\left\{f\right\}(x,\theta) \tag{2.50}$$

Also using the linearity of the Fourier transform and the slice theorem, the equivalence can be shown in the frequency domain:

$$\mathcal{F}_{1\mathrm{D}}\left\{\mathcal{R}\left\{f_{\mathrm{H}}\right\}\right\}(u,\theta) = (i,j)\,\mathbf{n}_\theta H_{1\mathrm{D}}(u)\,\mathcal{F}_{1\mathrm{D}}\left\{\mathcal{R}\left\{f\right\}\right\}(u,\theta) = (i,j)\,\mathbf{n}_\theta H_{1\mathrm{D}}(u)\,F(u\mathbf{n}_\theta) \tag{2.51}$$

The interpretation of the Riesz decomposition of intrinsically 1-D signals in multi-dimensional space is the same as the interpretation of the 1-D Hilbert decomposition in 1-D space. Moreover, for intrinsically multi-dimensional signals, it provides an interpretation by decomposing the signal into its intrinsically 1-D parts.

2.5 Conclusions

The multi-dimensional Fourier transform is a dot-product extension of the 1-D version, and as a result of the nature of the Fourier bases, an n-D Fourier transform is linear and isotropic. A wavelet is an example of a time/space-frequency atom which enables local analysis of time/space and frequency. It can be shown that the wavelet transform is in effect a convolution operation, and it inspires criteria for effective filtering such as dyadic frequency selection and compact support for filter design. This

will be useful for detecting prominent features across scales. For multi-dimensional signals, isotropic filtering is key. It can be shown that it is linked to the idea of steerability via the Radon transform but with greater accuracy and practicality. Finally, the Riesz transform, as an isotropic extension of the Hilbert transform, not only computes the imaginary part of a function $f(\mathbf{x}) \in \mathbf{L}(\mathbb{R}^n)$, but also decomposes the function into intrinsically 1-D components along any desired orientation. The first property will be the key for constructing the monogenic signal representation and the later will form the basis for decomposing and reconstructing images in the next chapter.

Chapter 3

Background on Image Segmentation

Image segmentation is an automatic process of finding the boundary of an object in an image. It usually consists of two parts. First is image based feature detection. Such feature detection often converts an image into a "feature" map, which is designed to highlight certain types of features, such as lines, ridges, or textures. The second is a boundary model which regulates the local artefacts which arise in feature detection by introducing some global constraints on the boundary. In this chapter we analyse a well-known boundary model - the active contour implemented in level sets, and introduce two local phase-based feature detection methods that are robust to intensity non-uniformity.

3.1 Level-set Segmentation Method

The level set concept involves implementing a parameterised n-D boundary (e.g. a contour in 2-D, or a surface in 3-D) in a non-parametric way such that the complexity

of the boundary representation is significantly reduced [44]. The level set approach has attracted much attention in recent years and is very well studied. The essence of a segmentation problem almost always reduces to an optimisation problem, in which the desired boundary is obtained when the optimisation criteria are satisfied. In this sub-section we investigate one of the most widely used level set models and the construction of its optimisation process.

3.1.1 Energy Minimising Active Contour Models

Active contour models (also known as "snakes") [23] are an energy-minimising spline that is guided by external constraint forces and influenced by image forces that pull it towards features such as lines and edges. As opposed to a feature detector that treats each image point independently, such as most local edge or line detectors, active contours facilitate the incorporation of global, higher-level constraints such as contour continuity and smoothness, which enables better interpretation of local events. This is because local feature detectors inevitably generate both false positives and false negatives - that is, they will miss some points that should be features and will label some points as features when they should not. The active contour model attempts to address these issues by adding smoothness and continuity. Kass et al. [23] parameterised the active contour as $C = \{\mathbf{v}(s) = (x(s), y(s))\}$, where s parameterises the curve C and goes from 0 to 1. The contour C converges, or "evolves", towards the desirable boundary while minimising its energy functional:

$$E^*_{\text{snake}} = \int_0^1 E_{\text{int}}\left(\mathbf{v}\left(s\right)\right) + E_{\text{img}}\left(\mathbf{v}\left(s\right)\right) + E_{\text{con}}\left(\mathbf{v}\left(s\right)\right) ds \tag{3.1}$$

The internal energy E_{int} consists of a first-order and a second-order term, which respectively encourage continuity and smoothness of the contour. The image term E_{img} consists of three terms; but in essence they are expressed in terms of image intensity, image gradient and image curvature. The third term E_{con} consists of additional constraint energies, for example, to provide interaction between multiple active contours.

One problem with the original active contour model is that it has to be initialised close to the target location in order to converge to the desired contour. To reduce the sensitivity of the final curve to the initial conditions, Cohen [10] suggested the use of a closed active contour with an inflating balloon force which makes the active contour behave like an edge seeking active model. Certainly, there is much variation of active contour models, some of which use global optimisation or dynamic programming and are even less sensitive to initialisation than the one described above. However, the work of this thesis focuses on feature detection, on which all contour models depend. In the next sub-section we investigate a natural extension of the active contour with balloon force and introduce the idea of level sets.

3.1.2 Level-set Based Contour Embedding

The active contour with balloon force can in practice be considered to be a closed, non-intersecting, hypersurface flowing along its gradient field with constant speed, or

with a speed that depends on the curvature. This was studied by Malladi et al. [31] using level sets as introduced by Osher and Sethian [36], which views the propagating surface C as a specific level set of a higher-dimensional function Φ, which is defined as $C(t) = \{\mathbf{x} \mid \Phi(t, \mathbf{x}) = 0\}$. Caselles et al. [5] provides a detailed discussion of the link between the active contour and the level set framework. Instead of evolving the parameterised active contour C by $C_t = F\mathbf{n}$ for a given function F and contour normals $\mathbf{n}$, we evolve the embedding function Φ according to $\Phi_t = F|\nabla\Phi|$. The biggest contribution of the level set embedding is that it converts the computational domain from $C :\to \mathbb{R}^n$ to $\Phi :\to \mathbb{R}$. Because C coincides with the set of points $\Phi = 0$, Φ is an implicit representation of the contour C. This representation is non-parametric, hence intrinsic. It is also topology-free since different topologies of the zero level-set do not imply different topologies of Φ.

3.1.3 Geodesic Active Contours

The geodesic active contours (GAC) model is a level set implementation of the simplified active contour model with the incorporation of an inflating balloon force originally proposed by Caselles et al. [5]. In the GAC model, the contour C in an image $I(\mathbf{x})$ is embedded in a higher-dimensional function Φ, which is chosen to be a signed distance function of C. The evolution equation of the GAC model is:

$$\Phi_t = \frac{\partial\Phi}{\partial t} = (F_{\text{snake}} + F_{\text{balloon}})|\nabla\Phi| = \left(\kappa \cdot \text{div}\left(g(I)\frac{\nabla\Phi}{|\nabla\Phi|}\right) + cg(I)\right)|\nabla\Phi| \quad (3.2)$$

where F_{snake} and F_{balloon} are the speed functions and are closely related to the active contours model and the balloon forces, respectively; while κ is the contour coefficient and c controls the magnitude of the balloon force. A key aspect of the GAC model is the selection of $g(I)$, the edge function, which represents "lower" level image information and controls the stopping criterion for the evolving contour. The level set model is generally independent of the choice of $g(I)$, so long as $g(I)$ is a positive strictly decreasing function and $g(I) \to 0$ as it approaches a feature. Typically, $g(I)$ is based on gradient-based edge detection methods. For example, in Caselles et al. [5], Malladi et al. [31], the authors choose:

$$g(\mathbf{x}) = \frac{1}{1 + |\nabla I^G(\mathbf{x})|} \tag{3.3}$$

where $I^G(\mathbf{x})$ is a Gaussian smoothed version of the image $I(\mathbf{x})$. $g(\mathbf{x})$ is chosen so that it approaches a minimum or zero at large intensity changes, and has a maximum value of one. More effective smoothing and edge detection methods have been studied. To this end, we present below a local phase-based method in designing $g(\mathbf{x})$.

3.2 Feature Symmetry/Asymmetry Measurement and Phase Congruency

Commonly-occurring intensity features such as step changes, lines (both ridges and valleys) and Mach bands (ramps) all give rise to points where the Fourier components of the signal are maximally in phase, or phase congruent [35]. The latter is a measure

of phase consistency through scales. The use of phase congruency is a very effective way to detect first order discontinuities such as a kink or non-smooth variation of an ideal noiseless signal. The method has significant advantages over gradient based methods. For example, it is a dimensionless quantity that is invariant to changes in image brightness or contrast. The major problem with this method is its sensitivity to noise. In this section, we present the phase congruency method proposed by Kovesi [28] with a noise compensation term. Under this framework, a later development that was specifically designed to detect symmetric features (ridges and valleys) and asymmetric features (step changes and ramps) is also presented.

3.2.1 Phase Congruency

The commonly-used gradient based edge detection methods are sensitive to many factors, including: image brightness, contrast, blurring and magnification. In particular, the gradient map $\left|\nabla I^G\right|$ of MRI is not a good feature indicator due to the MR bias field variations. Moreover, a gradient based edge detection process often needs an image specific threshold to avoid detecting noise and artefacts. Therefore, another formulation of edge detection is needed. An ideal feature measurement should be a dimensionless quantity that is invariant to change in image brightness or contrast, so that it offers an absolute measure of the significance of feature points. This allows the use of universal threshold values that can be applied over wide classes of images.

There is some evidence that the human visual system perceives a feature at points in an image where the Fourier components are maximally in phase (phase congru-

ency): Morrone and Owens [35] developed the local energy model, which leads to a dimensionless measure of phase congruency for feature detection. Venkatesh and Owens [53] later showed that energy is equal to phase congruency scaled by the sum of the Fourier amplitudes, $E(x) = PC(x)\sum_n A_n$. Therefore peaks in the energy function will correspond to peaks in phase congruency (PC), thus indicating consistent features. Instead of directly applying a Fourier transform to the image, a better approach is to use windowing (wavelets) for local analysis, in order to control the scale over which phase congruency is determined. Kovesi [28] used wavelets based on complex valued Gabor functions to decompose a 1-D signal $I(x)$ into its even (symmetric) and odd (anti-symmetric) part through scales n. That is:

$$even(x) = \sum_n even_n(x) = \sum_n I(x) * M_n^{even} \tag{3.4}$$

$$odd(x) = \sum_n odd_n(x) = \sum_n I(x) * M_n^{odd} \tag{3.5}$$

where M_n^{even} and M_n^{odd} are the pairs of even and odd Gabor filters in quadrature in scale n. Hence,

$$PC(x) = \frac{E(x)}{\varepsilon + \sum_n A_n(x)} = \frac{\sqrt{even^2(x) + odd^2(x)}}{\varepsilon + \sum_n A_n(x)} \tag{3.6}$$

where ε is a small positive constant that prevents the expression from becoming unstable when the denominator is small. However, now $A_n(x)$ is the amplitude of a signal component at a given wavelet scale, with $A_n(x) = \sqrt{even_n^2(x) + odd_n^2(x)}$.

One difficulty with phase congruency is its response to noise. Kovesi [28] addressed this problem by thresholding the local energy measurement:

$$PC\left(x\right) = \frac{\lfloor E\left(x\right) - T \rfloor^{+}}{\varepsilon + \sum_{n} A_{n}\left(x\right)} \tag{3.7}$$

where the operator $\lfloor . \rfloor^{+}$ only allows positive values, and T is an estimated upper bound of the noise responses over all scales.

3.2.2 Feature Symmetry and Asymmetry Measurement

The local energy $E\left(x\right)$ tells us that there is a feature at location x, but it is the relative value of the local energy components $even\left(x\right)$ and $odd\left(x\right)$ that tells us what kind of feature it is. Symmetry, in some sense, represents a generalisation of delta features such as lines and ridges in a 2-D image; and asymmetry represents a generalisation of step edges, or edges of a homogeneous region. At a point of local symmetry, the absolute value of the even component is large and the absolute value of the odd component is small. For this reason, Kovesi [27] proposed a measurement of local symmetry under his framework of calculating phase congruency by replacing $E\left(x\right)$ by $\left|even\left(x\right)\right| - \left|odd\left(x\right)\right|$ in the above equation. On the other hand, for the measurement of local asymmetry, it is instead replaced by the odd component minus the even component. Depending on the image modality and the type of features that we want to detect, local feature symmetry/asymmetry measurement can be used accordingly. Rajpoot et al. [38, 37] adopted the feature asymmetry (FA) measurement for delineating the 3-D left ventricle surface from noisy ultrasound images.

$$FA(\mathbf{x}) = \frac{\lfloor |odd(\mathbf{x})| - |even(\mathbf{x})| - T \rfloor^{+}}{\varepsilon + \sum_n A_n(\mathbf{x})} = \frac{\sum_n \lfloor |odd_n(\mathbf{x})| - |even_n(\mathbf{x})| - T_n \rfloor^{+}}{\varepsilon + \sum_n A_n(\mathbf{x})} \tag{3.8}$$

where $even_n(\mathbf{x})$ and $odd_n(\mathbf{x})$ are the local even and odd components of the image at scale n, respectively; $A_n(\mathbf{x})$ is the amplitude of local energy computed at the particular scale n; and T_n is a noise compensation term as defined in [28, 37, 38] as

$$T_n = \exp(\text{mean}(\log(A_n(\mathbf{x})))) \tag{3.9}$$

However, instead of using steerable 1-D filters as was suggested by Kovesi [28], Rajpoot et al. used the isotropic Riesz transform to construct multi-dimensional image components $even(\mathbf{x})$ and $odd(\mathbf{x})$. The difference is in the use of the quadrature filter pair M_n^{even} and M_n^{odd}; where Kovesi used a 1-D steerable version, Rajpoot et al. used an isotropic multi-dimensional filter pair, which is discussed in detail in the next section.

In order to enhance the feature measurement, alternatively a geometric product suggested by Rajpoot is used instead of the arithmetic sum in calculating $FA(\mathbf{x})$. Hence, only those locations where a feature appears exclusively in all scales will be highlighted. We write it as:

$$FA_{\Pi}(\mathbf{x}) = \frac{\prod_n \lfloor |odd_n(\mathbf{x})| - |even_n(\mathbf{x})| - T_n \rfloor^{+}}{\varepsilon + \prod_n A_n(\mathbf{x})} \tag{3.10}$$

An important note is that $FA(\mathbf{x})$ is a function that varies almost linearly with phase

deviation, while $FA_{\Pi}(\mathbf{x})$ varies geometrically, giving rise to a logarithmic feature scale.

The feature symmetry/asymmetry measurement proposed by Kovesi is very well defined for one-dimensional signal analysis. However, the orientation selectivity problem arises when it is applied to 2-D or higher-dimensional images. This is exactly the same problem we face when extending the wavelet transform and Hilbert transform from 1-D to n-D. Kovesi [28] originally proposed using steerability by applying the 1-D analysis $even_n(x)$ and $odd_n(x)$ in multiple orientations at each data point, but this produces a number of issues including artefacts introduced at junctions of features and high computational cost. In the following section, we review an alternative way to compute $even_n(\mathbf{x})$ and $odd_n(\mathbf{x})$ using monogenic signals.

3.3 Monogenic Signal Based Local Features

The monogenic signal is a natural extension of the 1-D analytic signal to n-dimensions. This is done via the Riesz transform, also known as the generalised Hilbert transform. The approach is derived analytically from irrotational and solenoidal vector fields, and it is connected with the 1-D analytic signal via the Radon transform. It can be applied in conjunction with phase congruency and other phase-based methods as an isotropic way to compute phase components of a multi-dimensional signal. Based on the monogenic local amplitude and local phase, there is a complete local signal representation that preserves the split of identity, one of the central properties of the 1-D analytic signal that decomposes a signal into structural and energetic information.

For a multi-dimensional signal the monogenic phase also provides additional geometric information as part of the decomposition. This section presents how the monogenic signal is constructed as well as its main interpretation of a signal via local energy, local phase and local orientation.

3.3.1 The Monogenic Signal

Local analysis of a 1-D signal can be done using analytic wavelets, which decompose a signal in two aspects - locality and analyticity. The locality of the analysis is determined by scale selection of the wavelet. At small scales, the analysis is highly localised, while at large scales the analysis is over a wide range and has a smoothing effect. The scale selection in the time/spatial domain is equivalent to frequency band selection in the Fourier domain, which has been described in Section 2.2 on the wavelet transform. The analyticity analysis constructs an analytic expression of the signal $I(x)$ via the Hilbert transform $I_{\mathrm{a}}(x) = I(x) + j\mathcal{H}(I(x))$, which enables the interpretation of phase, energy and orientation (for higher-dimensional signals). The generalised Hilbert transform (or Riesz transform) is an isotropic extension of the 1-D Hilbert transform. Combining these two ideas, the analytic expression of a signal $I(\mathbf{x}) \in \mathbf{L}(\mathbb{R}^n)$ at a certain scale s is:

$$I_{\mathrm{a}}^{s}(\mathbf{x}) = I(\mathbf{x}) * \psi^{s}(\mathbf{x}) + \mathbf{j}\mathcal{H}^{n}(I(\mathbf{x}) * \psi^{s}(\mathbf{x})) \tag{3.11}$$

where $\psi^{s}(\mathbf{x})$ is a wavelet function at scale s, $\mathbf{j}$ is a vector of quaternion units $\mathbf{j} = (i, j, k)$ while $i^2 = j^2 = k^2 = ijk = -1$, and $\mathcal{H}^n$ is the n-dimensional generalised

Hilbert transform. Due to the linearity of the convolution operator, this can be rewritten as:

$$I_{\mathrm{a}}^{s}(\mathbf{x}) = I(\mathbf{x}) * \psi^{s}(\mathbf{x}) + \mathbf{j} I(\mathbf{x}) * \mathcal{H}^{n}(\psi^{s}(\mathbf{x})) \tag{3.12}$$

A Hilbert pair is a pair consisting of the function and its Hilbert transformed components. In one-dimension a Hilbert pair is two components of 90 degree phase shift. $\psi^{s}(\mathbf{x})$ and $\mathcal{H}^{n}(\psi^{s}(\mathbf{x}))$ forms a Hilbert pair in multi-dimensions, and each two individual components are mutually in quadrature. Since the convention is to define $\psi^{s}(\mathbf{x})$ as a real valued even (point symmetric) wavelet function, i.e. $\psi^{s}(\mathbf{x}) = \psi^{s}(-\mathbf{x})$, its Hilbert transform $\mathcal{H}^{n}(\psi^{s}(\mathbf{x}))$ forms a set of real valued odd (plane asymmetric) functions along the principal directions. In image processing, the convolution product of the image and an even filter is usually called the even component, while the filtering with an odd filter is called the odd component. In this notation, $I_{\mathrm{a}}^{s}(\mathbf{x})$ can be denoted as:

$$I_{\mathrm{a}}^{s}(\mathbf{x}) = even_{s}(\mathbf{x}) + \mathbf{j} odd_{s}(\mathbf{x}) \tag{3.13}$$

The connection between the even and odd component is via the Hilbert transform as $odd(\mathbf{x}) = \mathcal{H}(even(\mathbf{x}))$. This expression for the analytic signal was studied by Felsberg and Sommer [13], in which the expression is called the monogenic signal. Originally, monogenic was another, somewhat archaic term for holomorphic, while monogenic was reused to express the multi-dimensional character of the functions. In mathematics, holomorphic is just another term for analytic. The terms

analytic function and holomorphic function are often interchangeable. Unser et al. [50] transposed the concept of monogenic signal to the wavelet domain, extended it with the higher order Hilbert transform [51], and formalised the framework for designing multi-dimensional tight wavelet frames [49], which have the combined properties of steerability and self-reversibility.

The n-D Hilbert transform is defined by a vector valued odd filter: the Riesz filter. It is represented in the Fourier domain as:

$$\mathbf{H}(\mathbf{u}) = \mathbf{j}\frac{\mathbf{u}}{|\mathbf{u}|} \tag{3.14}$$

In order to demonstrate the construction of the monogenic signal, let us take the 2-D case and define $\mathbf{j} = (i, j)$, where $i^2 = j^2 = -1$, and $\mathbf{u} = (u, v)$, where u, v are the 2-D variables in the Fourier domain. We have

$$\mathbf{H}(\mathbf{u}) = \mathbf{j}\frac{\mathbf{u}}{|\mathbf{u}|} = \left(i\frac{u}{\sqrt{u^2+v^2}}, j\frac{v}{\sqrt{u^2+v^2}}\right) = \left(H_1(u,v), H_2(u,v)\right) \tag{3.15}$$

Let us also call the bandpass filter $\Psi(\mathbf{u})$ and the image $I(\mathbf{u})$, both defined in the Fourier domain. The monogenic signal of the image is a vector valued representation:

$$I_{\mathrm{a}}(\mathbf{u}) = \begin{pmatrix} I(\mathbf{u})\Psi(\mathbf{u}) \\ I(\mathbf{u})\Psi(\mathbf{u})H_1(\mathbf{u}) \\ I(\mathbf{u})\Psi(\mathbf{u})H_2(\mathbf{u}) \end{pmatrix} \tag{3.16}$$

In the spatial domain this is equivalent to a series of convolution operations.

$$I_{\mathrm{a}}(\mathbf{x}) = \begin{pmatrix} I(\mathbf{x}) * \psi(\mathbf{x}) \\ I(\mathbf{x}) * \psi(\mathbf{x}) * h_1(\mathbf{x}) \\ I(\mathbf{x}) * \psi(\mathbf{x}) * h_2(\mathbf{x}) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ i \\ j \end{pmatrix} \tag{3.17}$$

where $\mathbf{x} = (x, y)$ are the 2-D variables in the spatial domain, and $I(\mathbf{x})$, $\psi(\mathbf{x})$, $h_1(\mathbf{x})$ and $h_2(\mathbf{x})$ are the inverse Fourier transforms of $I(\mathbf{u})$, $\Psi(\mathbf{u})$, $H_1(\mathbf{u})$ and $H_2(\mathbf{u})$, respectively. Usually, $\psi * h_i$ is computed in advance. In fact, $\psi * h_i$ is the corresponding quadrature counterpart of ψ, which is odd but in vector form for dimensions higher than one. Since the Riesz filters are also in quadrature, $\psi * h_1$ and $\psi * h_2$ are both odd and form a quadrature pair, which together make up the quadrature counterpart of the even bandpass filter ψ.

3.3.2 Monogenic Feature Symmetry/Asymmetry Measurement

To calculate the even and odd parts of an image, Kovesi [28, 27] used complex one-dimensional wavelets and steerability. Now, with the monogenic signal, the feature symmetry/asymmetry measurement can be extended to n-dimensions isotropically. Rajpoot et al. [37, 38] applied this to refine Kovesi's model. Instead of applying the 1-D Hilbert transform to generate the complex bandpass filter, the isotropic vector valued odd filter - the Riesz filter is used. In the monogenic representation, a real valued even component and a vector valued odd component are obtained. The magnitudes of the even and odd components are given by the $\mathbf{L}_2$-norm:

$$|even(\mathbf{x})| = |I * \psi| = I(\mathbf{x}) * \psi(\mathbf{x}) \tag{3.18}$$

$$|odd\,(\mathbf{x})| = |I * \psi * \mathbf{h}| = \sqrt{(I\,(\mathbf{x}) * \psi\,(\mathbf{x}) * h_1\,(\mathbf{x}))^2 + (I\,(\mathbf{x}) * \psi\,(\mathbf{x}) * h_2\,(\mathbf{x}))^2} \quad (3.19)$$

These new even and odd components can then be substituted into Kovesi's feature symmetry/asymmetry measurement formula. For example, the multiscale monogenic feature asymmetry is:

$$FA\,(\mathbf{x}) = \sum_i^n \frac{\lfloor |odd_i\,(\mathbf{x})| - |even_i\,(\mathbf{x})| - T_i \rfloor^+}{\varepsilon + \sum_n A_n\,(\mathbf{x})} = \sum_i^n \frac{\lfloor |odd_i\,(\mathbf{x})| - |even_i\,(\mathbf{x})| - T_i \rfloor^+}{\varepsilon + \sum_n \sqrt{even_n^2\,(\mathbf{x}) + odd_n^2\,(\mathbf{x})}} \quad (3.20)$$

The monogenic feature asymmetry is isotropic in n-dimensions compared to the original Kovesi's formulation using steerability. However, this method only uses the magnitude, or energetic information, of the decomposition. The magnitude of the even/odd component describes the even/odd featureness at a location, i.e. how prominent there is an even/odd feature. As far as featureness is concerned, the feature symmetry/asymmetry is a measurement of how much an even/odd feature is more prominent than an odd/even feature at the same location. However, this method does not provide geometric information of the feature, such as local orientation. This information is lost during the calculation of norm. Indeed, there is a fuller representation of local features under the monogenic signal framework, which is described in the following section.

3.3.3 Local Feature Representation

A phase approach is re-introduced to interpret the monogenic signal. The phase of a complex signal is a measure of the rotation of a real signal in the complex plane. In multi-dimensional space, the rotation between two vectors is specified by the plane formed by them. Taking the polar representation of a complex number $z = x + jy$, it is uniquely defined by $(r, \varphi) = \left(\sqrt{zz^*}, \arg(z)\right)$, where z^* is the complex conjugate of z and $\arg(z)$ is given by

$$\arg(z) = \arctan 2(y, x) = \operatorname{sign}(y) \arctan\left(\frac{|y|}{x}\right) \tag{3.21}$$

with $\arctan(\cdot) \in [0, \pi)$. The factor $\operatorname{sign}(y)$ indicates the direction of rotation in 1-D. Note that $\arg(z) = \pi$ is not defined, because the negative real numbers are singular in this definition. The angle π can represent both positive and negative rotations. This will become more obvious in n-dimensions.

In n-D space, the rotation direction is represented by an n-D unit vector as the axis of rotation. A straightforward generalisation of a multi-dimensional phase is then a vector with the length corresponding to the size of rotation angle and the direction corresponding to the rotation axis. In geometry this vector is the same concept as the rotation vector. Consequently, an n-D arctangent function is defined by the rotation vector that represents the rotation of $\operatorname{real}(\mathbf{z})$ into $\mathbf{z}$ (with $\mathbf{z} \neq 0$)

$$\arg(\mathbf{z}) = \arctan^{\mathrm{nD}}(\mathbf{z}) = \frac{\mathbf{z}_d}{|\mathbf{z}_d|} \arctan\left(\frac{|\operatorname{imag}(\mathbf{z})|}{\operatorname{real}(\mathbf{z})}\right) \tag{3.22}$$

where $\mathbf{z}_d$ is a direction vector of the rotation, which can be defined by the cross product with the real part $\mathbf{z}_d = \text{real}(\mathbf{z}) \times \mathbf{z}$. Since the phase is defined as the angle between a complex number $\mathbf{z}$ and the real axis, the cross product $\mathbf{z}_d$ is a vector perpendicular to the plane formed by $\mathbf{z}$ and its real part, hence the real axis. By normalising $\mathbf{z}_d$, the vector $\frac{\mathbf{z}_d}{|\mathbf{z}_d|}$ defines the direction of the phase of $\mathbf{z}$ in n-D space. The magnitude of the phase value is given by $\arctan\left(\frac{|\text{imag}(\mathbf{z})|}{\text{real}(\mathbf{z})}\right)$. Again, there is a singularity when $\mathbf{z}$ is equal to a negative real number, i.e. $\text{real}(\mathbf{z}) < 0$ and $\text{imag}(\mathbf{z}) = \mathbf{0}$. In this case, the magnitude of the phase is well defined as $|\arg(\mathbf{z})| = \arctan(0^-) = \pi$, but the rotation axis $\mathbf{z}_d$ is arbitrary in the imaginary subspace. Therefore, any rotation vector that is pure imaginary with a length of π is a solution. This will be even clearer in geometric algebra, or Clifford algebra [8, 9, 21, 22], where $\mathbf{z}_d$ is also called a blade of grade 2. When $\mathbf{z}$ is a real number, $\mathbf{z}_d$ is a zero blade, which is the dual of a blade of grade n. A zero blade does not have a direction, and by convention it is used to represent the phase of positive real numbers. Consequently, the phase of negative real numbers has to be undefined.

Using equation (3.22), the phase of the monogenic signal, or monogenic phase, is defined by:

$$\varphi\left(I_{\text{a}}(\mathbf{x})\right) = \arg\left(I_{\text{a}}(\mathbf{x})\right) = \arctan^{\text{nD}}\left(I_{\text{a}}(\mathbf{x})\right) \tag{3.23}$$

The rotation vector field $\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \varphi_3(\mathbf{x}))^{\mathsf{T}}$ represents the rotation of the real-valued signal $I(\mathbf{x})$ into the quaternionic-valued signal $I_{\text{a}}(\mathbf{x})$. φ always lies in the quaternion space orthogonal to the real axis since $\text{real}(\mathbf{z}_d) = 0$, and hence $\varphi_1 = 0$.

Comparable with phase wrapping in 1-D, there is a wrapping of the monogenic phase vector φ. If a vector in a certain direction would exceed the amplitude π, it is replaced by the vector minus 2π times the unit vector in that direction, i.e. it points in the opposite direction.

In practice, it is common that the presentation of $\varphi(\mathbf{x})$ is split into two parts - its magnitude $|\varphi(\mathbf{x})|$ and normalised direction $\hat{\varphi}(\mathbf{x})$. The magnitude of the monogenic phase at a certain point is called the local phase, while the direction of the monogenic phase is the local tangent pointing along the direction of a 1-D feature in the n-D space. The idea of the local orientation is the unit normal perpendicular to the feature, and it can be obtained by the cross product of $\hat{\varphi}(\mathbf{x})$ with the unit vector of the real axis. In fact, this is also equal to the direction of the imaginary part of the signal $I_{\mathrm{a}}(\mathbf{x})$. Therefore, instead of computing the cross product, the local orientation of an n-D real valued signal is obtained by the normalisation of the n-D Hilbert transform of the signal.

$$\varphi(I_{\mathrm{a}}(\mathbf{x})) = \begin{pmatrix} |\varphi(\mathbf{x})| \\ \hat{\varphi}(\mathbf{x}) \times \hat{\mathbf{1}} \end{pmatrix} = \begin{pmatrix} \arctan\left(\frac{|\mathcal{H}(I(\mathbf{x}))|}{I(\mathbf{x})}\right) \\ \frac{\mathcal{H}(I(\mathbf{x}))}{|\mathcal{H}(I(\mathbf{x}))|} \end{pmatrix} \tag{3.24}$$

where $\hat{\mathbf{1}}$ is the n-dimensional unit vector of the real axis, whose real part is equal to 1 and imaginary parts are equal to 0. For example, in the case of three dimensions $\hat{\mathbf{1}}$ is defined as $\hat{\mathbf{1}} = (1, 0, 0, 0)$.

Phase provides only half the information of a monogenic signal. To complete the representation, we need the norm of the quaternions for calculating its energy. Indeed, the norm is used for defining the local energy of $I_{\mathrm{a}}(\mathbf{x})$ by

$$|I_{\mathrm{a}}(\mathbf{x})| = \sqrt{I_{\mathrm{a}}(\mathbf{x})\, I_{\mathrm{a}}^{*}(\mathbf{x})} = \sqrt{I^2(\mathbf{x}) + |\mathcal{H}(I(\mathbf{x}))|^2} \tag{3.25}$$

The energy defined in monogenic signal is isotropic, which fulfills the desired property "split of identity", i.e. the invariance-equivariance property of signal decomposition [17]. This means that the split of information is independent to each other with a narrowly bandpassed version of the signal. Now, with the definition of the monogenic phase, it has been shown that energy and phase are indeed orthogonal. The local energy includes energetic information as a measure of featureness, and the magnitude of phase tells us what type of features there is. The combination of local energy and local phase includes structural information. In contrast to the 1-D case, the monogenic phase now includes a direction, which gives us additional geometric information. The geometric information describes the feature geometry in the n-D space and is often separated from the monogenic phase as the local orientation. In summary, the complete phase-based monogenic interpretation of a real valued signal $I(\mathbf{x})$ is:

$$\text{Local Energy} = |I_{\mathrm{a}}(\mathbf{x})| = \sqrt{I^2(\mathbf{x}) + |\mathcal{H}(I(\mathbf{x}))|^2} \tag{3.26}$$

$$\text{Local Phase} = |\varphi(\mathbf{x})| = \arctan\left(\frac{|\mathcal{H}(I(\mathbf{x}))|}{I(\mathbf{x})}\right) \tag{3.27}$$

$$\text{Local Orientation} = \hat{\varphi}(\mathbf{x}) \times \hat{\mathbf{1}} = \frac{\mathcal{H}(I(\mathbf{x}))}{|\mathcal{H}(I(\mathbf{x}))|} \tag{3.28}$$

3.4 Conclusions

In this section we investigate one of the most well-known edge based level set active contour models - the geodesic active contours (GAC). In fact, this model is the basis for the segmentation model we devised for this project. The major advantage of using the GAC model is that it is mathematically elegant and has an efficient implementation. By optimising the divergence of weighted gradient of the level set function, the GAC model essentially regulates the two most basic properties of a boundary - smoothness and continuity. The simplicity and efficiency of the GAC model makes it a very compelling segmentation model. Moreover, the GAC model splits the problem of feature representation completely from the modelling of object boundary. Therefore, we introduced two phase-based ideas for feature detection, namely, phase congruency and the monogenic signal, which are quite robust to intensity non-uniformity and suitable for our application. In the next chapter, we present our first idea to fuse multiple MR image sequences into one that has better representation of the patient's anatomy.

Chapter 4

Riesz Fusion

MRI scans generate images which consist of 3-D voxels. The measurement recorded in each voxel represents the average magnetic relaxation time of the tissues inside that voxel. However, these voxels are often not isotropic, i.e. the edge lengths for each dimension are not equal. Rather, one of its dimensions is typically five to eight times larger than the other two. Each slice of an MRI image is considered to be a 2-D image with thickness. The in-plane voxel dimensions are frequently isotropic. However, the magnitude of these voxels not only represents the tissue properties in the in-plane direction, but also the through-plane direction. This is called the partial volume effect. As a result, each "2-D" MRI slice is effectively blurred in the through-plane direction even though its in-plane resolution is high. This effect is most obvious when this thick slice cuts tangentially a boundary of two tissue types.

Figure 4.1 shows the effect of the partial volume effect on a binary sphere in a 3-D image (magnitude of 1 inside the sphere and 0 outside). The left figure is isotropic. The horizontal slice is located just above the sphere's top surface and tangential to

it. Therefore, nothing is seen on this slice. The right figure shows the three slices at exactly the same location but in an image that is anisotropically sampled in the through-plane direction. A blurred gray circle is shown where there should be zero magnitude. This is because the intensity of the adjacent voxels is extended in the through-plane direction due to the partial volume effect. In contrast, the partial volume effect is less obvious on clinical data, since in a real image there is rarely such sharp intensity change on a boundary. Even though medical signals can be assumed to be smooth, the effect can still be seen. Figure 4.2 shows a thick slice of a clinical axial scan. The cross-point of the blue dash lines indicates the location of a blurred boundary due to the partial volume effect. Although the in-plane resolution of the slice is high, the boundary between the bright (bladder) and dark (muscle) structure at the cursor is blurred, which would make it difficult to delineate the exact boundary location.

There is another consequence of the partial volume effect. Because slices are thick, 3-D feature detection/segmentation algorithms are prone to a pronounced staircase effect. This is shown in the side views of the scan (Figure 4.2 top-right and bottom-right sub-windows). In combination these pose real challenges to segmenting organs and structures in 3-D. Even if 2-D segmentation is done slice by slice separately, there remains a problem of fusing those 2-D segmentations from images of different views, which inevitably involves model guessing. Any processing on a scan in one orientation alone is not sufficiently good, and its 3-D model is prone to the staircase effect (see Figure 4.3). Therefore, in this dissertation a novel method is devised to first fuse these scans in different orientations into one single isotropic resolution 3-D

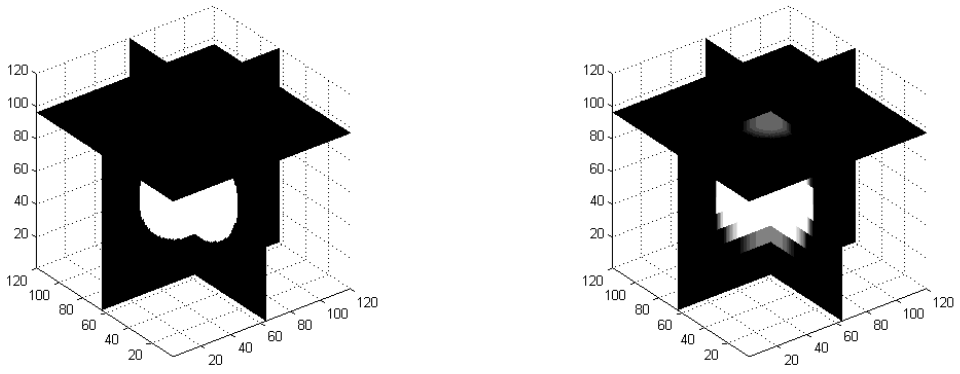

Figure 4.1: Illustration of the partial volume effect. Left: 3-D image of a sphere with isotropic image voxels. The tangential slice above the sphere (in the xy-plane) is displayed. Right: the same 3-D image is re-sampled by long-thin voxels along the z-direction. The same tangential slice now shows a gray circle due to the partial volume effect.

image. This method combines complementary information from these scans without application of any artificial modelling. Once a single isotropic 3-D image is obtained, further processing (such as image segmentation) can be applied to it directly.

In order to have a comprehensive view of the patient, radiologists usually take similar scans from different orientations. Figure 4.4 shows the human anatomy planes and their definition. The green, blue, red plane are the transverse/axial plane, coronal plane, and sagittal plane, respectively. Since these scans are often naturally orthogonal, we define a Cartesian coordinate system such that the axial slices are parallel to the xy-plane, coronal planes are parallel to the xz-plane, and sagittal slices are parallel to the yz-plane.

