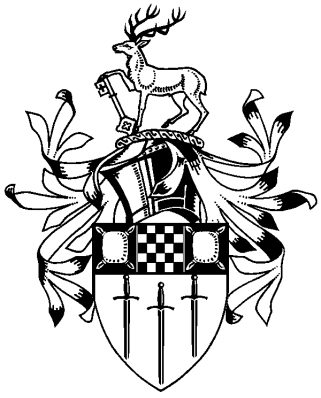
Probabilistic Partial Volume Modelling of Biomedical Tomographic Image Data

John P. Chiverton

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey



Centre for Vision, Speech and Signal Processing
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

August 2006

Summary

The partial volume effect is an imaging artefact associated with tomographic biomedical imaging data. Three-dimensional volumetric data points (voxels) enclose finite sized regions so that they may contain a mixture of signals which are then known as partial volume voxels. The limited spatial resolution of tomographic biomedical imaging data, due to the complex biomedical image acquisition processes, often results in large numbers of these partial volume voxels. Clinical applications of biomedical imaging data often require accurate estimates of tissues or metabolic activity, where many voxels in the data are partial volume voxels. Therefore accurate modelling of the partial volume effect can be very important for such quantitative applications.

The probabilistic models discussed and presented in this thesis provide a generic mathematically consistent framework in which the partial volume effect is modelled. Novel developments include an improved model of an intensity and gradient magnitude feature space to model the PV effect; a novel analytically derived formulation of the ground truth (prior) description of the PV effect; a novel gradient controlled spatially regulated classifier that utilises Markov Chain Monte Carlo simulations; and a fully automatic brain isolation technique that identifies brain voxels in neurological MRI data.

Simulated partial volume data and data from anatomical (MRI) and functional (PET) biomedical imaging modalities are utilized to assess the classification performance of the partial volume models. The data sets include: an imaged PET/CT phantom provided by the Royal Marsden Hospital, UK; publicly available simulated MR brain data together with the associated ground truths from the Montreal Neurological Institute, McGill University, Canada; and 20 normal MR data sets from the Center for Morphometric Analysis at Massachusetts General Hospital, USA. The performance of the developed classifiers were found to be competitive and in some cases superior to existing published quantitative estimation techniques.

Key words: Partial Volume, Biomedical, 3D, Parametric, Probabilistic, Physical Model

Email: j.chiverton@surrey.ac.uk/jpchiverton@theiet.org

WWW: http://www.eps.surrey.ac.uk/

Acknowledgements

I would like to thank the UK Engineering and Physical Sciences Research Council for funding this research. Most of all I would like to thank my supervisor, Kevin Wells for his patience, insight, help, dedication and many other traits that have made this research possible.

My work would not have been possible if many many people hadn't helped along the way. The list is probably endless, but hopefully these are the people that I should thank and acknowledge most of all. Some more than others but in no particular order, except Kevin of course. Kevin Wells (principal supervisor and head of group); Simon Doran (physics); Farzin Mokhtarian (background supervisor); Maria Petrou (previous head of group); Declan Johnson (neuroradiologist); Josef Kittler (head of centre). I would like to acknowledge and thank the people who supported the administration and management of the extensive information and computing infrastructure that is provided by CVSSP. In particular, Simon Alcott, Lee Gregory, Lesley Partridge, Rachel Gartshore and James Field. Thanks are also due to the wider administration and support provided by people in SEPS.

There are probably too many people to mention that have had some form of professional contribution to my work, in particular the people who I have shared an office with, other students who have worked under the same supervisors or similar projects and lecturers who have provided technical and or scientific assistance. Some of these people are: William Christmas; Theo Vlachos; Richard Bowden; Maria Faraklioti; Matt Deighton; Hataikan Porncharoensin; Daniel Rodriguez; Haval Kadhem; Rupin Dalvi; Mohammed Alnafaea; Miriam Barry; Chao Chen; Barbara Podda; Michelle (Qi) Wu; Naghi Jabari-Vasal; George Gregoriou; Natalia Arribas-Merino; Vassili Kovalev; Munir Zaman; Jean-Yves Guillemaut.

I would like to thank Mike Partridge, Bob Ott and Maggie Flower for their help and for the collaborative opportunities that were made available to my supervisor and his students (Miriam Barry and Haval Kadhem) to which I also benefitted. In particular I would like to thank them for the availability of the PET/CT imaging data and their insight and knowledge of the partial volume problem with this type of data. I am thankful to Miriam Barry and Haval Kadhem who were responsible, (as part of their undergraduate and postgraduate project work- supervised by Kevin and Mike), for the experimental acquisition of the PET/CT data which is used in this thesis.

Many thanks are also due to the people who have provided much of the MR medical imaging data that I have used in this thesis and in the papers that we have had published. These include the simulated BrainWeb phantom (MR simulated brain) from the Montreal Neurological Institute, McGill University, Canada and the publicly available 20 normal Internet Brain Segmentation Repository data sets from the Center for Morphometric Analysis at Massachusetts General Hosptial, USA. These publicly available data sets allow researchers across the world to assess the performance of their segmentation and classification algorithms on data sets that others can similarly compare their results.

I would also like to acknowledge and thank people from ISBE, Manchester for interesting discussions at various conferences and for publishing some very interesting work.

I would also like to thank the US National Institute of Health for partial funding of my attendance to IEEE NSS & MIC, Rome, October 2004 and the UK Institute of Physics and Engineering in Medicine for funding attendance at their annual scientific meeting in York, 2004.

Penultimately I would also like to thank the people involved in the examination of this thesis. Notably the examiners, John Illingworth and Neil Thacker for their time and hard work; and the chairperson, Adrian Hilton.

Lastly (but not least), I would also like to thank the dear friends who made CVSSP an enjoyable, life-changing environment to work in.

Dedication

To Ann and Mum.

Contents

Chapter 1

Introduction

The practice of medicine throughout the twentieth century has been greatly assisted by advances in biomedical imaging technologies. A biomedical image captures spatial information displaying some part of a patient's anatomy or their physiology. Currently, the information in a biomedical image is commonly used by a clinician to understand or extract qualitative information about the patient's condition. It is therefore important that the information contained in the image is accurately conveyed to the clinician in these qualitative diagnostic scenarios.

Increasingly, computers are being used to extract quantitative information, rather than, or in addition to, the qualitative information pertaining to the diagnosis and quite possibly the treatment of a patient. Quantitative information can further assist the clinician in the diagnosis of a patient's condition, where quantities of a tissue of interest or the amount of physiological activity in relation to the general population and or some part of the patient's body may be of particular interest. The move from qualitative to quantitative information will assist the clinician in the itemisation, designation and finalisation of a patient's treatment. It is therefore anticipated that this additional information will in many cases improve the prognosis of the patient. This thesis therefore develops probabilistic models that may be used in the extraction of this quantitative information.

The production life-cycle of a biomedical image typically involves acquisition, reconstruction, communication, storage, possibly further processing and finally presentation. At every stage accurate well-informed physical models of the processes involved should be incorporated into any subsequent stage of the life-cycle of a biomedical image. e.g. The reconstruction stage should possess knowledge about the preceding acquisition stage; similarly communication of the image data should not utilize compression algorithms that remove critical information that has been retained by the previous stages that include acquisition and reconstruction. Despite this, each stage of the production life-cycle of a biomedical image is rarely perfect. There are limitations in current knowledge, computational models and technology. These limitations prevent the accurate portrayal of the true biomedical information. Some of these limitations may sometimes be referred to as imaging artefacts. These artefacts may be superficial

annoyances under some circumstances, but other artefacts may affect the diagnostic quality of an image. This is particularly true for computer aided diagnostic applications, where quantitative information is gleaned from the imaging data.

This thesis is primarily concerned with a commonly found artefact known as the Partial Volume (PV) effect. This is addressed via probabilistic models of imaging data and subsequent development of appropriate analysis techniques. The PV effect refers to the possibility that a particular point in a biomedical image may have been sampled from a mixture of tissues or physiological activity concentrations. If a substantial number of data samples exhibit this PV effect, then it is important for a model of the imaging data to incorporate knowledge about such an artefact. This thesis utilizes PV models in probabilistic classification algorithms which attempt to determine the content of a particular data sample. Such PV models and the associated classification algorithms can be used as a means to an end in quantitative applications at the final presentation stage of the imaging life-cycle.

1.1 Imaging Systems

An imaging system is the term used here to refer to a generic system that acquires spatial information about some part of a patient's anatomy or physiology. Exemplar imaging systems may include Magnetic Resonance Imaging (MRI), Computed X-ray Tomography (CT), Single Photon Emission Tomography (SPECT) or Positron Emission Tomography (PET) image acquisition systems.

Some imaging systems initially subject the patient to a process that instigates the emission of a signal from within the patient (e.g. MRI, PET and SPECT). Other modalities, such as CT, are considered as transmission type modalities where X-rays are transmitted through the patient. The X-rays undergo variable amounts of attenuation dependent on the type of tissues through which they pass.

The emission type modalities can be further divided into two types of imaging system. Nuclear imaging modalities, such as PET or SPECT usually require the administration of a radioactive substance so that the subsequent radioactive decay of the substance can be detected as gamma-rays. MRI on the other hand relies on the external magnetisation, excitation and consequent relaxation of magnetic moments inside the patient that are a property of the nuclei in the atoms of the patient's tissues. The physics of the Magnetic Resonance (MR) and PET image acquisition processes are described further in chapter 2.

Due to the dependence on computing technology and the discrete nature of the signal acquisition process, most signals from imaging processes are sampled and stored in a digital format after detection. This process is illustrated in figure 1.1. Each stage of the image acquisition process has an affect on the resultant signal seen as an image. Acquiring information about the real world, particularly one which is not spatially accessible without non-invasive or destructive procedures, is a difficult and complicated task. This is the problem faced by medical imaging procedures, where the internal details of the subject are being imaged and minimal effect on the subject's system is

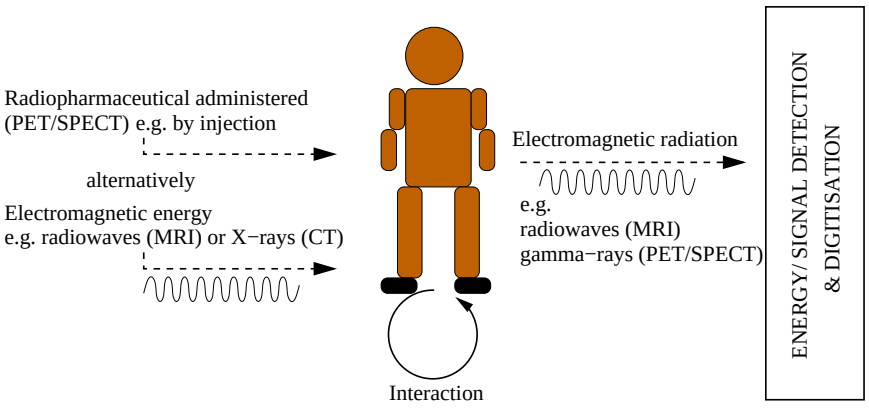


Figure 1.1: Illustration of the gross actions of some imaging systems.

desired. The difficulty of the imaging process is therefore reflected in the complexity of the physics of the devices used to produce the images. These complicated imaging processes produce clinically useful images but with the result of various compromises, limitations and assumptions in models of various physical principles and phenomena. Combining these compromises with limitations in modern technology produces images that are imperfect. These imperfections are often known as artefacts that the observer should be aware of, if they are to provide an informed interpretation of the information in the image. Information regarding exemplar clinical applications of medical imaging technology can be seen in chapter 2, section 2.4.

1.1.1 The Nature of the Partial Volume Effect

The final stage of an image acquisition process involves the detection and digitisation of the image signal (illustrated in figure 1.1). A finite number of three-dimensional (3-D) volumetric data points, (i.e. voxels), are utilised to represent the continuous information in the true object. The signal representing this information will have been subjected to a number of processes that predominantly only allow lower frequencies to pass, thus limiting the amount of higher frequency information. High frequency spatial information enables accurate portrayal of edges and other fine detailed image structures. Similarly, the digitisation process has to capture a limited number of voxels and hence limit the higher spatial frequencies, otherwise infinite amounts of computer memory would be required to store the images.

A concept known as the Point Spread Function (PSF) is often used to describe imaging and signal processing systems as it characterises a system's response to a point source. Once the PSF is known, responses to any other signal can be determined via mathematical operations such as convolution, see e.g. [46]. A Gaussian function is a typical imaging system PSF, where the signal at the centre of the PSF has the greatest weight, whilst signal from the surrounding areas also partially contribute toward the signal at the centre.

Therefore the PSF can be described as blurring the high frequency information, reducing the high frequency content and allowing lower frequencies to pass. This spreading in

the digital domain typically carries across multiple voxels. The wider the PSF the more the imaging information about an object will be spread. It is therefore logical to select a voxel's dimensions based on the width of a PSF. This helps to reduce redundancy in representation of the imaging data and to quantify the resolution of the imaging process. Fewer voxels are required to represent lower resolutions. Indeed, Haacke *et al.* [50] (for MRI) state that the optimal size of a voxel can be approximated by the Full-Width at Half Maximum (FWHM) of the PSF[1]. The FWHM is also a common measure of image resolution for other imaging modalities such as PET and SPECT, see e.g. [130].

The action of a PSF on a step edge is illustrated in figure 1.2. This figure also illustrates the resultant voxel intensity values, where voxels located over the step edge take values in between the continuous world ideal values. The result of this process is referred to as the Partial Volume (PV) effect.

The PV effect is therefore directly related to the amount of spreading induced by the action of the PSF of the image acquisition process. The PV effect describes an artefact that is common to all systems that attempt to represent a continuous signal with a finite number of samples, although it is described by various different terms for different fields of application. A simplified illustration of the PV effect in an image is given in figure 1.3.

Factors that contribute to more numerous PV voxels include the inherently 3-D nature of the medical image acquisition processes. A typical modern medical image possesses what is known as a "*slice thickness*", where PV voxels are not only affected by the in-slice PSF action, but also the inter-slice action of the 3-D PSF. It should be noted that early attempts at image quantification were limited to two-dimensional (2-D) analysis as the slice thickness was large (w.r.t. the x,y dimensions) and sometimes the slices possessed large inter-slice gaps, i.e. the slices were non-contiguous.

The effect that PV voxels have on the statistical properties of imaging data is to reduce the possible independence of statistical classes of interest and merge them into (conceptually) a single entity. For example, a statistical class in a neurological MR imaging application may include a particular range of intensities that might be typical for White Matter (WM). Similarly, a different range of intensities could be used to describe the Grey Matter (GM) voxels. If the PV effect was not present (i.e. an idealistic imaging device), then these classes would possess quite distinct intensities (ignoring noise). But due to factors such as noise and the PV effect, the boundary between the GM and WM intensities is usually quite ambiguous. A particular intensity may then originate from GM, WM or a mixture of the two classes. This is also true of functional imaging data where biological activity rather than morphological information is being imaged. For example, radionuclide based functional imaging typically produces inferior resolution data (larger voxel size, e.g. $4 \times 4 \times 4\text{mm}^3$) in comparison to MRI data (e.g. voxel dimensions of $1 \times 1 \times 2\text{mm}^3$) due in part, to the physical limitations of the signal measurement process described in section 2.2. In such cases larger voxel sizes at relatively the same scale produce even greater PV voxel populations.

[1] FWHM is an alternative measure of the width of a function such as a Gaussian function half way from the maximum point.

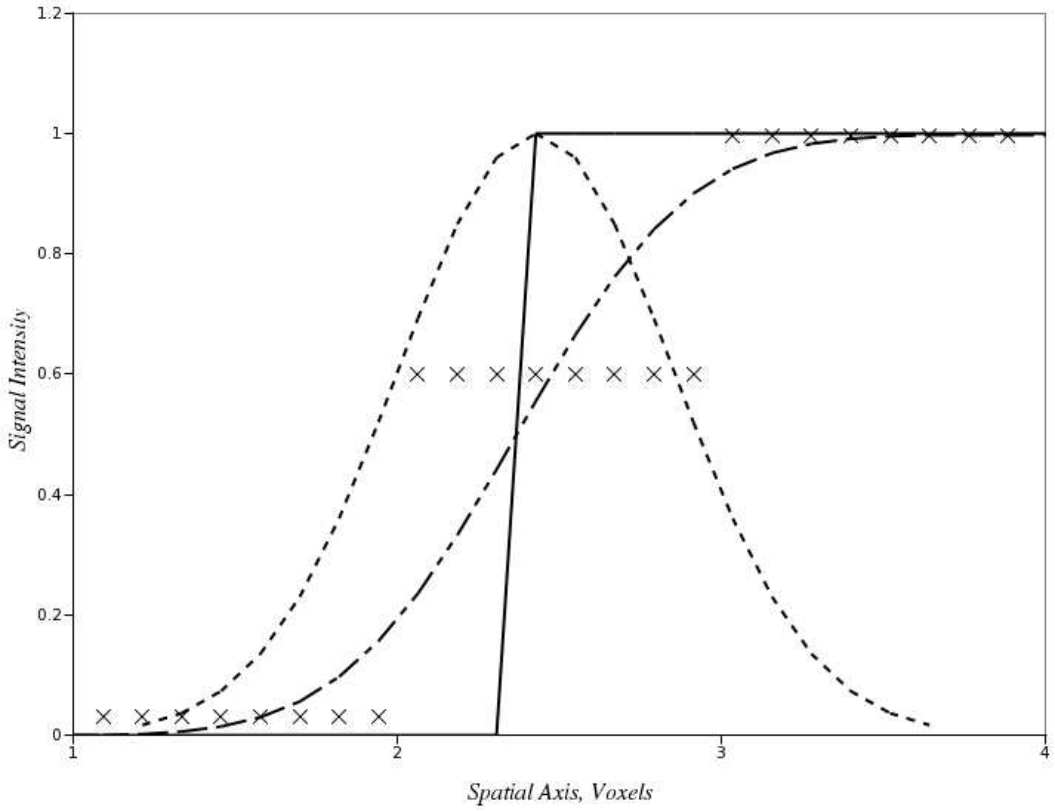


Figure 1.2: Illustration of the action of a PSF (dotted-and not to scale) on a step edge (solid line), producing PV voxels. The convolution result is the smoothed edge response (dot-dashed line). The sampling process produces the discrete voxel values which change at regular intervals so that the signal is no longer continuous on the spatial axis (represented as $\times$). These discrete levels occupy signal intensity values in between the original true signal levels. i.e. the original 0.0 signal intensity (prior to acquisition) is now represented by 0.025 intensity for voxel indices 1.0 to 2.0, 0.6 intensity for voxel indices 2.0 to 2.3 and the original 1.0 signal intensity is now represented by 0.6 intensity for voxel indices 2.3 to 3.0 and 1.0 for voxel indices 3.0 to 4.0. Therefore voxels with indices 1.0 to 3.0 can be referred to as PV voxels. Note the original continuous edge occurs at approximately voxel index 2.3, thus producing PV voxels with indices from 1.0 to 3.0.

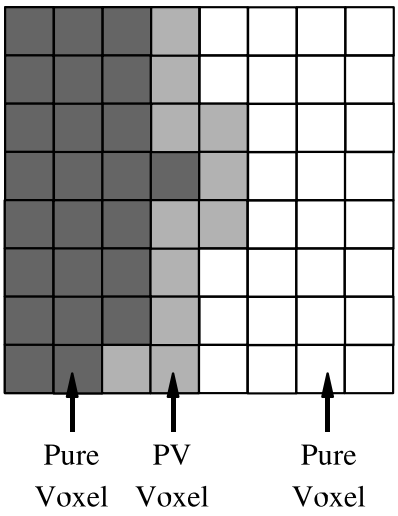


Figure 1.3: Illustration of the PV effect in image data.

The PV effect is therefore a significant factor in medical imaging data and quantitative estimates of tissues or physiological activities may therefore benefit by including modelling of the PV effect.

1.2 Contribution

The work in this thesis undertaken by the author has contributed to a number of developments. In particular these include:

- A PV mixing model that explicitly models the per voxel PV mixing as random vectors has been formulated utilising point estimates based on the expected mixing value for a given intensity. This has been found to be equivalent to an existing statistical model, [148], derived via an alternative interpretation of the probabilities, [19]. A lower bound on the voxel RMS classification error for intensity based PV classification has also been established. These developments are mainly presented in chapter 5 and partially published in [19].

- An existing 2-D PV model utilising intensity and gradient magnitude to identify likely PV voxels has been extended by improving and reformulating the PV model, [19], and reformulating the likelihood to model 3-D data, [18, 23]. These developments are mainly presented in chapter 6 and published in [18, 23, 19].

- The shape of global PV mixing priors has been demonstrated (analytically) to be invariant to the width of the PSF of the imaging system, [21]. This development is presented in chapter 7.

- A phenomenological law, known as Benford's Law has been related to the PV mixing process, [21, 152, 153]. It was found that this logarithmic law that describes the frequency of digits in arbitrary sources of information can also be

observed in the frequency of discretized PV mixing values. These developments are presented in chapter 7 and published in [21, 152, 153].

- A novel gradient controlled spatially regulated model of the PV effect is also proposed and found to achieve improved performance over all other models in this thesis. Comparison with other high performance classifiers via classification of publicly available MR brain data sets (with expert defined ground truths) was also undertaken and found to be competitive and in some instances superior to these alternative classifiers, [20]. These developments are published in [20] and presented in chapters 7 and 8.

- A fully automatic skull-stripping algorithm has been developed and tested on a number of MRI neurological data sets, [22, 17]. This technique was found to be as successful at removing non-Central Nervous System (CNS) tissues from image data as a commonly used existing skull-stripper (Brain Extraction Tool, BET, [131]). The automation of the skull-stripping algorithm required coarse estimates of the brain tissue intensity parameters, which were provided with a novel self-similarity measure. The measure identifies the dominant sample population (i.e. neurological tissues) by comparing the global histogram with individual slice histograms. Slice histograms most similar to the global histogram were found to contain the highest number of neurological tissue voxels. These developments are presented in chapter 3 and published in [22, 17].

1.3 Thesis Structure

This chapter has introduced the concept of an imaging system. The main topic of this thesis, namely the PV effect, has also been introduced using this imaging framework. The next chapter (2) goes on to describe two specific biomedical imaging modalities, MRI and PET, representative of the two main classes of biomedical imaging techniques: namely anatomical (morphological) and physiological (functional) imaging. This therefore provides an opportunity for a modality specific discussion of the PV effect. These two modalities provide state of the art morphological and functional information. The models developed within this thesis have therefore been tested using these two modalities to demonstrate their application across an anatomical and functional imaging context.

Chapter 3, published in [22, 17], focuses on a novel pre-processing step for neurological MRI data referred to here as skull-stripping. This represents the first piece of novel work developed and published by the author. This step is often seen as an essential first step prior to further processing of neurological MRI data. Skull-stripping is often used prior to classification or other neurological image processing steps (such as might be found in PET-based brain studies) to remove irrelevant non-CNS tissues. Therefore the number of classification classes that have to be incorporated in a subsequent classifier framework is reduced. A reduction in the number of classification classes relieves part of the computational burden and reduces uncertainty in the calculation of probabilities for individual classification classes that may share similar intensity values with non-CNS tissue classes.

Chapters 4 presents a review of existing PV classifiers, introducing some of the different approaches that have been taken to model and thereby quantify imaging data affected by the PV effect. This leads to the next three chapters, 5, 6, 7 that describe new PV models, and thus represent the main component of novelty in this thesis.

Chapter 5 focuses on intensity based PV models that utilize global PV mixing prior densities. Two popular core models that utilize the probabilities as analogous to the PV content of a voxel are described. A new PV model is then presented which models the PV mixture content as random vectors that can be explicitly incorporated into a probabilistic formulation. These three PV models are then used to classify simulated PV data, thus providing quantitative performance assessment.

The performance of these global mixing prior intensity based PV models is found to be limited for low Contrast to Noise Ratio (CNR) values. Therefore chapter 6 extends the intensity based feature space to include gradient magnitude. The first combined intensity and gradient magnitude PV feature space, [157], previously recommended to improve PV classifier performance, is described. This existing PV model is limited to a parametric 2-D gradient magnitude formulation and uses the analogical probability to PV content model. This chapter therefore introduces two further developments, firstly to introduce a novel intensity gradient magnitude likelihood, published in [19], and secondly to extend the parametric gradient magnitude formulation to 3-D, published in [18, 23]. These three intensity and gradient magnitude feature space PV models are then used to classify (using global PV mixing priors) simulated two class PV data. Although somewhat improved over the intensity based classifiers, the performances are still found to be limited for low CNR values. Therefore a third development is required to improve the performance of the PV classifiers further.

The PV models of chapters 5 and 6 utilise global PV mixing priors (implicitly for the analogous models). Chapter 7 extends the PV models to include a local prior instead of the global mixing prior. This local prior enables the spatial contextual information to be included in the classification process. The gradient magnitude is used to dynamically control the amount of regularisation, unlike the previous formulations that used the gradient magnitude simply to identify likely PV voxels, and is published in [20]. Performance assessment on the simulated PV data demonstrates this new formulation possesses improved performance on low CNR data. Therefore it is applied to further PV data, simulated and real, to assess its performance under more realistic conditions.

Chapter 8 introduces three sets of experiments. The local PV mixing prior model of the preceding chapter is assessed along with the most promising PV models from chapters 5 and 6. These experimental tests include classification of a series of simulated MRI brain data sets with varying amounts of noise. The simulated MRI brain data is available on line from McGill University Montreal Neurological Institute, [12]. These experiments determine the performance of the classifiers under more realistic conditions with particular reference to the classification of neurological MRI data. Two further sets of experiments are performed. Image data of a PET/CT phantom were provided, details of which can be seen in [5]. This PET phantom enables the performance of the classifiers to be assessed on real experimental PET imaging data. The high resolution CT data together with the PSF of the PET imaging process is used to derive a ground

truth representation of the PET imaging data. The PET data is then classified and the ground truth data is used to assess the performance of the classifiers. These experiments on real PET imaging data allow the performance of the classifiers to be determined and therefore implicitly test the validity of the PV model assumptions on real PET imaging data.

The final set of experiments apply the best performing classifier (using the local mixing prior) to 20 real neurological MRI data sets, assessed using ground truth image data available from the Center for Morphometric Analysis at Massachusetts General Hospital, [38]. The classification of these data sets extend the earlier experiments on the simulated MR brain data to conditions which are difficult to simulate such as severe MR imaging artefacts including gradual and sudden changing intensity inhomogeneities.

Chapter 2

Medical Imaging Background

The preceding chapter discussed the motivation for modelling of the Partial Volume (PV) effect with regard to a generic imaging system. This chapter presents some further information about anatomical and functional imaging systems, namely MRI and PET. This background information then leads on to possible clinical applications of the type of classifiers and PV models that are discussed in later chapters.

2.1 Anatomical Imaging with Magnetic Resonance

MRI is a popular anatomical imaging modality due to the fact that MR images possess high specificity and sensitivity to variations in soft tissue. Delineating small differences between soft tissues in the human body is particularly important for many clinical applications such as neurology where anatomical abnormalities are often indicators of particular neurological disorders or diseases. The PV effect is of particular importance with MRI data due to the complex anatomical structures being imaged and the clinical applications that require quantitative information to be extracted from these images. Therefore this thesis considers the PV effect with particular application to the MRI modality. Further information with regard to clinical applications will be discussed shortly, but first the physical principles of the MRI acquisition process are presented.

An MRI scanner utilizes electromagnetic subatomic processes to determine properties of the tissues at finite size volumetric locations (voxels) within the imaging subject. An explanation of these subatomic processes is now given.

2.1.1 Atomic Structure and External Magnetic Fields

An atom consists of a nucleus and electrons that surround the nucleus, where the nucleus consists of neutrons and protons. Each proton possesses an electric charge and is equal in magnitude but opposite in polarity to the charge of an electron. Particles are often described as possessing spin, an intrinsic property which is a convenient physical model.

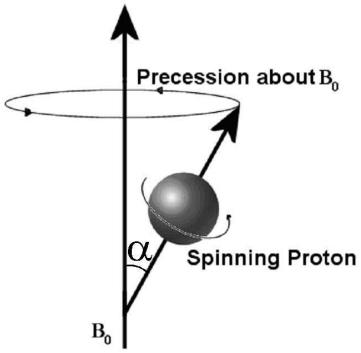


Figure 2.1: Illustration of a proton precessing about an externally applied magnetic field, B_0.

The protons also possess an intrinsic magnetic moment. The particles are therefore also affected by externally applied magnetic fields. A strong externally applied magnetic field, B_0, applied by the MRI scanner changes the direction of the spin of the protons so that they *precess* about the externally applied field. B_0 is said to *align* the magnetic moments of the protons. This is illustrated in figure 2.1.

There are two possible energy states for a proton that is precessing about B_0, a low energy state, E_1 or a higher energy state, E_2. The lower energy state is said to be where the protons are parallel with the magnetic field. The higher energy state is said to be where the protons are anti-parallel with the magnetic field. It is possible for the lower energy state protons to be excited or *kicked* into a higher energy state with the application of a radio-frequency magnetic field, B_1 that provides additional energy pushing the lower energy state protons into the higher energy state. This is illustrated in figure 2.2.

The change in energy, ΔE, is dependent on the frequency of precession of the protons, ω,

$$\Delta E = h.\omega, \tag{2.1}$$

where h is known as Planck's constant. The frequency of precession is given by

$$\omega = B_0.\gamma. \tag{2.2}$$

γ is known as the gyromagnetic ratio, the value of which is dependent on the environment of the protons being excited. For MRI, the human body possesses a large amount of water, a compound composed of hydrogen and oxygen atoms. Therefore MRI scanners are optimised so that they can easily detect small variations in the amount of hydrogen. Thus, in equations used to determine imaging parameters, γ is optimally given a value that corresponds to a hydrogen molecular environment.

2.1.2 Relaxation

The precession of the proton can be expressed using two vectors, the longitudinal vector, $\mathbf{M}_l$ and the transverse vector, $\mathbf{M}_t$. The longitudinal vector corresponds with the axis

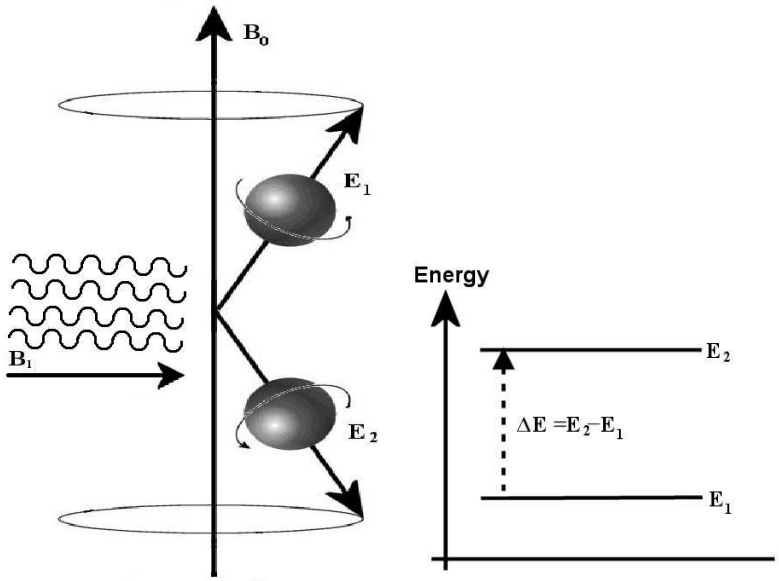


Figure 2.2: Illustration of high (E_2/anti-parallel) and low (E_1/parallel) energy states for a proton in a homogeneous externally applied magnetic field, B_0 and an oscillating magnetic field, B_1.

of precession (parallel with B_0) while the transverse vector is perpendicular to B_0, rotating at the frequency of precession. These vectors are illustrated in figure 2.3.

When B_1 is applied to the object being imaged, the magnitude of the longitudinal magnetisation vector, $\mathbf{M}_l$ decreases due to the subtractive nature of the net magnetisation of the excited nuclei in combination with the larger applied magnetic field, B_0. Prior to the application of B_1, $\mathbf{M}_t = 0$, due to the net magnetisation of the protons spinning out of phase. When the B_1 pulse has finished, the vectors return to their prior state at an exponential rate. The longitudinal relaxation exponential time constant is known as the T_1 recovery time and the transverse decay exponential time constant is known as the T_2 decay time.

The T_1 recovery time can be attributed to dependence on interactions with the surrounding molecular structure (the lattice). This longitudinal relaxation is primarily a result of the exchange of thermal energy with adjacent molecules; hence T_1 relaxation is due to spin-lattice interaction. With T_2 relaxation, the protons de-phase because of the influence of the surrounding local magnetic fields that are also de-phasing; hence T_2 relaxation is due to spin-spin interactions.

Table 2.1 lists some exemplar relaxation T_1 and T_2 time constants for a variety of common tissues imaged with MR.

External field inhomogeneities can also cause additional de-phasing of the transverse magnetisation vectors. Thus, the term $T_2{}^*$ refers to the commonly used time constant associated with the combined de-phasing from the external field inhomogeneities and the de-phasing associated with the spin-spin interaction.

The T_1 and T_2 relaxation times characterise the tissues being imaged together with the

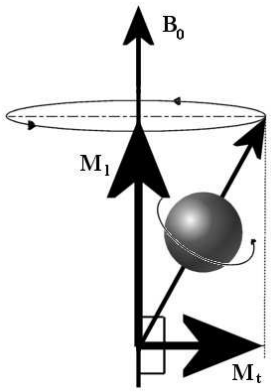


Figure 2.3: Illustration of the magnetisation vectors for an individual magnetic moment. The Transverse magnetisation vector is $\mathbf{M}_t$ and the longitudinal magnetisation vector is $\mathbf{M}_l$.

Table 2.1: Some exemplar MRI T_1 and T_2 relaxation time constant values for a variety of commonly imaged *in vivo* human tissues. Times quoted in milliseconds (ms) with a field strength, $B_0 = 1.5$ Tesla. Values taken from [50].

Tissue	T_1 **(ms)**	T_2 **(ms)**
Grey Matter	950	100
White Matter	600	80
Muscle	900	50
Cerebro-Spinal Fluid	4500	2200
Fat	250	60

density of hydrogen protons. Therefore image acquisition parameters, referred to here as image sequences are often optimised so that either T_1, T_2 or the density of protons become the distinguishing feature of the tissues found in a particular MRI acquisition. The density of the protons refers to the relative population of protons for a fixed volume of matter.

2.1.3 A Description of the MRI Signal

The signal for a particular voxel can now be seen to be dependent on the T_1, T_2 and proton density of the tissue located in a particular voxel, together with the applied magnetic fields, B_0 and B_1. Equations that describe the resulting signal are now described briefly, the description closely follows that of Liang and Lauterbur [82].

The relaxation processes of the transverse and longitudinal components can be described in terms of their magnetisations. For the transverse magnetisation using a rotating frame of reference denoted by $x'y'$[1].

$$M_{x'y'}(t) = M^{+}_{x'y'} \exp\left(-\frac{t}{T_2}\right); \tag{2.3}$$

and for the longitudinal magnetisation

$$M_{z'}(t) = M^{0}_{z}\left(1 - \exp\left(-\frac{t}{T_1}\right)\right) + M^{+}_{z'} \exp\left(-\frac{t}{T_1}\right), \tag{2.4}$$

again using a rotating frame of reference, denoted by z'. $M^{+}_{x'y'}$ and $M^{+}_{z'}$ are the transverse and longitudinal magnetisations respectively immediately after the application of the oscillating magnetic field, B_1. Similarly, M^{0}_{z} is the longitudinal magnetisation at thermal equilibrium. So that $M_{x'y'}(t = T_2) \approx 0.37.M^{+}_{x'y'}$ and $M_{z'}(t = T_1) \approx 0.63.M^{0}_{z'}$.

Equations 2.3 and 2.4 describe the relaxation of the magnetisation after an oscillating magnetic field has been applied. These equations can be used to determine the characteristics of the resulting image signal after the application of a variable number of variable duration oscillating magnetic fields, (B_1).

There are many different *image sequences* or combinations of different magnetic pulses characterised, in part, by an angle known as the flip angle. The flip angle denotes the angle in the rotating frame of reference to which the net magnetisation vector passes during the application of B_1. The flip angle, θ_{flip}, can be calculated from the product of the gyromagnetic ratio, γ, the oscillating magnetic field and the amount of time that the oscillating magnetic field is applied, t_{pulse}:

$$\theta_{flip} = \gamma.B_1.t_{pulse}. \tag{2.5}$$

One such image sequence is the **Spin Echo** imaging sequence. The Spin Echo sequence is characterised by the application of two pulses, the first being $\theta^{1^{st}}_{flip} = 90^o$ to place

[1]The rotating frame of reference is used to reduce the complexity of conceiving the magnetisation vector direction. In the real world frame of reference, the magnetisation vector is continually precessing. If a frame of reference is used which rotates at the same frequency as the precession, the complexity of calculations and mentally visualising the position of the magnetisation vector is reduced.

the protons in a high energy state; and the second being equal to $\theta_{flip}^{2^{nd}} = 180^o$ which helps to overcome effects from external magnetic field inhomogeneities [55]. The Free Induction Decay (FID) signal from a Spin Echo sequence can be described by (see e.g. [82, 88])

$$S_{SE} \propto \rho_{proton}. \left[1 - \exp\left(-\frac{T_R}{T_1}\right)\right] . \exp\left(-\frac{T_E}{T_2}\right); \tag{2.6}$$

where ρ_{proton} is the proton density. The echo delay time, T_E, is the time between the application of a pulse and the measurement of the *echo* from the signal. The repetition time, T_R is the time interval between successive pulses. The echo and repetition times are parameters that are controlled by design of a particular imaging sequence and are optimised so as to produce improved and desirable image properties.

2.1.4 MRI Signal Localisation

Slice Selection The MRI signal can be localised to an individual slice (any axis) of the imaging subject by varying the static magnetic field in steps along the slice axis of the imaging space. The result of this produces variations in the resonant frequency of each slice, therefore allowing the signal to be localised to individual slices.

Voxel Selection and Signal Detection Localisation of the signal within a slice requires two additional small gradients that vary along the width and height of the slice. The x-axis of the slice has a similar graduation of static magnetic field variation affecting the resultant resonant frequency. While the y-axis receives a graduated field for a short period of time that encodes phase information by forcing the protons to de-phase momentarily.

Signal Detection and Transformation of Imaging Data The precessing magnetisation is detected either by the same RF coils that create the oscillating magnetic field, B_1, or separate RF coils are used. Once the frequency encoded data have been collected, the data are said to exist in frequency space, $F(u, v)$. This frequency space is also known as 'k-space'. An inverse Fourier transform converts the data from frequency space to spatial space, $f(x, y)$. The result of such an operation is a vector with two elements for each voxel corresponding to the real and imaginary components, $a + j.b$. These two components for every voxel can be used to calculate a magnitude image that corresponds to the magnitude of the signal ($\sqrt{a^2 + b^2}$). The magnitude image is commonly used in clinical application for diagnostic purposes. Similarly a phase image can also be computed ($\tan^{-1}(b/a)$), although this is not as commonly used in clinical applications. Different types of phase images are used for imaging movement such as might be found in angiographic applications (imaging of blood).

2.1.5 MR Imaging Artefacts

The artefacts in the MR imaging process are not limited to PV type artefacts. Here are a few further artefacts that may be associated with a MRI acquisition process.

Chemical shift The frequency encode process can produce localisation errors which are often referred to as a Chemical shift artefact. For example, as fat and water possess different resonant (Larmor) frequencies, localising their signals in the frequency encode direction becomes problematic and therefore the scanner is not able to accurately identify their positions relative to each other. A second type of chemical shift artefact sometimes known as a black boundary artefact or chemical shift of the second kind appears as a black outline. This is due to voxels in the imaging subject being composed of both fat and water. Under some imaging conditions the variable frequency and hence decay of the different molecules produces a phase difference that can result, depending on the echo time, in a reduction of the NMR signal due to the signals from the fat and water possessing phase differences therefore cancelling each other out.

Gibbs ringing This occurs at high-contrast boundaries and is associated with difficulties in representing the high frequency information in a digital form. The effect can be reduced by increasing the resolution or by filtering to reduce the high frequency information prior to the inverse Fourier transform.

Susceptibility and metal artefacts Metal produces a loss of signal with a high intensity boundary and sometimes geometric distortions in the region surrounding the metal. Smaller susceptibility artefacts may occur at the boundary between tissues due to the different susceptibility of the tissues. This may result in a loss of signal in the boundary region. Most soft tissues however, such as Grey Matter (GM) and White Matter (WM) possess similar magnetic susceptibility [88] and would therefore not be significantly affected by such an artefact.

Field inhomogeneities Variations in the magnetic fields (B_0 and B_1), often produce variations in the signal intensities across the image. Small variations are difficult to remove completely at time of manufacture due to physical limitations and nearby ferromagnetic structures may reduce the homogeneity of the magnetic fields.

Motion artefacts These may be due to patient movement, such as movement of the head while being scanned, breathing artefacts, or flow effects from blood or other fluid in the body. Motion artefacts quite often result in ghosting where part of the anatomy can be (partially) observed in more than one place in the image data.

Other artefacts exist, further details of which can be found in text books such as [88]. These artefacts are quite often dependent on the patient, the anatomy being imaged and or the scanner. Careful control of the imaging conditions help to prevent some of the artefacts from dominating the critical information in the imaging data. For example, using an alternative scanning sequence for images with susceptibility artefacts or requesting the patient to limit their movement during the scanning procedure to limit motion artefacts.

2.1.6 MRI Resolution, PSF and PV Effect

The MRI data acquisition process produces data that is sampled or digital in form. As with any real-world digitisation processing of a continuous signal, the MRI image acquisition process is not capable of recording infinitely small samples, i.e. a finite volume is associated with each sample. The number of samples or sampling frequency is also limited. This is equivalent to a bandwidth being associated with the acquisition process, where frequencies above a certain cut-off point are not faithfully represented in the resulting digital images.

A sophisticated image acquisition process such as MRI has many contributing factors to limitations in the size of a voxel (resolution). Most of these factors can be described by a form of filtering process that reduces the higher-frequency content. These include

- Filtering operations ("artificially") introduced by design into the scanner such as:
  - A Gibbs filter to overcome a Fourier imaging artefact known as Gibbs ringing;
  - Analog filters to reduce noise prior to digitisation;
  - Cross-talk minimisation procedures (to minimize adjacent slices being 'excited' simultaneously, instead of a single slice, thereby minimising inter-slice interference).
- Filtering operations inherent to the MR image acquisition process including:
  - Transverse or T_2 relaxation decay envelope;
  - Frequency bandwidth on slice select gradient magnetic fields.

Attempts have been made to derive analytical expressions for various filtering effects. In particular Haacke *et al.*, [50], derive expressions describing the PSF due to the transverse relaxation decay envelope for gradient echo and spin echo imaging sequences. It is possible to combine these expressions with other filtering process expressions to produce a gross PSF estimate. Indeed one could hypothesise a simulation system that would provide a means to determine the PSF for particular scanners, image sequences, and conditions. However such a system would require substantial experimental confirmation of the PSF estimates.

Factors that limit the spatial resolution are not limited to filtering processes of the acquisition. The amount of noise in relation to the signal increases for smaller voxel sizes thereby reducing the benefit of higher resolutions. One should also note that the the amount of noise is highly dependent on imaging parameters that also contribute to filtering of the MR signal (and consequently the resultant PSF). The amount of noise in relation to mean signal is usually quantified using Signal to Noise Ratio (SNR), discussed shortly in section 2.3.

According to Haacke *et al.*, [50], the optimal resolution is

> "*...the smallest resolvable distance between two different objects, or two different features of the same object.*"

The PSF blurs the high frequency information, reducing the high frequency content. The action of the PSF carries across a number of voxels, as previously illustrated in figure 1.2. Therefore the optimal resolution of digital imaging data is dependent on the width of the PSF. Haacke *et al.* go on to state that the FWHM of the PSF can be used as a measure of the optimal resolution.

The PV effect is directly related to the amount of spreading induced by the action of the PSF of the MR image acquisition process. As discussed in chapter 1, the PV effect describes an artefact that is common to all systems that attempt to represent a continuous signal with a finite number of samples. While the gross action of the PSF in MR imaging is very similar to the PV effect seen in other imaging modalities, there are particular nuances that might affect the applicability of generic PV models.

On the most elementary level, most clinical MR image data is calculated from the magnitude of complex data values (see section 2.1.4). This does affect the way in which particular voxel constituents are represented as particular intensities, i.e. in a non-linear fashion. In particular assuming Gaussianity of the noise signals, the standard deviations of such mixtures may not be easily predicted. Similarly, the mixing process inherent to the action of the PSF is highly dependent on not only the proton density, but also the transverse and longitudinal relaxation times, which vary depending on the constituents of a voxel. A quick glance at a MR signal equation such as the Spin Echo expression given in equation 2.6 possibly indicates that the actual intensity and noise terms for particular voxel constituents would be very difficult to calculate analytically. Combining these difficulties with the numerous possible imaging conditions (e.g. different tissues) with the numerous different image sequences and imaging parameters would suggest any rigorous analysis would have to be very specific to a particular MR imaging application. It should also be noted that previous authors have found a "linear mixing" assumption to be valid based on phantom experiments, see [159].

Some models of the PV effect do not explicitly incorporate linear or non-linear mixing of the random variables. They may assume linear changes in intensity, but the variance terms are ignored, see e.g. [148]. These models appear to possess similar performance to models that explicitly utilize a linear mixing model, see e.g. chapter 5. It is therefore assumed that a linear mixing function provides a sufficiently good approximation to the mixing behaviour observed in MR images in relation to the CNR values that are found for such images.

2.2 Functional Imaging with PET

The preceding discussion was limited to the physics and signal properties associated with anatomical MRI, where tissue structures and anatomical detail is imaged. A functional imaging process, on the other hand, is primarily interested in capturing information about some sort of physiological activity, e.g. tracking the distribution of a particular type of molecule in the human body. Functional imaging can be undertaken with MRI, referred to as fMRI, which attempts to image blood oxygenation levels or the movement of a contrast medium through the body. Unfortunately fMRI suffers from

difficult to correct artefacts, large spatial localisation margins and a lack of absolute quantitative information about the activity being imaged. An alternative functional imaging technique known as PET provides higher sensitivity due to the use of positron emitting radioisotopes where pico-molar quantities can be detected. No other *in vivo* imaging technique can match this performance. Unfortunately PET images are affected by the PV effect due to the relatively large PSF associated with the imaging process. Therefore PET imaging data is also used to illustrate the application of the PV models developed in this thesis.

Images generated from PET image acquisition processes are generated with a need to quantify biological activity associated with a particular function in the human body. PET is a substantially different imaging modality in comparison to MRI. The most obvious difference is in the origin of the energy or signal measured by the scanner. MRI relies on exciting protons with an external magnetic field, where as PET requires the imaging subject to receive a radio-labelled pharmaceutical. A radio-labelled pharmaceutical is composed of a tracer and a radioisotope or radioactive label. The tracer is a compound that is either a target molecule to be studied, or an analogous compound (analogue) which behaves, at least partially, like the target molecule of interest. The radioisotope is an isotope with an unstable nucleus that, for PET, emits positrons. According to [130], the most commonly used radio-pharmaceutical is [^{18}F]fluoro-2-deoxy-D-glucose or ^{18}F FDG (FDG for short). FDG is a glucose analogue, i.e. it behaves like glucose, except that it remains trapped in the body's cells. This trapping is useful for imaging as the FDG accumulates therefore providing a measurable signal to be imaged. Its main use is in oncology and cardiology and has a half-life of 1.87 hours.

Once administered, the radioisotope continues to decay from within the physiology of the patient, emitting positrons. Due to the fact that positrons are the antiparticles of electrons, the emitted positrons interact with electrons in the surrounding matter forming a loosely bound state positronium. The positronium has a very short half-life and this interaction results in energy in the form of a pair of coincident photons or gamma rays travelling in almost anti-parallel (opposite) directions. This is sometimes referred to as an annihilation event of a positron and electron pair. This is illustrated in figure 2.4.

2.2.1 Gamma Ray Detection

The emitted gamma rays are then detected by gamma ray detectors that surround the patient (illustrated in figure 2.5). As these gamma rays are expected to be emitted in pairs, in opposite directions, the acquisition system can filter out any that do not occur in a pair. This helps to provide more accurate information about the spatial location of the decaying radioisotope from within the patient.

2.2.2 Photon Count Localisation

Many coincident event photons are detected by the crystals in the gamma ray detectors. Each pair of photons detected contribute toward a photon count that can then be used to determine the relative amount of activity in particular areas of the patient.

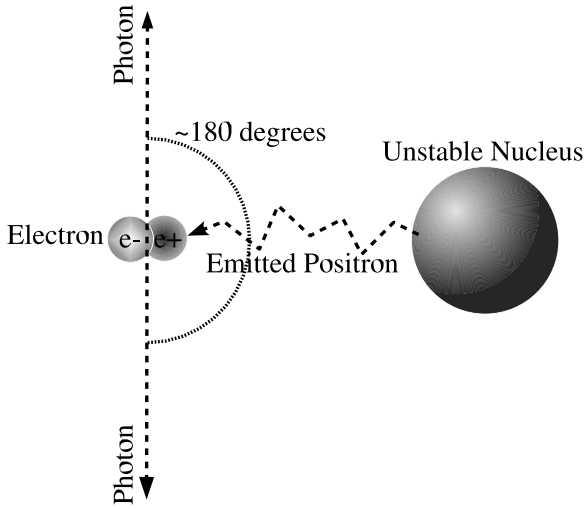


Figure 2.4: Illustration of the emission of a positron from an unstable nucleus, resulting in positronium (composed of an electron and a positron). The positronium annihilates very quickly resulting in two photons that can then be detected.

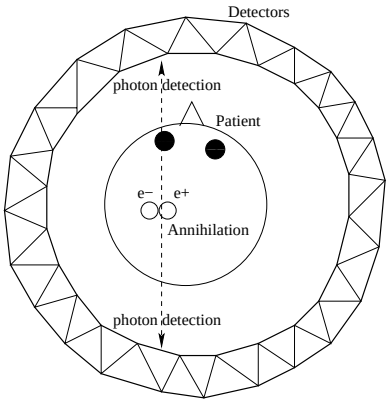


Figure 2.5: Illustration of the role of the gamma ray detectors used to detect the nearly coincident photons emitted from an annihilation event that has taken place inside the patient being imaged.

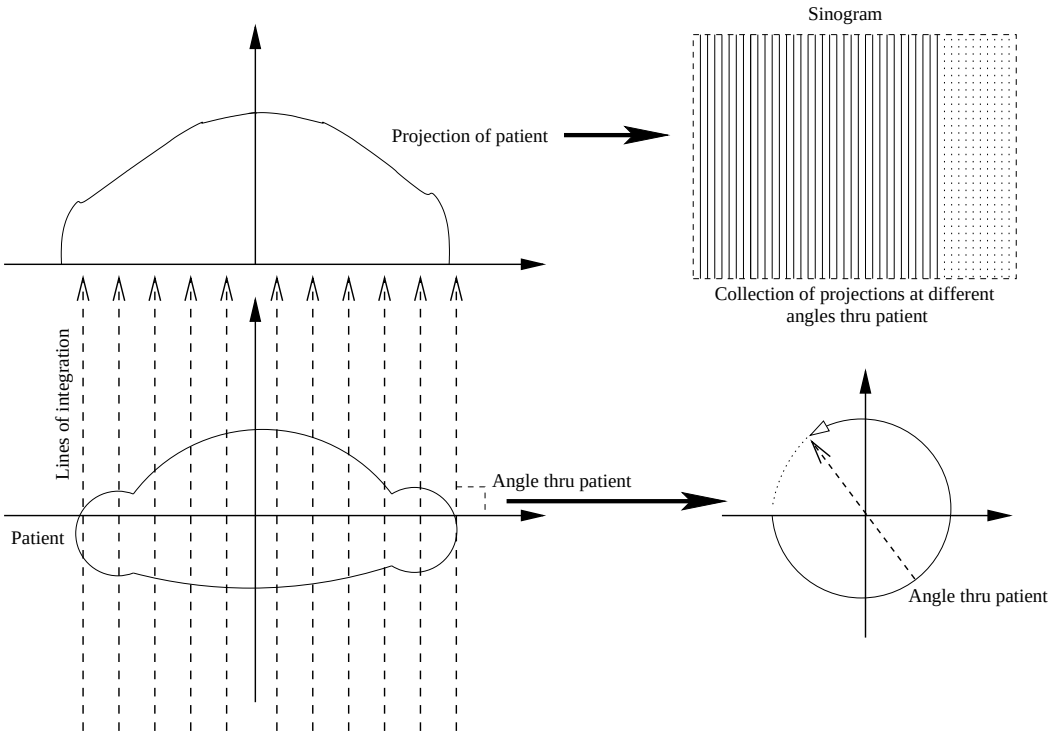
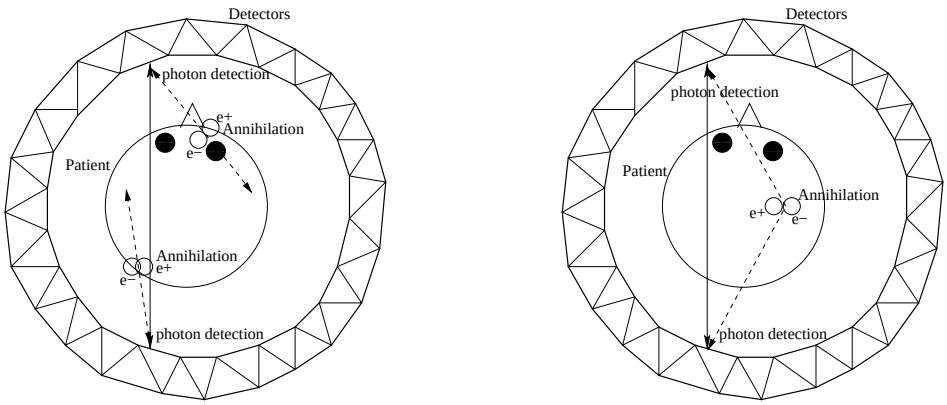


Figure 2.6: Illustration of the creation of a Sinogram from a series of projection signals through the patient.

Transformation of the photon count data is necessary as it is not in image form, but in a format known as a sinogram. The sinogram creation process is illustrated in figure 2.6. The sinogram data has to be reconstructed into image form, in order for a human to understand the data. The activity information collected by the detectors is in effect a summation of the activity over particular vectors through the imaging subject. This image reconstruction problem is an inverse problem where the activity at particular spatial locations is calculated. Various methods exist, some iterative and some analytical.

An exemplar classic analytical reconstruction technique is known as Filtered Back Projection (FBP). FBP is often related to the inverse Radon transform, a type of integral transform. Unfortunately analytical reconstruction methods produce images that are typically noisy. This is because they typically ignore the inherent randomness of the PET acquisition process. Also, iterative reconstruction methods can more readily incorporate various modelling strategies to overcome various imaging artefacts (described shortly) associated with the PET acquisition process. Additional steps can be used to improve the images produced through an analytical reconstruction process, but due to these limitations, iterative reconstruction techniques have grown in popularity.

Iterative methods are typically computationally expensive but produce better quality clinically useful image data. One such iterative reconstruction technique is known as Row-Action Maximum Likelihood Algorithm (RAMLA), [14]. RAMLA subdivides the data projections into subsets of projections. These subsets are then utilised iteratively



(a) Random Coincident Event (b) Scattered Coincident Event

Figure 2.7: Illustration of random and scattered erroneous annihilation coincident events. These two erroneous coincident events can be compared with a true coincident event that is illustrated in figure 2.5.

to reconstruct the image data via a Maximum Likelihood solution of the model.Further discussion is outside the scope of this thesis, but can be found in [14].

2.2.3 PET Imaging Artefacts

PET images suffer from a number of imaging artefacts, the effect of which can normally be reduced with various countermeasures.

Attenuation correction is necessary for gamma rays emitted from annihilation events located in the different tissues of the imaging subject. The tissues present different attenuating properties along a given ray path, and the amount of attenuation correction (AC) is dependent on the path that the gamma rays take through these tissues. Various methods exist but these usually rely on either a rotating scan of the source exterior to the imaging subject with similar energy or direct measurement of tissue attenuating properties with x-ray CT.

Random and scattered coincident errors refer to apparent but detected coincident events that are not true coincident events. These are illustrated in figure 2.7. The effect on the PET imaging data from scattered coincident errors can be minimised by various different techniques known as "Scatter Correction". Scatter correction typically involves the use of some form of *a priori* information such as the scatter response to a point source or information extracted from coincident events detected outside of the imaging subject. Correction for random coincident events is often achieved via a simple subtraction of estimated random events (utilizing a delayed coincidence channel during the acquisition process).

2.2.4 PET Resolution, PSF and PV Effect

As can be seen from figure 2.4, a finite distance is associated between the event of the unstable nucleus emitting the positron and the final annihilation event. The distance is a function of the amount of time it takes for the positron to reduce its energy to thermal levels due to its passage through the surrounding tissues. This distance is an intrinsic limiting factor for the spatial resolution of a PET scanner and is due to the uncertainty of the location of the unstable molecule that emitted the positron.

The spatial resolution of a PET imaging device is also limited by the fact that the photons are not necessarily emitted at exactly 180^o from each other. This angular uncertainty is referred to as noncolinearity and produces greater limitations on the spatial resolution for detectors with larger diameters. The resolution of a PET imaging device is also limited by the detector resolution limited by the size of the crystals used to detect the emitted photons.

These limiting factors in the spatial resolution of a PET imaging device also contribute toward the significant PV effect associated with PET imaging data and the associated PSF.

The PSF of PET imaging data is often cited to be shift-variant, so that the dimensions of the PSF vary depending on the location in the imaging data, see e.g. [155]. As can be seen shortly, chapter 7 analytically derives a prior distribution that describes the PV mixing probabilistically. The prior distribution is invariant to the size of the PSF and consequently it is also invariant to a shift-varying PSF.

The PET image acquisition process produces image intensities that possess a linear relationship with the actual activity concentration being imaged. This therefore suggests that a linear mixing assumption is also valid for PET imaging data, at least for the underlying image intensities.

2.3 Measures in Biomedical Images

2.3.1 Noise

Noise is a term used to describe the variability of a signal due to imperfect and non-predictable imaging conditions. The type of noise considered in this work typically results in a range of possible intensity values that might be associated with a given signal that would otherwise be represented by a single intensity value in the absence of noise. This spread of possible intensity values can be described by a function called a Probability Density Function (PDF) for a continuous random variable or a Probability Mass Function (PMF) for a discrete random variable. This function is usually concentrated about the true signal value and tails off for intensities further away.

Sources of noise in a MRI image acquisition process include many factors that affect the actual signal in a voxel. Examples include thermal noise in the electrical circuits of the MRI scanner and magnetic field effects of the imaging medium being studied. Other effects on the expected signal intensity may include natural variations in tissues

being imaged. For example, tissues possess many structures smaller than the voxel size that are not necessarily evenly distributed, such as blood vessels, and may therefore result in variations in the actual signal being imaged. Thus probabilistic modelling of an MRI signal can be used to reduce the effect of such variational phenomena in the images. Probability models will be discussed shortly.

Sources of noise in a PET image acquisition process also includes thermal noise in the electrical circuits, together with natural variation in activity levels particularly for variations that are difficult to resolve at the coarse PET image resolutions. The most dominant noise factor in PET imaging is associated with the scanner's count-rate capability. This is due to limitations on the maximum allowable radiation dose, along with limitations imposed at the data acquisition stage.

Measures of Noise

The Signal to Noise Ratio (SNR) and the Contrast to Noise Ratio (CNR) in images are important measures of image quality. SNR quantifies the amount of a signal in relation to the amount of noise, while CNR quantifies the dissimilarity of image regions or structures in the presence of noise.

SNR is given by

$$SNR = \frac{\mu}{\sigma}, \tag{2.7}$$

where σ is usually characterised by the standard deviation of the noise term and μ is usually characterised by the signal mean value. The SNR will be very small for relatively small signal levels or large noise levels. The SNR therefore provides an idea of the quality of an image, but it does not quantify the image properties that would allow a person or computer to distinguish individual parts of an image, unlike the CNR that summarises distinguishing parts of the image with a single value.