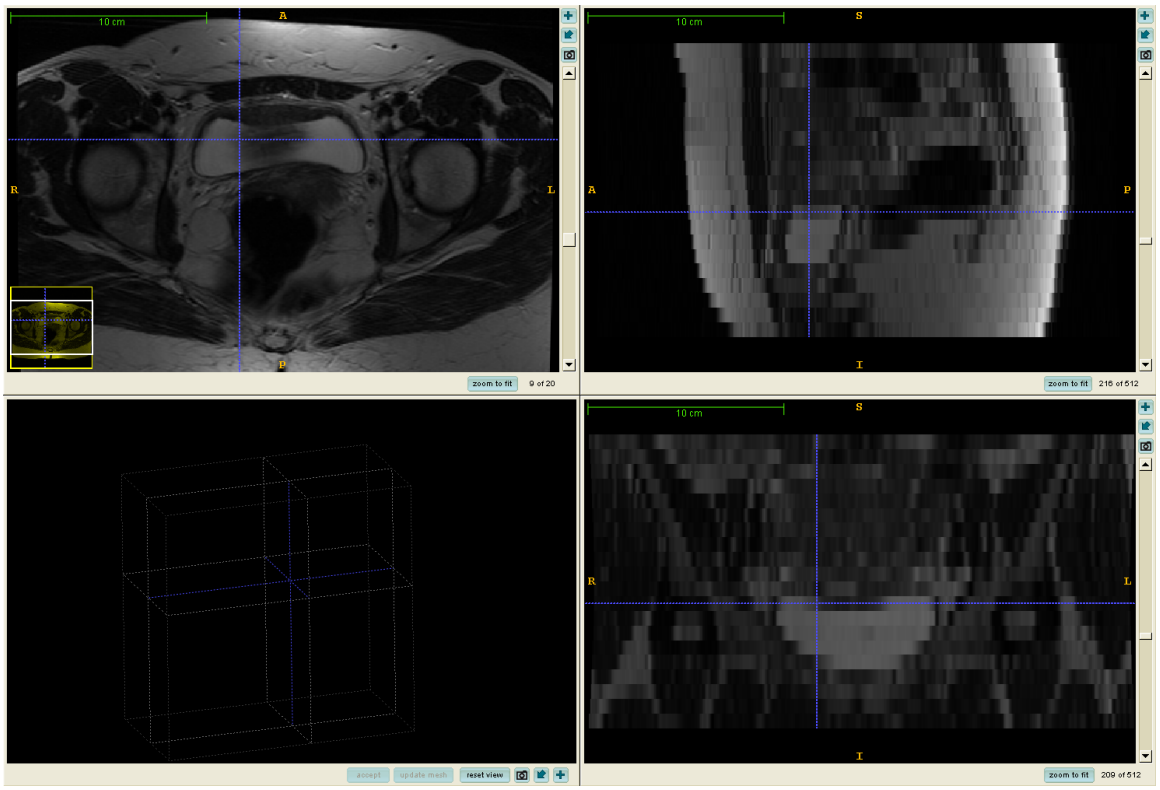

Figure 4.2: Partial volume effect in real clinical data. Top-left: the axial view of a 3-D MRI scan of the abdomen. The image slice tangential to the bladder is shown, where "shadow" is seen due to the partial volume effect. Top-right: the sagittal view of the same 3-D scan. Bottom-left: a plot showing the positions of the axial, sagittal and coronal slice images relative to the whole 3-D scan. Bottom-right: the coronal view of the 3-D MRI scan. The cursor is placed at the boundary of the "shadow" area in the tangential slice.

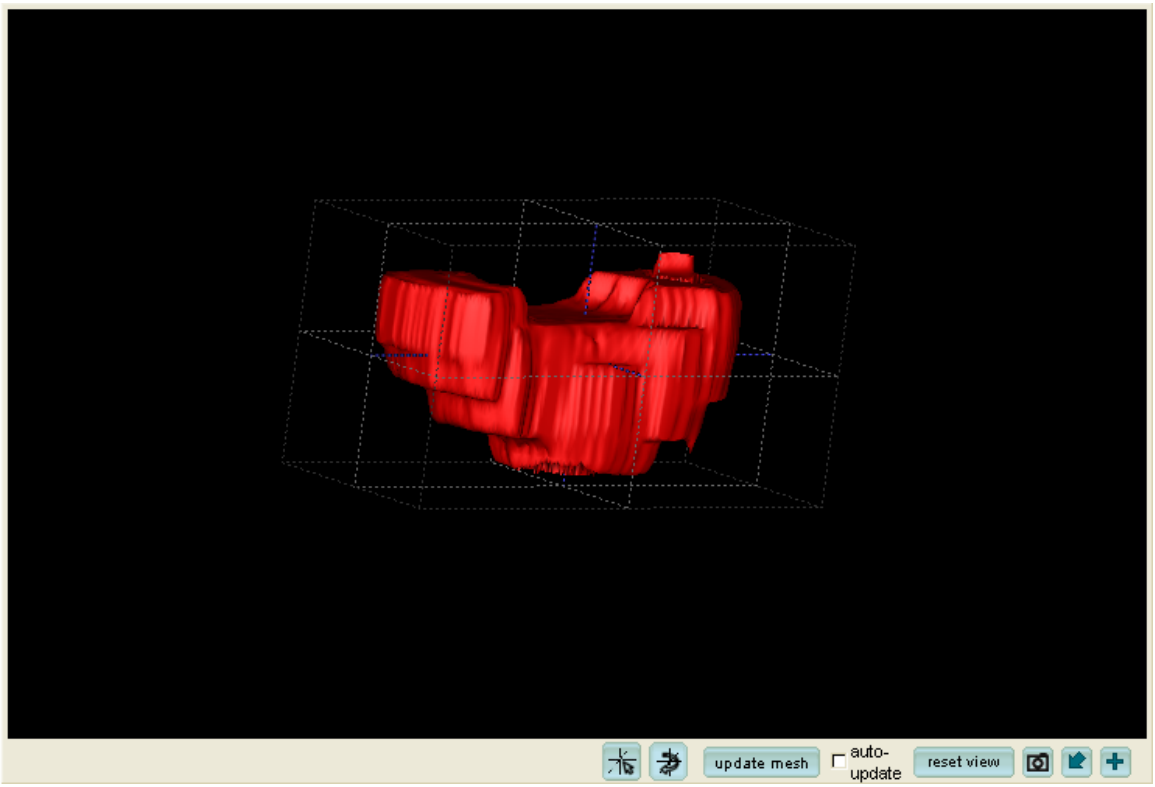

Figure 4.3: A 3-D segmentation of the bladder shown in Figure 4.2. It illustrates the "staircase effect" in the segmentation due to the thick image slices. Ideally it should have a smooth boundary which reflects the true boundary of the bladder.

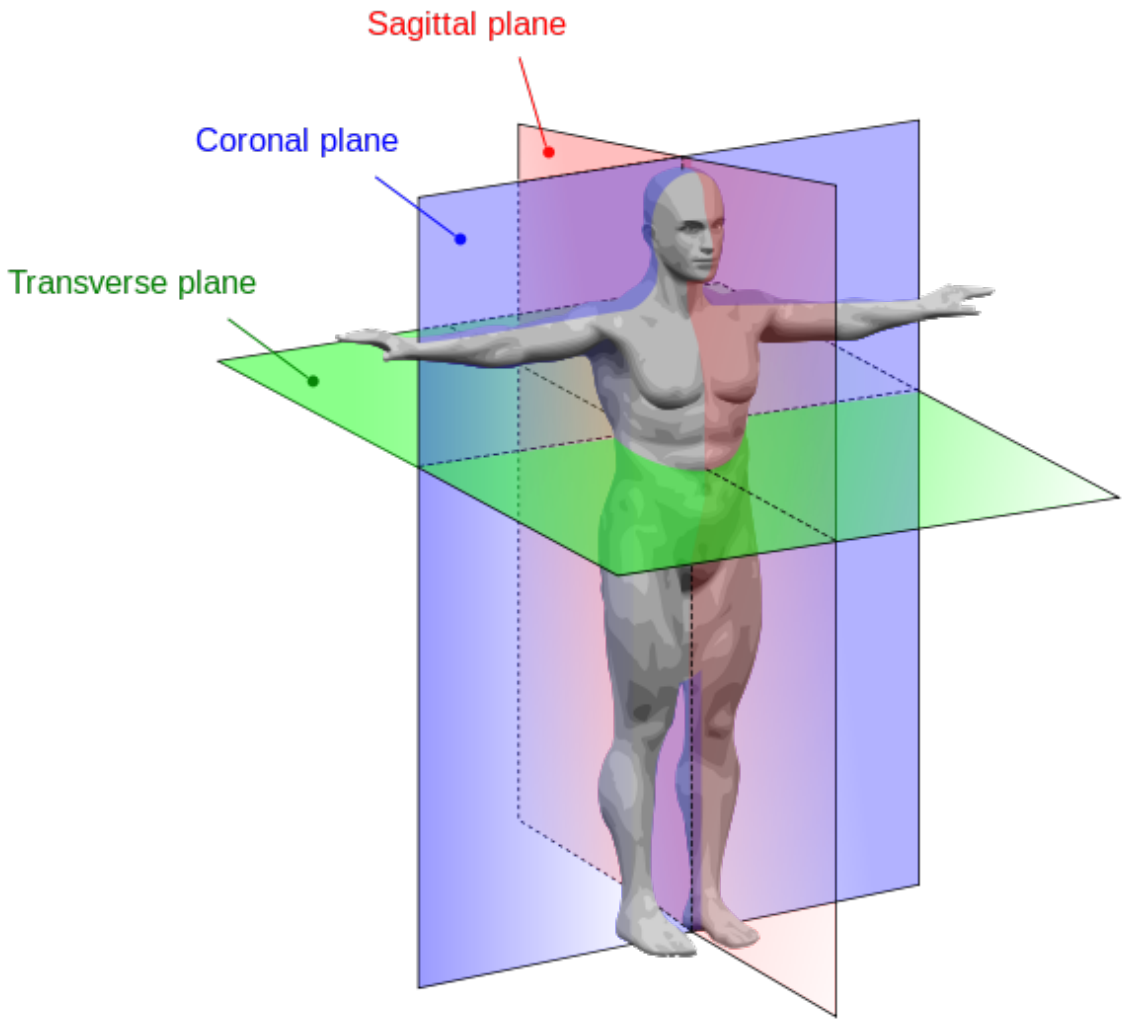

Figure 4.4: Human anatomy planes. This figure illustrates the positions of the anatomy planes relative to the patient's body in an MRI scan. Image is sourced from "http://www.my-ms.org/mri_plane_math.htm".

4.1 Riesz Decomposition

The Riesz transform decomposes an n-dimensional function along each individual dimension. It is a set of odd filters in the spatial domain with values approaching $\pm\infty$ around $0_{\pm}$, which makes it behave like a differentiator. Moreover, the Riesz filters are all-pass filters in the frequency domain and have unity weighting for every frequency component with the only exception for the zero frequency, where it suppresses the DC level of the signal to zero, which makes sense for a differentiator. Hence it can be taken as a set of all-pass differentiators along each individual dimension of the signal. In other words, it picks up features at all frequencies and preserves the signal's frequency spectrum. It is also isotropic, which means that it does not skew weightings along certain directions. It preserves the signal's frequency spectrum equally in different orientations. As a result, it is fully steerable. The Riesz transform can decompose a signal into an orthogonal set along any steerable direction. This is useful since we could apply this to decompose an image and selectively choose its frequency component. Recall that the Riesz filters are defined in the frequency domain as

$$\mathbf{H}(\mathbf{u}) = \mathbf{j}\frac{\mathbf{u}}{|\mathbf{u}|} = \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}(\mathbf{u}) = \begin{bmatrix} i\frac{u}{\sqrt{u^2+v^2+w^2}} \\ j\frac{v}{\sqrt{u^2+v^2+w^2}} \\ k\frac{w}{\sqrt{u^2+v^2+w^2}} \end{bmatrix} \tag{4.1}$$

where $\mathbf{j}$ is the quaternions $\mathbf{j} = (i, j, k)$ with basis elements i, j and k. In addition to the usual multiplication properties of vector basis defined in $\mathbb{R}^n$, i, j and k have the special property that $i^2 = j^2 = k^2 = ijk = -1$. By applying the Riesz filters to the

Fourier transform of a multi-dimensional function $f(\mathbf{x})$, we get the decomposition of the function:

$$\mathbf{H}F(\mathbf{u}) = \mathbf{j}\frac{\mathbf{u}}{|\mathbf{u}|}F = \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} F(\mathbf{u}) = \begin{bmatrix} i\frac{u}{\sqrt{u^2+v^2+w^2}}F(\mathbf{u}) \\ j\frac{v}{\sqrt{u^2+v^2+w^2}}F(\mathbf{u}) \\ k\frac{w}{\sqrt{u^2+v^2+w^2}}F(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} F_x(\mathbf{u}) \\ F_y(\mathbf{u}) \\ F_z(\mathbf{u}) \end{bmatrix} \tag{4.2}$$

Since the Riesz filters are all-pass filters except for the zero frequency component $F(0)$, i.e. the mean level of the function, it is included explicitly to ensure perfect decomposition and reconstruction:

$$\mathcal{R}(F)(\mathbf{u}) = \begin{bmatrix} F(0) \\ \mathbf{H}F(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} F(0) \\ i\frac{u}{\sqrt{u^2+v^2+w^2}}F(\mathbf{u}) \\ j\frac{v}{\sqrt{u^2+v^2+w^2}}F(\mathbf{u}) \\ k\frac{w}{\sqrt{u^2+v^2+w^2}}F(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} F(0) \\ F_x(\mathbf{u}) \\ F_y(\mathbf{u}) \\ F_z(\mathbf{u}) \end{bmatrix} \tag{4.3}$$

where $\mathcal{R}(F)$ denotes the Riesz decomposition of function $F(\mathbf{u})$. Since the Riesz transform is a uniform all-pass filter, isotropic in orientation, it has uniform weightings on all frequency components over all orientations. Along any particular orientation u_j, the decomposed component F_{u_i} is equal to zero when $i \neq j$. It is non-zero if and only if $i = j$. That is,

$$F_{u_i}\left(u_j, u_k = 0, \forall k \neq j\right) = \begin{cases} F_{u_i} & i = j \\ 0 & i \neq j \end{cases} \tag{4.4}$$

This means that the frequency components of $F(\mathbf{u})$ along an orientation u_i are captured entirely by the Riesz component F_{u_i} alone. This is important as it allows us to decompose a function onto a set of orthogonal basis elements u_i for all i. This property is called the split of identity. In the instance for 3-D, $u_i = [u, v, w]$.

4.2 Riesz Reconstruction

In order to reconstruct the signal from its individual Riesz components, an inverse Riesz transform is needed. It turns out that the inverse Riesz transform is defined by the conjugate of the Riesz filter set. That is,

$$\mathcal{R}^{-1}\left(\mathbf{F}_{\mathcal{R}}\right)(\mathbf{u}) = -\mathbf{H} \cdot \mathbf{F}_{\mathcal{R}}(\mathbf{u}) = -\mathbf{j}\frac{\mathbf{u}}{|\mathbf{u}|} \cdot \mathbf{F}_{\mathcal{R}}(\mathbf{u}) \equiv -\begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} \cdot \mathbf{F}_{\mathcal{R}}(\mathbf{u}) \tag{4.5}$$

The product of the Riesz filter and its conjugate collapses to one. Hence it inverses the Riesz decomposition,

$$-\mathbf{H}\cdot\mathbf{H} = -\begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} \cdot \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix} = -\left(i^2 \frac{u^2}{u^2+v^2+w^2} + j^2 \frac{v^2}{u^2+v^2+w^2} + k^2 \frac{w^2}{u^2+v^2+w^2} \right) = 1 \tag{4.6}$$

Including the zero frequency component for perfect reconstruction, the complete Riesz reconstruction becomes:

$$\mathcal{R}^{-1}\left(\mathbf{F}_{\mathcal{R}}\right)(\mathbf{u}) = \begin{bmatrix} F(0) \\ -\mathbf{H}\cdot\mathbf{F}_{\mathcal{R}}(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} F(0) \\ -H_x F_x(\mathbf{u}) \\ -H_y F_y(\mathbf{u}) \\ -H_z F_z(\mathbf{u}) \end{bmatrix} \tag{4.7}$$

4.3 Fusion Rules

Suppose we want to fuse two input images I_A and I_B. We propose a fusion scheme that is illustrated in Figure 4.5. First of all, the Riesz decomposition is applied to each input image. Then some fusion rules are applied to combine the Riesz components from the different images. Finally, Riesz reconstruction is used to reconstruct the combined image from the fused Riesz components.

For this fusion method to work, components with low frequency sampling rate are discarded and components with high frequency sampling rate are selected and re-used. For example, the axial images are in the xy-plane. They have a high sampling rate in the x- and y-directions, where anatomical details are contained; while the

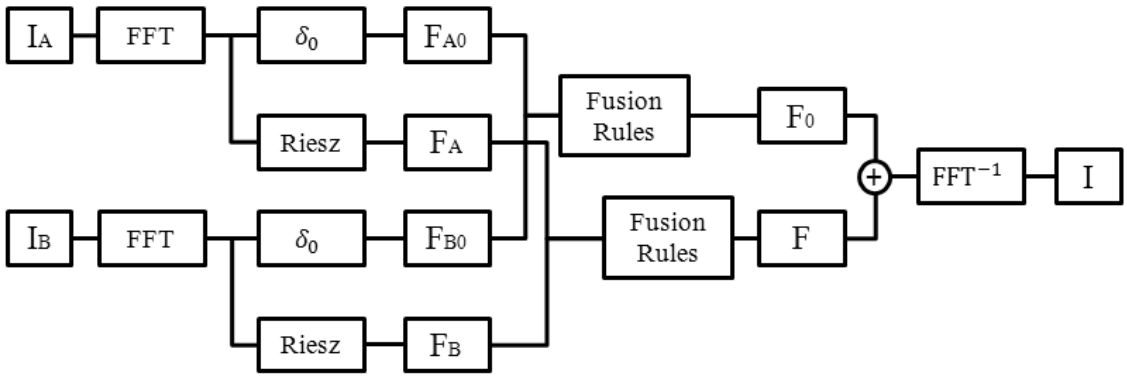

Figure 4.5: Riesz fusion scheme to fuse two images I_A and I_B. F_A and F_B are the Riesz transforms of I_A and I_B, respectively; and they are fused into F subject to some fusion rules. Similarly, F_{A0} and F_{B0} are equal to the zero frequency component of I_A and I_B and subsequently fused into F_0 by some fusion rules. The resulting fused image I is obtained by performing the inverse Fourier transform of $F + F_0$.

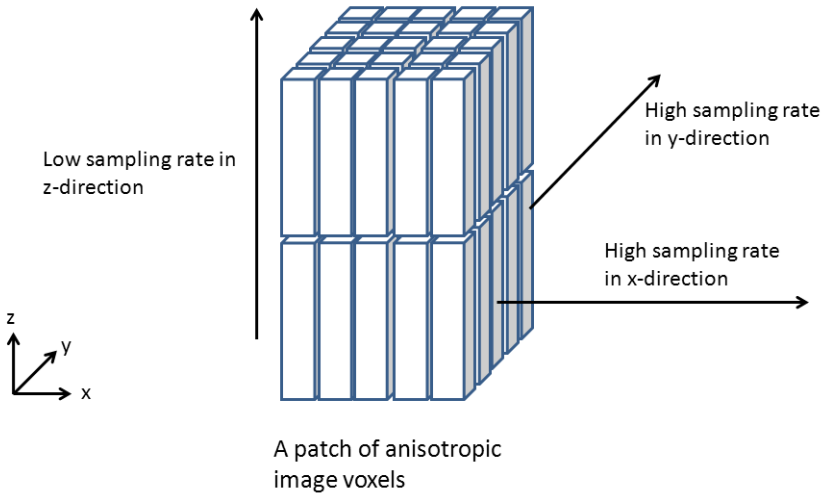

Figure 4.6: A schematic diagram that illustrates anisotropic image voxels in an MRI scan in relation with anisotropic sampling rates. Suppose an MRI scan is set up such that it has high frequency sampling rates in the x- and y-direction and low frequency sampling rate in the z-direction. The 3-D voxels of the resulting image would have short edges in the x- and y-dimension (high resolution) and long edge in the z-dimension (low resolution).

through-plane, or z-, direction has a low sampling rate and contains little detail (Figure 4.6). Therefore, the z-Riesz-component of the axial image is discarded and its x- and y-Riesz-component are selected and preserved. Similarly, the x-component of the sagittal image and y-component of the coronal image are discarded.

Further rules apply when there is more than one component selected for a particular direction. These rules are some functions that take input of multiple Riesz components and output a fused single component. We have devised and analysed three rules: minimum, maximum, and mean. As their names suggest, these rules compute the "minimum", "maximum", and "mean" of the input components as a fusion strategy, respectively. Since the function F is in the Fourier domain, F is in general complex. When the minimum/maximum rule is applied, it compares the magnitude of the real and imaginary part of F_A and F_B, individually.

$$\text{Minimum rule: } \begin{cases} \Re\left(F\right)\left(\mathbf{u}\right) = \text{MIN}\left(\Re\left(F_A\right), \Re\left(F_B\right)\right) \\ \Im\left(F\right)\left(\mathbf{u}\right) = \text{MIN}\left(\Im\left(F_A\right), \Im\left(F_B\right)\right) \end{cases} \tag{4.8}$$

$$\text{Maximum rule: } \begin{cases} \Re\left(F\right)\left(\mathbf{u}\right) = \text{MAX}\left(\Re\left(F_A\right), \Re\left(F_B\right)\right) \\ \Im\left(F\right)\left(\mathbf{u}\right) = \text{MAX}\left(\Im\left(F_A\right), \Im\left(F_B\right)\right) \end{cases} \tag{4.9}$$

The mean/average value of two complex numbers is computed for the real and imaginary part separately, i.e. $\text{MEAN}[(a+jb),(c+jd)] = \frac{a+c}{2} + j\frac{b+d}{2}$ for real numbers a, b, c and d. The formula for mean rule fusion can be written as

$$\text{Mean rule: } F = \text{MEAN}\left(F_A, F_B\right)\left(\mathbf{u}\right) = \frac{F_A\left(\mathbf{u}\right) + F_B\left(\mathbf{u}\right)}{2} \tag{4.10}$$

Similar fusion rules and fusion schemes have been proposed before by Rajpoot et al. [39] for fusing multi-view echocardiography images. The main purpose of multi-view echocardiography image fusion is to get the best image contrast from different views, which is not orientation dependent; while our purpose of fusing multi-view MR images is to resolve the problem of anisotropic frequency sampling in different views, which are orientation dependent. Therefore, selection rules are used in addition to fusion rules for our application. Moreover, the image decomposition method Rajpoot et al. used was the traditional multi-dimensional wavelet transform, which decomposes an n-D image into 2^n components; instead, we propose to use Riesz transform, which decomposes an image into n components where n is the number of dimensions of the image.

Note that computing the phase of the arithmetic mean of the vectors does not imply computing the average phase (Figure 4.7). Suppose $\mathbf{A} = a + jb$ and $\mathbf{B} = c + jd$ are two complex numbers and their arithmetic mean $\mathbf{C} = \frac{a+c}{2} + j\frac{b+d}{2}$ is also a complex number. The phase of $\mathbf{C}$ is not equal to the average phase of $\mathbf{A}$ and $\mathbf{B}$. A way to compute the third vector whose phase is actually equal to the mean phase is discussed in Section 5.4. Nonetheless, assuming the frequency spectra of the input images are smooth, which is the case for most medical images, the mean rule also smoothes the phase over the image. With less sharp phase changes, the reconstructed image would become smoother and less noisy.

Why do this in the frequency domain? All the processing suggested above - the Riesz decomposition, fusion rules and Riesz reconstruction are carried out entirely in the Fourier domain. There is a reason we want to do this. Due to the split

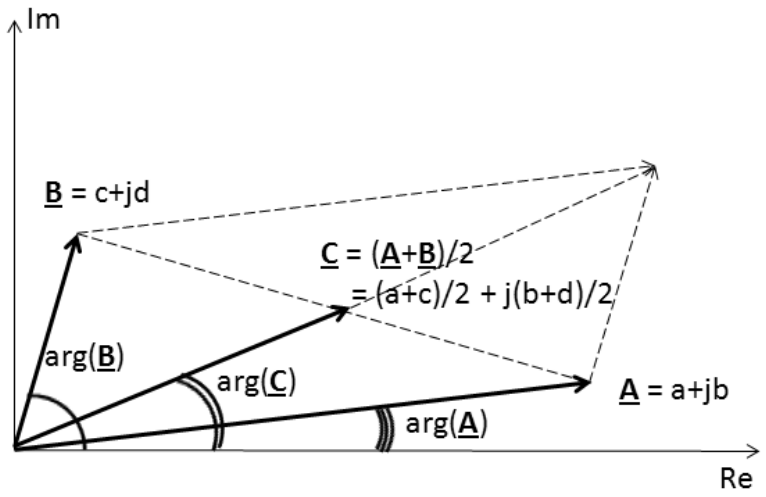

Figure 4.7: Arithmetic mean of two complex numbers $\mathbf{A} = a+jb$ and $\mathbf{B} = c+jd$. The resulting mean is given by complex number $\mathbf{C} = \frac{a+c}{2} + j\frac{b+d}{2}$. Note that the argument of $\mathbf{C}$ is not equal to the mean of arguments of $\mathbf{A}$ and $\mathbf{B}$.

of identity property of the Riesz transform, we may also split the input signal into components in orthogonal orientations and preserve the features of the original signal. Fundamentally we want to leverage the Riesz component from the orientation in which the input image has a high sampling rate. By selecting the Riesz components that have high sampling rate, which capture details of the image, it allows us to reconstruct those details from different scans in the final fused image. Combining the frequency components is the way to achieve this.

In contrast, if we apply the minimum/maximum fusion rules in the spatial domain, what it would do is to take the minimum/maximum brightness of the Riesz component of two images. For example, a voxel that is close to a boundary may suffer severely from the staircase effect in one image and less so in another. Hence this voxel has high brightness in one image due to the staircase effect but low brightness in the other. Our goal should be to suppress the brightness at this voxel in the fused image,

not to enhance. However, it is easy to see that the maximum rule in the spatial domain would keep the high value for this voxel and hence amplify the staircase effect. Similarly the minimum rule in the spatial domain would keep the low value at the voxel where brightness is artificially low due to the staircase effect, and hence fail to restore the desired boundary. Therefore, the minimum/maximum rule should be applied in the frequency domain, which enhances the frequency response in the fused image but not necessarily the staircase effect. For the mean rule, it does not matter whether it is applied in the spatial or frequency domain, because taking the mean is a linear operation and adding two signals in the Fourier domain is the same as adding them in the spatial domain: if

$$h(x) = af(x) + bg(x) \tag{4.11}$$

then

$$H(\omega) = aF(\omega) + bG(\omega) \tag{4.12}$$

for any complex numbers a and b. In the case of the mean rule, $a = b = 1/2$. To keep things simple, we intend to apply all rules in the frequency domain.

4.4 Implementation

The input images are not perfect. In practice, image pre-processing is necessary on the raw scans before fusing them into one isotropic 3-D image. Very often, the different

scans are not only in different orientations but also have different voxel resolutions (different in-plane resolution and different thickness of slice). Most importantly, they are not perfectly aligned, even though the MRI machine was calibrated and the scans were carried out for the same patient in one single sequence. For this reason, image registration is expected to be an essential pre-processing step.

The raw data used in this dissertation are in DICOM format, and were acquired on different machines with 1.5 or 3.0 Tesla magnetic field strength. Typically, the in-plane resolution of the image is about 0.5 mm per pixel, and the thickness of slices varies from 5 to 6 mm. In many cases, there are gaps between the slices which are about 1.5 to 2 mm wide. One of the limitations of our reconstruction method is that there are locations which fall in the gaps in all scans. Information at these locations are inevitably lost permanently. We will not be able to recover structures at those common-gap regions because information were never acquired in the first place. Instead of making a blind guess of what may go into those regions, the first adjustment we make is to extend the intensities of known adjacent regions into those gaps. In effect, we assume that the MRI slices are congruent.

After the adjustment for gaps, image registration is needed to correct for the mis-alignment between scans. Since each set of scans was done in a single sequence while the patient remained in a lying position, a rigid registration is used to re-align the global frame of different scans. Because the machine was calibrated before every patient's sequence, shearing of the image is not expected so affine registration is not needed. The second limitation is that the patient naturally breathed and her bladder was filling up during the time span of the scans, which typically lasted for one and a

half hours. Ideally, a non-rigid registration is also required to correct for local tissue movements. However, as different scans have different voxel dimensions, an off-the-shelf non-rigid registration algorithm often tries to warp the thick slices in one plane to fit the thick slices in another plane. As a result, a lot of undesired warping and foldings are generated. Rigid registration is the method of choice to correct the global misalignment of different scans. The method had successfully registered scans in all the 15 patient datasets we held. The choice of energy optimisation is normalised mutual information because it is unaffected by the MRI bias field effect; and the optimisation method is a standard downhill descent scheme. The implementation of the rigid registration algorithm is an off-the-shelf package called the Image Registration Toolkit (IRTK) developed by Rueckert and Schnabel [1].

After the scans are aligned, the common field of view is calculated since fusion can only be conducted in the region where information from at least two different scans are present. This process also significantly reduces memory needed for the calculations and increases computational speed.

The last pre-processing step is to interpolate the voxel intensity on a common cubic grid. The motivation for this step is two-fold. First, to prepare an isotropic grid for the fused image. Second, to smooth the structural changes between adjacent slices. Artificial sharp changes in intensity, e.g. the staircase effect between slices, increases the magnitude of high frequency components, which make the fused image noisier. By smoothing inter-slice intensity changes, the noise level and ripple effect are reduced in the reconstructed image. A linear interpolation method is used in this

[1] www.doc.ic.ac.uk/~dr/software/index.html

step.

4.5 Experiments and Validation

In order to validate the proposed fusion method, we ideally need a set of perfectly aligned images. When we apply rigid registration to the images for this project, we cannot fully ensure that the algorithm would perfectly align the images. Moreover, there may remain non-rigid deformations to correct for. An isotropic full body MRI scan is not available in our study. Therefore, we have adopted an alternative approach - to borrow isotropic 3-D images of a different anatomy. For the sole purpose of validating our fusion method, we downloaded a 3-D isotropic resolution (1 mm) T_1-weighted MRI brain image from the "BrainWeb", an online database of simulated MRI images[2]. It is an image of a normal brain (no anatomical abnormalities) with 3% noise level (calculated relative to the brightest tissue) and 20% intensity non-uniformity (mimicking the MR bias field effect). With this isotropic image, we simulated thick slices of the same image with long thin voxels by averaging voxel intensities along the x-, y- and z-axis. The original dimension of the brain image was 216 by 180 by 180 mm. The simulated slice thickness was 12 mm, while the in-plane resolution was 2 mm per pixel; i.e. the ratio of resulting voxel dimension of the simulated image was 1:1:6. Consequently, the simulated axial, sagittal, coronal sequence have 18, 15 and 15 slices, respectively. Finally, the chosen metric of similarity measure between the reconstructed image and the original image (ground truth) is the sum of squared

[2]http://www.bic.mni.mcgill.ca/brainweb/

differences (SSD), which is defined as:

$$SSD = \sum_{\text{all } x} \left(I_{\text{Riesz}}[x] - I_{\text{true}}[x] \right)^2 \tag{4.13}$$

where I_{Riesz} is the reconstructed image using the simulated thick slices and I_{true} is the original downloaded image with isotropic voxels.

4.5.1 Adaptive Fusion Rules

We first assessed the three fusion rules: minimum, maximum and mean. Recall our fusion strategy in Figure 4.5. Suppose the Riesz transform decomposes the true image into $\mathbf{F} = (F_x, F_y, F_z)$, which are complex Fourier components represented in vector form. Given the axial (in xy-plane) and coronal (in xz-plane) image to be fused, we have their Riesz decompositions as $\mathbf{F}^{\text{ax}} = \left(F_x^{\text{ax}}, F_y^{\text{ax}}, F_z^{\text{ax}}\right)$ and $\mathbf{F}^{\text{cor}} = \left(F_x^{\text{cor}}, F_y^{\text{cor}}, F_z^{\text{cor}}\right)$, respectively. Ultimately we want to reconstruct F_x from F_x^{ax} and F_x^{cor}. By defining fusion rules γ, we have $F_x^{'} = \gamma\left(F_x^{\text{ax}}, F_x^{\text{cor}}\right)$, where $F_x^{'}$ is the reconstructed x-Riesz component of the original image. In order to measure the reconstruction quality in the Fourier domain, we compute the sum of L_2 norm of the difference vectors.

$$SL2 = \sum_{\mathbf{u}} \left\| F_x(\mathbf{u}) - F_x^{'}(\mathbf{u}) \right\| \tag{4.14}$$

Ideally, we want to pick the fusion rule that minimises the sum of L_2 norm, or $SL2$. In addition, $SL2$ varies with frequency. We plot the $SL2$ against 20 frequency bins, ranging from zero to the maximum frequency. The $SL2$ value of each bin represents

the sum of L_2 norm between the reconstructed frequency components $F_x^{'}$ and the original true F_x within that frequency band. Figure 4.8 shows a plot of this for reconstructing the x-Riesz component. The vertical axis is the $SL2$ value of each bin of frequency band. The horizontal axis is normalised frequency of the image. At the value 1, it is the highest frequency along the shortest dimension (in this case it is x- and y-direction where there are 90 pixels). This is the highest frequency that is available in all directions. Above value 1, they are frequencies between the common highest frequency and the highest frequency of all. These frequencies are not available in all directions and they only exist because of the fact that our image is rectangular but the frequency space is spherical. The number of voxels falling in these frequency bands decreases exponentially; hence, their nominal $SL2$ values decreases exponentially as well (Figure 4.8 bottom). Moreover, the magnitude of the high frequency components are generally small compared to low frequencies, which leads to smaller sum of L_2 norms. However, low nominal $SL2$ values at high frequencies do not necessarily mean that the fusion of the high frequencies is significantly better than that of the low frequencies.

Figure 4.8 also shows that at low and mid frequency bands the Maximum rule generates the lowest $SL2$ fusion. The Minimum rule performs worst and the Mean rule gives only a modest result. The performance flips for higher frequencies (Figure 4.8 bottom), where the Mean rule outperforms the other two and the Maximum rule is the worst at some frequencies bands. This is expected because the Maximum rule can magnify noise, which is most significant at high frequencies.

Visual assessment of the images confirms this result. Figure 4.9-4.12 illustrates

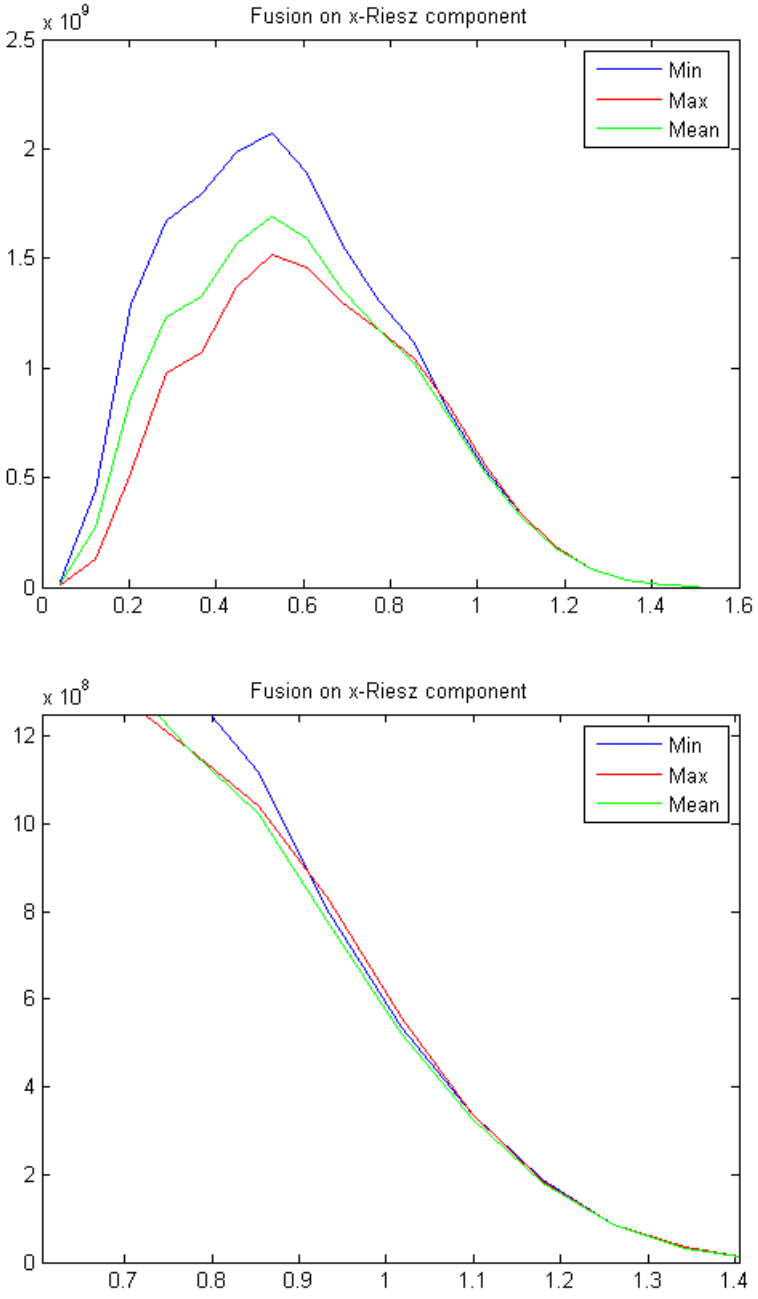

Figure 4.8: Top graph: sum of L_2 norm of the difference vector between the original image and reconstructed image by different fusion rules, plotted against frequencies. Bottom graph: a zoom-in segment of the top graph to show the high frequencies. X-axis is radial frequencies. Since the actual frequency scale (in Hz) depends on the actual image size, a normalised frequency is used here where 1 indicates the highest frequency in the shortest dimension. Y-axis is the sum of L_2 norm values.

the fusion on the simulated images. Figure 4.9 shows the isotropic ground truth image. Figure 4.10 is the simulated thick slices. All figures display the 3-D image at the same intersection. Here it illustrates that in a thick slice image the long voxels completely destroy the feature detectability from the side views (sagittal and coronal view in Figure 4.10). The partial volume effect has extended the blurring from the through-plane direction to in-plane and destroyed features in the image, despite its high in-plane resolution (Figure 4.9 and Figure 4.10 top-left). The reconstructions in Figure 4.11 and 4.12 are not perfect replicates of the original image, but they do recover most of the significant features in 3-D. In comparison to the thick axial image, where features are lost in side-views, many details are preserved in the constructions. Evidently, the Maximum rule reconstructed image (Figure 4.11) is noisier, especially in the back ground and region of homogeneous intensities; while the Mean rule reconstructed image (Figure 4.12) is smoother. However, the Maximum rule has outperformed the Mean rule overall because along with noise it also preserves sharp intensity changes better (Figure 4.11 and 4.12 yellow arrows). It was shown that the Maximum rule preserves the image integrity the best but is noisier, and the Mean rule is smooth but loses some sharp edges. Figure 4.13 and 4.14 show the difference image of the reconstructed images by the two fusion rules and the original ground truth image, respectively. It can be seen that the difference image of the Mean rule has higher values at sharp edges, such as the edge of the skull and corpus callosum.

The mean rule is smooth because effectively, it averages the frequency components and smooths phases. It smooths sharp edges with large intensity change, which contributes the most to the aggregate $SL2$ measurement. In order to support this

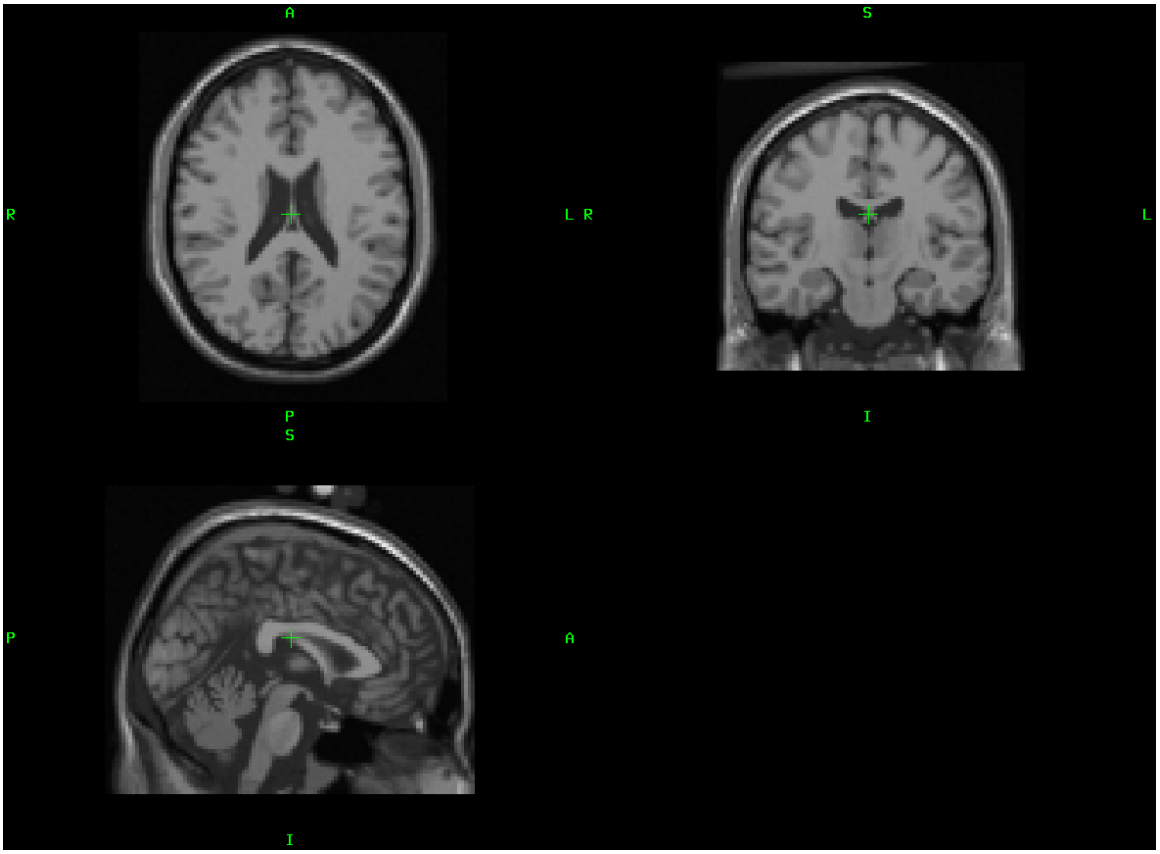

Figure 4.9: High quality isotropic T_1-weighted MR brain image used as the ground truth for validating the proposed image fusion method. Top-left: axial view. Top-right: coronal view. Bottom-left: sagittal view. The image is downloaded from the BrainWeb (http://brainweb.bic.mni.mcgill.ca/brainweb/. Configurations: 2:2:2 mm voxel dimension, 3% noise level, 20% intensity non-uniformity, healthy brain image).

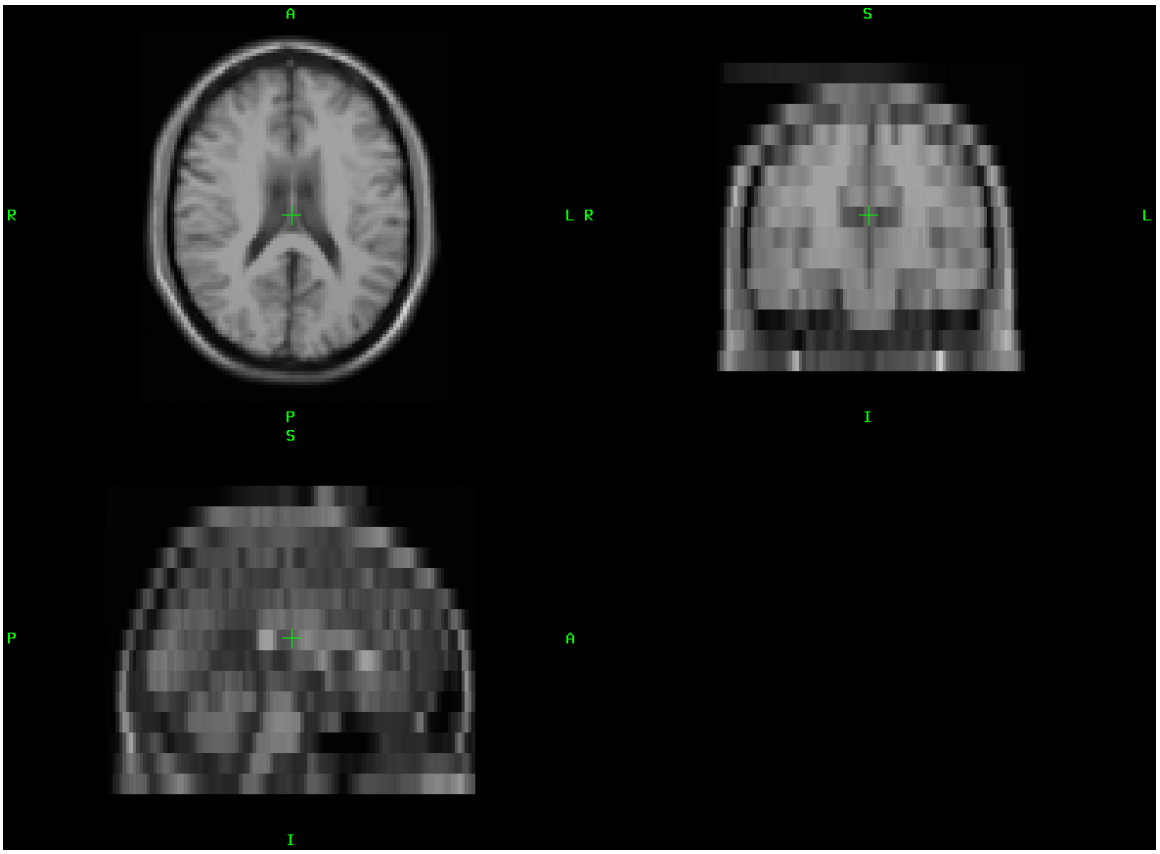

Figure 4.10: Simulated brain image of thick axial slices. Voxel dimension is 2:2:12 mm. Top-left: axial view. Top-right: coronal view. Bottom-left: sagittal view. Compared to Figure 4.9, thick slices lose most image features in the side-views - sagittal (bottom-left) and coronal (top-right) view for an axial image due to the partial volume effect. Even detail of the brain structures in the axial image (top-left) is significantly degraded compared to the original image.

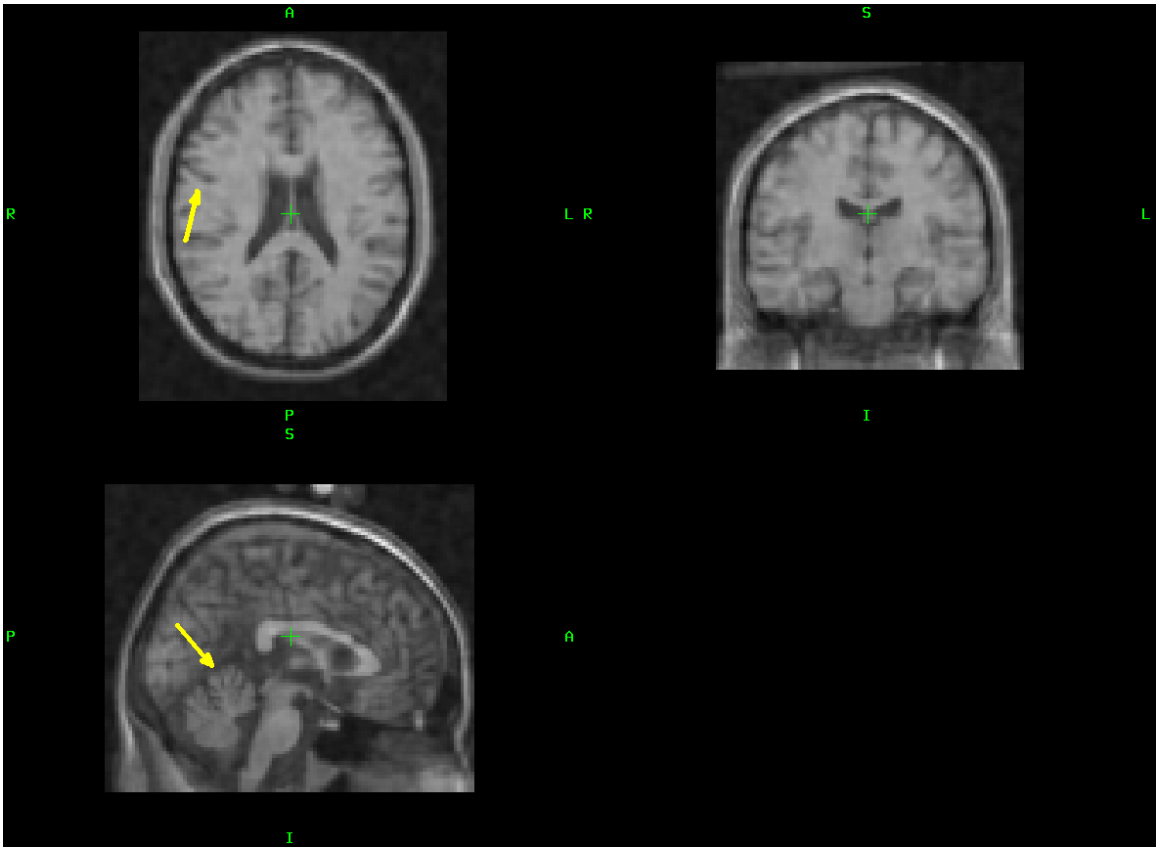

Figure 4.11: Reconstructed isotropic brain image using the proposed method with the Maximum rule. The reconstruction uses the axial image with thick slices in Figure 4.10 as input. The original image (ground truth) is shown in Figure 4.9. This experiment has successfully reconstructed much structure detail that does not present in the inputs such as the extensive foldings on the cerebral cortex (yellow arrow top-left) and structures of the cerebellum in the sagittal view (yellow arrow bottom-left).

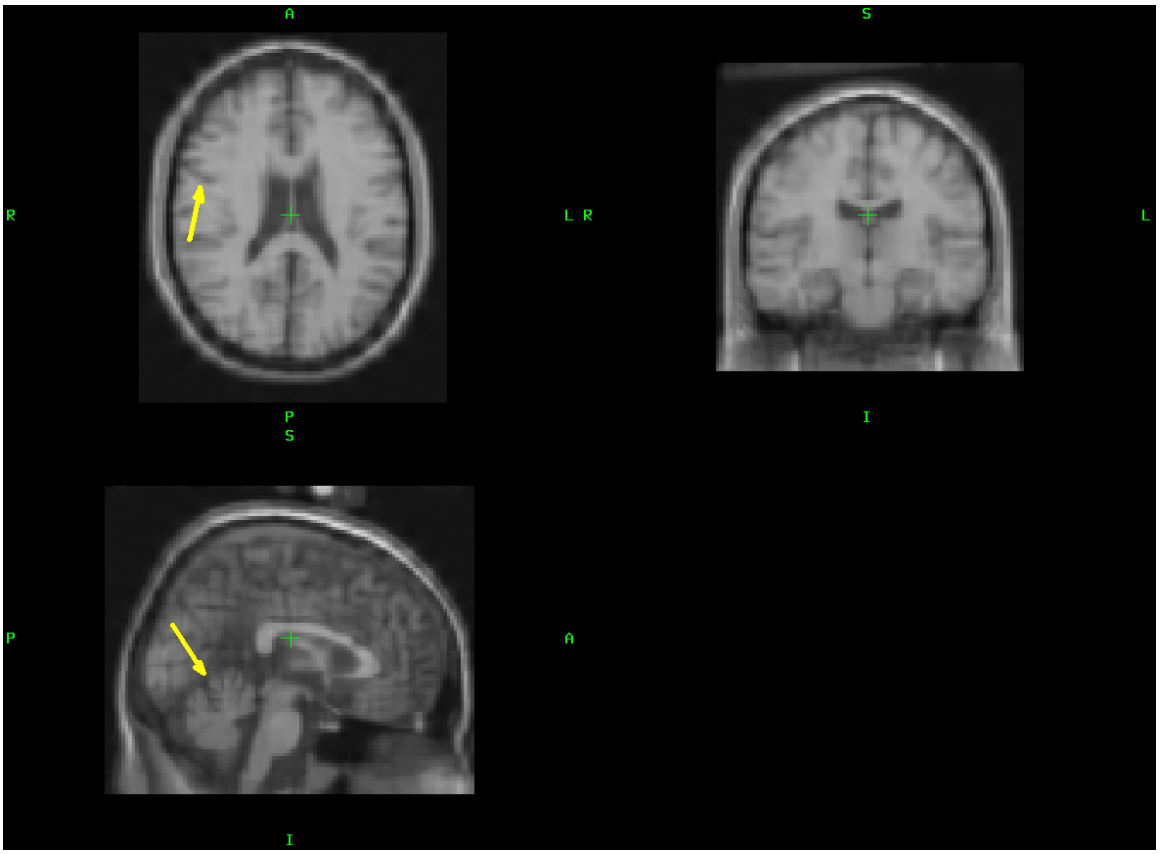

Figure 4.12: Reconstructed isotropic brain image using the proposed method with the Mean rule. The reconstruction uses the axial image with thick slices in Figure 4.10 as input. The original image (ground truth) is shown in Figure 4.9. This experiment has also successfully reconstructed much of the structure detail that does not present in the inputs, particularly in the sagittal and coronal view. Compared to the results by the Max rule in Figure 4.11, the structures on the cerebral cortex and the cerebellum (yellow arrows) are slightly blurrier but the overall image is perceived as being less noisy.

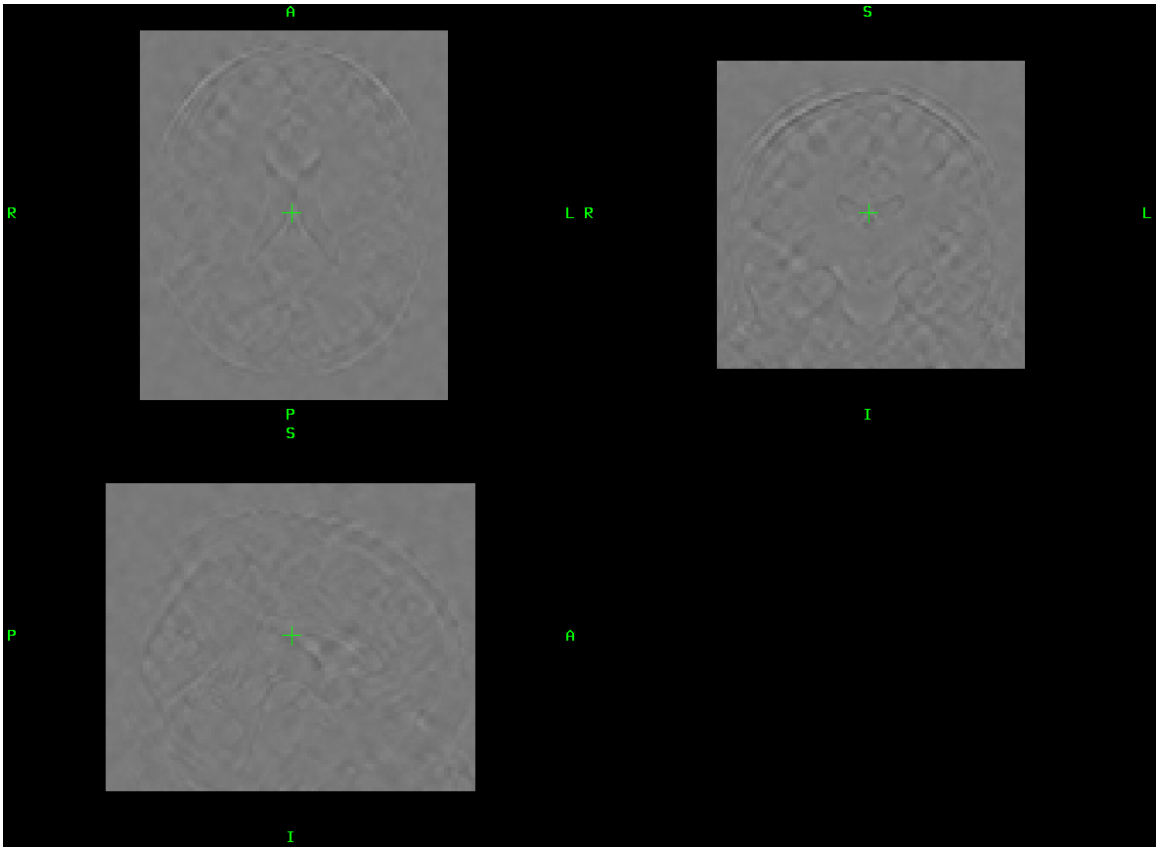

Figure 4.13: The difference image - the image generated by subtracting the Max rule reconstructed image (Figure 4.11) from the original ground truth image (Figure 4.9). Top-left: axial view. Top-right: coronal view. Bottom-left: sagittal view. Bright means large positive errors in the reconstructed image while dark indicates large negative errors. Zero error is represented in grey. This image shows that errors are generally low, and large errors are concentrated at edges with sharp intensity contrast such as the boundaries of the skull.

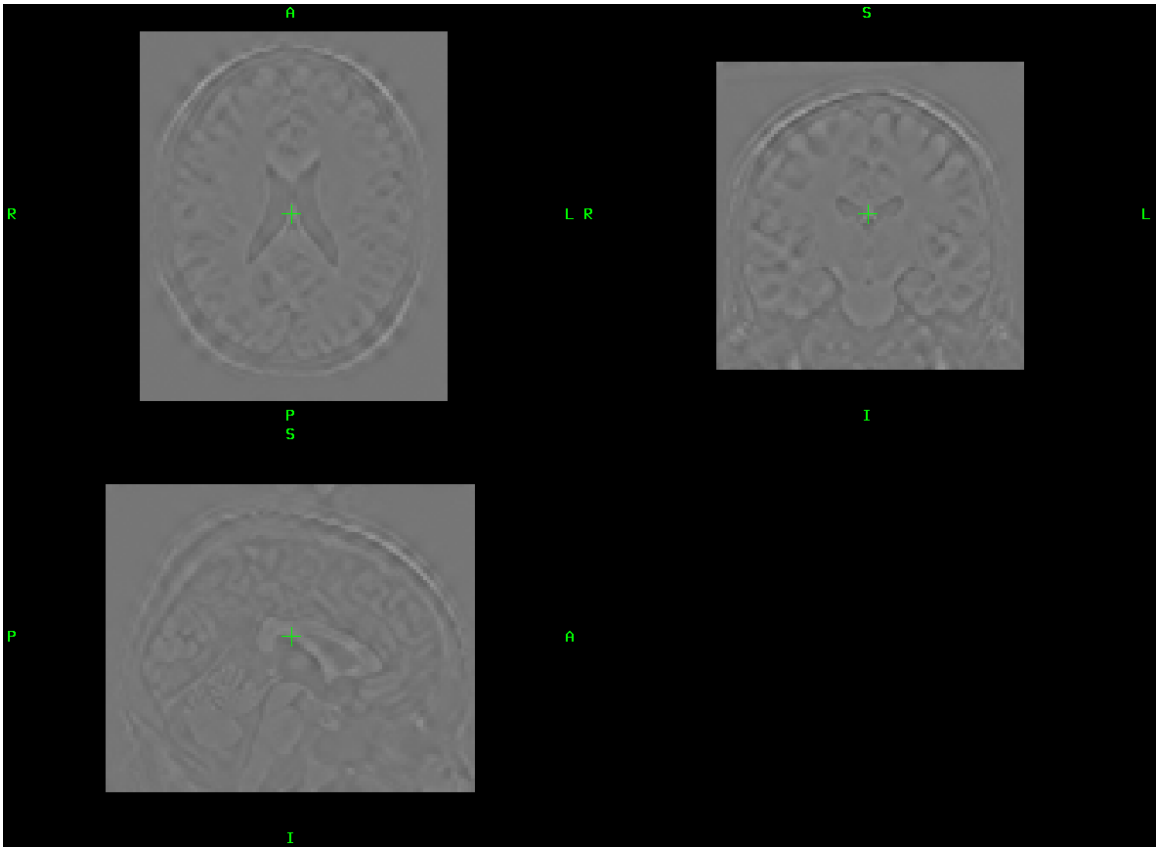

Figure 4.14: The difference image - the image generated by subtracting the Mean rule reconstructed image (Figure 4.12) from the original ground truth image (Figure 4.9). Top-left: axial view. Top-right: coronal view. Bottom-left: sagittal view. Bright means large positive errors in the reconstructed image while dark indicates large negative errors. Zero error is represented in grey. Compared to Figure 4.13, the errors here are greater concentrated at sharp edges. This is because the Mean rule effectively smoothens the image by taking the mean of input components. It also reflects the perception that the reconstructed image by the Mean rule is blurrier.

argument, we performed another set of experiments but this time instead of comparing the reconstructed image to the original image, we compared them to the blurred version of the original image. Before the comparison we passed an average filter to the original image. Two levels of blurring were used - a 2 by 2 by 2 and a 3 by 3 by 3 filter. The $SL2$ plot was then generated (Figure 4.15 and 4.16). When a blurred version of the original image is used as the benchmark, the mean rule outperforms the other rules, which indicates that there is a built-in averaging blurring effect inside the mean rule fusion. When the blurring level is high, the minimum rule performs the best at high frequencies (Figure 4.16) as it suppresses frequency components by design.

The plot in Figure 4.8 is for fusion of the x-Riesz component, where the max rule performs the best for low and mid frequencies and the mean rule overtakes for high frequencies. Similar results can be drawn from fusing the y- and z-Riesz component as well (Figure 4.17). Ideally, we want to track the plot with the lowest $SL2$. In all three directions, the performance flips around frequency 0.8 (measured in normalised frequency, where 1 represents the common highest frequency in all directions). In order to correct for this, we devise an adaptive approach to flip the fusion method according to frequency $|\mathbf{u}|$. We want to use the max rule when $|\mathbf{u}|$ is small and the mean rule when $|\mathbf{u}|$ is large; and the rule flips between each other around $|\mathbf{u}| = 0.8$. We propose an adaptive fusion rule, which is defined in the following:

$$F = \alpha F_{\text{MAX}} + (1 - \alpha) F_{\text{MEAN}} \tag{4.15}$$

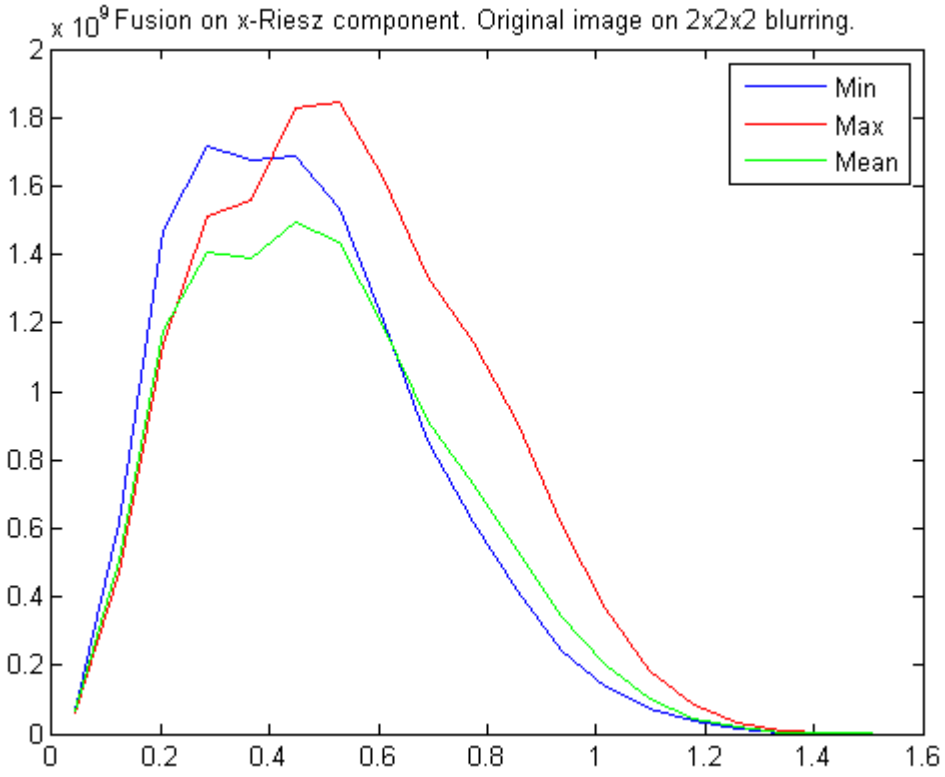

Figure 4.15: Error plots of the reconstructed image component in the x-direction by different fusion rules compared to the down-sampled ground truth image. Sum of L_2 norm of the error vectors are plotted against frequency. 2x2x2 average filter is applied to the original image. When comparing to an isotropic down-sampled ground truth, the Mean rule outperforms others in low frequencies because of its intrinsic blurring effect. In the high frequency region the Min rule gives the smallest errors as it suppresses high frequency components and hence intensity at sharp edges.

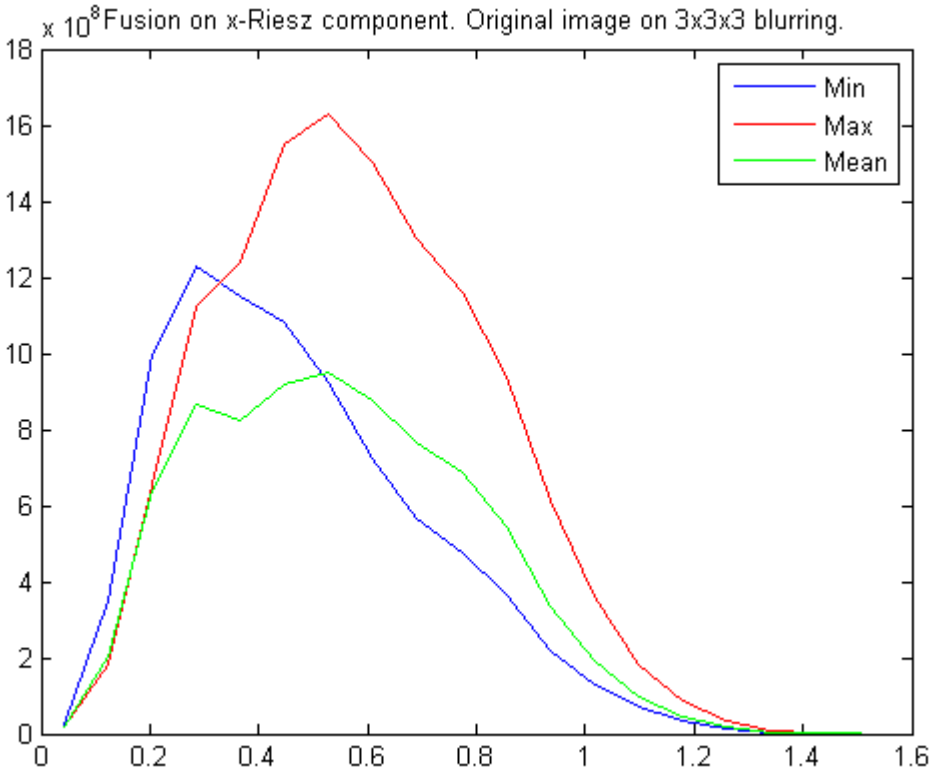

Figure 4.16: Error plots of the reconstructed image component in the x-direction by different fusion rules compared to the down-sampled ground truth image. Sum of L_2 norm of the error vectors are plotted against frequency. 3x3x3 average filter is applied to the original image. When the degree of down-sampling is higher (compared to 2x2x2 in Figure 4.15), the total error by the Mean rule is smaller but it is outperformed by the Min rule at a lower frequency. The Min rule also gives bigger advantage as a proportion to the other rules in the high frequencies.

where F_{MAX} and F_{MEAN} are the fused complex frequency components via the max and mean rule, respectively. The introduced parameter α governs the transition of the fusion rule and is frequency dependent,

$$\alpha = \frac{1}{1 + \exp\left(\frac{|\mathbf{u}| - a}{c}\right)} \tag{4.16}$$

When $|\mathbf{u}|$ is small (low frequencies), $\alpha \approx 1$ and the adaptive method is close to the max rule; when $|\mathbf{u}|$ is large (high frequencies), $\alpha \approx 0$ and the adaptive method is effectively using the mean rule. The parameter a defines the flipping point, i.e. when $|\mathbf{u}| = a$, $\alpha = 0.5$. Based on our previous observation, $a = 0.8$. The parameter c inside the exponent of α controls the range of the fusion rule transition, and $c = 0.05$ ensures that the transition is carried out in a relatively small frequency span (Figure 4.18).

The resulting fused brain image using the adaptive fusion rule is shown in Figure 4.19. Compared with the resulting image via the max rule, there is no significant visual difference. This is because the difference between the mean and max rule fusion for high frequencies is small. Moreover, the magnitudes of the high frequency components are themselves small relative to the lower frequencies. The effect of flipping between rules for those frequencies is not very significant visually. In theory, the adaptive method should output a smoother image than the pure max rule as it gives us the flexibility to trade off between image integrity (given by the max rule) and smoothness (given by the mean rule). The parameters a and c can be used to tune this trade-off for different applications when necessary.

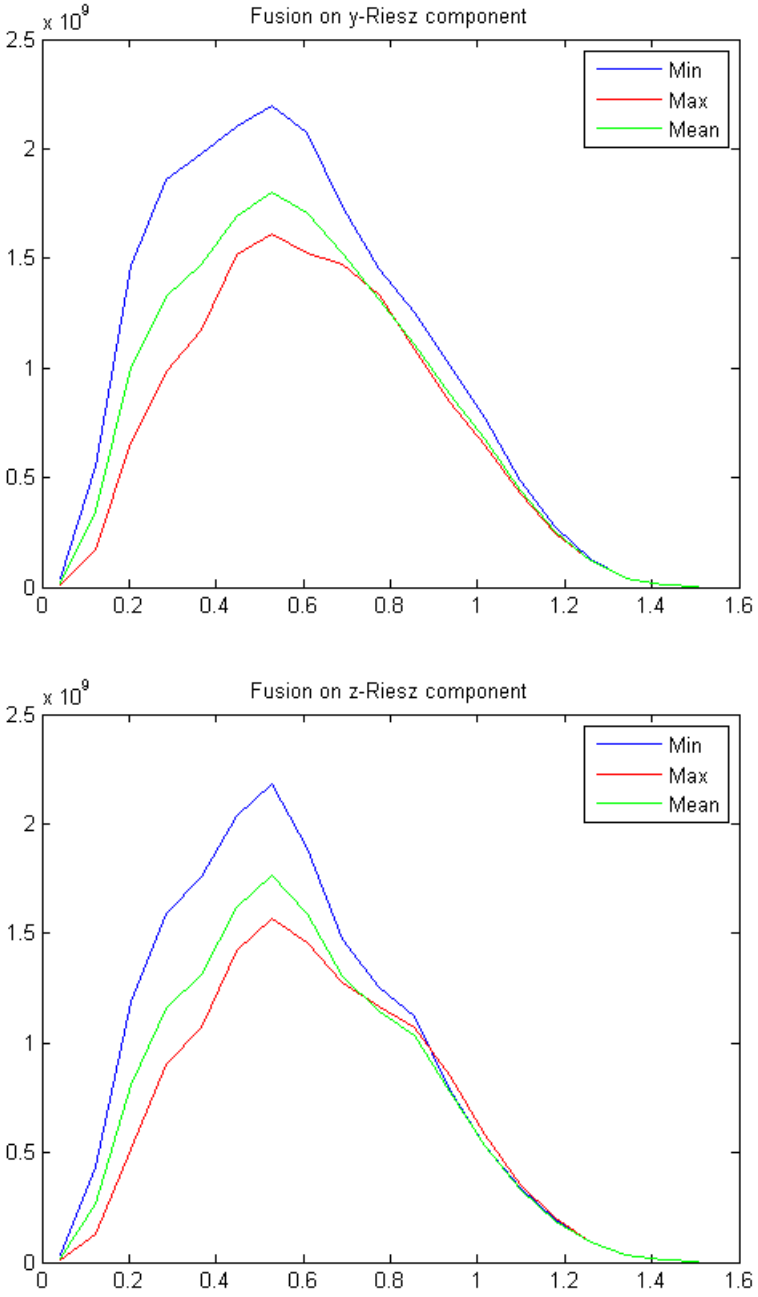

Figure 4.17: Error plots of the reconstructed image component in the y- (top chart) and z-direction (bottom chart) by different fusion rules compared to the ground truth image. Sum of L_2 norm of the error vectors are plotted against frequency. It shows that the Max rule is the method of choice for low frequencies but the best method may switch in the high frequency region. As the bottom chart shows, the Mean rule gives smaller error for fusing the z-component in high frequencies.

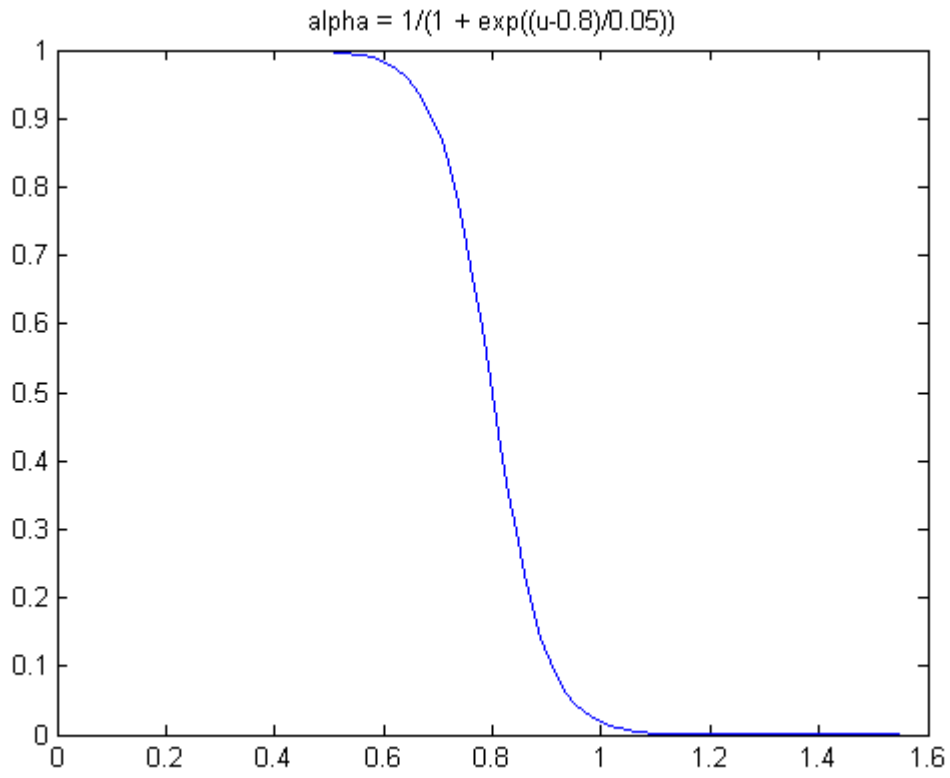

Figure 4.18: Plot of the proposed weighting factor α as a function of frequency (measured by normalised frequency where 1 represents the common highest frequency in all directions). In the low frequency band, $\alpha \to 1$ and the Max rule is applied; while in the high frequency band where $\alpha \to 0$ the Min rule is used.