CNR is given by

$$CNR = \frac{\mu_A - \mu_B}{\sigma} \tag{2.8}$$

where μ_A, μ_B correspond to the mean signal for tissues A and B respectively with common standard deviation, σ. The CNR value is greatest for small standard deviation values and large differences in tissue mean values, so that two tissues become highly distinguishable. For low CNR values, the tissues become less distinguishable for a human or a computer. Figure 2.8 illustrates some exemplar image slices from simulated PV data with various CNR values given in table 2.2.

2.3.2 Probabilistic Description of Biomedical Images

As briefly discussed above, noise results in a range of intensities where one might otherwise expect a single value. This range of intensities is often described with a special type of function known as a PDF for continuous measurement values or PMF

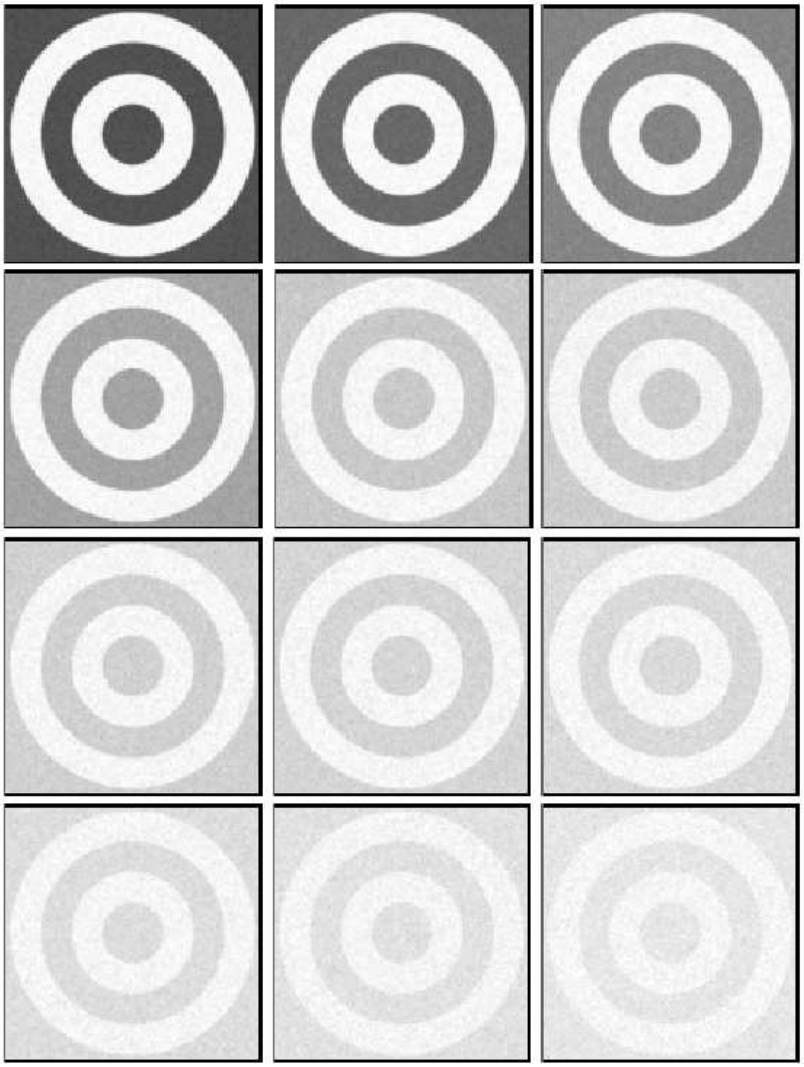


Figure 2.8: Illustration of the effect of different CNR values on the distinguish-ability of two classes using simulated PV data. Top left slice corresponds to the simulation with the greatest CNR value (50) and the remaining slices are displayed in CNR descending order from left to right and from top to bottom with values given in table 2.2. Notice how the two classes become very difficult to distinguish in the final simulation with only a CNR value of 3. Notice also that the individual SNR values are not helpful unless observed together. The image slices were produced with the aid of MRIcro, [114], where the contrast auto-balance feature was used, which scales the highest and lowest 1% of image intensities to 255 and 0 grey levels respectively.

Table 2.2: SNR and CNR values of the simulated PV data in figure 2.8. SNR values are for the background and spheroid regions of the simulated PV data.

SNR Background	*SNR Spheroid*	*CNR*
75.0	25.0	50.0
70.0	30.0	40.0
65.0	35.0	30.0
60.0	40.0	20.0
55.0	45.0	10.0
54.5	45.5	9.0
54.0	46.0	8.0
53.5	46.5	7.0
53.0	47.0	6.0
52.5	47.5	5.0
52.0	48.0	4.0
51.5	48.5	3.0
51.0	49.0	2.0

for measurements that are inherently discrete. For instance, it is common for photons arriving at the gamma ray detectors in a PET experiment to be described by the Poisson distribution which is a PMF. The MRI signal is usually described by a PDF due to the continuous nature of the measured magnetic fields.

Probabilistic Description of MRI Data

For a magnitude MRI image (where the real and imaginary components have been combined in quadrature), the probability distribution of the measured pixel or voxel intensities, M, is the Rician distribution and is given by [48]:

$$p_M(m) = \frac{m}{\sigma^2} \exp\left(-\frac{m^2 + A^2}{2.\sigma^2}\right) .I_0\left(\frac{m.A}{\sigma^2}\right), \tag{2.9}$$

where A is the signal intensity in the absence of noise, $I_0()$ is the zero order modified Bessel function of the first kind and σ is the standard deviation of the Gaussian noise in the real and imaginary images.

Experiments have confirmed that this distribution is a result of a Gaussian noise distribution in the complex domain being converted into a magnitude image[48]. Figure 2.9 illustrates the Rician distribution for $A/\sigma = 1, 6$.

A special case of the Rician distribution is when $A/\sigma = 0$ where the Rician distribution simplifies to a Rayleigh distribution. Therefore, equation 2.9 simplifies to

$$p_M(m) = \frac{m}{\sigma^2} \exp\left(-\frac{m^2}{2.\sigma^2}\right). \tag{2.10}$$

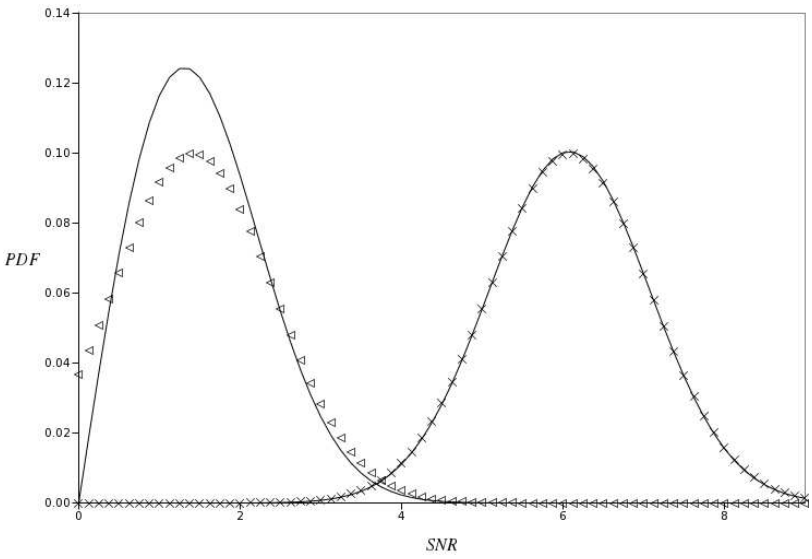


Figure 2.9: Illustration of the Rician noise distributions (equation 2.9) for magnitude MRI data for two A/σ, (SNR) values and their Gaussian approximations (equation 2.11). When $A/\sigma = 1$, (left most line), the distribution is not very well approximated by a Gaussian distribution ($\triangleleft$). When $A/\sigma > 3$, (right most line, $A/\sigma = 6$), the distribution approximates a Gaussian distribution, ($\times$).

This Rayleigh distribution corresponds to the areas in an MRI image where zero or very low NMR signal is present, (i.e. regions in the image where the SNR is low). For values of $A/\sigma > 3$, equation 2.9 approximates to a Gaussian distribution, (i.e. areas in the image where the SNR is high), as expressed by

$$p_M(m) \approx \frac{1}{\sqrt{2.\pi.\sigma^2}} \exp\left(-\frac{(m-\sqrt{A^2+\sigma^2})^2}{2.\sigma^2}\right). \tag{2.11}$$

This illustrates that the Gaussian approximation has a mean of $\sqrt{A^2+\sigma^2}$. This shows that the Gaussian approximation is always affected by a bias, and only when the SNR is large does the value of σ^2 become insignificant in comparison to the value of A.

Probabilistic Description of PET Imaging Data

The emission and counting of photons are represented by a Poisson random process, described by a Poisson PMF:

$$P(g) = \frac{\overline{g}^g}{g!}.\exp\left(-\overline{g}\right), \tag{2.12}$$

where g is the number of photons and $\overline{g}$ is the mean photon count. This Poisson PMF is illustrated in figure 2.10.

However, PET data is rarely defined by a simple Poisson process due to various other processes that affect the imaging data during the acquisition, such as the reconstruction process and corrections for scatter.

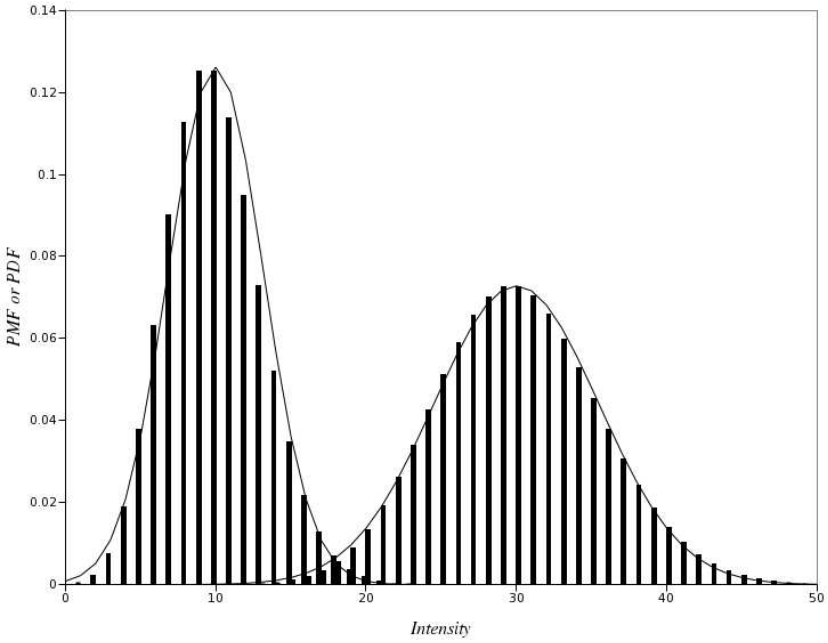


Figure 2.10: Illustration of two Poisson distributions (bars) with associated Gaussian distributions (lines) with same parameter values ($\overline{g} = 10, 30$). Notice for the left most Poisson and Gaussian distributions, the Gaussian distribution is a worse approximation in comparison to the right most pair.

Despite this, the Poisson process can be approximated by a Gaussian distribution for large enough mean values. If $\overline{g}$ is too small (less than 20), then the distribution is asymmetric, which a Gaussian distribution does not model appropriately. Two Poisson distributions together with the Gaussian approximations are illustrated in figure 2.10. Also, a Gaussian distribution has many desirable mathematical properties, that allow simpler calculations and manipulations to be undertaken. In particular it allows the models of the PV effect in this thesis to be generalised to PET data as well as MRI data.

2.4 The PV Effect in Biomedical Images

Post-acquisition computer processing of biomedical images may often only involve simple 2-D measurements of structures in the 3-D data. A clinician may be responsible for determining the size of a particular structure in the image. Before the popularisation of computers for use in *"reading"* (viewing) medical images, a clinician may simply have been presented with a hard copy version of the image often produced on film, and viewed with a light box. The clinician would then have measured the structure using traditional measuring objects such as a measuring ruler. Today, many biomedical images are viewed directly on a specialised computer workstation with software specifically designed to take care of communication with a centralised biomedical im-

age computer database[2]. This database may be connected to MRI, PET and or other biomedical imaging systems in a hospital. The radiology software will also usually allow the clinician to use tools that simulate the real world measuring instruments, perhaps even in 3-D. Sometimes it may also include semi-automatic tools to segment particular structures in the image data. These tools often rely on each structure being spatially contiguous and possessing different ranges of intensities from surrounding structures and or well defined edges so that the algorithm can determine the bounds of the structure automatically. Such algorithms could include models of the PV effect which might improve quantitative volumetric analysis.

The PV effect is of particular interest in applications where quantitative information is being measured from biomedical image data. Models of the PV effect are quite different from image models found in typical pattern recognition systems. A typical pattern recognition or decision theoretic approach may choose to assign the contents of a voxel to a single categorisation (i.e. a single tissue or functional class):

- An optimal decision - one class is considered the best over all the alternative options;
- A set of probabilities - each available class is assigned a probability.

Alternatively, assignment to a mixture of categories is also possible and is required in classification of PV affected image data:

- An optimal decision - one particular mixture of tissues or activities is considered the best over all the other alternative options;
- A PDF - associated with a continuous range of possible mixture values.

If each voxel is assigned a membership to a single class, then the model is stating that a voxel may only contain a single type of tissue or activity. This assignment is common in the broader pattern recognition field, where relatively few data points may have been generated from a mixture of signals. This approach can be compared to an approach where an individual voxel is considered to originate from more than one tissue or activity. This is considered to be more relevant in the biomedical imaging field due to the resolution effects of the comparatively large PSF in the biomedical image acquisition systems, as discussed here.

This thesis therefore uses and develops two of the mixture type approaches, namely an optimal decision on a mixture and a range of probabilities on a mixture.

Some specific application areas are now discussed to illustrate the importance of the PV effect in a clinical context.

[2]A biomedical image computer database together with the associated networking is often referred to by the term "PACS"-Picture Archival and Communication System.

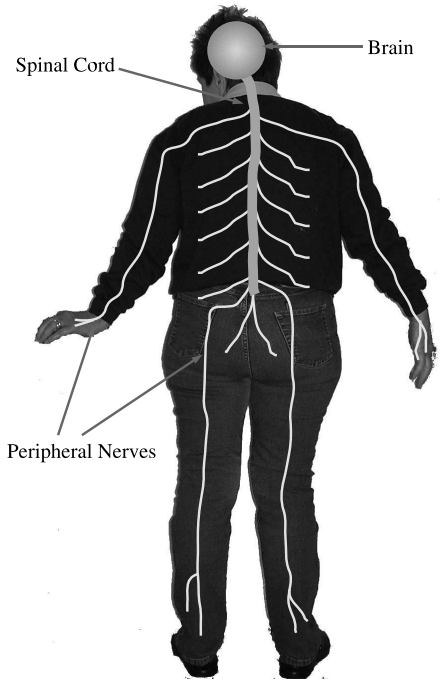


Figure 2.11: A rough illustration of the human nervous system, consisting of the brain and spinal cord that are described as the CNS and the peripheral nerves that are described as the peripheral nervous system.

2.4.1 Anatomical Imaging: Neurology

> **Neurology** *"the study of the structure, functioning, and diseases of the nervous system..."*
> Oxford's Concise Colour Dictionary, [86].

This chapter has so far discussed the PV effect in the context of the physical processes of biomedical imaging modalities. These discussions have taken place independent of any clinical background, examples of which are now discussed.

Neuro-anatomical imaging with MRI is a prevalent area of clinical importance and research interest. The diagnosis and staging of neurological disorders and diseases is often assisted with MRI data due to the high sensitivity and specificity of MRI for imaging soft tissues, especially as soft tissues are the major components of the human brain. The brain is part of the nervous system that consists of the CNS (that also includes the spinal cord) together with the peripheral nervous system (consisting of nerves not in the CNS). The peripheral and central nervous systems are illustrated in figure 2.11.

The nervous system consists of neurones that are cells interconnected with each other via axons and dendrites. A neurone is illustrated in figure 2.12. There are many different types of neurones. In general one can say that the axons are usually the output medium for neuronal signals and dendrites are typically the input medium for

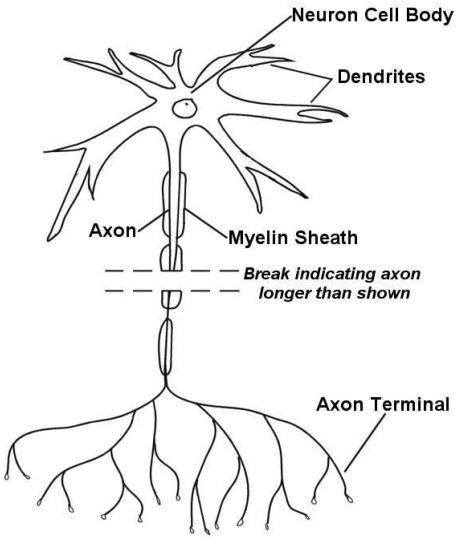


Figure 2.12: Illustration of a neurone together with an axon surrounded by a myelin sheath acting as an electrical insulating material. This insulating material assists the conduction of electrical energy which therefore allows more rapid communication of neural signals along the (particularly variable) length of an axon to other neurones.

neuronal signals. Neurones communicate with each other via neurotransmitters which are released by relative changes in electrical potential across specialised adjacent cell membranes known as synapses. Many axons are surrounded by myelin sheaths that increase the speed at which neuronal signals are communicated between each cell.

These myelin sheaths are quite significant in MR imaging of the human brain due to their effect on the resultant MR image signal. In particular, if one considers the brain component of the CNS, then it is quite often grossly divided into two types of tissue, GM and WM. The WM is typically quite different in appearance from GM due to the axons of the WM being myelinated in comparison to GM which is considered to have un-myelinated axons, or at least a considerably lower density of myelination. This is illustrated using an exemplar MR neurological image slice in figure 2.13.

The volume of space occupied by the brain, Cerebro-Spinal Fluid (CSF), GM and or WM are sometimes affected by diseases and disorders of the CNS that may affect the quantity of these tissues. Therefore neurological MR imaging applications sometimes require accurate estimates of whole brain, CSF, GM and or WM volumes. Diseases and disorders of the CNS that may affect the GM, WM, CSF and or brain volumes include, for example, schizophrenia [63, 84], post-traumatic stress disorder [147], Huntington's disease [98] and bipolar disorder [124]. The effect of these disorders on the gross anatomy help to indicate that the diagnosis of patients affected by these disorders may see benefit from accurate estimates of the anatomical volumes.

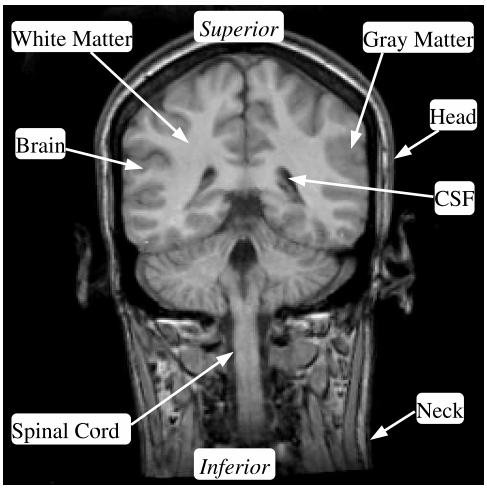


Figure 2.13: Illustration of WM and GM using an exemplar coronal T1 MR image slice from [38]. The ventricular Cerebro-Spinal Fluid (CSF), brain and spinal chord are also labelled within the bounds of the head and neck of the imaging subject.

Different types of cancer such as a vestibular schwannoma also affect the neurology of the brain and PV modelling of these cancerous growths has helped with diagnosis [148]. Furthermore, as a model of the PV effect may be applied across different stages of the biomedical image acquisition process, so can a PV classification strategy be used in diverse advanced computational imaging strategies. Many biomedical image processing applications do, or could possibly be improved with accurate estimates of voxel content, rather than just processing raw image data. This can be seen with algorithms that attempt to identify individual anatomical sub-structures within the brain (e.g. [89]). The brain is divided into empirically defined regions and given anatomical labels. These regions may have one or more cognitive functions [145] and particular regions may be more adversely affected than others in particular disease states. This means that a diagnosis may be aided with identifying visually abnormal structures. This is quite often done "manually" by a clinician, but computationally intensive techniques are becoming more popular such as techniques reliant on "registering" the MR images of a patient's brain to a standard brain space where neuroanatomical structures can be identified automatically. This technique is quite often criticised due to the natural variation found in humans, and in particular their brains [143]. Furthermore, human brains go through several stages of development in the normal duration of a human life, (see e.g. [49]), and these developments dramatically affect the structures of the brain and the resultant NMR signal. These changes are most prominent in the first two years of life and are associated with a process of myelination, where at birth, only part of the WM to be is myelinated (referred to here as "un-myelinated WM"). This myelination process occurs most rapidly during the first two years of life and continues more slowly until the end of twenty or more years [9, 118, 99, 51, 87, 61].

These facts and observations suggest that accurate per voxel identification and quantitation of gross anatomical structures, such as the GM, WM and CSF would assist in diagnostic research procedures and further computationally intensive neurological applications.

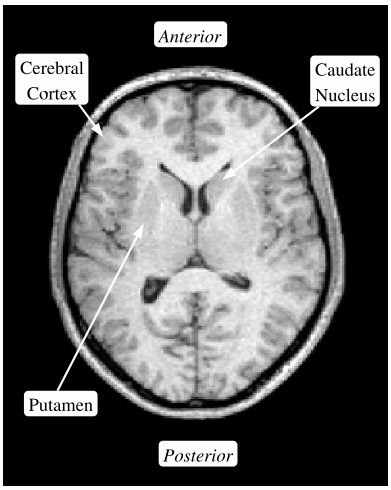


Figure 2.14: Illustration of the variability of myelination density of the GM found in the human brain. An exemplar transverse slice from a T1 MR scan (volunteer from University of Surrey) is used. Notice the difference in intensities between the GM of the Basal Ganglia (Putamen and Caudate Nucleus) in comparison to the GM of the Cerebral Cortex. This difference in intensity is due to the variability of the amount of myelin in these different regions of the brain.

Peculiarities of the PV Effect in MR Images of the Brain

The PV effect has been considered extensively in computational techniques applied to biomedical imaging data, in particular to MR images of the brain, see e.g. [3, 24, 101, 122, 157].

The brain is a complex organ that is often grossly defined as being composed of CSF, GM and WM. These are gross anatomical terms and some parts of the brain may not be conveniently categorised as such. WM and GM are defined by the density of axon myelination (high for WM) and density of neuronal cells (high for GM) [105]. However, the density of myelination is inhomogeneous. Some regions of GM, such as the Basal Ganglia (located in the centre of the brain - illustrated in figure 2.14) appear to have a higher density of myelination in comparison to the cortical GM (Cerebral Cortex) which is located on the surface of the brain. The Basal Ganglia is classed as GM by the medical literature, (e.g. [31]), but a PV classifier would possibly identify it as a mixture of GM and WM.

Interestingly, simulated MR images of the brain, described in [72], do in fact simulate the Basal Ganglia as PV voxels composed of variable amounts of WM and GM. Therefore a PV classifier whose classifier performance was quantified using the ground truth to this simulated MR brain would probably be unaffected by the variable myelin density. However there are publicly available MR brain data sets, see e.g. [38] whose ground truth was produced by human operators. Regions of the brains are therefore assigned to either GM, WM or CSF with no PV voxels. So it would be interesting to determine the performance of a PV classifier, such as those described in this thesis, on these two types of data set.

2.4.2 Functional Imaging: Oncology

> **Oncology** *"the study and practise of treating tumours."*
> Oxford's Concise Colour Dictionary, [86].

This chapter has, up until this point, only considered the physical principles of PET functional imaging. Quantitative functional imaging techniques, including PET and accurate PET image models, are especially useful in oncology. Cancer is a major cause of ill-health and premature death in humans and accurate quantitation of a tumour is essential to the accurate diagnosis and staging of cancer, where well-informed treatments can be prescribed helping to improve patient quality of life. Therefore a brief discussion of PET with application to oncology is now given.

A tumour is an unusual growth of tissue that may be benign (not harmful) or malignant (harmful). A malignant tumour presents a danger to the patient as it may destroy normal tissue that surrounds it but, often more importantly, its growth interferes with the normal function of the surrounding tissue. A malignant tumour may also spread to other parts of the patient's body.

Imaging of a tumour may occur at many stages of medical diagnosis. It is sometimes used as part of a preliminary investigation, staging, diagnosis and or treatment planning.

PET and MRI are important imaging modalities in oncology. MRI is useful to assess size and location of a tumour. PET is particularly useful prior to radiotherapy and other oncological treatments as quantitative information about the malignancy of the tumour can be acquired. PET imaging with the radioisotope FDG is particularly useful as most types of malignant growth can be detected due to their high metabolic activity in comparison to normal tissue.

There are principally three types of treatment options for cancer:

- Radiotherapy:
  - Targeted ionising radiation is administered to the patient which is used to kill the cancerous tissue cells. The radiation may be administered via external beam or via molecular pathways.
- Chemotherapy
  - anticancer drugs that target and inhibit particular processes in a cell's life-cycle are administered
- Hormonal therapy
  - some cancers are under hormonal control and can be controlled by adjusting the relevant hormone(s).

Imaging plays a vital role in localising and quantifying the cancer in the patient. It is particularly useful in the radiotherapy treatment planning process where it is often

required by the radiotherapist to identify the location of the tumour and the location of other radiation sensitive organs in order to produce a radiotherapy treatment plan. Imaging is therefore used as an integral part of the treatment process of the oncological patient.

2.5 In Conclusion

This chapter has discussed a number of background topics pertaining to the topic of this thesis, namely modelling and classification of PV biomedical imaging data. The next chapter presents a novel automatic pre-processing step, known as skull-stripping, often required prior to PV classification of neurological MR data. Skull-stripping may also be used in PET-based brain studies, where registered skull-stripped MR data may be used to localise the resulting PET signal. The chapters following then go on to discuss modelling of the PV effect.

Chapter 3

Pre-processing of Neurological MRI Data: Skull Stripping

3.1 Introduction

The statistical analysis of neurological Magnetic Resonance Imaging (MRI) data is often made easier when voxels corresponding to non-Central Nervous System (CNS) tissue compartments are removed from the MRI data set, as illustrated in figure 3.1. This figure (3.1) illustrates how the CNS tissue compartments of White Matter (WM), Grey Matter (GM) and Cerebro-Spinal Fluid (CSF) can be identified in the intensity histogram of the skull-stripped data volume. It also illustrates how the non-CNS tissue compartment voxels contribute towards similar ranges of intensities to the CNS tissue compartment voxels, thereby confounding the identification of individual CNS tissue compartment voxels when using intensity based analysis.

Skull stripping techniques have previously included the use of surface deformation techniques [131, 85, 2] which are computationally complex. Others have utilized the watershed transform [52] or techniques mostly reliant on morphological operations [135, 79, 127, 91, 15] which are often dependent on the definition of a 3-D brain mask. The definition of a 3-D brain mask is problematic due to the inherent three dimensional connectivity of the brain with other anatomical structures in MRI images of the human head. This has led other researchers to introduce more sophisticated techniques that combine surface deformation techniques with morphological operations [65, 108, 120]; watershed transforms with surface deformation techniques [126]; and techniques based on a consensus of publicly available skull-stripping techniques such as [110] that used the publicly available code associated with [131, 128, 32, 29]. References [128, 10] take a slightly different approach, both of which use a combination of edge detection and morphological processing, although [128] introduced a denoising step in the form of anisotropic diffusion filtering.

Artificial Intelligence techniques, such as the use of multiple software agents have also been used in [43], where the agents specialize in differing image features including region, edge and intensity features. Lee *et al.* in [75] automatically isolate a midsagittal

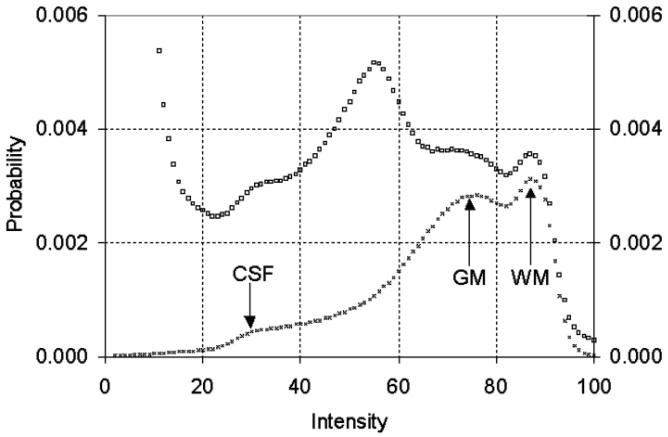


Figure 3.1: Illustration of the benefits of pre-processing neurological MRI data. The histograms correspond to an un-processed T1 MRI head scan (□), and a skull stripped T1 MRI head scan (×). The histogram corresponding to the skull-stripped data illustrates the benefit of removing the irrelevant information from the data set, where the peaks corresponding to GM, WM and CSF can be seen more easily. This is due to the removal of unwanted non-CNS tissue component voxels in an MRI scan of a human head that share a similar intensity range as CNS tissue compartment voxels.

slice from sagittally acquired data sets utilizing landmarks such as the tip of the nose. The technique in [75] is extended to skull-strip the entire CNS in 3-D in [59]. The use of landmarks may be hampered by neurological data sets that do not necessarily include facial features such as the tip of the nose. A technique by Soltanian Zadeh and Windham in 1997, [132], was unusual as it used a multi-resolution contour tracking algorithm to extract the contours of the human brain in both X-ray Computed Tomography (CT) and MRI neurological data sets. But contour tracking can be a computationally complex task for volumetric imaging data and quite dependent on the quality of the imaging data.

The new work presented here describes a novel fully automatic methodology utilizing statistical techniques including fitting of probabilistic functions and thresholding. Further image processing intensive operations include the use of region-growing and mathematical morphological operations. A fully automatic approach has been developed because of the need in some studies for unsupervised processing where large numbers of data sets are utilized, for example, in longitudinal studies of brain atrophy [4]. The technique is deemed to be flexible in contrast to other skull-stripping techniques dominated by mathematical morphological operations due to the use of self-similarity functions, a 2-D brain mask and non-dependency on landmarks or the plane of acquisition.

A unique development in the Statistical Morphological Skull-Stripper (SMSS) is the use of self-similarity functions that utilize volumetric and slice specific image statistics to identify dominant components in the imaging data. A well known problem with MRI data is the variability of statistical descriptors such as means and standard deviations of voxel intensities corresponding to particular tissue classes, such as GM or WM.

SMSS uniquely identifies these dominant components by acknowledging the fact that these components will be the most numerous in neurological imaging data. This makes possible the identification of image slices with these dominant components without reference to unreliable statistical descriptors, leading to slices that are plausible 2-D brain masks. The use of a 2-D brain mask for template matching the entire volume is an approach based on a manually assisted segmentation method previously described in [112]. The 2-D brain mask enables a variable number of erosions and dilations to be performed. Thus reducing the dependence of the technique on the relatively smaller sizes of the anatomical structures of interest (relative to adults for infant data). This is in contrast to techniques such as [91] that use a fixed number of morphological operations. Similar work in [135] utilized a variable number of erosions and dilations, but these were applied to a thresholded 3-D data volume, where the iterations ceased when a major disconnection event occurred. Our approach differs significantly from the approach taken in [135] due to the use of self-similarity functions and a 2-D brain mask.

SMSS has also been designed so as to include CSF voxels. This enables better subsequent analysis to be undertaken; for example, probabilistic models of the three main CNS component classes can be built for the purposes of Partial Volume (PV) modelling, using techniques such as those described in chapters 5 to 7.

This newly fully automated technique (SMSS) is then applied to a number of neurological MRI scans that include infant and adult data sets (T1 and T2 weighted). These results are then compared with the results of an alternative popular skull stripping technique known as BET, as discussed in [131]. BET is a simple but effective and popular deformable surface skull-stripping technique.

As far as the author are aware, no other skull-stripping work has previously investigated the skull-stripping problem of infant MRI data. Infant neurological MRI data is considered to be quite different from adult MRI data due to the relative difference in sizes of anatomy, which is often acquired at coarser imaging resolutions. The Nuclear Magnetic Resonance (NMR) signal from infant neurological data is also quite different from adult data due to the developing nature of the human brain, see for example [119, 9]. These factors perhaps suggest that the skull-stripping of infant neurological data may be more problematic in comparison to the already complicated task of skull-stripping adult data. SMSS together with BET, is therefore applied to both infant and adult neurological data sets to determine whether SMSS and BET are as capable of skull-stripping infant and adult neurological data sets.

3.2 Methodology

Skull stripping techniques based solely on mathematical morphological operations usually require some user interaction. In contrast, the methodology presented here is primarily based on mathematical morphological operations with additional statistical techniques that fully automate the skull stripping process.

The main contribution of this chapter lies in the methodology described below and its demonstrated application to adult as well as infant brain data.

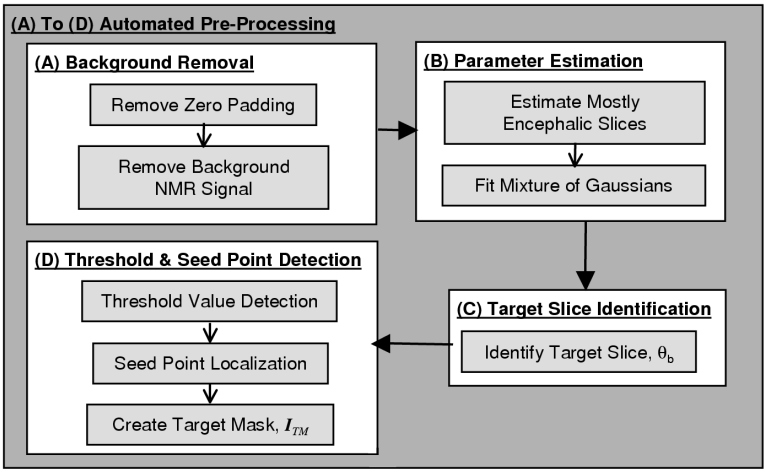


Figure 3.2: Diagram of the steps involved in the first part of the skull-stripping algorithm, sections 3.2.1 to 3.2.5. The algorithmic steps are continued in figure 3.6.

In summary the method proceeds as follows: initially, background voxels are removed via automated region growing using statistical termination criteria; parameter estimation is then undertaken to fit an intensity Gaussian mixture model to the histogram of a subset of the data defined by a further fitting operation; from this a transverse target slice is selected which is segmented by automatic thresholding and region growing to create a 2-D target mask. The entire 3-D volume is then subjected to a series of 3-D morphological operations. The process halts when all the brain voxels have been disconnected from the unwanted surrounding non-brain voxels in the transverse plane target slice using the transverse 2-D target mask (initially described in [112]). A set of further 3-D morphological operations are then undertaken to compensate for the over-erosion of the brain volume. The methodology is described in more detail in the sections below (also summarized in figures 3.2 and 3.6).

3.2.1 Background Removal

An initial first step in this skull stripping algorithm is the removal of voxels that contain zero or very little NMR signal. These voxels usually arise either due to the air surrounding a patient or from additional voxels that have been added to the data volume to obtain a convenient set of dimensions for processing with Fourier methods (zero-padding). Initially, individual slices are considered independently to remove zero-padded voxels, via a simple region-growing operation. This is automatically initiated by selecting seed points corresponding to the four corners of each slice, terminated when no connected zero value voxels remain (using 2-D eight way connectivity).

The second stage is a more complex region-growing operation to remove voxels with little NMR signal present that correspond to air surrounding the patient. These voxel intensities are usually considered to possess a Rayleigh Probability Density Function (PDF) and voxels that correspond to non-air regions with a larger NMR signal can be approximated with a Gaussian PDF [48]. Therefore, during the background region

growing process, the boundary of non-air voxels can be detected when the statistics of the border region changes. From this, it was decided that a region growing operation that iteratively grew the background region using a process that increased the intensity threshold but halted when the standard deviation of the border region goes through an inflection. This would signal the point at which the surrounding background region "floods" into the foreground object corresponding to the point at which the statistics of the background Rayleigh density are affected by the relative increase in brightness of the Gaussian distributed foreground. Once the termination criterion has been met, then the values of the voxels defined as background, through being included in the region grown, are set to a zero intensity value. This enables subsequent operations to exclude those voxels based solely on their zero intensity value *label*.

3.2.2 Parameter Estimation

This stage attempts to determine approximate values for the parameters of the PDFs of the GM and WM CNS tissue component voxels. Approximate values for these parameters will aid subsequent stages in the skull-stripping algorithm.

Initially, slices in the data volume need to be identified that contain a substantial number of CNS tissue component voxels. These slices are identified by comparing the Root Mean Square (RMS) error between the normalized histogram for the entire data volume following the background removal stage (described previously), and the normalized histogram for individual image slices. The schema relies on the assumption that the CNS tissue component voxels are the most populous tissue component voxels within the data volume. This implies that the RMS error will decrease for image slices with a substantial population of CNS tissue component voxels.

It was initially proposed that an appropriate populous CNS slice could be found to be defined by the region where the first RMS histogram error minima occurs (below the mean error for the entire data volume). In practise, it transpires that further appropriate slices could be found from this minima and up to where the RMS error became greater than the mean error as illustrated in figure (3.3).

Once these slices have been identified, a finite Gaussian mixture model is fitted to the aggregate histogram, $f_{data}(g)$ of these slices. The number of Gaussian components, K was set to three for adult data and one for infant data. The use of three components for adult data can be understood from the fact that adult neurological MRI data typically possesses three distinct ranges of intensities, representative of the intensities associated with GM, WM and CSF. This is in contrast to the overlapping nature of the intensities from these same components for infant neurological MRI data where $K = 1$. This is because infant GM and WM are not physiologically well differentiated.

The Nelder-Mead downhill simplex algorithm [104] is then used to fit the K class finite Gaussian mixture model to the histogram (excluding the background voxels) to estimate the parameters of each major Gaussian component that contributed to the histogram of the entire set of selected 2-D slices. The Nelder-Mead downhill simplex

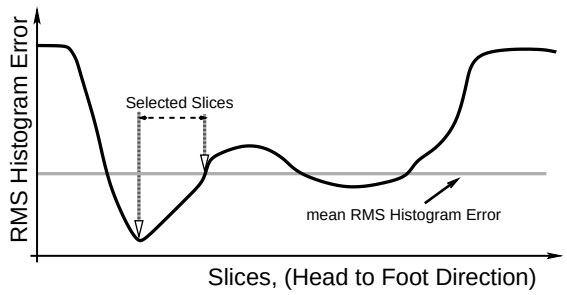


Figure 3.3: Illustration of the process of determining image slices that contain a substantial population of CNS tissue component voxels. The mean RMS histogram error is calculated for the entire data volume after the background removal stage, while the solid black line corresponds to the histogram errors for individual image slices for the same volume.

algorithm attempts to minimize the mean square error, $\mathcal{E}(\mathcal{I})$ defined by:

$$\mathcal{E}(\mathcal{I}) = \sqrt{\sum_{g=1}^{g_{max}} \left(f_{data}(g) - f_{model}(g|\mathcal{I})\right)^2}, \tag{3.1}$$

where $g = 1$ to g_{max} represents the voxel intensity range. $\mathcal{I}$ is the set of parameters of the K class Gaussian mixture model:

$$\mathcal{I} = \{\mu_v, \sigma_v, P(\tau_v) | 1 \leq v \leq K\}, \tag{3.2}$$

where μ_v, σ_v and $P(\tau_v)$ are the mean, standard deviation and *a priori* probability for the v^{th} class component of the mixture model respectively. τ_v is indicative of the v^{th} class component event from a set of possible class component events. τ_v is valid for $1 \leq v \leq K$, so that for adult data, $K = 3$ resulting in a probability space divided into 3 possible events, with prior probabilities, $P(\tau_1)$, $P(\tau_2)$ and $P(\tau_3)$. Thus the mixture model, $f_{model}(g|\mathcal{I})$, is defined by:

$$f_{model}(g|\mathcal{I}) = \sum_{v=1}^{K} \left(\frac{P(\tau_v)}{\sqrt{2.\pi.\sigma_v^2}} . \exp\left(-\frac{(g-\mu_v)^2}{2.\sigma_v^2} \right) \right). \tag{3.3}$$

The initial values for the set of parameters in $\mathcal{I}$ were calculated from $f_{data}(g)$. For each class component, τ_v, empirically determined points in the Cumulative Distribution Function (CDF) of $f_{data}(g)$ were determined and used to initialize the mean values of each class, μ_v. To initialize the standard deviation values, σ_v, the histogram is divided into K regions of equal intensity width and the values taken by each σ_v are a third of the size of each of these regions. Similarly, the *a priori* probability values, $P(\tau_v)$ are initialized by determining the total population of non-background voxels, dividing this value by K and normalizing those values with respect to the total voxel population. The algorithm is terminated when a minima was found by the optimization algorithm.

3.2.3 Automated Slice Selection

Following removal of background voxels where no NMR signal is present, a suitable *transverse* slice (the target slice) has to be selected utilizing results from the preceding parameter estimation step. In previous work, [112], this has required manual intervention in the segmentation process, to select the slice, and undertake the required thresholding. However, following the background isolation and parameter estimation steps described above, this process can now be achieved automatically. There are some unique properties associated with a suitable transverse target slice that can be utilized in the automation of this process: (i) the CNS tissue component voxels represent the largest group of all the tissue components in the ideal target slice; (ii) the frequency of occurrence of all tissue component voxels (CNS and non-CNS) are at a local maxima in the region of the target slice; and (iii) the WM voxels represent the largest CNS component. Using these observations the formulation will now be discussed in detail.

A data set that corresponds to an MRI scan of the human head, Θ, composed of a number of transverse slices, θ_a, ordered consecutively so that the first slice, $a = 1$, corresponds to the top most slice (furthest from the feet in the direction of the head) is represented by the following expression:

$$\Theta = \{\theta_a | 1 \leq a \leq M\}, \tag{3.4}$$

where M is the number of transverse slices in the data set. An image slice, θ_a, is a mapping from a 2-D point, $\boldsymbol{\omega} = (x\ y)^T$, to a scalar intensity value, i.e. $\theta_a(\boldsymbol{\omega}) = g$. A suitable target slice, θ_b, (to be used in subsequent morphological processing steps), is then defined as the slice that maximizes the accumulated tissue class, τ_v, probabilities, ψ_b:

$$\psi_b \geq \psi_a \ \forall a \neq b, \tag{3.5}$$

where

$$\psi_a = \sum_{\forall \boldsymbol{\omega}} \sum_{\forall j} P(\tau_v).p(g = \theta_a(\boldsymbol{\omega})|\tau_v). \tag{3.6}$$

$p(g|\tau_v)$ is the tacitly assumed Gaussian PDF associated with tissue class τ_v, the parameters of which are taken from the preceding parameter estimation step, $\mathcal{I}$ defined by equation (3.2), except the tissue class priors, $P(\tau_v)$, which were set to empirically estimated values. The prior values were selected so as to maximize the probability of finding a slice with a large number of GM and in particular, WM voxels.

3.2.4 Seed Point and Threshold Value Detection

After the 2-D target slice has been identified, a binary 2-D target mask is created representing the CNS tissue classes using an automated process. This target mask is used in subsequent steps in the algorithm to determine whether the CNS tissue component voxels have been successfully disconnected from the extraneous non-CNS tissue component voxels, e.g. at the optic nerve. To create the mask, the CNS tissue voxels in the target slice were isolated using region growing operations. Intensity based region growing operations usually require user-specified initial seed points, and intensity

values that act as thresholds to terminate the growth of the region. This section describes how the seed points and threshold values used in this work can be determined automatically. First, the method used to determine threshold values needed for region growing is described below.

The CNS tissue component voxels within the target slice represent the largest single class of voxels. This observation leads to the following formalized approach for identifying suitable values for automatically initializing threshold and seed point intensities:

The target slice, θ_b, containing intensities, g, can be formally described by a normalized histogram:

$$h(g|\theta_b) = \frac{|f(g|\theta_b)|\,.\nu}{D_h.D_w}, \tag{3.7}$$

where

$$f(g|\theta_b) = \{\boldsymbol{\omega}|\theta_b(\boldsymbol{\omega}) = g, \forall\boldsymbol{\omega}\}, \tag{3.8}$$

D_w and D_h are the width and height, in voxels, of the image slice, θ_b and ν is the bin width (usually $\nu = 1$).

The histogram is then used to identify the largest cluster of voxel intensities representing the majority of points in θ_b; these possess a limited range of intensity values, which will correspond to the vast majority of CNS tissue component voxel values. A center-of-cluster function, $X(g'|\theta_b)$, is calculated using a moving window over $h(g|\theta_b)$:

$$X(g'|\theta_b) = \frac{1}{(\mathfrak{T}_2 + \mathfrak{T}_1)} \sum_{g=g'-\mathfrak{T}_1}^{g'+\mathfrak{T}_2} h(g|\theta_b) \text{ for } 1 \leq g, g' \leq g_{max} \tag{3.9}$$

where $(\mathfrak{T}_2 + \mathfrak{T}_1)$ is the width of the moving window and is calculated from the distance between the two means of the GM and WM CNS tissue component classes, μ_{GM} and μ_{WM} respectively, selected from $\mathcal{I}$. ie. $(\mathfrak{T}_2 + \mathfrak{T}_1) = \chi.|\mu_{WM} - \mu_{GM}|$, where χ is a window width parameter and $\mathfrak{T}_2 = \mathfrak{T}_1$. The voxel intensity range is the same as given for equation 3.1. In the experiments that follow, an optimal value of $\chi = 3.0$ was empirically determined. A suitable value for the seed point voxel in θ_b, is given when the maximum for (3.9) is found at $g' = g_{peak}$.

The region growing operation that follows also requires upper, $\mathfrak{T}_U$ and lower, $\mathfrak{T}_L$ threshold intensity values to limit region growth. $\mathfrak{T}_U$ can be determined from $h(g|\theta_b)$: a suitable value for $\mathfrak{T}_U$ occurs at the highest intensity *knee* of the right most CNS tissue component class in the target slice histogram, $h(g|\theta_b)$. This is illustrated in figure (3.4) and can be defined when the following becomes true:

$$h(\mathfrak{T}_U = g|\theta_b) \leq \alpha.X(g' = g_{peak}|\theta_b) \text{ for } g \geq g_{peak}, \tag{3.10}$$

where α defines the location of the *knee* and whose value can be obtained empirically. In practise $\alpha = 0.10$ was found to give reliable results.

The lower threshold, $\mathfrak{T}_L$ can then be assumed to be symmetric about g_{peak}, therefore, $\mathfrak{T}_L = 2.g_{peak} - \mathfrak{T}_U$, which will enable the region growing algorithm to include the vast majority of the population of CNS tissue component voxels in the target slice, θ_b. Having determined appropriate threshold values, a region growing process is automatically

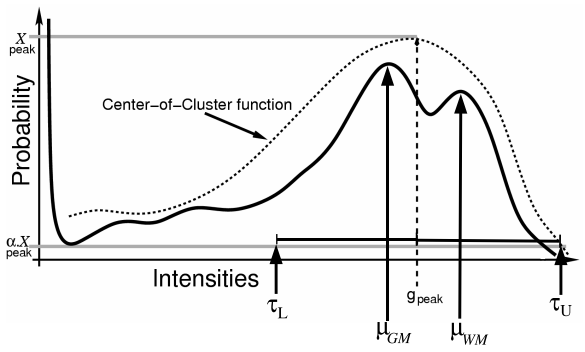


Figure 3.4: Illustration of the determination of the upper threshold intensity value, $\mathfrak{T}_U$. The histogram, $h(g|\theta_b)$ (plane line) and the center-of-cluster function, $X(g'|\theta_b)$, (dotted line) are illustrated. $\mathfrak{T}_U$ can be determined close to the *knee* of the right most peak of $h(g|\theta_b)$, occurring at approximately α of the peak value of $X(g'|\theta_b)$. The approximate means, μ_{GM} and μ_{WM} are also illustrated. These are used in the calculation of the window width, $(\mathfrak{T}_2 + \mathfrak{T}_1) = \chi.|\mu_{WM} - \mu_{GM}|$, for equation 3.9.

(a) (b)

Figure 3.5: Illustration of (a) 2-D eight way connectivity and (b) 3-D twenty-six way connectivity, used by functions Connected$_{2D}$() and Connected$_{3D}$() in equations 3.11 and 3.16, respectively.

initiated using the following formalism: a 2-D target mask, I_{TM}, (a set of 3-D points that correspond to a single image slice, b), is defined by:

$$I_{TM} = \text{Connected}_{2D}(C, I_{TS}), \tag{3.11}$$

where Connected$_{2D}(\Lambda_1, \Lambda_2)$ is a function that takes a set of 3-D seed points in set Λ_1 and using 2-D eight-neighborhood connectivity (illustrated in figure 3.5(a)), determines the corresponding connected points in set Λ_2. I_{TS} is a set of 3-D points and can be considered as a binary representation of the target slice:

$$I_{TS} = \{(x\ y\ b)^T | \mathfrak{T}_L \leq \theta_b(\boldsymbol{\omega}) \leq \mathfrak{T}_U, \forall \boldsymbol{\omega} = (x\ y)^T\}, \tag{3.12}$$

and C is a set of seed points, defined by points taking a grey level value equal to g_{peak}, (as illustrated in figure 3.4 and described by equation 3.10), and located surrounding the centroid of the points in I_{TS}.

3.2.5 Morphological Operations

After the target mask has been defined, a set of morphological operations are executed to disconnect the CNS tissue component voxels from non-CNS tissue component voxels. These connections arise due to the various anatomical parts of the head such as the optic nerve and the meninges that possess relatively high NMR signal and are located between the brain and the surrounding non-CNS tissues. The configuration of morphological operations used here is similar to the configurations initially suggested in [112], but with some modifications, primarily to include the majority of the CSF tissue component voxels.

For this stage of the algorithm, the reader is reminded to refer to figure 3.6 to obtain a detailed overview of the steps in this latter stage. Symbols used in the various equations that follow can be cross-referenced with the symbols in figure 3.6. Initially, the entire 3-D data set, Θ, is transformed to a binary representation using the previously determined intensity threshold values, $\mathfrak{T}_L$ and $\mathfrak{T}_U$, to produce a binary mask volume L_0:

$$L_0 = \{(x\ y\ a)^T | \mathfrak{T}_L \leq \theta_a(\boldsymbol{\omega}) \leq \mathfrak{T}_U, \forall\ a, \forall \boldsymbol{\omega} = (x\ y)^T\}. \tag{3.13}$$

A number of iterations are then performed to disconnect the set of voxels in L_0, that correspond to CNS tissue compartment voxels from non-CNS voxels. This process starts with a series of erosion operations to remove small links between major compartments.

First, a 3-D mathematical morphology erosion operation is applied to the binary mask volume, commencing at iteration $r = 1$:

$$L_1^{[r]} = L_2^{[r-1]} \ominus B, \tag{3.14}$$

where initially,

$$L_2^{[r=0]} = L_0, \tag{3.15}$$

and B is a cubic morphological structuring element of fixed size, $3 \times 3 \times 3$ voxels. The next stage in the current iteration utilizes a 26-way connectivity analysis to determine the set of points that are still connected to the points in the target mask. Essentially, all the points in $L_1^{[r]}$, following the erosion, that are connected to any of the points that are jointly in the target mask are kept, that is:

$$L_2^{[r]} = \text{Connected}_{3D}(I_{TM}, L_1^{[r]}), \tag{3.16}$$

where $\text{Connected}_{3D}(-,-)$ is a similar connectivity function to the 2-D connectivity function as used in equation 3.11, but using 3-D 26-way connectivity instead of 2-D 8-way connectivity (illustrated in figure 3.5(b)).

A test is now performed to determine whether the current iteration should terminate. If the membership of the current $L_2^{[r]}$ results in a set of points for the target slice, I_{TS}, (corresponding to image slice b) that are a subset of the set of points in the target mask, I_{TM}, then it is deemed that the algorithm has disconnected the entire set of CNS tissue component voxels from the non-CNS tissue component voxels. This termination criterion can be formalized as when:

$$(I_{TS} \cap L_2^{[r=w]}) \subseteq I_{TM}, \tag{3.17}$$

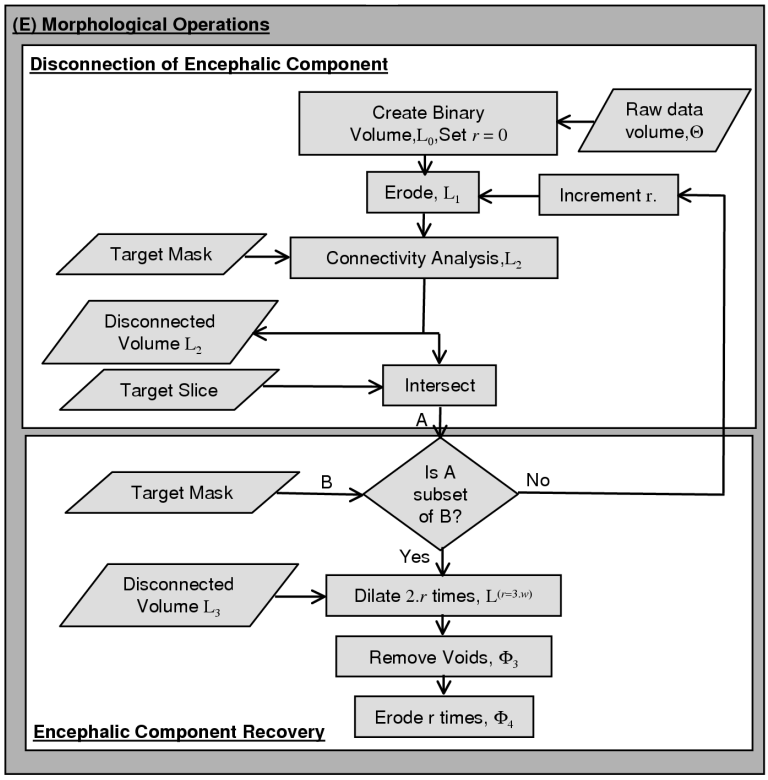


Figure 3.6: Diagram of the steps involved in the second part of the skull-stripping algorithm, utilizing morphological operations, step E, following on from steps A to D in figure 3.2.

becomes true, where the target slice, I_{TS}, is defined in equation 3.12. Otherwise r is incremented and the steps from equation (3.14) are re-applied.

Once this process has completed, a number of dilations ($2.w$ times) of the resultant set of points $L_2^{[r=w]}$ are performed. This recovers voxels corresponding to CNS tissue component voxels inadvertently lost in the disconnection process including additional CSF component voxels:

$$L_2^{[r+1]} = L_2^{[r]} \oplus B,\ w < r \leq 3.w. \tag{3.18}$$

After the dilations have been performed at step, $r = 3.w$, the set of voxels $L_2^{[r=3.w]}$, will still contain a significant number of residual voids. E.g. for T1 MRI data, the CSF component voxels (such as the ventricles) will usually possess intensity values outside the threshold range $(\mathfrak{T}_L, \mathfrak{T}_U)$ and therefore will not have been included in the target mask. Some skull-stripping techniques are not concerned with CSF regions, but the statistical modelling of the entire CNS tissue component volume can benefit from the inclusion of the CSF tissue component voxels, especially when modelling the PV effect. Therefore, these voids are filled using the following steps. A set of 3-D points, Φ_1, within the image space and not in $L_2^{[r=3.w]}$ is defined by the following expression:

$$\Phi_1 = \{(x\ y\ a)^T | (x\ y\ a)^T \in I_a, \notin L_2^{[r=3.w]}, \forall a\}. \tag{3.19}$$

This effectively results in an inversion of the segmented binary data volume. The previously defined background voxels can now be used in a region growing operation to grow around the outside of the CNS tissue component region. This process avoids "holes" such as the ventricles.

Therefore a set of points, Φ_2, that are connected to the set of background seed points, S (as discussed in the first section of this chapter), within Φ_1, have to be determined:

$$\Phi_2 = \text{Connected}_{3D}(\Phi_1, S). \tag{3.20}$$

The segmented CNS tissue component voxels, Φ_3, can then be defined as the inverse of Φ_2:

$$\Phi_3 = \{\boldsymbol{\omega} = (x\ y\ a)^T | \boldsymbol{\omega} \in I_a, \notin \Phi_2, \forall a\}. \tag{3.21}$$

Finally, a set of erosion operations are then performed. These erosions remove any non-CNS tissue component voxels that might have been included within the segmented data volume during the application of equation (3.18):

$$\Phi_4^{[r+1]} = \Phi_4^{[r]} \ominus B,\ 3.w < r \leq 4.w; \tag{3.22}$$

where $\Phi_4^{[r=3.w]} = \Phi_3 \ominus B$. The resulting volume, $\Phi_4^{[r=3.w]}$, then represents the derived skull-stripped CNS voxels component.

Table 3.1: Summary of MRI developmental and evaluation test data sets used to assess the performance of the skull-stripping algorithm. (C), (T) and (S) indicate acquisition plane i.e. Coronal, Transverse and sagittal respectively.

ID	**Description**	**Voxel Dimensions** mm^3
V1,V2	Development,Adult,T1,(S)	1.00×1.50×1.00
V3	Development,Adult,T1,(T)	1.00×1.00×1.00
V4	Development,Infant(10 months),T2,(S)	0.86×6.00×0.86
V5	Development,Infant(10 months),T1,(S)	0.86×6.00×0.86
V6	Development,Infant(10 months),T2,(C)	0.86×6.80×0.86
V7-V9	Development,Adult, T1,(C)	1.00×3.00×1.00
V10	Evaluation,Infant(10 months),T1,(C)	0.86×6.80×0.86
V11	Evaluation,Infant(10 months),T2,(T)	0.86×0.86×6.00
V12	Evaluation,Infant(10 months),T1,(T)	0.86×0.86×6.00
V13-V20	Evaluation,Adult,T1,(C)	1.00×3.00×1.00

3.3 Performance Assessment

The methodology presented here was developed using 9 MRI developmental data sets, V1 to V9. Evaluation test data sets V10 to V20 were used purely for evaluation purposes, after the complete development of the algorithm. Data sets V1 to V20 have different dimensions and are of variable quality. These data sets and their properties are summarized in table 3.1. Adult data set V1 is from a volunteer who was at University of Surrey, U.K. The infant data set volumes (V4-V6 and V10-V12) are from a single subject and are illustrated in figure 3.7. The age of the infant was 10 months at the time of image acquisition. Adult data set volume V3 is publicly available with a copyright notice[1]. Adult data set volumes (V7-V9 and V13-V20) are also publicly available[2]. Adult data set V2 is part of the Chapel Hill Volume Rendering Test Data Sets.

The performance of the automatic skull stripper was compared with the performance of BET [131]. BET can be summarized as a skull-stripping technique primarily based upon a surface deformation algorithm. Initially intensity thresholds for the CNS image data are estimated from the intensity frequency histogram. Secondly, the center-of-gravity for voxels with these intensity threshold values is calculated which is then used to initialize the deformable surface consisting of a spherical triangular tessellation. The tessellated sphere starts from a point that should be within the CNS spatial limits and is then allowed to deform with smoothness constraints until the CNS limits have been

[1] V3 copyright notice: "Courtesy of, and ©by, Mark Bentum, bentum@wsrt00.nfra.nl, Netherlands Foundation for Research in Astronomy".

[2] Data sets V7-V9 and V13-V20 come from a set of 20 normal MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/.