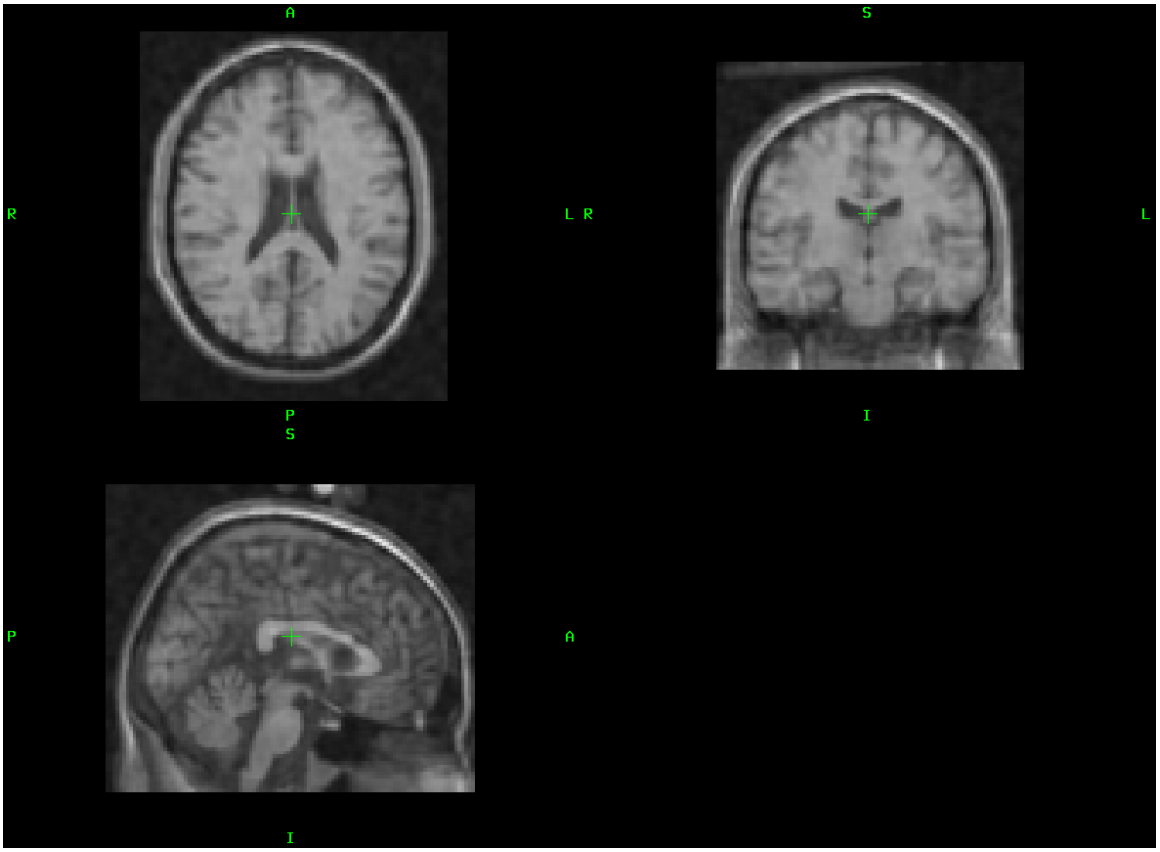

Figure 4.19: Final reconstructed isotropic brain image using the proposed method with adaptive fusion rules. The reconstruction uses the axial image with thick slices in Figure 4.10 as input. The original image (ground truth) is shown in Figure 4.9. This shows that the proposed adaptive fusion method successfully reconstructs most of the structure detail that does not present in the inputs. Compared to the results by the single Max or Mean rule in Figure 4.11 and 4.12, the adaptive fusion rule takes the best of both.

Image blurring	Slice Thickness	Min	Mean	Max	Adaptive
No blurring (ground truth)	4	2.910×10^9	1.405×10^9	0.907×10^9	0.907×10^9
	6	6.049×10^9	3.027×10^9	1.983×10^9	1.984×10^9
	8	9.089×10^9	4.677×10^9	3.101×10^9	3.103×10^9
2x2x2 average filter	4	3.670×10^9	3.434×10^9	4.206×10^9	4.180×10^9
	6	5.355×10^9	3.944×10^9	4.511×10^9	4.489×10^9
	8	7.439×10^9	4.711×10^9	4.818×10^9	4.799×10^9
3x3x3 average filter	4	1.032×10^9	1.067×10^9	2.111×10^9	2.083×10^9
	6	2.285×10^9	1.246×10^9	2.185×10^9	2.161×10^9
	8	4.138×10^9	1.855×10^9	2.408×10^9	2.388×10^9

Table 4.1: Fusion rule comparison. Performance of different fusion rules are quantified by the sum of L_2 norm of the difference vector between the reconstructed and original image. The smaller the value the better. Without blurring of the original image, the Max rule performs the best; when blurring is applied, the Mean rule outperforms others. The Adaptive method is a mix between the Mean and Max rule.

4.5.2 Validation

All four proposed fusion methods (using Minimum, Maximum, Mean, and Adaptive fusion rule) were applied to fuse the simulated axial, sagittal and coronal brain image. The SSD between the reconstructed image and the original image was then computed. Three levels of slice thickness were chosen, which were 4, 6 and 8 mm (the original isotropic image has a dimension of 1x1x1 mm per voxel). In addition, SSD was also calculated between the reconstructed image and a blurred version of the original image. Two levels of blurring were chosen - a 2x2x2 and a 3x3x3 isotropic averaging blurring filter was applied. The results are listed in Table 4.1.

Compared with the original image without blurring, the max rule performs the best. This is as expected since it has the lowest $SL2$ plot in the frequency domain. By the SSD measure, it also outperforms the adaptive rule by a small margin. This supports our hypothesis that the max rule preserves the image integrity as it restores

sharp edges. The mean rule gives higher SSD measure, which means the sum of errors of the mean rule reconstruction is higher, even though it looks smooth and noise free by visual assessment. This is because the mean rule smooths sharp edges. When the noise level is low (3% of the brightest tissue in our brain image), the SSD over a smoothed edge can be significantly larger than the SSD due to pure noise. This effect is best demonstrated with a simulated ideal image. Figure 4.20 shows an intensity profile of a line going through an ideal binary sphere (1 inside the sphere and 0 outside) and its reconstructed versions. Here it can be seen that the mean rule (red line) reconstructs the image (blue line) almost noise free (flat in homogeneous regions) except that it smoothes the image around sharp edges. In contrast, the max rule (green line) restores the sharp edges better, however, it also introduces artificial noise over the whole image. In aggregate, the mean rule still contributes to a larger SSD, 0.8682 versus 0.7609 for the max rule, because the noise level is low relative to the intensity contrast. This leads to a higher SSD reading for the mean rule even though it gives smoother image by visual assessment.

When the original image is simulated into thick slices by averaging the voxels in through-plane direction, the sampling rate is effectively reduced. Hence high frequency components, or details of the image are lost. Therefore, it is not possible for the reconstruction method to restore the original image fully, because the total information contained by the axial, sagittal and coronal images is less than the total original information. Assuming the fusion method restores the isotropy of the anisotropic thick slices but not necessarily the full frequency spectrum, it is reasonable to compare the reconstructed image to the original image with isotropic blurring. By

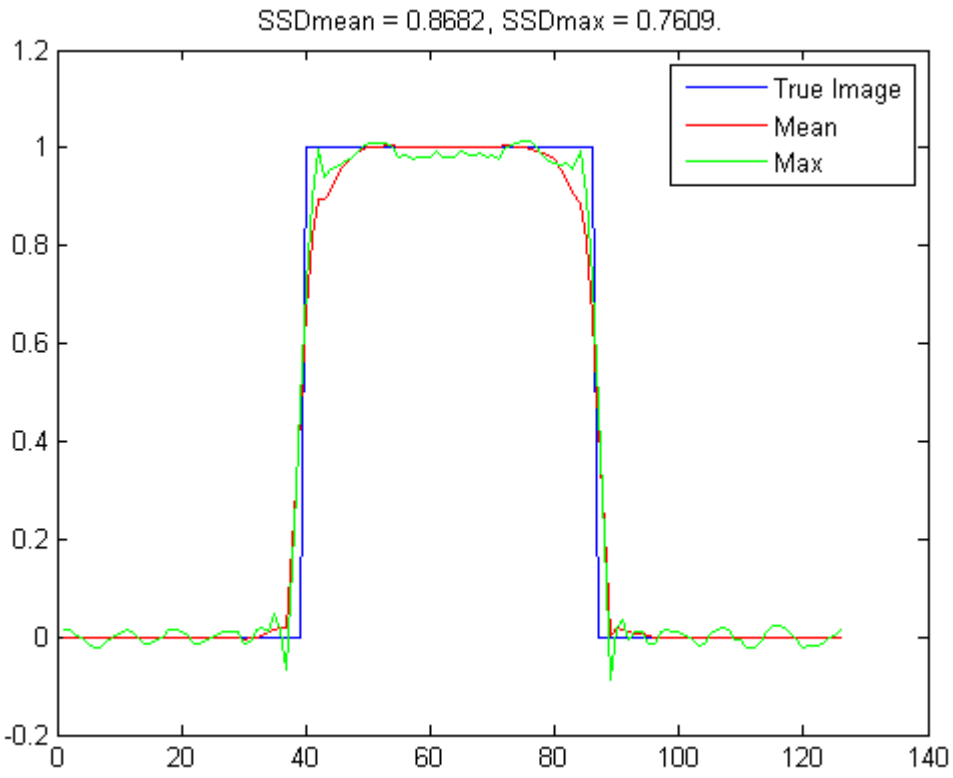

Figure 4.20: A line profile of a binary sphere (1 inside the sphere and 0 outside) and the reconstruction of it using the Mean and Max rule. The Mean rule gives a nice smooth reconstruction of a simple binary sphere. However, it loses significant amount of intensity at the sharp edge as a result of its blurring effect. In contrast, the Max rule reconstruction is more responsive to the sharp edge but it is noisier and more prone to reconstruction artifacts at supposedly steady regions. The Max rule preserves image integrity better and the Mean rule gives smoother reconstructions.

this standard, the mean rule yields the lowest SSD (Table 4.1), which suggests that the reconstructed image by the mean rule is the closest to the blurred version of the original image. When blurring increases from 2x2x2 to 3x3x3, the SSD of the mean rule reduces even further. This supports our previous observation that the mean rule smoothes sharp edges while introducing little artificial noise, which is the same effect of increase in blurring.

For completeness the SSD measures of the min rule are also given but most of the time it performs worse than other methods. Finally, the adaptive approach fares very close to the max rule, which may be due to our choice on parameters a and c. It may also be the case for this particular image. The difference between the adaptive approach and the max rule would be larger if the image has bigger magnitude in high frequencies, or a higher noise level. After all, the adaptive fusion method is a flexible approach and Table 4.1 shows that its performance is between the max and mean rule, which is expected.

4.5.3 Riesz Transform versus Discrete Wavelet Transform

In order to assess the performance of the Riesz fusion method, we developed a method that is similar in all respects except that it decomposes the signal using the discrete wavelet transform (DWT) instead of the Riesz transform. In addition to decomposition in orthogonal basis orientations, the DWT also consists of a set of bandpass filters that decompose the signal in scales. A 3-D DWT decomposes a 3-D image into 8 components in a tensor product construction - LLL, LLH, LHL, HLL, LHH,

Image	DWT SSD	DWT time (s)	Riesz SSD	Riesz time (s)	Compare SSD (%)	Compare time (%)
3-D brain image	1.819	4.924	1.443	2.222	20.67%	54.87%

Table 4.2: Riesz vs. DWT fusion. Sum of squared difference and computational time. The Riesz fusion method performs both better and faster compared with the fusion using DWT.

HLH, HHL, and HHH (a 3-D extension of equation 2.43). Here "L" and "H" stands for a lowpass and highpass operation in the DWT, respectively. For instance, LHL represents the frequency component that is a lowpass in x-direction, highpass in y-direction, and lowpass in z-direction. Scale selection for the DWT is done in a dyadic fashion. The number of scales is chosen to be three. In fact, the number of scales does not matter in this application as long as all frequency components are covered equally. The mother wavelet is chosen from the Daubechies, symlets and Coiflets families. The wavelet that reconstructs the simulated test image with the smallest SSD is then chosen for the purpose of comparison. In the case of the brain image, the Coiflets 5 was used. In addition to similarity measure, the computational time was also recorded and is listed in Table 4.2.

According to this measure, the Riesz fusion method outperforms the DWT fusion method by 20.67% (smaller SSD) while it takes 54.87% less time than DWT fusion. To fuse three 3-D images each of size 109 by 91 by 91 voxels the Riesz fusion method requires slightly more than 2 seconds.

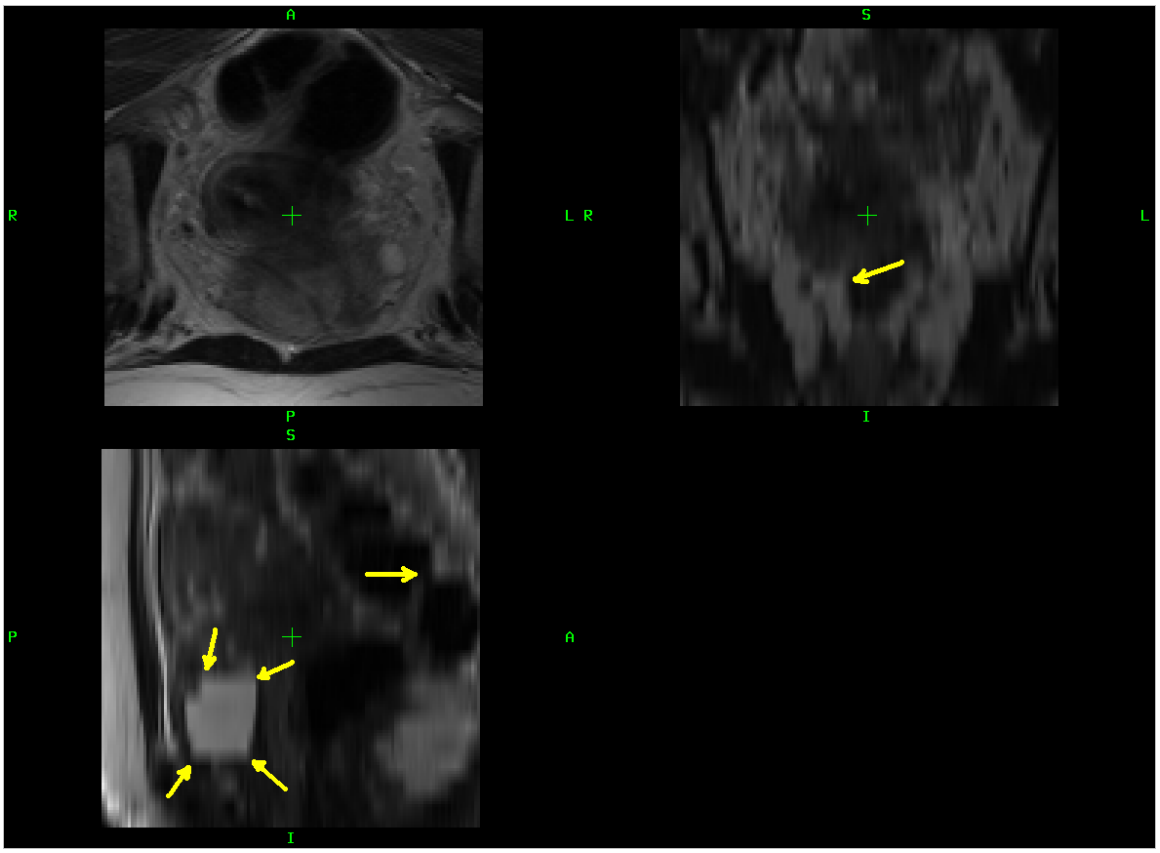

Figure 4.21: The abdominal MRI scan with thick axial slices are linearly interpolated as a pre-processing step. The linear interpolation smoothens abrupt intensity changes between slices; however, structures of the organs in the sagittal and coronal view are still heavily affected by the staircase effect (yellow arrows).

4.6 Fusion of Clinical Images

Figure 4.21 shows a 3-D slice view of the axial scan of a patient. Slices are linearly interpolated in the through-plane (superior-inferior) direction. From this figure we can see that although in-plane resolution is high, the through-plane resolution is low and non-isotropic (Figure 4.21 yellow arrows). For instance, in the bladder, the upper and lower tangential slices distort its boundary due to the partial volume effect. A 3-D segmentation of the bladder in this image would result in a flat upper/lower boundary and "staircase" effect between adjacent slices as shown before in Figure 4.3.

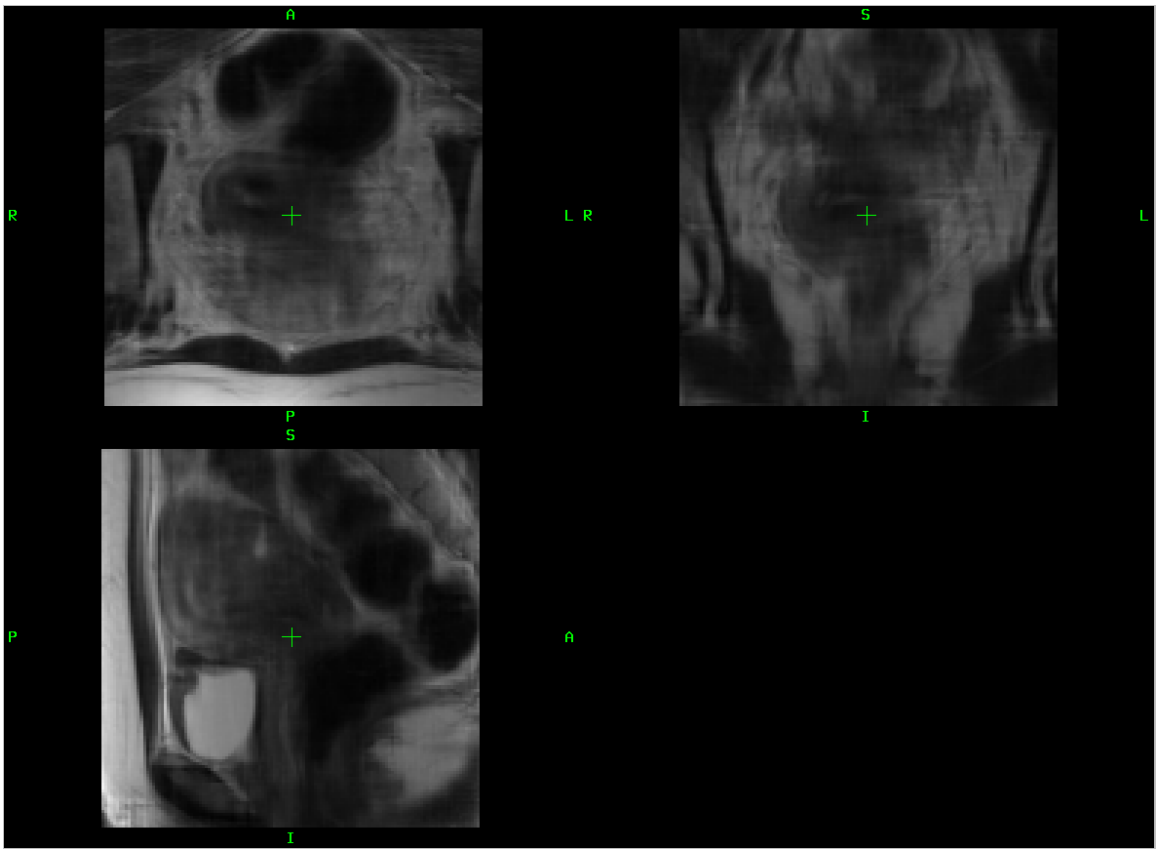

Figure 4.22: Reconstructed 3-D isotropic abdominal image using the proposed fusion method with three clinical scans (axial, coronal and sagittal) as input. Compared to the single axial scan in Figure 4.21, much of the structures in the sagittal and coronal view are successfully reconstructed. Boundaries of the bladder and some of the bowels are better defined for segmentation.

Figure 4.22 shows the fused image after the Riesz fusion method is applied. The image quality in different views are much improved. Details of the structures are more visible in all three orientation as their frequency components are complemented by different scans. As a result, a better 3-D model of the organs can be drawn from this fused image.

4.7 Conclusions

In this chapter we present a method to fuse multi-view MRI images. The purpose of this fusion is two fold: 1), to correct for the intensity distortions that are caused by the partial volume effect of long-thin voxels as a result of the anisotropic frequency sampling during image acquisition; and 2), to reconstruct an image that combines details in multi-view scans and has isotropic voxel dimensions for more accurate further image processing. We propose to use the Riesz transform as an isotropic allpass filter to decompose the images; then suggest several fusion rules to re-combine image components. We validate the method with simulated images, where the ground truth is known; and also show the improvement after application of the method to real clinical abdominal images. In the next chapter, we build on the existing level set models and present a method that computes phase congruency but from a different perspective.

Chapter 5

Level Set Segmentation

Segmentation is the process of automatically finding a desired boundary of an object. In our example, the object of interest is the bladder in the patient's abdominal MRI scans. The boundary in an image $I(\mathbf{x})$ can be represented as a set $C = \{\mathbf{x} \mid p(\mathbf{x})\}$, where $p(\mathbf{x})$ are the points on the boundary with coordinates $\mathbf{x}$. The automatic boundary seeking process is achieved by an optimization approach, where a pre-defined energy is minimised. In the level set representation of contours, the number of points $p(\mathbf{x})$ changes at each iteration during the optimization process. In addition, each point coordinate $\mathbf{x}$ is an n-dimension vector, which makes C a set of n-dimension vectors. The level set is a contour/boundary representation to simplify this optimization process. Instead of using C (a vector field) to represent a contour, a scalar function ϕ is defined such that $C = \{\mathbf{x} \mid \phi(\mathbf{x}) = 0\}$. By doing so, the vector field optimization is reduced to a scalar field optimization problem. As a result, the level set method is a very convenient and powerful approach to solve evolving contours in multi-dimensions. Most commonly, $\phi(\mathbf{x})$ is defined as the signed distance function of

the binary evolving contour, which preserves useful properties such as curvature and enables narrow band estimation to avoid unnecessary computations.

As for any optimization problem, there is a meaningful energy term to be optimised. This energy term is a function of ϕ, denoted by $E(\phi)$. Broadly speaking, there are two main types of energies - region and edge based energy. A region based energy is stationary (at maximum/minimum) when a distinctive region in the image is identified. Similarly, an edge based energy is stationary when the contour coincides with edges, including asymmetric (step changes, ramps) and symmetric (ridges, valleys) edges. The effectiveness of these two energy formulations is very much image dependent. Evidently, region and edge based energies are good for different types of images. The effectiveness of either of these two types of energies for our application is discussed in this chapter. Once an energy $E(\phi)$ is formulated, the iteration step $\frac{\partial \phi}{\partial t}$ of the optimisation is derived from $E(\phi)$ via the Euler-Lagrange equations. Ultimately $\frac{\partial \phi}{\partial t}$ is all it needs to solve the energy optimization problem. For any given $\phi_k(\mathbf{x})$, $\phi_{k+1}(\mathbf{x})$ can be found via $\phi_{k+1}(\mathbf{x}) = \phi_k(\mathbf{x}) + \frac{\partial \phi}{\partial t} \triangle t$ as a closer estimation to the final solution when $E(\phi)$ is stationary.

In this chapter, we apply this level sets framework to our segmentation problem and focus our effort on assessing the effectiveness of different energy formulations. In addition, we incorporate a scale factor into both the region and edge based methods.

5.1 Region Based Energy Minimisation

The region based method is ideal for segmenting statistically homogeneous regions where clearly defined edges between the region and the background may not exist. Chan and Vese [6] were inspired by the Mumford-Shah functional for segmentation and introduced their energy minimisation for level sets that is closely related to the minimum partition problem. The energy that Chan and Vese proposed is

$$E = E_1(C) + E_2(C) = \int_{\text{inside}(C)} |I(\mathbf{x}) - c_1|^2 \, d\mathbf{x} + \int_{\text{outside}(C)} |I(\mathbf{x}) - c_2|^2 \, d\mathbf{x} \quad (5.1)$$

where C is the evolving contour, $I(\mathbf{x})$ is the image intensity, c_1 and c_2 are the mean intensity inside and outside C, respectively. Note that the integrals E_1 and E_2 are the sum of squared differences inside and outside the evolving contour. The differences are between the image intensity and the mean intensity pixel by pixel. This formulation is very closely related to the variance of intensity inside and outside the evolving contour. Effectively, the Chan and Vese model aims to segment an image into two parts - the foreground and background, in such a way that the intensity variances of the foreground and background are minimised at the same time. The great strength of this model is that it does not require an explicit edge to partition the foreground and background, and hence the model is called "active contours without edges". This model works particularly well in segmenting highly homogeneous regions with or without sharp edges and images with smooth intensity variation such as a pure slope

or a Gaussian image (Figure 5.1). Since the Chan and Vese model is based on a statistical measure, or variance, and not on edge detection, it is more resilient to noise than many edge based methods. However, there are two fundamental problems with the method. First, region based methods cannot differentiate between objects and boundaries. For example, some objects have a similar intensity appearance to the background but may be partitioned by distinct boundaries. This model will only recognise the edges as the foreground and everything else, including the object and background, as the background (Figure 5.2). In many cases, segmenting the object, despite its relative appearance to the background is more desirable. The second problem with the region based method is more significant in clinical applications. In the case that the object (desirable foreground) has a varying appearance, the Chan and Vese model may fail to differentiate it from the background even though there are edges between the object and the background. Lankton and Tannenbaum [29] recognised this problem and illustrated it in Figure 5.3, where both the object (foreground) and the background have heterogeneous intensity. Since the Chan and Vese model minimises the intensity variance on a global basis, it inevitably fails to differentiate high intensities in the object and in the background. This is particularly a problem for real MRI images, where, for example, the magnetic bias field effect artificially increases image intensity near the body surface and causes non-uniform intensity appearance of the same tissue. In order to overcome this issue, Lankton and Tannenbaum proposed to apply the region based energy minimisation locally. A local mask $B(\mathbf{x})$ was introduced:

$$B(\mathbf{x},\mathbf{p}) = \begin{cases} 1, & \|\mathbf{x}-\mathbf{p}\| < r \\ 0, & \text{otherwise.} \end{cases} \tag{5.2}$$

where r is a pre-defined radius of the circular local mask and $\mathbf{p}$ is its centre. The purpose of this mask $B(\mathbf{x})$ is to mask out a local region of the image where the Chan-Vese model is then applied locally. The minimisation energy of the localised Chan-Vese model becomes:

$$\begin{aligned} E_{\text{local}}(\mathbf{p}) &= E_1(C) + E_2(C) \\ &= \int_{\text{inside}(C)} B(\mathbf{x},\mathbf{p}) |I(\mathbf{x}) - c_1|^2 \, d\mathbf{x} + \int_{\text{outside}(C)} B(\mathbf{x},\mathbf{p}) |I(\mathbf{x}) - c_2|^2 \, d\mathbf{x} \end{aligned} \tag{5.3}$$

Then the global energy is assembled by $E = \int E_{\text{local}}(\mathbf{p}) \, d\mathbf{p}$. This method effectively solves the minimum partition problem in a local patch - that is, it identifies piece-wise constant regions locally; and the global energy reaches the minimum when every local energy is minimised. This method generally solves the problem of the traditional region based approaches in segmenting objects with non-uniform intensity, and offers some controls over the scale of intensity contrast. In contrast to the traditional scale which picks up a particular bandwidth of the frequency spectrum, the localisation in region based energy controls the range of enhancement of intensity contrasts in a histogram.

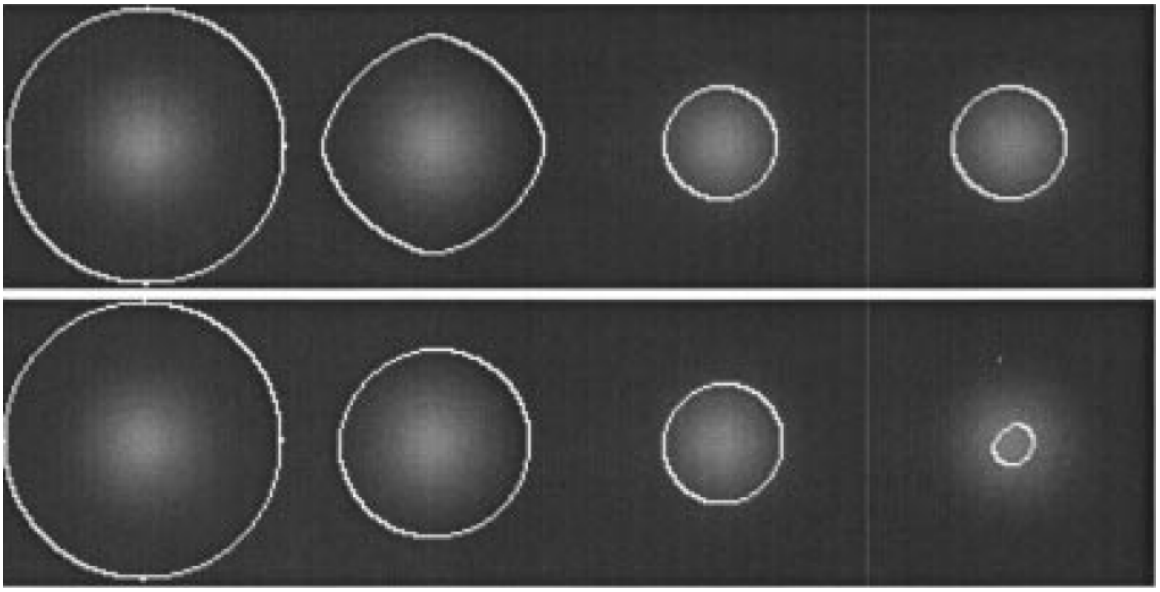

Figure 5.1: Segmentation of an object with gradual changing contour. Top-row: evolution process of the active contour without edges method. Bottom-row: evolution process of the classic edge based active contour model. Left: initialisation of the evolution process, which is set the same for both methods. Left-to-right: intermediate evolution snapshots. Right: final segmentation results for the object. Source: Chan and Vese [6] Fig. 9 Page 272.

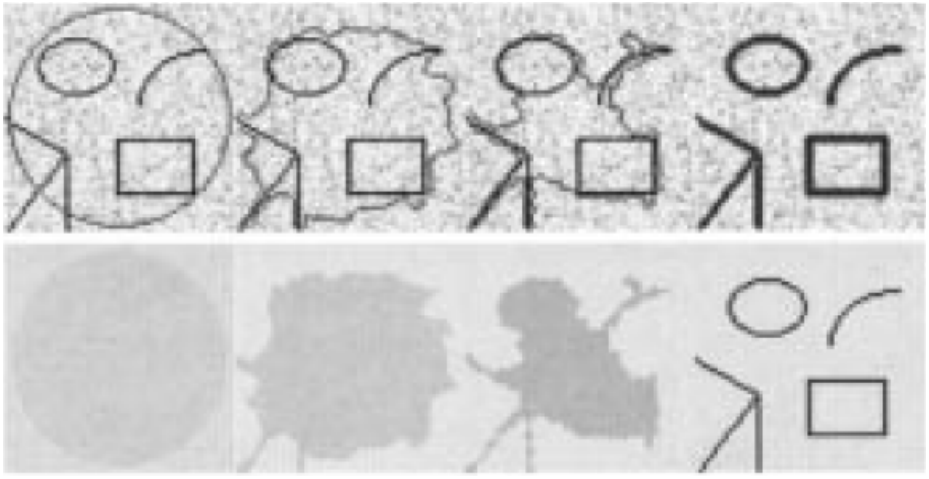

Figure 5.2: Segmentation of edges and curves using the Chan and Vese model, which separates the edges as foreground from the background. Top-row: the evolution of the segmentation contour on top of the original image. Bottom-row: the evolution of the segmentation mask. Source: Chan and Vese 2001 paper [6] Fig 6 Page 271.

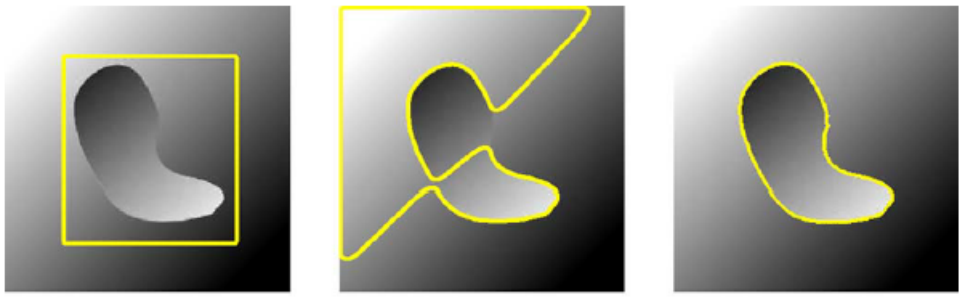

Figure 5.3: Segmentation of a blob object with heterogeneous intensity. (a). Initialisation of the segmentation problem. (b). Unsuccessful result of the classic region-based segmentation method, which is confused with the foreground/background intensities. (c). Successful result of edge-based segmentation technique, which is optimised at sharp local intensity change. Source: Lankton and Tannenhaum 2008 paper [29] Fig. 1 Page 2029.

5.2 Edge Based Energy Minimisation

The fundamental governing equation of the edge based method is defined by the level sets' evolution step:

$$\frac{\partial \phi}{\partial t} = |\nabla \phi| \operatorname{div}\left(g\left(\mathbf{x}\right) \frac{\nabla \phi}{|\nabla \phi|}\right) \tag{5.4}$$

The step $\frac{\partial \phi}{\partial t}$ tells us at each evolution step the size and direction of the level set function ϕ. $\nabla \phi$ is the gradient of ϕ - a vector field and the normals of the level sets. $|\nabla \phi|$ is the magnitude of these vectors and represents the slope of the level sets. In the case of ϕ being the Euclidean distance function of the zero level set, $|\nabla \phi| = 1$ because each level set $\phi = c_1$ and another level set $\phi = c_2$ have the same Euclidean distance between them. The term $\frac{\nabla \phi}{|\nabla \phi|}$ is the normalised gradient of ϕ, hence unit normals of the level sets. The divergence of unit normals $\operatorname{div}\left(\frac{\nabla \phi}{|\nabla \phi|}\right)$ are equal to curvatures κ along the level sets. Instead, the unit normals are weighted by an edge function $g\left(\mathbf{x}\right)$,

which is ideally smooth and equal to zero at a desirable edge and one otherwise. As a result, $\operatorname{div}\left(g\left(\mathbf{x}\right)\frac{\nabla\phi}{|\nabla\phi|}\right)$ is equal to zero at edges and κ otherwise. Therefore, the level set function ϕ settles on edges presented by $g\left(\mathbf{x}\right)$ and propagates forwards according to the curvature κ of the zero level set. This can be seen more clearly by expanding the divergence term by the product rule of divergence:

$$\operatorname{div}\left(g\left(\mathbf{x}\right)\frac{\nabla\phi}{|\nabla\phi|}\right)=g\left(\mathbf{x}\right)\cdot\operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right)+\nabla g\left(\mathbf{x}\right)\cdot\frac{\nabla\phi}{|\nabla\phi|} \tag{5.5}$$

The first term is the curvature of the evolving curve weighted by the edge function. The second term is the dot product of the gradient of the edge function $\nabla g\left(\mathbf{x}\right)$ and the unit normal of the evolving curve $\frac{\nabla\phi}{|\nabla\phi|}$, i.e. the projection of $\nabla g\left(\mathbf{x}\right)$ to the normal direction of the evolving curve. The importance of this term is that it attracts the evolving curve to the desired edges from either direction when the curve is near those edges. Because of this term $\nabla g\left(\mathbf{x}\right)$, it is important that the edge function $g\left(\mathbf{x}\right)$ is smooth. If $g\left(\mathbf{x}\right)$ consists of sharp edges, its gradient $\nabla g\left(\mathbf{x}\right)$ will be large and lead to unstable evolving steps.

In order to make the curve evolving model more flexible, it is common practice to apply a regularisation term $cg\left(\mathbf{x}\right)$ and the updating step becomes

$$\frac{\partial\phi}{\partial t}=|\nabla\phi|\left(\operatorname{div}\left(g\left(\mathbf{x}\right)\frac{\nabla\phi}{|\nabla\phi|}\right)+cg\left(\mathbf{x}\right)\right) \tag{5.6}$$

The second term is a constant weighted by the edge function. This constant is similar to Cohen's [10] balloon force on active contours, which pushes the curve outwards/inwards when it is not near the edges. The constant c defines the direction

and magnitude of the curve's evolving velocity. It is weighted with the edge function, so the balloon force goes to zero when the curve reaches the edges. With this balloon force, segmentation can be initialised inside/outside the object and the curve grows outwards/inwards towards the object boundary.

This is a classic edge based model and is proven to model surface evolution well. The key is the design of the edge function $g(\mathbf{x})$, which is supposed to represent edges. The early authors who invented this model originally used gradient based methods to represent edges. Mostly commonly it is the gradient of the Gaussian blurred image $|\nabla G * I(\mathbf{x})|$, where $G(\mathbf{x})$ is a pre-defined scale-controlling Gaussian filter. However, as we have discussed in earlier chapters, gradient based edge detection has many known problems, including intensity dependency, which we want to avoid for segmenting objects with non-uniform intensities. In the following section we investigate and propose alternative models.

5.3 Local Phase-Based Edge Detection

As opposed to the gradient based method, local phase-based edge detection has several advantages. First, it is less intensity dependent. Since local phase is a local measure of relative strength of both even (lines, ridges, valleys) and odd (step changes, slopes) features, it is a more consistent edge representation in general. Edges with variable intensity may have the same phase while its gradient map appears differently. Second, as one of the consequences of that observation, local phase is a more sensitive edge model. It is capable of capturing a small kink in a smoothly varying function, which

would be hard to achieve by the gradient based method. However, along with its high sensitivity to features, it is also sensitive to noise. This is a problem with the gradient based method as well, apart from the fact that the image is usually Gaussian blurred before computing its gradient. This leads to the third advantage of the local phase-based method - flexibility of filter choice. Since a bandpass filter is used, a specific frequency band can be chosen appropriately, unlike Gaussian blurring, which is intrinsically a lowpass process. Moreover, bandpass filtering can simulate frequency band selection from wavelets and construct the filter in a cascade fashion. We analysed scale and local phase in Chapter 2 and 3, respectively; here, we recall the salient features for convenience.

The single scale local phase of an n-dimension function is defined by the Riesz transform as:

$$\text{local phase} = LP = \arctan\left(\frac{\left|\mathcal{H}\left(I^B\right)\right|}{I^B}\right) \tag{5.7}$$

where $\mathcal{H}$ represents the Riesz transform and I^B is a bandpass version of the image as $I^B = B * I(\mathbf{x})$. The bandpass filter $B(\mathbf{x})$ plays a role for frequency selection and controls the scale of analysis. In addition, this bandpass filter $B(\mathbf{x})$ is assumed to filter out the zero-frequency component, i.e. no DC level, which is effectively subtracting its mean value from the original function. This is necessary because the DC level of the function will distort the local phase calculation. After all, local phase is a measure of the change of function, i.e. non-zero frequency components. The arctan function is one measure of the relative strength of $\left|\mathcal{H}\left(I^B\right)\right|$ and I^B, which

are also known as the odd and even part of the function representing odd and even features, respectively. In homogeneous regions, the local phase is equal to 0 or π as there is little change of the function and $\left|\mathcal{H}\left(I^B\right)\right| \approx 0$. At sharp edges $\left|\mathcal{H}\left(I^B\right)\right|$ is large and I^B tends to have negative values on one side of the edge and positive values on the other due to the elimination of the DC level. As a result, the local phase tends to flip signs and show large phase changes around edges. This is useful because it is guaranteed to have significant $\nabla g\left(\mathbf{x}\right)$ in the edge based level set formulation.

The local phase measure of features works exceptionally well in synthetic noiseless images. However, a frequently-encountered problem of this approach is its sensitivity to noise. One solution is to apply the idea of phase congruency. Morrone and Owens [35] observed that at most of the important features, the signal's frequency components are mostly in phase. For example, the Fourier series of a step change consists of primarily a set of sine waves. While these sine waves are out-of-phase everywhere in the signal except at the place the step change occurs where the sine waves become in-phase. In fact, each of these sine waves can be considered to be a super narrowly bandpassed version of the original signal. The idea of phase congruency states that different frequency bands of the signal are in-phase locally at a genuine feature. That is to say, the local phases in different scales of a signal are mostly in-phase at a feature. In practice, replicating the infinitesimal narrow bandpass filters is neither feasible nor desirable as we pursue reliability as well as efficiency of the method. However, we could simulate this with a finite set of bandpass filters. Having computed the local phase in different frequency bands, we propose to use the variance to measure their consistency:

$$\text{Phase Variance} = \text{Var}\,(LP) = \frac{1}{N}\sum_{k=1}^{N}\left(LP_k - LP_{\text{mean}}\right)^2 \tag{5.8}$$

where N is the total number of scales. By calculating the variance of local phases through different scales, we explicitly measure the phase consistency at every local spatial point. As opposed to Kovesi's version of local energy weighted phase congruency, which gives higher weighting for phases with higher local energy, our model computes phase "congruency" more accurately by not biasing towards components with high energy. Consider Morrone and Owens' original observation as an example, the Fourier components of a step change are in-phase, however, the phase of a component with large magnitude does not necessarily contribute more than a component with small magnitude. In fact, in-phaseness at low-frequencies give us confidence in featureness while in-phaseness at high-frequencies provide us with increased localisation of the feature. Our model treats all phases equally at all different scales.

In order to address the problem of noise, Kovesi proposed an energy thresholding approach to suppress noise by first estimating a universal noise energy level and then subtracting this amount from the local energy (equation 3.7). We are sceptical about both this noise estimation method and the thresholding operation. We argue that a universal noise estimation only works well in estimating the type of uniform background noise. For real medical image processing, the challenges usually lie in other artefacts, such as the MR bias field effect, Gibbs ringing in filtering, and fusion artefacts. In theory, if the number of scales is sufficiently large, the changes induced by noise should be diluted. Moreover, this approach naturally combines individual

feature detection at different scales.

Two questions remain: how to compute the mean phase LP_{mean}, and how to choose the bandpass filter with appropriate scale parameters. We discuss these two issues one by one in the following sections.

5.4 Phase Wrapping and Mean Phase Computation

When computing the arctan function in equation 5.7, phase wrapping is unavoidable. By convention, $\arctan\left(\frac{y}{x}\right) = \theta + k\pi$, where $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ and $k \in \mathbb{Z}$. By examining the sign of x and y separately, we could extend the range of θ to $(-\pi, \pi)$. However, there is not an easy way to recover the integer k. As a result, every time k switches from one integer to another θ flips sign, e.g. from $\frac{\pi}{2}$ to $-\frac{\pi}{2}$. This results in a zigzag shape in the measure of local phases, which would lead to false feature detection (the local phase value also flips sign at line features). Some solutions have been suggested. One is to avoid the arctan function by generating a look-up table instead. Another is to shift the phase vectors by 90 degree before calculating their phase. This assumes that the phase values of feature points cluster around $-\pi$ and π, where most of the phase wrapping happens. Therefore, offsetting the phase by 90 degree could prevent excessive phase wrapping there. However, this requires pre-knowledge about the spatial distribution of phase and this is difficult to define.

The main problem with Kovesi's method of phase congruency is that each scale-

phase is amplitude-weighted. A straightforward answer to this is to normalise each of these phase vectors. Suppose there are three complex numbers $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, whose phases are θ_a, θ_b and θ_c, respectively. We could write them in polar form:

$$\mathbf{a} = |\mathbf{a}| \cos\theta_a + j\,|\mathbf{a}| \sin\theta_a \tag{5.9}$$

$$\mathbf{b} = |\mathbf{b}| \cos\theta_b + j\,|\mathbf{b}| \sin\theta_b \tag{5.10}$$

$$\mathbf{c} = |\mathbf{c}| \cos\theta_c + j\,|\mathbf{c}| \sin\theta_c \tag{5.11}$$

The question is to find $\mathbf{c}$ such that θ_c is the mean phase of θ_a and θ_b (Figure 5.4). Suppose $\mathbf{a}$ and $\mathbf{b}$ have the same length, i.e. $|\mathbf{a}| = |\mathbf{b}|$, which can be achieved by normalising both of them. Then they form a rhombus in Figure 5.4. By geometry, it is obvious that $\mathbf{c} = \mathbf{a} + \mathbf{b}$, and its phase

$$\theta_c = \theta_a + \frac{\theta_b - \theta_a}{2} = \frac{\theta_a + \theta_b}{2} \tag{5.12}$$

Normalisation is the key to this method. It is needed for each phase vector before every single summation. If there are N scales, the number of normalisations needed would be $2\,(N-1)$. Each normalisation requires 5 multiplications, divisions, or square root operations. Hence the total number of operations for this approach would be $10\,(N-1)$. In practice, it is intended to preserve the original phase vector in each scale. It is not desirable to overwrite the normalised vector to the original one,

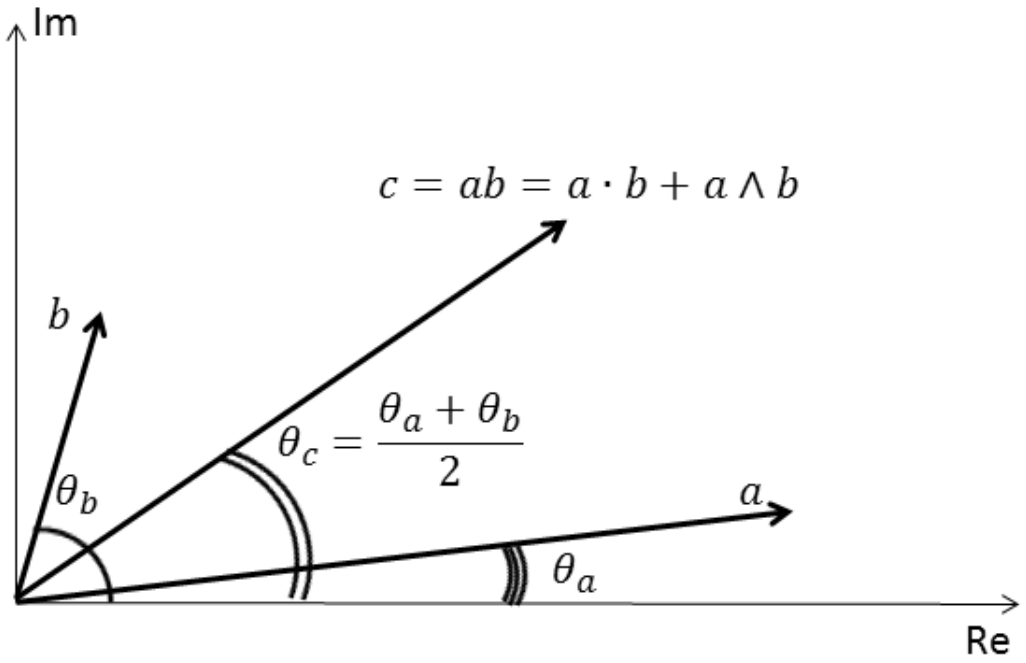

Figure 5.4: Mean phase of two vectors. Given two vectors **a** and **b** with phase θ_a and θ_b, respectively, the vector **c** whose phase is the mean of θ_a and θ_b is given by the geometric product of **a** and **b**.

therefore, extra memory would need to be allocated for this calculation. Moreover, the final resulting vector **c** has no physical meaning except for its phase.

Alternatively, we propose to compute the mean phase vector **c** via geometric algebra, also known as Clifford algebra [22]. Geometric algebra defines vectors with a set of standard basis elements $\mathbf{e}_i$ where $i \in \{1, 2, 3, \ldots, n\}$, that is similar to quaternions, which are used to defined the n-dimensional Riesz filter and monogenic signal. From the set of standard basis it builds up higher order algebraic elements such as $\{1, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_{12}, \mathbf{e}_{13}, \mathbf{e}_{23}, \mathbf{e}_{123}\}$ to form the basis for multivectors. In fact, geometric algebra is more general and can be used as a framework to analyse multi-dimensional monogenic signals and even higher order Riesz transforms. In geometric algebra, two basic vector operations are defined: the inner and outer product. Un-

der this framework, these two products are general algebraic operations which, for example, need not result in a scalar, and is also not positive definite. Now a simple complex number $\mathbf{a} = a_0 + ja_1$ (as in equation 5.9) can be defined in the geometric algebra framework as a multivector in 2-D space:

$$\mathbf{a} = a_0 + a_1\mathbf{e}_1\mathbf{e}_2 \tag{5.13}$$

Since $(\mathbf{e}_1\mathbf{e}_2)^2 = -1$, this takes the role of j in the complex plane. The number of standard basis used to construct the algebraic basis for a vector is called its *grade*. For example, if a multivector consists of only bivector components, its grade will be 2; similarly if it has only trivector components the grade will be 3. For vectors (grade 1) $\mathbf{a}$ and $\mathbf{b}$, the geometric product is the sum of the inner and outer products:

$$\mathbf{ab} = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \wedge \mathbf{b} \tag{5.14}$$

In 3-D the outer product (second term) is analogous to the cross product for vectors. Since phase vectors have only two components - real (scalar) and imaginary ($\mathbf{e}_1\mathbf{e}_2$), we can define $\mathbf{c}$ as the geometric product of $\mathbf{a}$ and $\mathbf{b}$, so that:

$$\begin{aligned}
\mathbf{c} &= \mathbf{ab} = \mathbf{a}\cdot\mathbf{b} + \mathbf{a}\wedge\mathbf{b} && (5.15)\\
&= |\mathbf{a}|\,|\mathbf{b}|\,(\cos\theta_a\cos\theta_b - \sin\theta_a\sin\theta_b) && (5.16)\\
&\quad + \mathbf{e}_1\mathbf{e}_2\,|\mathbf{a}|\,|\mathbf{b}|\,(\sin\theta_a\cos\theta_b + \cos\theta_a\sin\theta_b) && (5.17)\\
&= |\mathbf{a}|\,|\mathbf{b}|\cos(\theta_a+\theta_b) + \mathbf{e}_1\mathbf{e}_2\,|\mathbf{a}|\,|\mathbf{b}|\sin(\theta_a+\theta_b) && (5.18)
\end{aligned}$$

which leads to $|\mathbf{c}| = |\mathbf{a}|\,|\mathbf{b}|$, and $\theta_c = \theta_a + \theta_b$. Although in this round θ_c is not equal to the mean phase $\frac{\theta_a+\theta_b}{2}$ directly, it can be obtained easily by de Moivre's theorem $\cos\theta + j\sin\theta = (\cos n\theta + j\sin n\theta)^{\frac{1}{n}}$. In this case, the mean phase of $\mathbf{c}$ is equal to the phase of its square root, or $\frac{\theta_c}{2} = \frac{\theta_a+\theta_b}{2} = \arg\left(\sqrt{\mathbf{c}}\right)$.

In our application to real images, $\mathbf{a}$ would be the complex image value, whose real and imaginary part are derived from a monogenic signal at a particular scale; and $\mathbf{b}$ would be the complex image value at another scale. We call it a complex image value, or monogenic signal value, because it is a complex representation of the image, from which the monogenic local energy, local phase and local orientation can be derived. The problem can be simplified further by combining all its Riesz components into one single imaginary part; hence both $\mathbf{a}$ and $\mathbf{b}$ are 1-D multivectors with a scalar component. We denote these by $\mathbf{m}_k$ at a particular scale k. For a given N number of scales and $\mathbf{m}_k$, $k \in \mathbb{Z}$ from 1 to N, our proposed method computes the exact mean local phase by

$$LP_{\text{mean}} = \arg\left(\sqrt[N]{\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \ldots, \mathbf{m}_N}\right) = \arg\left(\sqrt[N]{\prod_{k=1}^{N} \mathbf{m}_k}\right) \tag{5.19}$$

where the operator $\prod$ indicates a geometric product series. Note that this approach does not require each of the monogenic values $\mathbf{m}_k$ to have the same amplitude, nor is normalisation needed. Implementation-wise, this method simply multiplies the monogenic values of a new scale to the product of the previous scales and does not require extra memory to store normalised values. It is also more efficient in terms of the number of operations to be executed. Each geometric product consists of 4 multiplication operations, and there are $N-1$ geometric product operations. The total number of operations needed for this method is $4\,(N-1)$, which is less than half the number of operations for the normalisation approach. In addition, the final product $\prod_{k=1}^{N} \mathbf{m}_k$ preserves its physical meaning, as its amplitude is equal to the product of $\mathbf{m}_k$ for all k and its phase is equal to the sum of all phase of $\mathbf{m}_k$.

5.5 Bandpass Filters

The importance of the bandpass filter in our method is to control scale and for selection of the frequency band. There is always a trade-off between frequency band selectivity and spatial smoothness of the filter, for which a number of filters have been designed. In this section, we investigate a number of well-known filters for the application of feature detection in clinical MRI images. In monogenic signals, scale is controlled by applying a bandpass filter to the original function before using the Riesz transform. Generally, the bandpassed version of the function can be written as

$$I^B(\mathbf{x}) = B(\mathbf{x}) * I(\mathbf{x}) \tag{5.20}$$

and in the frequency domain as

$$I^B(\mathbf{u}) = B(\mathbf{u})\, I(\mathbf{u}) \tag{5.21}$$

For isotropic filtering, the frequency band selection should be the same in all directions. That is to say, in polar coordinates the bandpass function $B(|\mathbf{u}|, \theta_u)$ should not depend on the orientation θ_u at all. Hence the multi-dimensional filter function $B(\mathbf{u})$ can be reduced to a one-dimension function $B(|\mathbf{u}|)$ in the frequency domain. In this section, we simplify the writing to $B(u)$. Once the bandpass filter is defined in the frequency domain, its spatial counterpart can be derived by the n-dimensional inverse Fourier transform. In the derivation of the monogenic signal (equation 3.16), sometimes the Riesz filters and the bandpass filter are combined to form a filter set. These filters are in quadrature pairs, i.e. they are mutually orthogonal to each other. In order to make the analysis easier, the convention is to define $B(u)$ as an even filter, i.e. $B(-u) = B(u)$; and hence its quadrature counter-parts obtained by the Riesz filter will be odd, or $B_H(-u) = -B_H(u)$. Since the generalisation to n-dimensions is naturally dealt with by the Riesz transform and the multi-dimension Fourier transform, we may concentrate our analysis on the frequency band selection and wave form in one-dimension.

5.5.1 The Gabor Filter

The Gabor filter is one of the earliest wavelets and the most studied. In the spatial domain it is a multiplication of a Gaussian and a sinusoid. It has two - even and odd counterparts, as noted before. The even part reads:

$$b^{\mathrm{e}}_{\mathrm{gabor}}(x) = aG_{\sigma_x}(x)\cos(kx) \tag{5.22}$$

where $G_{\sigma_x}(x) = \exp\left(-\frac{x^2}{2\sigma_x^2}\right)$ is the Gaussian function with a spread defined by σ. The parameter k is the frequency of the sinusoidal, and a is a constant which normalises the filter. Multiplication in the spatial domain is convolution in the frequency domain. Since a sinusoidal in the frequency domain is a pair of shifted Dirac delta function, a Gabor filter in the frequency domain is a pair of shifted Gaussians. Again, it has two - an even and an odd part. The even part in frequency domain reads:

$$B^{\mathrm{e}}_{\mathrm{gabor}}(u) = A\left(G_{\sigma_u}(u-k) + G_{\sigma_u}(u+k)\right) \tag{5.23}$$

where $\sigma_u = \frac{1}{\sigma_x}$ is the spread in frequency domain, and A is a normalisation constant. The Gabor odd counterpart is:

$$b^{\mathrm{o}}_{\mathrm{gabor}}(x) = aG_{\sigma_x}(x)\sin(kx) \tag{5.24}$$

$$B^{\mathrm{o}}_{\mathrm{gabor}}(u) = A\left(G_{\sigma_u}(u-k) - G_{\sigma_u}(u+k)\right) \tag{5.25}$$

In complex notation, the even and odd Gabor filters can be unified by the complex exponent, $\exp(jkx) = \cos kx + j \cdot \sin kx$. Hence,

$$b_{\text{gabor}}(x) = aG_{\sigma_x}(x) \exp(jkx) = b^{\text{e}}_{\text{gabor}}(x) + jb^{\text{o}}_{\text{gabor}}(x) \tag{5.26}$$

$$B_{\text{gabor}}(u) = B^{\text{e}}_{\text{gabor}}(u) + jB^{\text{o}}_{\text{gabor}}(u) \tag{5.27}$$

Since by convention we have defined our real filter as an even function, its imaginary counterpart must be odd. This is true in both the spatial and frequency domain. These even and odd filters are connected by the Riesz transform (multi-dimensional Hilbert transform).

$$b^{\text{o}}_{\text{gabor}}(x) = h(x) * b^{\text{e}}_{\text{gabor}}(x) \tag{5.28}$$

$$B^{\text{o}}_{\text{gabor}}(u) = H(u) B^{\text{e}}_{\text{gabor}}(u) \tag{5.29}$$

where $h(x) = \frac{1}{\pi x}$ and $H(u) = j\frac{u}{|u|}$. Figure 5.5 demonstrates the Gabor filter.

The local phase of the filter is defined as $LP(x) = \arctan \frac{b_{\text{imag}}(x)}{b_{\text{real}}(x)}$. For the Gabor filter, it reads:

$$LP_{\text{gabor}}(x) = \arctan \frac{b^{\text{o}}_{\text{gabor}}(x)}{b^{\text{e}}_{\text{gabor}}(x)} = \arctan \frac{G_{\sigma_x}(x) \sin(kx)}{G_{\sigma_x}(x) \cos(kx)} = kx \tag{5.30}$$

while $kx \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. When kx is beyond this range, phase wrapping occurs as dis-

cussed in the previous section. This is shown in Figure 5.5 (e). The massive phase wrapping is due to the frequent oscillations in $b^{\mathrm{e}}_{\mathrm{gabor}}(x)$ and $b^{\mathrm{o}}_{\mathrm{gabor}}(x)$, or the intrinsic oscillation property of the sinusoidals, whose frequency is defined by k. Since its local phase is equal to kx, the higher frequency of the sinusoidal, the steeper the local phase curve, and hence more frequent phase wrapping as it exceeds the $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ range faster. This excessive phase wrapping would cause problems in local phase-based feature detection as it leads to artificial "ripples" near the detected features. The number of ripples can be reduced by lowering the frequency of the sinusoidals, however, this also means shifting the filter frequency band towards the lower frequencies, which violates our goal of free frequency selection. In fact, this is the case because the Gabor function is so narrowly banded in the frequency domain. As the bandpass filter shifts towards the higher frequencies, it becomes more like a short burst of noise. Indeed, the Gabor filter is very selective in frequency bandpassing, however, it sacrifices its usability along the way towards higher frequencies as it loses "transitional" low frequency components.

For multi-scale image analysis, a filter bank can be constructed (Figure 5.5-(f)). The spread parameter σ_u controls the width of the bandpass filter. In theory, if σ_u is small the filter will be very selective in frequency, but then the filter would be very wide in the spatial domain and eventually becomes the Fourier transform. On the other hand, if σ_u is large then frequency selectivity will be lost, σ_x will be small and there will be fewer oscillations captured in a smaller envelope. In order to give a sense of how an aggregate filter of the filter bank might look like, the sum of them is computed and plotted as a dotted line. Generally, it is desirable that the combined

filter covers the complete frequency spectrum evenly, so that no features at a certain scale are missed or over-stated.

5.5.2 The Log-Gabor Filter

The log-Gabor filter is defined in the frequency domain as the following:

$$B_{\mathrm{lg}}^{\mathrm{e}}(u) = A \exp\left(-\frac{\log^2\left(\frac{u}{k}\right)}{2\log^2(\sigma_u)}\right) \tag{5.31}$$

where k is the centre frequency (at which the power spectrum is at its maximum), and σ_u controls the spread of the frequency spectrum in a logarithmic fashion. Again, this is the definition of the even log-Gabor filter and it is real. Its odd (imaginary) counterpart is obtained via the Hilbert (Riesz) transform.

$$B_{\mathrm{lg}}^{\mathrm{o}}(u) = H(u)\, B_{\mathrm{lg}}^{\mathrm{e}}(u) \tag{5.32}$$

The respective convolution filter in the spatial domain is computed by the inverse Fourier transform (Figure 5.6). The first difference noted between log-Gabor and Gabor is that it is no longer a shifted bandpass filter. Instead, it is a log-scale Gaussian bandpass function, whose amplitude, despite its parameter k and σ_u, always builds up exponentially from zero frequency, reaches its peak at $u = k$, and then decays relatively slowly toward the high frequencies. Because of this continuity with more "transitional" frequency components between $u = 0$ and $u = k$, the number of oscillations occurring in the spatial convolution kernel is significantly reduced. In fact, the even log-Gabor filter $b_{\mathrm{lg}}^{\mathrm{e}}(x)$ looks more like a Mexican hat and does not

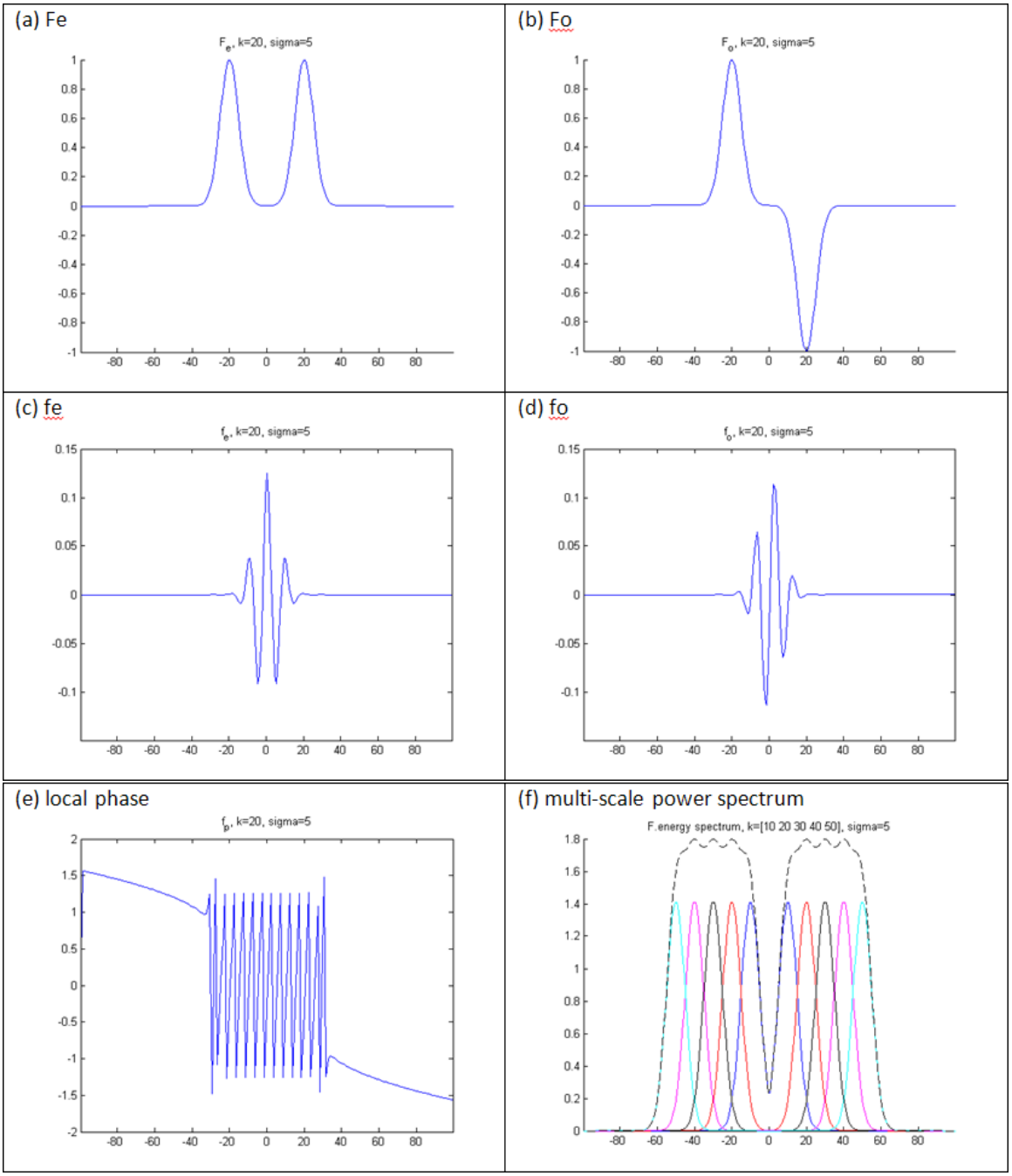

Figure 5.5: Gabor quadrature filter pair. (a) even part in frequency domain. (b) odd part in frequency domain. (c) even part in spatial domain. (d) odd part in spatial domain. (e) filter local phase. (f) multi-scale power spectrum.

have periodic oscillations at all, along with which the number of phase wrappings is reduced significantly (Figure 5.6-(e)).

The potential drawback of the log-Gabor filter is its long tail into high frequencies once it has reached the peak. This limits the frequency selectivity the filter can have. For example, suppose bandpassing the mid-frequency band is desirable with $k = 40$ and keeping $\sigma_u = 2$ unchanged, the resulting bandpass filter will have a long tail extending a long way into the high frequencies. In this case, at the outer bound of the frequency range when $u = 100$, $B^{\mathrm{e}}_{\mathrm{lg}}(u)$ is equal to 0.4174. That is 41.74% of the peak value, which implies a very sharp cut-off of frequencies. This would lead to large number of oscillations in the spatial convolution kernels as well as its local phase. This is called Gibbs' ringing. Even with $k = 10$ in Figure 5.6, it already starts to show some Gibbs' ringing in the local phase plot near both ends. In order to avoid heavy Gibbs' ringing artefacts, the range of choice for k is likely to be limited. Overall, the log-Gabor filter provides moderate localisation in spatial filtering and has reasonable local phase wrapping. However, its frequency selectivity is relatively poor compared to other filters due to its long tail of the log-Gaussian function. It generally serves well for some low-range frequency selections but does not fit the purpose of even and flexible bandpass processing.

5.5.3 The Gaussian-Derivative Filter

The Gaussian-derivative filter is defined as the derivative of a Gaussian function $G_{\sigma_u}(u) = \exp\left(-\frac{u^2}{2\sigma_u^2}\right)$. In the frequency domain the filter reads:

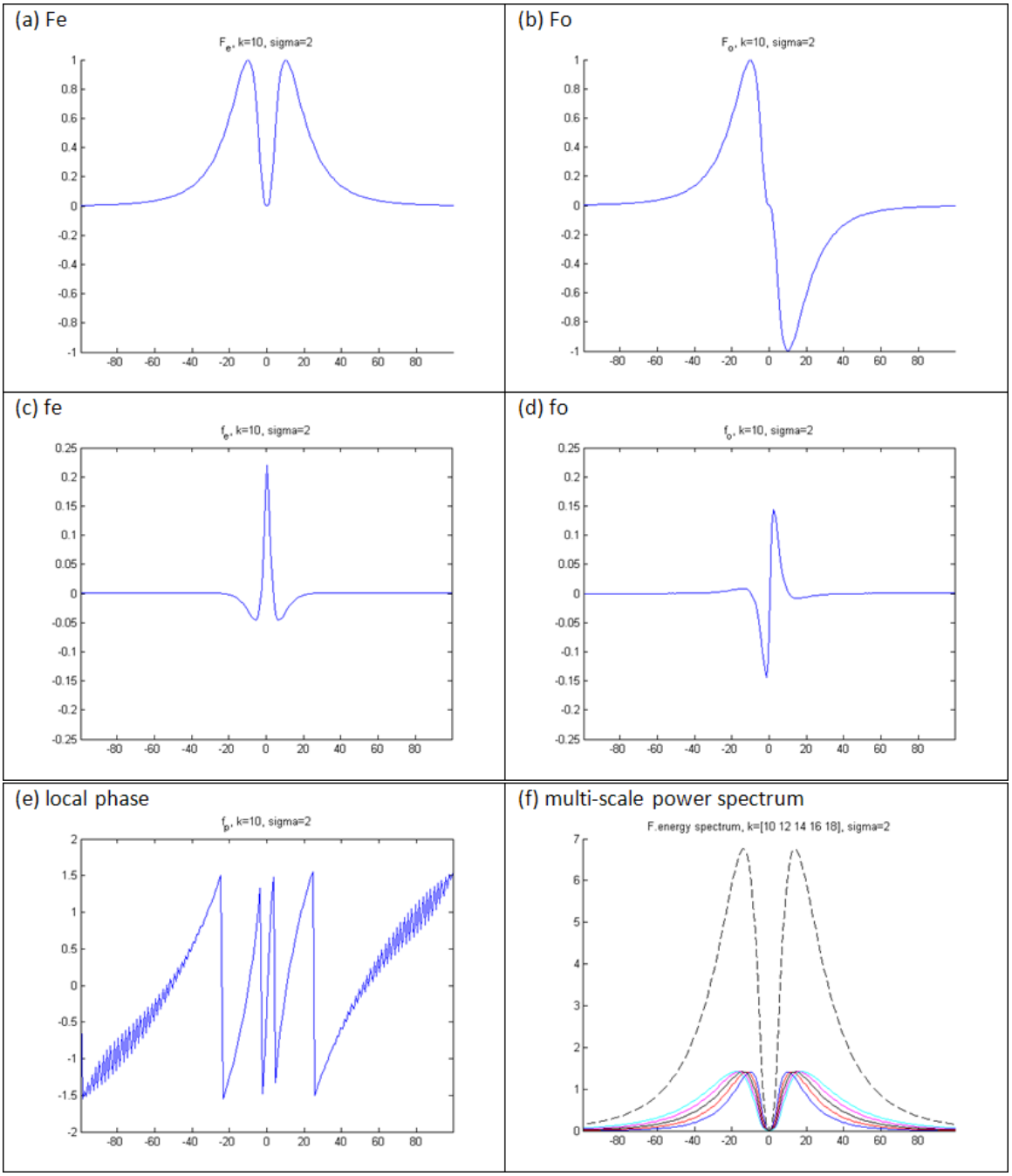

Figure 5.6: Log-Gabor quadrature filter pair. (a) even part in frequency domain. (b) odd part in frequency domain. (c) even part in spatial domain. (d) odd part in spatial domain. (e) filter local phase. (f) multi-scale power spectrum.

$$B^{\mathrm{e}}_{\mathrm{gd}}(u) = \frac{d}{du} G_{\sigma_u}(u) = Au \exp\left(-\frac{u^2}{2\sigma_u^2}\right) = Au \exp\left(-\frac{u^2}{2k^2}\right) \tag{5.33}$$

Here the denotation of parameter σ_u is replaced by k. Note that σ_u is the spread of the original Gaussian function G_{σ_u}, and it defines the point at which the gradient of G_{σ_u} is the greatest. Therefore, when the derivative is used as a filter, σ_u also defines the centre frequency, or $k = \sigma_u$. Figure 5.7 shows an instance of this filter.

The first observation is that the Gaussian-derivative filter has many desirable properties. For the purpose of band selection, its frequency band is well bounded and does not have a long tail. Its spatial convolution kernel is sharp, well localised, and does not have redundant oscillations. The odd kernel shown in Figure 5.7-(d) is a very neat differentiator. The amount of phase wrapping is kept to a minimum - only two instances. Outside the localised range, phase is equal to zero and stable. It indicates that the amplitude of the odd kernel is also equal to zero outside the centre range, which confirms that the Gaussian-derivative filter provides very good localisation. The filter bank constructed by a set of k values evenly spread over the entire frequency spectrum. Finally, the only parameter that needs to be set is the centre frequency k, which is very easy to use, unlike the difference-of-Gaussian filter (to be discussed in next subsection) that requires two parameters to define and is not obvious as to where the centre frequency is.

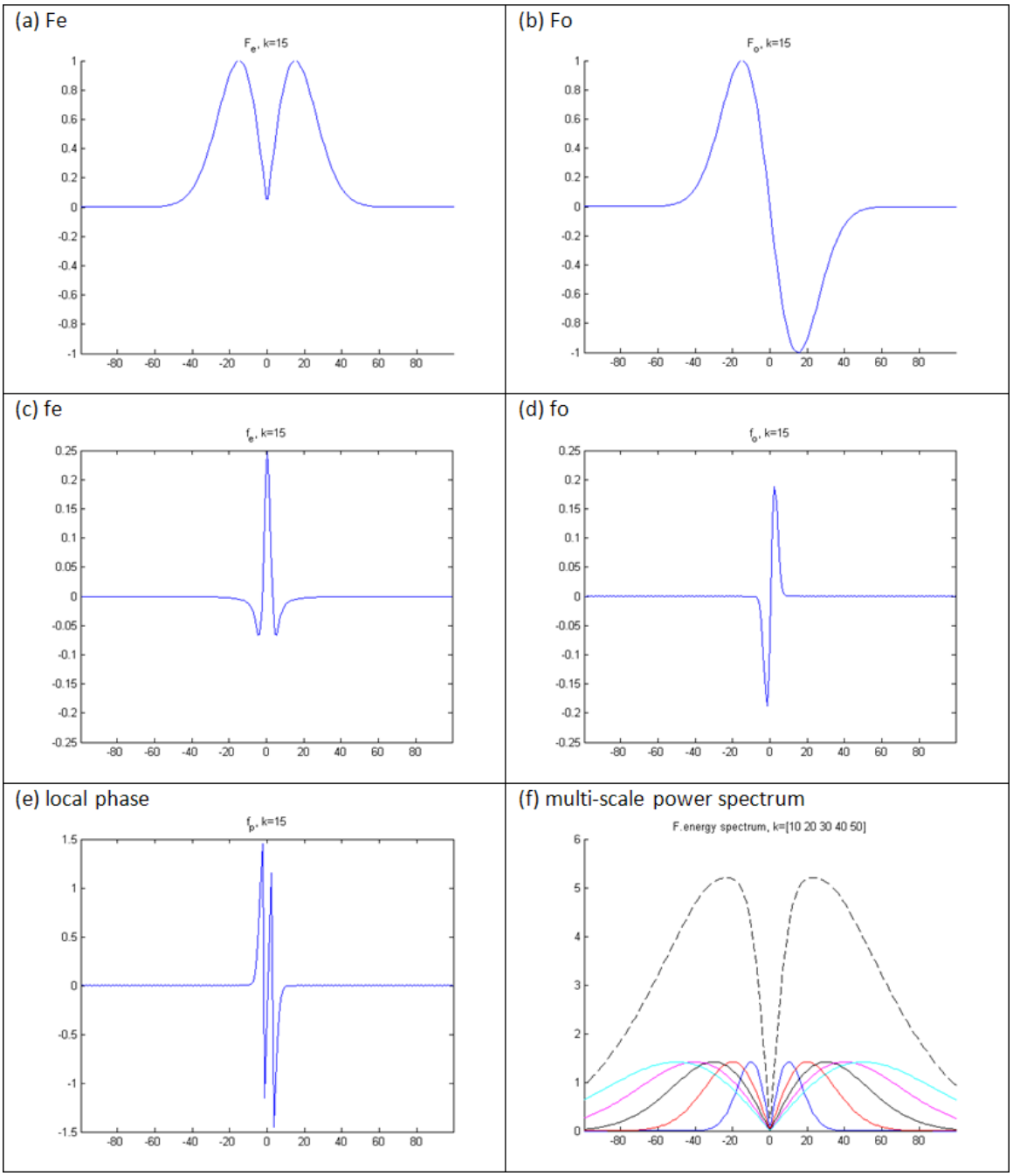

Figure 5.7: Gaussian-derivative quadrature filter pair. (a) even part in frequency domain. (b) odd part in frequency domain. (c) even part in spatial domain. (d) odd part in spatial domain. (e) filter local phase. (f) multi-scale power spectrum.