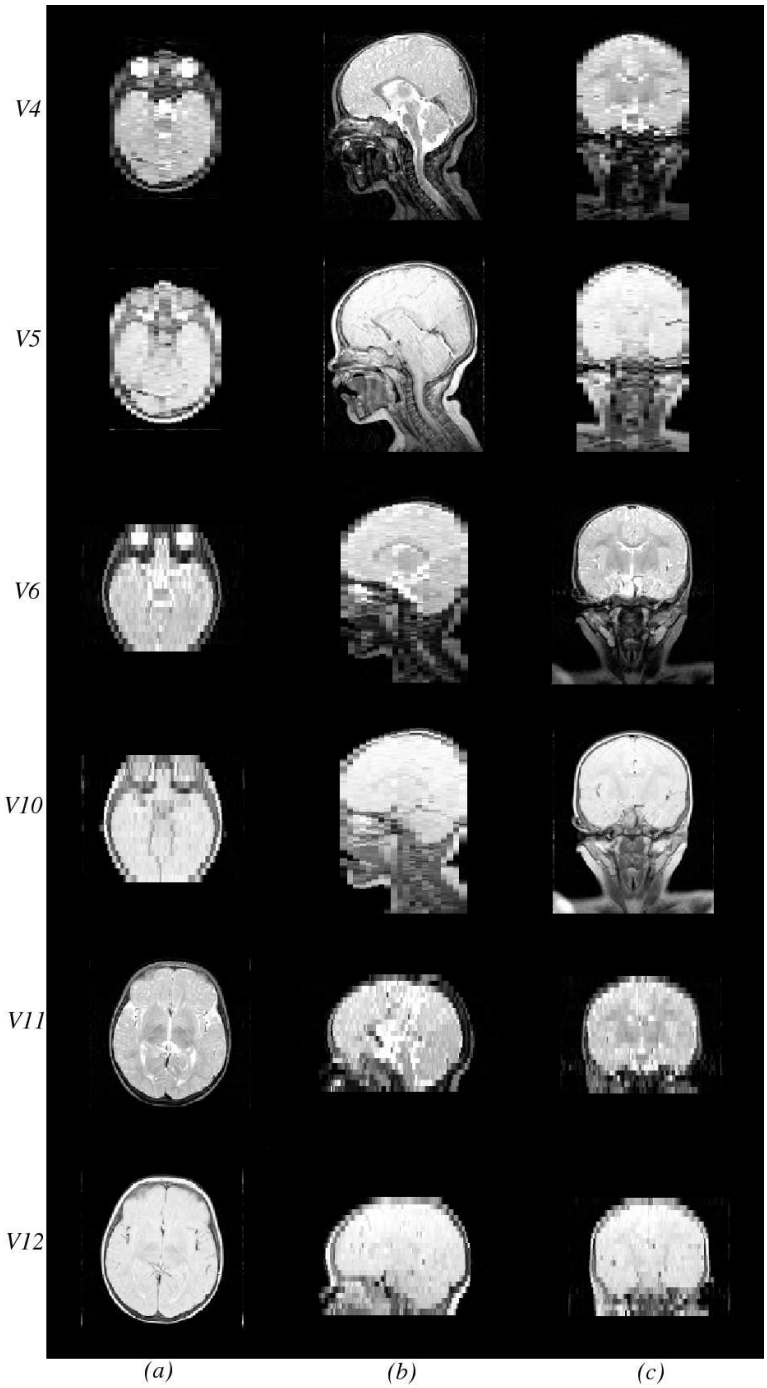


Figure 3.7: *(a)* Mid-transverse, *(b)* mid-sagittal and *(c)* mid-coronal exemplar unprocessed image slices taken from approximate mid-data set points in the infant data set volumes (V4-V6 and V10-V12). The age of the subject at date of acquisition was 10 months. These exemplar slices illustrate the quality of the infant data sets and the limited available anatomical image information due to the poor resolution in the various different non-acquisition planes.

Table 3.2: Arithmetic mean skull stripping results, divided into infant and adult developmental and evaluation test data sets.

Test Data Sets	SMSS Dice	BET Dice	SMSS FP	SMSS FN	BET FP	BET FN
Infants (Development)	0.93	0.90	0.10	0.04	0.19	0.00
Adults (Development)	0.92	0.83	0.15	0.01	0.29	0.00
Infants (Evaluation)	0.89	0.89	0.11	0.08	0.20	0.00
Adults (Evaluation)	0.86	0.86	0.22	0.00	0.25	0.00

determined. This process is sometimes repeated if pre-defined smoothness constraint constants of the surface have not been met.

Quantitative assessment of the skull-stripping performance was undertaken with the use of ground truth masks generated by an expert segmenter using software described in [39] for data sets V1-V12. Further ground truths utilized for data sets V13 to V20 are publicly available. The performance metrics included the Dice coefficient, [35, 126, 110] as this gives a scalar measure $[0, 1]$ for the performance of a segmentation overall (where 1 is representative of a perfect match):

$$\text{Dice} = 2.\frac{|A \cap B|}{|A| + |B|}, \tag{3.23}$$

where A and B are sets containing the segmentation result voxels and the ground truth voxels respectively, and $|A|$ represents the cardinality of elements in set A. False Positives (FP) and False Negatives (FN) were also calculated as these give an indication of the relative amount of under and over segmentation respectively. These metrics were determined for the BET segmentations and the SMSS segmentations in relation to the ground truth.

Parameter settings for SMSS included the window width parameter, $\chi = 3.00$, and the knee parameter, $\alpha = 0.10$ both empirically determined to provide reliable results for developmental data sets V1 to V9. Default parameter settings were used for BET.

3.4 Results and Discussion

The results of applying the skull stripping methodology to the data sets summarized in table 3.1 can be seen in figure 3.8 and summarized, using mean values, in table 3.2. Figures 3.10 and 3.12 show exemplar image slices from each of the resulting skull stripped data volumes utilizing SMSS, while figures 3.11 and 3.13 show the corresponding results obtained for BET [131].

Through visual comparison of the segmentation results with the quantitative results, one can state that the quantitative results reflect well the variable ability of the skull-stripping techniques to segment the CNS from the non-CNS tissue voxels.

If a segmentation result can be classed as high quality with an arbitrarily high threshold on the Dice coefficient, such as Dice$>$ 0.90, then SMSS demonstrates good performance

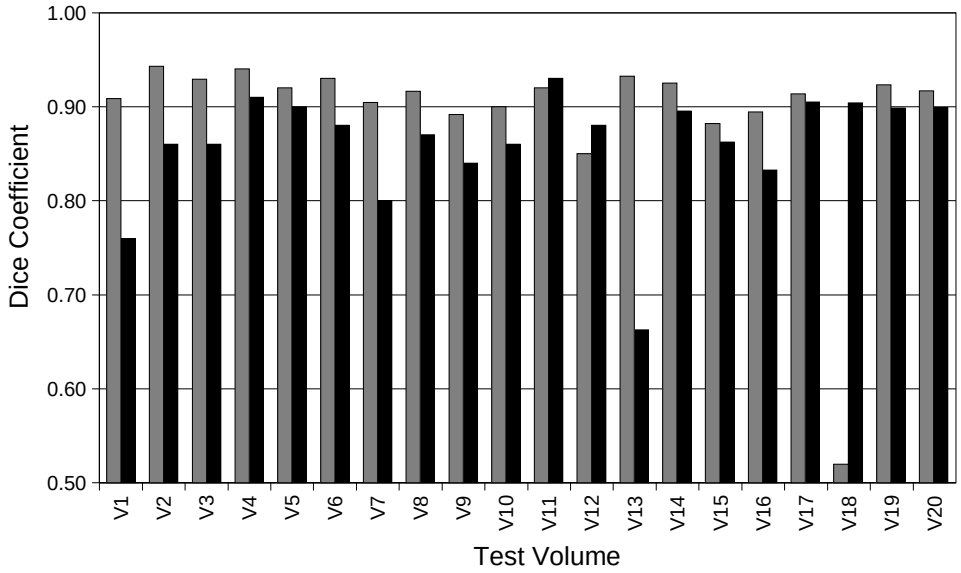


Figure 3.8: Dice coefficient results for SMSS (grey) and BET (black) skull stripped test volumes, V1 to V20. SMSS was developed with the aid of developmental data sets V1 to V9. Evaluation test data sets V10 to V20 were then used to evaluate the performance of SMSS.

for 6 out of the possible 11 evaluation test data sets. This can be compared with BET that shows high performance for only 3 evaluation test data sets. The numerical results obtained for BET are in general agreement with previously published results, e.g. [76].

BET consistently under-segments the CNS from the non-CNS voxels (demonstrated by a high number of FPs). Under-segmentation is where non-CNS voxels are not removed from the segmented data volume. SMSS provides segmentation results that tend to also under-segment, whilst also over-segmenting a small amount (higher number of FNs). BET provides two parameters that can be adjusted to control the quality of the segmentation, but it was found that these parameters increased over-segmentation whilst only reducing the under-segmentation by a small amount.

In contrast to the results obtained for evaluation test data sets V10-V17,V19 and V20, SMSS has produced poor results for evaluation test data set V18. This can be seen in the results presented in figure 3.8 and the sagittal, coronal and transverse images of the segmentation result obtained for V18 in figure 3.12. The primary reason for the poor segmentation result is due to under-estimation of the GM mean in the parameter estimation stage (stage (B) in figure 3.2). Therefore, if the window width parameter, χ from equation 3.9 is adjusted to 2.5, then the segmentation result improves from a Dice coefficient value of 0.52 to a value of 0.91. Exemplar transverse, coronal and sagittal image slices of this improved segmentation can be seen in figure 3.9. Improved parameter estimates could be determined automatically with the use of more advanced models that take account of PV effects and other imaging artefacts.

Skull stripping of infant data presents its own unique set of challenges because of the relatively poor voxel dimensions of the structure under study. Another important

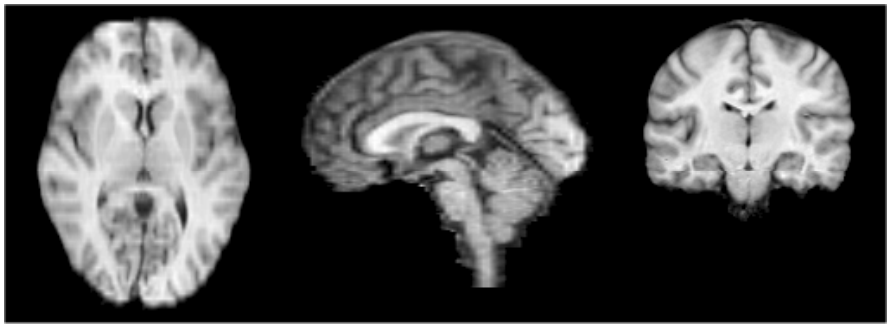


Figure 3.9: Improved segmentation result for evaluation test data set V18 with SMSS. A modified window width parameter, $\chi = 2.5$ was used, in place of the default value of 3.0, thus improving the segmentation result with a Dice coefficient value of 0.52 to a value of 0.91.

consideration with infant data is the dramatic changes in the WM and GM CNS tissue composition that occur in the first 2 years of life. Also, the CNS components of infant data appear to share very similar ranges of intensities. For this reason, the window $(\mathfrak{T}_2 + \mathfrak{T}_1)$ utilized by equation (3.9) defined by the means of the GM and WM CNS tissue components could not be determined. This range of intensities was then modelled as a single component Gaussian during the fitting process and as a result of this, the window width was derived from the standard deviation of this fitted Gaussian. The effect of this on the level of automation of the skull stripping algorithm is minimal, as programmatically, the age of the patient is available as part of the medical imaging file format data structure. The mean performance quantities for adults and infants in table 3.2, are presented independently, illustrating that the mean Dice coefficient for SMSS is approximately equal for both adults and infants. This is similarly so for BET, suggesting that both BET and SMSS are as capable of overcoming the challenges of skull-stripping infant neurological data.

As can be seen from table 3.1, the developmental and evaluation data sets consisted of data acquired in either the coronal, sagittal or transverse planes. SMSS utilizes a 2-D transverse plane brain target mask, I_{TM} in equation 3.17 to control the number of morphological operations. The use of a transverse plane target mask may be problematic when the data has not been acquired in the transverse plane. This is due to MRI data usually possessing coarser resolutions in the non-acquisition planes. Table 3.3 presents mean performance quantities for the developmental and evaluation test data sets divided into the three different acquisition planes. The results in table 3.3 appear to suggest that the performance of SMSS is not affected by the plane of acquisition. BET also appears to perform equally well for data that has been acquired in any acquisition plane. The results in table 3.3 were calculated for all 20 developmental and evaluation data sets to increase the statistical viability of the mean errors. This was due to the limited number of data sets acquired in the sagittal (4 data sets) and transverse (3 data sets) planes.

Table 3.3: Arithmetic mean skull stripping results, divided into data sets for different planes of acquisition.

Acquisition Plane	**SMSS Dice**	**BET Dice**	**SMSS FP**	**SMSS FN**	**BET FP**	**BET FN**
Transverse	0.91	0.86	0.16	0.00	0.26	0.00
Coronal	0.88	0.86	0.18	0.02	0.24	0.00
Sagittal	0.93	0.86	0.11	0.03	0.24	0.00

3.5 In Conclusion

The work in this chapter has demonstrated a newly developed fully automatic skull stripping methodology, SMSS. The results obtained using SMSS have shown that the technique could be used as part of a wider neurological tissue analysis framework, possibly as part of a consensus of skull-strippers, particularly as BET appears to consistently under-segment whereas SMSS under-segments less and over-segments more. This is confirmed by the quantitative analysis in relation to the expert segmented ground truth obtained for each of the test data sets. The quantitative performance of SMSS and BET appear to illustrate that SMSS and BET are adaptable to the various relative sizes in anatomy found in CNS of the infant in contrast to the usual problem of skull-stripping of adult MRI T1 weighted data sets. SMSS also appears to perform consistently regardless of the plane of acquisition. This is despite the use of a transverse 2-D target mask which results in a flexible and effective segmentation algorithm.

This chapter has developed an automatic methodology for isolating the entire brain component from neurological MR data. However imaging artefacts such as the Partial Volume (PV) effect have not been explicitly modelled in this skull-stripping methodology. Such modelling may improve the segmentation results. Furthermore many quantitative neurological applications require accurate estimates of the volume of individual neurological tissues, namely GM, WM or CSF. Therefore the remaining chapters of this thesis investigate models of the PV effect. These models are then used to classify neurological MR data in chapter 8 using a brain mask which excludes non-neurological tissue voxels. A brain mask could be obtained with the methodology developed in this chapter. The absence of non-neurological tissues reduces the number of classification components that have to be included in the classification procedure thus potentially reducing the classification error.

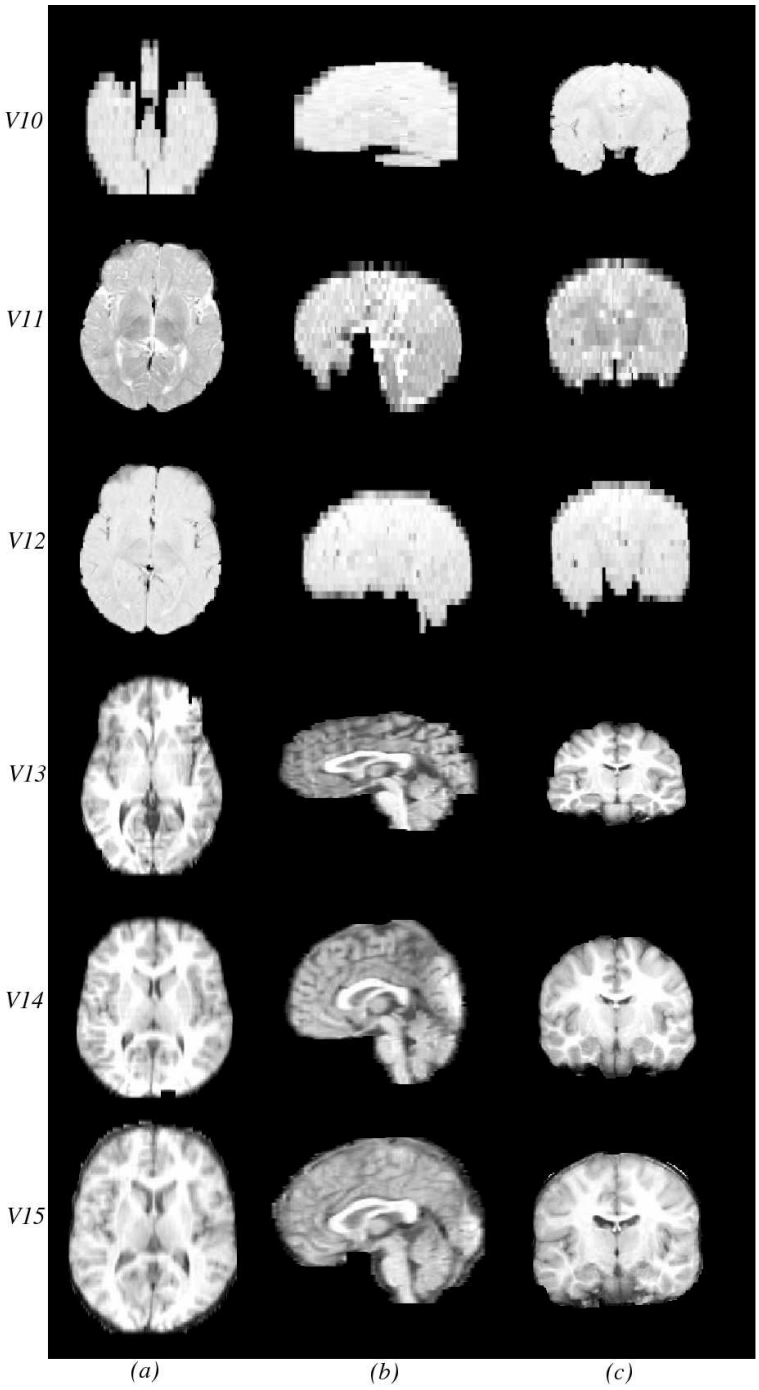


Figure 3.10: *(a)* Mid-transverse, *(b)* mid-sagittal and *(c)* mid-coronal image slices taken from the results of SMSS for evaluation test data sets V10 to V15. Each row represents a different data set, starting from V10 for row one and finishing with V15 for the final row (see table 3.1 for further details). Data sets V10 to V12 correspond to MRI scans of an infant.

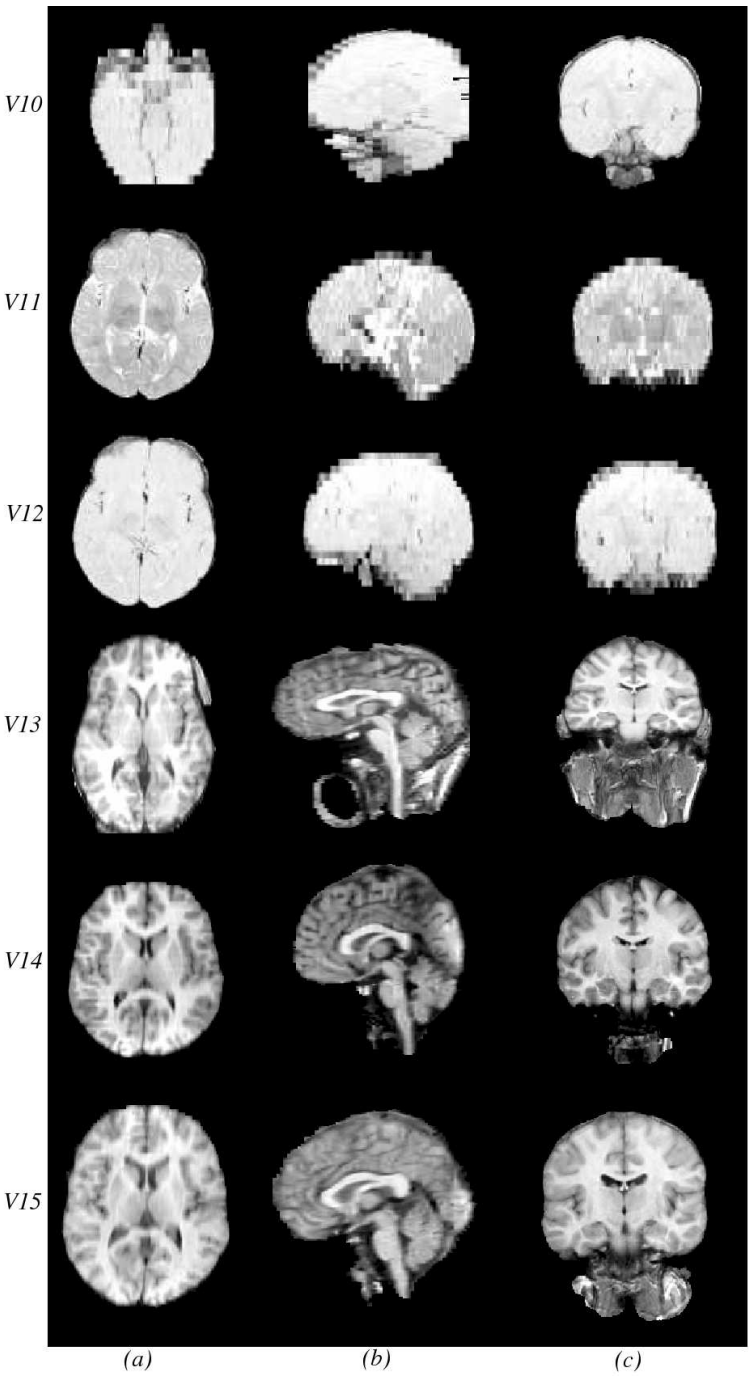


Figure 3.11: *(a)* Mid-transverse, *(b)* mid-sagittal and *(c)* mid-coronal image slices taken from the results of BET for evaluation test data sets V10 to V15. Each row represents a different data set, starting from V10 for row one and finishing with V15 for the final row (see table 3.1 for further details). Data sets V10 to V12 correspond to MRI scans of an infant.

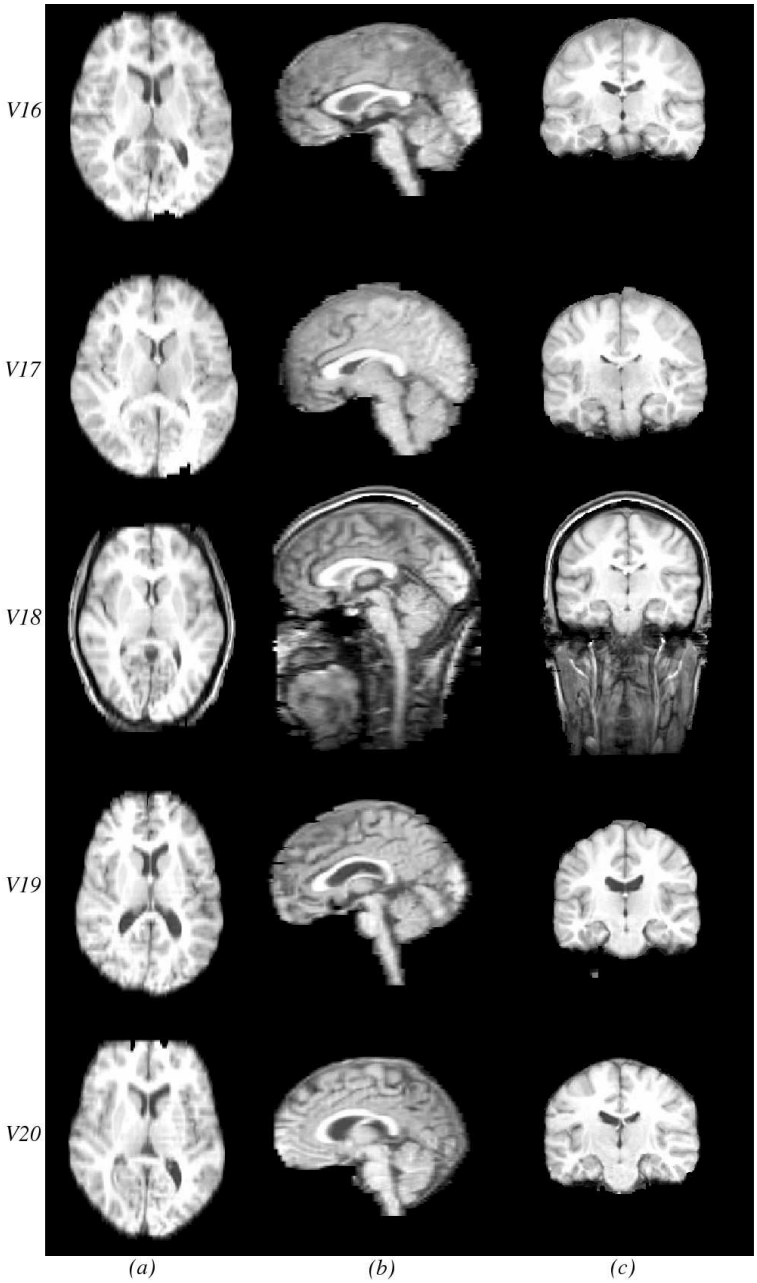


Figure 3.12: *(a)* Mid-transverse, *(b)* mid-sagittal and *(c)* mid-coronal image slices taken from the results of SMSS for evaluation test data sets V16 to V20. Each row represents a different data set, starting from V16 for row one and finishing with V20 for the final row (see table 3.1 for further details).

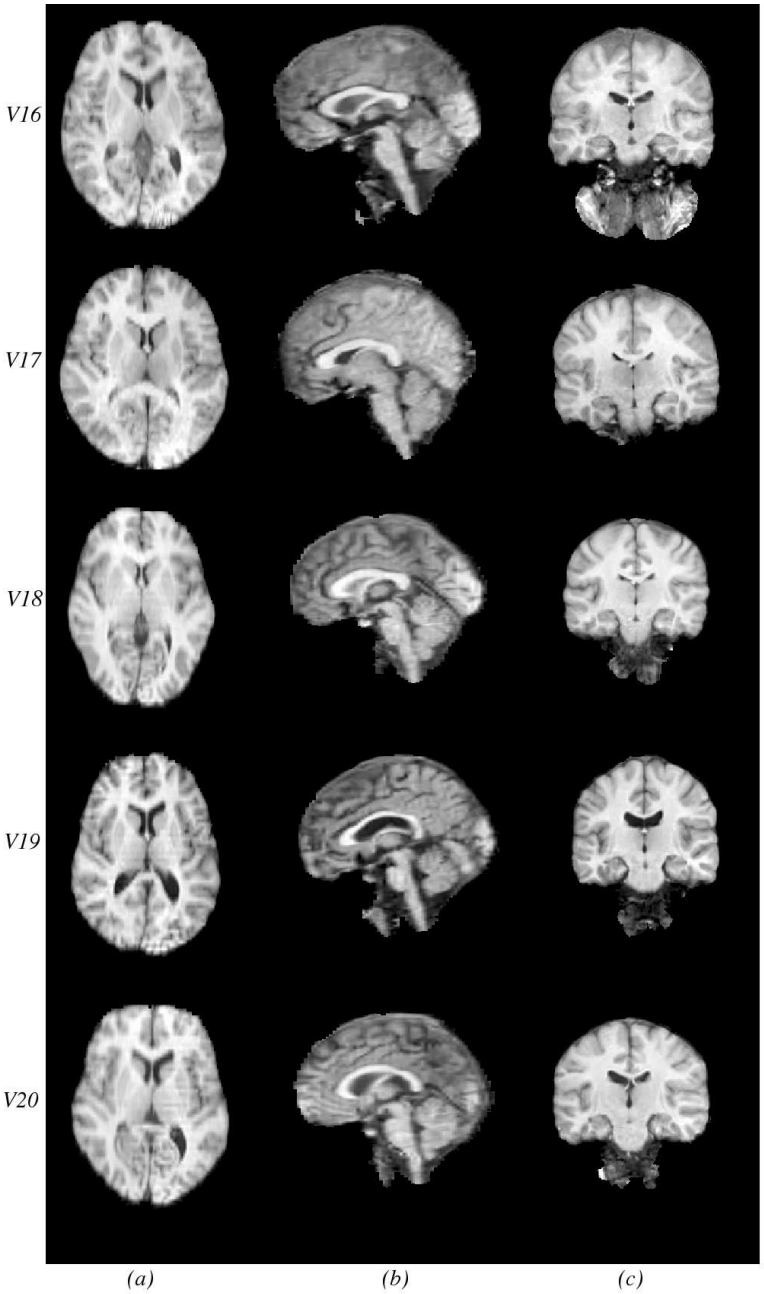


Figure 3.13: *(a)* Mid-transverse, *(b)* mid-sagittal and *(c)* mid-coronal image slices taken from the results of BET for evaluation test data sets V16 to V20. Each row represents a different data set, starting from V16 for row one and finishing with V20 for the final row (see table 3.1 for further details).

Chapter 4

Current Techniques in Probabilistic Partial Volume Modelling

Many people have investigated the Partial Volume (PV) effect with variable levels of rigour and or practicality. This chapter attempts to review some of the essential or seminal developments of modelling of the PV effect in probabilistic formulations. The chapter has been divided into three main topics: PV likelihood models, contextual information and estimation and inference. These topics have many inter-dependencies, but it is felt that their division helps to provide a cognate overview of modelling of the PV effect.

Likelihood models are often the central focus of modelling the PV effect, describing the probable range of frequencies of the measured information such as image intensity or image localised gradient magnitude information. Likelihood models make inference and estimation possible, where the tissue or material content of a voxel and the parameters of the likelihood models can be estimated or inferred upon using various different methods. Contextual information models such as Markov Random Fields (MRFs) enable the spatial information in an image to be utilized, often constraining an impossibly ill-posed problem to one that is well regulated.

Much classification work based on probabilistic techniques attempt to divide the imaging data into distinct regions each possessing distinct properties. These image regions are often described using likelihood models, with each distinct image region possessing different parameters such as mean intensities and standard deviations for Gaussian likelihoods. Such a model is often described by finite mixture theory, where individual likelihood components describe the statistical properties of each unique image region. A finite mixture model can be understood via Bayes theory. Bayes theorem provides a method for calculating the posterior probability, $P(\tau_a|\mathbf{x})$ of a particular image region, τ_a, i.e. an event, occurring given a particular voxel measurement, $\mathbf{x}$, such as an intensity:

$$P(\tau_a|\mathbf{x}) = \frac{p(\mathbf{x}|\tau_a).P(\tau_a)}{p(\mathbf{x})}, \tag{4.1}$$

where $p(\mathbf{x}|\tau_a)$ is the likelihood of a particular voxel measurement, $\mathbf{x}$ given an event τ_a. Intuitively, an image composed of a number of regions, one of which is denoted by the event τ_a will have a range of voxel measurements, $\mathbf{x}$. $P(\tau_a)$ is a scalar prior probability of the event τ_a occurring, which can be understood as the probability that the image modeller assigns to voxels in the image space being part of a particular image region, τ_a. The denominator, $p(g)$ is known as the marginal distribution because it can be calculated by marginalising out unwanted variables, like so:

$$P(\tau_a|\mathbf{x}) = \frac{p(\mathbf{x}|\tau_a).P(\tau_a)}{\sum_{\forall b} p(\mathbf{x}|\tau_b).P(\tau_b)}. \tag{4.2}$$

The marginal distribution, $p(g)$, provides a specification of the finite mixture model, where each likelihood, $p(\mathbf{x}|\tau_b)$, is scaled by the prior probabilities, $P(\tau_b)$.

Much finite mixture theory is not directly applicable to PV affected imaging data. Chapters 1 and 2 illustrated that the classification classes in PV affected data are not distinct entities and therefore simple finite mixture models do not accurately model PV affected data. Similarly, classical decision functions that assign a voxel to a particular classification class, e.g. τ_a, based on the the maximum posterior probability do not accurately model the continuous nature of PV affected data. Nevertheless, this has not prevented some authors from using the popular and well developed techniques of finite mixture theory in application to the PV effect problem. These approaches are discussed in a subsequent section, (4.1.2). But first, seminal work on probabilistic models of the PV effect are discussed, where the PV effect is modelled as a continuous classification problem.

4.1 Partial Volume Likelihood Models

4.1.1 Seminal Work

The seminal work on probabilistic classification of data that contains PV voxels was published by Choi *et al.* in 1991, [24]. This work presented a likelihood that could be used to model multi-channel Magnetic Resonance Imaging (MRI) data. The likelihood modelled the probability density of an m element vector of intensity levels, $\mathbf{x}$, given the set of parameters, θ and a mixing vector, $\boldsymbol{\alpha}$ ('label' vector):

$$p(\mathbf{x}|\theta, \boldsymbol{\alpha}) = \frac{1}{\sqrt{(2.\pi)^m.|\boldsymbol{\sigma}|}} \exp\left(-\frac{1}{2}\left(\mathbf{x} - \mathbf{M}^T\boldsymbol{\alpha}\right)\boldsymbol{\sigma}^{-1}\left(\mathbf{x} - \mathbf{M}^T\boldsymbol{\alpha}\right)\right). \tag{4.3}$$

The set of parameters,$\theta = \{\mathbf{M}\ \boldsymbol{\sigma}\}$ contains the multivariate mean matrix, $\mathbf{M}$, containing means for each image channel and tissue classification class and the noise co-variance matrix, $\boldsymbol{\sigma}$. Choi *et al.* assume $\boldsymbol{\sigma}$ arises purely from scanner noise and does not vary with the material being imaged (ie. $\boldsymbol{\sigma}$ is not influenced by such factors as tissue heterogeneity). Therefore $\boldsymbol{\sigma}$ is not tissue dependent, only image channel dependent. Also, this model is a multivariate Gaussian, and is therefore limited to applications where one assumes Gaussian distributed noise sources are applicable.

This model and all the following likelihood PV models assume what is known as linear mixing of the classification classes in the PV voxels. This assumption is now discussed.

The Linear Mixing Assumption in Biomedical Imaging Data

The assumption of linear mixing in PV likelihood models assumes that if a voxel is composed of a particular mixture or fraction of tissue or material classification classes, then the resultant signal detected by the image acquisition device will reflect this proportion linearly. Mathematically, the result of the linear combination of two independently distributed random variables or signals, X_a and X_b, is:

$$X_{a,b} = \alpha_1.X_a + \alpha_2.X_b, \tag{4.4}$$

so that the PV distributed random variable, $X_{a,b}$ is a linear mixture of the pure components, X_a and X_b, where $\alpha_1 \in [0, 1]$ and $\alpha_2 = 1 - \alpha_1$.

Windham *et al.* in 1988, [159], performed an experiment to determine whether the assumption of linear mixing is correct for their Principal Components Analysis (PCA) of the PV effect. Windham *et al.* provided the following explanation:

"If it is assumed that the magnetic resonance signal from a voxel containing more than one material is given by the volume weighted summation of the individual signals from the different materials, then the gray level of the corresponding pixel is the summation of the volume weighted gray levels of the different materials that would be obtained for voxels containing pure samples of the different materials. This is a reasonable assumption since the signal from a voxel is directly proportional to the net magnetization; the net magnetization is the sum of all of the individual magnetic moments provided that the frequency bandwidth across the voxel is larger than the chemical shift of the different materials in the voxel."

The NMR signal seen in a magnitude MR image has undergone a number of complicated physical processes, (see e.g. section 2.1). These physical processes together with the fact that MR images are usually calculated with a magnitude operation suggest that the actual functional relation between the random variables may not be linear. Indeed, as quoted above, Windham *et al.* do not refer to random variables. Despite this, even the grey level may not have a linear functional relationship with the voxel's constituents, dependent on the tissues being imaged and the image acquisition protocol.

4.1.2 Further Partial Volume Likelihood Models

An alternative probabilistic likelihood model of the PV effect was offered by Santago and Gage in 1993 [122]. Initially Santago and Gage's PV model was for PV voxels composed of a maximum of two Gaussian distributed classification classes [122]. This was then developed further in 1995 for multiple tissue PV voxels from Gaussian or Poisson noise sources [123]. The premise of this work was that the PV effect contributed an additional classification class to a finite mixture model, thus taking advantage of established finite mixture pattern recognition theory, (see e.g. [142]).

In [122, 123], pure tissue classes are described by the pure tissue distribution such as a Gaussian or a Poisson distribution. Mixture distributions can be thought of as consisting of an infinite number of these pure class distributions stretched over a uniform distribution that represents the range of intensities existing between the class means.

Thus, mixture distributions are determined via an integral of the product of a likelihood similar to equation 4.3 but for a single tissue class and a uniform distribution:

$$p(x|\mu_{a,b}(\mu_a,\mu_b)) = \int_0^1 p(x|\mu = \alpha.\mu_a + (1-\alpha).\mu_b).d\alpha. \tag{4.5}$$

This result is in fact equivalent to a convolution type operation, see e.g. [134]. Exemplar results of this operation are shown in figure 4.1.

This equation is conditioned on discrete events, unlike the phenomena that the equation is attempting to model and in contrast to equation 4.3 which is conditioned on a continuous event space.

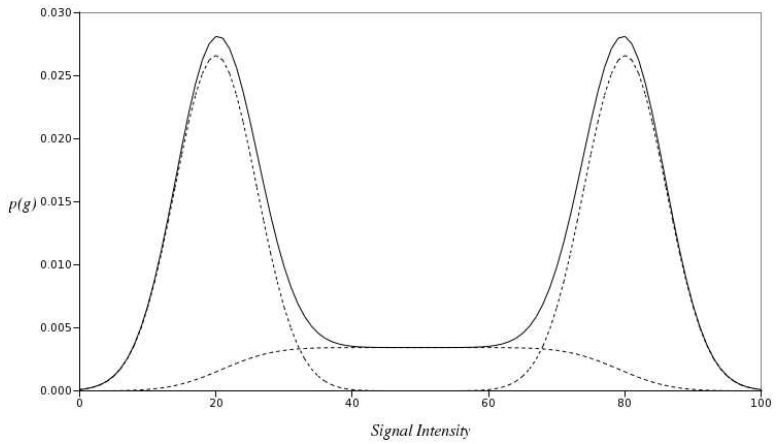
A second observation that can be made with regards to equation 4.5, is that the individual classification classes possess co-dependency between the parameters. i.e. the pure classification classes, $p(x|\mu_a)$ and $p(x|\mu_b)$ possess independent parameters, but the mixture class, $p(x|\mu_{a,b}(\mu_a,\mu_b))$ is dependent on the parameters from both pure classes, indicating that these classes should not really be treated as separate entities. At this stage in the discussion this might not seem to be particularly important, but many algorithms and techniques, especially of the finite mixture model nature assume the classes are independent entities.

A third observation can be made with regards to the formulation suggested by [122, 123] relates the convolution in equation 4.5 to Bayesian probability theory. The convolution is over a uniform function of α. This, as will be seen later (chapter 7), is very similar to the concept of an improper prior where the probability of obtaining a particular mixture at any particular intensity level is constant. An improper prior is one which does not specify any information about the expected mixture configurations that might occur as a result of the PV effect.

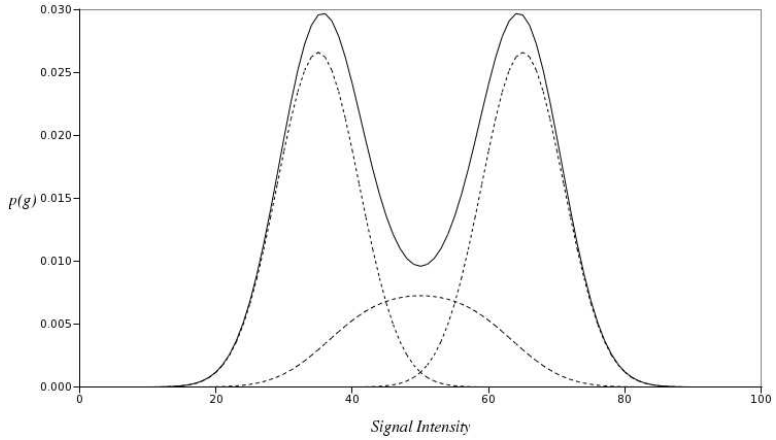
A further observation with regard to the approach taken by Santago and Gage is that the PV distribution between the two pure components requires a common width parameter for the two pure components. This therefore prevents accurate modelling of PV distributions that may arise from pure distributions with unequal width parameters.

Vokurka *et al.* in 2002, [148], presented an interesting development of equations 4.5 and 4.3. The formulation suggested in [148] replaced the explicit modelling of the probability density of the PV mixture value, α, or the probability of a voxel being from either a pure classification class or a PV mixture classification class with a probability analogy. In [148] the mixture content of a voxel is assumed to be equivalent to the probabilities of each of the classification classes for that voxel. A further novel feature of this model replaced the uniform prior convolved with Gaussians as used in 4.5 with two triangle priors convolved with Gaussians for each class. These are illustrated in figure 4.2. Each triangular function is convolved with each pure tissue Gaussian class individually to create two new mixture Gauss-triangle classification densities, $p_{PV}(x|\mu_a,\mu_b,\sigma_a)$ and $p_{PV}(x|\mu_b,\mu_a,\sigma_b)$:

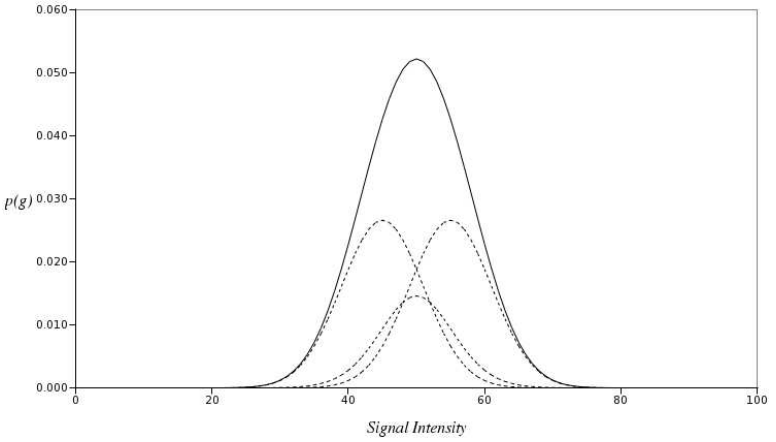
$$p_{PV}(x|\mu_a,\mu_b,\sigma_a) = \int_{\mu_a}^{\mu_b} p(x-\tau|\mu=0,\sigma_a).p_T^{a,b}(x=\tau|\mu_a,\mu_b).d\tau, \tag{4.6}$$



(a) means: 20, 80; standard deviations: 6, 6



(b) means: 35, 65; standard deviations: 6, 6



(c) means: 45, 55; standard deviations: 6, 6

Figure 4.1: Scaled PV PDF classification classes used by Santago and Gage in [122, 123] to model the distribution of intensities that occur as a result of the PV effect. Each figure illustrates two scaled pure distributions, the corresponding scaled PV distribution and the (normalised, i.e. unscaled) overall distribution (solid line). The mean values and the common standard deviation are given in each sub-figures' caption (a-c) and were used to obtain the PV distributions as a result of the convolution in equation 4.5.

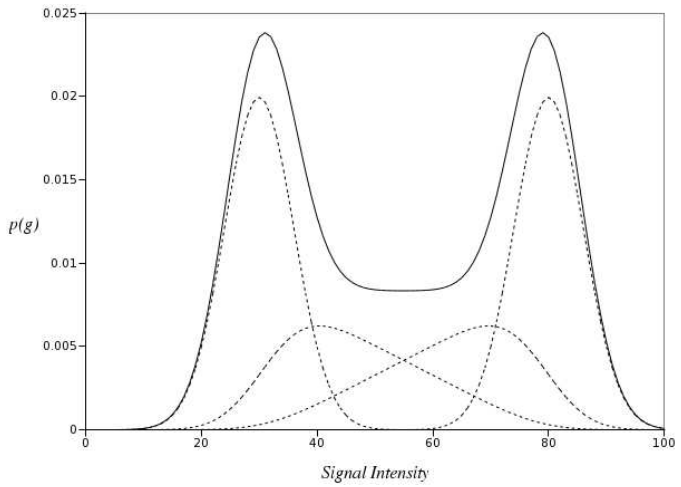


Figure 4.2: Pure functions (Gaussians), PV functions (rounded Triangular functions) and overall two class PDF (solid line). The PV functions are the results of the convolution with the Gaussian pure classification classes. These functions are used for the classification of volumetric data with the PV effect, as suggested in [148].

similarly:

$$p_{PV}(x|\mu_b,\mu_a,\sigma_b) = \int_{\mu_a}^{\mu_b} p(x-\tau|\mu=0,\sigma_b).p_T^{b,a}(x=\tau|\mu_b,\mu_a).d\tau. \tag{4.7}$$

Using equations 4.6 and 4.7 and pure classification class likelihoods, $p(x|\mu_a,\sigma_a)$ and $p(x|\mu_b,\sigma_b)$, it is possible to calculate the expected PV mixture composition via (for classification class a):

$$p(a|x,\mu_a,\mu_b,\sigma_a,\sigma_b) = \frac{P(a).p(x|\mu_a,\sigma_a) + P_{PV}(a,b).p_{PV}(x|\mu_a,\mu_b,\sigma_a)}{P(a).p(x|\mu_a,\sigma_a) + P(b).p(x|\mu_b,\sigma_b) + P_{PV}(a,b).(p_{PV}(x|\mu_b,\mu_a,\sigma_b) + p_{PV}(x|\mu_a,\mu_b,\sigma_a))}. \tag{4.8}$$

This equation represents the topic of inference which is described in more detail for the other models in a later section of this chapter (section 4.3).

In common with the models proposed by Santago and Gage, [122, 123], the convolution in equations 4.6 and 4.7 is analogous to an uniform PV prior density where the two equal but opposite triangular functions reflected about the point $\mu_a + (\mu_b - \mu_a)/2$ is equivalent to a flat PV prior density.

4.1.3 Partial Volume Mixture Prior Densities

PV mixture prior densities are used to describe the probability densities of PV mixtures in the absence of noise sources such as Gaussian or Poisson noise that might be found as a result of the medical image acquisition process. As has already been discussed, a number of people have assumed that this PV mixture prior density is or can at least be

assumed to be uniform (including, but not limited to [74, 3, 117, 128, 95]). In contrast to this approach a number of other researchers have used non-uniform mixture PV prior densities [6, 67, 78].

Bello *et al.* in 1998, [6] utilized Monte Carlo simulations of geometric shapes similar to Multiple Sclerosis lesions found in the human brain. These semi-ideal shapes were placed in a discretized space, representative of the sampling space that might be encountered in the MRI acquisition process. A distribution function was then simulated by counting the degrees of filling of the shapes' boundary voxels. No attempt was made to parameterize this PV mixture distribution function. The discretized sampling space was also uniform, which is not representative of the typical Point Spread Function (PSF) encountered in medical imaging modalities such as MRI [50].

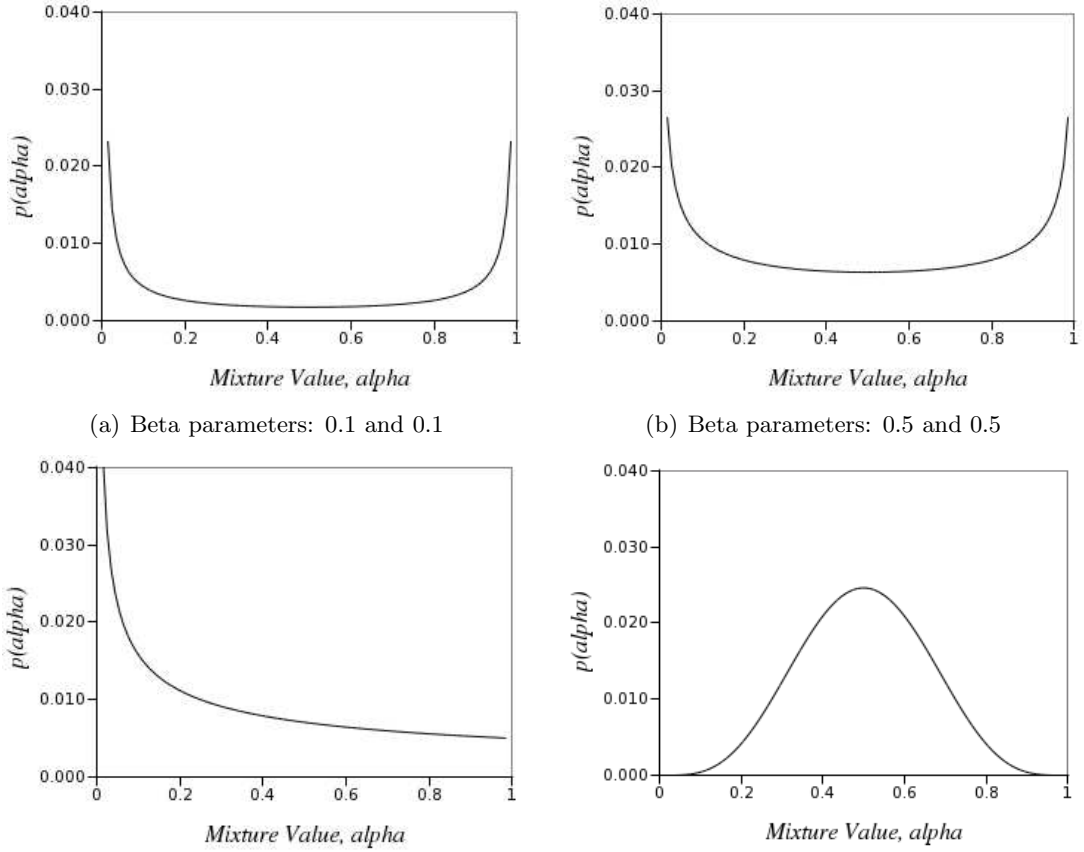
Links *et al.* in 1998, [83] performed a number of simulations of the PV effect with respect to MRI neurological data and found, for their particular simulations, that the assumption of a boxcar PSF instead of one characterised by a sinc PSF might not be of particular importance in classification algorithms. This statement was a result of observations made on the appearance of difference images between the various simulations of the brain data. One might choose to investigate this topic further so as to determine whether accurate modelling of the PSF and consequently the PV mixture prior distribution might improve classifier performance of PV mixture voxels. A companion paper by Reiss *et al.*, [109], utilize the software phantom as described in [83], with the measured PSF of the scanner. However, comparative classification performance with the two different PSFs within the simulated phantom is not investigated.

Related Mixture Prior Densities in Alternative Application Areas

Kitamoto and Takagi in 1999, [67], investigated the mixture prior density with application to remotely sensed imagery. Similar to [6], Kitamoto and Takagi assumed a uniform sampling space. Kitamoto and Takagi utilized statistical geometry to derive many different formulae for the various scenarios that they thought were relevant to their imaging domain. Kitamoto and Takagi found that the numerous formulae were impractical in real modelling applications, where a number of the formulae might be equally applicable for different regions of the image data. For example, one of the formula might describe a rectangular image pixel (boxcar PSF) with a size that is smaller than the rectangular object being imaged, thus creating edge effect mixed pixels. Another area of the image might be affected by mixed pixels that arose because the image pixel was larger than the object of interest. Therefore Kitamoto and Takagi decided to utilize the Beta distribution as they found that this distribution could approximate the variety of mixture distributions that their derived formulae might model. The Beta distribution is often used in probabilistic applications where a variable is the fraction of two other variables, which is the case for mixed pixel and PV problems [62]. The Beta distribution is given by:

$$p(\alpha|u_1, u_2) = \frac{1}{B(u_1, u_2)}.\alpha^{(u_1-1)}.(1-\alpha)^{(u_2-1)}, \tag{4.9}$$

where $B(u_1, u_2)$ is the Beta function. The parameters, u_1 and u_2 control the shape of this function. If $u_1 = u_2 = 1$, the result is a uniform distribution. If $u_1 = u_2 = 0.5$, the



(a) Beta parameters: 0.1 and 0.1

(b) Beta parameters: 0.5 and 0.5

(c) Beta parameters: 1.0 and 0.5

(d) Beta parameters: 5.0 and 5.0

Figure 4.3: Typical exemplar shapes of the Beta distribution for various parameter values (in captions for (a-d)). The Beta distribution was used by Kitamoto and Takagi in 1999, [67] to model various mixed pixel prior distributions in satellite imagery.

result is a concave distribution, similar to the mixture distributions given by Kitamoto and Takagi. Typical exemplar shapes of the Beta distribution for various parameter values can be seen in figure 4.3.

Kitamoto and Takagi provided no assessment of the impact of providing an improved modelling schema for the mixture prior distribution. The emphasis of their work was on providing a methodology for determining the formulae as a result of different sized (square) uniformly sampled pixels and regular shaped imaging objects of interest (such as squares or crosses). The PSF of the image acquisition process was not modelled.

Further Biomedical Imaging Prior Models

Van Leemput *et al.* in 2003, [78], compared classification performances of classifiers with uniform and non-uniform prior distributions. The non-uniform prior distribution has no functional constraint, taking values from what is estimated from the discretized data. This estimation step is described in a later section, 4.3. Figure 4.4 illustrates two pure

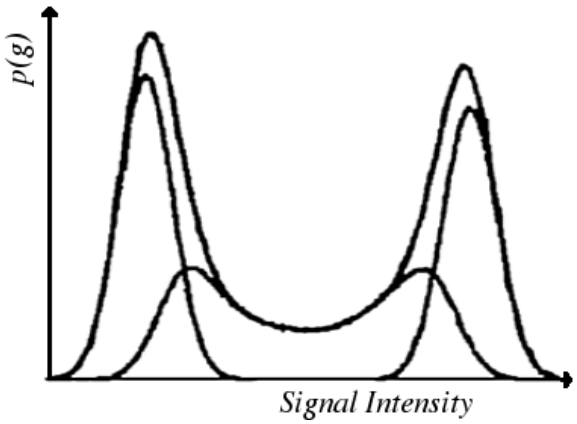


Figure 4.4: The Pure and PV distributions obtained by Van Leemput *et al.* in 2003, [78], by simulating the PV effect by down-sampling synthetically generated images. As can be seen from this figure, the centre distribution corresponding to PV voxels is not the result of a convolution with a flat distribution as used by other authors, (e.g. [123], compare with figure 4.1).

distributions and a single PV distribution obtained by Van Leemput by simulating the PV effect by down-sampling synthetically generated images (using a uniform sampling space-similar to [67] and [6]). The result is a non-uniform PV prior distribution.

Van Leemput *et al.* found that the use of a non-uniform PV prior distribution improves the classification performance when the overlap of the two pure distributions is small. The converse was found when the two pure distributions overlapped each other significantly. This could possibly be a result of the estimation step used by Van Leemput that imposed no prior functional constraint on the shape of this distribution, so that when the two pure distributions overlapped significantly, the estimation step was unable to determine the true shape of the underlying PV prior distribution.

An interesting conference paper by Joshi and Brady in 2005, [64], presented a methodology that attempts to model the down-sampling procedure inherent in any image or signal acquisition process, similar to the approach taken by Van Leemput *et al.* The methodology is formulated so as to allow for the high resolution data to be composed of random variables of any PDF. The down-sampling process is modelled as a uniformly weighted convolution of these basis PDFs, therefore implicitly assuming a uniform image acquisition PSF. Despite this the equivalent underlying prior distribution from a uniform PSF is not uniform and probably more closely approximates the true mixture prior distribution in biomedical imaging data in comparison to a uniform distribution, (see e.g. [78, 67]).

The Joshi and Brady methodology also utilizes a finite number of these basis PDFs, in contrast to the continuous nature of the underlying possible mixture configurations in any individual PV voxel. The argument being that the image data is discretized and therefore only a finite number of mixtures are inherently possible. Training is performed on user selected regions of pure classification class in the image data, but the suggested application is on MRI neurological data that has been found to be composed of 50%

voxels that might be affected by the PV effect (e.g. [102]).

4.1.4 Localised Image Gradient Magnitude Likelihood Models

In the preceding sections, all of the models discussed only utilized first order intensity information. Some models of the PV effect incorporate further information such as second order information in the likelihood models. An extension of the work in [148], by Williamson *et al.* in 2002, [157] utilized an approximation of the two-dimensional gradient magnitude of the image intensity data. The MRI data was assumed to possess voxels that have Gaussian distributed noise, thereby allowing the result of Rice in 1938, [111] to describe the probabilistic localised image gradient magnitude data. Rice found that the sum of two random variables that are governed by Gaussian PDFs results in a Rice PDF [1]. This density was introduced in section 2.1 as MRI data is often considered to possess a Rician PDF. The Rician density approaches a Gaussian density for high signal to noise ratios and a Rayleigh density for low signal to noise ratios.

Williamson *et al.* found that it was useful to incorporate extra probabilistic second order information in the form of the intensity gradient magnitude into their likelihood equation describing pure and PV distributed voxels. Pure material or tissue regions were deemed to possess a Gaussian distributed intensity PDF and a Rayleigh distributed intensity gradient magnitude PDF. For voxels that are considered to be PV voxels, the PV Gauss-Triangle densities as described in [148] (see equation 4.6) are also used by Williamson *et al.* However, Williamson *et al.* also utilize the gradient magnitude information. Voxels that possess relatively higher gradient magnitudes are deemed to more likely contain a mixture of tissues, i.e. PV voxels. The justification for this is that PV voxels are often located on boundary regions. In terms of the mathematics, the joint intensity, g and intensity gradient magnitude, z, likelihood for PV voxels is given by:

$$p_{PV}(g,z) = p_{PV}(g).p_{PV}(z|g), \tag{4.10}$$

or by making explicit some of the parameters required by both PDFs:

$$p_{PV}(g,z|\mu_a,\mu_b,\sigma_a) = p_{PV}(g|\mu_a,\mu_b,\sigma_a).p_{PV}(z|\mu_z = A(g|\mu_a,\mu_b),\sigma_z = A_\sigma(\sigma_a,\sigma_b)), \tag{4.11}$$

where $A(g|\mu_a,\mu_b)$ is given as a half-circular function by Williamson *et al.* μ_a, μ_b and σ_a, σ_b are the means and standard deviations for classification classes a and b. μ_z is a centering parameter for $p(z|\mu_z,\sigma_z)$ which is a complicated function, especially due to its dependence on $A_m(g|\mu_a,\mu_b)$ and $A_\sigma(\sigma_a,\sigma_b)$:

$$p_{PV}(z|\mu_z,\sigma_z) = \frac{z}{\sigma_z^2}.\exp\left(-\frac{z^2+\mu_z^2}{2.\sigma_z^2}\right).I_0\left(\frac{\mu_z.z}{\sigma_z^2}\right), \tag{4.12}$$

where $I_0(x)$ is the modified Bessel function of the first kind, order zero.

An idealised plot of the pure and PV equations can be seen in figure 4.5.

Critical points about this method can be made that were also made about the methods of Santago and Gage [122, 123] and separately about the methodology of Vokurka *et*

[1] The Rice PDF is named after S.O. Rice. Another name for this density is a Rician PDF

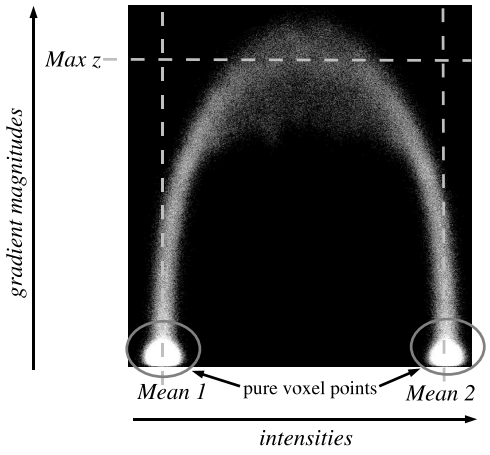


Figure 4.5: Exemplar plot of the intensity and intensity gradient magnitude feature space. The central arc, whose mean value is given by $A(g|\mu_a, \mu_b)$, describes likely PV values while the rounded regions with lower intensity gradient magnitude values correspond to pure classification class voxel data.

al., [148]. Each of these methods create separate classification classes for the PV voxels, but the PV classification classes are not really separate classification classes because the PV classification classes share parameters with the pure voxel classification classes. These methods also implicitly assume uniform PV prior distributions, a fact that is obscured by their analogical forms.

The work in [157] was extended from singular acquired MRI image sequences to multiple image sequence data by Thacker *et al.* in 2004 in [139]. They use a heuristically defined density to describe the joint intensity gradient magnitude for multiple image sequences, given by:

$$p_{PV}(z'|\mu_z) = \frac{z^\gamma}{\mu_z^{(\gamma+1)}}.\exp\left(-\frac{z^2}{2.\mu_z^2}\right), \tag{4.13}$$

where z' is the sum of the 2-D gradient magnitudes calculated for each image sequence.

In [157] and [139], both methods utilize localised image gradient magnitude likelihoods that are dependent on the localised image intensity to calculate μ_z. The localised gradient magnitude could be dependent upon the true PV mixture composition of the voxel rather than a noise affected voxel intensity. Another factor not considered in the models in [157] or [139] is the data is inherently three dimensional, but both models only utilize 2-D gradient magnitude information. In [139], the authors suggest that relatively larger inter-slice distances relative to in-plane voxel distances present difficulties in determining a theoretical description of the PDF that might arise, despite [139] utilizing a heuristically defined density.