5.5.4 The Difference-of-Gaussian Filter

The difference-of-Gaussian (DoG) filter is defined as the difference of two Gaussians with different spreads,

$$B^{\mathrm{e}}_{\mathrm{DoG}}(u) = G_{\sigma_{u1}}(u) - G_{\sigma_{u2}}(u) = \exp\left(-\frac{u^2}{2\sigma^2_{u1}}\right) - \exp\left(-\frac{u^2}{2\sigma^2_{u2}}\right) \tag{5.34}$$

where $\sigma_{u1} > \sigma_{u2}$. The DoG has a lot of similarity with the Gaussian-derivative filter (Figure 5.8). It guarantees zero DC level (zero frequency component). It is relatively narrowly band, therefore good for band selection. Its convolution kernel is also sharp and well localised. It does not have much phase wrapping in the central region, however, a lot is seen outside that region. These may be caused by numerical errors especially when the amplitudes there are small. Another issue is that it is not easy to parameterise. Given a desired centre frequency k, it is not obvious as to how to set parameters σ_{u1} and σ_{u2} to achieve a particular bandpass filter. Firstly, we take the derivative of the filter function and set it to zero:

$$\frac{dB^{\mathrm{e}}_{\mathrm{DoG}}}{du}(u) = -\frac{u}{\sigma^2_{u1}}\exp\left(-\frac{u^2}{2\sigma^2_{u1}}\right) + \frac{u}{\sigma^2_{u2}}\exp\left(-\frac{u^2}{2\sigma^2_{u2}}\right) = 0 \tag{5.35}$$

Rearranging, we get:

$$\frac{\sigma^2_{u1}}{\sigma^2_{u2}} = \exp\left(\frac{u^2}{2\sigma^2_{u2}} - \frac{u^2}{2\sigma^2_{u1}}\right) = \exp\left(\frac{u^2}{2}\left(\frac{1}{\sigma^2_{u2}} - \frac{1}{\sigma^2_{u1}}\right)\right) \tag{5.36}$$

Taking natural logs on both sides,

$$\ln\left(\frac{\sigma_{u1}^2}{\sigma_{u2}^2}\right) = \frac{u^2}{2}\left(\frac{1}{\sigma_{u2}^2} - \frac{1}{\sigma_{u1}^2}\right) = \frac{u^2}{2}\cdot\frac{\sigma_{u1}^2 - \sigma_{u2}^2}{\sigma_{u1}^2\sigma_{u2}^2} = \frac{u^2}{2}\cdot\frac{\sigma_{u1}^2/\sigma_{u2}^2 - 1}{\sigma_{u1}^2} \tag{5.37}$$

Let the desired centre frequency be $u = k$, and the ratio $\frac{\sigma_{u1}^2}{\sigma_{u2}^2} = 2$. We can then solve for the corresponding σ_{u1}:

$$\sigma_{u1} = \frac{k}{\sqrt{2\ln 2}} \tag{5.38}$$

and $\sigma_{u2} = \frac{\sigma_{u1}}{\sqrt{2}} = \frac{k}{2\sqrt{\ln 2}}$.

5.5.5 The Mellor-Brady Filter

The Mellor-Brady filter [34] is defined rather differently. It is not an analytic filter; rather, it exists only in discrete form. However, its formulation has drawn some similarities to the Hilbert transform and yields interesting properties. Therefore we include it for analysis. First of all, the Mellor-Brady filter is defined in the spatial domain as the following:

$$b_{\mathrm{MB}}^{\mathrm{e}}(x) = f_1(x) - f_2(x) = \frac{A}{x^{\alpha+\beta(k+1)}} - \frac{B}{x^{\alpha+\beta k}} \tag{5.39}$$

where k is a scale parameter and can be any positive integer, $k \in \mathbb{Z}^+$. The parameters α and β are constants that control the spread of the kernel and their values are set to be $\alpha = 2$ and $\beta = 0.25$ by default. The coefficients A and B are normalisation factors which make the area under the curve f_1 and f_2 exactly equal to 1. However,

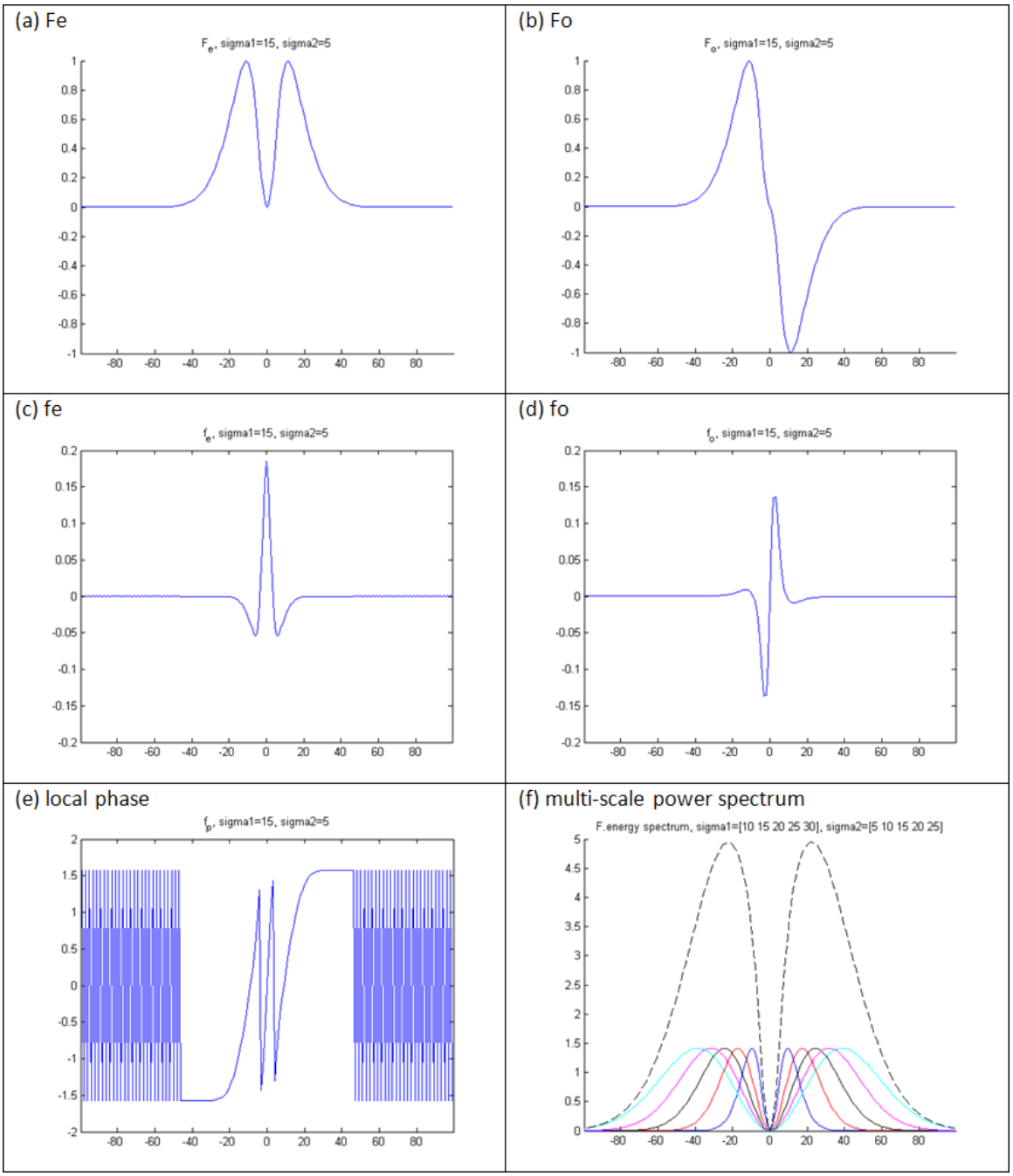

Figure 5.8: Difference-of-Gaussian quadrature filter pair. (a) even part in frequency domain. (b) odd part in frequency domain. (c) even part in spatial domain. (d) odd part in spatial domain. (e) filter local phase. (f) multi-scale power spectrum.

these two coefficients cannot be analytically derived, simply because the integrals of f_1 and f_2 are not finite, or $\int_0^\infty f_1 dx = \infty$. Moreover, there is a singularity at $x = 0$, where $b^{\mathrm{e}}_{\mathrm{MB}} \to \infty$. However, this filter can still be valid in the discrete world with a slight modification. One option is to force the value at a singularity to be equal to twice the nearest adjacent value,

$$b^{\mathrm{e}}_{\mathrm{MB}}(0) = 2b^{\mathrm{e}}_{\mathrm{MB}}(0_+) \tag{5.40}$$

where $x = 0_+$ represents the next grid point in discrete coordinates. Suppose x is a set of monotonically increasing discrete points $[x_0, x_1, x_2, \ldots, x_N]$ and $x_0 = 0$. Then $b^{\mathrm{e}}_{\mathrm{MB}}$ becomes $[b_0, b_1, b_2, \ldots, b_N]$. The sum of $b^{\mathrm{e}}_{\mathrm{MB}}$, or $\sum_i b_i$ is finite as long as $b_0 \neq \infty$. Here we artificially let $b_0 = 2b_1$. Consequently, a finite A and B can be calculated and hence $b^{\mathrm{e}}_{\mathrm{MB}}$ is ensured to sum to zero, which fulfills the zero DC level criterion for bandpass filtering. A discrete Fourier transform is then performed to compute its frequency spectrum, and its odd counterpart is obtained by a subsequent Riesz transform. These are plotted in Figure 5.9.

The design of the Mellor-Brady filter is intriguing because in its formulation the terms f_1 and f_2 are a generalised version of the Hilbert filter $\frac{1}{\pi x}$, which is also in the form $A/x^{\alpha+\beta}$. This filter uses the exponent $\alpha + \beta$ to control scale and can be viewed as the difference-of-Hilbert filter. Recall that the Hilbert filter behaves like a differentiator that shifts a function's underlying sinusoidals by 90 degrees (that is, it turns every sine wave into a cosine, and vice versa). In the spatial domain, the Mellor-Brady convolution kernel seems to inherit some of these properties - extremely

sharp and localised, and no excessive oscillations outside the localised region. The ringing in the local phase are likely due to the Gibbs' phenomenon as the frequency spectrum is sharply cut-off. This cannot be eliminated because it is a consequence of the artificial hold-out at $x = 0$. Since the peak value $b^{\mathrm{e}}_{\mathrm{MB}}(0)$ has been artificially held low, it induces negative high frequency components to regulate the irregularity. If the value at $b^{\mathrm{e}}_{\mathrm{MB}}(0)$ is increased, e.g. equal to 4 times or 10 times of the value $b^{\mathrm{e}}_{\mathrm{MB}}(0_+)$, the negative amplitudes at high frequencies will decrease.

5.6 Conclusions

In this chapter we propose an alternative way to measure phase congruency, namely computing phase variance over scales. At a location where phase variance is small, it means that the local phase there does not vary much at different scales, which is another way of saying phases are congruent at that location. Conversely, when phase variance is large, it means phases are not consistent and hence indicates a location with no significant feature. The main difference between the proposed phase variance measure and the original phase congruency is that the original method measures phases for each individual frequency. Conceptually, this is equivalent to assuming at each scale level the signal is filtered by an infinitesimally narrow bandpass filter, which is in practice difficult to implement, and the method is prone to noise. In contrast, our approach uses analytic wavelets whose frequency spectra are smooth, and it is proven to be much more practical to apply to real clinical data than the original phase congruency method. We have analysed five well-known wavelet filters and concluded

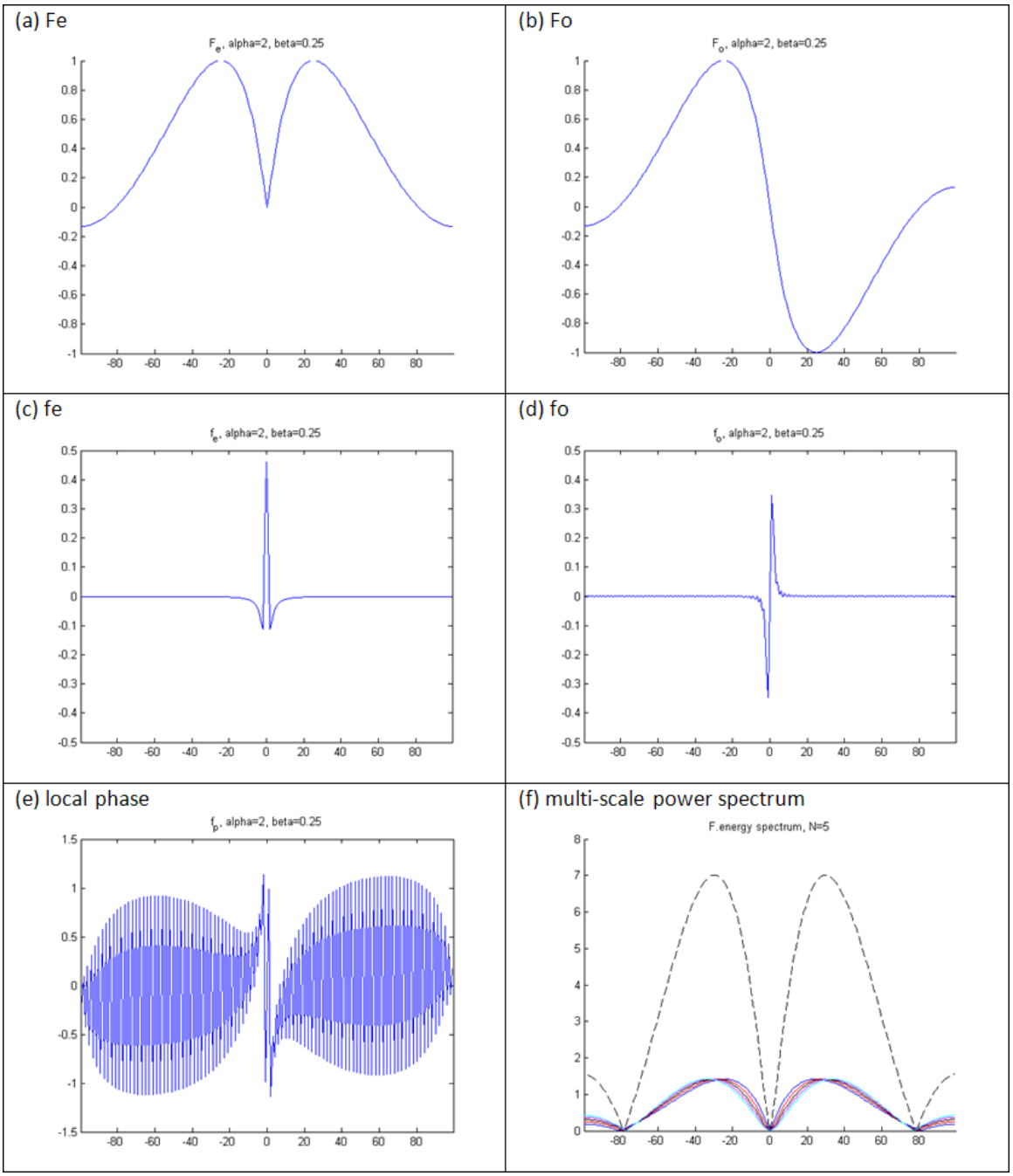

Figure 5.9: Mellor-Brady quadrature filter pair. (a) even part in frequency domain. (b) odd part in frequency domain. (c) even part in spatial domain. (d) odd part in spatial domain. (e) filter local phase. (f) multi-scale power spectrum.

that the Gaussian-derivative filter may be more suitable for our application. We also offer a neat and accurate way to compute the mean phase using geometric algebra, which completely avoids the phase wrapping problem that usually goes along with inverse trigonometric functions. In the next chapter, we review the processing steps on clinical data in detail and reveal the issues that arise along the way.

Chapter 6

Processing of Clinical Data

The project that formed the background to this thesis is a collaboration with the MRI centre of the Oxford Radcliffe Hospital Trust (ORHT), where images of the patients were taken. Patients with endometriosis symptoms such as hematochezia and dyspareunia were first received and identified by gynaecologists at the Oxford Radcliffe Hospital. They were then sent to the MRI centre, where abdominal MRI scans were taken as part of the diagnosis. Images of a total of 15 patients with confirmed bladder endometriosis and/or endometriomas were provided by the MRI centre. These images were taken between 2008 and 2011. There is no international common protocol for imaging abdominal endometriosis. The existing imaging protocol was mainly developed in-house by the MRI centre, and it is the subject of regular revision and modification on a case by case basis. As a result, the setup for each individual set of images is not exactly the same. Table 6.1 lists all the scans that were available for this study. The patient data sets are anonymised (including the actual date when the images were acquired) and numbered by the order in which they were received from

the MRI centre. Usually, five to six scans are taken of a patient. The most common scans are those in the axial and sagittal direction. For example, the T_2-weighted, T_1-weighted, T_1-weighted fat saturated axial scan and the T_2-weighted sagittal scan are available for all patient sets. The other common ones are T_1-weighted fat saturated sagittal scan and the T_2-weighted short axis scan, which refers to the orientation of the uterus (short axis opposed to the long axis along the patient's body). Coronal scans are not commonly taken. There is only one patient set with a T_2-weighted coronal scan. The dimensions of the images also vary. The in-plane dimensions of T_2- and T_1-weighted images are usually 512 by 512 pixels, while the T_1-fat-saturated images are, apart from a couple of exceptions, often 256 by 256. The gap spacing between slices in the sagittal images is smaller than in the axial images, although axial images have higher in-plane resolution. Table 6.2 shows the image dimension (the number of voxels in 3-D), voxel dimension (size of each voxel in 3-D) and gap spacing between slices in the T_2-weighted axial images and T_1-fat-saturated sagittal images. Those figures illustrate just how different the scans can be. It may be noted that this is a challenge that rarely occurs in medical image analysis, and makes the analysis undertaken in this thesis even more complex.

In the following sections, we present and discuss results of the three main steps in processing the real clinical images, namely: image pre-processing (including registration and interpolation of common region); multi-view fusion; and finally bladder segmentation.

	T_2-weighted				T_1-weighted		T_1-FatSat	
	Axial	Sagittal	Coronal	S-Axis	Axial	Sagittal	Axial	Sagittal
P01	Y	Y	Y		Y		Y	
P02	Y	Y		Y	Y		Y	Y
P03	Y	Y		Y	Y		Y	Y
P04	Y	Y		Y	Y		Y	Y
P05	Y	Y		Y	Y		Y	Y
P06	Y	Y		Y	Y		Y	Y
P07	Y	Y		Y	Y		Y	Y
P08	Y	Y		Y	Y		Y	Y
P09	Y	Y		Y	Y		Y	Y
P10	Y	Y		Y	Y		Y	Y
P11	Y	Y		Y	Y	Y	Y	Y
P12	Y	Y		Y	Y	Y	Y	
P13	Y	Y		Y	Y		Y	Y
P14	Y	Y		Y	Y		Y	
P15	Y	Y		Y	Y		Y	Y

Table 6.1: Table of available scans by patient number. There are total of 15 individual cases.

	T_2Axial			T_1-FatSat Sagittal		
	Image Dim	Voxel Dim	Gap Spc	Image Dim	Voxel Dim	Gap Spc
P01	512x512x20	.55 x .55 x 6	1.5	N/A	N/A	N/A
P02	512x512x26	.47 x .47 x 6	1.5	256x256x20	1.25x1.25x5	1.5
P03	512x512x20	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5
P04	512x512x22	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5
P05	512x512x22	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5
P06	512x512x22	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5
P07	512x512x20	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5
P08	512x512x22	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5
P09	512x512x22	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5
P10	512x512x26	.47 x .47 x 5	1.5	512x512x22	.51 x .51 x 5	0.5
P11	512x512x24	.43 x .43 x 5	1.0	256x256x72	1.25x1.25x4	- 2.0
P12	512x512x24	.47 x .47 x 5	1.5	N/A	N/A	N/A
P13	512x512x22	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5
P14	512x512x23	.47 x .47 x 6	1.5	N/A	N/A	N/A
P15	512x512x22	.47 x .47 x 5	1.5	256x256x22	1.02x1.02x5	0.5

Table 6.2: Table of image dimension, voxel dimension and gap spacing in T_2-weighted axial images and T_1-fat-saturated sagittal images. All figures are in mm.

6.1 Image Pre-Processing

As discussed earlier in the thesis, we first extend the voxels in the through-plane direction to fill the gaps between slices. For example, instead of the actual voxel size of 0.47 by 0.47 by 5 mm (Table 6.2) we assume it to be 0.47 by 0.47 by 6.5 mm (voxel length plus gap spacing). This is done by simply changing the grid spacing for the image when registration is performed. For most of the images, the patient remained stationary during the entire scan and no major movement was observed between different scans of the same patient. Since the main purpose of the registration process is to align multi-view images and prepare them for fusion, we considered that rigid registration would be sufficient. Despite the type of the scan (such as T_2-weighted or T_1-weighted), most images have high contrast between the patient's skin and air. Therefore, aligning the patient's skin and air boundary would have the biggest impact on the similarity between the two images. This makes the registration of these images straightforward. As long as the patient remained stationary during the scans and there were no large movements on her skin, and this serves for our purpose of aligning the patient's body. In most cases this method works well. Figure 6.1 and 6.2 shows the overlay of the T_2-weighted axial and sagittal images of patient P01 before and after registration, respectively. The overlay image is produced in such a way that its odd pixels display the intensity from one source image and its even pixels display the intensity from the other. When the images are unregistered, substantial discrepancies are clearly seen along the boundary of the body. However, when the body's boundary is aligned (after registration) even the organs inside the body are

observed to be aligned reasonably well. This is one of the ideal cases when movement inside the body is small even though it is subject to breathing motions. Fortunately, most of the registration cases between the same type of image (that is, for example, T_2-weighted or T_1-weighted) work well without additional adjustment. Occasionally, registration between images of two different types needs to be attended to manually by adjusting their initial positions such that they are close to the desired registration state. This happened for example in registering the T_1-weighted fat saturated sagittal image to the T_2-weighted axial image for patients P12 and P14. This is because the fat-saturated image and the target image, which is the T_2-weighted axial image in our case, often have "opposite" intensity appearances for the same tissue. For example, the intensity in the layer of fat is suppressed in the fat-saturated image, so appears dark; while it appears bright in the T_2-weighted and T_1-weighted images; bland fluid in the bladder has the highest intensity in the T_2-weighted image but it is completely dark in both T_1-weighted and T_1-fat-saturated image (Figure 6.3). These differences in tissue appearance prompt us to use normalised mutual information [46], which is designed to measure the similarity between images of different modalities for registration. Manual adjustment of their initial positions also prevents the registration algorithm from being trapped in a local minimum. The result of a pair of registered images of different types is shown in Figure 6.3.

All the images are registered to a single reference image, which is chosen to be the T_2-weighted axial scan. They are then truncated to a field that corresponds to the region that they have in common. Images are also resampled to an isotropic uniform grid of size 0.5 by 0.5 by 0.5 mm. Consequently, all the image sets have the same

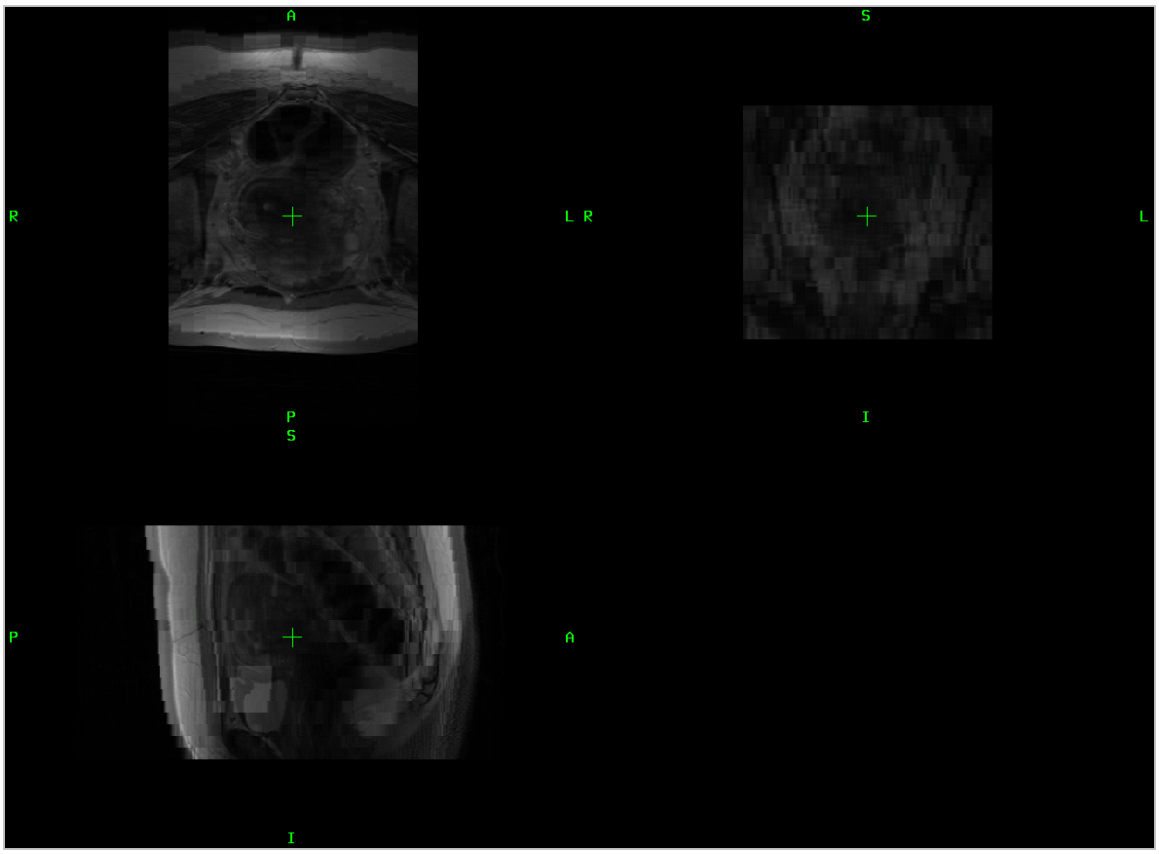

Figure 6.1: View of the overlay of unregistered T_2-weighted axial and sagittal scan of patient P01. Mis-alignment of the two unregistered images is illustrated as overlapping shadows.

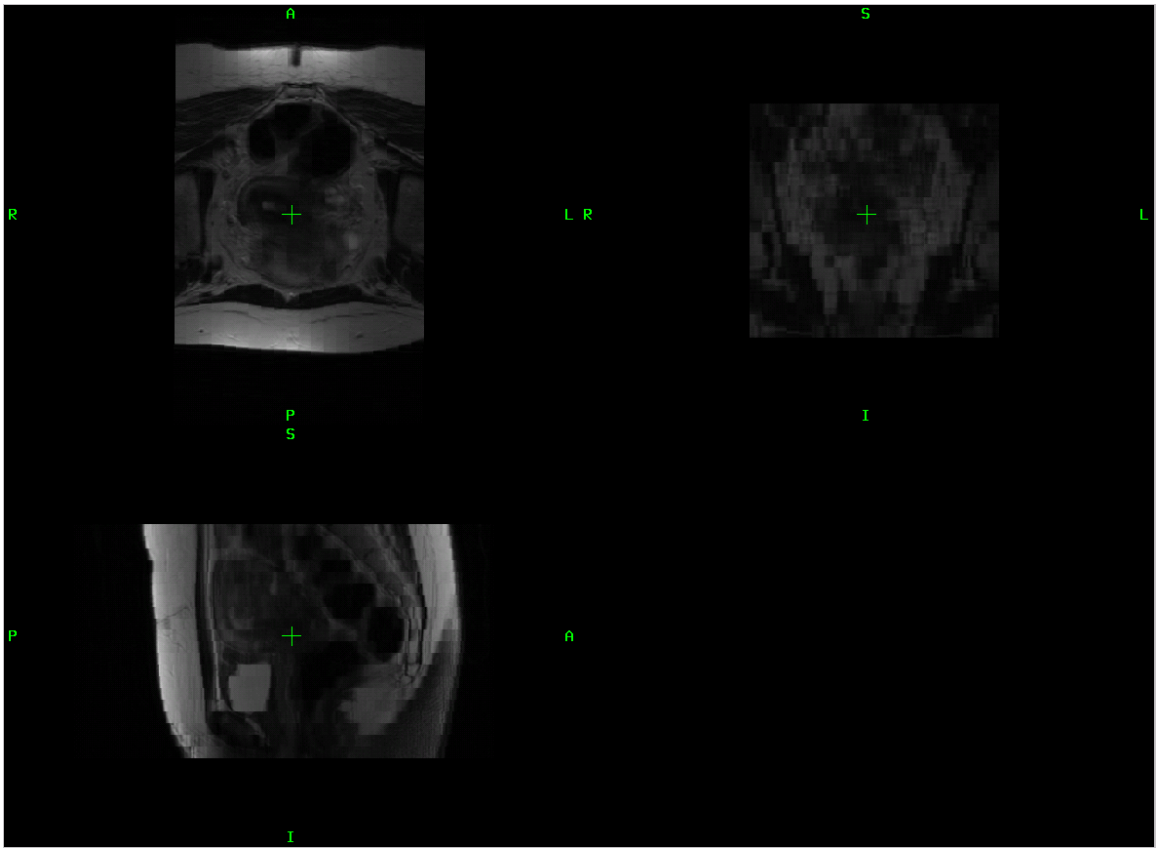

Figure 6.2: The same overlay view of registered T_2-weighted axial and sagittal scan of patient P01. As compared to the unregistered overlay in Figure 6.1, overlapping shadows are not seen in this registered view, which indicates that the images are properly aligned.

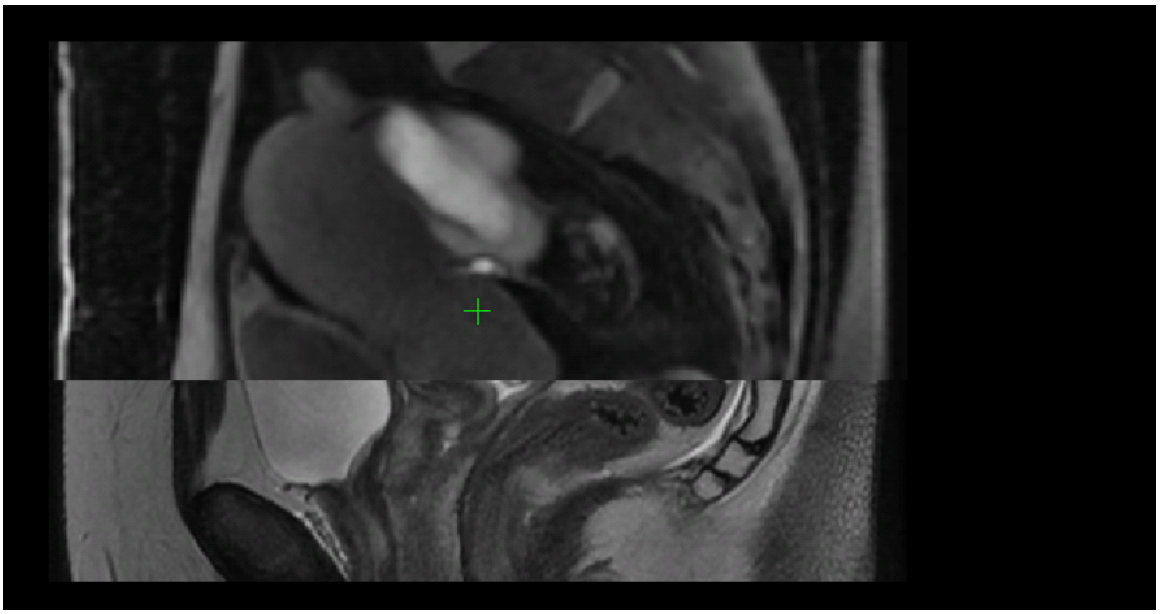

Figure 6.3: Shutter view of the registered T_2-weighted and T_1-weighted fat saturated image. This figure illustrates how good the registration is even for structures appearing in different intensities. Top shutter: T_1-weighted fat saturated sagittal image of patient P12, where fat and fluid (bladder) is dark. Bottom shutter: T_2-weighted sagittal image of patient P12, where fat and fluid (bladder) is bright.

image dimension and voxel dimension. The interpolation between slices also reduces the staircase effect between thick slices. Figure 6.4 illustrates the effects of these pre-processing steps. At the end of these processing steps the images are passed on for fusion, which is discussed in the following section.

6.2 Multi-view Fusion

The fusion method that we proposed in Chapter 4 is designed to fuse three component images that are mutually perpendicular - namely, the axial, sagittal and coronal views. Recall that we presented four fusion rules that combine multiple frequency coefficients in the same direction. In this section, we first investigate the data set of patient P01, which contains the axial, sagittal and coronal T_2-weighted scans. The main difficulty

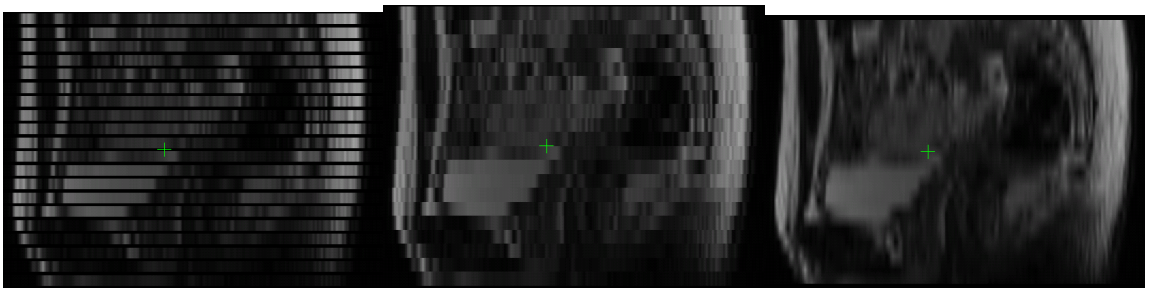

Figure 6.4: Pre-processing of the T_2-weighted axial image of patient P03. They are shown in a sagittal perspective. Left: original axial slices with gaps. Middle: extending voxels into the gaps (no interpolation). Right: linear interpolation between slices.

of assessing the fusion method on real clinical data such as these is that there is no longer any ground truth/gold standard of what the result should be. Therefore, there is no explicit benchmark that we can compare our results to. Instead, we superimpose plots of all the images available to us, namely, plots of the original axial, sagittal and coronal images, as well as the images generated by the proposed fusion methods with different fusion rules (the mean, max, adapt and min rule). We then assess the difference between them so that we can make a sound judgement on which fusion rule may be a better representation of the "true" underlying data.

6.2.1 Comparison of Fusion Rules

Since we are primarily interested in the bladder, we extract a sub-region of the image and focus on that part. Note that this reduces the computational complexity of our method (so it yields results much more quickly) since the size of the working image is smaller. This also helps validation of the method as the bladder has a relatively simple and predictable structure - high intensities inside and low intensities in the surrounding bladder muscles. To begin with a simple case, we select a line profile of

different bladder images, which can then be plotted in one graph for easy comparison. Figure 6.5 shows the bladder sub-region of the fused image of patient P01 using the mean rule. A line is chosen to be the crossing of the mid-axial and mid-sagittal slice and it is shown as a white line in Figure 6.5. Since it is a line going in the anterior-posterior direction (i.e. along the y-axis), it appears only as a small white dot in the coronal view.

Intensities along this line are plotted in Figure 6.6 and 6.7. This line passes through the bladder, therefore it is expected to have high intensity in the middle of the line (inside the bladder) and low intensity for the surrounding muscle. The first figure shows that this is the case except for the line profile in the coronal scan. This is because this line is in the through-plane direction of the coronal images, where high frequency features are not expected due to the partial volume effect. Apart from that the line profiles from the axial and sagittal image both show the expected features. The two plots are also reasonably well aligned, which indicates the registration of the axial and sagittal images is working well and that there was not much movement between the scans. As a reference, the line profile from the fused image using the mean rule is also plotted on this figure. Apart from the plot of the coronal scan, the line profile of the mean rule fused image sits between those of the axial and sagittal images. This is exactly what is expected, because the mean rule is linear and taking the mean in the frequency domain is the same as taking the mean in the spatial domain. However, the most promising part is that the line of the fused image does not seem to be affected by the line of the coronal image. This means that the intensity variation of an image in its through-plane direction is completely discarded, and the

intensity variation of the fused image in one particular direction is a combination of those in the component images that have high frequency features in that direction.

With the line profile of mean rule fusion as a reference, which is plotted in both Figure 6.6 and 6.7, the second figure compares the line profiles by different fusion rules. Again the mean rule line profile sits in the middle of other profiles. The profile of the min rule has slower response to edges and less details, therefore it produces a less desirable response. The max rule is the opposite to the min rule and generally has a large response at edges, and sometimes it even overshoots. The high intensity variation also makes it appear noisier. The proposed adaptive rule is a combination of the max and mean rule as it switches rules depending on its frequency. In this instance, the line profiles of the max rule and adaptive rule are almost identical (the green line is completely overlaid on top of the blue line in Figure 6.7). This is because real clinical images tend to be smooth and to have fewer sharp edges; and our pre-defined frequency threshold parameter α is high. In fact, during the pre-processing steps such as the interpolation after registration and resampling of grid points, the images are blurred/smoothed to a certain extent. This reduces the magnitudes of the high frequency coefficients, which leads to small differences between what rules to use for such frequencies. Certainly, the adaptive rule for fusion can be fine-tuned by the frequency threshold parameter α, but the exact value for α would depend on the images we are working on. After all, the adaptive rule would result an intermediate response between the mean rule and max rule.

The choice of fusion rule can also differ based on the post-processing steps. For example, if the fused image is used for segmentation that is edge based, a fusion

method that better preserves boundaries, or an adaptive rule with parameter $\alpha \to 1$ may be desirable. In contrast, if the image is fused for use by a radiologist for further examination, a noise suppressed version (or $\alpha \to 0$) would be desirable because the human visual system is already a very good edge detector and the human brain is excellent in object recognition. In this project, fusion serves both purposes. At this stage, we conclude that the mean rule and adaptive rule fusion will be subject to further investigation in the following sections under the context of segmentation.

6.2.2 Fusing Short-Axis Image

Ideally, the fusion consists of at least three component images whose orientations are mutually perpendicular, so that high frequency features are recorded in one of the images or another. However, in practice, this is not always the case, as is shown in Table 6.1. In that Table, only patient P01 has three perpendicular views of the T_2-weighted scan. Instead, axial, sagittal and short-axis images were recorded for other patients. The short axis is referred to the orientational axis of the uterus, as opposed to the long axis of the patient's whole body. The purpose of the short axis image is for the radiologist to better examine the patient's uterus. The radiologist usually picks a short axis that is roughly in line with the orientation of the uterus from the sagittal view and then performs a short-axis scan in that direction. It follows that the short-axis image is always a rotation about the x-axis and provides a view between the axial and coronal images. Depending on the orientation of the patient's uterus during the scan, the short-axis view could be either closer to the axial, or the

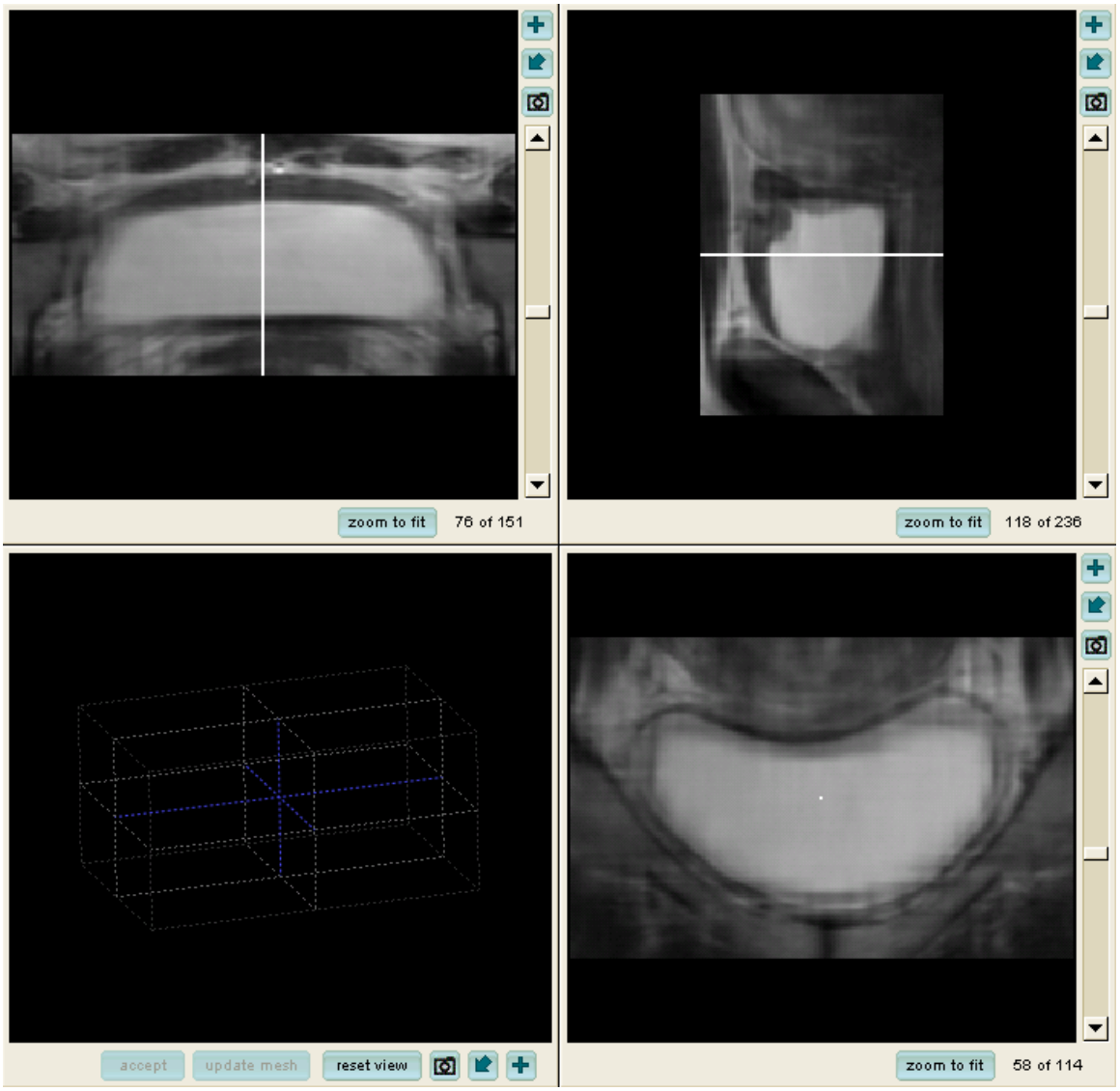

Figure 6.5: A chosen line profile of the bladder image of patient P01. The position of the chosen line is shown as a white line in the mid-axial and mid-sagittal slice, and as a white dot in the coronal view (in the anterior-posterior direction). The intensity profile of this line is plot in Figure 6.6.

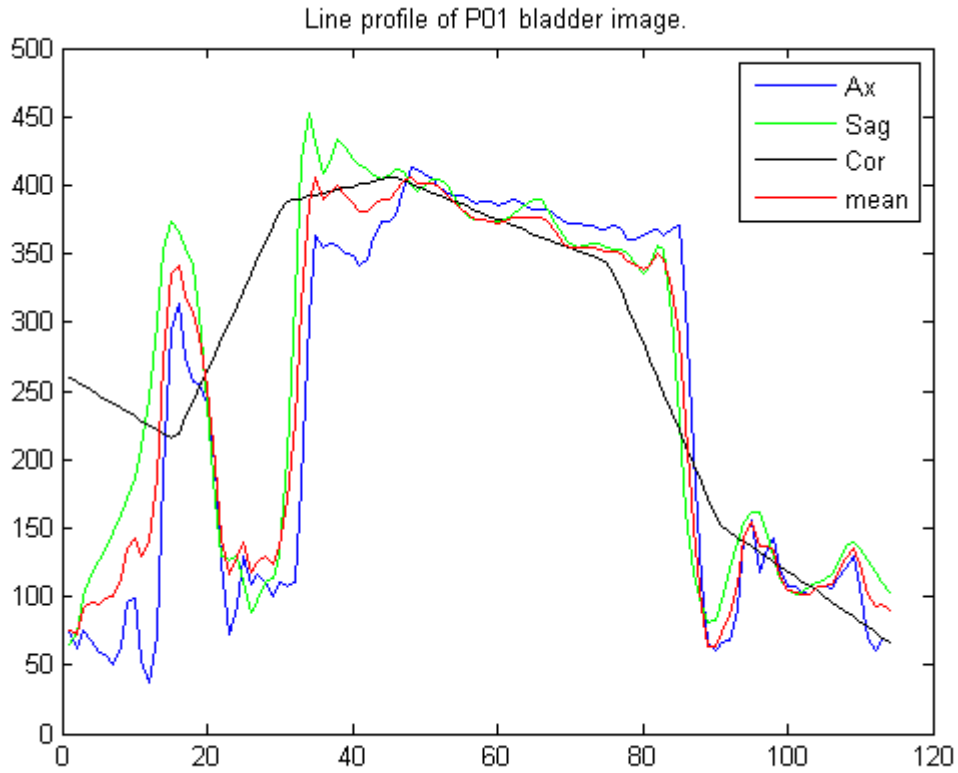

Figure 6.6: Intensity profiles of the line illustrated in Figure 6.5 in the axial (blue), sagittal (green), coronal (black) and the fused image using the mean rule (red). Since the line is going in the anterior-posterior direction (the low sampling direction in the coronal image), the coronal intensity profile (black) is very rough. Detail of the image along this line is given by the axial and sagittal image, and the fused image (red) reconstructs the features.

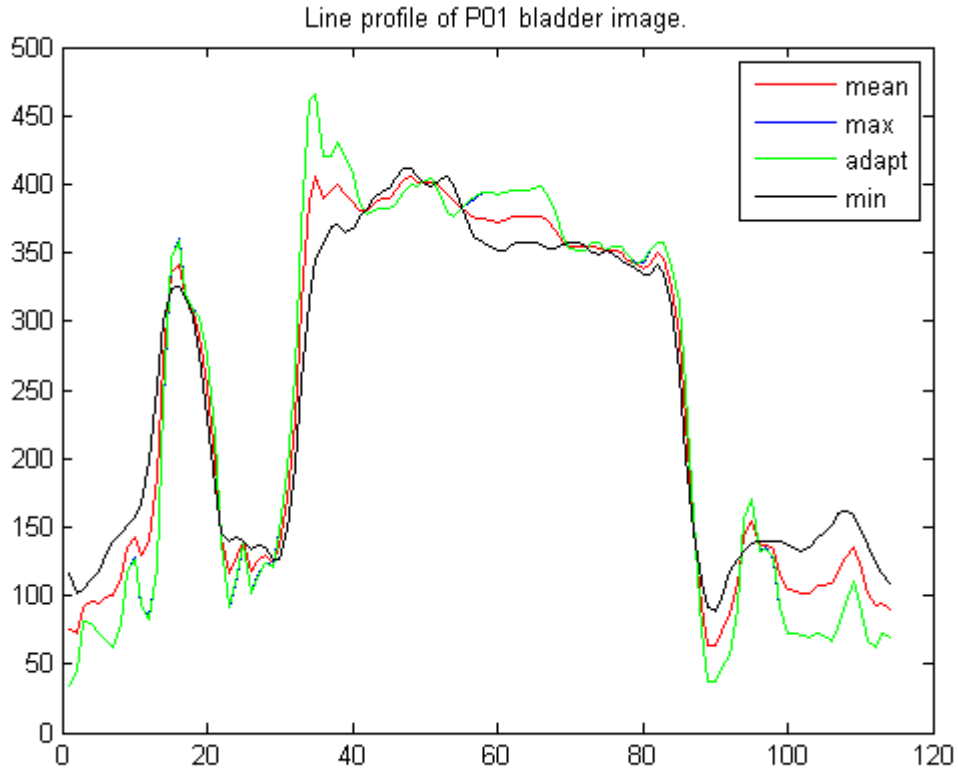

Figure 6.7: Intensity profiles of the line illustrated in Figure 6.5 in the fused bladder images using the mean rule (red), max rule (blue), adaptive rule (green) and min rule (black), respectively. Take the mean rule line profile as a benchmark, the min rule reconstruction is smoother because of its suppression on all frequencies. The max/adaptive reconstruction has sharper response at edges but also noisier.

coronal view. Since the uterus has an elongated pear shape and the MR images have high resolution in-plane, short-axis scanning is the best way to examine the uterus muscle as images are taken slice by slice along the elongated direction. Because its purpose is to examine abnormality in muscles, short-axis images are only taken in a T_2 relaxation time weighted fashion, which is the most sensitive in response to muscle changes. The textures inside the uterus are rarely visible in a T_1-weighted or T_1-fat-saturated image.

In order to integrate the short-axis image into our fusion method, it is useful to recall that the Riesz filter is steerable, as is the Riesz decomposition. Suppose we want to steer the Riesz filter in a way that is defined by a rotation matrix $\mathbf{R}$. Then the steered Riesz filters may be written as $\mathbf{H}(\mathbf{Ru})$. If we expand the terms, we have

$$\mathbf{H}(\mathbf{Ru}) = \mathbf{j}\frac{\mathbf{Ru}}{|\mathbf{Ru}|} \tag{6.1}$$

One important and crucial property of the Riesz filters is that they are isotropic, which means that their magnitude is equal in all directions. More precisely, in this case this means that $|\mathbf{Ru}| = |\mathbf{u}|$. If we write out the terms, expand it, and use trigonometry, it is not hard to see that this is the case. Because of this special property, the rotation matrix can be taken out:

$$\mathbf{H}(\mathbf{Ru}) = \mathbf{j}\frac{\mathbf{Ru}}{|\mathbf{Ru}|} = \mathbf{j}\frac{\mathbf{Ru}}{|\mathbf{u}|} = \mathbf{Rj}\frac{\mathbf{u}}{|\mathbf{u}|} = \mathbf{RH}(\mathbf{u}) \tag{6.2}$$

This makes the implementation of steerable Riesz decomposition considerably easier, as it is simply a rotation matrix multiplication to the components obtained by the

usual Riesz decomposition. A modified steerable Riesz decomposition is written as

$$\mathcal{R}\left(F\right)\left(\mathbf{u}\right)=\begin{bmatrix} F\left(0\right) \\ \mathbf{RH}F\left(\mathbf{u}\right) \end{bmatrix}=\mathbf{j}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & R_{1,1} & R_{1,2} & R_{1,3} \\ 0 & R_{2,1} & R_{2,2} & R_{2,3} \\ 0 & R_{3,1} & R_{3,2} & R_{3,3} \end{bmatrix}\begin{bmatrix} F\left(0\right) \\ \frac{u}{\sqrt{u^2+v^2+w^2}}F\left(\mathbf{u}\right) \\ \frac{v}{\sqrt{u^2+v^2+w^2}}F\left(\mathbf{u}\right) \\ \frac{w}{\sqrt{u^2+v^2+w^2}}F\left(\mathbf{u}\right) \end{bmatrix}=\begin{bmatrix} F\left(0\right) \\ F_{Rx}\left(\mathbf{u}\right) \\ F_{Ry}\left(\mathbf{u}\right) \\ F_{Rz}\left(\mathbf{u}\right) \end{bmatrix} \tag{6.3}$$

where $\mathbf{j}$ is the quaternion unit defined as $\begin{bmatrix} 1 & i & j & k \end{bmatrix}^{\mathrm{T}}$. The corresponding reconstruction would be:

$$\mathcal{R}^{-1}\left(\mathbf{F}_{\mathcal{R}}\right)\left(\mathbf{u}\right)=\begin{bmatrix} F\left(0\right) \\ -\mathbf{H}\cdot\mathbf{R}^{-1}\mathbf{F}_{\mathcal{R}}\left(\mathbf{u}\right) \end{bmatrix}=\begin{bmatrix} 1 \\ -H_x \\ -H_y \\ -H_z \end{bmatrix}\cdot\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & R_{1,1}^{-1} & R_{1,2}^{-1} & R_{1,3}^{-1} \\ 0 & R_{2,1}^{-1} & R_{2,2}^{-1} & R_{2,3}^{-1} \\ 0 & R_{3,1}^{-1} & R_{3,2}^{-1} & R_{3,3}^{-1} \end{bmatrix}\begin{bmatrix} F\left(0\right) \\ F_{Rx}\left(\mathbf{u}\right) \\ F_{Ry}\left(\mathbf{u}\right) \\ F_{Rz}\left(\mathbf{u}\right) \end{bmatrix} \tag{6.4}$$

where $\mathbf{R}^{-1}$ is the inverse rotation matrix, i.e. $\mathbf{R}^{-1}\mathbf{R}=I$, the identity matrix.

This is useful because the rotation matrix effectively steers the decomposition orientation to align the local frame of the short-axis image. Therefore, the resulting rotated member components $F_{Rx}\left(\mathbf{u}\right)$, $F_{Ry}\left(\mathbf{u}\right)$ and $F_{Rz}\left(\mathbf{u}\right)$ are decomposed frequency components along the local x-, y- and z-directions. Recall that the purpose of the decomposition is to collect the frequency component in the through-plane direction of the scan and dispose of it. Suppose $\mathbf{R}$ is defined such that it steers the z-component

of the Riesz filter H_z along the through-plane direction of the short-axis image. Then all the useful high frequencies would be stored in $F_{Rx}(\mathbf{u})$ and $F_{Ry}(\mathbf{u})$. We could replace $F_{Rz}(\mathbf{u})$ with zeros and then rotate the rest of the components back to our global frame, in which the fusion is performed. So the complete flow of fusing the short-axis image is:

1. Riesz filters are rotated such that H_z is aligned with the normal to the short-axis images (i.e. through-plane direction).

2. The image is decomposed in local coordinates in the frequency domain: $F(\mathbf{u}) \rightarrow \begin{bmatrix} F(0) & F_{Rx}(\mathbf{u}) & F_{Ry}(\mathbf{u}) & F_{Rz}(\mathbf{u}) \end{bmatrix}^{\mathrm{T}}$.

3. The image component in the through-plane direction is discarded: $F_{Rz}(\mathbf{u}) \rightarrow 0$.

4. Rotate the components back to the global frame by applying the inverse rotation matrix $\mathbf{R}^{-1}$.

5. Perform fusion in the global coordinates with the Riesz components from the other scans (axial and sagittal).

6. Reconstruct the image from the fused components.

However, several technical problems arise in a real implementation of this method. Although we emphasise the importance of the isotropic property of the Riesz transform, in practice true isotropy is never achieved. This is because the images were never recorded isotropically at the first place - the images conventionally have a rectangular shape. When the Riesz filters are rotated off the image grid, an imbalance is

induced. Depending on the shape of the image, the more elongated it is the higher the imbalance is induced when it is steered. Of course, this imbalance can be restored after the inverse rotation matrix is applied, but only if the image frequency coefficients are not altered. For most images it is reasonable to assume that they are approximately isotropic. It is worth regularly checking that this condition holds, especially when steered filters are used for further processing.

A second tedious problem derives from the fact that the field of view of the short-axis image is very limited. As opposed to a typical full abdominal axial/sagittal scan, which covers all organs in the abdomen, the short-axis scans rarely go beyond the field of view of the uterus. This limits use of the short-axis scan to fusing images just for the uterus. For fusing images for the bladder region, only the axial and sagittal images are available and used. Fortunately, even with only the axial and sagittal scan, it has shown big improvement in restoring details of the image. Figure 6.8 shows the resulting fused image (middle column), comparing with the pre-fused axial (left column) and sagittal (right column) image, respectively. It has shown that the fused image has the best of both pre-fused scans. Moreover, the fused image can also fix some ambiguities at boundaries due to the partial volume effect. For example, by comparing the top-left and top-middle graphs in Figure 6.8 it can be seen that the lining of bladder is less blurry and distinct in the fused image, which is good for segmentation.

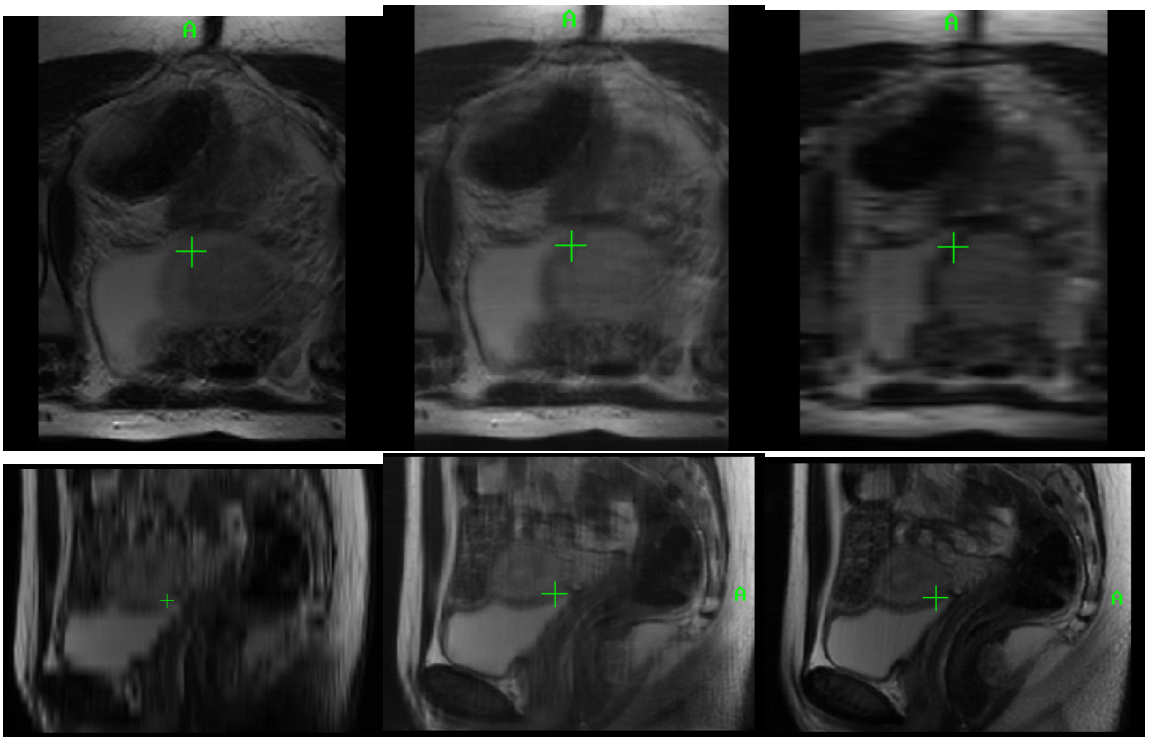

Figure 6.8: Comparison of the fusion result. Top-left: axial view of the pre-fused axial scan. Bottom-left: sagittal view of the pre-fused axial scan. Top-middle: axial view of the resulting fused image. Bottom-middle: sagittal view of the resulting fused image. Top-right: axial view of the pre-fused sagittal scan. Bottom-right: sagittal view of the pre-fused sagittal scan.

6.3 Phase Variance Feature Detection

The feature detection method is based on local phase - an interpretation of the analytic signal representation. While the Hilbert transform is well-known for constructing the imaginary counterpart of a 1-D real signal, the Riesz transform is used as an isotropic Hilbert transform extension to n-dimensions. Since features are mostly scale dependent, a scale-space representation is therefore recommended and controlled by a pre-defined bandpass filter. Conventionally, multi-scale features are considered separately, or combined by simple addition or local phases/energy. A novel contribution of this thesis is the application of the geometric product as a mean to combine multi-scale image representations. One property of the geometric product is that the phase of the product is equal to the sum of the phases of the member vectors. As a result, the sum of multi-scale phases, hence the mean phase, can be computed reliably without worrying about the classic problem of phase wrapping, from which inverse trigonometric functions such as arcsine, arccosine and arctangent all suffer. These have already been presented and discussed in Chapter 3 and 5.

6.3.1 Mean Phase Computation Using De Moivre's Theorem

In order to compute the mean phase without using inverse trigonometry, phase manipulations have to be done using the sine/cosine representation. For example, we may want to find $\cos\theta_m$ knowing $\cos n\theta_m$, where θ_m is the mean phase over scales. Their relationship is given by de Moivre's theorem:

$$(\cos\theta + j\sin\theta)^n = \cos n\theta + j\sin n\theta \tag{6.5}$$

In our application, n is the number of scales we are considering, and the right-hand-side of the equation is the geometric product of multi-scale monogenic image representation, while θ on the left-hand-side is the mean phase that we want to solve the equation for. Mathematically, the geometric mean of the geometric product is written as

$$\cos\theta_m + j\sin\theta_m = \sqrt[n]{\cos n\theta_m + j\sin n\theta_m} = \left(\prod_{k=1}^{n}\mathbf{m}_k\right)^{\frac{1}{n}} \tag{6.6}$$

where the operator $\prod$ represents geometric product series and $\mathbf{m}_k$ is the monogenic image representation at scale k. If needed the mean phase can be computed via the angle of the complex geometric mean:

$$\theta_m = \arg\left(\left(\prod_{k=1}^{n}\mathbf{m}_k\right)^{\frac{1}{n}}\right) = \arctan\left(\frac{\text{imag}\left(\left(\prod_{k=1}^{n}\mathbf{m}_k\right)^{\frac{1}{n}}\right)}{\text{real}\left(\left(\prod_{k=1}^{n}\mathbf{m}_k\right)^{\frac{1}{n}}\right)}\right) \tag{6.7}$$

In practice, when implementing the above method, in particular formula 6.6, there is a technical difficulty involved in computing the inverse powers in MATLAB. This is indeed a problem for computing the inverse of power that is above 2. The fundamental cause of this problem derives from the fact that when the inverse power is required for a complex number, MATLAB blindly applies de Moivre's theorem but only in a specific way. Suppose a complex number $z = a + jb$ to the power of $\frac{1}{5}$ is to be computed. What MATLAB does is to calculate the phase of the complex number via

$\theta = \arctan\frac{b}{a}$, then divides θ by 5, and then returns the resulting complex number as $\sqrt[5]{z} = \sqrt[5]{|z|}\cos\frac{\theta}{5} + j\sqrt[5]{|z|}\sin\frac{\theta}{5}$. This works, but only over the limited range of 2π for θ, because that is the range which the arctangent function implicitly assumes. For instance, MATLAB uses the "atan2" function, which assumes $\theta \in [\pi, -\pi)$ to calculate the phase of z. Suppose θ is in fact outside this range, for example is equal to $\frac{3\pi}{2}$, or 270°, instead of $-\frac{\pi}{2}$, or −90°, which is assumed by the arctangent function. This subtle difference leads to completely different results, as MATLAB would give an answer of the new phase $\frac{-90°}{5} = -18°$, while it is quite conceivable that $\frac{270°}{5} = 54°$ is what we really want! To see the problem directly linked to our case, now suppose our mean phase is in fact 54°, which is well inside the supposedly $[\pi, -\pi)$ range, and we have 5 levels of scales. Let $z = \cos 54° + j\sin 54°$ be the true geometric mean complex vector and z^5 be the geometric product we have (inside the brackets in equation 6.6). When we compute the mean phase by computing $z = \left((\cos 54° + j\sin 54°)^5\right)^{\frac{1}{5}}$, we do not get back to $\cos 54° + j\sin 54°$, i.e. $(z^5)^{\frac{1}{5}} \neq z$. Instead MATLAB returns $\cos -18° + j\sin -18°$, which is a completely different vector in the complex plane!

In fact, solving $\cos n\theta_m + j\sin n\theta_m$ for $\cos\theta_m + j\sin\theta_m$ is a multi-solution problem. The algorithm that MATLAB uses only gives one of the solutions that is within the π and $-\pi$ range. To demonstrate the problem, suppose $n = 2$ and let $z = \cos n\theta_m + j\sin n\theta_m = a + jb$ and $z_m = \cos\theta_m + j\sin\theta_m = p + jq$. By de Moivre's theorem we have,

$$z = a + jb = (z_m)^2 = (p + jq)^2 \tag{6.8}$$

$$= \left(p^2 - q^2\right) + j2pq \tag{6.9}$$

Equating the real and imaginary parts we have a set of simultaneous equations in two unknowns p and q:

$$p^2 - q^2 = a \tag{6.10}$$

$$2pq = b \tag{6.11}$$

Here it is useful to recall the identities between sine and cosine (hence p and q) that $p^2 + q^2 = 1$. The first equation is then reduced to $2p^2 - 1 = a$, and p is solved to be $p = \pm\sqrt{\frac{a+1}{2}}$, which is substituted to the second equation to solve for $q = \pm\frac{b}{2}\sqrt{\frac{2}{a+1}}$. In short, we have two sets of solutions:

$$\begin{aligned} p_1 &= \sqrt{\tfrac{a+1}{2}}, \quad q_1 = \tfrac{b}{\sqrt{2(a+1)}} \\ p_2 &= -\sqrt{\tfrac{a+1}{2}}, \quad q_2 = \tfrac{-b}{\sqrt{2(a+1)}} \end{aligned} \tag{6.12}$$

Suppose $a = \cos 270^\circ = 0$ and $b = \sin 270^\circ = -1$. The square root function in MATLAB would assume $\theta_{m1} = -90^\circ$ and return $p_1 = \frac{\sqrt{2}}{2}$ and $q_1 = -\frac{\sqrt{2}}{2}$. However, in fact there is another solution that $p_2 = -\frac{\sqrt{2}}{2}$ and $q_2 = \frac{\sqrt{2}}{2}$, representing $\theta_{m2} = 135^\circ$.

When $n = 3$, the same method can be applied by writing $a + jb = (p + jq)^3$,

expanding the terms on the right-hand-side, and equating the real and imaginary parts. A simultaneous equation set is written as

$$p\left(p^2 - 3q^2\right) = a \tag{6.13}$$

$$q\left(3p^2 - q^2\right) = b \tag{6.14}$$

Again by applying the trigonometric identity $p^2 + q^2 = 1$, the equation set can be reduced to

$$p\left(4p^2 - 3\right) = a \tag{6.15}$$

$$q\left(4p^2 - 1\right) = b \tag{6.16}$$

Since equation 6.15 is a third order polynomial, it is obvious that it has three roots, or solutions of p. Although it is tedious to write out the three solutions of p explicitly, it is easy to solve for them numerically by optimisation with a pre-defined initialisation value p_0. Generally speaking, the specific solution we get by optimisation relies solely on the value p_0 to initialise the search. Once one solution of p is found, q can be derived by substitution as $q = \frac{b}{(4p^2-1)}$.

6.3.2 Chebyshev Polynomials

To generalise this problem, we need to have an expression for a in terms of p for any number of levels n. In fact, this expression is called a Chebyshev polynomial. There are two kinds of Chebyshev polynomials mostly known as T_n (the first kind) and U_n (the second kind). The first kind T_n is closely related to the relationship between $\cos n\theta$ and $\cos\theta$, and noticeably $\cos n\theta$ is indeed equal to the Chebyshev polynomial of $\cos\theta$ of the first kind, or $\cos n\theta = T_n(\cos\theta)$. In our representation, it means $T_n(p) = a$ for n levels. The second kind of Chebyshev U_n relates $\sin n\theta$ and $\sin\theta$. Explicitly, it is written as $\sin n\theta = U_{n-1}(\cos\theta) \cdot \sin\theta$. Putting them together, we have:

$$T_n(p) = a \tag{6.17}$$

$$q \cdot U_{n-1}(p) = b \tag{6.18}$$

From this equation set, we notice that equation 6.17 is a single variable polynomial, from which p can be solved. Since the n-th Chebyshev polynomial has the order of power n, there will be n solutions for p, though they are not necessarily all distinctive roots. Once a root of equation 6.17 is found, the corresponding q can be derived from $q = \frac{b}{U_{n-1}(p)}$.

A compact definition of the first kind of the Chebyshev polynomials is given in a recurrence fashion:

$$T_0(x) = 1 \tag{6.19}$$

$$T_1(x) = x \tag{6.20}$$

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x) \tag{6.21}$$

from which we observe that $T_2(x) = 2x^2 - 1$ and $T_3(x) = 4x^3 - 3x$, which match equation 6.10 and 6.15 which we derived above for $n = 2$ and $n = 3$, respectively. Similarly, the second kind of Chebyshev can be defined recurrently:

$$U_0(x) = 1 \tag{6.22}$$

$$U_1(x) = 2x \tag{6.23}$$

$$U_n(x) = 2xU_{n-1}(x) - U_{n-2}(x) \tag{6.24}$$

One can easily see that $U_1(x) = 2x$ and $U_2(x) = 4x^2 - 1$, which is also confirmed by our derivations for equation 6.11 and 6.16, respectively; and q is indeed equal to $q = \frac{b}{U_{n-1}(p)}$. As the above recursive definition is compact, sometimes it is useful to have an explicit expression of the polynomials. Fortunately, we can write T_n and U_n as follows:

$$T_n(x) = \frac{\left(x+\sqrt{x^2-1}\right)^n + \left(x-\sqrt{x^2-1}\right)^n}{2} \tag{6.25}$$

$$U_n(x) = \frac{\left(x+\sqrt{x^2-1}\right)^{n+1} - \left(x-\sqrt{x^2-1}\right)^{n+1}}{2\sqrt{x^2-1}} \tag{6.26}$$

Now the remaining question is how to set p_0 - the value to initialise the optimisation process. Generally speaking, suppose we know the roots of ${}^{dT_n(x)}/_{dx}$ are, say $[x_1, x_2, x_3, \ldots, x_{n-1}]$, which are equal to the position of stationary points (local minimum, maximum, inflection points) of T_n. Since T_n is an nth order polynomial, it is expected to be a monotonic function between its stationary points. Therefore, any root of $T_n(x)$ (if exists) must lie in one of the intervals between $[x_1, x_2, x_3, \ldots, x_{n-1}]$, i.e. $[-\infty, x_1)$, $[x_1, x_2)$, $[x_2, x_3)$, ..., $[x_{n-2}, x_{n-1})$, and $[x_{n-1}, +\infty]$. To simplify the problem further specifically for our application, recall that when $x = \cos\theta$, x is bound to the range $[-1, 1]$. Hence our range of root searching is reduced from an infinite $[-\infty, +\infty]$ to a finite range $[-1, 1]$. Moreover, recall that the Chebyshev polynomial $T_n(x)$ is also bound to $[-1, 1]$ because $T_n(\cos\theta) = \cos n\theta$. In fact, all the stationary points of T_n are either maximum or minimum points (no inflection points) and exactly equal to 1 or -1, respectively. This implies that when solving polynomial $T_n(p) - a = 0$ for p, there will be exactly n distinctive roots, except the cases when a is equal to -1, or 1. These two very special cases happen when pairs of roots collapse to only one distinctive root at local minima/maxima. This is best demonstrated by a plot of $T_n(x)$ for different n (Figure 6.9). For the desired root p_{desired} to be found by optimisation, there is one strict condition - the initialisation p_0 is posed within the

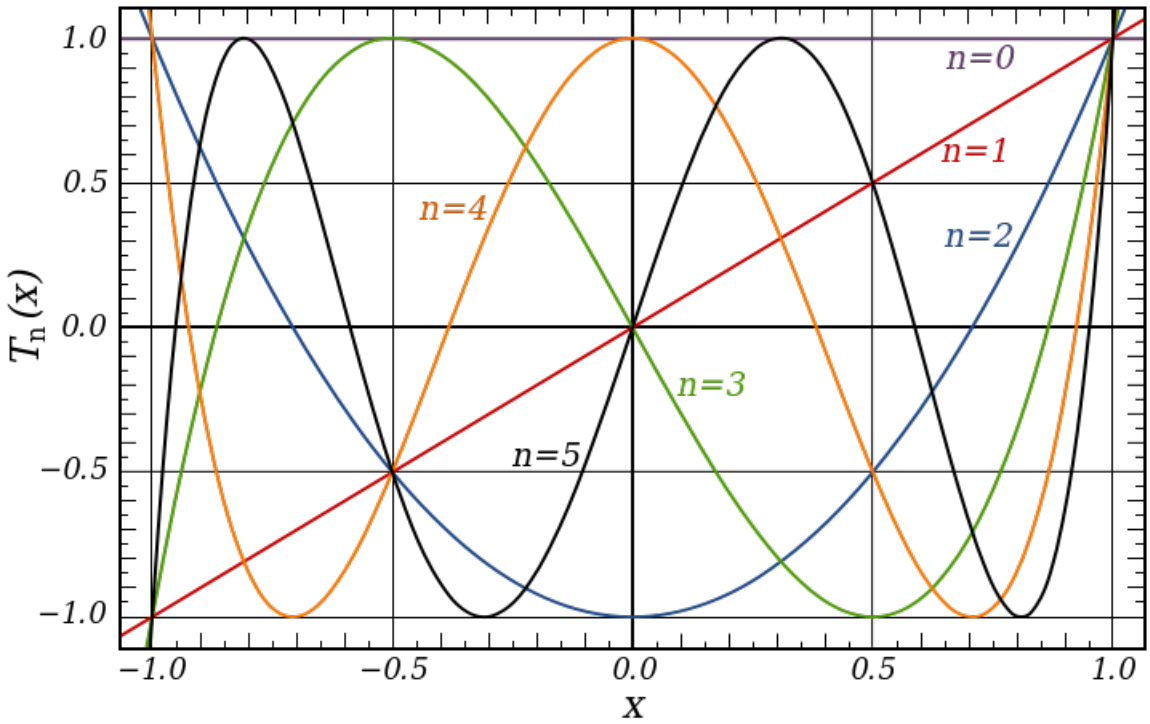

Figure 6.9: Plot of the Chebyshev polynomial of the first kind $T_n(x)$ when $x \in [-1, 1]$. Solving equation $T_n(q) - a = 0$ for q will always have n distinctive roots, except the cases when $a = -1$ or 1.

same range in which p_{desired} is (for example, say, $[x_2, x_3)$ if $p_{\text{desired}} \in [x_2, x_3)$) because of the monotonic property of T_n within that range. Results may be improved if another weaker condition is also met - a is not too close to -1 or 1, because curvatures near the stationary points are small and the optimiser may mistakenly jump into the adjacent interval.