4.2 Contextual Information

Contextual information in pattern recognition might refer to any number sources of available information, but in the imaging domain it often refers to the fact that pixels

or voxels are inherently spatially located and any neighbouring voxels can aid in the classification process of the individual voxel in question.

As discussed in the previous section, the localised gradient magnitude likelihood does indeed incorporate some contextual information, in the form of the localised magnitude of the gradient. This allows it to determine if a voxel might be more likely to be from a PV voxel but not if a voxel has been correctly classified by comparing its classification assignment with the classification assignments of its neighbouring voxels. Thus, it does not impose explicitly modelled spatial regularisation on the aforementioned likelihood models.

4.2.1 Voxel Local Histograms

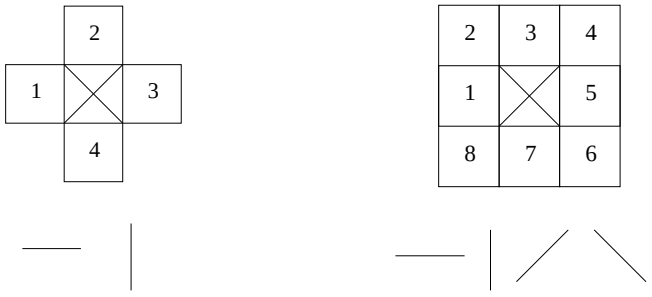
An interesting methodology in a paper by Laidlaw *et al.* in 1998, [74], incorporates spatial information into the classification of individual voxels affected by the PV effect through the creation of localised image histograms. Some similarity can be found with this work and the work of Joshi and Brady, [64]. Both methods derive types of histogram basis functions, except Laidlaw *et al.* fit their basis functions to histograms of a neighbourhood of voxels rather than Joshi and Brady who calculate the probability of a point measurement coming from their different basis functions.

The work of Laidlaw *et al.* attempts to reconstruct localised image histogram functions by taking voxel measurements from a cubic region in the image data and up-sampling this cubic region using tricubic B-splines. Laidlaw *et al.* then fit their pure and PV basis functions to this up-sampled region. A possible problem that is not discussed in this work could possibly arise from the high sensitivity of the localised image histograms to particularly high levels of voxel intensity noise. Therefore this might often prevent accurate localised image histograms from being created, and therefore preventing the correct basis functions from being fitted.

Laidlaw *et al.* also assume the image acquisition process can be characterised by a boxcar PSF. They state that this is sufficiently accurate in practise in comparison to other unpublished calculations and results obtained with other PSFs. Their chosen application is volume rendering and they evaluated their methodology on simulated PV data, although no information was given about how the data was simulated, such as, for example, the PSF used to perform the down-sampling of the high resolution data. Their performance assessment utilizes multiple simulated image channels, therefore reducing the sensitivity of the methodology to individual image channel voxels with high levels of image intensity noise.

4.2.2 Markov Random Fields

The work of Laidlaw *et al.*, [74], is interesting as it creates a probabilistic framework that takes account of the inherent spatial information available in imaging data. Probability theory does however have an established theory to take into account contextual spatial information, namely MRFs.



(a) 1^{st} order (b) 2^{nd} order

Figure 4.6: Cliques, $\mathfrak{N}_{\boldsymbol{\omega}_i}$, utilized by Choi *et al.* [24]. (a) illustrates first order neighbourhood system (top) and cliques (bottom). (b) illustrates second order neighbourhood system (top) and cliques (bottom). Cliques illustrated are first and second order pair-wise neighbourhood interaction cliques as utilized by Choi et al.

As with the likelihood models of the previous section, the seminal paper by Choi *et al.* in 1991 not only treated the likelihood aspects of the PV effect but also combined MRF theory into the classification framework [24]. The MRF approach taken by Choi *et al.* utilized second order pair-wise neighbourhood interaction. This neighbourhood action produces pairs of pixels known as *cliques*, $\mathfrak{N}_{\boldsymbol{\omega}_i} = \{\boldsymbol{\omega}_j | \text{where } \boldsymbol{\omega}_j \text{ is a neighbour of point } \boldsymbol{\omega}_i\}$. These cliques help to define what is known as a MRF, where the conditional probabilities are defined over these neighbouring regions, this is discussed in more detail shortly. The cliques utilized by Choi *et al.* can be seen in figure 4.6.

Only two dimensional contextual information was taken into account despite Choi *et al.* utilizing inherently three dimensional data. One must note however, that the theory developed by Choi *et al.* was seminal and at the time, the distance between individual slices in the MRI data that Choi *et al.* chose to apply their methodology was probably significantly greater than the inter-pixel distances in plane. This was corrected in 1997 when Choi *et al.* extended the two-dimensional MRF to three dimensions [25], thus improving the segmentation.

In MRF theory, each clique possesses what is known as a clique potential. This clique potential attempts to quantify the dissimilarity between the voxel in question in comparison to the other voxels in the clique, i.e. neighbouring voxels in this case. Choi *et al.* opted for a very simple clique potential:

$$V_c(\boldsymbol{\omega}_i) \equiv V_{\boldsymbol{\omega}_i,\boldsymbol{\omega}_j} = \kappa_{\boldsymbol{\omega}_j} \|\boldsymbol{\alpha}_{\boldsymbol{\omega}_j} - \boldsymbol{\alpha}_{\boldsymbol{\omega}_i}\|^2, \tag{4.14}$$

where $\kappa_{\boldsymbol{\omega}_j} = \frac{1}{\sqrt{2}}$ when pixel $\boldsymbol{\omega}_j$ is a diagonal neighbour of pixel $\boldsymbol{\omega}_i$ or $\kappa_{\boldsymbol{\omega}_j} = 1$ when pixel $\boldsymbol{\omega}_j$ is an orthogonal neighbour to pixel $\boldsymbol{\omega}_i$ and c is a clique, from a set of available cliques, $c \in \mathcal{C}$.

A MRF is a type of random field with probabilities that are only conditionally dependent on the neighbouring pixels or voxels, not on the pixels or voxels outside of these neighbouring points or region. This property is known as the dependence Markov property. The locally defined conditional probabilities and the full probability of the entire field also have to conform to a positivity condition in order to be classed as Markovian.

This positivity condition states that all the configurations have to possess non-zero probability, i.e. to be possible. These two conditions, positivity and dependence, are sufficient for a random field to be classed as Markovian. The dependency property is especially useful as it allows the probability of individual voxels to be calculated without having to utilize the state of the entire random field.

The Gibbs Distribution

The Gibbs distribution from statistical physics, [44], is usually cited as fulfilling these Markovian conditions, where the Gibbs distribution for a particular state, η, is given by:

$$P(E_\eta|T) = \frac{1}{Z}\exp\left(-\frac{E_\eta}{k_B.T}\right), \tag{4.15}$$

where E_η is the energy of the system at state η, k_B is known as Boltzmann's constant and T is the absolute temperature (i.e. in degrees kelvin). Z is the normalisation factor, sometimes referred to as the *partition function*,

$$Z = \sum_\eta \exp\left(-\frac{E_\eta}{k_B.T}\right); \tag{4.16}$$

so that

$$\sum_\eta P(E_\eta|T) = 1. \tag{4.17}$$

The Gibbs distribution is used in statistical physics to determine the probability that a system is in a particular state, with energy, E_η for a given temperature, T. Intuitively, a system at high temperature is likely to possess greater energy, whereas at lower temperatures, such a high energy state is not as likely. The Gibbs distribution can be used to describe the probability of the underlying voxel states of an image or data volume where the states may refer to tissues or activities. If the image is highly disordered (i.e. many voxels have different states from each other) then the image can be considered to possess high energy. But if the temperature is relatively low, then such a state of disorder is not as likely. Optimization of the system parameters may reduce the disorder of the system by altering the states of individual voxels so that they become more similar to each other. This optimization will therefore lead to a less disordered system and thereby a more probable state.

The more common form of the Gibbs distribution as used in computational imaging science is given by:

$$P(x) = \frac{1}{Z}\exp\left(-H(x)\right), \tag{4.18}$$

where $H(x)$ is sometimes referred to as the energy function, see e.g. [160]. The Boltzmann constant and absolute temperature parameter are usually replaced by other parameters that can be directly related to the imaging problem. Imaging science typically defines these energy functions in terms of clique potentials, such as the clique potential used by Choi *et al.*, [24], described earlier (equation 4.14). These clique potentials measure the dissimilarity of a voxel's state with the state of the neighbouring voxels.

A Gibbs distribution can be shown to be equivalent to a MRF with the following argument (similar to an example given by Won and Gray in [161]).

The conditional probability of the state of voxel $\boldsymbol{\omega}_i$ given the state of all the other voxels is given by:

$$P(x_{\boldsymbol{\omega}_i}|x_{\boldsymbol{\omega}_j}\ \forall\ j \neq i) = \frac{P(x_{\boldsymbol{\omega}_k}\ \forall\ k)}{P(x_{\boldsymbol{\omega}_j}\ \forall\ j \neq i)} = \frac{\frac{1}{Z}\exp\left(-H(x_{\boldsymbol{\omega}_k}\ \forall\ k)\right)}{\frac{1}{Z}\exp\left(-H(x_{\boldsymbol{\omega}_j}\ \forall\ j \neq i)\right)}, \tag{4.19}$$

where the energy functions take the form, e.g.

$$H(x_{\boldsymbol{\omega}_k}\ \forall\ \boldsymbol{\omega}_k) = \sum_{\forall\ k} H(x_{\boldsymbol{\omega}_k}). \tag{4.20}$$

Cancelling the common factors and noting that the denominator can be found by marginalising over all values for voxel $\boldsymbol{\omega}_i$:

$$P(x_{\boldsymbol{\omega}_i}|x_{\boldsymbol{\omega}_j}\ \forall\ j \neq i) = \frac{\exp\left(-H(x_{\boldsymbol{\omega}_k}\ \forall\ k)\right)}{\sum\limits_{\forall x_{\boldsymbol{\omega}_i}} \exp\left(-H(x_{\boldsymbol{\omega}_k}\ \forall\ k)\right)}. \tag{4.21}$$

The denominator and numerator can then be split into two, the first terms referring to the energy functionals not dependent on $x_{\boldsymbol{\omega}_i}$, so that:

$$\begin{aligned} P(x_{\boldsymbol{\omega}_i}|x_{\boldsymbol{\omega}_j}\ \forall\ j \neq i) &= \frac{\exp\left(-H(x_{\boldsymbol{\omega}_j}\ \forall\ j \neq i)\right).\exp\left(-H(x_{\boldsymbol{\omega}_i})\right)}{\exp\left(-H(x_{\boldsymbol{\omega}_j}\ \forall\ j \neq i)\right).\sum\limits_{\forall x_{\boldsymbol{\omega}_i}} \exp\left(-H(x_{\boldsymbol{\omega}_i})\right)} \\ &= \frac{\exp\left(-H(x_{\boldsymbol{\omega}_i})\right)}{\sum\limits_{\forall x_{\boldsymbol{\omega}_i}} \exp\left(-H(x_{\boldsymbol{\omega}_i})\right)}. \end{aligned} \tag{4.22}$$

This states that the conditional probability of voxel $\boldsymbol{\omega}_i$ is only dependent on the energy functionals that utilize $\boldsymbol{\omega}_i$. So that for the clique potential used by Choi *et al.* in 1991, described earlier (equation 4.14) allows for the conditional probability of a voxel's state to be fully specified from itself and over the neighbouring voxels. This greatly simplifies any calculation of image 'roughness' or other dissimilarity measure that might otherwise be conditionally dependent on the states of the voxels from the entire data volume.

The preceding argument has illustrated that the Gibbs distribution possesses the dependency Markovian property. Further evidence is usually required by the contextual probability theory community as to the suitability of this approach and to provide further details of the relationship between the probabilistic description of the entire data and how it relates to the probabilistic description on a per voxel, i.e. conditional basis. A number of approaches are given in the literature, see e.g. [8, 26, 160, 161], which typically reference what is commonly known as the Hammersley and Clifford theorem, after an unpublished document in 1971, [53]. There appears to be a number of interpretations of this theorem. For example, in [8, 30], the theorem is stated in terms of the cliques and how these govern the resulting global probability distribution. But [161] state the theorem in terms of the Gibbs distribution being the only valid statistical

description of a MRF. [8, 26] state this was the problem that Hammersley and Clifford originally set out to solve, i.e. what is the most general form for the probabilistic description of a MRF.

In any case, most of these texts utilise a similar set of theorems, corollaries and lemmas surrounding the Hammersley Clifford theorem to establish a probabilistic form that implements the MRF conditions. The conclusion of which is that the Gibbs distribution is the most general form that implements the Markovian properties.

Choi *et al.*, [24], therefore assume that the clique potential in equation 4.14 is governed by a Gibbs distribution that can be used to implement a MRF, (letting $H(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\boldsymbol{\omega}_i)$):

$$P(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}) = \frac{1}{Z_{\boldsymbol{\omega}_i}} \exp\left(- \sum_{c \in \mathcal{C}_{\boldsymbol{\omega}_i}} V_c(\boldsymbol{\omega}_i) \right) = \frac{1}{Z_{\boldsymbol{\omega}_i}} \exp\left(-\beta . \sum_{\forall \boldsymbol{\omega}_j \in \mathfrak{N}_{\boldsymbol{\omega}_i}} \kappa_{\boldsymbol{\omega}_j} \| \boldsymbol{\alpha}_{\boldsymbol{\omega}_j} - \boldsymbol{\alpha}_{\boldsymbol{\omega}_i} \|^2 \right), \tag{4.23}$$

where Z is a normalizing factor known as the *partition function* and

$$\mathfrak{N}_{\boldsymbol{\omega}_i} = \{ \boldsymbol{\omega}_j | \text{where } \boldsymbol{\omega}_j \text{ is a neighbour of point } \boldsymbol{\omega}_i \}. \tag{4.24}$$

Choi *et al.* also assumed that the MRF is isotropic and homogeneous, so that the clique regularisation parameter, β, is independent of location $\boldsymbol{\omega}_i$ and cliques, $c \in \mathcal{C}_{\boldsymbol{\omega}_i}$ respectively [80]. These assumptions simplify the model but one could argue that the model might be improved (depending on the data) by not considering the data to be isotropic or homogeneous. Neurological MRI data may not be isotropic if one considers some regions of the brain to be composed of a variety of densities of WM and GM such as the Putamen in relation to other regions of the brain composed of GM or WM [31].

In terms of the homogeneous property of the MRF, it is clear that at boundary regions in image data, this homogeneity criterion is not valid.

Inhomogeneous MRF (Allowing for Edge or Transition Regions in the Image Data)

The work of Wang *et al.* in 2001 introduced an inhomogeneous MRF applied to MRI data [150], therefore allowing for variable spatial regularisation to be applied to the imaging data. The inhomogeneous MRF was modelled using discrete classification class labels, $l_{\boldsymbol{\omega}_i}$, with the following Gibbs distribution:

$$P(l_{\boldsymbol{\omega}_i}) = \frac{1}{Z_{\boldsymbol{\omega}_i}} \exp\left(-\beta_{\boldsymbol{\omega}_i} . \sum_{c \in \mathcal{C}_{\boldsymbol{\omega}_i}} V_c(\boldsymbol{\omega}_i) \right), \tag{4.25}$$

where

$$V_c(\boldsymbol{\omega}_i) \equiv V_{\boldsymbol{\omega}_i, \boldsymbol{\omega}_j} = \begin{cases} 1 & \text{if} \quad l_{\boldsymbol{\omega}_i} = l_{\boldsymbol{\omega}_j} \\ 0 & \text{if} \quad l_{\boldsymbol{\omega}_i} \neq l_{\boldsymbol{\omega}_j} \end{cases} . \tag{4.26}$$

The value of the Gibbs regularisation parameter, $\beta_{\boldsymbol{\omega}_i}$ for voxel $\boldsymbol{\omega}_i$ was determined using a locally defined entropy rate, an information theoretic concept that attempts

to measure local changes in entropy. Boundary regions are defined as more likely to occur when the entropy rate is greater than surrounding regions. The entropy rate was calculated over an initial labelling of the data without the use of the MRF (equations 4.25 and 4.26). $\beta_{\boldsymbol{\omega}_i}$ was then calculated using this entropy rate measure in conjunction with some empirically defined parameters. Therefore the calculation of $\beta_{\boldsymbol{\omega}_i}$ is based partially on theoretically sound principles, while the use of empirically defined parameters might cause problems in the application of the methodology to previously un-seen MRI data sets.

Another problem with equations 4.25 and 4.26 is the use of discrete pixel or voxel labels, $l_{\boldsymbol{\omega}_i} \in \{l^1 \;\; l^2 \; ... \; l^K\}$ (for K classification classes) in comparison to the continuous pixel labels of equations 4.14 and 4.23, $\boldsymbol{\alpha}_{\boldsymbol{\omega}_i} = ((\alpha_{\boldsymbol{\omega}_i,1} \in [0,1]) \;\; (\alpha_{\boldsymbol{\omega}_i,2} \in [0,1]) \; ... \; (\alpha_{\boldsymbol{\omega}_i,K} \in [0,1]))^T$ (which more accurately models the true nature of the PV effect). Wang *et al.* utilize Gaussian representations of classification classes that represent both pure and PV classification classes. The use of a Gaussian for the PV classification classes vastly simplifies the development of the model, especially with regards to the optimization and estimation process, but is rather inaccurate with regards to the true distribution of the PV voxels (e.g. see figures 4.1, 4.2 and 4.4). Also the continuous nature of the PV effect can not be accurately represented by discrete classification classes.

Bias Field Modelling in MRI Data

Many methods utilizing MRF theory that have been applied to medical data published since the seminal work of Choi *et al.* in 1991 have not explicitly modelled the PV effect. A popular example is the work of Held *et al.* [57]. Held *et al.* built on the work of Wells *et al.* [154]. Both methods utilize a type of kernel density estimation, often known as Parzen window estimation ([151, 97]). The work of Wells *et al.* and Held *et al.* was motivated by the effect of magnetic field inhomogeneities resulting in non-stationary image statistics across the data (known as the bias field from here). This was also the motivation of the work of another popular paper by Zhang *et al.*, [163]. Zhang *et al.* utilized a Hidden Markov Random Field (HMRF) to model the underlying bias field in the MRI data. It could be argued that if explicit modelling of the PV effect was used within the models of these papers then it could result in particularly complicated models. These complicated models could prove to be difficult to perform parameter estimation of the true parameters particularly as both the bias field and the PV effect produce a continuum between otherwise discrete tissue labels (e.g. GM and WM for neurological MRI data).

Additional Regularisation Terms

Ruan *et al.* in 2000 [117] used a similar model to that used by Wang *et al.* in 2001 [150], except Ruan *et al.* assumed a semi-homogeneous MRF, so that the MRF regularisation parameter, $\beta_{\boldsymbol{\omega}_i} = \beta_{l_{\boldsymbol{\omega}_i}}$, was constant for all voxels, $\boldsymbol{\omega}_i$, for a particular image classification class label, $l_{\boldsymbol{\omega}_i}$ at voxel $\boldsymbol{\omega}_i$, (see equations 4.25 and 4.26). Even though Ruan *et al.* assumed a semi-homogeneous MRF (when considering edge regions), they included a further regularisation term to measure the amount of spatial variations in

intensity. This was primarily provided to model the within image class cluster variations rather than variation in the amount of required spatial regularisation that occurs along image class cluster borders (edge regions) - as with the work of Wang et al. The additional regularisation term was based on a fractal measure called the Hölder exponent. This fractal regularisation term provided additional distinct features for various regions in the image data, enabling the PV voxels to be differentiated from pure voxels even though they might share the same grey level value. This therefore allows for a form of inhomogeneous spatial regularisation.

While the inclusion of an additional regularisation parameter is interesting, particularly one based on a fractal measure, this additional complexity moves the statistical or probabilistic modelling of the image data further towards a heuristic formulation and therefore possibly further away from a framework that allows for wider understanding and generalisation improving applicability to new and previously un-seen problems. In addition to this, if a formulation is based on consistent rules based on easily verifiable assumptions it becomes easier to understand why performance is not optimal given further new and previously un-seen problems.

Automatic Determination of Partition Function, Z

An interesting paper by Woolrich *et al.* in 2005 [162] utilised a number of PV model formulations, one of which explicitly modelled the continuous nature of the PV effect, similar to the formulation used by Choi *et al.* in 1991, [24] (see equations 4.14 and 4.23). Woolrich *et al.*, however, proposed a novel way of analytically determining the partition function. The Gibbs distribution for the joint likelihood over the entire label space, $\boldsymbol{A}$ with the clique potential was given by Woolrich *et al.* as:

$$p(\boldsymbol{A}|\beta) = \frac{1}{Z(\beta)}.\exp\left(-\frac{\beta}{4}.\sum_{\forall \boldsymbol{\omega}_i}\sum_{\forall v}\sum_{\boldsymbol{\omega}_j \in \mathfrak{N}_{\boldsymbol{\omega}_i}} (\alpha_{\boldsymbol{\omega}_i,v} - \alpha_{\boldsymbol{\omega}_j,v})^2\right), \tag{4.27}$$

where β is again the spatial regularisation parameter and v is indicative of each classification class. For other models that utilize discrete classification classes, $l_{\boldsymbol{\omega}_i}$, it was not immediately obvious how Z could be determined analytically, but for equation 4.27 its form can be determined because:

$$Z(\beta) = \int_{\boldsymbol{A}} p(\boldsymbol{A}|\beta).\mathrm{d}\boldsymbol{A}, \tag{4.28}$$

which due to the Gaussian form of equation 4.27 has the convenient and simple form:

$$Z(\beta) \propto \frac{1}{\beta^{N.K}}, \tag{4.29}$$

where N is the number of voxels in the data space and K is the number of classification classes. Thus, a simple formulation is provided to determine the regularisation parameter, β and the partition function, $Z(\beta)$. The elegance in this solution could possibly prevent the reader from being reminded that, for this model, β is not dependent on the voxel location, $\boldsymbol{\omega}_i$.

4.2.3 Related Contextual Classification Work - Probabilistic Relaxation

Probabilistic relaxation has also been extensively used in the image analysis field to incorporate contextual information into a probabilistic formulation, see e.g. [115, 70, 68, 69, 54]. Probabilistic relaxation applied to images aims to improve the pixel or voxel labelling by optimising or improving a quality measure of the pixel or voxel labels.

Authors have found that probabilistic relaxation can be used to improve the labelling of the data, especially over the first few iterations. However after a number of iterations the quality of the labelling often degrades, see e.g. [133]. Poole in 1990, [103] noted this and proposed an alternative formulation that utilised a *"tailored sequence of updating functions"*. These updating functions were learnt via training on test images using a graph based tree pruning approach, so that after the optimal number of iterations the data labelling quality ceased to improve but would also not degrade.

The probabilistic relaxation algorithm discussed by Poole is different from many MRF approaches as the probabilistic relaxation only utilises the original data values at the start of the algorithm. The relaxation of these posterior probabilities is repeatedly applied without reference to the original data. This is quite different from an MRF approach which will often explicitly incorporate the original data into every iteration, see e.g. [24, 162]. Probabilistic relaxation and a MRF approach are conceptually very different. The framework proposed by Poole requires prior training to learn the geometric changes in the posterior probabilities after every iteration. Such prior knowledge may not be desirable for much medical data. However, if once such a formulation has been found and learnt, then this information can possibly be thought of as an alternative route to adaptive and hopefully optimal relaxation steps.

The approach proposed by Poole implicitly incorporates estimation and inference into the contextual classification schema known as probabilistic relaxation. This is also true for many of the MRF formulations. However estimation and inference often requires some sort of dedicated step, some of which are discussed below.

4.3 Estimation & Inference

The preceding discussions were primarily limited to the likelihoods that have been used to describe PV distributions. A likelihood or PDF is typically utilized within a conditional probability formula, usually referred to as Bayes theorem, see e.g. [77], resulting in a posterior density for the continuous form of Bayes or a probability for the discrete case (see equation 4.1).

The likelihoods utilize parameters such as means and standard deviations. These parameters are usually unknown for biomedical imaging data, where prior training on different data sets is usually inappropriate due to the innate variability of the data from person to person, scanner to scanner and even image to image.

Each data point or voxel in the image data will typically have an associated posterior density, $p(\boldsymbol{\alpha}|\boldsymbol{x})$ or probability, $P(\tau_a|\boldsymbol{x})$, associated with it. The product of these

densities or probabilities can then form a further density or probability:

$$p(\boldsymbol{A}|\boldsymbol{X}) = \prod_{\forall \boldsymbol{\alpha}, \boldsymbol{x}} p(\boldsymbol{\alpha}|\boldsymbol{x}); \tag{4.30}$$

or for the discrete case

$$P(\tau_a|\boldsymbol{X}) = \prod_{\forall \boldsymbol{x}} P(\tau_a|\boldsymbol{x}). \tag{4.31}$$

This single probability or density for the entire data can then be used in algorithms to determine the relative *correctness* of the current parameter estimates in relation to further estimated values. These newly estimated values may come from a deterministic methodology that is defined analytically or via a stochastic procedure (i.e. a random but possibly well informed guess). These parameter estimates are usually designed to maximize the posterior probabilities and may sometimes attempt to model correlations that might exist between each data point or voxel.

4.3.1 Deterministic Strategies

Deterministic parameter estimation and inference techniques applied to non-PV related statistical fields can sometimes utilize a convenient solution or even, in some instances, an analytical result. However, this is not usually the case for medical imaging data due, in part, to the added complexity that PV voxels contribute to image models.

Seminal Work

The seminal work on probabilistic modelling of the PV effect by Choi *et al.* in 1991, [24], also provided appropriate techniques for estimation and inference using the PV model likelihood previously described in equation 4.3. Choi *et al.* chose to utilize an algorithm that iteratively estimates the mode of the posterior distribution for each voxel, often known as the Iterated Conditional Modes (ICM) algorithm, described in algorithm 1.

Algorithm 1 (Iterated Conditional Modes Algorithm: Mixture Images)

1.Given an initial voxel labelling, $\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}$ $\forall \boldsymbol{\omega}_i$.
2.For each $\boldsymbol{\omega}_i$ *in the data space:*
  a.Find maximum of $p(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|\boldsymbol{\alpha}_{\Re_{\boldsymbol{\omega}_i}})$, *w.r.t.* $\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}$.
  b.Return to (a) if every $\boldsymbol{\omega}_i$ *not visited.*
3.Check for convergence, i.e. no changes in $\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}$ $\forall \boldsymbol{\omega}_i$.
4.If not converged return to (2).

The posterior distribution for each voxel, $p(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|g, \boldsymbol{\alpha}_{\Re_{\boldsymbol{\omega}_i}})$ was then maximised by reducing the problem to a quadratic optimization procedure. This was made possible by a number of simplifying steps. Firstly, noting that

$$p(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|g, \boldsymbol{\alpha}_{\Re_{\boldsymbol{\omega}_i}}) \propto p(g|\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}).p(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|\boldsymbol{\alpha}_{\Re_{\boldsymbol{\omega}_i}}) \tag{4.32}$$

. Secondly, it is well known that a monotonically increasing function that is to be maximised can also be maximised indirectly by minimising its (negative) natural logarithm. So that taking the negative logarithm of the r.h.s. (recalling the definitions given for equations 4.3 and 4.14):

$$-\ln\left[p(g|\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}).p(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|\boldsymbol{\alpha}_{\mathfrak{N}_{\boldsymbol{\omega}_i}})\right] =$$

$$\frac{1}{2}\left(\mathbf{x}_{\boldsymbol{\omega}_i}-\mathbf{M}^T.\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}\right)^T\boldsymbol{\sigma}_N^{-1}\left(\mathbf{x}_{\boldsymbol{\omega}_i}-\mathbf{M}^T.\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}\right)+\sum_{\boldsymbol{\omega}_j\in\mathfrak{N}_{\boldsymbol{\omega}_i}}\beta.\kappa_{\boldsymbol{\omega}_j}||\boldsymbol{\alpha}_{\boldsymbol{\omega}_j}-\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}||^2. \quad (4.33)$$

This can then be minimised via quadratic minimisation procedures, see e.g. [104]. Unfortunately, this limits the number of possible mixtures to one more than there are image channels, i.e. multivariate data is required for more than two possible mixture constituents. Choi *et al.* do however go on to describe a method that takes advantage of the fact that neurological MR data may usually only be expected to have PV voxels composed of a maximum of two tissues. The technique identifies the most likely tissues based on their posterior probabilities using mixture values from the preceding estimation step. These most likely tissues can then be used in the minimisation procedure thereby limiting the classification component count.

The noise covariance matrix is estimated interactively by a user during the initialisation process, where the mean matrix is initialised also interactively by a user. The mean matrix is then updated via a further optimization process, but the covariance matrix is not. The assumption of negligible inter-tissue variation in the covariance together with only allowing for it to be initialised and not updated could reduce the accuracy of the model for estimation of PV mixtures in imaging data.

The mean matrix is estimated by calculating the pure component means based on voxels that have been labelled as almost pure voxels. This may also present difficulties due to the possibility of the lack of pure voxels for some data sets. Also, the thresholds are set in order to include approximately equal numbers from each of the classification components which may also be sub-optimal for some images.

The model of Choi *et al.* also requires the estimation of the spatial regularisation parameter, β. This regularisation parameter was estimated by maximising the logarithm of the conditional posterior distribution conditioned on β. Choi *et al.* made the important observation that the Gibbs distribution in combination with the clique potential defined in equation 4.14 results in a Gaussian distribution. This observation was used in combination with simulations with various numbers of mixel constituents to determine optimal values of β under these different conditions. Unfortunately, this approach does not allow for variation of β within the image data. This is not optimal for images that are composed of more than a single classification component or perhaps even regions of mixtures that occur in neurological MR images of the brain, (previously discussed in chapter 2).

The Expectation-Maximisation Algorithm

The EM algorithm is a well-known iterative technique used to estimate *missing* data by finding the modes of the posterior distribution. The algorithm was popularised

by Dempster *et al.* in 1977, [33]. It has been applied quite extensively to medical image analysis problems where the pixel or voxel labels form the missing data, see e.g. [13, 78, 81, 121, 125, 144, 154, 163].

The EM algorithm consists of iterating over two main steps: the expectation step; and the maximisation step. The expectation step refers to the calculation of the expectation of the *complete* likelihood, i.e. the likelihood using the previously estimated voxel labels and parameter values. The maximisation step refers to the maximisation of the likelihood given the data which results in new estimates that can be used in the expectation step again, until convergence. Algorithm 2 summarises the EM steps.

Algorithm 2 (EM Algorithm)

1.Given an initial voxel labelling, $\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}$ $\forall \boldsymbol{\omega}_i$ and other conventional parameter values.
2.Calculate the conditional expectation of the complete data log-likelihood.
3.Calculate new parameter values by maximising the complete data log-likelihood.
4.Check for convergence
5.If not converged return to (2).

Quite often optimal parameter or voxel label estimates are not available analytically and a combination of analytical maximisation parameter estimation calculations are combined with stochastic estimation techniques, see e.g. [78]. Alternatively [163] used the ICM algorithm to find optimal voxel labels and the EM algorithm to calculate the optimal parameter estimates.

Some techniques appear to derive the EM algorithm so that it is consistent with the assumptions of the original paper by Dempster et al in 1977 [33], see e.g. [78]. The PV problem requires careful consideration of the assumptions used in the formulation of the EM algorithm in order to provide a theoretical explanation of the convergence properties of the algorithm. Other authors have applied the EM algorithm to data containing PV voxels, see e.g. [95]. The EM algorithm is a popular iterative technique and prior to 1977 there were many iterative EM like algorithms without theoretical justification. Dempster *et al.* provided a theoretical framework to which many of the previously existing EM like algorithms could find explanation, see e.g. the peer-group discussion at the end of [33].

4.3.2 Stochastic Techniques

A stochastic estimation and inference technique can be thought of as one that relies on random guesses to estimate a parameter or to simulate a distribution of interest. In the general literature such techniques include general Monte-Carlo simulation [3, 78], Gibbs sampling [42, 116, 162] and full Markov Chain Monte Carlo (MCMC) simulations [162, 90, 113].

In simple terms a general Monte Carlo simulation utilizes random samples drawn from a particular distribution. These random samples are often used to evaluate an integral

numerically. Various techniques exist to reduce the number of samples required while still achieving an accurate result. Gonzalez Ballester *et al.* in 2002, [3], however only utilized direct sampling of the posterior distribution of a collection of PV affected voxels to simulate the collective posterior distribution of those voxels combined. This proved to be computationally effective as otherwise the tissue volume (and confidence bounds) of each voxel would have to have been estimated individually thereby reducing the computational burden.

Other techniques utilizing similar concepts as the simple Monte Carlo simulation technique include Gibbs sampling which is a particular type of a MCMC simulation technique. Gibbs sampling was originally applied to image data by Geman and Geman in 1984, [42], although they did not include any modelling of the PV effect. The essential idea with Gibbs sampling is that the posterior density of an entire image (or any other object consisting of multiple parameters such as voxel labels) can sometimes be divided into per-voxel conditional densities. These individual conditional densities can then be simulated individually rather than attempting to simulate samples for the entire image (or system) with a huge number of free parameters such as the voxel labels. These conditional densities are often parameterized in a form that can be easily sampled, see e.g. [116, 162].

Sometimes, however, particularly with Bayesian problems, the posterior density may not be available analytically. If this is the case, as is often found with the continuous form of Bayes, then a full MCMC simulation may be used, [113, 80, 162], such as the Metropolis-Hastings algorithm. For example, the posterior density is calculated via Bayes theorem, i.e.:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}).p(\mathbf{x})}{p(\mathbf{y})} \tag{4.34}$$

where $\mathbf{y}$ might represent the data such as the voxel intensities and $\mathbf{x}$ might represent a parameter(s) of interest. The denominator, often referred to as the marginal density, is calculated via a marginalisation integral, i.e.

$$p(\mathbf{y}) = \int_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}).p(\mathbf{x}).d\mathbf{x}. \tag{4.35}$$

In many situations the result of this integral may not be available, leaving the posterior density only partially available:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}).p(\mathbf{x}). \tag{4.36}$$

The Metropolis-Hastings algorithm utilizes the ratio of the posterior distributions, so that the denominator is no longer required. This therefore means that the posterior density can be easily simulated despite the marginal density being unknown.

The result of these stochastic simulation type strategies is often a series of samples of the posterior of interest, such as the posterior density of parameter that is being estimated. When the simulation has converged (Woolrich *et al.*, [162] run the simulation for a fixed number of samples) the parameter can be estimated from the Maximum *a posteriori* (MAP) value, the expected value or some other representative point or interval estimate, see e.g. [77].

4.4 In Conclusion

This chapter has discussed many of the existing approaches to probabilistic modelling of the PV effect. The two seminal approaches included modelling of the PV effect using a continuous mixture random variable, [24] and a technique that categorises the PV voxels as additional classification classes, [122, 123]. Many of the techniques derived from these seminal works have improved on the formulations, accuracy and introduced additional, sometimes *ad hoc* image measures in an attempt to improve classifier performance and or reduce computational complexity. Many of the techniques utilize spatial contextual information by modelling the neighbourhood information as a discrete MRF. This however is not really appropriate for a model of the PV effect, especially when using a continuous mixture random variable. Many of the techniques also assume homogeneous spatial regularisation, this is also not appropriate due to the edge information and other inhomogeneous voxel labelling usually associated with imaging data. Many of the techniques assume a uniform PSF and or PV prior mixing distributions. These assumptions are probably not representative of the true physical image model.

The importance of accurate modelling of the PV effect was discussed in the earlier chapters 1 and 2. It would therefore appear that improved modelling of the PV effect, resulting in improved PV classifier performance and a greater understanding of the underlying PV physical process would be useful. This information together with the apparent possibilities to improvement of the existing PV models that have been discussed here provide ample opportunity for further PV model investigations. This is the topic of the remainder of this thesis.

Chapter 5

Intensity Based Finite PV Mixture Models

This chapter compares a number of likelihood models that have been used to model data that have been affected by the PV effect. The models are first described and then their respective performances are assessed with the use of simulated data. The models are proposed as different refinements of modelling the PV mixing process. This provides a novel presentation of the inter-relationship of relevant PV models. In particular the final model described here is found to be approximately equivalent to an alternative PV model that utilizes analogy of the probabilities as mixture constituents of the voxels. This alternative model is found to be approximately equivalent under particular conditions relating to the data and model configuration as will be seen shortly.

Finally, it is intended that this chapter will provide a systematic description of intensity PV models that underpin the developments in PV modelling described in subsequent chapters 6, 7 and 8.

5.1 Methodology

Biomedical volumetric imaging data is usually considered to be composed of a number of classification classes, which are referred to in this thesis by the symbol τ_v, where the subscript v is indicative of the particular classification class. Classification classes most often correspond to a particular tissue or object type. In probability theory, the symbol, τ_v, represents a classification class specific event.

Classification classes for neurological MR data may include, for example, GM or WM. For oncological PET data the classification classes may include statistical descriptions of areas of high physiological activity in contrast to areas of low activity, such as a classification class corresponding to tumour metabolic activity that is surrounded by a second classification class indicative of normal biological metabolic activity. For a more in depth discussion of the PV effect and the origin of the PV effect see chapter 2, sections 2.1 and 2.2.

5.1.1 Model A: Finite Gaussian Mixture Model

For a statistical model to describe PET or MR image data, it should consist of a number of classification classes. Mixture models provide a convenient framework that can accommodate statistical descriptions of a number of classification classes. A classification class can be described by a PDF, $p_G(g|\tau_v)$ and a prior probability $P(\tau_v)$. The PDF provides a method to calculate the probability of obtaining a particular grey level, g given the classification class τ_v being true. The prior probability, $P(\tau_v)$ provides a probability associated with classification class τ_v being true, independent of any other variables such as the grey level.

Thus, the mixture model can be formulated via a weighted summation of the PDFs, where the weights are the prior probabilities:

$$p(g) = \sum_{\forall v} p_G(g|\tau_v).P(\tau_v), \tag{5.1}$$

where $\sum_{\forall v} P(\tau_v) = 1$ and g is a grey level associated with the imaging data. $p(g)$ is known as a marginal density because the summation over the classification class events, τ_v produces $p(g)$ which expresses the PDF of g irrespective of the classification class events, hence the class events are marginalised.

Notice how the prior probabilities, $P(\tau_v)$ act as scaling parameters to the individual classification class PDFs, $p_G(g|\tau_v)$. This can be related to imaging data where the prior probabilities describe the relative amounts of the different classification classes present in the imaging data. The PDFs describe the variation of the intensities associated with an individual classification class due to effects such as thermal noise in the image acquisition device and or natural variations in a particular class. An example of natural variations in tissue classification classes in anatomical MR imaging data may include small blood vessels or variations in myelination density of WM in the human brain. For functional PET data, natural variations in tracer uptake are often associated with variation in physiological activity.

$p_G(g|\tau_v)$ is often known as a likelihood model, and for a Finite Gaussian Mixture Model (FGMM) it is governed by a Gaussian PDF:

$$p_G(g|\tau_v) = \frac{1}{\sqrt{2.\pi.\sigma_v^2}}.\exp\left(-\frac{(g-\mu_v)^2}{2.\sigma_v^2}\right), \tag{5.2}$$

where μ_v and σ_v are the mean and standard deviation for classification class τ_v respectively. An exemplar FGMM is illustrated in figure 5.1.

Bayes formula can be used to calculate the probability that an individual voxel, $\boldsymbol{\omega}$ with grey level, $g_{\boldsymbol{\omega}}$ belongs to a particular classification class, τ_v:

$$P(\tau_v|g_{\boldsymbol{\omega}}) = \frac{p(g_{\boldsymbol{\omega}}|\tau_v).P(\tau_v)}{p(g_{\boldsymbol{\omega}})}, \tag{5.3}$$

where $p(g_{\boldsymbol{\omega}})$ is the marginal density given by equation 5.1.

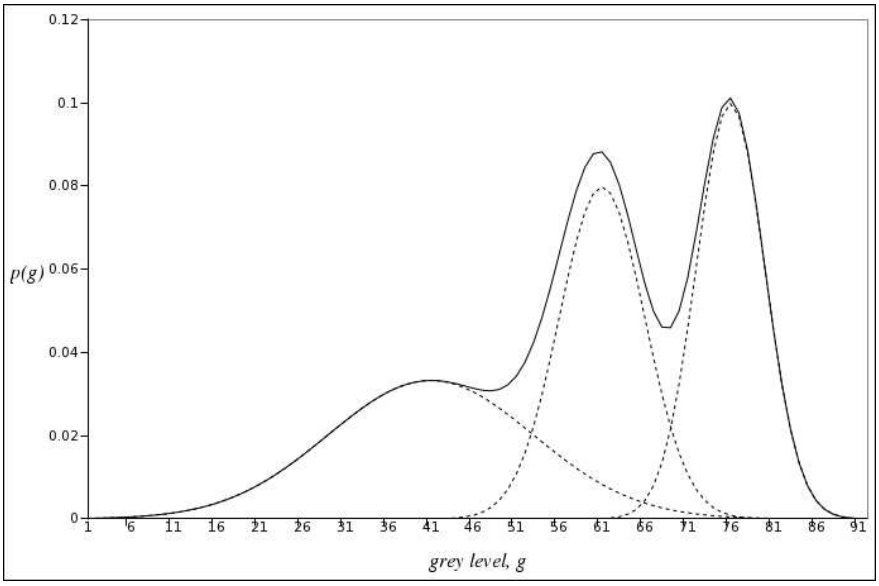


Figure 5.1: Example of a Finite Gaussian Mixture Model (FGMM) with 3 Gaussian likelihood classification classes.

Often in the decision theoretic field a decision as to whether a pixel belongs to one of the discrete classification classes has to be made and is often based on the classification class with the highest posterior probability [34, 151]. This approach is not taken here as this model is concerned with PV voxels where a single volume element may contain signals from more than a single classification class. Therefore, for the finite Gaussian mixture model, an analogy is taken with the probability for a particular classification class as being representative of the amount of that particular classification class present in a voxel. The use of the probabilities in this way is consistent with empirical probability, e.g. Collins Dictionary of Mathematics, [11], provide the following definition for empirical probability:

> *"...the proportion observed in a sample..."*.

Thus, as a simple example, given a two classification class mixture model, with classification classes τ_a and τ_b and a voxel with grey level, $g_{\boldsymbol{\omega}}$ where $P(\tau_a|g_{\boldsymbol{\omega}}) = 0.5$ and $P(\tau_b|g_{\boldsymbol{\omega}}) = 0.5$ then the voxel, $\boldsymbol{\omega}$, will be classified as being composed of 50% of each classification class.

Assumption of Gaussian Likelihood for MR and PET Imaging Data

Finite Gaussian mixture models have previously been used to describe PET and MR imaging data, e.g. [81, 121, 125]. The following discussion helps to justify the use of Gaussian PDFs.

Gudbjartsson and Patz in 1995, [48] considered the theoretical density of magnitude MR data (most MR data is calculated from the magnitude of the two Fourier components

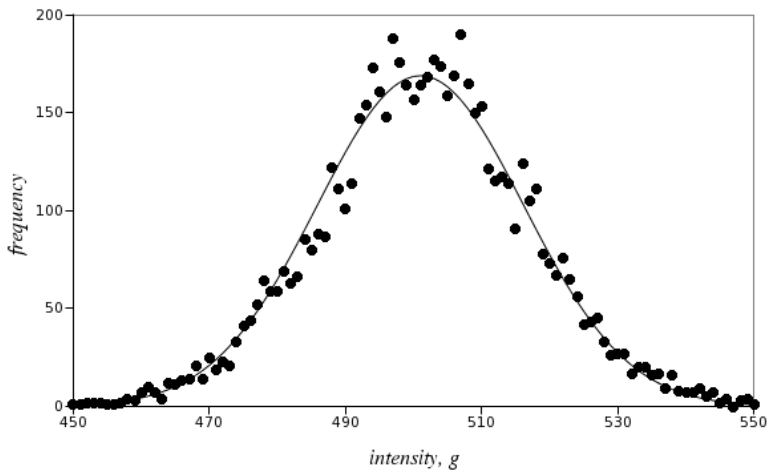


Figure 5.2: Histogram of fully corrected RAMLA reconstruction PET data (•) with Gaussian model (line) superimposed. This figure illustrates the empirical relevance of a Gaussian model choice to describe PET data statistically. The PET frequency data is from a PET phantom, described in chapter 8.

as the reconstruction process is based on the Fourier transform) and found it to be Rician distributed. A Rician density tends towards a Gaussian density at high SNRs. Most MR imaging data of human anatomy possess SNRs high enough so that they can be described as a Gaussian density. This fact therefore validates, in part, the use of Gaussian PDFs to describe the classification classes found in MR imaging data.

PET projection data are intrinsically Poisson distributed. However once the data has been subjected to a reconstruction process and various other processes, the Poisson distribution is usually no longer a valid choice for a statistical description of the data. Gaussians are often applied empirically when there is no theoretical basis for a particular statistical model selection. Also, a Poisson distribution is well approximated by a Gaussian distribution for large enough mean values details of which can be seen in chapter 2, section 2.3. A histogram of PET data can be observed to be approximately Gaussian in shape, see for example figure 5.2.

For further information about the MR and PET image formation processes see chapter 2, sections 2.1, 2.2 and 2.3.

5.1.2 Model B: Finite Gaussian Mixture Model with PV Distributions

A well known problem with using a finite Gaussian mixture models for the description of data that has voxels affected by the PV effect (PV voxels) is that it does not take into account the PV effect on the resulting mixture model. The PV effect produces voxels that consist of signals originating from more than a single classification class. These mixture or PV voxels can be considered to come from an additional classification class in a Finite Mixture Model (FMM). An example from such a PV FMM is illustrated in

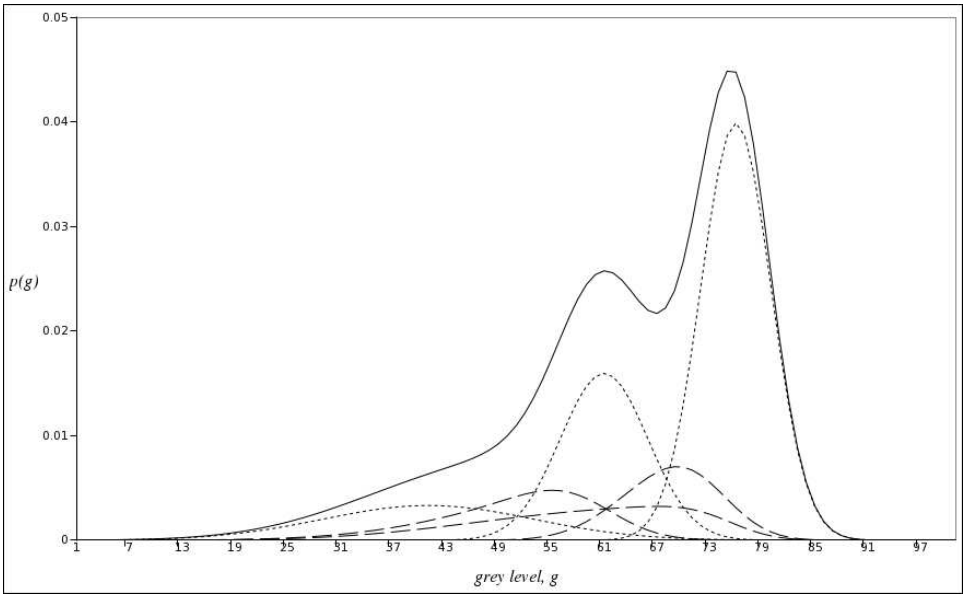


Figure 5.3: Example of a PV finite mixture model with 3 pure and 3 PV classification classes. With reference to equation 5.4: dotted lines indicate pure tissue distributions, $p(g|\tau_v).P(\tau_v)$; dashed lines indicate PV tissue distributions, $p(g|\tau_{v,a}).P(\tau_{v,a})$; solid line represents distribution for the entire data, $p(g)$.

figure 5.3 and this can be compared with the PV data densities for real clinical imaging data presented and discussed in more detail in chapter 2, sections 2.1 and 2.2.

The PV finite mixture model is composed of pure and PV components where each PV component is modelled from a mixture of two pure components. The possibility of three components in a PV voxel is considered insignificant as most PV voxels will be limited to mixtures of two components. Also the worst case volumetric error for the case of three components misrepresented with a two component model is $33.\dot{3}\%$. Furthermore, despite the PV models explicitly modelling only two component mixtures, the PV models described shortly do not preclude the posterior probabilities taking values representative of mixtures for more than two components. Thus the PV finite mixture model is given by the following formulation:

$$p(g) = \sum_{v=1}^{K} p(g|\tau_v).P(\tau_v) + \sum_{v=1}^{K-1} \sum_{a=v+1}^{K} p(g|\tau_{v,a}).P(\tau_{v,a}), \tag{5.4}$$

where $P(\tau_{v,a})$ is the PV prior probability of classification classes τ_v and τ_a. The PV distributions for this model, $p(g|\tau_{v,a})$ are often modelled as the convolution of a Gaussian with a uniform prior distribution, (see e.g. [128]), resulting in functions of the form that can be seen in figure 5.3. The PV events, $\tau_{v,a}$ and $\tau_{a,v}$ are equivalent, i.e. $\tau_{v,a} \equiv \tau_{a,v}$, and the sum of all the unique events sum to 1:

$$\sum_{v=1}^{K} P(\tau_v) + \sum_{v=1}^{K-1} \sum_{a=v+1}^{K} P(\tau_{v,a}) = 1, \tag{5.5}$$

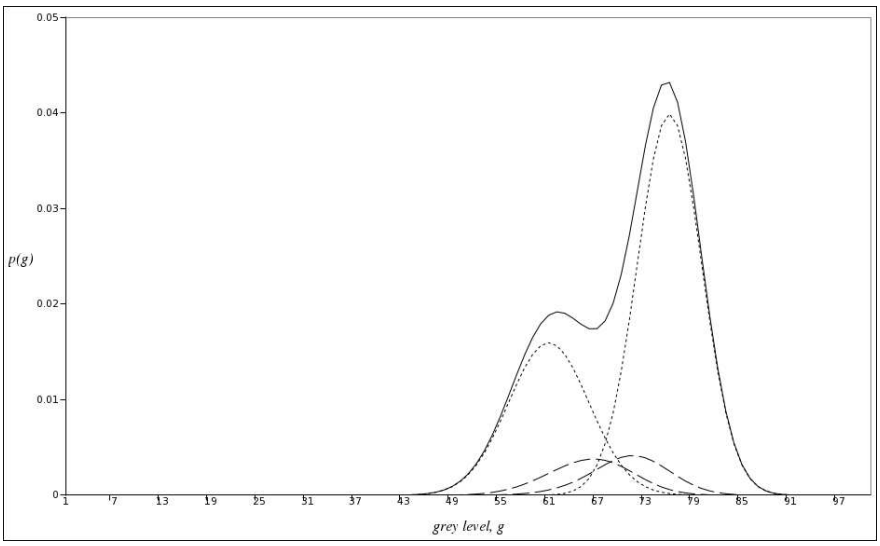


Figure 5.4: Example of a PV finite mixture model with 2 pure and 2 PV classification classes. The same parameters have been used as were used for the mixture model in figure 5.3, but one of the pure classification classes has been removed so as not to complicate the illustration. Note that only a single PV intensity density between the two pure classes was used in figure 5.3. Now each PV class has been split into two PV classification classes so as to allow the posterior probability to be calculated for each classification class instead of an additional, somewhat artificial PV density.

where K is the number of classification classes. This formulation allows the posterior probability via Bayes to be calculated for individual pure voxel classification classes, τ_v, *or* PV voxel classification classes, $\tau_{v,a}$. Variants of this type of PV model can be seen in [122, 123], which were previously discussed in chapter 4.

An improvement to equation 5.4 is possible that allows for the posterior probability to be calculated for individual pure *and* PV voxel classification classes combined. This improvement was proposed by Vokurka *et al.* in 2002, [148] and consists of separating the PV PDFs or PDF components into two separate components, corresponding to the two pure components from whence they came. This is illustrated in figure 5.4. The PV PDF is now separated into two classification classes. In this case these mixture classes were modelled as triangular PDFs, $p_T(g|\tau_{v,a}^{GT})$ convolved with the respective pure classification class Gaussian PDF, $p_G(g|\tau_v)$. i.e. The convolution of a triangle or ramp with a Gaussian from the pure classification (tissue or activity) class with characteristic width σ_v. This results in $p_{GT}(g|\tau_{v,a}^{GT})$, where GT is indicative of the Gaussian-Triangle convolution. This convolution is discussed shortly.

The mixture model for such a PV finite mixture model is given by:

$$p(g) = \sum_{v=1}^{K} p(g|\tau_v).P(\tau_v) + \sum_{v=1}^{K-1} \sum_{a=v+1}^{K} \left\{ p_{GT}(g|\tau_{v,a}^{GT}).P_{GT}(\tau_{v,a}^{GT}) + p_{GT}(g|\tau_{a,v}^{GT}).P_{GT}(\tau_{a,v}^{GT}) \right\}, \tag{5.6}$$

where $\tau_{v,a}^{GT}$ is not equivalent to $\tau_{a,v}^{GT}$. i.e. These PV events are seen as distinct, $\tau_{v,a}^{GT} \ncong$

$\tau_{a,v}^{GT}$, unlike the PV events for equation 5.4, $\tau_{v,a}^{GT} \equiv \tau_{a,v}^{GT}$. It should also be noted that,

$$\sum_{v=1}^{K} P(\tau_v) + \sum_{v=1}^{K-1} \sum_{a=v+1}^{K} (P_{GT}(\tau_{v,a}^{GT}) + P_{GT}(\tau_{a,v}^{GT})) = 1. \tag{5.7}$$

This results in the posterior probability using Bayes theorem for an individual classification class, τ_v to be given by:

$$p(\tau_v|g) = \frac{p(g|\tau_v).P(\tau_v) + \sum\limits_{\forall a \neq v} p_{GT}(g|\tau_{v,a}^{GT}).P_{GT}(\tau_{v,a}^{GT})}{p(g)}. \tag{5.8}$$

The composition of a particular voxel is then given by analogy of the computed posterior probability, as previously described for the finite Gaussian mixture model with equation 5.3.

Gaussian-Triangle Convolution Density

As has already been discussed and originally suggested by Vokurka *et al.* in 2002 [148], the PDF for a PV classification class, $p_{GT}(g|\tau_{v,a}^{GT})$ in equation 5.8 is given by the convolution of a triangular PDF, $p_T(g|\tau_{v,a})$ with the respective pure Gaussian density $p(g|\tau_v)$. This section discusses the convolution, its evaluation and its form. For a complete evaluation of the convolution see appendix B.

The equation for the triangle can be formulated in terms of the formula for a straight line, $y = M.x + C$:

$$p_T(g|\tau_{v,a}) = \begin{cases} \frac{1}{Q}(M.g + C) & \text{for} \quad \mu_v \leq g \leq \mu_a \\ 0 & \text{for} \quad \text{elsewhere} \end{cases}, \tag{5.9}$$

where μ_v and μ_a are the mean values for classification classes τ_v and τ_a respectively, from which the PV density belongs. Q is a normalising parameter, M is the slope and C is the intercept of the straight line. Q is easily calculated from the integral of this function (see appendix B) and $M = -1$ or $+1$ and $C = \mu_a$ or $-\mu_v$ depending on whether $\tau_{v,a}$ is being considered or $\tau_{a,v}$.

Now consider the pure classification class PDF:

$$p_G(g|\tau_v) = \frac{1}{\sqrt{2.\pi.\sigma_v^2}}.\exp\left(-\frac{g^2}{2.\sigma_v^2}\right), \tag{5.10}$$

where μ_v has been dropped because this PDF is now convolved with $p_T(g|\tau_{v,a})$ which is defined between the two means of the two classification classes, τ_v and τ_a:

$$\begin{aligned} p_{GT}(g|\tau_{v,a}^{GT}) &= p_G(g|\tau_v) * p_T(g|\tau_{v,a}^{GT}) \\ &= \int_{\mu_a}^{\mu_v} p_G(g-\gamma|\tau_v).p_T(\gamma|\tau_{v,a}^{GT}).d\gamma, \end{aligned} \tag{5.11}$$

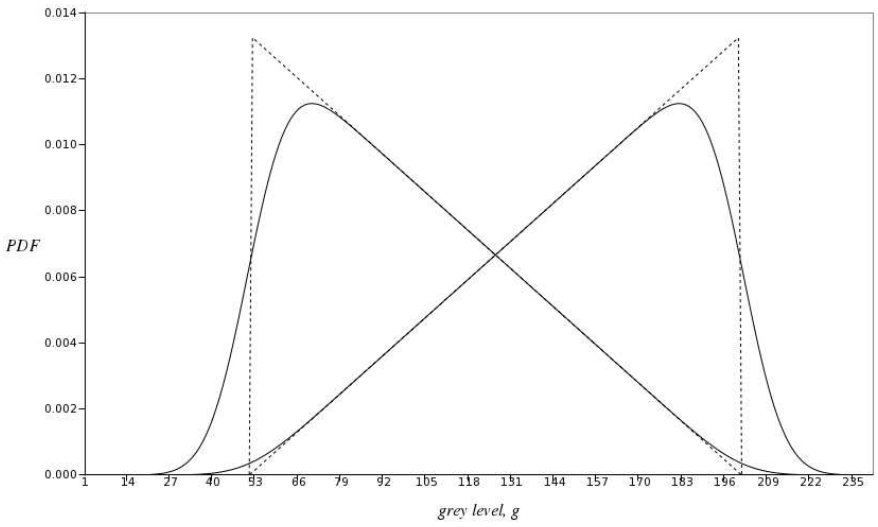


Figure 5.5: Illustration of exemplar Gaussian-Triangle PV density plots given by $p_{GT}(g|\tau_{v,a}^{GT})$ and $p_{GT}(g|\tau_{a,v}^{GT})$, (see equation 5.12). Dotted lines are the original triangular densities, $p_T(g|\tau_{v,a})$, $p_T(g|\tau_{a,v})$ and the solid lines correspond to $p_{GT}(g|\tau_{v,a}^{GT})$, $p_{GT}(g|\tau_{a,v}^{GT})$. The parameters used to generate these example densities were $\mu_v = 50$, $\mu_a = 200$ and $\sigma_v, \sigma_a = 4$.

where γ is the variable of the convolution integral. For more information about convolution see for example [134]. The result of this integral is given by:

$$p_{GT}(g|\tau_{v,a}^{GT}) = \frac{(M.g + C)}{-2.Q}\left\{\text{erf}\left(\frac{g-\mu_v}{\sqrt{2.\sigma_v^2}}\right) - \text{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_v^2}}\right)\right\} + \frac{-M.\sigma_v}{Q.\sqrt{2.\pi}}\left\{\exp\left(\frac{(g-\mu_v)^2}{2.\sigma_v^2}\right) - \exp\left(\frac{(g-\mu_a)^2}{2.\sigma_v^2}\right)\right\}. \quad (5.12)$$

Figure 5.5 illustrates a pair of example Gaussian-triangle PV densities from equation 5.12.

5.1.3 Model C: Continuous PV Mixture Model

The posterior probabilities of models A and B are representative of the composition of individual voxels, utilising discrete classification classes. The use of Bayes formula in this way is very convenient and computationally efficient but it also obscures some of the more subtle aspects of the PV problem. Chapters 1 and 2 illustrated that the effect of the action of the PSF was to merge the independent classification classes into a single entity. An explicit model of PV composition of individual voxels is therefore proposed and in this way a more mathematically consistent model of the PV effect can be realised. This is the approach taken for model C, described shortly. It is then compared to models A and B.

Model C incorporates the voxel content explicitly as a random vector of continuous random mixture variables (in the range of 0 to 1-where 1 is total pure voxel content and 0 is zero voxel content). Each element in this mixture vector represent the content of the voxels as random variables. This mixture vector can then be used explicitly in Bayes theorem as variables of conditional densities, together with other variables such as image intensity. Thus, to clarify the change in reasoning, Models A and B described in this thesis utilised discrete PV voxel label assignments, but model C utilises continuous voxel label assignments. The use of continuous voxel labels to model the PV effect is akin to the approach taken by, for example [24, 162]. These PV models often incorporate a spatial smoothness constraint (see chapter 7 for more information on spatial smoothness). But they do not necessarily model the global likelihood of the possible mixture configurations which is in contrast to the approach taken in this chapter, similar to that also taken in [162, 3, 78]. Classification with a global mixture likelihood can be used as a pre-cursor to the application of a model that incorporates a smoothness constraint which will be addressed in chapter 7.