Unfortunately, there is no simple mechanism for the initial guess of p_0 that guarantees the return of p_{desired}. However, we propose one way to achieve this under the assumption that the features of the image change slowly over different scales. Since the local phase is a proxy for image features, if features (such as edges and textures) change slowly over scales so does the local phase. Suppose an image is filtered by

a series of bandpass kernels, each of which broadly selects a particular scale k, and k is chosen to increase linearly and slowly. It would not be hard for one to verify that the intensity of a particular pixel in the image changes smoothly according to k. This is because as the scale parameter k changes, the scale selecting bandpassed version of the image is also switching slowly from a blurred image to another image that reflects the high frequencies. The cosine of the mean phase $\cos\theta_m$ that we are ultimately solving for is in fact a mean bandpass representation of the image, since the mean monogenic image vector can be thought as

$$\begin{aligned}
\mathbf{m}_m(\mathbf{x}) &= B_m * I(\mathbf{x}) + j\,|B_m * \mathbf{H} * I(\mathbf{x})| && (6.27)\\
&= I_m(\cos\theta_m + j\sin\theta_m) && (6.28)\\
&= \left(\prod_{k=1}^{n}\mathbf{m}_k\right)^{\frac{1}{n}} = \left(\prod_{k=1}^{n}(I_k\cos\theta_k + jI_k\sin\theta_k)\right)^{\frac{1}{n}} && (6.29)
\end{aligned}$$

where B_m is the mean bandpass filter, I_m is the geometric mean of the local energy over different scales, and $\prod$ is a geometric product operator. Suppose $\cos\theta_k$ changes smoothly as k moves from 1 to n, it is reasonable to assume that the cosine of the mid-scale is closer to the cosine of the mean phase. Therefore, for the search for $p_{\text{desired}} = \cos\theta_m$ we propose to use the normalised real part of the mid-scale monogenic image vector to initialise the optimiser, or written as

$$p_0 = \cos\theta_{\lfloor\frac{n}{2}\rfloor} \tag{6.30}$$

where the operator $\lfloor . \rfloor$ means rounding to the nearest integer. With this initial guess, the optimization algorithm guarantees to solve for a solution p_{desired} that is in the same monotonic interval as p_0. The result is a consistent smooth mean phase map without excessive phase wrapping - the problem that we intend to solve at the first place.

6.4 Bladder Wall Segmentation

The inner wall of the bladder is best seen in the T_2-weighted MR images due to the strong contrast between the low density fluid inside the bladder and the high density muscle tissues. Bland fluid inside the bladder has high intensities on the T_2-weighted image while bladder wall muscles have very low intensities. This makes the edge based level sets method work best. The problem of segmenting the inner bladder using the Chan-Vese region based method is that the Chan-Vese method is based on a *global* intensity statistical measure, whose effectiveness depends critically on the uniformity of the intensity distribution. Unfortunately, the well-known MR bias field effect distorts that uniformity. Especially at regions of high intensities like fat and fluids in T_2-weighted images, the distortion is particularly intense. This significantly reduces the reliability of the Chan-Vese method for segmenting the inner bladder wall. A modified version of the Chan-Vese region based method such as the Lankton's local region level sets could improve the results. However, this is affected by the choice of the size of the local region. If the size of local region is too large, it also suffers from the bias field intensity distortion. However, if the size is too small, it suffers from noise or

fusion artefacts at a small scale. In contrast, the edge based method is designed and well known for picking out edges of high intensity changes. Moreover, the proposed local phase feature detection method is based on the relationship between local and global structures and does not depend on the simple intensity change, which can be induced by the MR bias field effect. Therefore we suggest the use of the local phase edge based level sets for segmenting the inner bladder wall as it is neat and produces more reliable results.

When segmenting the bladder in 3-D, however, a high definition and isotropic scan of the bladder does not exist, despite the ones that we reconstructed using the proposed fusion method. Instead, the raw images acquired are only isotropic in 2-D slices. Before applying the method directly to the fused 3-D images, we suggest to test the approach on 2-D slices first. This is because we want to separate the segmentation failure that is due to the segmentation method itself from that which is merely due to fusion artefacts. Since errors would be cumulative, any artefact induced by the fusion would distort the image and propagate the error to the segmentation. By testing the segmentation method on the 2-D raw image first, we are confident that no fusion or registration artefact is there and the results would truly reflect the performance of the segmentation algorithm alone.

Figure 6.10 shows a typical pair of T_2- and T_1-weighted MR images of the bladder, as well as their intensity histograms. The bladder wall is segmented manually, where the inner wall boundary is denoted in green and the outer wall boundary is denoted in yellow. The histogram of the entire image is plotted in blue; and the histograms of the manually segmented regions are plotted in the corresponding colour

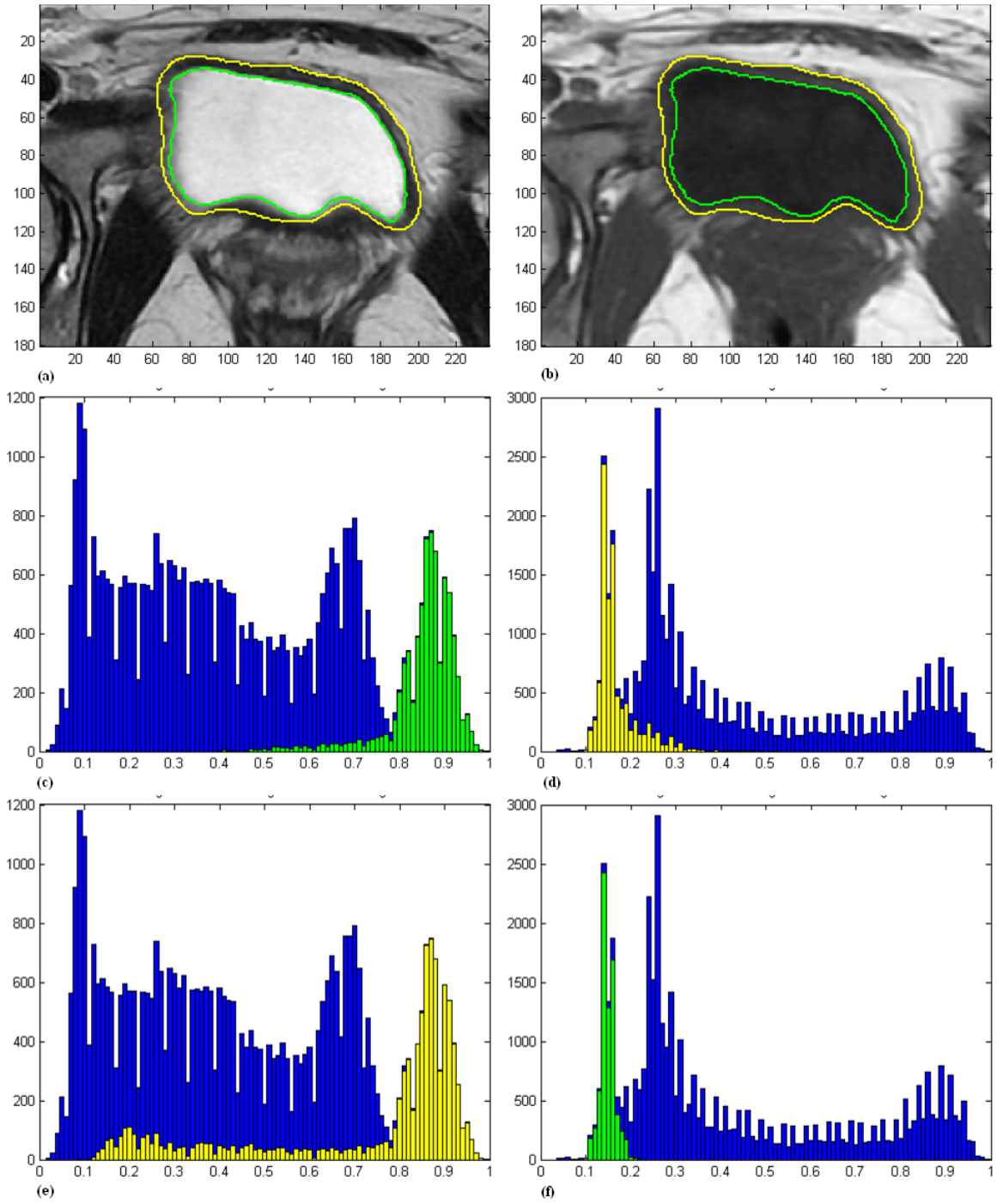

Figure 6.10: Intensity histogram of the bladder segmentations in T_2-weighted and T_1-weighted axial image. Chart (a) and (b): A typical T_2-weighted and T_1-weighted axial image of the bladder, respectively, with manual segmentation of the inner wall (green contour) and the outer wall (yellow contour). Chart (c) and (e): Histogram of the T_2 image with highlighted inner wall region (in green) and outer wall region (in yellow). Chart (d) and (f): Histogram of the T_1 image with highlighted inner wall region (in green) and outer wall region (in yellow)

and overlaid on the image histogram. Ultimately our aim is to automatically segment the area between the two regions (marked in green and yellow). In a T_2-weighted image, the inner wall boundary region is bright because spin-spin interaction among the hydrogen protons is minimal due to the structure and the sparsity of the water molecules. However, the molecules in the bladder wall tissues can have different sparsity because of muscle contraction; hence their intensities vary (Figure 6.10, a-3). Moreover, the T_1 characteristics depend on the proton's ability to transfer energy between the surrounding lattice, and the most efficient energy transfer occurs when the natural motional frequencies of the protons are at the Larmor frequency, which is proportional to the magnetic field strength. It turns out that the natural motional frequency of hydrogen protons in water is much higher than the Larmor frequency, whereas the natural motional frequencies of hydrogen protons in muscles are lower. Therefore both water inside the bladder and the bladder wall tissues are dark on the T_1 image (Figure 6.10, b-2).

6.4.1 Coupled Level Sets Segmentation

The segmentation quality in the T_2-weighted image is limited by the variation of the muscle intensity, and the segmentation in the T_1-weighted image is difficult because of the lack of clear boundaries between different tissues. With this in mind, we have developed a coupled level set method to segment the bladder wall, in which the inner wall boundary is obtained from the T_2 image and the outer wall boundary is obtained from the T_1 image. The inner and outer bladder wall boundaries are obtained in a

sequential approach. The geodesic active contour model described in chapter 5 is first applied to the T_2-weighted bladder image with a local phase-based edge function $g(\mathbf{x})$ and with an initialisation inside the bladder. The resulting segmentation of the inner wall boundary then forms the initialisation of the segmentation in the T_1 image, while the level set model in the T_1 image is coupled with the resulting contour in the T_2 image. We have developed two methods for coupling: one employs an explicit stopping criterion ("switch"), while the other has a spring force term ("spring"). In both methods, a dual level set representation is used, where the level set function for the inner wall is called ϕ_{in} and the level set function for the outer wall is called ϕ_{out}. ϕ_{in} is obtained from the segmentation in the T_2 image and fixed. For the "switch" method, ϕ_{out} evolves under the governing equation:

$$\frac{\partial \phi_{\text{out}}}{\partial t} = \left(\kappa \text{div} \left(g_{\text{T1}}(\mathbf{x}) \frac{\nabla \phi_{\text{out}}}{|\nabla \phi_{\text{out}}|} \right) + (1 - H(\phi_{\text{in}} - d)) c g_{\text{T1}}(\mathbf{x}) \right) |\nabla \phi_{\text{out}}| \qquad (6.31)$$

where κ is the contour coefficient, c controls the magnitude of the balloon force, $g_{\text{T1}}(\mathbf{x})$ is the edge function of the T_1-weighted image, and the factor $(1 - H(\phi_{\text{in}} - d))$ is the stopping criterion of the balloon (inflation) force as $H(\ldots)$ is the Heaviside function and d is the maximum distance allowed for the balloon force to act away from the inner wall boundary $C_{\text{in}} = \{\mathbf{x} \mid \phi_{\text{in}}(\mathbf{x}) = 0\}$. Note that d represents prior knowledge of the maximum bladder wall thickness. For the "spring" method, the evolution is formulated as

$$\frac{\partial \phi_{\text{out}}}{\partial t} = \left(\kappa \text{div} \left(g_{\text{T1}} \left(\mathbf{x} \right) \frac{\nabla \phi_{\text{out}}}{|\nabla \phi_{\text{out}}|} \right) + c g_{\text{T1}} \left(\mathbf{x} \right) - k \phi_{\text{in}} \right) |\nabla \phi_{\text{out}}| \tag{6.32}$$

where k is the spring constant. As a signed distance function, the values of $\phi_{\text{in}}(\mathbf{x})$ represent the shortest distance from the point $\mathbf{x} = (x, y, z)$ to the inner wall contour C_{in}. In this way, the term $k\phi_{\text{in}}$ can be thought of as a spring force that acts between the inner and outer contours (hence k corresponds to the spring constant).

In order to measure the performance of the methods, segmentation results are compared with manual segmentations obtained by the authors using a tablet computer. Segmentation quality is then assessed using the Dice similarity coefficient (DSC) between manual and automated results:

$$DSC = \frac{2N \left(R_{\text{man}} \cap R_{\text{alg}} \right)}{N \left(R_{\text{man}} \right) + N \left(R_{\text{alg}} \right)} \tag{6.33}$$

where $N(\ldots)$ is the number of pixels/voxels in a region, and R_{man} and R_{alg}represents the labelled regions of the manual and algorithmic segmentation, respectively.

6.4.2 Validation

Data sets of 11 patients were acquired. For each patient data set, the axial T_1-weighted image was rigidly registered to the axial T_2-weighted image, from which 54 axial slices of the bladder were extracted. Since we are using both the T_2- and T_1-weighted image for segmentation, we have to assume that the deformation of the bladder was small and negligible. The local phase-based feature detection model is performed for all images to produce the edge function $g(\mathbf{x})$. Then $g(\mathbf{x})$ is passed to

the level set formulations and three segmentations are performed for each image pair: no-coupling (level sets on both T_2- and T_1-weighted image), "switch" coupling and "spring" coupling. For the "switch" method, d is set to be 5 pixels, which corresponds to the expected bladder wall thickness of 2.5 mm. k for the "spring" coupling is set to be 0.1. The expansion coefficient c is equal to -0.8, and the contour coefficient κ is 4 and 20 for the T_2 and T_1 segmentation, respectively.

Figure 6.11 shows the DSC values measured between the resulting and manual segmentation; and Figure 6.12 shows the results of 3 segmentation trials. For 72% of the total 54 trials for segmenting the inner bladder wall boundary in the T_2-weighted image, the DSC is over 95%. The 10 worst performing cases are subject to three types of failures: (a) a half-full/collapsed bladder causing creased wall boundaries and inhomogeneous intensities inside; (b) ambiguous boundaries caused by another water containing object (typically a cyst) adjacent to the bladder; and (c) vague boundaries in the top/bottom image slice of the bladder. We note that all of these three types of failures are due to the partial volume effect, which is what the proposed fusion method is supposed to solve. When segmenting in a fused 3-D image, the effect of these problems would be reduced to minimum. For the segmentation of the outer bladder wall boundary in the T_1-weighted image, when such failures happen, there is no way to recover the outer wall boundary as it relies on the result from the T_2 image. Despite the T_2 segmentation, if a single edge based level set model is run on a T_1-weighted image without coupling, almost in every trial the active contour leaks into other organs due to the similarity of intensities with adjacent organs (Figure 6.12). By applying the proposed coupled method, the number of leakages is significantly

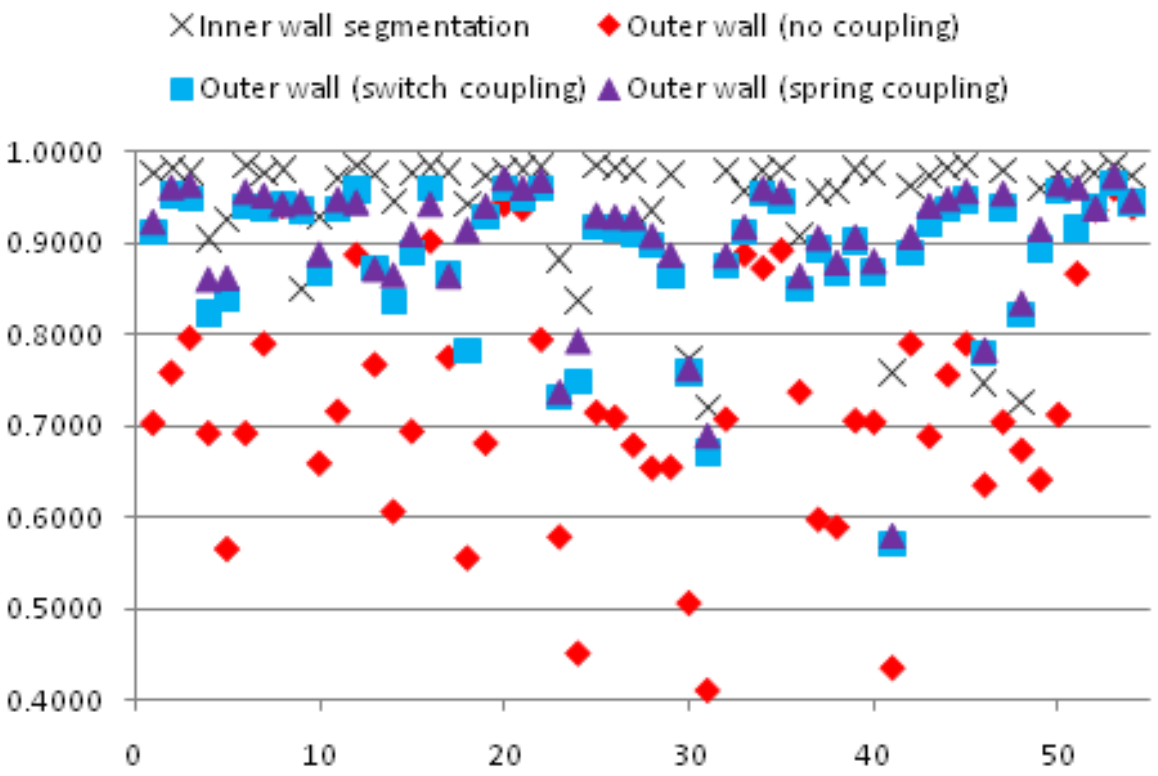

Figure 6.11: DSC plot of total 54 segmentation experiments; x-axis is experiment number, y-axis is DSC value (range from 0 to 1, 1 is best).

reduced (from 41 to 7) and the number of cases of DSC over 0.95 increases from 1 to 14 for segmenting the outer wall.

As shown in Figure 6.12, currently there is little difference between the "switch" and "spring" coupling methods, probably as a result of the parameter settings, but they have some different properties. The "switch" coupling acts to eliminate the balloon force when C_{out} is a certain distant away from C_{in}. This method only guarantees to stop leakage when the maximum allowable distance is reached, and it does not help to attract the contour to a weak edge that is within this distance; while the "spring" method somewhat helps the contour attract to weak edges by the additional spring force acting against the balloon force. However, the trade-off of this is that

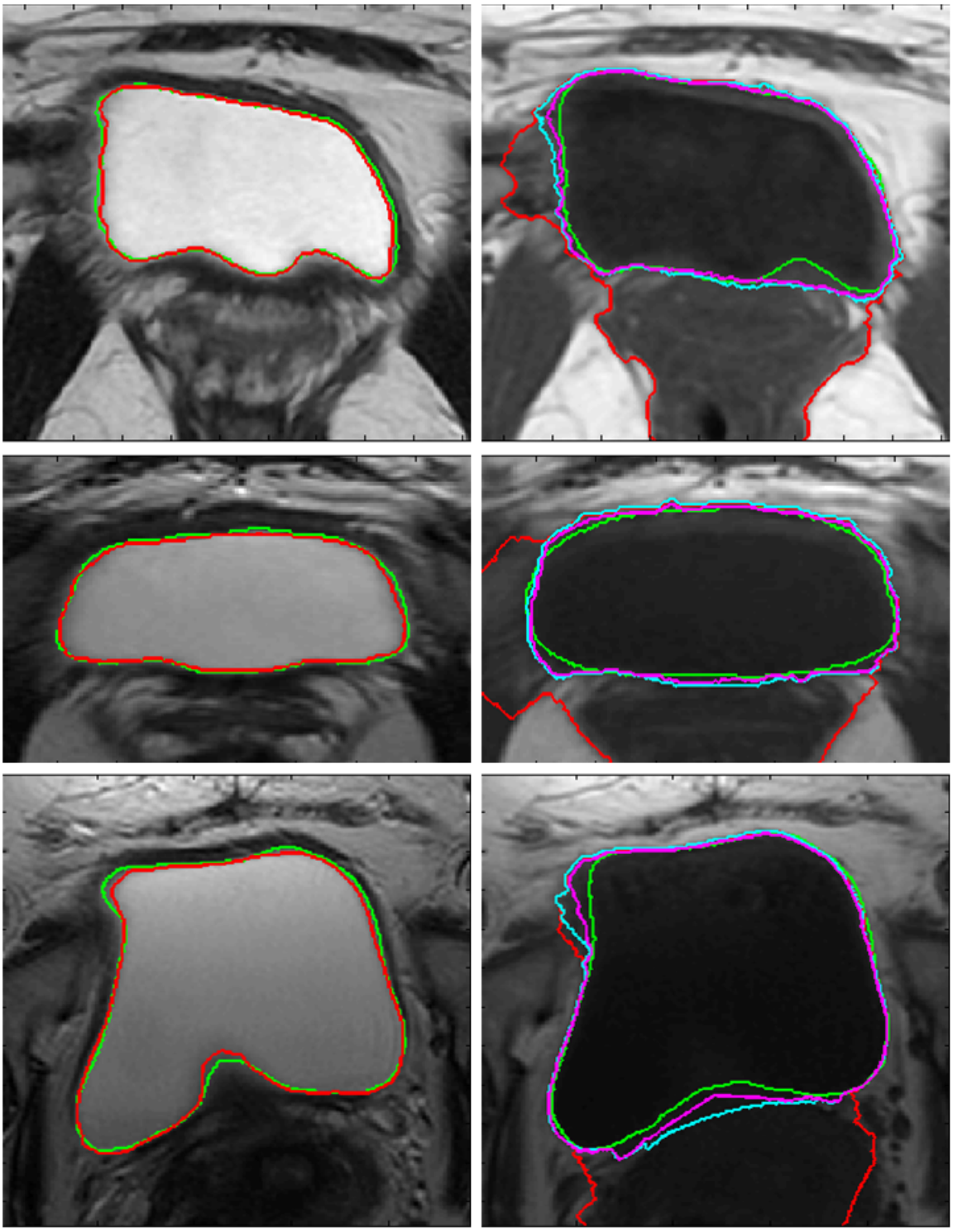

Figure 6.12: Representative segmentation result of 3 example data sets. Left column: inner wall segmentation. Right column: outer wall segmentation. Green: manual segmentation contour. Red: automatic segmentation with no coupling. Light blue: with switch coupling. Purple: with spring coupling.

the bladder wall thickness measured by this method will be slightly lower than for the "switch" method because of the additional spring force. Another limitation is due to the assumption that the two images are perfectly registered. In practice, there are small deformations due to patient movement and the bladder filling-up.

6.5 Conclusions

In this chapter we describe image processing for real clinical data. Following the order of processing steps, the main components are image registration, image interpolation/resampling of the common field of view, image fusion, and segmentation of the bladder wall. Although the ground truth of the patient image in 3-D is unknown (since an isotropic high definition 3-D image of the patient was not recorded), we compare the resulting fused images by different fusion rules qualitatively, and conclude that the Max rule is likely to preserve localisation of features better, especially for sharp edges, while images generated by the Mean rule is better for visual assessment, such as for radiologists to examine the patient's anatomy. The Adaptive rule is a parameterised fusion rule that switches between the two. We also introduce a steerable Riesz fusion method to make use of short-axis, or other image that is acquired not along the Euclidean axes. In order to solve the problem of the n-th root of a complex number when computing the vector mean for mean phase, we present a solution of solving the Chebyshev polynomials. As this is a classic multi-solution problem, we manage to solve it numerically by making sensible initialisations (initial guess of the desired solution). The result is a reliable way to calculate the mean phase

of multi-scale processing and hence an accurate measure of phase variance. Finally, in order to utilise the appearance discrepancy of the bladder inner and outer wall boundary in different MR echo sequences, we devise a coupled level set method to segment the bladder wall, and validate it with manual segmentations in 2-D.

Chapter 7

Conclusion And Future Work

The work described in this thesis addresses a very challenging and broad question: diagnosing endometriosis, in particular the deep infiltrating endometriosis lesions, using image processing techniques. This is not easy because on its own it is a difficult disease to diagnose by looking at the images alone even for the best trained radiologists. Endometriosis is generally a diffused disease, and lesions can be anywhere in the patient's abdomen. The sensitivity and specificity of diagnosing endometriomas (a cystic type of endometriosis lesion at the ovaries) is high by MRI but they can be achieved by the cheaper ultrasound imaging as well; and the localisation of endometriomas provided by MRI is less important because its variation does not require different surgical procedure. In contrast, deep infiltrating endometriosis lesions that cause fibrosis in the bladder wall or the bowel wall muscles are of particular interest to us. Although MRI already offers the best sensitivity and specificity for examining these lesions, the rate of false negatives is still high. The reasons for this probably comprise several factors:

- we are trying to identify the lesions from sparsely sampled data. Even with high in-plane resolution, there is typically a 2 mm gap between each MRI slices, which is also a similar size of a lesion. Lesions can be missed in these gaps.

- the images are acquired from an anisotropic frequency sampling process. This directly leads to long-thin voxels in the images (or thick slices), whose intensities are distorted by partial volume effect. Fibrotic nodules exist in the bladder/bowel wall, whose thickness is about 2 - 3 mm; while the MR image slice thickness is even bigger at 3.5 - 4.5 mm. The partial volume effect may distorts the intensity and texture significantly at the wall structure, which makes identifying nodules in these images particularly hard.

- unlike a brain image, which is the image of a rigid structure, abdominal images are subject to body movements - breathing motion, bowel contraction, and bladder filling up during the course of the MRI scan. These motions can easily displace organs by several millimetres. Moreover, these motions take place during a frequency sampling process. The result is not a pure dislocated boundary; it is the inconsistency among sampled frequency components that distort textures and induce artefacts around the bladder/bowel walls.

In response to these problems, we devised a method that upsamples the image to one with isotropic voxel size without losing its high in-plane resolution. In addition, the method fuses multi-view MR images to restore a high frequency sampling rate in all three directions. The fusion is conducted in the frequency domain, as we want to reverse the process of sparse MR frequency sampling. By using the Riesz transform

the frequency spectrum of each single view MR image is decomposed into x-, y- and z-components. These components are then selected and fused to become the combined spectrum. The result is a fused image that is corrected for most of the distortion and artefacts. We have validated the method with simulated brain images, where the ground truth is known. We have also assessed different fusion rules on real clinical data qualitatively. We concluded that choosing between the Mean and Max fusion rule is a trade-off between preserving feature certainty and feature localisation. The Mean rule presents features with certainty while it is a low-pass operation, which inevitably blurs sharp edges. On the other hand, the Max rule preserves very sharp edges and hence accurate localisation but induces extra image artefacts and noise. Therefore, we provided the Adaptive rule that switches between the two. In order to cope with a member image that is not on the main axes, such as the short-axis image, we also presented a steerable version of the proposed fusion method. With the steerability, we are able to fuse images that are in any orientation.

Since the main difficulty of assessing the disease in MR images is the distortion and artefacts at wall structures, we decided that having a good feature representation is paramount. We were inspired by the idea of phase congruency, but also recognised its problem of high sensitivity to noise. We use wavelets for scale selection, which is equivalent to frequency band selection, and apply monogenic signal analysis to compute local phases. We proposed phase variance as a measure of phase congruency. To calculate the phase variance accurately and efficiently, we propose to compute the mean phase using geometric algebra. Instead of using inverse trigonometric functions, which are prone to the phase wrapping problem, we suggest first computing the

geometric product of the complex image representation at each scale. Then the geometric mean is calculated by taking the n-th root of the product. It leads to solving for the n-th order Chebyshev polynomials, which we have provided an effective way to solve numerically. We have also investigated and compared five well-known filters and suggested the Gaussian-derivative filter for our application. The result is a robust feature detector that manages to identify the bladder wall boundaries.

Two different implementations of level sets were considered for image segmentation - the edged based geodesic active contour model and the localised region based approach. Both methods provide some form of control over scales. However, the difference is that edged based feature detection defines scale selection as a frequency bandpass processing, while the scale control in the localised region based method is related to a band selection in a histogram instead. Although both methods are resistant to intensity non-uniformity, the region based method is less robust to artefacts around boundaries. This is because that the size of localisation in the region based method only changes the scope for the optimisation process but does not make the artefacts go away. On the other hand, these artefacts can be diluted via frequency filtering. For segmenting the bladder wall, we suggested two options of coupled level sets - "switch" and "spring" coupling, and validated the segmentations by comparing the results to manual segmentations. The results show that the coupling method effectively reduces the number of cases that suffer from boundary leaking into other adjacent structures, and improves the bladder wall segmentation significantly.

7.1 Future Work

The problem with clinical MRI image acquisition is that time is limited. Depending on the MRI scanner, there is a shortest repetition time that the machine can use without reducing the image quality significantly. Moreover, the maximum amount of time that we can leave the patient in the MRI scanner is limited, as too long a period of time will cause discomfort to the patient. As a result, there is a physical limit on the number of samplings that the machine can make; hence it constrains us to continue having sparse data acquisitions. Therefore, further improvements on the currently proposed fusion method would add real value to this project. The leading remaining limitation to the quality of the reconstructed image is the deformation due to body motions during the scan. An image registration step is needed to correct for these deformations. In addition to the conventional non-rigid image registration where the deformation field is computed in the spatial domain, we may consider a method that deforms the source image in the frequency domain instead. At each iteration, the fused image is reconstructed, and the overall phase variation of the fused image is computed as a measure of featureness. The goal is to optimise for the sharpest fused image (analogous to finding the focus of an optic lens) by deforming the frequency spectrum of the source image.

Once a high resolution fused image is obtained, image segmentation can be applied to delineate individual organs. In this thesis we present a method to segment the bladder. For future work, attempts to segment other structures such as the uterus, rectum and uterosacral region will follow. While an edge based method works well

in segmenting the bladder, a region based method may be more suitable for the bowel wall, since there is no clearly defined boundary between the bowel inner wall and the substance inside. Instead, its texture-like structure inside the bowel can be dealt with by region based method. In addition to individual segmentation of structures, we believe shape models of organs and a model of their relative positions inside the abdomen would help identify adhesive endometriosis (another severe form of endometriosis), which may distort organ locations dramatically.

Bibliography

[1] D. Bartosik, S.L. Jacobs, L.J. Kelly, et al. Endometrial tissue in peritoneal fluid. *Fertility and sterility*, 46(5):796, 1986.

[2] M. Bazot, E. Darai, R. Hourani, I. Thomassin, A. Cortez, S. Uzan, and J.N. Buy. Deep pelvic endometriosis: MR imaging for diagnosis and prediction of extension of disease1. *Radiology*, 232(2):379–389, 2004.

[3] M. Bazot, R. Detchev, A. Cortez, P. Amouyal, S. Uzan, and E. Daraï. Transvaginal sonography and rectal endoscopic sonography for the assessment of pelvic endometriosis: a preliminary comparison. *Human Reproduction*, 18(8):1686–1692, 2003.

[4] A Calderón. Intermediate spaces and interpolation, the complex method. *Studia Mathematica*, 24(2):113–190, 1964.

[5] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International journal of computer vision*, 22(1):61–79, 1997.

[6] T.F. Chan and L.A. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266–277, 2001.

[7] C. Chapron, J.B. Dubuisson, V. Pansini, M. Vieira, A. Fauconnier, H. Barakat, and B. Dousset. Routine clinical examination is not sufficient for diagnosing and locating deeply infiltrating endometriosis. *The Journal of the American Association of Gynecologic Laparoscopists*, 9(2):115–119, 2002.

[8] W.K. Clifford. Applications of Grassmann's extensive algebra. *American Journal of Mathematics*, 1(4):350–358, 1878.

[9] W.K. Clifford. On the classification of geometric algebras. *Macmillan, London*, pages 397–401, 1882.

[10] L.D. Cohen. On active contour models and balloons. *CVGIP: Image understanding*, 53(2):211–218, 1991.

[11] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7):909–996, 1988.

[12] L. Fedele, E. Piazzola, R. Raffaelli, and S. Bianchi. Bladder endometriosis: deep infiltrating endometriosis or adenomyosis? *Fertility and sterility*, 69(5):972–975, 1998.

[13] M. Felsberg and G. Sommer. The monogenic signal. *Signal Processing, IEEE Transactions on*, 49(12):3136–3144, 2001.

[14] D. Gabor. Theory of communication. part 1: The analysis of information. *Electrical Engineers-Part III: Radio and Communication Engineering, Journal of the Institution of*, 93(26):429–441, 1946.

[15] R. Gazvani and A. Templeton. New considerations for the pathogenesis of endometriosis. *International Journal of Gynecology & Obstetrics*, 76(2):117–126, 2002.

[16] R. Gazvani and A. Templeton. Peritoneal environment, cytokines and angiogenesis in the pathophysiology of endometriosis. *Reproduction*, 123(2):217–226, 2002.

[17] G.H. Granlund and H. Knutsson. *Signal processing for computer vision*. Kluwer Academic Pub, 1995.

[18] A. Grossmann and J. Morlet. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM journal on mathematical analysis*, 15(4):723–736, 1984.

[19] S.W. Guo and Y. Wang. The prevalence of endometriosis in women with chronic pelvic pain. *Gynecologic and obstetric investigation*, 62(3):121–130, 2006.

[20] J. Halme, M.G. Hammond, J.F. Hulka, S.G. Raj, L.M. Talbert, et al. Retrograde menstruation in healthy women and in patients with endometriosis. *Obstetrics and gynecology*, 64(2):151, 1984.

[21] D. Hestenes. *Space-time algebra*, volume 1. Gordon and Breach New York, 1966.

[22] D. Hestenes and G. Sobczyk. *Clifford Algebra to Geometric Calculus*, volume 5 of *Fundamental Theories of Physics*. Springer Netherlands, 1984.

[23] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988.

[24] N.G. Kingsbury. The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement. In *Proc. EUSIPCO*, volume 98, pages 319–322, 1998.

[25] K. Kinkel, K.A. Frei, C. Balleyguier, and C. Chapron. Diagnosis of endometriosis with imaging: a review. *European radiology*, 16(2):285–298, 2006.

[26] P.R. Koninckx and D.C. Martin. Deep endometriosis: a consequence of infiltration or retraction or possibly adenomyosis externa? *Fertility and sterility*, 58(5):924–928, 1992.

[27] P. Kovesi. Symmetry and asymmetry from local phase. *Tenth Australian Joint Converence on Artificial Intelligence*, pages 2–4, 1997.

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

[29] S. Lankton and A. Tannenbaum. Localizing region-based active contours. *Image Processing, IEEE Transactions on*, 17(11):2029–2039, 2008.

[30] M.H. Levitt. *Spin dynamics: basics of nuclear magnetic resonance*. Wiley, 2008.

[31] R. Malladi, J.A. Sethian, and B.C. Vemuri. Shape modeling with front propagation: A level set approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(2):158–175, 1995.

[32] S.G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693, 1989.

[33] S.G. Mallat. *A wavelet tour of signal processing*. Academic Press, 1999.

[34] M. Mellor and M. Brady. Phase mutual information as a similarity measure for registration. *Medical Image Analysis*, 9(4):330–343, 2005.

[35] M.C. Morrone and R.A. Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6(5):303–313, 1987.

[36] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of computational physics*, 79(1):12–49, 1988.

[37] K. Rajpoot, V. Grau, and J.A. Noble. Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3-D echocardiography images. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 783–786. IEEE, 2009.

[38] K. Rajpoot, A. Noble, V. Grau, and N.M. Rajpoot. Feature detection from echocardiography images using local phase information. 2008.

[39] K. Rajpoot, J. Noble, V. Grau, C. Szmigielski, and H. Becher. Multiview RT3D echocardiography image fusion. *Functional Imaging and Modeling of the Heart*, pages 134–143, 2009.

[40] J.M. Rawson. Prevalence of endometriosis in asymptomatic women. *The Journal of reproductive medicine*, 36(7):513, 1991.

[41] V.M. Rice. Conventional medical therapies for endometriosis. *Annals of the New York Academy of Sciences*, 955(1):343–352, 2002.

[42] J.A. Sampson. Intestinal adenomas of endometrial type: their importance and their relation to ovarian hematomas of endometrial type (perforating hemorrhagic cysts of the ovary). *Archives of Surgery*, 5(2):217, 1922.

[43] I.W. Selesnick, R.G. Baraniuk, and N.G. Kingsbury. The dual-tree complex wavelet transform. *Signal Processing Magazine, IEEE*, 22(6):123–151, 2005.

[44] J.A. Sethian. *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, volume 3. Cambridge university press, 1999.

[45] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *Medical Imaging, IEEE Transactions on*, 17(1):87–97, 1998.

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

[47] K. Togashi, K. Nishimura, I. Kimura, Y. Tsuda, K. Yamashita, T. Shibata, Y. Nakano, J. Konishi, I. Konishi, and T. Mori. Endometrial cysts: diagnosis with MR imaging. *Radiology*, 180(1):73–78, 1991.

[48] N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, and J.C. Gee. N4ITK: improved N3 bias correction. *Medical Imaging, IEEE Transactions on*, 29(6):1310–1320, 2010.

[49] M. Unser, N. Chenouard, and D. Van De Ville. Steerable pyramids and tight wavelet frames in l2(bbr-d). *Image Processing, IEEE Transactions on*, 20(10):2705–2721, 2011.

[50] M. Unser, D. Sage, and D. Van De Ville. Multiresolution monogenic signal analysis using the Riesz–Laplace wavelet transform. *Image Processing, IEEE Transactions on*, 18(11):2402–2418, 2009.

[51] M. Unser and D. Van De Ville. Wavelet steerability and the higher-order Riesz transform. *Image Processing, IEEE Transactions on*, 19(3):636–652, 2010.

[52] R.F. Valle and J.J. Sciarra. Endometriosis: treatment strategies. *Annals of the New York Academy of Sciences*, 997(1):229–239, 2003.

[53] S. Venkatesh and R.A. Owens. An energy feature detection scheme. In *the international conference on image processing*, pages 553–557. Singapore, 1989.