An important distinction between the work in this chapter, described by model C and the work in the existing literature that emphasises the use of a global mixture likelihood prior is the method of inference. Many techniques infer utilising the mode of the posterior, i.e. the most probable mixture configuration of a voxel. This includes techniques that rely on the EM algorithm, [78, 95] or Monte Carlo techniques that have been designed to find the mode [3]. Model C of this section utilises the posterior mean, i.e. the expected mixture configuration given the data and the model, which has also been used in [162] via a fully Bayesian technique utilising Gibbs sampling. As noted in texts on Bayesian techniques such as [77], the expected posterior value is equivalent to a point estimation with a minimum squared error loss. This is the quantity that many authors attempt to minimise as part of performance characterisation of their techniques, see e.g. [74, 128] and the work in this thesis.

Related work, such as might be found in the remote sensing literature, have used the posterior mean, for example Kitamoto and Takagi 2000, [66]. There are subtle differences between the work here and in [66], notably where Kitamoto and Takagi make *hard decisions* about whether their satellite imagery pixels are either pure or mixed. If a decision is made about a particular pixel being mixed, then the pixel's constituents are then estimated via the posterior mean. This strategy appears to be rather inconsistent, combining optimal mixture estimates from posterior modes *and* posterior means. This is in contrast to model C (and all subsequent models in this thesis) where the optimal posterior point estimate is given by the posterior mean.

The importance of the posterior mean for model B is also highlighted through an analytical comparison of model B with the model of this section, model C.

Voxel Labels

For this model each voxel, $\boldsymbol{\omega}$, is composed of signals from a number of classification classes, $\alpha_{\boldsymbol{\omega},v} \in [0,1]$ where v is indicative of the classification class, as for τ_v. If then, for example, $\alpha_{\boldsymbol{\omega},v} = 1$ then that voxel is considered to be composed of only classification

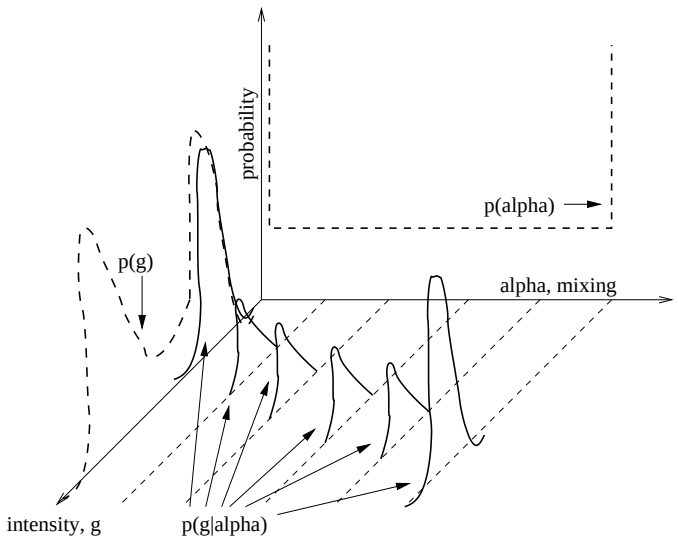


Figure 5.6: Sketch of the model space described by equation 5.15 for two pure components.

class τ_v. If there are K classification classes then:

$$\sum_{v=1}^{K} \alpha_{\boldsymbol{\omega},v} = 1 \ \forall \ \boldsymbol{\omega}. \tag{5.13}$$

A PV vector for each voxel is then composed of these individual PV components:

$$\boldsymbol{\alpha}_{\boldsymbol{\omega}} = (\alpha_{\boldsymbol{\omega},1} \ \alpha_{\boldsymbol{\omega},2} \ ... \ \alpha_{\boldsymbol{\omega},K})^T. \tag{5.14}$$

The probability density of obtaining a particular mixture vector when a particular intensity measurement occurs is given by $p(\boldsymbol{\alpha}_{\boldsymbol{\omega}}|g_{\boldsymbol{\omega}})$. This probability density is known as the posterior probability density of the mixture vector and can be determined using Bayes theorem:

$$p(\boldsymbol{\alpha}_{\boldsymbol{\omega}}|g_{\boldsymbol{\omega}}) = \frac{p(g_{\boldsymbol{\omega}}|\boldsymbol{\alpha}_{\boldsymbol{\omega}}).p(\boldsymbol{\alpha}_{\boldsymbol{\omega}})}{p(g_{\boldsymbol{\omega}})}, \tag{5.15}$$

where $p(\boldsymbol{\alpha}_{\boldsymbol{\omega}})$ is the prior probability density for the mixture vector; $p(g_{\boldsymbol{\omega}}|\boldsymbol{\alpha}_{\boldsymbol{\omega}})$ is the grey level likelihood or PDF given a particular mixture vector and $p(g_{\boldsymbol{\omega}})$ is known as the marginal grey level density. Each of these terms are sketched in figure 5.6 and will now be discussed.

Grey Level Likelihood

The grey level likelihood that was central to the previous models, and is also here, can be considered. But now it is dependent on the mixture vector, $\boldsymbol{\alpha}$, (dropping the point specific notation):

$$p(g|\boldsymbol{\alpha}) = \frac{1}{\sqrt{2.\pi.\sigma_{\alpha}^2}}.\exp\left(-\frac{(g-\mu_{\alpha})^2}{2.\sigma_{\alpha}^2}\right), \tag{5.16}$$

where $\mu_\alpha = \boldsymbol{\alpha}^T.\boldsymbol{\mu}$ is the mixture mean, $\sigma_\alpha^2 = \sum_{v=1}^{K}(\alpha_v.\sigma_v)^2$ is the mixture variance calculated by taking the product of each class mixture variable (squared) with the respective class variance, and $\boldsymbol{\mu} = (\mu_1\ \mu_2\ ...\ \mu_K)^T$ is a vector of the classification class means. Equation 5.16 and the calculation of the mixture parameters effectively assumes linear uncorrelated mixing of the individual classification components. This signifies that each classification class signal is uncorrelated and Gaussian distributed, as has been used in many other models of the PV effect, e.g. [24, 74, 158, 64, 162]. There are some potentially open questions about the validity of assuming linear mixing, especially for MR imaging data for particular tissue type combinations. This was discussed in the review of existing work, chapter 4 and is also examined in more detail in chapter 2. The mixture parameters, μ_α and σ_α^2 can be calculated in this way due to the result of the sum of K Gaussian distributed random variables as demonstrated in appendix C. The assumption of Gaussianity has already been discussed in section 5.1.1.

The Mixture Vector Prior Probability

The prior density, or global mixture likelihood, $p(\boldsymbol{\alpha})$, takes a similar form to the mixture combinations utilised in models A and B although the prior densities were not explicitly defined for these previous models. The prior density limits the PV model to possibilities of PV mixtures containing a maximum of two classification classes (as for model B), thus ignoring the very small possibility of a voxel containing more than a single mixing component:

$$p(\boldsymbol{\alpha}) = \sum_{v=1}^{K} p(\alpha_v).P(\tau_v) + \sum_{v=1}^{K-1}\sum_{a=v+1}^{K} p(\alpha_v, \alpha_a).P(\tau_{v,a}). \tag{5.17}$$

This simplifies for the case of $K = 2$, i.e. two classification classes, so that:

$$p(\boldsymbol{\alpha}) = p(\alpha_1).P(\tau_1) + p(\alpha_2).P(\tau_2) + p(\alpha_1, \alpha_2).P(\tau_{1,2}). \tag{5.18}$$

The continuous prior probability density for each pure classification class is given by,

$$p(\alpha_v) = \delta(1 - \alpha_v).f(\alpha_v). \tag{5.19}$$

The Dirac delta function, $\delta()$ is used to limit the pure terms in equation 5.17 to particular mixture vector values that are representative of a single component being present, i.e. when $\alpha_v = 1.0$. This is due to the sifting property of the Dirac delta function, $\delta()$, as will be seen shortly. $f(\alpha_v)$ is a prior mixing density discussed below.

The PV prior probability densities are given by,

$$p(\alpha_v, \alpha_a) = U(\alpha_v, \alpha_a).f(\alpha_v, \alpha_a) \tag{5.20}$$

where the indicator function, $U(\alpha_v, \alpha_a) = 1$ iff $0 < \alpha_v, \alpha_a < 1$, otherwise $U(\alpha_v, \alpha_a) = 0$, thus limiting each PV term to be valid when only two mixture components are present (i.e. non-zero) and all other mixture components are zero in value.

For this particular model, a very simple approach is adopted to address the mixing prior functions. These are, in this case, assumed to be given by uniform densities, $f(\alpha_v) = 1$

for $\alpha_v \in [0,1]$ and $f(\alpha_v, \alpha_a) = 1$ for $\alpha_v, \alpha_a \in (0,1)$. Uniform prior mixing densities have been used in most models of the PV effect (either implied or explicitly assumed), see for example [3, 117, 122]. A uniform density implies that all PV combinations are considered to be equally likely, but this simplifying assumption is not strictly true as will be seen shortly. The topic of PV mixture densities is investigated in chapter 7 where more informative forms of the two classification class mixing prior functions, $f(\alpha_v, \alpha_a)$ are investigated.

Marginal Grey Level Density

The marginal grey level density as used in equation 5.15 as a normalisation term is defined for this model as:

$$p(g) = \int_{\boldsymbol{\alpha}} p(g|\boldsymbol{\alpha}).p(\boldsymbol{\alpha}).d\boldsymbol{\alpha}, \tag{5.21}$$

so that the mixture vector is integrated out of the expression (or marginalised) resulting in a density that describes the marginal grey level density, i.e. irrespective of the values that the mixture vector may take.

Expanding out the definitions for the grey level likelihood, $p(g|\boldsymbol{\alpha})$ (from equation 5.16) and the PV prior density, $p(\boldsymbol{\alpha})$, (from equation 5.17) the argument of the integrand in equation 5.21 becomes:

$$p(g|\boldsymbol{\alpha}).p(\boldsymbol{\alpha}) = \sum_{v=1}^{K} p(g|\boldsymbol{\alpha}).p(\alpha_v).P(\tau_v) + \sum_{v=1}^{K-1}\sum_{a=v+1}^{K} p(g|\boldsymbol{\alpha}).p(\alpha_v, \alpha_a).P(\tau_{v,a}). \tag{5.22}$$

So that equation 5.21 results in:

$$p(g) = \int_{\boldsymbol{\alpha}} \left\{ \sum_{v=1}^{K} p(g|\boldsymbol{\alpha}).p(\alpha_v).P(\tau_v) + \sum_{v=1}^{K-1}\sum_{a=v+1}^{K} p(g|\boldsymbol{\alpha}).p(\alpha_v, \alpha_a).P(\tau_{v,a}). \right\} d\boldsymbol{\alpha}. \tag{5.23}$$

The mixture model is limited to PV voxels composed of two components and pure voxels. As a result of this, the marginalisation process consists of integrating with respect to $\boldsymbol{\alpha}$ for each PV component where each PV component consists of two PV variables, α_v and α_a. As noted by Santago and Gage in 1995 [123], these two PV variables are equivalent to a single random variable because, for a two class problem, $\alpha_v = 1 - \alpha_a$, so that the marginalisation of each PV component results in

$$\int_{\boldsymbol{\alpha}} p(g|\boldsymbol{\alpha}).p(\alpha_v, \alpha_a).P(\tau_{v,a}).d\boldsymbol{\alpha} \Rightarrow \int_{\alpha_v} p(g|\alpha_v, \alpha_a = 1-\alpha_v).p(\alpha_v, \alpha_a = 1-\alpha_v).P(\tau_{v,a}).d\alpha_v$$

So that the marginal density is given by

$$p(g) = \sum_{v=1}^{K} \int_{\alpha_v} p(g|\boldsymbol{\alpha}).p(\alpha_v).P(\tau_v).d\alpha_v + \sum_{v=1}^{K-1}\sum_{a=v+1}^{K} \int_{\alpha_v} p(g|\boldsymbol{\alpha}).p(\alpha_v, \alpha_a).P(\tau_{v,a}).d\alpha_v. \tag{5.24}$$

Noting that, for an arbitrary function, $f()$:

$$\int_{-\infty}^{\infty} \delta(x-q).f(x).dx = f(q), \tag{5.25}$$

where $\delta(0) = \infty$ and $\int_{-\infty}^{\infty} \delta(x).dx = 1$, then the integration over the pure mixture components takes advantage of the sifting property of the Dirac delta function in the pure classification class priors so that

$$p(g) = \sum_{v=1}^{K} p(g|\tau_v).P(\tau_v) + \sum_{v=1}^{K-1} \sum_{a=v+1}^{K} P(\tau_{v,a}). \int_{\alpha_v} p(g|\boldsymbol{\alpha}).p(\alpha_v, \alpha_a).d\alpha_v, \tag{5.26}$$

remembering that $p(g|\boldsymbol{\alpha})$ is the grey level likelihood, a Gaussian distribution defined in equation 5.16 which is a function of the means, $\boldsymbol{\mu}$ and standard deviations, $\boldsymbol{\sigma}$. This demonstrates that the marginal density is now composed of the sum of the scaled pure components

$$\sum_{v=1}^{K} p(g|\tau_v).P(\tau_v) \tag{5.27}$$

and the sum of the scaled PV components

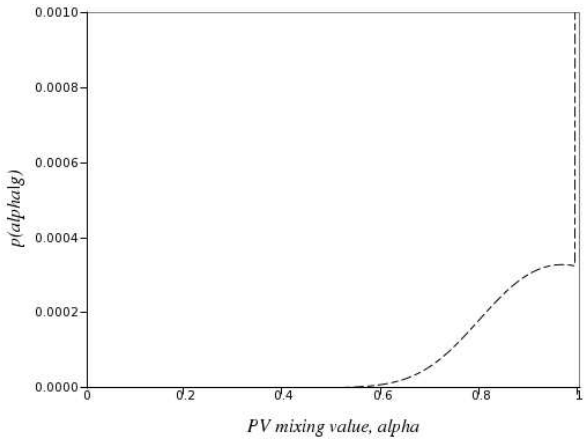
$$\sum_{v=1}^{K-1} \sum_{a=v+1}^{K} P(\tau_{v,a}). \int_{\alpha_v} p(g|\boldsymbol{\alpha}).p(\alpha_v, \alpha_a).d\alpha_v. \tag{5.28}$$

The integral in this marginal distribution is evaluated in the experiments that follow shortly using numerical integration methods. As will be seen in chapter 7, evaluation of the marginal distribution is often not required.

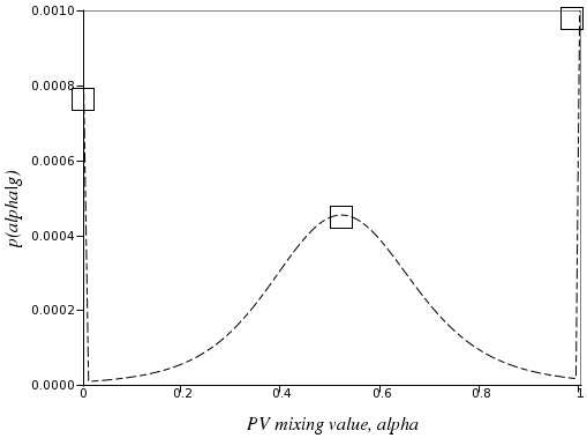
Inference from the Continuous PV Mixture Model

A different approach is required to make a decision as to the composition of a PV voxel for the model of this section, model C. In the previous models, A and B, voxel composition inference was performed by analogy with the probabilities obtained from those mixture model posterior probabilities for the discrete classification class memberships. In contrast to this, model C, the continuous PV mixture model explicitly describes the continuous mixture composition of the individual voxels. So now it is possible to make PV inferences without analogy from the posterior density, $p(\boldsymbol{\alpha}|g)$ given by Bayes theorem previously described by equation 5.15. The posterior density results in a continuously defined range of probabilities for any combination of PV mixture values for any particular grey level. Figure 5.7 illustrates three possible posterior densities for a particular exemplar mixture model.

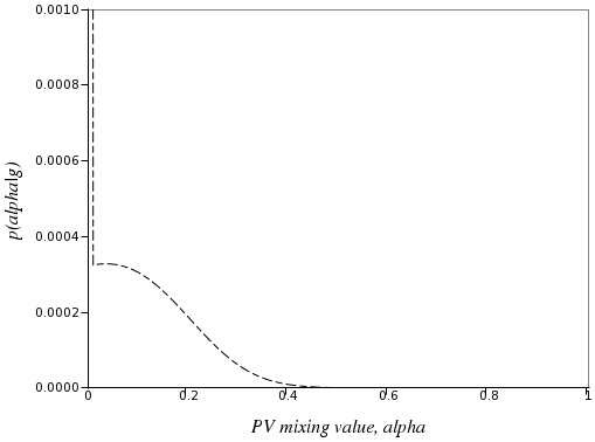
Usually some instance of the PV mixture vector, $\boldsymbol{\alpha}$ has to be inferred from the range of possible mixture values, resulting in a decision as to the PV composition of a voxel. Two approaches are often cited in the Bayesian literature, the posterior mode or the posterior mean, [116]. The mode is often used (e.g. [3, 24, 128, 95]) to infer the most



(a) $p(\alpha_{v,a}|g = 75)$



(b) $p(\alpha_{v,a}|g = 87)$



(c) $p(\alpha_{v,a}|g = 100)$

Figure 5.7: Comparison of three posterior densities, for a mixture density with parameters, $\mu_v = 75, \mu_a = 100$ and $\sigma_v, \sigma_a = 5$. The feature space is sketched in the earlier figure 5.6 which helps to explain the sharp peaks at $\alpha = 1$ or 0 in the sub-figures illustrated here. The posterior density may have more than a single mode ('□' in (b)), where the lowest probability mode corresponds to the appropriate mixture value. The relative parameter values are similar to what might be found in MR imaging data of the human brain.

likely PV mixture, usually when neighbourhood information is also being modelled, (e.g. [24, 128, 95]).

For this particular formulation where neighbourhood information is not modelled, it is clear that the mode is not a suitable choice due to the possible multi-modality of the posterior density (see figure 5.7). Therefore the mean, or expected posterior value is a more reasonable choice:

$$\mathbf{E}\left[\alpha_v | p(\boldsymbol{\alpha}|g)\right] = \int_0^1 \alpha_v . p(\boldsymbol{\alpha}|g) . d\alpha_v, \tag{5.29}$$

so that (abbreviating the conditional expectation notation, $\mathbf{E}\left[\alpha_v | p(\boldsymbol{\alpha}|g)\right]$ to $\mathbf{E}\left[\alpha_v|g\right]$)

$$\boldsymbol{\alpha}_{\boldsymbol{\omega}} = \begin{pmatrix} \mathbf{E}\left[\alpha_1|g_{\boldsymbol{\omega}}\right] \\ \mathbf{E}\left[\alpha_2|g_{\boldsymbol{\omega}}\right] \\ \ldots \\ \mathbf{E}\left[\alpha_K|g_{\boldsymbol{\omega}}\right] \end{pmatrix}.$$

Where each expectation takes the form, following on from equations 5.29 and 5.15:

$$\mathbf{E}\left[\alpha_v|g\right] = \frac{1}{p(g)} \int_0^1 \alpha_v . p(g|\boldsymbol{\alpha}) . p(\boldsymbol{\alpha}) . d\alpha_v, \tag{5.30}$$

this results in, (utilising equation 5.22):

$$\mathbf{E}\left[\alpha_v|g\right] = \frac{1}{p(g)} \left(p(g|\tau_v).P(\tau_v) + \sum_{\forall a \neq v} \int_0^1 \alpha_v . p(g|\boldsymbol{\alpha}) . p(\alpha_v, \alpha_a) . P(\tau_{v,a}) . d\alpha_v \right), \tag{5.31}$$

where $\alpha_a = 1 - \alpha_v$.

The integrals are calculated with numerical methods for the experiments that follow shortly.

Comparison of the Continuous PV density with the Gaussian-Triangle PV density of model B

It is insightful to make comparisons between this expression and equation 5.8, the posterior for model B that utilises the Gaussian-Triangle PV densities as illustrated in figures 5.5. The result of each integration in equation 5.31 can be thought of as approximately equivalent to the Gaussian-Triangle PV densities in 5.8:

$$p_{GT}(g|\tau_{v,a}^{GT}).P_{GT}(\tau_{v,a}^{GT}) \approxeq \int_0^1 \alpha_v . p(g|\boldsymbol{\alpha}) . p(\alpha_v, \alpha_a) . P(\tau_{v,a}) . d\alpha_v. \tag{5.32}$$

Remembering that $2.P_{GT}(\tau_{v,a}^{GT}) = P(\tau_{v,a})$ and from equation 5.11:

$$p_{GT}(\tau_{v,a}^{GT}) = \int_{\mu_a}^{\mu_v} p_G(g - \gamma|\tau_v) . p_T(\gamma|\tau_{v,a}^{GT}) . d\gamma, \tag{5.33}$$

which means:

$$\int_{\mu_a}^{\mu_v} p_G(g-\gamma|\tau_v).p_T(\gamma|\tau_{v,a}^{GT}).d\gamma \approxeq \frac{1}{2}\int_0^1 \alpha_v.p(g|\boldsymbol{\alpha}).p(\alpha_v,\alpha_a).d\alpha_v. \tag{5.34}$$

Expanding out the function definitions yields (using $p(\alpha_v, \alpha_a = 1-\alpha_v) = 1$, from equation 5.17):

$$\begin{aligned}\int_{\mu_a}^{\mu_v} \frac{1}{\sqrt{2.\pi.\sigma_v^2}} \exp\left(-\frac{(g-\gamma)^2}{2.\sigma_v^2}\right).\frac{1}{2.(\mu_v-\mu_a)^2}(\gamma-\mu_a).d\gamma \\ \approxeq \frac{1}{2}\int_0^1 \alpha_v \frac{1}{\sqrt{2.\pi.\sigma_\alpha^2}}.\exp\left(-\frac{(g-\mu_\alpha)^2}{2.\sigma_\alpha^2}\right).d\alpha_v.\end{aligned} \tag{5.35}$$

Due to the two class mixing and the direct correspondence between the mixing of the mean values and the mixing parameter, α_v, the mean values can be thought to correspond to pure and zero content of classification class τ_v. Therefore a change of variables is performed on the right hand side, using

$$\mu_\alpha = \alpha_v.\mu_v + (1-\alpha_v).\mu_a, \tag{5.36}$$

so that

$$\frac{d\mu_{\alpha_v}}{d\alpha_v} = \mu_v - \mu_a, \tag{5.37}$$

resulting in (assuming σ_α is not dependent on α_v):

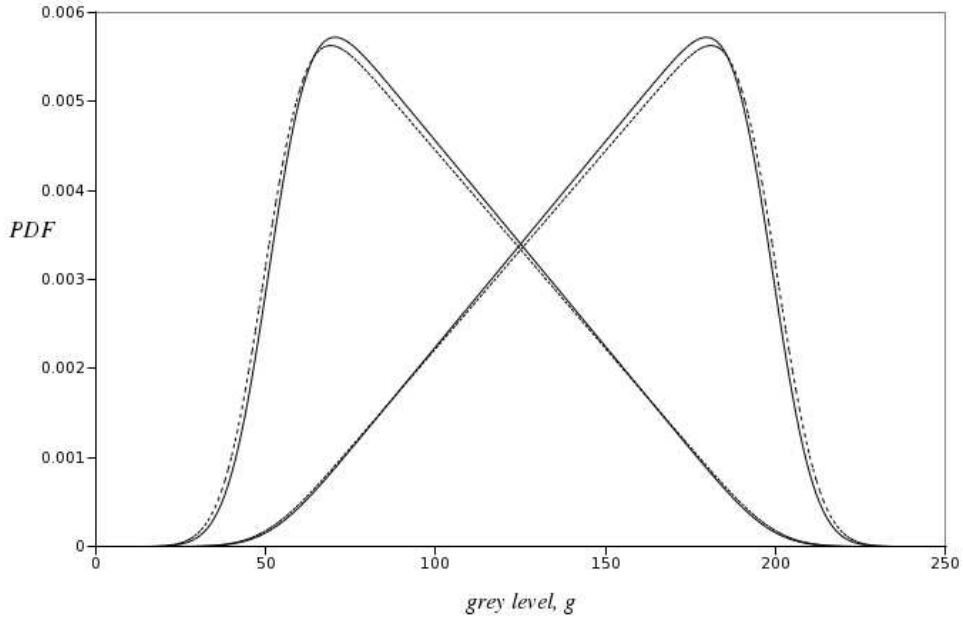
$$\begin{aligned}\int_{\mu_a}^{\mu_v} \frac{1}{\sqrt{2.\pi.\sigma_v^2}} \exp\left(-\frac{(g-\gamma)^2}{2.\sigma_v^2}\right).\frac{1}{2.(\mu_v-\mu_a)^2}(\gamma-\mu_a).d\gamma \\ \approxeq \frac{1}{2}\int_{\mu_a}^{\mu_v} \frac{\mu_\alpha-\mu_a}{\mu_v-\mu_a}.\frac{1}{\sqrt{2.\pi.\sigma_\alpha^2}}.\exp\left(-\frac{(g-\mu_\alpha)^2}{2.\sigma_\alpha^2}\right).\frac{d\mu_\alpha}{\mu_v-\mu_a}.\end{aligned} \tag{5.38}$$

This simplifies to:

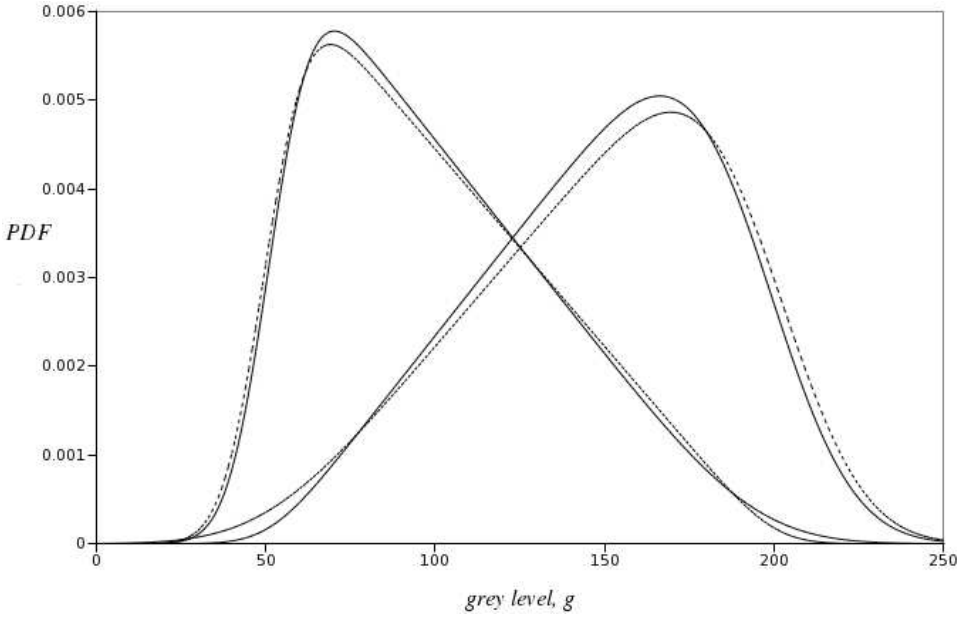
$$\begin{aligned}\int_{\mu_a}^{\mu_v} \frac{1}{\sqrt{2.\pi.\sigma_v^2}} \exp\left(-\frac{(g-\gamma)^2}{2.\sigma_v^2}\right).\frac{(\gamma-\mu_a)}{2.(\mu_v-\mu_a)^2}.d\gamma \\ \approxeq \int_{\mu_a}^{\mu_v} \frac{1}{\sqrt{2.\pi.\sigma_\alpha^2}}.\exp\left(-\frac{(g-\mu_\alpha)^2}{2.\sigma_\alpha^2}\right).\frac{(\mu_\alpha-\mu_a)}{2.(\mu_v-\mu_a)^2}.d\mu_\alpha\end{aligned} \tag{5.39}$$

Clearly they would be equivalent except for the fact that $\sigma_\alpha \neq \sigma_v$ and σ_α is not constant because σ_α varies according to:

$$\sigma_\alpha^2 = \alpha_v^2.\sigma_v^2 + (1-\alpha_v)^2.\sigma_a^2. \tag{5.40}$$

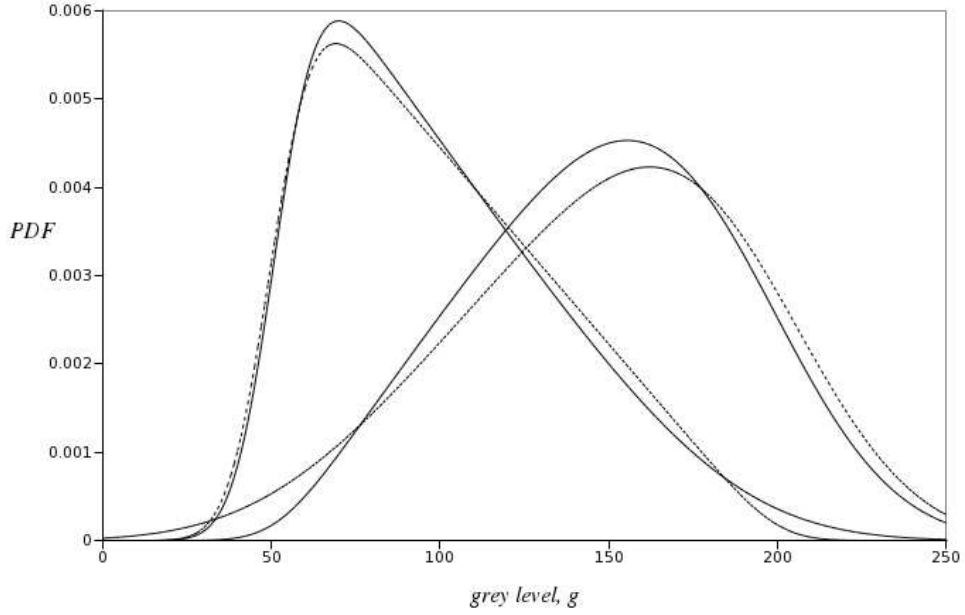


(a)

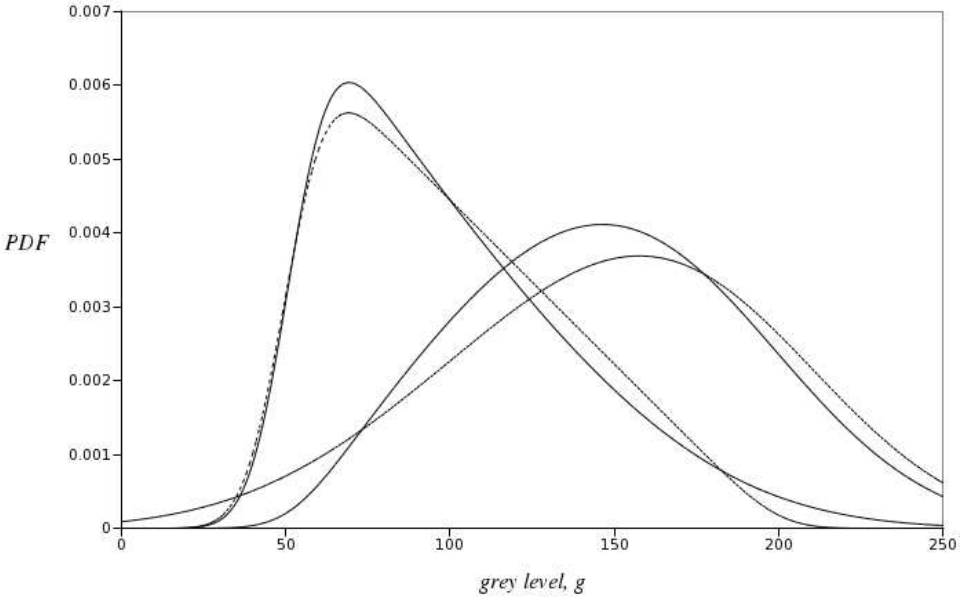


(b)

Figure 5.8: Comparisons of posterior mean point estimates of model C (complete higher lines) with the Gaussian Triangle convolution PV densities of model B (dotted lower lines) for increasing difference between the standard deviations of each of the pure classes. These plots (see also overleaf) illustrate that for equal standard deviations, the posterior mean of model C is approximately equivalent to the Gaussian Triangle Convolution PV densities of model B. When the standard deviations are quite different then the two models diverge. Parameters used were, $\mu_v = 50$, $\mu_a = 200$ and (a) $\sigma_v, \sigma_a = 10$; (b) $\sigma_v = 10, \sigma_a = 20$; (c) $\sigma_v = 10, \sigma_a = 30$; and (d) $\sigma_v = 10, \sigma_a = 40$. *Continued on next page.*



(c)



(d)

Figure 5.8: *Continued.*

Figure 5.8 illustrates the similarity of these two functions for exemplar sets of parameters. For equal standard deviations the two models appear to be approximately equivalent but when the standard deviations are not equal the two models diverge. This would also appear to be the case for the model found in [102] which also uses a linear combination of variances, similar to Vokurka *et al.* in [148].

So the result of equation 5.39 shows that models B and C are approximately equivalent for equal standard deviations. Model B was previously described in [140] as

"...an estimate of the mean volumetric contribution to the formation of a voxel with grey level g."

However this statement was made without further justification. The result of equation 5.39 now provides evidence to show that model B does provide an estimate of the mean volumetric contribution when the standard deviations are very similar, but when the standard deviations are quite different then the models are not equivalent.

This result is interesting and experiments that are discussed shortly will also compare the performances of the classifiers based on models B and C.

5.2 Experiments, Results and Discussions

The procedures for a series of experiments are now discussed and their results are also presented in this section.

5.2.1 Classification of Two Class Simulated Data

This experiment utilises the Gaussian mixture model (A), the Gauss-Triangle PV model (B), and the explicit mixture model (C) to classify two classification class simulated PV data. The results of each classification are then quantified by comparison with the simulated ground truth.

These simulations and their respective classifications with PV models A, B and C allow the performance of each of the PV classifiers to be assessed under controlled conditions where the parameters of the simulated data are known *a priori*. The simulated data is generated with a range of CNRs that might be realistically found to occur in medical imaging data.

Experimental Procedure

Two class simulated PV data with a series of ellipsoid bands or concentric spheroids were generated. These concentric spheroids maximize the surface area thereby maximising the number of PV voxels that are generated after convolution with the PSF. Table 5.1 lists the simulations and the respective parameters used by the software simulator to create the simulated data. Exemplar central transverse image slices can be seen in figure 5.9. Each simulation was downsampled from a high resolution $512 \times 512 \times 512$

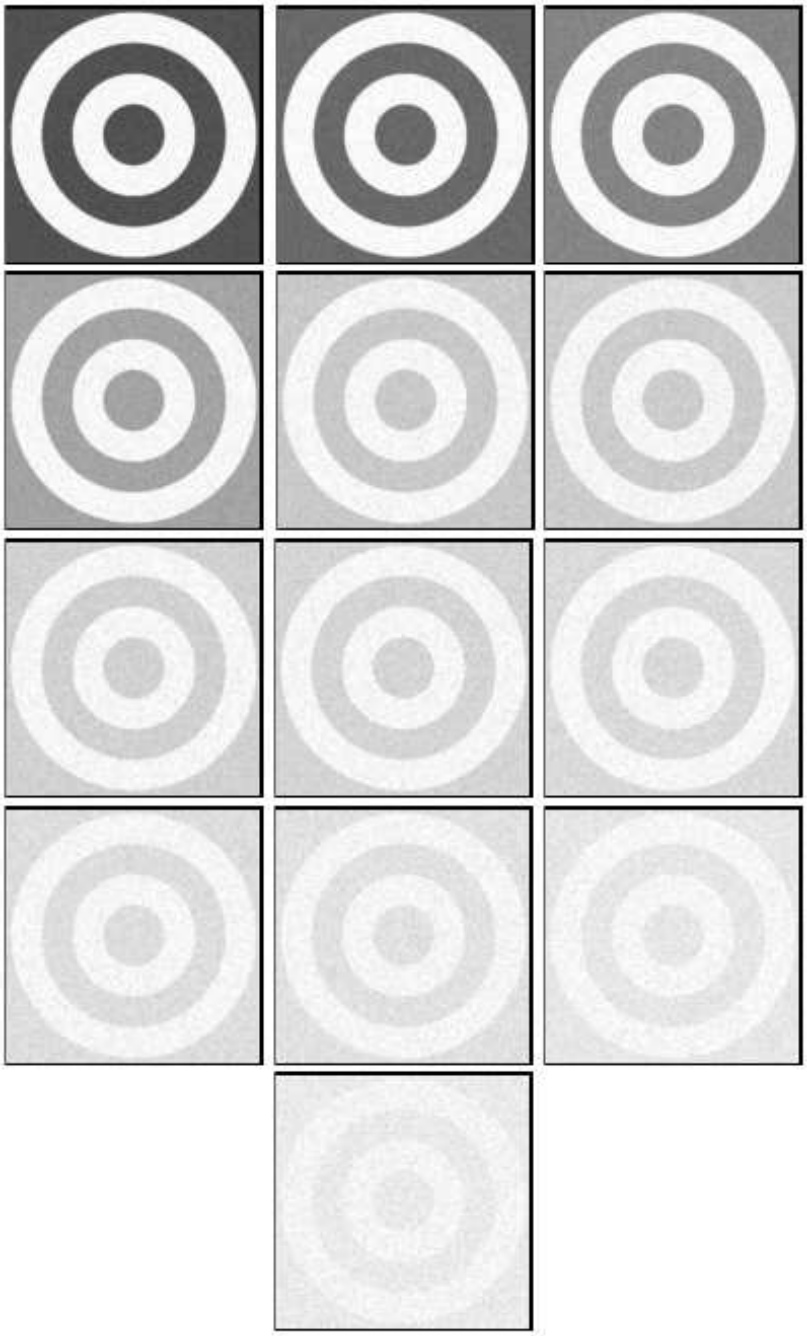


Figure 5.9: Exemplar middle transverse image slices from each two class simulated data volume (details in table 5.1). Top left slice corresponds to the simulation with the greatest mean distance (1000) and the remaining slices are displayed in CNR descending order from left to right and from top to bottom corresponding to table 5.1. Notice how the contrast to noise ratio gradually becomes worse until the two classes become very difficult to distinguish in the final simulation with only a mean distance of 40. These image slices were produced with the aid of MRIcro, [114], where the contrast auto-balance feature was used, which scales the highest and lowest 1% of image intensities to 255 and 0 grey levels respectively.

Table 5.1: Summary of the two class simulations' parameters.

CNR	**Mean Distance**	**Mean 1**	**Mean 2**	**Stan. Dev. 1**	**Stan. Dev. 2**	**No Bands**
50	1000	500	1500	20	20	2
40	800	600	1400	20	20	2
30	600	700	1300	20	20	2
20	400	800	1200	20	20	2
10	200	900	1100	20	20	2
9	180	910	1090	20	20	2
8	160	920	1080	20	20	2
7	140	930	1070	20	20	2
6	120	940	1060	20	20	2
5	100	950	1050	20	20	2
4	80	960	1040	20	20	2
3	60	970	1030	20	20	2
2	40	980	1020	20	20	2

sized volume. A symmetric Gaussian PSF was selected so that the FWHM of the PSF corresponded to voxel sizes that would produce a $128 \times 128 \times 128$ size downsampled.

Each simulation was then classified with a PV classifier utilising each of the models described thus far, models A, B and C. The parameters for each model were obtained from either the simulation parameters or for the case of the prior terms, $P(\tau_v)$ and $P(\tau_{v,a})$, these were estimated utilising the ground truth associated with each simulation. Knowledge of the true parameter values allowed for the performance of the classifier with the respective model to be assessed, independent of any parameter estimation algorithm.

Performance assessment was undertaken with the voxel Root Mean Square (RMS) error metric, in common with the quantitative assessments previously undertaken in the literature on PV classifier performances, (see e.g. [74, 128]). Each voxel of the classifier output is compared with the noiseless ground truth data that has been convolved with the PSF. The mean of the squared differences for all the voxels in the data volumes is calculated and the square root is taken.

Results and Discussion

The voxel RMS error results for the classification of the two class PV simulations can be seen in figures 5.10 and 5.11 for PV classifiers utilising models A, B and C.

The voxel RMS error measures the disagreement between the classifier PV estimate and the ground truth for every voxel that has been classified in the imaging data. These figures appear to demonstrate that the classifiers utilising models B and C possess similar PV classification performances. This is consistent with findings in section 5.1.3, equation 5.39 when the standard deviations for the two pure classes are equal, which is true for these simulations.

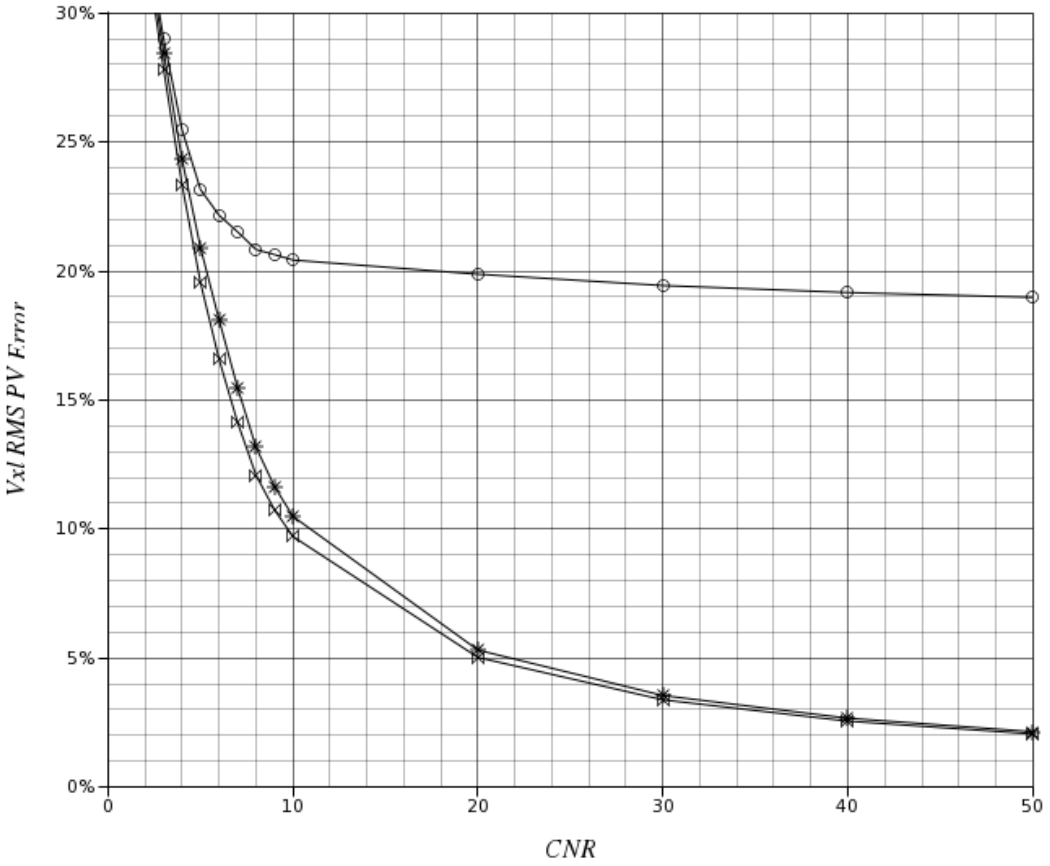


Figure 5.10: Plot of RMS PV voxel errors obtained for the classification of two class simulated data. $\circ$ correspond to data points obtained for the finite Gaussian mixture model (model A); $\bowtie$ correspond to data points obtained for the Gaussian-Triangle convolution formulation (model B); and $*$ correspond to data points obtained for the explicitly modelled mixture values (model C). Notice the approximate reciprocal relationship between the RMS voxel error and the CNR value.

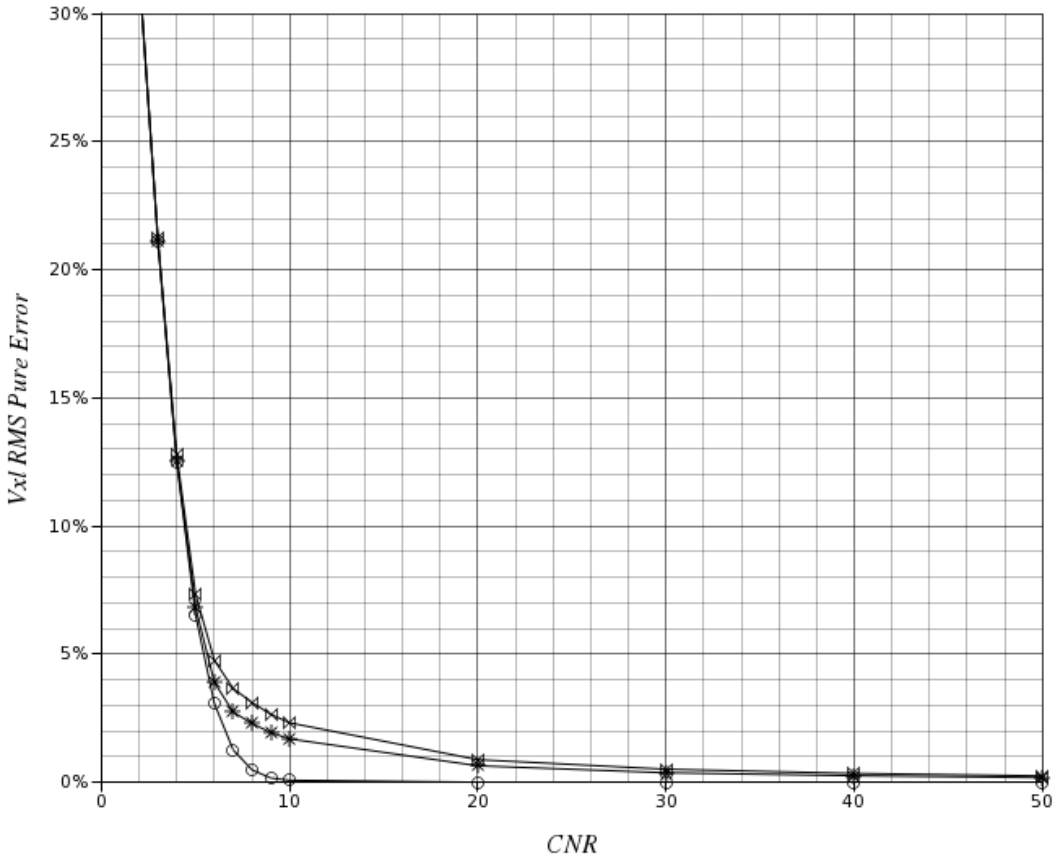


Figure 5.11: Plot of RMS pure voxel errors obtained for the classification of two class simulated data. $\circ$ correspond to data points obtained for the finite Gaussian mixture model (model A); $\bowtie$ correspond to data points obtained for the Gaussian-Triangle convolution formulation (model B); and $*$ correspond to data points obtained for the explicitly modelled mixture value model (model C).

Each of the PV classifiers' performances appear to degrade as the mean distance decreases, as expected, especially when one considers the especially low CNR for some of the simulations as illustrated by figure 5.9.

The finite Gaussian mixture model, described by model A, where the PV effect was not modelled by additional mixture components has the worst PV voxel classification performance for the majority of the simulations displayed in figure 5.10. But figure 5.11 illustrates that the finite Gaussian mixture model actually performs slightly better on classifying pure voxels in some of the simulated data volumes when compared with the two models that do model the PV effect. One might argue that the differences between inter-classifiers' performance presented in figure 5.11 is slight. This is especially true given that for the lower CNRs (CNR<6), the RMS voxel error is very much worse.

Furthermore, the pure voxels are identified exactly for this simulated PV data with the simulation parameter information. For real imaging data the true or underlying parameter values will not be known. Also the PV simulation process described here is a simplification of the true data acquisition process. Real imaging data will have many other minor artefacts or deviations from an expected model thus complicating the somewhat artificial notion of being able to identify a voxel as consisting of a single component or a mixture of components.

Reciprocal PV RMS Error Rule

The degradation of the PV RMS error appears to follow the reciprocal of the CNR. This observation is consistent with other similar intensity based classifiers and simulations, see [153]. This can be understood by inspection and comparison of the formulae for the CNR and the voxel RMS error. Recall that the CNR is defined as:

$$\mathrm{CNR}_{v,a} = \frac{\|\mu_v - \mu_a\|}{\sigma_{v,a}} \tag{5.41}$$

and the intensity normalised RMS error is defined as:

$$\mathrm{RMS} = \frac{\sqrt{\frac{1}{K}\sum_{\forall \boldsymbol{\omega}} \left(\mu_{T_{\boldsymbol{\omega}}} - \mu_{\boldsymbol{\alpha}_{\boldsymbol{\omega}}}\right)^2}}{\|\mu_v - \mu_a\|}, \tag{5.42}$$

where $\mu_{T_{\boldsymbol{\omega}}}$ is the ground truth value for voxel $\boldsymbol{\omega}$ and $\mu_{\boldsymbol{\alpha}_{\boldsymbol{\omega}}}$ is the classifier estimate for voxel $\boldsymbol{\omega}$. If the CNR and RMS error do possess a reciprocal relationship, then

$$\frac{\sigma_{v,a}}{\|\mu_v - \mu_a\|} = \frac{\sqrt{\frac{1}{K}\sum_{\forall \boldsymbol{\omega}} \left(\mu_{T_{\boldsymbol{\omega}}} - \mu_{\boldsymbol{\alpha}_{\boldsymbol{\omega}}}\right)^2}}{\|\mu_v - \mu_a\|}, \tag{5.43}$$

which simplifies to

$$\sigma_{v,a}^2 = \frac{1}{K}\sum_{\forall \boldsymbol{\omega}} \left(\mu_{T_{\boldsymbol{\omega}}} - \mu_{\boldsymbol{\alpha}_{\boldsymbol{\omega}}}\right)^2. \tag{5.44}$$

The r.h.s. is clearly a variance measure which helps to illustrate why the CNR value appears to share a reciprocal relationship with the PV RMS voxel error.

Table 5.2: Summary of the two class simulations' parameters with unequal standard deviations.

CNR	**Mean 1**	**Mean 2**	**Stan. Dev. 1**	**Stan. Dev. 2**
20.0	900	1100	10	10
15.0	900	1050	10	10
12.6	900	1100	10	20
10.0	900	1000	10	10
9.5	900	1050	10	20
8.9	900	1100	10	30
6.9	900	1100	10	40
6.7	900	1050	10	30
6.3	900	1000	10	20
5.5	900	1100	10	50
5.1	900	1050	10	40
4.5	900	1000	10	30
4.2	900	1050	10	50
3.4	900	1000	10	40
2.8	900	1000	10	50

5.2.2 Classification of Two Class Simulated PV Data with Unequal Standard Deviations

At the end of section 5.1.3, models B and C were compared analytically. It was shown that model B approximates model C when the data possesses equal standard deviations for all the classification classes. Figure 5.8 illustrated that for greater differences in the standard deviations of the classes being classified, the further model B diverges away from a linear mixture PV model. The previous section presented experiments on simulated PV data with equal standard deviations, which confirmed that models B and C possess approximately equal classification performance. Therefore a further set of simulated PV data was generated but with unequal standard deviations for the two classification classes.

Experimental Procedure

The parameters of the simulated PV data with unequal variances can be seen in table 5.2.

The parameters were selected so as to concentrate on the CNR values most likely to be applicable to real medical imaging data and where greatest change was found for the voxel RMS error in the preceding set of experiments, i.e. for CNR values between 0 and 20. This was undertaken by selecting three mean distances (900,1050; 900, 1100; and 900, 1150). The standard deviation for the lower mean intensity class (900) was kept constant with a value of 10 and the high mean intensity classes, utilized standard deviation values of 10,20,30,40 and 50. These simulations therefore produced a reasonable number of data points to enable the comparative assessment of models

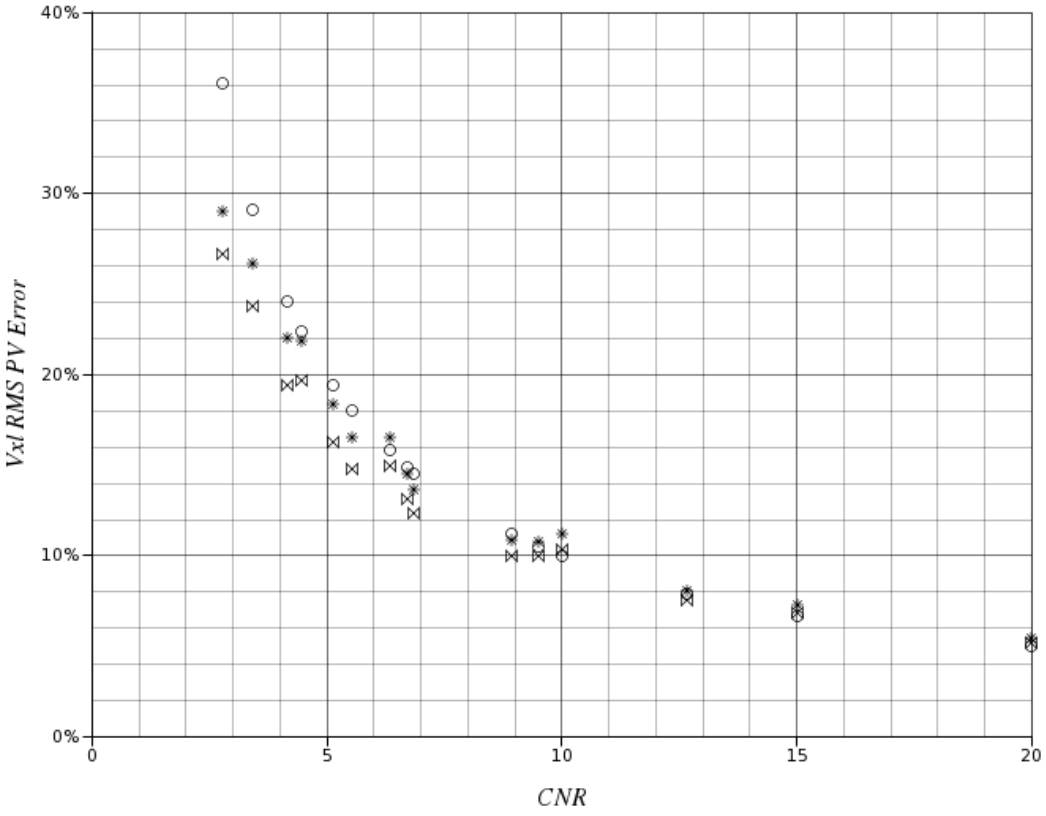


Figure 5.12: Comparison of Models B and C for classifying two class simulated PV data with unequal standard deviations. RMS PV voxel errors are shown. $\bowtie$ correspond to data points obtained for the Gaussian-Triangle convolution formulation (model B); and $*$ correspond to data points obtained for the explicitly modelled mixture values (model C). Also shown is the reciprocal of the CNR, $\circ$, illustrating the reciprocal relationship of the PV error in relation to the CNR value.

B and C PV classifiers for a variety of CNR values resulting from unequal standard deviations. The CNR for unequal standard deviations is calculated here using the mean variance, i.e.

$$CNR'_{v,a} = \frac{\|\mu_v - \mu_a\|}{\sqrt{\frac{\sigma_v^2+\sigma_a^2}{2}}}. \tag{5.45}$$

Performance assessment was again undertaken with the voxel RMS error between the true ground truth values (noiseless) and the output of the classifiers.

Results and Discussion

The voxel RMS errors for the PV, pure and gross (PV and pure) voxels can be seen in figures 5.12, 5.13 and 5.14 respectively. Figure 5.12 also includes points plotted from the reciprocal of the CNR value and can therefore be compared with the PV voxel RMS values in this figure.

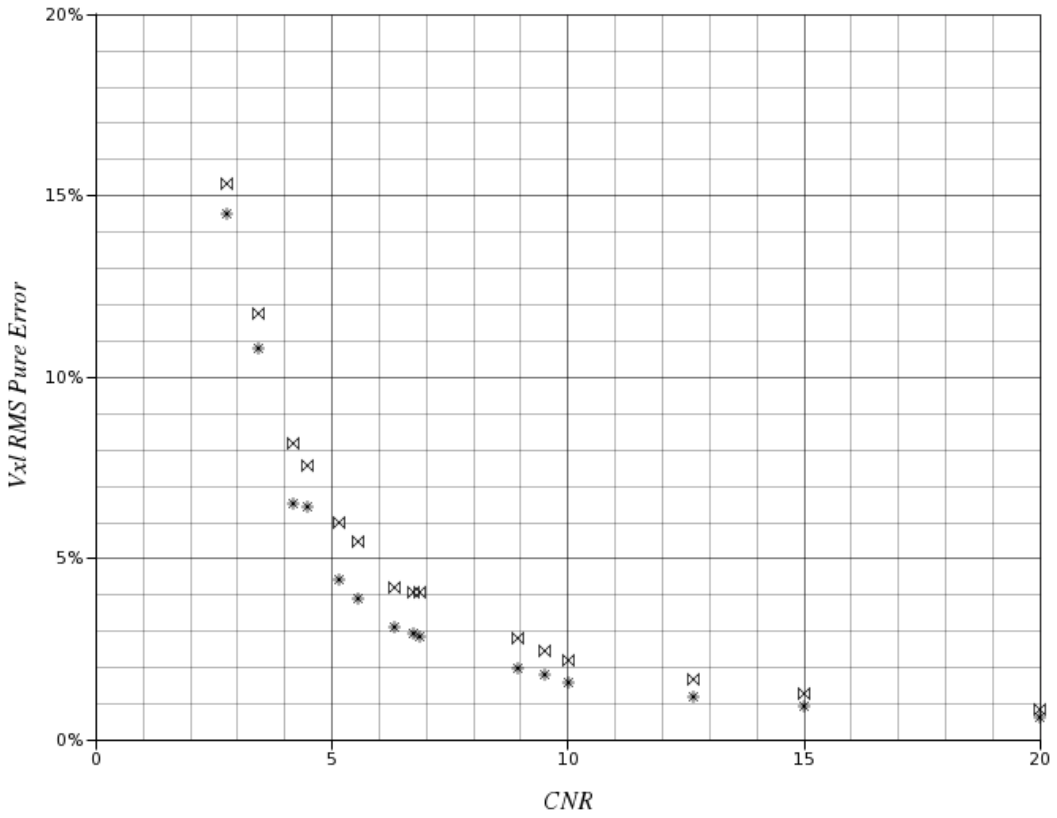


Figure 5.13: Comparison of Models B and C for classifying two class simulated PV data with unequal standard deviations. RMS pure voxel errors are shown. $\bowtie$ correspond to data points obtained for the Gaussian-Triangle convolution formulation (model B); and $*$ correspond to data points obtained for the explicitly modelled mixture values (model C).

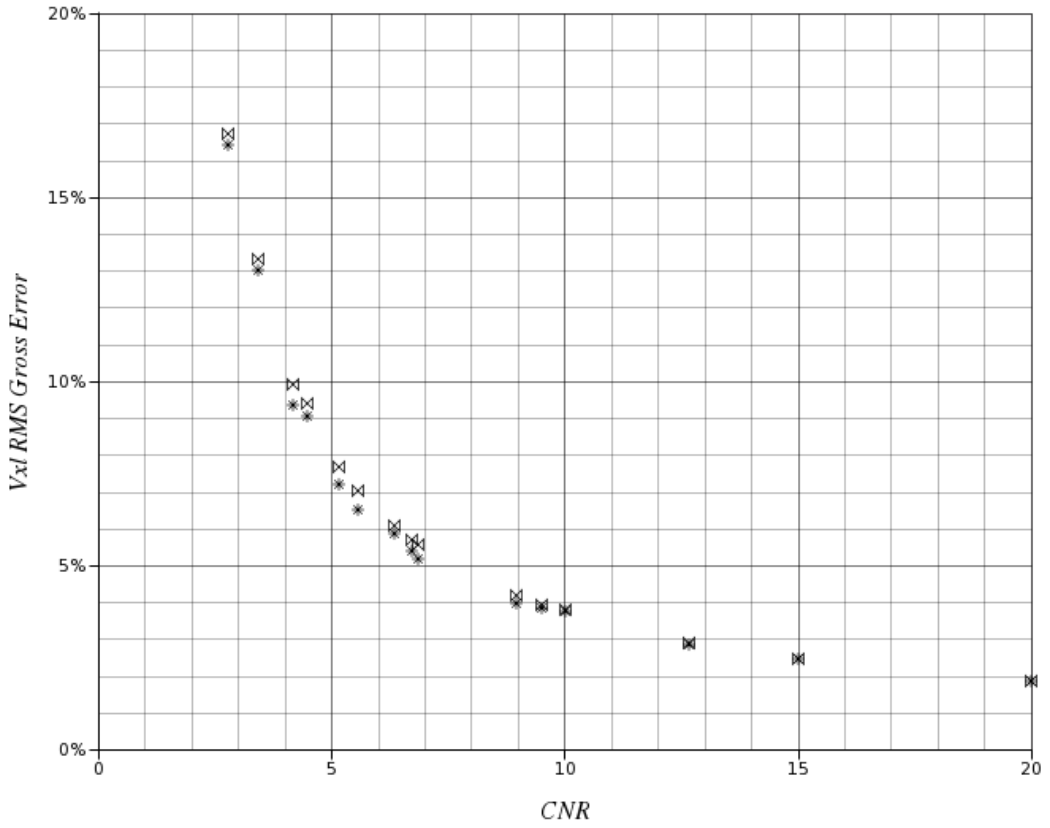


Figure 5.14: Comparison of Models B and C for classifying two class simulated PV data with unequal standard deviations. RMS gross voxel errors are shown. $\bowtie$ correspond to data points obtained for the Gaussian-Triangle convolution formulation (model B); and $*$ correspond to data points obtained for the explicitly modelled mixture values (model C).

Each of the figures demonstrate generally good agreement between the performance of the PV classifier utilizing model B and the performance of the PV classifier utilizing model C. This is especially true for the voxel RMS error for the pure and PV voxels combined (gross, see figure 5.14). The difference in the PV error for models B and C appears to decrease for higher CNR values (> 10). This is due to the data points with CNR values of 15.0 and 20.0 possessing equal standard deviations. The PV voxel RMS error disparities appear to complement the pure voxel RMS error disparities, which result in the almost equivalent gross voxel RMS error. These disparities are likely to have arisen from the differences in the distributions, previously illustrated in figure 5.8. For larger differences in standard deviation value, the spread of the PV distribution for model B is wider in comparison to the spread of what would be the PV distributions for model C. For model B, this reduces the relative weight of the pure components for the intensities surrounding the (pure component) mean values. This has the effect of reducing the classifier performance for pure components, unless the CNR is particularly low, i.e. where the pure distribution components start to overlap each other. This can be observed in figure 5.13. The wider spread of the PV components for model B also increases the probability that some PV voxels may possess increased posterior probabilities for model B in relation to the usually dominating effect of the pure components. The effect on the gross error is therefore small due to this complementary behaviour.

The theoretical CNR-reciprocal PV error also illustrated in figure 5.12 demonstrates reasonably good agreement with the experimental values. This is especially true for CNR values > 6. For the very low CNR values, the effect of the overlap of the pure components would result in this theoretical PV error no longer being valid. This is due to dominating effect of the pure components and dominating effect of the error from the overlap of the pure distributions. This observation is supported with the following argument.

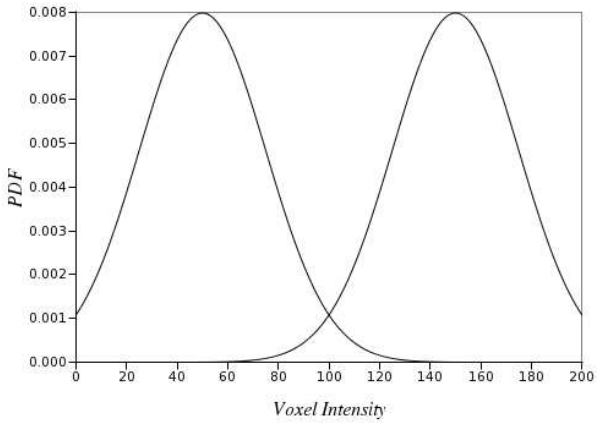
If there is some imaging data that can be described with a two class mixture model with Gaussian distributed components, with means μ_1, μ_2 and equal standard deviations σ, then it is desired that the two distributions have zero overlap which occurs at some multiple, κ of the standard deviation. This can be expressed by (where $\mu_2 > \mu_1$):

$$(\mu_2 - \kappa.\sigma) - (\mu_1 + \kappa.\sigma) = 0. \tag{5.46}$$

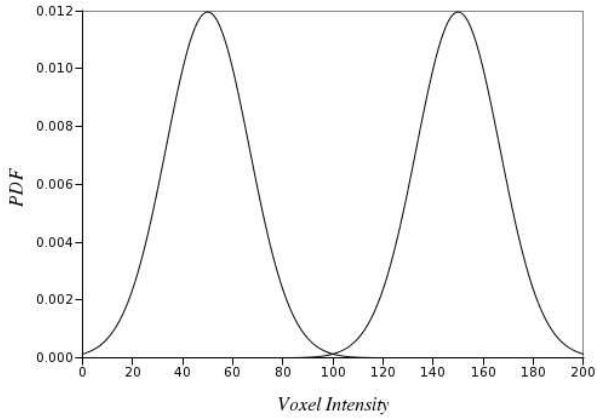
Solving for multiples of κ reveals:

$$2.\kappa = \frac{(\mu_2 - \mu_1)}{\sigma}, \tag{5.47}$$

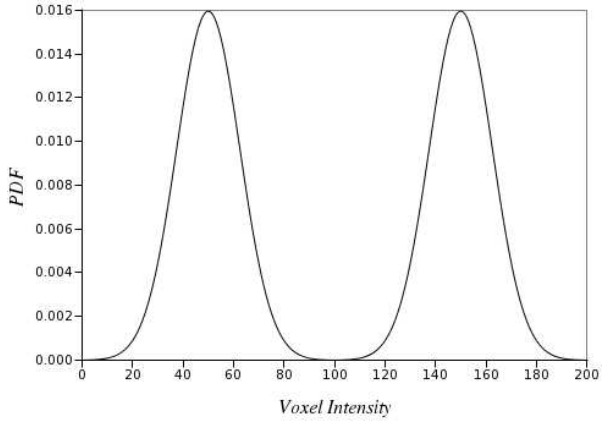
which is now expressed in terms of the CNR value for this imaging data. Consider the fact that Gaussian distributions possess very little distribution at a distance of $3.\sigma$ or greater from the mean value. Then this would also suggest (from equation 5.47) that CNR values greater than 6, i.e. where $\kappa = 3$, would result in imaging data with non-overlapping intensity distributions. This therefore helps to explain the greater disparities, seen in figure 5.12, between the theoretical PV error and the PV errors obtained for PV models B and C for CNR values smaller than 6. This is further illustrated with exemplar Gaussian component plots in figure 5.15.



(a) CNR=4



(b) CNR=6



(c) CNR=8

Figure 5.15: Illustration of the amount of overlap for different CNR values. Two Gaussian components, mean values of 50 and 150 and equal standard deviations, are illustrated in each sub-figure. The CNR value is adjusted by varying the standard deviations. Notice for CNR$<$ 6 the overlap becomes significant.

5.3 In Conclusion

This chapter has described two PV models that infer the constituency of a PV voxel by analogy of probability to the contents (A and B). It has also described a PV model (C) that has been found to be approximately equivalent to one of the analogous PV models (B). The performance of the classifiers were limited for smaller CNR values. Observations from real imaging data suggest that clinical CNR values are usually below 10, (for example some neurological MR imaging data). If a contrast medium is being imaged within the patient, or a longer acquisition time is used or larger voxels are imaged, then higher CNR values might be found.

These observations together with the observed performances of the classifiers in this chapter suggest that models B and C might be immediately suitable for application to some types of clinical data. Except, in clinical practise, one should also remember that other sources of error, such as issues of parameter estimation and limited numbers of voxel data samples may produce inferior results to those shown. Similarly, for quantifying the performance of real imaging data, the separation of the PV and pure components is a somewhat artificial step due to a number of reasons. These reasons include the difficulty in knowing whether to assign a ground truth voxel to pure or a PV mixture, due to an inherent error in any ground truth creation process; and the number of discretization steps used to divide the possibly continuous intensity scale of the imaging data may further confuse or complicate the division into pure and PV voxels. Thus the performance characterisation of real imaging data requires a gross RMS voxel error that combines the pure and PV errors.

Interestingly the PV RMS voxel error obtained for models B and C was found to be approximately equal to the reciprocal of the CNR for values greater than 6. This was not the case for the simpler finite Gaussian mixture model (A), which does not model the PV effect.

The framework used in model C is to be preferred for further development as it will provide for the inclusion of new model developments using techniques that are consistent with conventional probability theory. As will be seen in the chapters that follow, these developments produce improvements to the classification performance of the models discussed in this chapter.

Chapter 6

Feature Space Enhancement with Image Gradient Magnitude

Chapter 5 found that the classification performance of intensity based PV classifiers was limited for low CNR values. This can be attributed to the large amount of overlap of the individual class densities on the single intensity axis, especially when one considers the PV components of the model and how these overlap with the other components. This means that given a particular intensity value, if the data has low CNR, then that intensity value could equally likely have been generated by either of the pure signal components in the mixture model or by the PV component in between the two classes.

Even if the data has a high CNR(> 6) other sources of error may arise. For example, if there are three or more pure components the classification process may become even harder. An intensity located at a position close to the mean value of the central pure or PV components could have been generated from the central pure component or from the PV components of the exterior pure components. This situation is illustrated in figure 6.1

Many PV models of MR imaging data have incorporated multiple image sequences. These models extend the single intensity feature space to a multiple intensity feature space (e.g. [24, 74, 138]). These multiple channels allow the vectorial mean distance between individual pure components to be increased and allow unique vectorial intensities to be associated with the individual PV components. Sometimes however multiple image channels may not be available and additional alternative features may be sought to extend the limited feature space.

6.1 Methodology

6.1.1 Model D: 2-D Gradient Magnitude PV Classifier

One such possible feature for biomedical imaging data is a localised image gradient magnitude measure, or edge strength as suggested by Williamson *et al.* in 2002, [157].

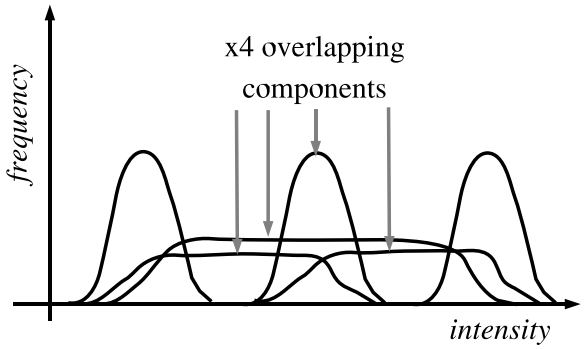


Figure 6.1: Illustration of the dilemma that may occur when attempting to classify single-channel imaging data with more than two classification class components. The central pure component shares intensity values from the PV components that are shared between itself and the adjacent pure components. In addition a third PV component also shares the same range of intensity values and is generated from the two adjacent pure components.

In [157] regions of PV voxels are assumed to be associated with higher values of gradient magnitude. Conversely, regions of pure voxels are assumed to be associated with lower values of gradient magnitude.

To understand these assertions, one might like to consider the scenario where PV voxels are created at the boundary between two classification classes; these PV voxels will be associated with higher levels of gradient magnitude in comparison to the pure voxel regions. Clearly not all pure voxels will have low gradient magnitude due to the inherent noise of the imaging data and not all PV voxels will have high gradient magnitude values, but on average these assumptions will be true.

Therefore a two dimensional feature space can be created, combining intensity and gradient magnitude measurements or features from data consisting of pure and PV component voxels. An exemplar feature space for PV simulated data can be seen in figure 6.2.

To model this feature space, Williamson *et al.* chose to incorporate a probabilistic description for 2-D gradient magnitude within a PV mixture model formulation that was previously utilised by the same research group [148], already described in this work, in the preceding chapter, section 5.1.2.

Williamson *et al.* continued to assume that pure classification classes for MR data are governed by Gaussian PDFs. With this assumption a result for the combination of two Gaussian distributed random variables combined in quadrature was taken, originally given by Rice in 1938, [111]. The result is known as the Rice or Rician density after Rice. This density is also used to describe magnitude MR imaging data, [48]. As discussed in chapter 2, at high SNR values, the Rician density tends towards a Gaussian density, while at low SNR values it tends towards a Rayleigh density.

Initially a probabilistic description of the gradient magnitude applied to PV affected imaging data will be evaluated and then in the following section this probabilistic

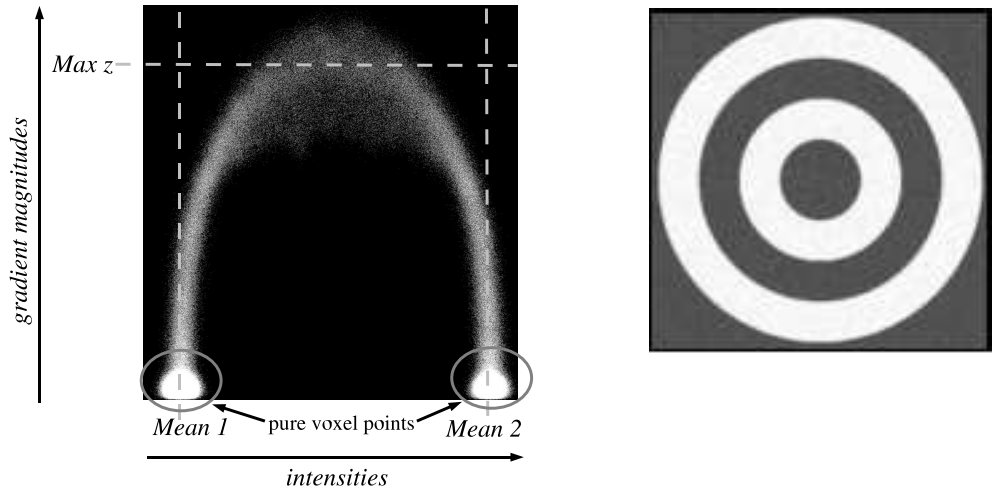


Figure 6.2: Illustrative example of (a) intensity vs localised image gradient magnitude scatter plot or feature space and (b) image slice from simulated PV data from which the scatter plot was generated. A 3-D gradient magnitude kernel was used to calculate the gradient magnitude. Notice how the gradient magnitude is larger for the PV points located between the mean values (mean 1 and mean 2) of the two pure mixture components.

description will be incorporated into the PV mixture model of [148], thus reviewing the methodology presented in [157].

Derivation of a Gradient Magnitude Random Variable

The process of calculating the 2-D gradient magnitude of imaging data consists of the following steps:

1. Convolve imaging data, I, with gradient masks, M_x and M_y for the x and y data dimensions, creating gradient channels, Z_x and Z_y:

$$Z_x = I * M_x, \tag{6.1}$$

and

$$Z_y = I * M_y. \tag{6.2}$$

2. Calculate magnitude, Z_{2D} of the two gradient channels, Z_x and Z_y:

$$Z_{2D} = \sqrt{Z_x^2 + Z_y^2}. \tag{6.3}$$

Convolving the Image Data
For the first step, an important idea and one that was made apparent by the third model, model C, in the preceding chapter is that every voxel,$\boldsymbol{\omega}$, in the imaging data, I,

results in an intensity, $I(\boldsymbol{\omega}) = g_{\boldsymbol{\omega}}$ drawn from a Gaussian distributed random variable so that (for 2-D):

$$I = \begin{pmatrix} X_{(0,0)^T} = g_{(0,0)^T} & X_{(1,0)^T} = g_{(1,0)^T} & \dots & X_{(n_x,0)^T} = g_{(n_x,0)^T} \\ X_{(0,1)^T} = g_{(0,1)^T} & X_{(1,1)^T} = g_{(1,1)^T} & \dots & X_{(n_x,1)^T} = g_{(n_x,1)^T} \\ \dots & \dots & \dots & \dots \\ X_{(0,n_y)^T} = g_{(0,n_y)^T} & X_{(1,n_y)^T} = g_{(1,n_y)^T} & \dots & X_{(n_x,n_y)^T} = g_{(n_x,n_y)^T} \end{pmatrix},$$

where n_x and n_y are the dimensions of the image and each voxel intensity is an instance of a Gaussian distributed random variable, i.e.

$$X_{\boldsymbol{\omega}} : \mathcal{N}(\mu_{\boldsymbol{\alpha}_{\boldsymbol{\omega}}}, \sigma^2_{\boldsymbol{\alpha}_{\boldsymbol{\omega}}}). \tag{6.4}$$

The convolution masks, M_x and M_y are ordered sets or matrices of real valued kernel weights, for example:

$$M_x = \begin{pmatrix} +1 \\ 0 \\ -1 \end{pmatrix}, \; M_y = \begin{pmatrix} +1 & 0 & -1 \end{pmatrix};$$

whose elements are indexed by 2-D points, $\boldsymbol{\omega}_{m_x}$ and $\boldsymbol{\omega}_{m_y}$.

These definitions allow for the convolutions in equations 6.1 and 6.2 to be further defined as:

$$Z_x(\boldsymbol{\omega}) = \sum_{\forall \boldsymbol{\omega}_{m_x}} I(\boldsymbol{\omega}'_x).M_x(\boldsymbol{\omega}_{m_x}); \tag{6.5}$$

and

$$Z_y(\boldsymbol{\omega}) = \sum_{\forall \boldsymbol{\omega}_{m_y}} I(\boldsymbol{\omega}'_y).M_y(\boldsymbol{\omega}_{m_y}); \tag{6.6}$$

where vector arithmetic is used to calculate the points at which the image data is convolved, i.e. $\boldsymbol{\omega}'_x = \boldsymbol{\omega} + \boldsymbol{\omega}_{m_x} - \boldsymbol{\omega}_{c_x}$ and $\boldsymbol{\omega}'_y = \boldsymbol{\omega} + \boldsymbol{\omega}_{m_y} - \boldsymbol{\omega}_{c_y}$. $\boldsymbol{\omega}_{c_x}$ and $\boldsymbol{\omega}_{c_y}$ are the central points of each of the kernel masks.

As can be seen from equations 6.5, 6.6 and 6.4, the elements of $Z_x(\boldsymbol{\omega}) = z_{x,\boldsymbol{\omega}}$ and $Z_y(\boldsymbol{\omega}) = z_{y,\boldsymbol{\omega}}$ are the results of sums of weighted Gaussian distributed random variables.

The sum of weighted Gaussian distributed random variables are also Gaussian distributed (see appendix C), therefore:

$$Z_x = \begin{pmatrix} S_{x,(0,0)^T} = z_{x,(0,0)^T} & S_{x,(1,0)^T} = z_{x,(1,0)^T} & \dots & S_{x,(n_x,0)^T} = z_{x,(n_x,0)^T} \\ S_{x,(0,1)^T} = z_{x,(0,1)^T} & S_{x,(1,1)^T} = z_{x,(1,1)^T} & \dots & S_{x,(n_x,1)^T} = z_{x,(n_x,1)^T} \\ \dots & \dots & \dots & \dots \\ S_{x,(0,n_y)^T} = z_{x,(0,n_y)^T} & S_{x,(1,n_y)^T} = z_{x,(1,n_y)^T} & \dots & S_{x,(n_x,n_y)^T} = z_{x,(n_x,n_y)^T} \end{pmatrix},$$

where

$$S_{x,\boldsymbol{\omega}} : \mathcal{N}(\mu_{x,\boldsymbol{\omega}}, \sigma^2_{x,\boldsymbol{\omega}}). \tag{6.7}$$

Similarly:

$$Z_y = \begin{pmatrix} S_{y,(0,0)^T} = z_{y,(0,0)^T} & S_{y,(1,0)^T} = z_{y,(1,0)^T} & \dots & S_{y,(n_x,0)^T} = z_{y,(n_x,0)^T} \\ S_{y,(0,1)^T} = z_{y,(0,1)^T} & S_{y,(1,1)^T} = z_{y,(1,1)^T} & \dots & S_{y,(n_x,1)^T} = z_{y,(n_x,1)^T} \\ \dots & \dots & \dots & \dots \\ S_{y,(0,n_y)^T} = z_{y,(0,n_y)^T} & S_{y,(1,n_y)^T} = z_{y,(1,n_y)^T} & \dots & S_{y,(n_x,n_y)^T} = z_{y,(n_x,n_y)^T} \end{pmatrix},$$

where

$$S_{y,\boldsymbol{\omega}} : \mathcal{N}(\mu_{y,\boldsymbol{\omega}}, \sigma^2_{y,\boldsymbol{\omega}}). \tag{6.8}$$

Magnitude Calculation
Now that the results of the convolutions, Z_x and Z_y have been determined as Gaussian distributed random variables, it is possible together with equation 6.3 to define the Cumulative Distribution Function (CDF) of the gradient magnitude random variable, $P_{Z_{2D}}(z_{2D})$ in terms of the PDFs for the two convolution results, $p_{Z_x}(z_x)$ and $p_{Z_y}(z_y)$[1]:

$$P_{Z_{2D}}(z_{2D}) = \iint\limits_{z_{2D} \geq \sqrt{z_x^2 + z_y^2}} p_{Z_x,Z_y}(z_x, z_y).dz_x.dz_y; \tag{6.9}$$

and assuming that z_x and z_y are independent results in[2]:

$$P_{Z_{2D}}(z_{2D}) = \iint\limits_{z_{2D} \geq \sqrt{z_x^2 + z_y^2}} p_{Z_x}(z_x).p_{Z_y}(z_y).dz_x.dz_y. \tag{6.10}$$

To perform this calculation Williamson *et al.* in [157] suggested that the result due to Rice in [111] was applicable. Rice derived the gradient magnitude of two Gaussian distributed random variables for application to modelling of noise characteristics in electrical circuits and utilised the following parameter values: $\mu_{x,\boldsymbol{\omega}} = 0$ and $\sigma^2_{x,\boldsymbol{\omega}} = \sigma^2_{y,\boldsymbol{\omega}}$.

Relating these parameter values to modelling of the image gradient magnitude allows for the x gradient mask being located in a homogeneous region of the image data under which all elements of I have approximately the same parameter values[3], while the y gradient mask is located either:

- Over a boundary or heterogeneous image region, $\mu_{y,\boldsymbol{\omega}} \neq 0$;
- Over a homogeneous image region , $\mu_{y,\boldsymbol{\omega}} = 0$.

Figure 6.3 helps to illustrate these conditions further. In all cases, $\sigma^2_{x,\boldsymbol{\omega}}$ and $\sigma^2_{y,\boldsymbol{\omega}}$ are assumed to be equal. For homogeneous image regions this assumption is true, but for heterogeneous image regions it might not be true. The assumption of equal variances greatly simplifies and makes possible the derivation of the approximate image gradient magnitude PDF.

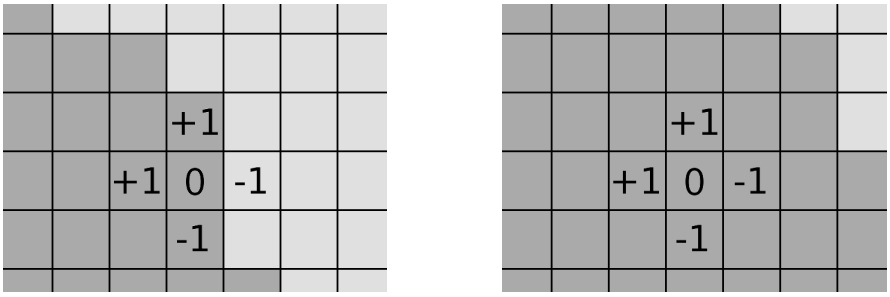
Therefore:

$$P_{Z_{2D}}(z_{2D}) = \iint\limits_{z_{2D} \geq \sqrt{z_x^2 + z_y^2}} \frac{1}{2.\pi.\sigma_z^2}. \exp\left(-\frac{1}{2.\sigma_z^2}(z_x^2 + (z_y - \mu_z)^2)\right) .dz_x.dz_y \tag{6.11}$$

[1]For detailed information about functions of random variables see for example [134]

[2]The two gradient convolution results, Z_x and Z_y, are probably not independently distributed particularly if large gradient convolution masks are used, but for the purposes of this work this assumption is deemed to be acceptable.

[3]Recall from the sum of a number of Gaussian distributed random variables (appendix C) that the resultant mean is given by the weighted sum of each of the means of the random variables. Also, each gradient mask has *weights* that sum to zero, thus reducing the mean of the resultant Gaussian to zero. This is not the case if the mask is over a heterogeneous region, resulting in non-zero mean values.



(a) Heterogeneous Case (b) Homogeneous Case

Figure 6.3: Illustration of the two possible conditions of the location of the gradient masks, M_x and M_y.

$$= \frac{1}{2.\pi.\sigma_z^2}. \iint\limits_{z_{2D} \geq \sqrt{z_x^2+z_y^2}} \exp\left(-\frac{1}{2.\sigma_z^2}(z_x^2 + z_y^2 - 2.z_y.\mu_z + \mu_z^2)\right).dz_x.dz_y. \tag{6.12}$$

where $\mu_z = \mu_{y,\boldsymbol{\omega}}$ and $\sigma_z = \sigma_{x,\boldsymbol{\omega}} = \sigma_{y,\boldsymbol{\omega}}$. A change of variables is necessary to remove the double integrals on the single function. Taking advantage of the circular symmetry of the function, where $z_{2D}^2 = z_x^2 + z_y^2$, $z_x = z_{2D}.\sin(\theta)$, $z_y = z_{2D}.\cos(\theta)$ and $dz_x.dz_y = z_{2D}.d\theta.dz_{2D}$, (for further information see for example [136]), so that:

$$P_{Z_{2D}}(z_{2D}) = \frac{1}{2.\pi.\sigma_z^2}. \int\limits_{-\infty}^{z_{2D}} \int\limits_{0}^{2.\pi} \exp\left(-\frac{z_{2D}^2 - 2.z_{2D}.\cos(\theta).\mu_z + \mu_z^2}{2.\sigma_z^2}\right).z_{2D}.d\theta.dz_{2D}, \tag{6.13}$$

$$= \frac{1}{2.\pi.\sigma_z^2}.\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right) \int\limits_{-\infty}^{z_{2D}} \exp\left(-\frac{z_{2D}^2}{2.\sigma_z^2}\right).z_{2D} \int\limits_{0}^{2.\pi} \exp\left(\frac{z_{2D}.\cos(\theta).\mu_z}{\sigma_z^2}\right).d\theta.dz_{2D}. \tag{6.14}$$

As the inner integral is of the same form as a special function, known as the modified Bessel function of the first kind, order zero, the inner integral can be replaced by using, [47],

$$\int\limits_{0}^{2.\pi} \exp\left(\frac{z_{2D}.\cos(\theta).\mu_z}{\sigma_z^2}\right).d\theta = 2.\pi.I_0\left(\frac{z_{2D}.\mu_z}{\sigma_z^2}\right), \tag{6.15}$$

the CDF then becomes:

$$P_{Z_{2D}}(z_{2D}) = \frac{1}{\sigma_z^2}.\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right) \int\limits_{-\infty}^{z_{2D}} \exp\left(-\frac{z_{2D}^2}{2.\sigma_z^2}\right).I_0\left(\frac{z_{2D}.\mu_z}{\sigma_z^2}\right).z_{2D}.dz_{2D}. \tag{6.16}$$

This CDF, $P_{Z_{2D}}(z_{2D})$ is now described in terms of the PDF of Z_{2D}:

$$P_{Z_{2D}}(z_{2D}) = \int\limits_{-\infty}^{z_{2D}} p_{Z_{2D}}(z_{2D}).dz_{2D}, \tag{6.17}$$

so the PDF of Z_{2D} can be obtained from the CDF by removing the integral:

$$p_{Z_{2D}}(z_{2D}) = \frac{z_{2D}}{\sigma_z^2}.\exp\left(-\frac{\mu_z^2 + z_{2D}^2}{2.\sigma_z^2}\right).I_0\left(\frac{z_{2D}.\mu_z}{\sigma_z^2}\right). \tag{6.18}$$

This PDF is known as the Rician PDF after Rice, [111], and is an approximate description of the behaviour of the gradient magnitude, Z_{2D}, of the image data, I.

For homogeneous image regions where $\mu_z = 0$, equation 6.18 reduces to:

$$p_{Z_{2Dpure}}(z_{2D}) = \frac{z_{2D}}{\sigma_z^2}.\exp\left(-\frac{z_{2D}^2}{2.\sigma_z^2}\right). \tag{6.19}$$

which is known as a Rayleigh PDF.

Equations 6.18 and 6.19 are valid for when $Z_x : \mathcal{N}(0, \sigma_z^2)$ and $Z_y : \mathcal{N}(\mu_y, \sigma_z^2)$, where $\mu_z = \mu_y$. Clearly other situations may occur, in particular when $Z_x : \mathcal{N}(\mu_x, \sigma_z^2)$. Appendix D demonstrates that when $\mu_x \neq 0$, equations 6.18 and 6.19 are still valid, where $\mu_z = \sqrt{\mu_x^2 + \mu_y^2}$.

Incorporation of the Gradient Magnitude into a PV Mixture Formulation

Initially abstract likelihood terms are considered for the intensity and gradient magnitude. Once these abstract likelihoods are considered, it will become possible to incorporate the gradient magnitude of equation 6.18 into the PV mixture model formulation that Williamson *et al.* utilised in [157].

If a joint intensity, g, and gradient magnitude, z_{2D}, likelihood, $p(g, z_{2D})$ is considered then from conditional dependence:

$$p(g, z_{2D}) = p(z_{2D}|g).p(g). \tag{6.20}$$

So this states that the intensity likelihood formulation can be incorporated without modification or inclusion of a gradient magnitude variable. But the gradient magnitude likelihood must be made dependent on an intensity variable.

Once Williamson *et al.* had established a suitable likelihood for the 2-D localised image gradient magnitude, (as given by equation 6.18), and the dependencies of the likelihoods, they adapted the mixture model of Vokurka *et al.*, [148], to include the gradient magnitude likelihood. The intensity based mixture model from [148] has already been discussed in chapter 5, section 5.1.2 and is referred to here as model B.

For convenience, the equation for the mixture model is repeated here (from equation 5.6):

$$p(g) = \sum_{v=1}^{K} p(g|\tau_v).P(\tau_v) + \sum_{v=1}^{K-1}\sum_{a=v+1}^{K} \left\{p_{GT}(g|\tau_{v,a}).P_{GT}(\tau_{v,a}) + p_{GT}(g|\tau_{a,v}).P_{GT}(\tau_{a,v})\right\}. \tag{6.21}$$

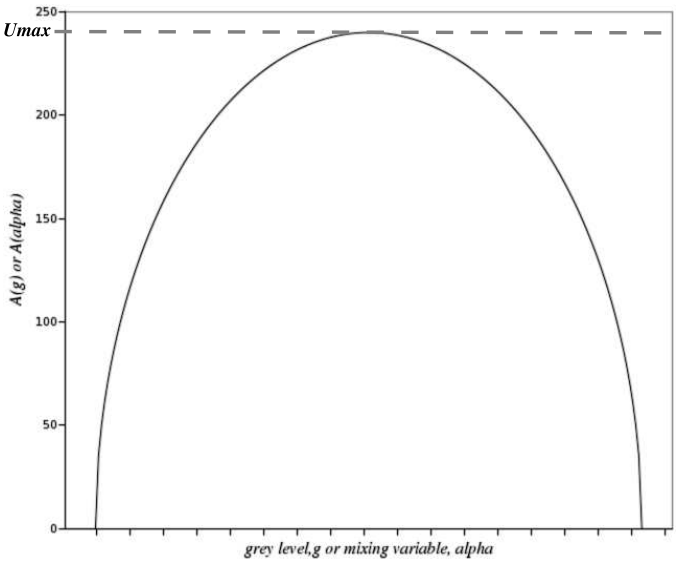


Figure 6.4: Illustration of the half ellipse function, $A(g|\tau_{v,a})$ or $A(\boldsymbol{\alpha}|\tau_{v,a})$ used in the gradient magnitude likelihoods of models D, E and F. This illustration can be compared with the scatter plot given in figure 6.2 as $A()$ is used to describe the location of the gradient magnitude as the PV mixtures vary from composition of one class to another. U_{max} is the maximum mean gradient magnitude and determines the height of the half ellipse.

Incorporating the gradient magnitude likelihood:

$$\begin{aligned} p(g, z_{2D}) = &\sum_{v=1}^{K} p(z_{2D}|g, \tau_v).p(g|\tau_v).P(\tau_v) \\ &+ \sum_{v=1}^{K-1} \sum_{a=v+1}^{K} \{p(z_{2D}|g, \tau_{v,a}).p_{GT}(g|\tau_{v,a}).P_{GT}(\tau_{v,a}) \\ &+ p(z_{2D}|g, \tau_{a,v}).p_{GT}(g|\tau_{a,v}).P_{GT}(\tau_{a,v})\} . \end{aligned} \tag{6.22}$$

Williamson *et al.* chose to modify the gradient magnitude likelihood by setting the gradient magnitude mean, $\mu_z = A(g|\tau_{v,a})$, where $A(g|\tau_{v,a})$ is a half ellipsoid function defined between the means, μ_v and μ_a, (see figure 6.4 and compare to figure 6.2) so that equation 6.18 becomes:

$$p(z_{2D}|g, \tau_{v,a}) = \frac{z_{2D}}{\sigma_z^2}. \exp\left(-\frac{A(g|\tau_{v,a})^2 + z_{2D}^2}{2.\sigma_z^2}\right) .I_0\left(\frac{z_{2D}.A(g|\tau_{v,a})}{\sigma_z^2}\right). \tag{6.23}$$

Clearly, for pure classification class regions of the image data, $A(g|\tau_{v,a}) \to 0$ as $g \to \mu_v$ or μ_a, and as seen by equation 6.19, the 2-D gradient magnitude for pure classification class image regions where $\mu_z = 0$, the gradient magnitude PDF is Rayleigh distributed. Therefore the gradient magnitude term of the first summation term in equation 6.22 corresponds to the Rayleigh PDF (equation 6.19) which is no longer dependent on g, so that:

$$p_{pure}(g, z_{2D}) = p_{pure}(g).p_{pure}(z_{2D}), \tag{6.24}$$

resulting in:

$$\begin{aligned} p(g, z_{2D}) = &\sum_{v=1}^{K} p(z_{2D}|\tau_v).p(g|\tau_v).P(\tau_v) \\ &+ \sum_{v=1}^{K-1} \sum_{a=v+1}^{K} \{p(z_{2D}|g, \tau_{v,a}).p_{GT}(g|\tau_{v,a}).P_{GT}(\tau_{v,a}) \\ &+ p(z_{2D}|g, \tau_{a,v}).p_{GT}(g|\tau_{a,v}).P_{GT}(\tau_{a,v})\} . \end{aligned} \tag{6.25}$$

So this is now a PV mixture model incorporating a localised intensity based gradient magnitude feature. From this, the posterior probability for a particular classification class can be formulated, following the similar formulation of equation 5.8 of model B (i.e. using Bayes theorem):

$$p(\tau_v|g, z_{2D}) = \frac{p(g|\tau_v).p(z_{2D}|\tau_v).P(\tau_v) + \sum_{\forall j \neq i} p_{GT}(g|\tau_{v,a}).p(z_{2D}|g, \tau_{v,a}).P_{GT}(\tau_{v,a})}{p(g, z_{2D})}. \tag{6.26}$$

This equation expresses the probability of obtaining a particular classification class, τ_v, given an intensity value and a gradient magnitude value. As for model B, the voxel mixture composition is again taken by analogy to the probabilities obtained from these *posterior probabilities*.

6.1.2 Model E: Reformulated 2-D Gradient Magnitude PV Classifier

Following the model progressions discussed in the previous chapter, 5, it became apparent that model B was in fact approximately equivalent to model C. However, the formulation in model C explicitly modelled the PV composition, rather than by analogy as in models B and D. The same progression in model development is now proposed for model D to create model E. Model E is mixture model C incorporating a localised image based gradient magnitude feature.

Recall from section 5.1.3, equation 5.16 described the variation of intensities, g, given a particular mixture vector, $\boldsymbol{\alpha}$, repeated here for convenience:

$$p(g|\boldsymbol{\alpha}) = \frac{1}{\sqrt{2.\pi.\sigma_\alpha^2}}. \exp\left(-\frac{(g-\mu_\alpha)^2}{2.\sigma_\alpha^2}\right). \tag{6.27}$$

Also recall from this chapter the conditional dependence of the intensity and gradient magnitudes in equation 6.20, but they are now modified to include the mixing parameter, $\boldsymbol{\alpha}$:

$$p(g, z_{2D}|\boldsymbol{\alpha}) = p(z_{2D}|g, \boldsymbol{\alpha}).p(g|\boldsymbol{\alpha}). \tag{6.28}$$

The intensity likelihood is again not dependent on the gradient magnitude. But for this formulation, the gradient magnitude likelihood is dependent on the intensity *and* the mixing vector, $\boldsymbol{\alpha}$. However, the mixing parameter can now take the role that g took in the half ellipsoid function, $A(g|\tau_{v,a})$. This can be understood from the fact that the half ellipsoid was initially proposed as a model of the gradient magnitude centering

parameter, μ_z. The dependence on the intensity to determine this parameter results in a direct dependence on the noise that is also associated with the voxel intensities. This direct dependence is removed with the use of the mixing parameter vector, $\boldsymbol{\alpha}$, thus modifying equation 6.23 to:

$$p(z_{2D}|\boldsymbol{\alpha}, \tau_{v,a}) = \frac{z_{2D}}{\sigma_z^2}. \exp\left(-\frac{A(\boldsymbol{\alpha}|\tau_{v,a})^2 + z_{2D}^2}{2.\sigma_z^2}\right) .I_0\left(\frac{z_{2D}.A(\boldsymbol{\alpha}|\tau_{v,a})}{\sigma_z^2}\right). \tag{6.29}$$

The gradient magnitude is therefore no longer dependent on the intensity feature, g:

$$p(g, z_{2D}|\boldsymbol{\alpha}) = p(g|\boldsymbol{\alpha}).p(z_{2D}|\boldsymbol{\alpha}). \tag{6.30}$$

The introduction of the mixing vector, $\boldsymbol{\alpha}$ thus provides a further useful role. It not only incorporates a conceptually useful variable which can be related to the actual underlying physical processes, such as voxel content, but it also removes the conditional dependence of the likelihood terms. This therefore reduces the direct dependence on the intensity variable, g and consequently the effect of system noise in this gradient magnitude term.

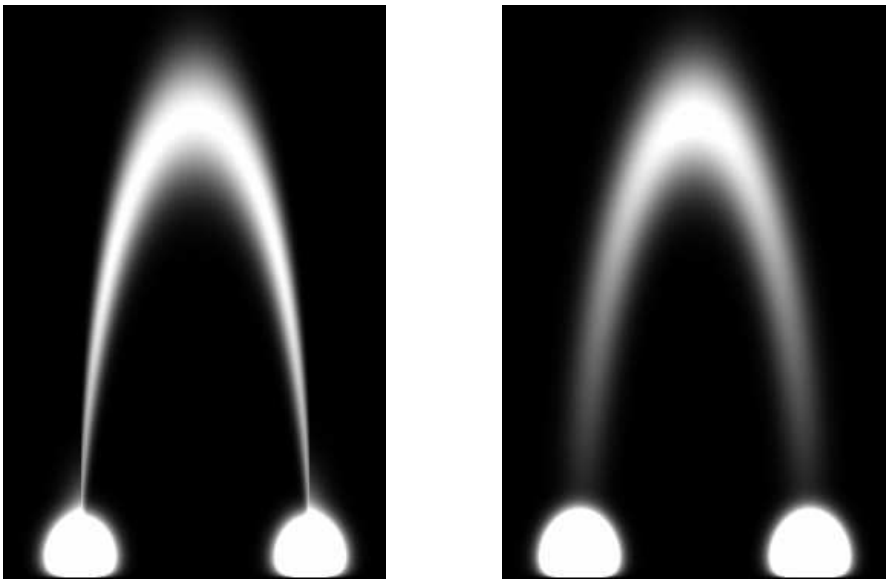
Inference from the posterior, $p(\boldsymbol{\alpha}|g, z_{2D})$, follows the same formulation and prior, $p(\boldsymbol{\alpha})$, as for model C (see chapter 5), utilising the expected posterior:

$$\begin{aligned}\mathbf{E}\left[\alpha_v|g, z_{2D}\right] = \\ &\frac{1}{p(g, z_{2D})}.p(g|\tau_v).p(z_{2D}|\tau_v).P(\tau_v) \\ &+\frac{1}{p(g, z_{2D})}.\sum_{\forall j \neq i}\int_0^1 \alpha_v.p(g|\boldsymbol{\alpha}).p(z_{2D}|\boldsymbol{\alpha}, \tau_{v,a}).p(\alpha_v, \alpha_a).P(\tau_{v,a}).d\alpha_v.\end{aligned} \tag{6.31}$$

The use of $A(g)$ in the preceding formulation results in the gradient feature being directly dependent on the intensity feature, g. This dependency affects the inference of the PV content. $A(\boldsymbol{\alpha})$ removes this dependency, therefore allowing the posterior mean to be (correctly) evaluated over $\boldsymbol{\alpha}$. Possible alternatives may have included $A(\bar{g})$, where $\bar{g}$ is a locally calculated mean intensity value, but this would be attempting to estimate the appropriate value of $\mu_{\boldsymbol{\alpha}}$. In short, g is a crude estimate of $\bar{g}$, which is also a crude estimate of the optimal value of $\mu_{\boldsymbol{\alpha}}$. Neither g nor $\bar{g}$ is a function of the mixing parameter, $\boldsymbol{\alpha}$. Therefore, this new reformulation of the intensity gradient magnitude model is not approximately equivalent to the preceding formulation (model D). This is in contrast to the approximate equivalence of the intensity based models (under particular model and data conditions), previously discussed in chapter 5, for models B and C.

Qualitative Model Based Feature Space Examination

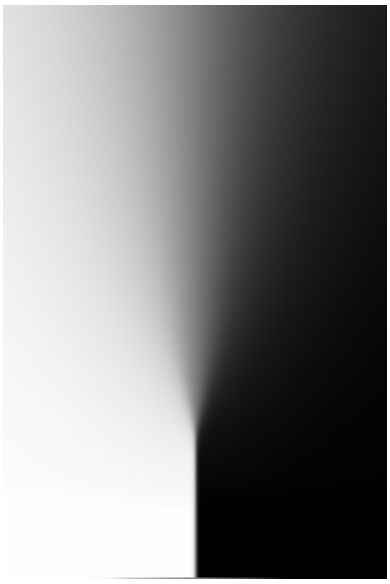
The significance of this new formulation at first may not seem to be of particular importance. Figure 6.5 illustrates exemplar feature spaces and decision surfaces described by model D and the new model E. As can be seen from these two illustrations, the two



(a) Model D: Feature Space

(b) Model E: Feature Space

(c) Model D: Decision Surface



(d) Model E: Decision Surface

Figure 6.5: Comparison of feature spaces and their associated decision surfaces for intensity and localised image gradient magnitude PV models D and E. For each space or surface the horizontal axes represents intensity and the vertical axes represent localised image gradient magnitude. For the feature spaces, the brightness of a point in pictures (a) and (b) represents the relative strength of the combined marginal probability, $p(g, z_{2D})$. For the decision surfaces, the brightness of a point in pictures (c) and (d) represents the posterior probabilities for the left most classification class, where white represents probability 1 and black, probability 0.

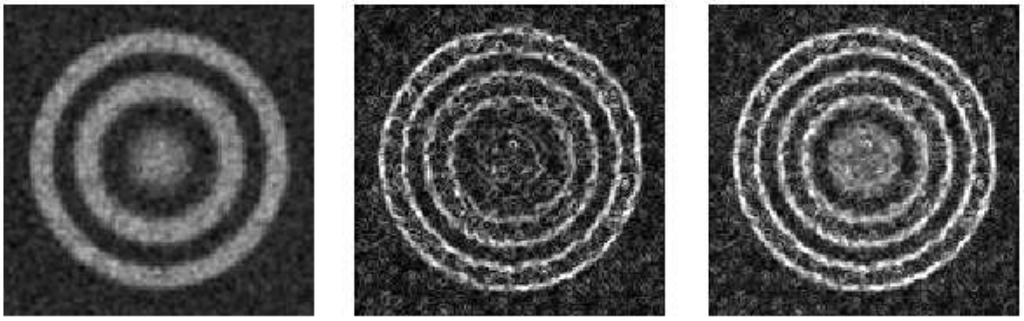


Figure 6.6: Illustration of benefit of using 3-D gradient magnitude over 2-D gradient magnitude for the detection of PV voxels whose component lie mainly in the z direction. The first image corresponds to an exemplar slice through a simulated 3-D PV data set, while the second image corresponds to the same slice but through a data volume consisting of the result of a 2-D gradient magnitude operation over the data in the first slice. Similarly for the third image slice, except a 3-D gradient operator was used. The third and second image slices illustrate for the central region how the partially clipped spheroid in the z direction has been detected using a 3-D gradient magnitude kernel, where as the result of using a 2-D gradient magnitude kernel has failed to detect this partially clipped spheroid.

feature spaces and consequently the decision surfaces for the same exemplar parameters are quite different. The new formulation appears to provide a smoother decision surface, where no hard borders differentiate between a PV voxel and a pure voxel for particular pairs of intensity and gradient magnitude values. Both models provide a rapid mid-point change between pure classification classes for low gradient magnitude measurements; this is similarly so for the transition between pure and PV voxels for model D at greater gradient magnitude values. But for model E, if a voxel has a higher gradient magnitude value then it gradually becomes more and more likely to be generated by a PV classification class.

6.1.3 Model F: 3-D Gradient Magnitude PV Classifier

Williamson *et al.* in 2002, [157], developed their PV classifier utilising a 2-D gradient magnitude measure and consequently a probabilistic description of 2-D gradient magnitude was utilised in their likelihood model. An improvement to this model can be made by observing that voxels are inherently 3-D structures due to the image acquisition processes of the biomedical imaging devices and the domain from which the data is acquired. Therefore, it was thought that a 3-D gradient magnitude measure would model the data more accurately. The result of a 2-D gradient magnitude filtering operation is compared with a 3-D gradient magnitude filtering operation, both over 3-D simulated data in figure 6.6.

Model F takes the reformulated gradient magnitude PV classifier (model E) and replaces the 2-D gradient magnitude likelihood with a 3-D gradient magnitude likelihood.

It should be noted that biomedical imaging data, particularly MRI data, is often acquired with slice thicknesses and inter-slice gaps that may be greater than the in-plane voxel dimensions. The 3-D model described here may therefore become less valid if the distance between each slice is very much greater than the in-plane voxel dimensions, particularly if the interslice gap occupies the majority of the inter-plane distance. Nevertheless, the classification performance should become no worse than the performance of the intensity only models of the previous chapter (5).

Derivation of a 3-D Gradient Magnitude Likelihood

The process of calculating the 3-D gradient magnitude of imaging data consists of the following steps (compare to section 6.1.1, equations 6.1, 6.2 and 6.3):

1. Convolve imaging data, I, with gradient masks, M_x, M_y and M_z for the x, y and z data dimensions, creating gradient channels, Z_x, Z_y and Z_z:

$$Z_x = I * M_x; \tag{6.32}$$

$$Z_y = I * M_y; \tag{6.33}$$

$$Z_z = I * M_z. \tag{6.34}$$

2. Calculate magnitude, Z_{3D} of the three gradient channels, Z_x, Z_y and Z_z:

$$Z_{3D} = \sqrt{Z_x^2 + Z_y^2 + Z_z^2}. \tag{6.35}$$

Convolving the 3-D Image Data
Similar to section 6.1.1, the result of the convolutions, Z_x, Z_y and Z_z are 3-D image matrices whose elements, $S_{x,\boldsymbol{\omega}}$, $S_{y,\boldsymbol{\omega}}$ and $S_{z,\boldsymbol{\omega}}$ are sums of Gaussian distributed random variables which are also Gaussian distributed(from appendix C), so that:

$$S_{x,\boldsymbol{\omega}} : \mathcal{N}(\mu_{x,\boldsymbol{\omega}}, \sigma^2_{x,\boldsymbol{\omega}});$$

$$S_{y,\boldsymbol{\omega}} : \mathcal{N}(\mu_{y,\boldsymbol{\omega}}, \sigma^2_{y,\boldsymbol{\omega}});$$

$$S_{z,\boldsymbol{\omega}} : \mathcal{N}(\mu_{z,\boldsymbol{\omega}}, \sigma^2_{z,\boldsymbol{\omega}}).$$

3-D Magnitude Calculation
This section presents a derivation of the PDF, $p_{Z_{3D}}(z_{3D})$, of the 3-D gradient magnitude random variable, Z_{3D} (equation 6.35).

The CDF of Z_{3D}, $P_{Z_{3D}}(z_{3D})$ is given by

$$P_{Z_{3D}}(z_{3D}) = \iiint\limits_{z_{3D} \geq \sqrt{z_x^2+z_y^2+z_z^2}} p_{Z_x,Z_y,Z_z}(z_x, z_y, z_z).dz_x.dz_y.dz_z; \tag{6.36}$$

and assuming independence of each gradient magnitude component

$$P_{Z_{3D}}(z_{3D}) = \iiint\limits_{z_{3D} \geq \sqrt{z_x^2+z_y^2+z_z^2}} p_{Z_x}(z_x).p_{Z_y}(z_y).p_{Z_z}(z_z).dz_x.dz_y.dz_z. \tag{6.37}$$

Now consider the case where $Z_x, Z_y : \mathcal{N}(0, \sigma_z^2)$ and $Z_z : \mathcal{N}(\mu_z, \sigma_z^2)$, this signifies the situation where the kernel masks M_x and M_y are covering homogeneous regions of the volumetric data and M_z covers a heterogeneous region of the volumetric data. This is very similar to the scenario described in section 6.1.1 for model D, and illustrated in figure 6.3, except model E is applicable to 3-D volumetric data.

For these parameter values, the 3-D gradient magnitude CDF in equation 6.37 becomes

$$P_{Z_{3D}}(z_{3D}) = \iiint\limits_{z_{3D} \geq \sqrt{z_x^2+z_y^2+z_z^2}} \frac{1}{(2.\pi)^{3/2}.\sigma_z^3} \exp\left(-\frac{z_x^2 + z_y^2 + (z_z - \mu_z)^2}{2.\sigma_z^2}\right) .dz_x.dz_y.dz_z$$
$$= \iiint\limits_{z_{3D} \geq \sqrt{z_x^2+z_y^2+z_z^2}} \frac{1}{(2.\pi)^{3/2}.\sigma_z^3} \exp\left(-\frac{z_x^2 + z_y^2 + z_z^2 - 2.z_z.\mu_z + \mu_z^2}{2.\sigma_z^2}\right) .dz_x.dz_y.dz_z. \quad (6.38)$$

Utilising a change of variables, from a Cartesian coordinate system to a spherical coordinate system where $z_{3D}^2 = z_x^2 + z_y^2 + z_z^2$, $z_z = z_{3D}.\cos(\phi)$ and $dz_x.dz_y.dz_z \rightarrow z_{3D}^2.\sin(\phi).d\theta.d\phi.dz_{3D}$ (see for example [136]):

$$P_{Z_{3D}}(z_{3D}) = \frac{1}{(2.\pi)^{3/2}.\sigma_z^3} \int\limits_{-\infty}^{z_{3D}} \int\limits_{0}^{\pi} \int\limits_{0}^{2.\pi} \exp\left(-\frac{z_{3D}^2 - 2.\mu_z.z_{3D}.\cos(\phi) + \mu_z^2}{2.\sigma_z^2}\right) .z_{3D}^2.\sin(\phi).d\theta.d\phi.dz_{3D}. \quad (6.39)$$

The variables of integration have now been separated out. Rearranging for the dependent and non-dependent variables

$$P_{Z_{3D}}(z_{3D}) = \frac{\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right)}{(2.\pi)^{3/2}.\sigma_z^3} \int\limits_{-\infty}^{z_{3D}} \int\limits_{0}^{\pi} \exp\left(-\frac{z_{3D}^2 - 2.\mu_z.z_{3D}.\cos(\phi)}{2.\sigma_z^2}\right) .z_{3D}^2.\sin(\phi). \int\limits_{0}^{2.\pi} .d\theta.d\phi.dz_{3D} \quad (6.40)$$

$$= \frac{\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right)}{(2.\pi)^{3/2}.\sigma_z^3} \int\limits_{-\infty}^{z_{3D}} \exp\left(-\frac{z_{3D}^2}{2.\sigma_z^2}\right) .z_{3D}^2 \int\limits_{0}^{\pi} \exp\left(\frac{\mu_z.z_{3D}.\cos(\phi)}{\sigma_z^2}\right) .\sin(\phi) \int\limits_{0}^{2.\pi} d\theta.d\phi.dz_{3D},$$

thus simplifying the integrands. Consider the inner most integral which can be evaluated as $\int\limits_{0}^{2.\pi} .d\theta = 2.\pi$, resulting in:

$$= \frac{2.\pi.\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right)}{(2.\pi)^{3/2}.\sigma_z^3} \int\limits_{-\infty}^{z_{3D}} \exp\left(-\frac{z_{3D}^2}{2.\sigma_z^2}\right) .z_{3D}^2 \int\limits_{0}^{\pi} \exp\left(\frac{\mu_z.z_{3D}.\cos(\phi)}{\sigma_z^2}\right) .\sin(\phi).d\phi.dz_{3D}$$

$$= \frac{\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right)}{(2.\pi)^{1/2}.\sigma_z^3} \int_{-\infty}^{z_{3D}} \exp\left(-\frac{z_{3D}^2}{2.\sigma_z^2}\right).z_{3D}^2 \int_0^{\pi} \exp\left(\frac{\mu_z.z_{3D}.\cos(\phi)}{\sigma_z^2}\right).\sin(\phi).d\phi.dz_{3D}. \tag{6.41}$$

Now if the next inner most integral is considered

$$\int_0^{\pi} \exp\left(\frac{\mu_z.z_{3D}.\cos(\phi)}{\sigma_z^2}\right).\sin(\phi)d\phi, \tag{6.42}$$

then a result from Gradshteyn and Ryzhik [47] can be utilised:

$$I_v(a) = \frac{\left(\frac{a}{2}\right)^v}{\Gamma(v+1/2).\Gamma(1/2)} \int_0^{\pi} \exp(\pm a.\cos(\zeta)).\sin^{2.v}(\zeta).d\zeta, \tag{6.43}$$

where $\Gamma()$ is the Gamma function and $I_v(a)$ is a modified Bessel function of order v, where v maybe non-integer. Letting $v = 1/2$, $a = \mu_z.z_{3D}/\sigma_z^2$ and $\zeta = \phi$, for equation 6.42 results in:

$$\int_0^{\pi} \exp\left(\frac{\mu_z.z_{3D}}{\sigma_z^2}.\cos(\phi)\right).\sin(\phi).d\phi = \frac{\Gamma(1).\Gamma(1/2)}{\left(\frac{\mu_z.z_{3D}}{2.\sigma_z^2}\right)^{\frac{1}{2}}}.I_{\frac{1}{2}}\left(\frac{\mu_z.z_{3D}}{\sigma_z^2}\right), \tag{6.44}$$

and as $\Gamma(1) = 1$ and $\Gamma(1/2) = \sqrt{\pi}$:

$$\int_0^{\pi} \exp\left(\frac{\mu_z.z_{3D}}{\sigma_z^2}.\cos(\phi)\right).\sin(\phi).d\phi = \left(\frac{2.\pi.\sigma_z^2}{\mu_z.z_{3D}}\right)^{\frac{1}{2}}.I_{\frac{1}{2}}\left(\frac{\mu_z.z_{3D}}{\sigma_z^2}\right). \tag{6.45}$$

Placing this result into the 3-D gradient magnitude CDF, of equation 6.41:

$$P_{Z_{3D}}(z_{3D}) = \frac{\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right)}{(2.\pi)^{1/2}.\sigma_z^3} \int_{-\infty}^{z_{3D}} \exp\left(-\frac{z_{3D}^2}{2.\sigma_z^2}\right).z_{3D}^2 \left(\frac{2.\pi.\sigma_z^2}{\mu_z.z_{3D}}\right)^{\frac{1}{2}}.I_{\frac{1}{2}}\left(\frac{\mu_z.z_{3D}}{\sigma_z^2}\right).dz_{3D} \tag{6.46}$$

$$= \frac{\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right)}{\sigma_z^2.\sqrt{\mu_z}} \int_{-\infty}^{z_{3D}} \exp\left(-\frac{z_{3D}^2}{2.\sigma_z^2}\right).z_{3D}^{3/2} I_{\frac{1}{2}}\left(\frac{\mu_z.z_{3D}}{\sigma_z^2}\right).dz_{3D}.$$

This equation now describes the CDF, $P_{Z_{3D}}(z_{3D})$ in terms of the PDF, $p_{Z_{3D}}(z_{3D})$:

$$P_{Z_{3D}}(z_{3D}) = \int_{-\infty}^{z_{3D}} p_{Z_{3D}}(z_{3D}).dz_{3D}, \tag{6.47}$$

so the PDF of Z_{3D} can be obtained from the CDF by removing the integral:

$$p_{Z_{3D}}(z_{3D}) = \frac{\exp\left(-\frac{\mu_z^2}{2.\sigma_z^2}\right)}{\sigma_z^2.\sqrt{\mu_z}} \exp\left(-\frac{z_{3D}^2}{2.\sigma_z^2}\right).z_{3D}^{3/2} I_{\frac{1}{2}}\left(\frac{\mu_z.z_{3D}}{\sigma_z^2}\right). \tag{6.48}$$

This equation describes the PDF of the 3-D gradient magnitude for Gaussian distributed data for heterogeneous volumetric image regions. This PDF can be extended to the scenario where each gradient kernel is covering a heterogeneous region, (as is done for the 2-D case in appendix D), by replacing μ_z with $U = \sqrt{\mu_x^2 + \mu_y^2 + \mu_z^2}$, resulting in:

$$p_{Z_{3D}}(z_{3D}) = \frac{\exp\left(-\frac{U^2}{2.\sigma_z^2}\right)}{\sigma_z^2.\sqrt{U}} \exp\left(-\frac{z_{3D}^2}{2.\sigma_z^2}\right).z_{3D}^{3/2} I_{\frac{1}{2}}\left(\frac{U.z_{3D}}{\sigma_z^2}\right). \tag{6.49}$$

This is confirmed by the results obtained by different authors with different derivations for disparate applications in [1, 164].

Equation 6.49 and the 2-D gradient magnitude Rician density (equation 6.18) are plotted with data points from simulated PV data in figure 6.8.

The 2-D gradient magnitude density of equation 6.18 for homogeneous image regions, $\lim_{U\to 0} p_{Z_{2D}}(z)$, tends towards a Rayleigh density (equation 6.19). However, for the 3-D gradient magnitude, equation 6.49, the limiting density as $\lim_{U\to 0} p_{Z_{3D}}(z_{3D})$ is not quite as obvious. If $U = 0$, then following a similar set of steps outlined by equations 6.37 to 6.48 results in the Maxwell density:

$$p_{Z_{3Dpure}}(z_{3D}) = \frac{2.z_{3D}^2}{\sigma^3\sqrt{2.\pi}}. \exp\left(-\frac{z_{3D}^2}{2.\sigma^2}\right). \tag{6.50}$$

This Maxwell density together with a Rayleigh density (equation 6.19) and data points from simulated PV data are illustrated in figure 6.7.

Incorporation of the Gradient Magnitude into a PV Mixture Formulation

The 3-D gradient magnitude density described by equation 6.49 can now be incorporated into a model of the PV effect. The formulation described by model E is adapted to create a new 3-D formulation. Little adaptation is required, except to replace the 2-D gradient magnitude likelihoods with 3-D gradient magnitude likelihoods.

The same arguments that were used to adapt the 2-D gradient magnitude likelihood to utilise the mixture vector, $\boldsymbol{\alpha}$ resulting in equation 6.29 can be similarly used for the 3-D gradient magnitude likelihood. So that the 3-D gradient magnitude PV likelihood (from equation 6.49) is given by

$$p_{3D}(z_{3D}|\boldsymbol{\alpha}, \tau_{v,a}) = \frac{z_{3D}^{3/2}}{\sigma_z^2.\sqrt{A(\boldsymbol{\alpha}|\tau_{v,a})}} \exp\left(-\frac{A(\boldsymbol{\alpha}|\tau_{v,a})^2 + z_{3D}^2}{2.\sigma_z^2}\right).I_{\frac{1}{2}}\left(\frac{A(\boldsymbol{\alpha}|\tau_{v,a}).z_{3D}}{\sigma_z^2}\right). \tag{6.51}$$

and using the arguments given in section 6.1.2, the inference from the posterior, $p_{3D}(\boldsymbol{\alpha}|g, z_{3D})$, (see equation 6.31), follows the same formulation and utilises the same prior, $p(\boldsymbol{\alpha})$, as for model C (see chapter 5). Thus, the point estimate, utilising the

posterior mean for the 3-D gradient magnitude model is given by:

$$\begin{aligned}\mathbf{E}\left[\alpha_v|g, z_{3D}\right] = \\ &\frac{1}{p(g, z_{3D})}.p(g|\tau_v).p_{3Dpure}(z_{3D}|\tau_v).P(\tau_v) \\ &+\frac{1}{p(g, z_{3D})}.\sum_{\forall j\neq i}\int_0^1 \alpha_v.p(g|\boldsymbol{\alpha}).p_{3D}(z_{3D}|\boldsymbol{\alpha}, \tau_{v,a}).p(\alpha_v, \alpha_a).P(\tau_{v,a}).d\alpha_v. \quad (6.52)\end{aligned}$$

So this PV model, model F, should provide improved modelling of the PV effect as it not only utilises the 3-D gradient magnitude (in comparison to the 2-D gradient magnitude of model D), but it also utilises the amended feature space of model E. It is hoped that these developments will provide improved classification performance over the PV intensity models of chapter 5.

6.2 Experiments, Results and Discussions

This chapter has described possible improvements to the basic intensity based PV models of the previous chapter. These improvements utilise gradient magnitude to identify voxels that are more likely to be a mixture of two components. This section presents a series of experiments utilising simulated data as was done for the previous chapter in order to assess the performances of the novel model developments utilising gradient magnitude under controlled conditions.

The first set of experiments applies the models to isotropic simulated PV data, where the inter-slice spacing is the same as the inter-voxel spacing in plane.The second set of experiments applies the models to anisotropic simulated PV data, where the inter-slice spacing is greater than the inter-voxel spacing in plane. These two sets of experiments imitate the scenarios that are often encountered with biomedical imaging data. Some biomedical imaging data is acquired in an isotropic format but it is often acquired in an anisotropic format. Anisotropic data helps to increase the SNR without reducing the in-plane resolution, which unfortunately decreases the inter-plane resolution.

6.2.1 Classification of Isotropic Simulated PV Data

The experiments in this section utilise the intensity and localised gradient magnitude models (D, E and F) within PV classifiers to classify two classification class simulated PV data.

These simulations and their respective classifications with the PV classifiers utilising models D, E and F allow the performances of the PV classifiers to be assessed under controlled conditions where the parameters of the simulated data are known *a priori*.

Table 6.1: Summary of the two class simulations' parameters.

CNR	**Mean Distance**	**Mean 1**	**Mean 2**	**Stan. Dev. 1**	**Stan. Dev. 2**	**No Bands**
50	1000	500	1500	20	20	2
40	800	600	1400	20	20	2
30	600	700	1300	20	20	2
20	400	800	1200	20	20	2
10	200	900	1100	20	20	2
9	180	910	1090	20	20	2
8	160	920	1080	20	20	2
7	140	930	1070	20	20	2
6	120	940	1060	20	20	2
5	100	950	1050	20	20	2
4	80	960	1040	20	20	2
3	60	970	1030	20	20	2
2	40	980	1020	20	20	2

Experimental Procedure

The simulated PV data that was generated for the experiments in the previous chapter (5) utilised a symmetric 3-D Gaussian PSF, i.e. the width of the PSF was the same in all directions. The data was subsequently isotropically sampled so that each voxel possessed isotropic sampling ratios (1:1:1). This means that the simulated PV data could again be used in this set of experiments to assess the performances of the classifiers for isotropically sampled simulated PV data. The use of the same simulated PV data will also enable the performances of the basic intensity based models of chapter 5 to be compared with the models of this chapter under the same controlled conditions.

The respective parameters are repeated for convenience in table 6.1.

The intensity based parameters for each simulation were previously determined from either the simulation parameters or for the case of the prior terms, $P(\tau_v)$ and $P(\tau_{v,a})$ from the ground truth data. The parameters of the gradient magnitude models, i.e.

- The maximum gradient magnitude centralisation parameter, U_{max}, that is used in the calculation of $A()$ in equations 6.23, 6.29 and 6.51 (for models D,E and F respectively). U_{max} determines the maximum height of the half ellipse function, calculated with $A()$ as illustrated in figure 6.4;
- The width parameters for pure, σ_z^{pure} and PV, σ_z^{pv} regions, used in the calculation of equations 6.19 (2-D pure) and 6.23 (2-D PV).

The maximum gradient magnitude centralisation parameter, U_{max} can be determined by calculating the mean gradient magnitude for voxels composed of exactly 50% of each classification component.

Exemplar plots of the PV gradient magnitude equations, given by equations 6.18, (2-D) and 6.49, (3-D) can be seen in figures 6.8(a) and 6.8(b). These figures illustrate that

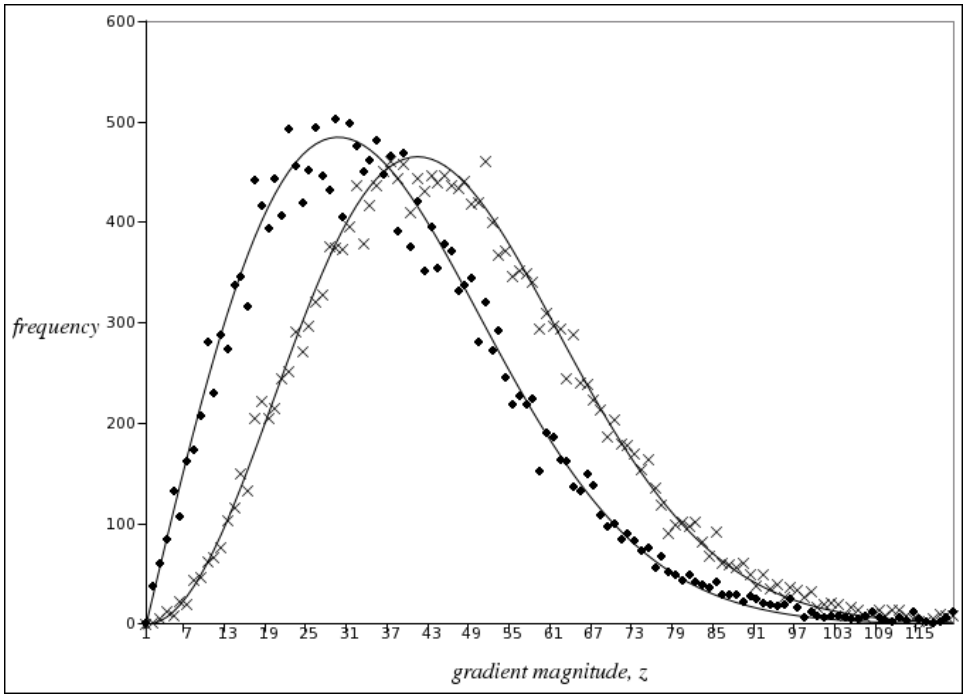
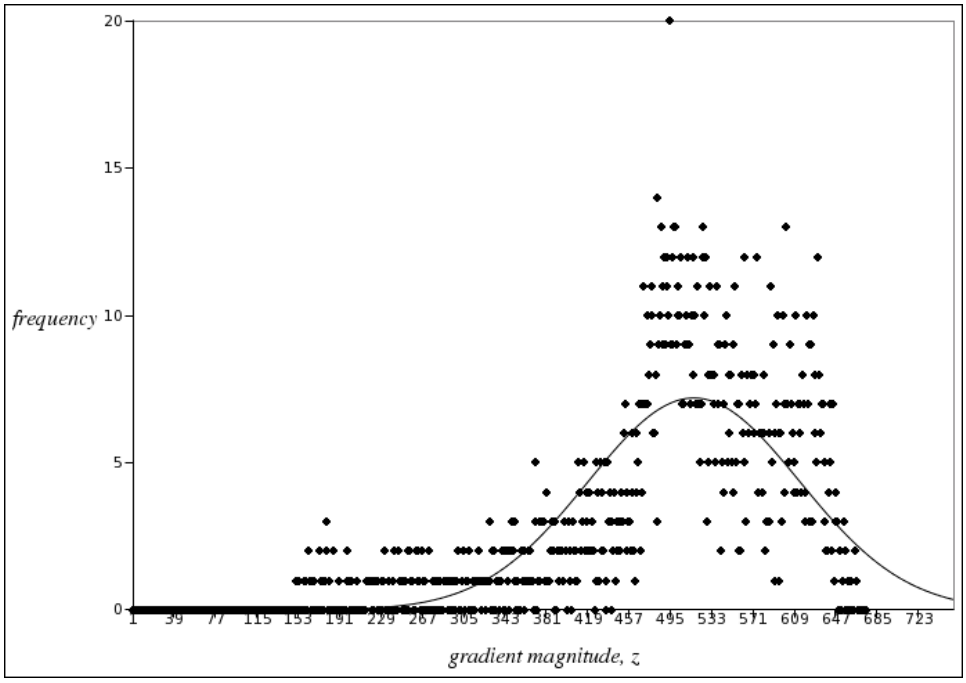


Figure 6.7: 2-D (•) and 3-D (×) gradient magnitude histogram data points from simulated data for pure voxels plotted along with the model PDFs of equations 6.19 (2-D) and 6.50 (3-D).

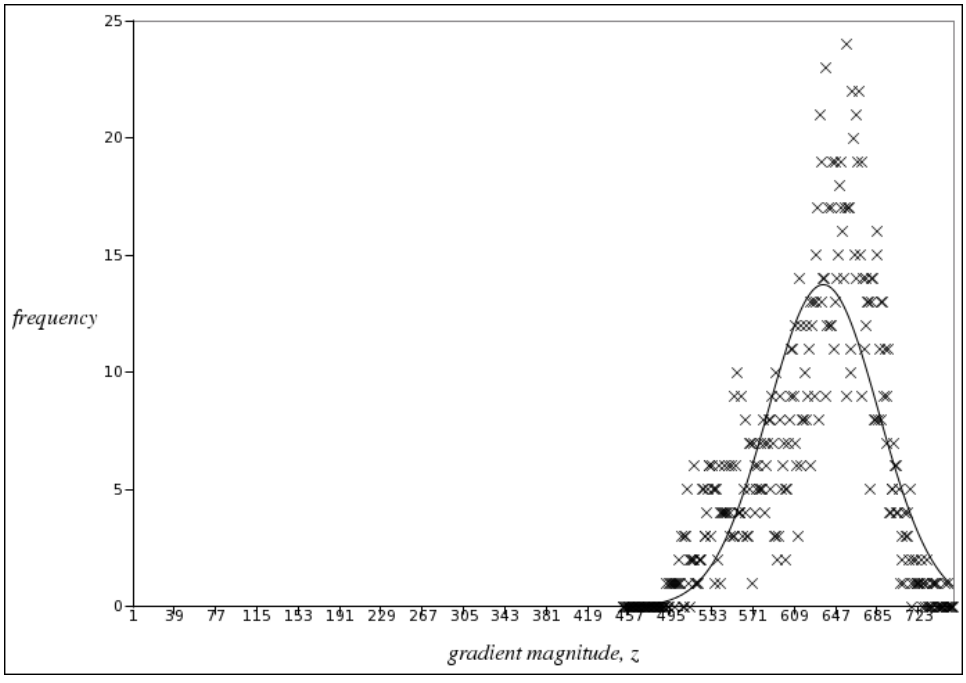
the densities are located correctly across the gradient magnitude data points with the calculated value of U_{max}. The width of these densities are discussed shortly.

The width parameter for pure gradient magnitude regions, σ_z^{pure} is available analytically. For this, the gradient mask calculation is calculated over pure image regions, where each voxels' intensity value is generated from an identically and independently distributed Gaussian random variable with known and equal standard deviations. Also, as the gradient mask elements are also known, the elements form weights in the calculation of the gradient magnitude for each direction and can be used in the calculation of σ_z^{pure} (see appendix C). Equations 6.19 (2-D) and 6.50 (3-D) are plotted using the analytically calculated width parameter along with exemplar data profiles from the mixture gradient magnitude scatter plot in figure 6.7. The 2-D model and 3-D model curves appear to follow the respective data points with good proximity. Also the 2-D model and 3-D model curves appear to share very similar shapes. The small dissimilarity would present very little difference in a resultant posterior probability calculation if one function were used in the place of the other. If this were the case then the gradient magnitude parameters would have to be estimated rather than being derived directly from the intensity domain parameters.

The width parameter for PV gradient magnitude regions could not be calculated analytically due to the discrete nature of the data and the large number of possible PV configurations that the gradient mask may cover. However, this does not prevent the width parameter of the high gradient magnitude densities described by equations 6.18

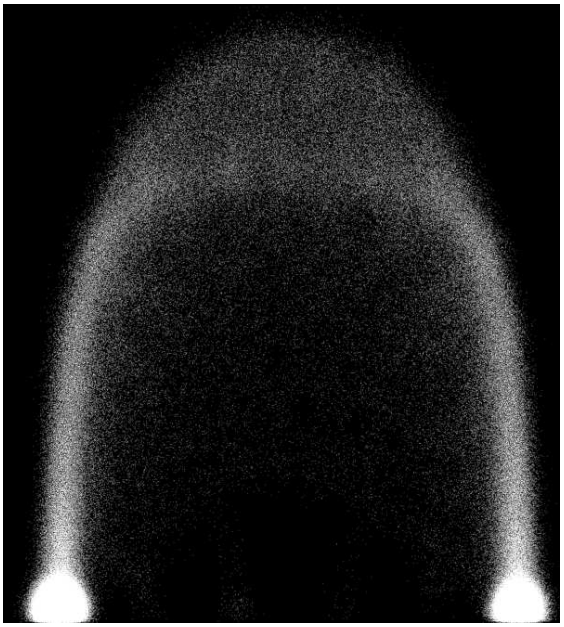


(a)

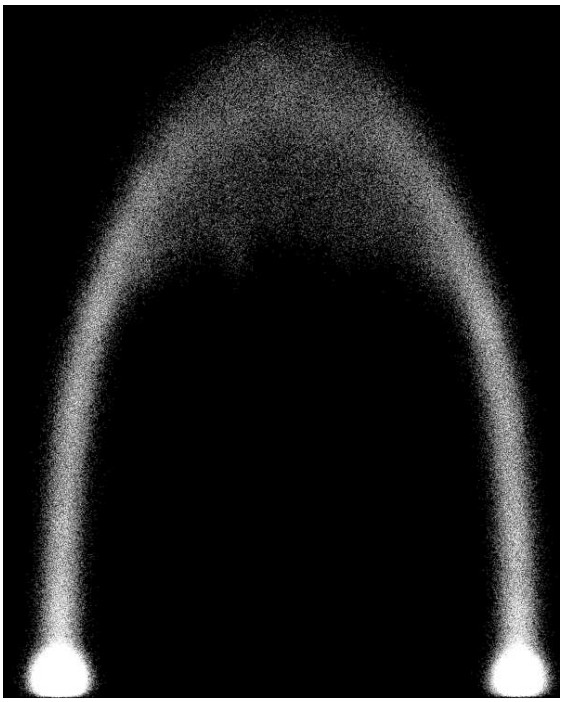


(b)

Figure 6.8: 2-D (•) and 3-D (×) gradient magnitude data points from simulated data for a given mean value plotted along with the model PDFs of equations 6.18 and 6.49.



(a) 2-D Gradient Magnitude Feature Space



(b) 3-D Gradient Magnitude Feature Space

Figure 6.9: Comparison of intensity (x-axis) versus (a)2-D and (b)3-D gradient magnitude feature spaces of 3-D simulated PV data. These scatter plots help to illustrate the benefit of using a 3-D gradient magnitude kernel over a 2-D gradient magnitude kernel for 3-D simulated PV data. The 2-D feature space has many low gradient magnitude PV data points located between the two pure components. This can be compared with the 3-D feature space where PV data points between the two pure components have high gradient magnitude values, consistent with the proposed model E.

and 6.49 from being calculated using the ground truth data. Equations 6.18 and 6.49 both tend towards Gaussian densities for high SNR values, as noted in [1]. The standard deviation of gradient magnitude data points for particular mixture combinations were calculated. These standard deviations were then used as the width parameters for equations 6.18 and 6.49. As can be seen from figures 6.8(a) and 6.8(b), these figures illustrate that the 3-D mixture gradient magnitude data points fit the 3-D model very well using this method. However the 2-D mixture gradient magnitude data points do not fit the 2-D model quite so well. This high-lights the potential inadequacies of modelling 3-D PV data with a 2-D feature and probabilistic description.

Exemplar scatter plots are given in figure 6.9. The 2-D scatter plot appears to illustrate that there are many data points that do not conform to either of the theoretical models D and E illustrated in figure 6.5. This is in contrast to the 3-D scatter plot that does appear to conform to the feature space of model E. This can be understood from the 2-D gradient magnitude calculation. This calculation may result in low gradient magnitude data points from volumetric image regions even with PV voxels. Such PV voxels would have been generated from the PSF filtering action in the z-axis. This also helps to explain why the 2-D model in figure 6.8(a) does not appear to fit the data points as well as the 3-D model for the 3-D data in figure 6.8(b).

A simple gradient kernel was used for all the experiments that follow and the calculations that have been discussed, for 2-D

$$M_x = \begin{pmatrix} +1 \\ 0 \\ -1 \end{pmatrix}, \; M_y = \begin{pmatrix} +1 & 0 & -1 \end{pmatrix};$$

and for 3-D

$$M_x = \begin{pmatrix} +1 \\ 0 \\ -1 \end{pmatrix}, \; M_y = \begin{pmatrix} +1 & 0 & -1 \end{pmatrix}, \; M_z = \begin{pmatrix} +1 & 0 & -1 \end{pmatrix}.$$

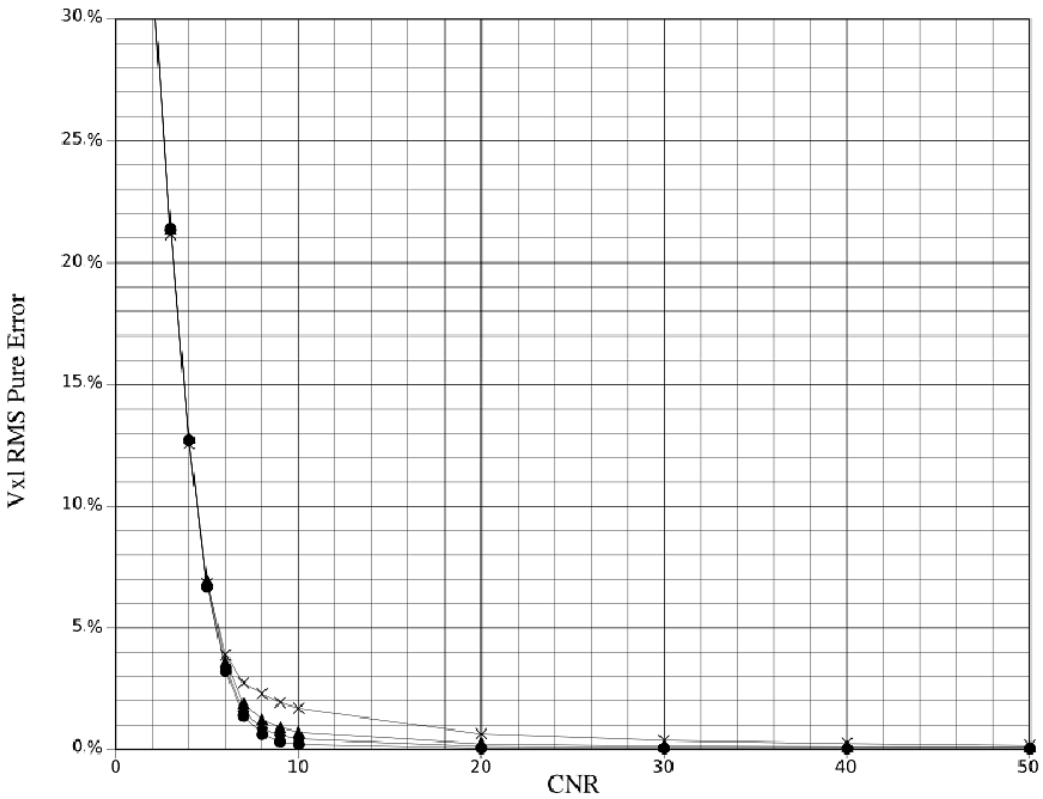
where the M_z kernel has elements defined on the z-axis.

Performance assessment was undertaken with the voxel RMS error metric, as defined in chapter 5.

Results and Discussion

The voxel RMS errors for each of the simulations in table 6.1 and each of the models (C,D,E and F) can be seen in figures 6.10 (pure voxels) and 6.11 (PV voxels).

The results for the classifier based on model C, the non-gradient magnitude based classifier defined in chapter 5 are also plotted along side the gradient magnitude classifier results. As can be seen from figures 6.10 and 6.11, the intensity based classifier has the highest classification error. So the inclusion of gradient magnitude in the likelihood model of the PV effect appears to improve the classification performance of simulated PV data.



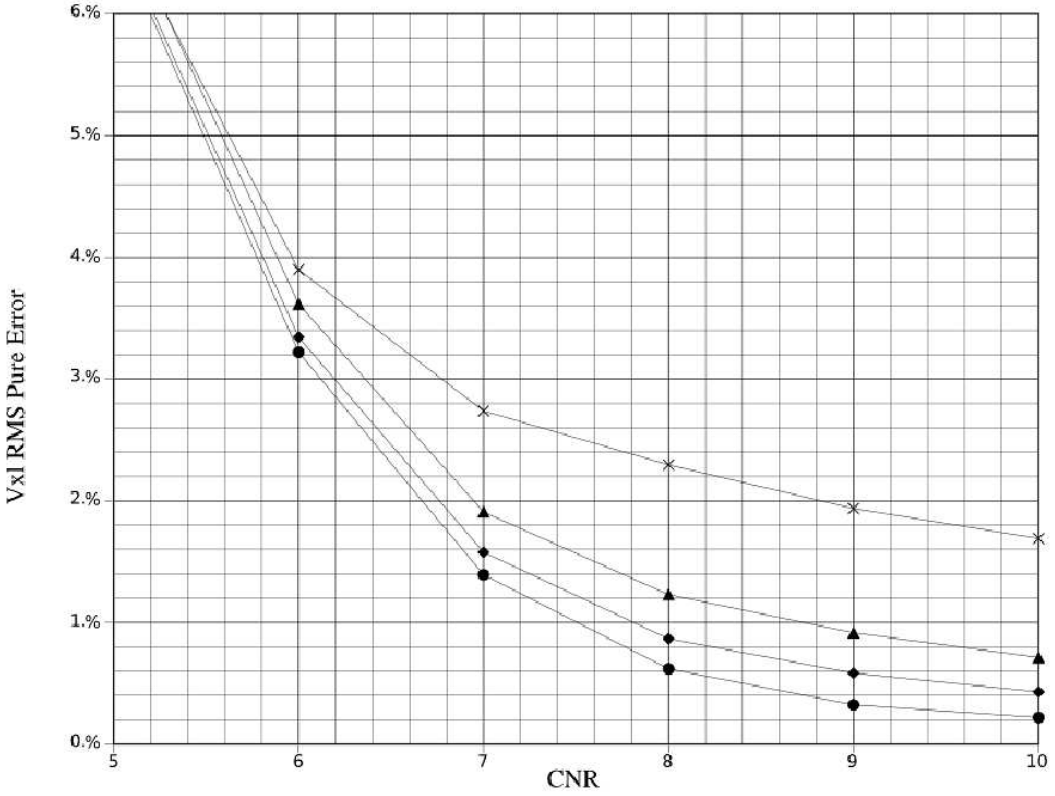
(a)

Figure 6.10: Pure voxel RMS errors obtained for classifying isotropic simulated PV data with classifier models C (×),D (▲),E (⧫) and F (•). (a) is for CNR values from 2 to 50; (b) is for CNR values from 5 to 10. Continued on next page.

Also, one can observe that model E has slightly better performance over model D for many of the pure and PV data points. This suggests that the improved feature space modelling (see figure 6.5), even with the use of 2-D gradient magnitude, improves the classification performance.

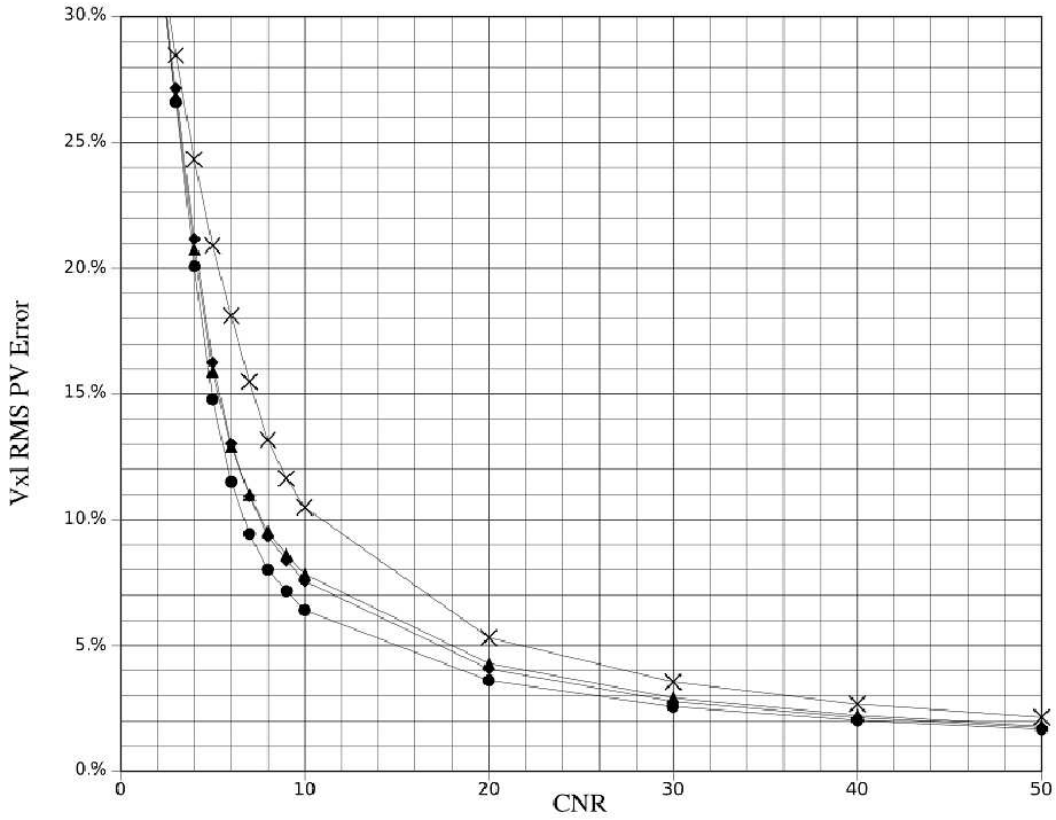
Additionally, the errors given in figures 6.10 and 6.11 appear to suggest that the use of a 3-D gradient magnitude model in the improved feature space described by model F, helps to improve the classification performance for all of the data points.

These results are interesting and help to justify the methodology and theory described in this chapter. As briefly mentioned earlier, these experiments have involved the classification of isotropically sampled simulated PV data volumes. The inventors of model D, using 2-D gradient magnitude have suggested that there is no benefit in the use of 3-D gradient magnitude. They suggested this was due to the prevalence of anisotropic volumetric imaging data [139]. The next set of experiments attempt to illustrate that there is some benefit in the utilisation of 3-D gradient magnitude on anisotropically sampled data.



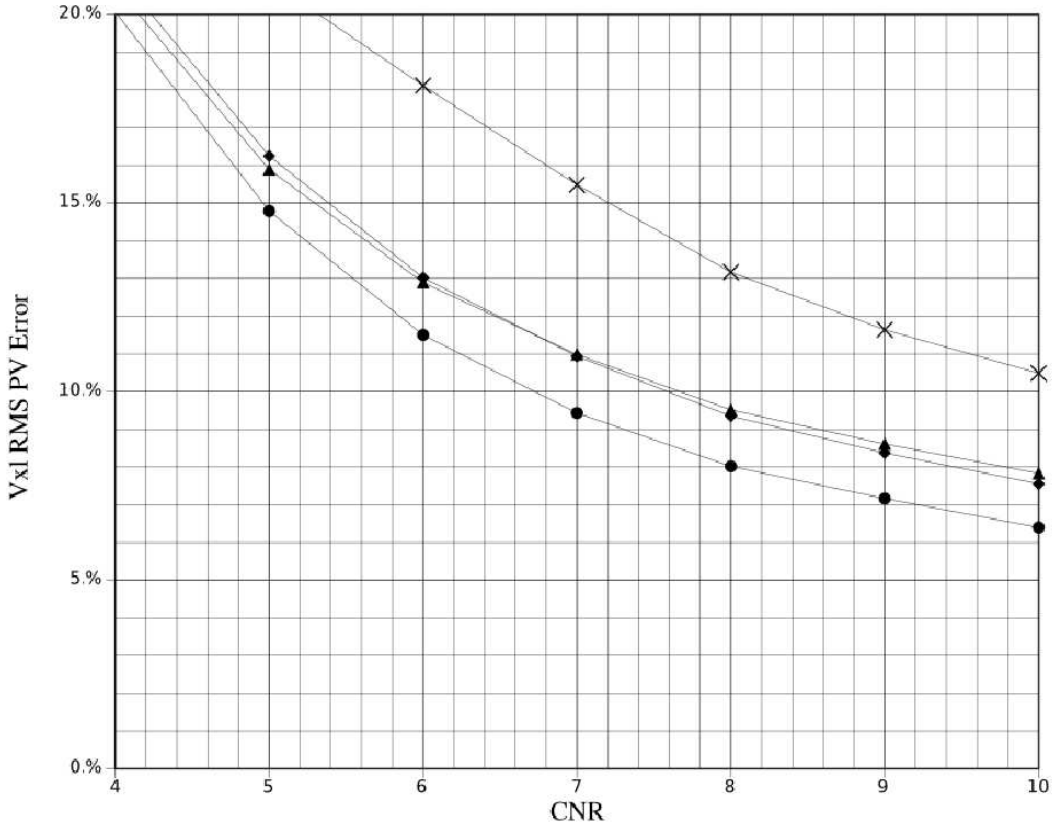
(b)

Figure 6.10: *continued*



(a)

Figure 6.11: PV voxel RMS errors obtained for classifying isotropic simulated PV data with classifier models C (×),D (▲),E (⧫) and F (●). (a) is for CNR values from 2 to 50; (b) is for CNR values from 4 to 10. Continued on next page.



(b)

Figure 6.11: *continued*

6.2.2 Classification of Anisotropic Simulated PV Data

As shown in the previous set of experiments, the utilisation of a 3-D gradient magnitude model improves the performance of the classification of PV data over a model utilising only 2-D gradient magnitude. The simulated PV data of the previous set of experiments was isotropic and while some medical imaging data is sometimes found to be isotropic, most MRI data is anisotropic. The experiments in this section involve the classification of anisotropic simulated PV data to determine whether the use of a 3-D gradient magnitude model still provides improved classification over a 2-D gradient magnitude model of 3-D simulated PV data.

Experimental Procedure

For this set of experiments new simulated PV data was generated to simulate anisotropic slice sampling. This involved modifying the PSF. In the preceding set of experiments, a symmetric Gaussian PSF was utilised to filter the high-resolution data prior to down-sampling. For the anisotropic simulations of this set of experiments, the anisotropic slices are assumed to possess a greater slice thickness. A greater slice thickness is often associated with biomedical imaging data to reduce the SNR without reducing the in-plane resolution. To simulate the greater slice thickness, the PSF is modified to have a greater width in the z-axis. Table 6.2 provides a list of the simulations together with their associated sampling ratios.

As can be seen from table 6.2, the same parameters that were used for the experiments in chapter 5 and the isotropic gradient magnitude simulations of this chapter are again used here. The only modifications are the simulated slice thicknesses, which result in sampling ratios of 1:1:2 and 1:1:4, in effect doubling the slice thickness for each set of simulations. These slice thicknesses simulate possible sampling ratios that may be encountered with biomedical imaging data and can be thought as representative of other similar sampling ratios.

The calculation of the 3-D gradient magnitude has to be modified slightly. In the preceding sets of experiments, the 3-D gradient magnitude was calculated according to equation 6.35 and repeated here for convenience:

$$Z_{3D} = \sqrt{Z_x^2 + Z_y^2 + Z_z^2} \tag{6.53}$$

where Z_x, Z_y and Z_z are the results of the gradient mask convolutions with the image data for the x, y and z axes respectively.

The gradient masks, M_x, M_y and M_z assume the data points in each axis exist on an isotropic grid but the data is anisotropically sampled for these experiments. The isotropic grid has to be estimated. As each 2-D plane forms an isotropic 2-D grid, the only affected data points are on the z axis and consequently only affect the z gradient mask.

The gradient mask for the z axis is given by

$$M_z = \begin{pmatrix} +1 & 0 & -1 \end{pmatrix}.$$

Table 6.2: Summary of anisotropic simulations' parameters.

CNR	**Mean Distance**	**Mean 1**	**Mean 2**	**Stan. Dev. 1**	**Stan. Dev. 2**	**Sampling Ratio**
50	1000	500	1500	20	20	1:1:2
40	800	600	1400	20	20	1:1:2
30	600	700	1300	20	20	1:1:2
20	400	800	1200	20	20	1:1:2
10	200	900	1100	20	20	1:1:2
9	180	910	1090	20	20	1:1:2
8	160	920	1080	20	20	1:1:2
7	140	930	1070	20	20	1:1:2
6	120	940	1060	20	20	1:1:2
5	100	950	1050	20	20	1:1:2
4	80	960	1040	20	20	1:1:2
3	60	970	1030	20	20	1:1:2
2	40	980	1020	20	20	1:1:2
50	1000	500	1500	20	20	1:1:4
40	800	600	1400	20	20	1:1:4
30	600	700	1300	20	20	1:1:4
20	400	800	1200	20	20	1:1:4
10	200	900	1100	20	20	1:1:4
9	180	910	1090	20	20	1:1:4
8	160	920	1080	20	20	1:1:4
7	140	930	1070	20	20	1:1:4
6	120	940	1060	20	20	1:1:4
5	100	950	1050	20	20	1:1:4
4	80	960	1040	20	20	1:1:4
3	60	970	1030	20	20	1:1:4
2	40	980	1020	20	20	1:1:4

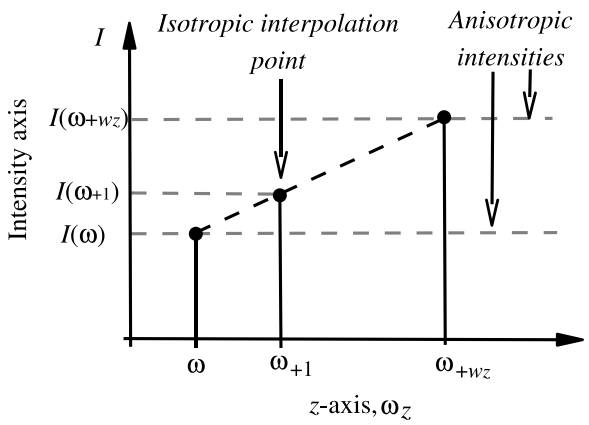


Figure 6.12: Illustration of the assumption of linear variation in intensities when performing linear interpolation to calculate an isotropically located sample, $\boldsymbol{\omega}_{+1}$, located between two anisotropically sampled points, $\boldsymbol{\omega}$ and $\boldsymbol{\omega}_{+w_z}$.

The gradient mask only requires data points either side of the x-y plane in question. Using linear interpolation, a data point either side of this x-y plane can be calculated for a sampling ratio of 1:1:1. Linear interpolation assumes the intensity values vary linearly as a function of distance. For an anisotropic sampling ratio of $1:1:w_z$, the intensity value of the data at an isotropic z axis distance, $I(\boldsymbol{\omega}_{+1})$ away from a particular voxel, $\boldsymbol{\omega}$ is given by

$$\begin{aligned} I(\boldsymbol{\omega}_{+1}) &= (I(\boldsymbol{\omega}_{+w_z}) - I(\boldsymbol{\omega})).\frac{1}{w_z} + I(\boldsymbol{\omega}) \\ &= I(\boldsymbol{\omega}_{+w_z})/w_z + I(\boldsymbol{\omega}).\left(\frac{1-w_z}{w_z}\right), \end{aligned} \tag{6.54}$$

where $\boldsymbol{\omega}_{+w_z}$ is the value of the image data on the anisotropic grid. This situation is illustrated in figure 6.12.

The result of convolution with the simple gradient mask is therefore given by:

$$Z_z(\boldsymbol{\omega}) = I(\boldsymbol{\omega}_{+1}) - I(\boldsymbol{\omega}_{-1}) \tag{6.55}$$

where $\boldsymbol{\omega}_{-1}$ is the point on the isotropic grid in the opposite direction to $\boldsymbol{\omega}_{+1}$, so that:

$$Z_z(\boldsymbol{\omega}) = I(\boldsymbol{\omega}_{+w_z})/w_z + I(\boldsymbol{\omega}).\left(\frac{1-w_z}{w_z}\right) - I(\boldsymbol{\omega}_{-w_z})/w_z - I(\boldsymbol{\omega}).\left(\frac{1-w_z}{w_z}\right). \tag{6.56}$$

resulting with

$$Z_z(\boldsymbol{\omega}) = \frac{1}{w_z}.(I(\boldsymbol{\omega}_{+w_z}) - I(\boldsymbol{\omega}_{-w_z})). \tag{6.57}$$

This illustrates that the calculation of the z axis gradient for anisotropic data is a weighted version of the usual z axis gradient calculation. From earlier discussions about model F in this chapter and the content of appendix C, this result does not appear to alter any of the assumptions about Gaussianity of the gradient components. It does however alter the standard deviation of the z axis convolutions, by a scaling factor of w_z.

It remains to be seen whether this will have an impact on the classification performance of the 3-D classifier. It is expected that due to the presence of two other gradient components for each calculation, the standard deviations of these two components will dominate the width of the resulting gradient magnitude density of equation 6.49. Therefore model F will be applied unaltered for classification of the anisotropic imaging data experiments that follow shortly.

As with the preceding sets of experiments, the intensity domain parameters were known *a priori* and the same techniques as described in section 6.2.1 were used to determine the gradient magnitude domain parameters from the simulations and the ground truth data volumes.

The gradient magnitude for the 2-D models and the 3-D models were also calculated using the methods described in section 6.2.1.

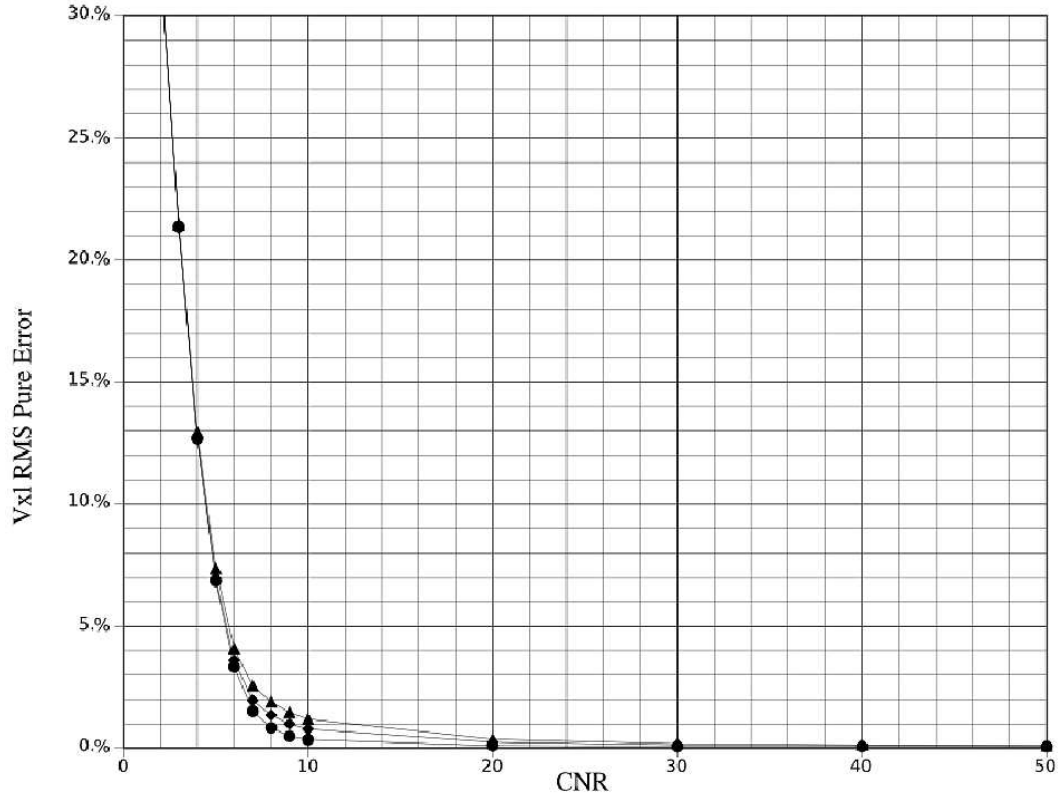
Results and Discussion

The results of classifying the anisotropic simulated PV data as shown in table 6.2 with models D,E and F can be seen in figures 6.13 to 6.16. Each figure is divided into more than one sub-figure, one of which plots the entire results for a particular sampling ratio and the second and perhaps third sub-figures display the results for a limited range of CNR values, so as to high-light differences between the performances of each of the classifiers.

Figures 6.13 and 6.14 display the pure and PV errors respectively for classifiers using models D, E and F for the simulations displayed in table 6.2 with a sampling ratio of $1 : 1 : 2$. These figures illustrate that the errors obtained for the 3-D classifier (model F) are superior to the errors obtained for the 2-D models D and E, particularly for the pure RMS voxel errors for CNR values between 5 and 20. The PV RMS voxel errors appear to be very similar for models E and F, both of which appear to have superior performance over the former 2-D classifier, (model D described in [157]).

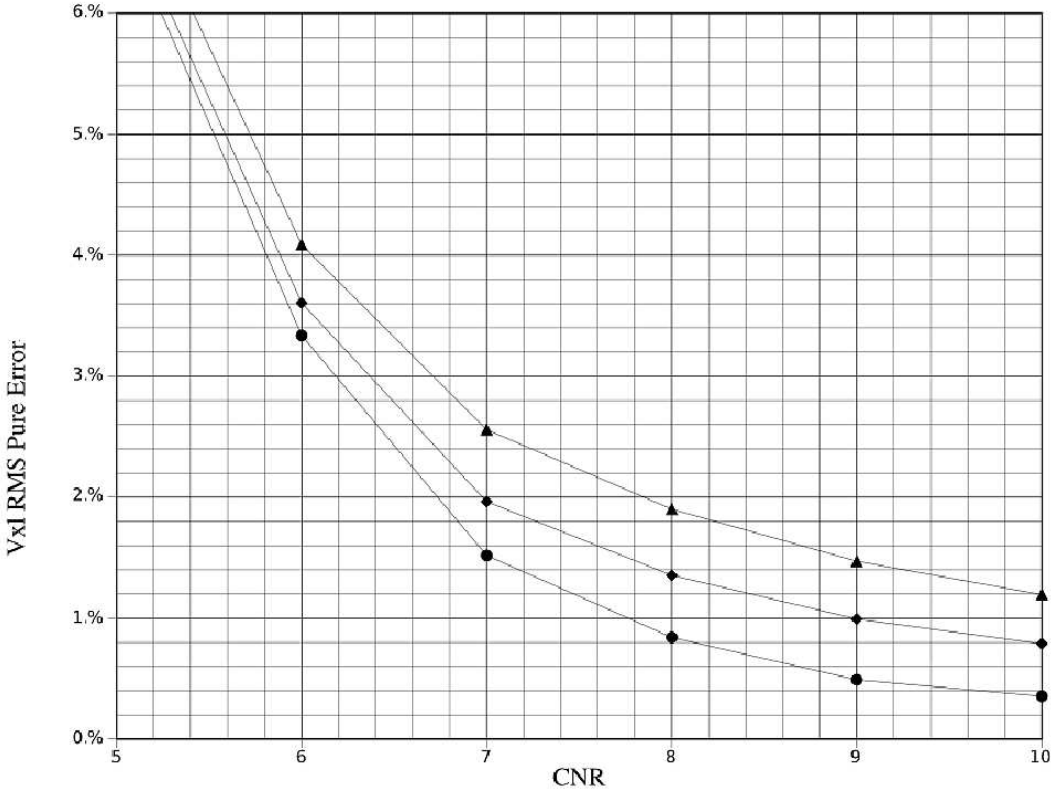
Figures 6.15 and 6.16 display the pure and PV errors respectively for classifiers using models D,E and F for the simulations displayed in table 6.2 with a sampling ratio of $1 : 1 : 4$. These figures also illustrate that the errors obtained for the 3-D classifier (model F) are superior for the classification of pure voxels (figure 6.15). For the classification of PV voxels, all the models appear to possess similar classification performances. For CNR values over 10, the 3-D classifier (model F) appears to be somewhat superior. For CNR values smaller than 8, the former 2-D classifier (model D) appears to possess slightly better classification performance over the two new classifiers.

These results appear to suggest that the inclusion of the gradient magnitude from the third dimension and correct modelling of the gradient magnitude improves the classification rate. In situations where performance has not been improved (such as for PV voxels on anisotropic data with sampling ratios of $1 : 1 : 4$), the classification rate for all the models appear to be very similar. These two observations from the results of these experiments seem to indicate that 3-D gradient magnitude is a useful feature over 2-D gradient magnitude when classifying data with anisotropic PV voxels.



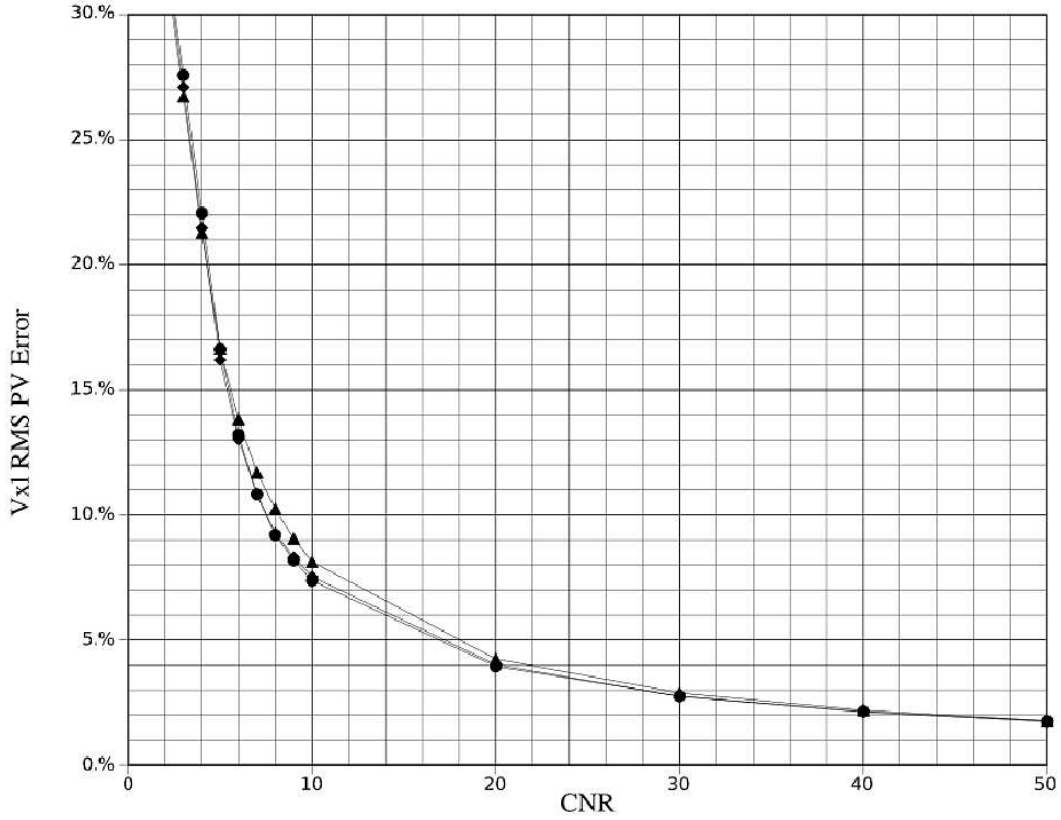
(a)

Figure 6.13: Pure voxel RMS errors obtained for classifier models D (▲),E (⧫) and F (●) for anisotropic simulated PV data with sampling ratios of 1 : 1 : 2. (a) is for CNR values from 2 to 50; (b) is for CNR values from 5 to 10. *Continued on next page.*



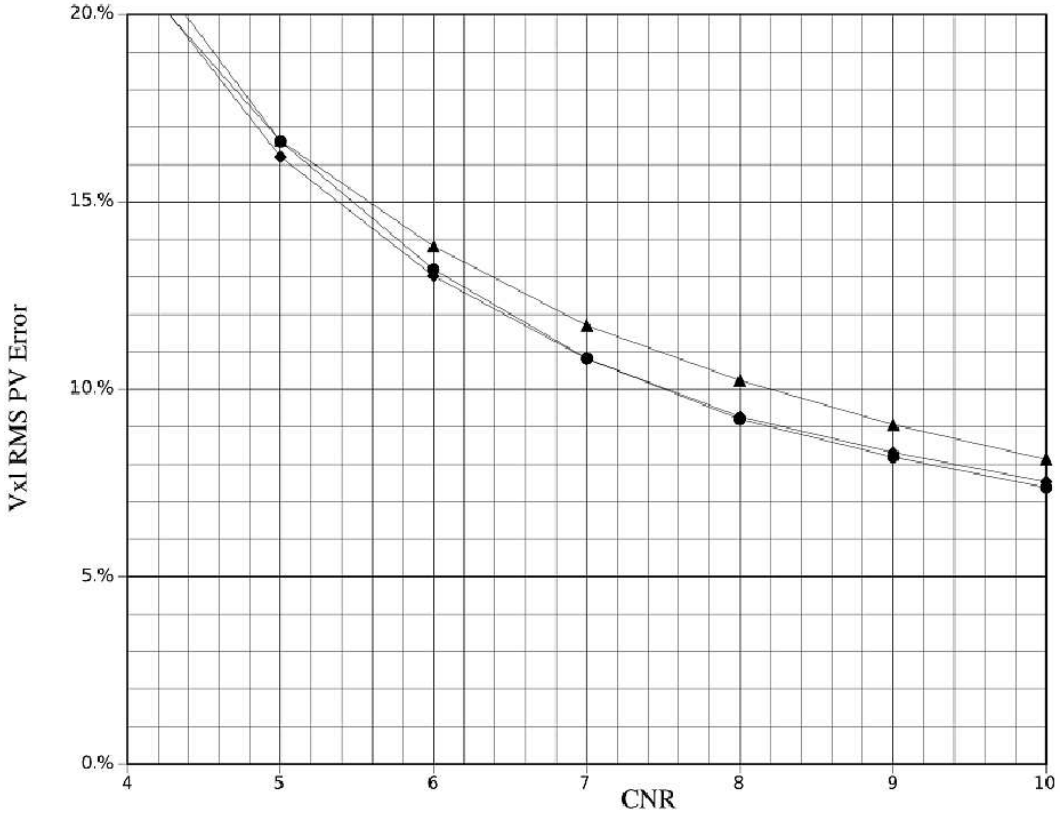
(b)

Figure 6.13: *continued*



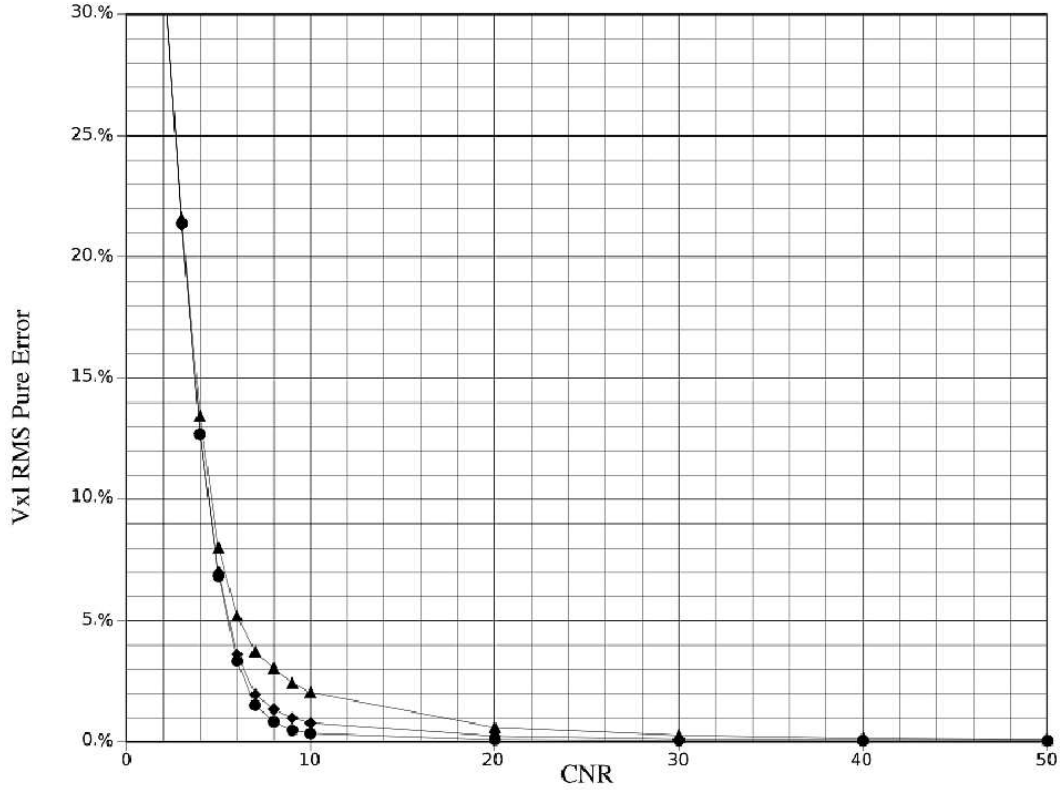
(a)

Figure 6.14: PV voxel RMS errors obtained for classifier models D (▲),E (⧫) and F (●) for anisotropic simulated PV data with sampling ratios of $1:1:2$. (a) is for CNR values from 2 to 50; (b) is for CNR values from 4 to 10. *Continued on next page.*



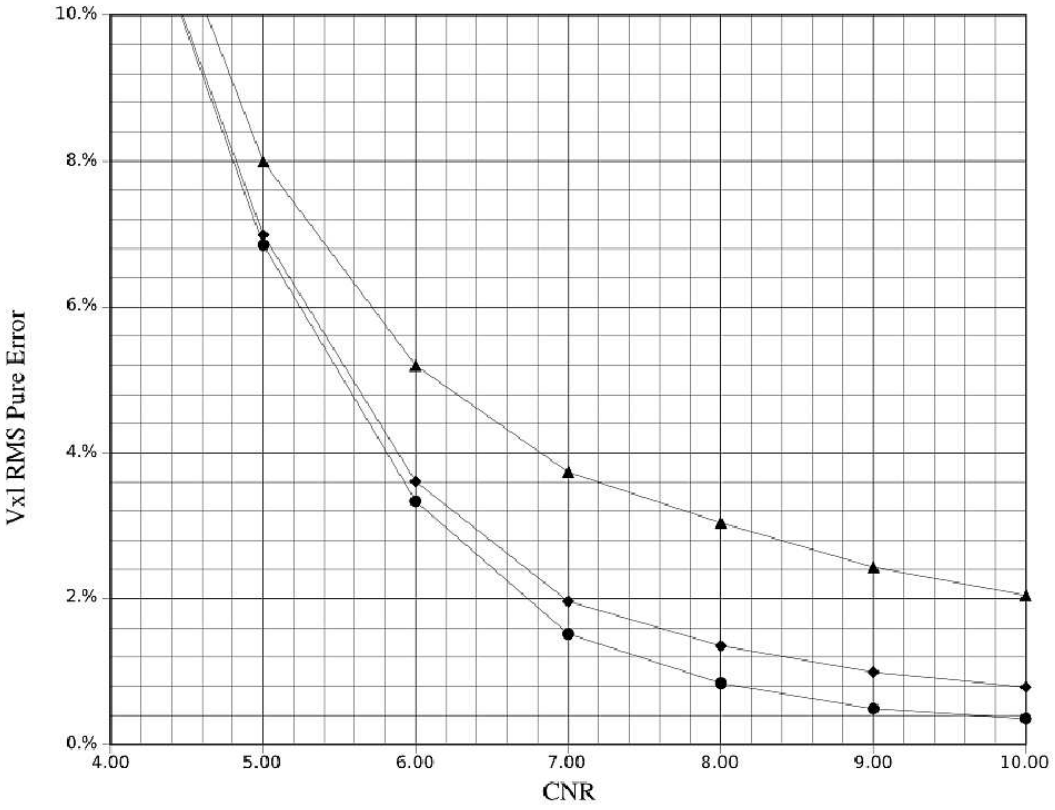
(b)

Figure 6.14: *continued*



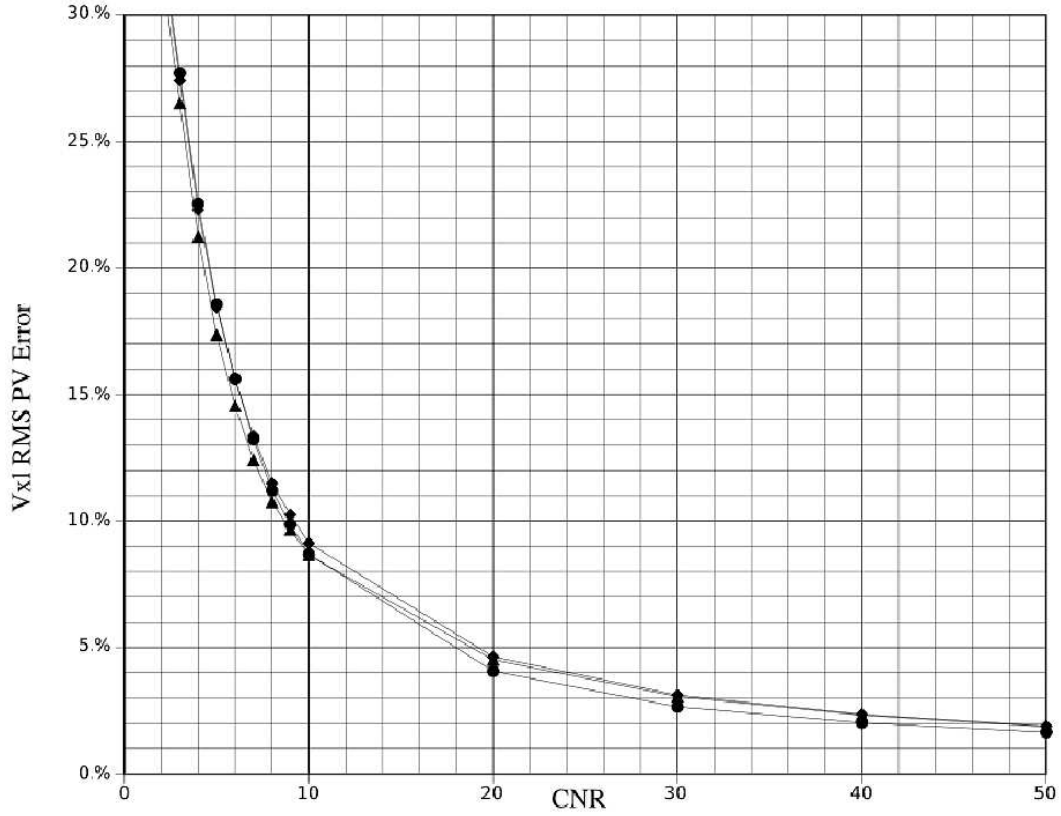
(a)

Figure 6.15: Pure voxel RMS errors obtained for classifier models D (▲),E (⧫) and F (●) for anisotropic simulated PV data with sampling ratios of 1 : 1 : 4. (a) is for CNR values from 2 to 50; (b) is for CNR values from 4 to 10. *Continued on next page.*



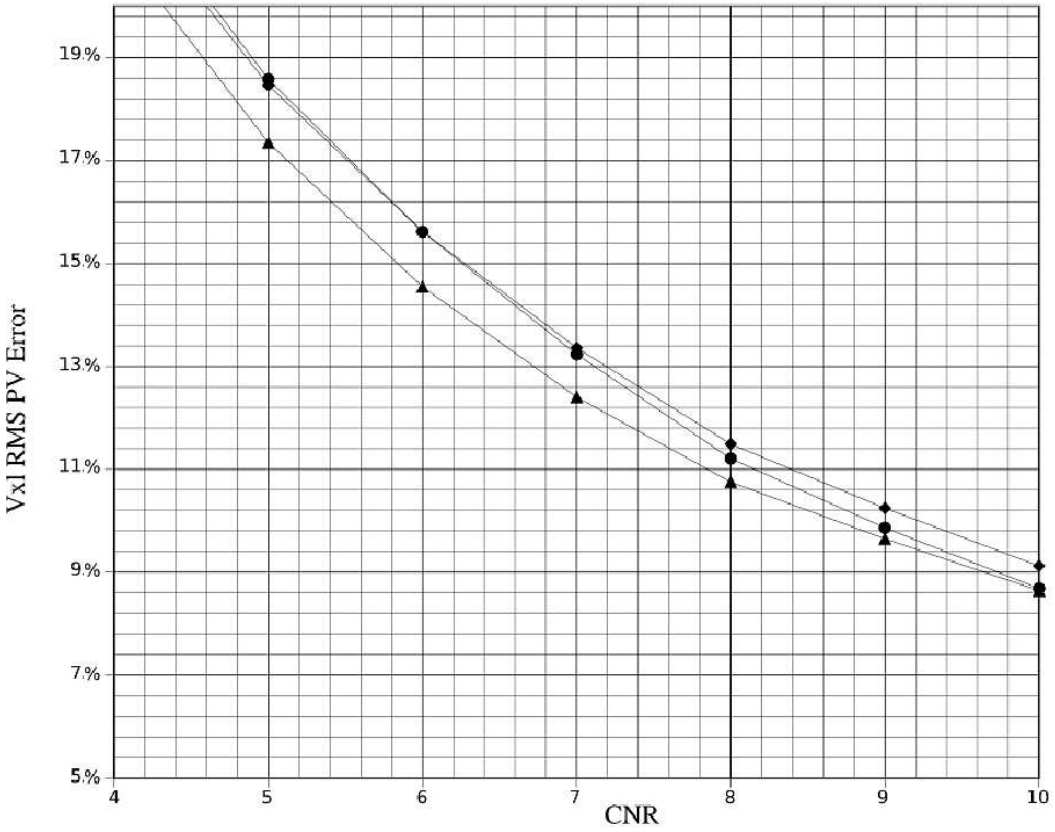
(b)

Figure 6.15: *continued*



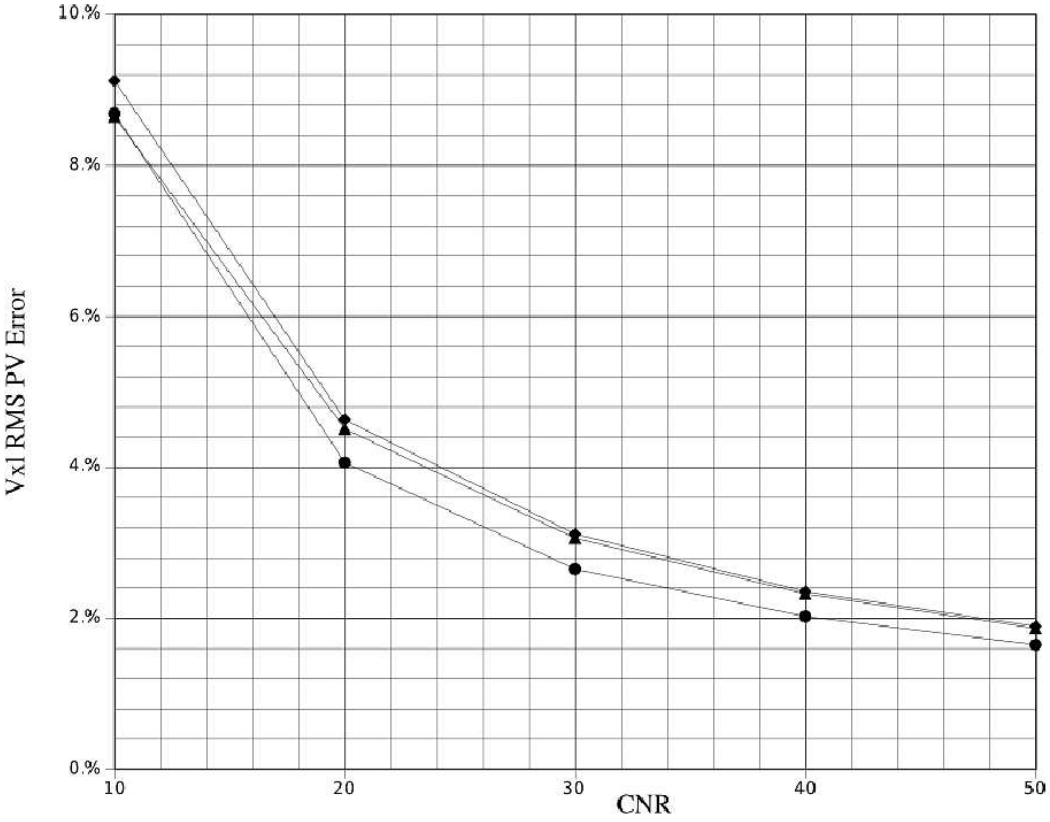
(a)

Figure 6.16: PV voxel RMS errors obtained for classifier models D (▲),E (⧫) and F (●) for anisotropic simulated PV data with sampling ratios of 1 : 1 : 4. (a) is for CNR values from 2 to 50; (b) is for CNR values from 4 to 10; (c) is for CNR values from 10 to 50. *Continued on next two pages.*



(b)

Figure 6.16: *continued*



(c)

Figure 6.16: *continued*

6.3 In Conclusion

Improvements to the basic intensity models presented in chapter 5 have been investigated. These improvements have involved modelling of 2-D and 3-D gradient magnitude and the effect on the combined intensity and gradient magnitude feature spaces.

An initial model based on an intensity and 2-D gradient magnitude feature space, initially suggested by Williamson *et al.* in [157] has been modified in this chapter. Firstly, an improved and refined feature space was modelled, utilising an alternative formulation. The results obtained from classifying simulated two class PV data suggest that the improved feature space model improves the classifiers ability to distinguish the composition of voxels. In addition to this feature space improvement, it was recognised that the 2-D gradient magnitude description of inherently 3-D data could be improved with the correct modelling of 3-D gradient magnitude. This further development was also demonstrated to improve the classifier's performance.

Classification performance for pure voxels in the simulated PV data for CNR values above 5 possessed RMS errors below 5%. For PV voxels, the RMS error fell below 5% for CNR above 15. As discussed in chapter 5, biomedical imaging data usually has CNR values below 10, so if a data volume possessed a large number of PV voxels, the classifiers' performance may not be adequate for clinical application. Also, these performances are under controlled conditions where all parameters are known *a priori*, due to the availability of the ground truth. Clearly, further improvements and experiments are required to (i) improve the classification performance and (ii) determine the performance of the classifiers where parameters are not known *a priori*. These are the topics of the next two chapters.

Chapter 7

Informative Partial Volume Prior Models

The preceding chapters have presented PV models some of which have been developed as part of this thesis. Part of the new developments has included the proposal of a formulation that explicitly models the PV mixing via a per voxel PV random vector, $\boldsymbol{\alpha}$. Each PV random vector corresponds to an individual voxel with vector elements representative of the amount of a particular tissue or activity in that particular voxel, $\alpha_1, \alpha_2, ..., \alpha_K$. This random vector can therefore be directly included in the probabilistic formulation. In the previous chapters the PV random vector was assigned a uniform prior, $p(\alpha) = 1$, (when $K = 2$), for the PV values, i.e. $\alpha \in (0, 1)$. As will be seen, this uniform prior distribution is not, in fact, representative of the true PV prior distribution. However, Bayesian models benefit from well informed prior distributions and the development of this global PV prior distribution is the topic of the first part of this chapter. The global PV prior is then used as the basis for a proposal distribution in a more sophisticated and better performing probabilistic formulation of the PV effect where the PV posterior distribution is conditionally dependent on a locally defined prior PV distribution. It will be seen that this locally defined prior distribution enables the Bayesian formulation to adapt to local PV behaviour and thereby improve the classification performance.

7.1 Global PV Priors

7.1.1 Introduction

This section of the work is concerned with the estimation of global prior mixture distributions, $p(\boldsymbol{\alpha})$, as used in the PV models of the preceding chapters. The global prior mixture distribution describes the distribution of intensities in the absence of noise and can therefore be analysed also in the absence of a noise component, which is the approach taken here. An assumption used here, is that the PV effect predominantly arises from signals that have been significantly affected by a low-pass filtering process. This filtering process was discussed in chapters 1 and 2. The filtering action is a natural

effect in limitations of the image acquisition process and further processes that are used to condition the image signal prior and post acquisition.

A new method for modelling the global PV prior density given a system PSF is presented. Similarly an old phenomenological observation, namely Benford's Law is also related to the global PV mixing distribution.

Linear mixing is also assumed, where, for two classes, a and b, $\alpha_a = 1 - \alpha_b$, (as was also assumed in chapter 5). Thus for two PV mixing components, $\boldsymbol{\alpha}$ can be fully specified as a scalar value, α, resulting in a single-variate prior probability, $p(\alpha)$. Initially $p(\alpha)$ is estimated analytically. The result of this analysis is then compared with the proposed application of Benford's Law, and a third related distribution proposed by Kitamoto and Takagi, [66], previously used to describe the mixing processes that occur in remotely sensed satellite image data. The applicability of these new results are then tested on synthetic volumetric image data, where the true prior densities are known.

7.1.2 Theory

Gaussian Based Derivation of the Prior Density

If one considers an idealised bimodal signal to be characterised by a step edge composed of an infinite number of frequencies, then the result of a band-limited signal acquisition process with a low-pass frequency response will smooth the step-edge, thus reducing the higher-frequency components in the signal. To illustrate this, an idealised noiseless edge with intensity values arbitrarily assigned to values of 0 and 1, illustrated in figure 7.1(a) can be described by:

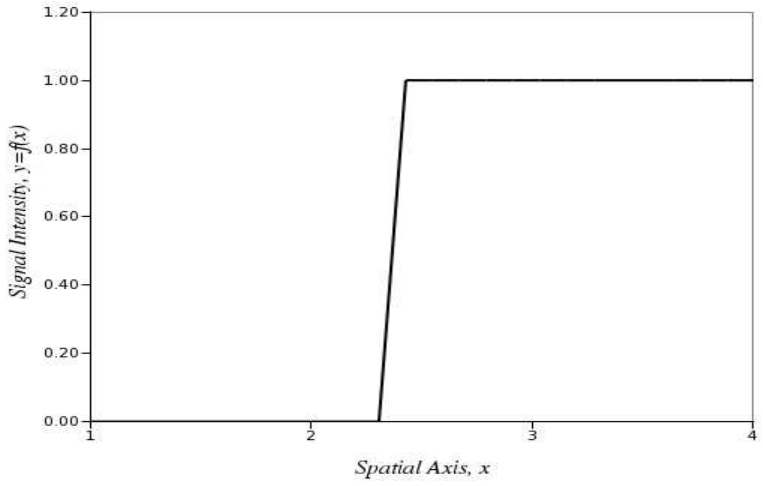
$$f(x) = \begin{cases} 1 & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x \leq 0 \end{cases} . \tag{7.1}$$

The low-pass response of the signal acquisition process can sometimes be characterised by a Gaussian PSF, $h(x)$, as is often found in image or signal acquisition systems, illustrated in figure 7.1(b). In instances where the PSF is not Gaussian, a Gaussian PSF is still a better approximation to most real PSFs of imaging systems in comparison to a boxcar type PSF that has been used in models of the PV effect, see e.g. [74, 83]:

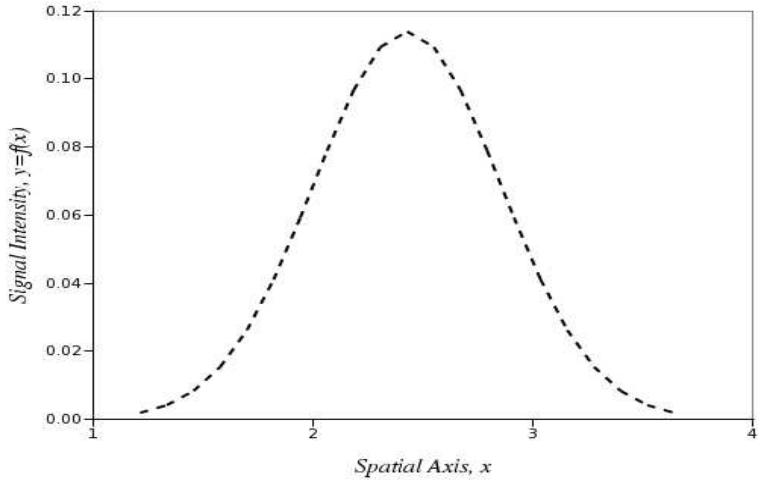
$$h(x) = \frac{1}{\sqrt{2.\pi.\sigma^2}} . \exp\left(-\frac{x^2}{2.\sigma^2}\right) . \tag{7.2}$$

where σ^2 is the variance or measure of width of the PSF. The idealised step edge in equation 7.1 can then be convolved with this kernel to obtain the idealised noiseless representation of the signal post-acquisition:

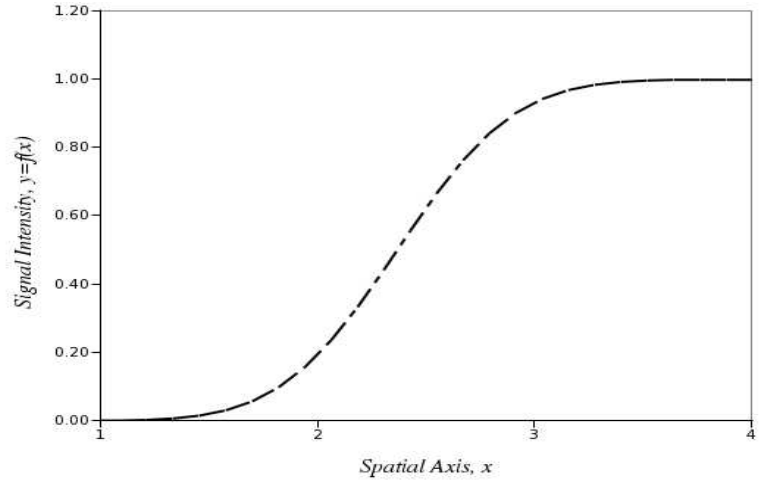
$$\begin{aligned} y(x) &= f(x) * h(x) \\ &= \int_{-\infty}^{\infty} f(\tau).h(x-\tau).d\tau \\ &= \int_{0}^{\infty} \frac{1}{\sqrt{2.\pi.\sigma^2}} . \exp\left(-\frac{(x-\tau)^2}{2.\sigma^2}\right) .d\tau. \end{aligned} \tag{7.3}$$



(a)



(b)



(c)

Figure 7.1: Illustration of the result of the convolution (c) of the step edge (a) with the Gaussian PSF kernel (b).

Noting that $(x-\tau)^2 \equiv (\tau-x)^2$ and letting $u = (\tau-x)/\sqrt{2.\sigma^2}$ so that $du/d\tau = 1/\sqrt{2.\sigma^2}$:

$$y(x) = \frac{1}{\sqrt{2.\pi.\sigma^2}}. \int_{-x/\sqrt{2.\sigma^2}}^{\infty} \exp\left(-u^2\right).du.\sqrt{2.\sigma^2}$$

$$= \frac{1}{\sqrt{\pi}}. \int_{-x/\sqrt{2.\sigma^2}}^{\infty} \exp\left(-u^2\right).du. \tag{7.4}$$

This integral has no closed form solution so a special function known as the Gaussian error function, $\text{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp\left(-u^2\right).du$ can be used:

$$y(x) = \frac{1}{\sqrt{\pi}}.\frac{\sqrt{\pi}}{2}\left(\text{erf}(\infty) - \text{erf}\left(-\frac{x}{\sqrt{2.\sigma^2}}\right)\right), \tag{7.5}$$

resulting in, 7.3 is given by:

$$y(x) = \frac{1}{2}\left(1 + \text{erf}\left(\frac{x}{\sqrt{2.\sigma^2}}\right)\right). \tag{7.6}$$

This convolution result is illustrated in figure 7.1(c). It is now desirable to obtain a continuous histogram model of the signal represented by equation 7.6. This can be considered as a function that describes the proportion of the signal that occupies finite ranges of $y(x)$. It can be determined by initially finding the inverse of $y(x)$, $x(y) = y^{-1}(x)$, thus:

$$x(y) = \sqrt{2.\sigma^2}.\text{erf}^{-1}\left(2.y - 1\right), \tag{7.7}$$

where $\text{erf}^{-1}(t)$ is the inverse error function, so that $\text{erf}(\text{erf}^{-1}(t)) = t$. Due to the global monotonicity of $y(x)$, the frequency histogram information of $y(x)$ can be found by determining the distance $x(y)$ travels in a small interval, represented by $D(y) \approx x(y+\delta y) - x(y)$, and taking the limit of $\delta y \to 0$ results in the first derivative of $x(y)$, $D(y) \approx \frac{\mathrm{d}x(y)}{\mathrm{d}y}$. Utilizing the result of Carlitz [16], the first derivative of the inverse error function is taken to be $\frac{\mathrm{derf}^{-1}(z)}{\mathrm{d}z} = \frac{\sqrt{\pi}}{2}\exp\left(\text{erf}^{-1}(z)^2\right)$, and letting $\alpha = y$ so that:

$$p_{Gauss}(\alpha = y) = C_{G_{a,b}}. \left.\frac{\mathrm{d}x(y)}{\mathrm{d}y}\right|_{y=\alpha}$$

$$= C_{G_{a,b}}.\sqrt{2.\pi.\sigma^2}.\exp\left(\text{erf}^{-1}\left(2.\alpha - 1\right)^2\right), \tag{7.8}$$

where $C_{G_{a,b}}$ is a normalizing term for the above Gaussian based model. This result illustrates that the standard deviation of the original convolution PSF, σ, has no effect on the *shape* of the resulting mixture density other than to contribute a scaling parameter. This Inverse Cumulative Gaussian (ICG) density is illustrated in figure 7.2.

Benford's Law

A phenomenological law, known as Benford's Law, has previously been used to describe the natural ordering of frequency data. It was originally discovered by Newcomb in 1888

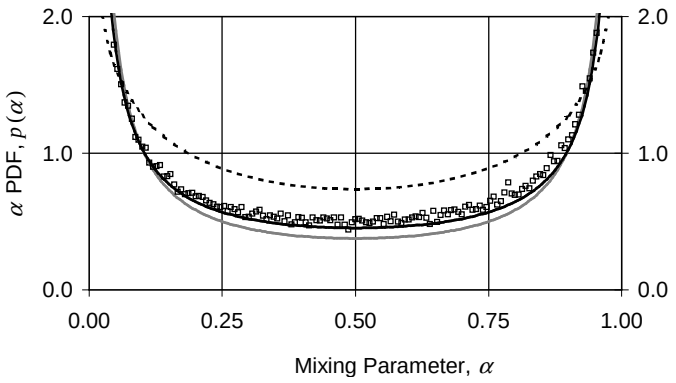


Figure 7.2: Comparison of the three theoretical prior mixing densities, Benford mixing density (gray line), Kitamoto and Takagi mixing density (dotted line) and the ICG mixing density (black line). Exemplar data points are also shown (squares) for simulated PV data. These exemplar data points are typical of the data points produced via a PV filtering process.

[93], but it was popularised by Frank Benford who re-discovered the phenomenological law in 1938. Both authors discovered it by observing that books of logarithm tables contained pages that were more warn at the beginning. They discovered that it could be related to the frequency of the leading numerical digits in other sources of data such as newspapers, atomic weights and black body radiation. The digit 1 was found to be more likely than all the other digits as the leading digit, resulting in, quite un-intuitively, a non-uniform distribution for this data. The following equation was suggested to describe the leading digits, [7]:

$$P(\beta) = \log_{10}\left(1 + \frac{1}{\beta}\right), \tag{7.9}$$

where $\beta \in \{b : 0 < b < 10,\ b \in \mathbb{Z}^+\}$, and $\mathbb{Z}^+$ is the set of positive integers. Equation 7.9 is illustrated in figure 7.3. This law can be extended to any number of significant digits, $\boldsymbol{\beta} = \{\beta_1, ..., \beta_Q\}$, with specific order, by the following [7, 58]:

$$P(\boldsymbol{\beta}) = \log_{10}\left[1 + \left(\sum_{q=1}^{Q} \beta_q . 10^{Q-q}\right)^{-1}\right]. \tag{7.10}$$

e.g. given a number, 134, then $\boldsymbol{\beta} = \{1, 3, 4\}$ and $P(\boldsymbol{\beta}) = \log_{10}(1 + 1/134)$, but if the number has a different decimal place, such as 1.34, then $P(\boldsymbol{\beta}) = \log_{10}(1 + 1/134)$ also.

Hill in 1996 [58], provided an explanation of Benford's Law by showing that if data comes from many different distributions and is scale and base invariant then the overall distribution will tend towards this Benford distribution-like behaviour.

Similarly, the result of the filtering process on noiseless data produces data points that are governed by many different distributions due to the large number of different edge or boundary configurations and the result of their convolutions with the PSF. The result of this filtering operation also produces data with a histogram that is scale invariant.

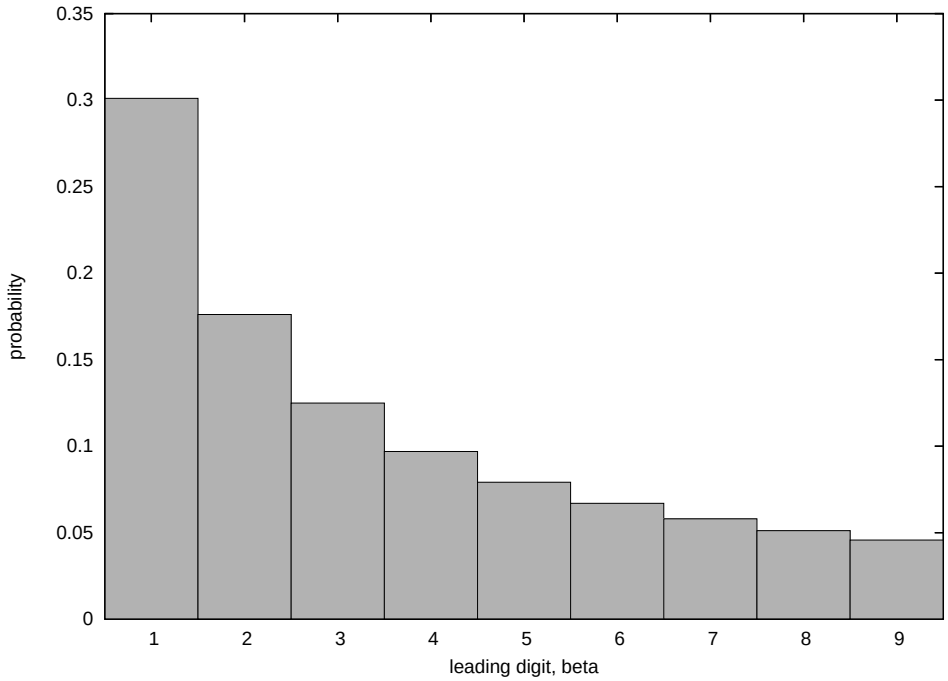


Figure 7.3: Illustration of Benford's Law for a single digit, described by equation 7.9.

This can be understood from the shape of the histogram which is not dependent on the size of the pixels or voxels, demonstrated by the earlier derivation of the ICG density, the result of which is given by equation 7.8. The histogram of this filtered data is also base invariant, where the histogram's shape is independent of the number of signal levels between the signal components j and k.

As a result, one can consider the prior mixing density to be composed of two equal, but symmetrically opposite, ordered, frequency components that are not independently directly observable. Also note that Benford's law provides a discrete probability distribution, i.e. a PMF which describes the random distribution of a discrete random variable or vector, in comparison to a PDF that describes the random distribution of a continuous random variable. See chapter 2 for further information regarding distribution functions of discrete and continuous random variables. Therefore the prior mixing density, using Benford's law is actually described by a PMF, composed of two Benford distributions, thus describing a prior mixing distribution not density. This will be explained shortly, but first let us consider the general form of this PMF, composed of two Benford distributions:

$$P(\zeta_a, \zeta_b) = C_{B_{a,b}}. \left[\log_{10}\left(1 + \frac{1}{\zeta_a}\right) + \log_{10}\left(1 + \frac{1}{\zeta_b}\right) \right] \tag{7.11}$$

where ζ_a and ζ_b are variables, referred to here as *Benford variables*, that have to be related via some sort of mapping to the continuous mixing variables, α_a and α_b, respectively. a, b are indicative of particular classification components or classes under consideration and $C_{B_{a,b}}$ is a normalizing factor. The Benford variables, ζ_a and ζ_b can

be observed in equation 7.10 as taking values according to

$$\sum_{q=1}^{Q} \beta_q.10^{Q-q}, \tag{7.12}$$

where β_q is the q^{th} significant digit in the observation. This therefore implies a type of precision invariance, where Q may take any value, with greater values providing greater precision. Noting that the mixing variables, referred to here as, α can also be quoted to any number of significant digits or precision, therefore a mapping between α and the Benford variable, v can be found with the use of an integerisation type function, int(). i.e.

$$v = \text{int}(\alpha.10^{Q}). \tag{7.13}$$

This therefore provides a means to describing the continuous random variable in the discrete domain, where the PMF of the prior mixing is given by

$$P(\alpha_a, \alpha_b) = C_{B_{a,b}}. \left[\log_{10}\left(1 + \frac{1}{\alpha_a.10^{Q}}\right) + \log_{10}\left(1 + \frac{1}{\alpha_b.10^{Q}}\right) \right]. \tag{7.14}$$

This result is similar to the previously defined ICG density in equation 7.8, in that equation 7.14 is scale invariant, i.e. there is no parameter to control the shape. This Benford prior distribution is also illustrated in figure 7.2.

7.1.3 Experimental Methodology

As with the preceding chapters, the models proposed here are assessed using simulated PV data. The simulated data consists of concentric spheroids with alternating intensities, μ_a and μ_b. This has then been filtered under a variety of conditions, ranging from boxcar-like to Gaussian-like PSFs. These filters have been designed using the Kaiser window method [71].

A χ^2 Goodness-of-Fit (GoF) test of the two prior mixing models described thus far, $p_{Gauss}(\alpha)$, the ICG prior density and $p_{Benford}(\alpha)$, the Benford prior density and a further density given in [66] is performed on histograms produced from this simulated PV data.

Simulated Filter Design

The Kaiser window method calculates the signal domain filter coefficients, w_δ by:

$$w_\delta(n) = \frac{1}{I_0(\delta)}.I_0\left(\delta.\sqrt{1 - \frac{2.n}{\Upsilon - 1}}\right) \tag{7.15}$$

where $n = \left[-\frac{\Upsilon}{2}, \frac{\Upsilon}{2}\right]$ is the co-efficient index. Υ is the window size and δ is used to express the maximal ripple or side lobe magnitude as a fraction of the maximal pass band gain. $I_0(u)$ is the modified Bessel function of the first kind, order 0. Υ and δ are calculated according to a well known algorithm (given in algorithm 3, see for example

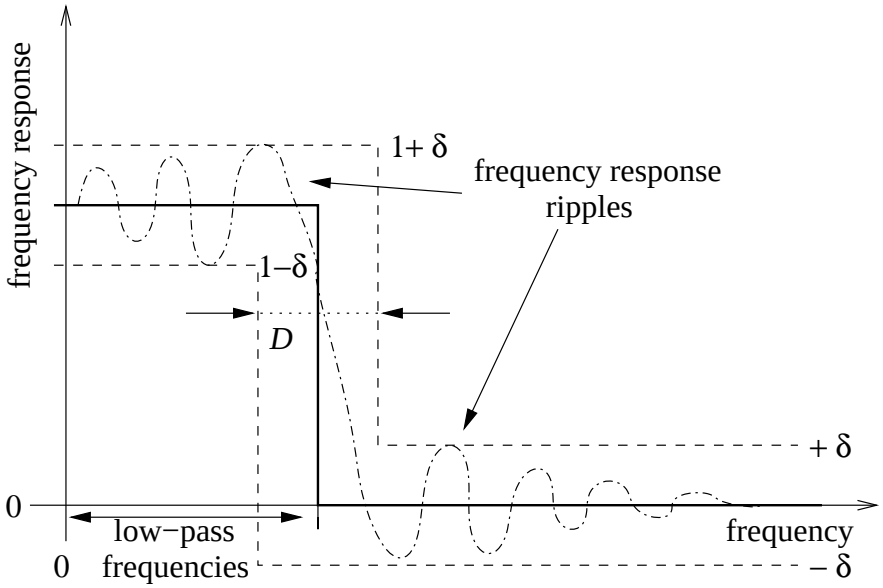


Figure 7.4: Illustration of the Kaiser window design method parameters, transition width D and ripple magnitude, δ in the frequency domain.

[37]), where a further parameter, the transition width, D, is required. These parameters are illustrated in figure 7.4. A compromise has to be made between transition width and ripple size. If a narrow transition width is required then it becomes more difficult to realise a filter with smaller ripples, i.e. a smoother frequency response. These properties are of particular interest to designers of systems where the theoretical frequency domain properties are important properties of their work. However, these frequency domain properties are not of particular interest here, except as a means to an end in the flexible design of a variety of spatial domain window shapes.

Algorithm 3 (Kaiser Filter Design Algorithm)

1.Define maximum pass band and stop band ripple sizes, r_p *and* r_s*;*
2.Determine $r = min(r_p, r_s)$*;*
3.Convert to decibels by calculating, $A = -20.\log_{10}(r)$*;*
4.Calculate: $a = \begin{cases} 0.1102.(A - 8.7) & \text{if } A \geq 50 \\ 0.5842.(A - 21)^{0.4} + 0.07886.(A - 21) & \text{if } 21 < A < 50 \\ 0 & \text{if } A \leq 21 \end{cases}$ *;*
5.Define a transition width, D*;*
6.Calculate the window width, $\Upsilon \geq \frac{A-7.95}{28.72.D}$.

This algorithm therefore enables design of a low-pass filter with a variety of frequency and signal domain properties. The parameters of the windows that were designed for this work can be seen in table 7.1, where the transition width, D, is expressed as a fraction of the sampling rate.

Table 7.1: Kaiser Window Parameter Values: Window Widths, Υ for a given ripple value, δ (fraction of maximum pass band gain) and transition width, D,(fraction of sample rate).

Ripple	Transition Widths, D					
δ	0.100	0.125	0.156	0.195	0.244	0.305
9.0E-02	9	7	6	5	4	3
3.0E-02	16	13	10	8	6	5
1.0E-02	22	18	14	11	9	7
3.3E-03	29	23	19	15	12	9
1.1E-03	36	28	23	18	15	12
3.7E-03	42	34	27	22	17	14
1.2E-04	49	39	31	25	20	16
4.1E-05	56	44	36	28	23	18
1.4E-05	62	50	40	32	25	20

For the first row in table 7.1, i.e. $\delta = 9.0\text{E}{-}02$, the filter function is equivalent to a boxcar [37]. A boxcar filter is often used in synthetic data, such as computer graphics or simulation of the PV effect [83]. The fifth row in table 7.1, i.e. $\delta = 1.1\text{E}{-}03$, is closest to a Hamming window, $w(n) = 0.54 + 0.46.\cos(\pi.n/\Upsilon)$, [37]. For the filter functions in rows 6 to 9 ($\delta > 1.1E - 03$), the Kaiser window can be thought to approximate a Gaussian window, which is a commonly assumed PSF in many imaging and signal processing applications. Figure 7.5 illustrates the Kaiser window function in the signal intensity domain, for the above three examples representing the boxcar, Hamming and Gaussian window functions.

Prior Density Evaluation

The densities evaluated were the ICG mixture prior density given by equation 7.8; the Benford mixture prior density given by equation 7.14; and for comparison, a density derived by Kitamoto and Takagi [66] referred to here as the Kitamoto and Takagi density. Kitamoto and Takagi reasoned that the pixels in satellite images could be described by squares, without any consideration of the PSF of the imaging system. The authors then commenced with a derivation based on the area occupied by the pixels over possible border regions (for this particular density). The Kitamoto and Takagi density is given by:

$$p_{Kita}(\alpha_a, \alpha_b) = -\frac{2.c^2}{(b+c)^2}.\left(\log(\alpha_a) + \log(\alpha_b)\right), \tag{7.16}$$

where Kitamoto and Takagi defined c as the size of a pixel and b as the size of the foreground object that has been subjected to a smoothing process. Note that this density also models the mixing variable as an explicit random vector in the same manner

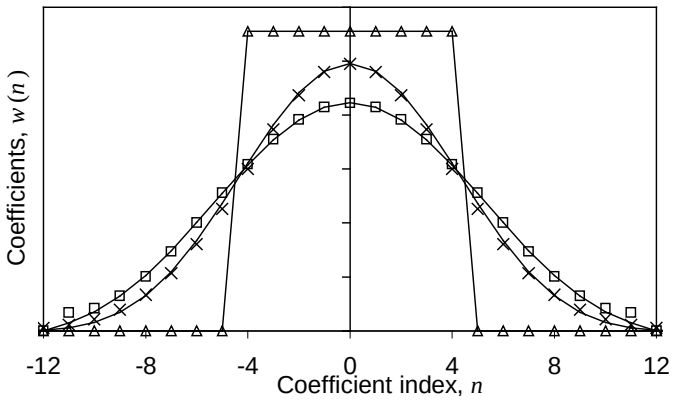


Figure 7.5: Three window functions (lines) and their approximations using the Kaiser window design method (points superimposed). The Boxcar window is given by the line joining the triangles; the Hamming window is given by the smooth line joining the squares; and a Gaussian window joining the crosses.

as has been proposed for the Benford and ICG densities. For this particular density, Kitamoto and Takagi asserted that the pixel size should be smaller than the foreground object size, i.e. $b > c$. An observation with regard to this density is the fact that the shape of the density is not affected by the parameters of the imaging process, i.e. c and b, similar to the densities proposed in this work (equations 7.8 and 7.14). However, as can be seen from figure 7.2, the densities possess different shapes in the tail regions and consequently different probabilities in the center region of each density, when fitted to filtered data. For the Kitamoto and Takagi density, this can be attributed to the fact that the density was derived without reference to a PSF that is intrinsic to the action of the image acquisition process.

7.1.4 Results

An exemplar set of data points drawn from histogram data smoothed by a Kaiser window with a transition width, $D = 0.305$ and maximum ripple size, $\delta = 1.4E - 05$, can be seen in figure 7.2. This figure also plots the three theoretical densities. The χ^2 GoF results seen in figure 7.6, illustrate that the ICG prior density possesses the lowest χ^2 errors with respect to the Kaiser filtered data for the majority of cases. One can also observe that the ICG error decreases as the filter functions tend towards a Gaussian window function, characteristic of the ICG density.

The Benford prior density also results in smaller χ^2 error terms when compared with the Kitamoto and Takagi prior density and appears to closely follow the ICG prior density χ^2 term, although every Benford error for the particular test data used is greater than the corresponding ICG error.

The GoF error for all three prior models significantly increases under two circumstances. The first circumstance being when the window filter function is significantly larger than the objects of interest in the data. The concentric spheroid bands in the simulated data have a finite width (30 voxels), and for particular sets of parameters for the Kaiser

window method, the window size is significantly larger than the widths of the bands (up to window widths of 62). These occur from the lower left of the windows in table 7.1, thus for these cases, the error is significantly greater.

The second circumstance for large GoF errors for all three models is when the window filter function becomes very small (less than 3 voxels wide). The result of a very small window is to cause digitisation effects on the resulting densities where a large number of the bins in the digital histograms are empty, thus contributing to a larger error.

7.1.5 Discussion

A methodology for modelling a density from the signal intensity domain PSF has been presented. This has been illustrated with a Gaussian PSF model resulting in improved χ^2 GoF for PSFs that approach a Gaussian over two other approaches. This investigation has also demonstrated that the Benford prior density can be used as a plausible description of the mixing that results from a signal acquisition process characterised by a finite window width. Each of the theoretical prior mixing densities possess functional shape invariance given different sized window functions.

This investigation into theoretical prior densities for the PV effect has explicitly examined cases where: the width of the kernel for the acquisition process is smaller than the objects of interest affected by the mixing process; and the PSF can be approximated by one of the window filters designed using the Kaiser window method.

This new information about the form of the global prior distribution will now be used in the simulation of a PV model together with a locally defined PV mixing prior.

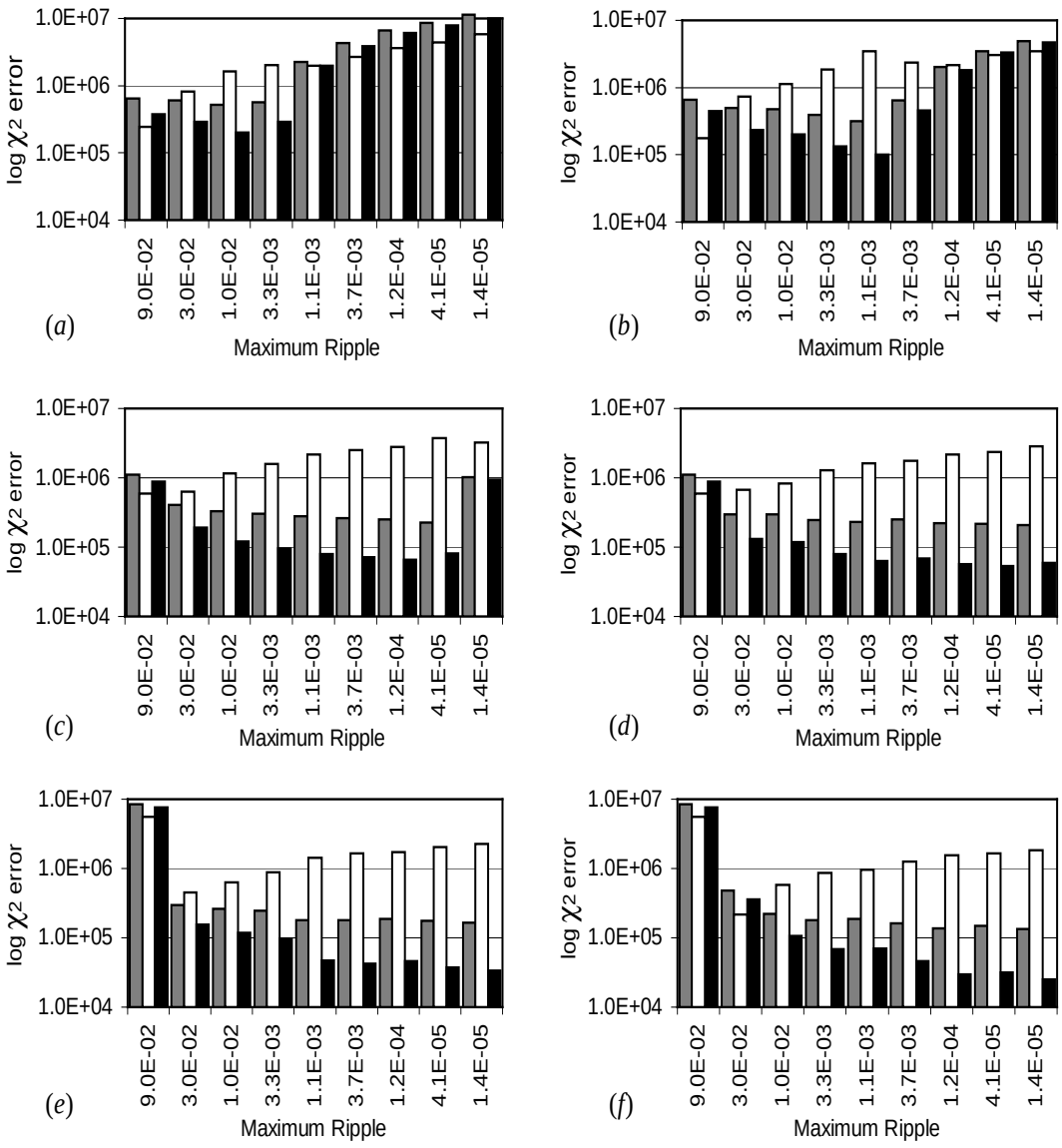


Figure 7.6: χ^2 GoF results for Benford (gray), Kitamoto and Takagi (white) and ICG (black) prior mixture densities when compared with the experimental mixture densities obtained by convolving simulated volumetric data sets with window functions designed using the Kaiser window design methodology with parameters given in table 7.1. (a) is for transition width, D=0.100, (a Boxcar function); (b) is for transition width, D=0.125; (c) is for transition width, D=0.156; (d) is for transition width, D=0.195; (e) is for transition width, D=0.244, similar to a Hamming window; and (f) is for transition width, D=0.305, similar to a Gaussian function.

7.2 A Contextual PV Prior

Often data samples may have what is often known as *context*. For spatially located data samples this context is space. Intuitively, if a data sample in a particular location is classified with a particular class label then any data samples located near to that data sample are likely to have a similar class label. This prior knowledge is often useful in improving the classification performance of otherwise ambiguous data samples. Also, if the joint likelihood for a number of data samples is calculated, then the value of this joint likelihood may not be fully specified if the contextual information is not included.

As discussed in the conclusions of chapter 6, further model developments are required to improve the classification performance of the PV classifier that has been developed thus far. The available spatially derived contextual information is one possible type of information that could be harnessed to improve classification performance and this is the topic of this part of the thesis.

This work demonstrates that the contextual information can be correctly modelled and incorporated into the classifier developments described in chapters 5 and 6.

7.3 Model G: Methodology

In chapter 6 a probabilistic description using Bayes theorem of the PV content of a voxel, $\boldsymbol{\alpha}$, for an intensity value, g and locally calculated gradient magnitude, z, was given by:

$$p(\boldsymbol{\alpha}|g,z) = \frac{p(g|\boldsymbol{\alpha}).p(z|\boldsymbol{\alpha}).p(\boldsymbol{\alpha})}{p(g,z)}, \tag{7.17}$$

where it was assumed that g and z are independent. This formulation was used to determine the expected mixture content for a particular intensity level, g and local gradient magnitude, z in equations 6.31 (2-D) and 6.52 (3-D).

A further variable, $\boldsymbol{\alpha}_{\mathfrak{N}_{\boldsymbol{\omega}_i}}$ is also now considered that models the mixing in the neighbourhood of a particular voxel, $\boldsymbol{\omega}_i$. Instead of adapting equation 7.17, the posterior density of the mixing vector, $\boldsymbol{\alpha}$, (ignoring the voxel specific notation), is derived by re-considering the basic Bayesian formulation. i.e. The posterior density of the mixing for a particular voxel, $\boldsymbol{\alpha}$ via Bayes theorem is given by

$$p(\boldsymbol{\alpha}|g,z,\boldsymbol{\alpha}_{\mathfrak{N}}) = \frac{p(g|\boldsymbol{\alpha},z,\boldsymbol{\alpha}_{\mathfrak{N}}).p(\boldsymbol{\alpha},z,\boldsymbol{\alpha}_{\mathfrak{N}})}{p(g,z,\boldsymbol{\alpha}_{\mathfrak{N}})}, \tag{7.18}$$

where the reader should note, $p(g,\boldsymbol{\alpha},z,\boldsymbol{\alpha}_{\mathfrak{N}})$ is equal to $p(g|\boldsymbol{\alpha},z,\boldsymbol{\alpha}_{\mathfrak{N}}).p(\boldsymbol{\alpha},z,\boldsymbol{\alpha}_{\mathfrak{N}})$ *or* $p(\boldsymbol{\alpha}|g,z,\boldsymbol{\alpha}_{\mathfrak{N}}).p(g,z,\boldsymbol{\alpha}_{\mathfrak{N}})$. Assuming little direct conditional dependence of g on z and g on $\boldsymbol{\alpha}_{\mathfrak{N}}$ due to the indicating mixing vector, $\boldsymbol{\alpha}$:

$$\begin{aligned} p(\boldsymbol{\alpha}|g,z,\boldsymbol{\alpha}_{\mathfrak{N}}) &= \frac{p(g|\boldsymbol{\alpha}).p(\boldsymbol{\alpha}_{\mathfrak{N}}|z,\boldsymbol{\alpha}).p(z,\boldsymbol{\alpha})}{p(g).p(z,\boldsymbol{\alpha}_{\mathfrak{N}})}, \\ &= \frac{p(g|\boldsymbol{\alpha}).p(\boldsymbol{\alpha}_{\mathfrak{N}}|z,\boldsymbol{\alpha}).p(\boldsymbol{\alpha})}{p(g).p(\boldsymbol{\alpha}_{\mathfrak{N}}|z)}; \end{aligned} \tag{7.19}$$

so that the probability of the mixing for a particular voxel with contextual information no longer requires the gradient magnitude densities discussed in chapter 6.

The only unknown term in the numerator of equation 7.19 is the localised prior mixing term, $p(\boldsymbol{\alpha}_{\mathfrak{N}}|z,\boldsymbol{\alpha})$, which, as will be seen shortly, is well described by a Gaussian form and thus possesses symmetry where $p(\boldsymbol{\alpha}_{\mathfrak{N}}|z,\boldsymbol{\alpha}) = p(\boldsymbol{\alpha}|z,\boldsymbol{\alpha}_{\mathfrak{N}})$. So that $p(\boldsymbol{\alpha}|z,\boldsymbol{\alpha}_{\mathfrak{N}})$ can be used interchangeably with $p(\boldsymbol{\alpha}_{\mathfrak{N}}|z,\boldsymbol{\alpha})$. Conveniently, $p(\boldsymbol{\alpha}|z,\boldsymbol{\alpha}_{\mathfrak{N}})$ is dependent on the neighbourhood information *and* the gradient magnitude.

7.3.1 Form of the Localised Mixing Prior

The neighbourhood information may be succinctly described by the vectorial mean of the mixture values of the neighbouring voxels to point $\boldsymbol{\omega}_i$:

$$\boldsymbol{\alpha}_{\mathfrak{N}_{\boldsymbol{\omega}_i}} = \frac{1}{|\mathfrak{N}_{\boldsymbol{\omega}_i}|} \sum_{\boldsymbol{\omega}_j \in \mathfrak{N}_{\boldsymbol{\omega}_i}} \boldsymbol{\alpha}_{\boldsymbol{\omega}_j} \tag{7.20}$$

where $|\mathfrak{N}_{\boldsymbol{\omega}_i}|$ is the cardinality of the set of neighbours to point $\boldsymbol{\omega}_i$:

$$\mathfrak{N}_{\boldsymbol{\omega}_i} = \{\boldsymbol{\omega}_j | \forall\ \boldsymbol{\omega}_j \text{ that are neighbours of voxel } \boldsymbol{\omega}_i\} \tag{7.21}$$

i.e. $\mathfrak{N}_{\boldsymbol{\omega}_i}$ is a set of points located within a certain distance, $\mathcal{C}_D$, of $\boldsymbol{\omega}_i$. The squared difference between this mean and the actual voxel's label is equivalent to the sum over what is known as the clique potentials, $V_{\boldsymbol{\omega}_i,\boldsymbol{\omega}_j}$:

$$\sum_{\boldsymbol{\omega}_j \in \mathfrak{N}_{\boldsymbol{\omega}_i}} V_{\boldsymbol{\omega}_i,\boldsymbol{\omega}_j} = \sum_{\boldsymbol{\omega}_j \in \mathfrak{N}_{\boldsymbol{\omega}_i}} \kappa_j.(\boldsymbol{\alpha}_{\boldsymbol{\omega}_j} - \boldsymbol{\alpha}_{\boldsymbol{\omega}_i})^2 \equiv C.(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i} - \boldsymbol{\alpha}_{\mathfrak{N}_{\boldsymbol{\omega}_i}})^2, \tag{7.22}$$

where C is a constant. A similar clique potential, $V_{\boldsymbol{\omega}_i,\boldsymbol{\omega}_j}$ was previously seen in chapter 4, equation 4.14 (using notation, V_{ij}), where the seminal paper on probabilistic PV modelling was described, published by Choi *et al.* in 1991, [24]. The clique potential used here is the same as the clique potential used by Woolrich *et al.* in 2005, [162], which was also previously described in chapter 4, on page 76.

PV models incorporating local mixture information using this clique potential or similar have previously utilized a Gaussian distribution, [24, 162] as the PV mixing prior distribution. The Gaussian distribution can be seen as a specific instance of the Gibbs distribution when using this particular clique potential. The Gibbs distribution was given in chapter 4 by:

$$P(\boldsymbol{x}) = \frac{1}{Z} \exp\left(-H(\boldsymbol{x})\right) \tag{7.23}$$

where $H(\boldsymbol{x})$ is known as the energy function, $\boldsymbol{x}$ refers to the state of the voxels, which for PV voxels, will be given by the mixing vectors, $\boldsymbol{\alpha}$ and Z normalises the Gibbs distribution and is known as the partition function. The energy function, $H(\boldsymbol{x})$ measures the amount of disparity of the state of the voxels, $\boldsymbol{x}$. So that if the PV mixing prior density is locally defined as:

$$p(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|\boldsymbol{\alpha}_{\mathfrak{N}_{\boldsymbol{\omega}_i}}, \sigma^2_{\mathfrak{N}_{\boldsymbol{\omega}_i}}) = \frac{1}{(2.\pi.\sigma^2_{\mathfrak{N}_{\boldsymbol{\omega}_i}})^{K/2}}. \exp\left(-\frac{1}{2.\sigma^2_{\mathfrak{N}_{\boldsymbol{\omega}_i}}} \sum_{\forall v} (\alpha_{\boldsymbol{\omega}_i,v} - \alpha_{\mathfrak{N}_{\boldsymbol{\omega}_i},v})^2\right), \tag{7.24}$$

where $\boldsymbol{\alpha}_{\boldsymbol{\omega}_i} = (\alpha_{\boldsymbol{\omega}_i,1}\ \alpha_{\boldsymbol{\omega}_i,2}\ ...\ \alpha_{\boldsymbol{\omega}_i,K})^T$ and $\boldsymbol{\alpha}_{\mathfrak{N}_{\boldsymbol{\omega}_i}} = (\alpha_{\mathfrak{N}_{\boldsymbol{\omega}_i},1}\ \alpha_{\mathfrak{N}_{\boldsymbol{\omega}_i},2}\ ...\ \alpha_{\mathfrak{N}_{\boldsymbol{\omega}_i},K})^T$, so that there are n classification classes and v in equation 7.24 is therefore limited to $1 \leq v \leq n$. $\sigma_{\mathfrak{N}_{\boldsymbol{\omega}_i}}$ is the standard deviation of the distribution that can be used to control the amount of spatial regularisation. Most authors, such as Choi *et al.*, [24], utilize a regularisation parameter, which takes the form of (for the entire data)

$$\beta \propto \frac{1}{\prod\limits_{\forall \boldsymbol{\omega}_i} \sigma^2_{\mathfrak{N}_{\boldsymbol{\omega}_i}}} \tag{7.25}$$

but for the Gaussian form of the Gibbs distribution, a standard deviation parameter is more appropriate. Assuming this Gaussian form, the prior distribution would then result in a Gaussian instance of the Gibbs distribution for the entire data, given by:

$$p(\mathbf{A}|\Sigma_{\mathfrak{N}}) = \prod_{\forall \boldsymbol{\omega}_i} p(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|\boldsymbol{\alpha}_{\mathfrak{N}_{\boldsymbol{\omega}_i}}, \sigma^2_{\mathfrak{N}_{\boldsymbol{\omega}_i}}), \tag{7.26}$$

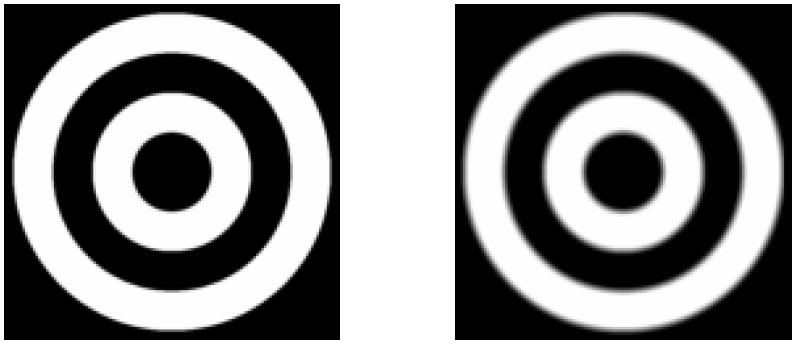
where $\mathbf{A} = \{\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|\boldsymbol{\omega}_i \in \boldsymbol{\Omega}\}$, $\Sigma_{\mathfrak{N}} = \{\sigma_{\mathfrak{N}_{\boldsymbol{\omega}_i}}|\boldsymbol{\omega}_i \in \boldsymbol{\Omega}\}$ and $\boldsymbol{\Omega}$ is the set of all points in the image space. The conditional independence of the mixture vectors from all other mixture vectors in the image space, except the mixture vectors of the neighbouring voxels is a property of the Gibbs distribution. This was illustrated earlier, in chapter 4.

As previously stated, Choi *et al.* [24] noted the Gaussian like distribution of the data with respect to the clique potential described in equation 4.14. Choi *et al.* also confirmed this observation experimentally with a Quantile-Quantile (QQ) plot. A QQ plot is a scatter plot of frequency components of two Cumulative Distribution Functions (CDF)s, thus enabling the similarity of the two CDFs to be observed graphically. The Quantile refers to a fraction of the CDF probability and a series of these quantiles for the two CDFs (theoretical and experimental) should therefore present a linear relationship if the two distributions are equivalent.

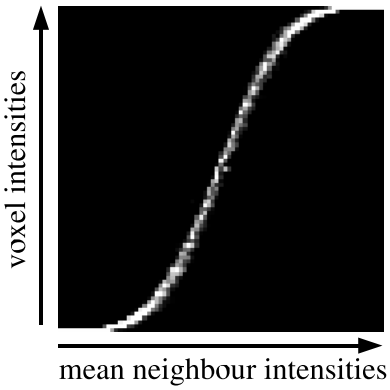
Woolrich *et al.* [162] also observed that their neighbourhood measure can also be approximated with what is known as a conditionally specified Gaussian model, as described in, for example, [8, 30].

To confirm the Gaussian nature of the measure in equation 7.20, an exemplar scatter plot of the various values of $\boldsymbol{\alpha}$ for particular values of $\boldsymbol{\alpha}_{\mathfrak{N}}$ can be seen in figure 7.7(c). Following the procedure adopted by Choi *et al.* in 1991, a QQ plot is also produced for the data in the scatter plot also given in figure 7.7(d). The QQ plot illustrates a good agreement between a Gaussian form for the locality measure in equation 7.20.

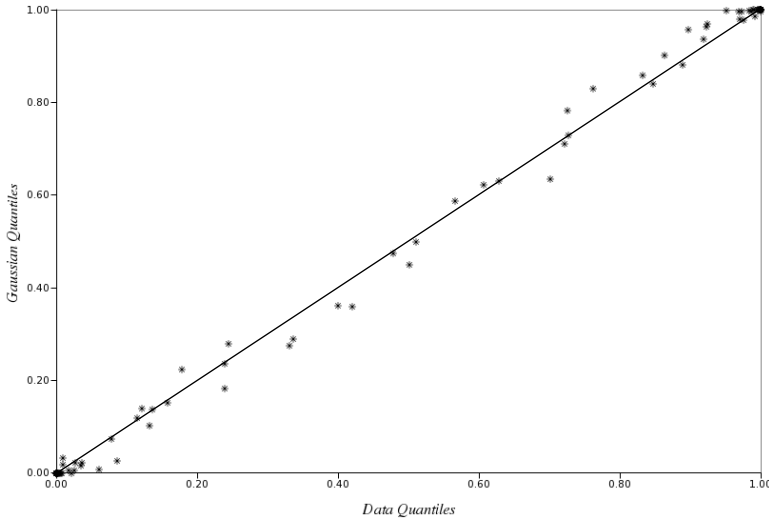
Choi *et al.*, [24], determined the optimal value of their variance measure by the value that maximizes the log-likelihood function of their image model. Woolrich *et al.*, [162], assigned a prior density to describe their variance measure that was subsequently incorporated into the image model, therefore enabling the variance measure to be adapted to the data. One should note however, the heuristic manner in which Woolrich *et al.* selected their prior density, apparently based on conjugacy rather than observation. A conjugate density in Bayesian type methods relies on selecting a density that results in a convenient form of the posterior, to assist in sampling or analytical point estimation methods, e.g. see [77].



(a) (b)



(c)



(d)

Figure 7.7: (a)Exemplar slice through simulated PV data, with mean values 960 and 1040; (b) Exemplar slice through the volume consisting of mean values of neighbouring voxel intensities of the volume illustrated in (a); (c) Scatter plot of (a) versus (b), i.e. ground truth voxel intensities versus mean of neighbouring ground truth voxel intensities; (d) A Quantile to Quantile plot for simulated PV data (y-axis) versus Gaussian functions (x-axis). (d) helps to illustrate that a Gaussian function is a good description of the variation of ground truth intensities as a function of the mean of the neighbouring ground truth intensities.

Choi *et al.*, [24], set their variance measure to be constant for the entire image data. This is equivalent to asserting that there should be an equal amount of spatial regularisation through out the image data, i.e. homogeneity of the imaging space. At the boundaries of classification components such a condition is not optimal. Equation 7.24 describes the variation of the mixture mean as a function of the mean of the local mixture means, without the gradient magnitude measure as required by $p(\boldsymbol{\alpha}|\boldsymbol{\alpha}_{\mathfrak{N}}, z)$ in equation 7.19. As the homogeneity assumption is less valid at classification component boundaries, together with the idea that gradient magnitude varies at a boundary, a scatter plot was created to determine whether there is a functional relationship between $\sigma_{\mathfrak{N}}$ and z. This is illustrated in figure 7.8(e) using simulated PV data.

Figure 7.8(e) illustrates that there is a functional relationship between $\sigma_{\mathfrak{N}}$ and z. Interestingly many data points with high gradient magnitude values actually possess smaller standard deviation values. This can be confirmed with close inspection of the exemplar standard deviation image slice. This is due to the form of the locality measure. The locality measure, equation 7.20, is an averaging operation that results in little variation for a voxel that is located in the middle of a slope such as a boundary between classification classes, composed of PV voxels. This situation is depicted in figure 7.9.

Therefore, the probability of obtaining a particular mixture, $\boldsymbol{\alpha}$, is a function of the local variation measure, $\boldsymbol{\alpha}_{\mathfrak{N}}$ and the variance of this measure, $\sigma_{\mathfrak{N}}(z)^2$ is now a function of the gradient magnitude. This results in equation 7.24 being modified to

$$p(\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}|\boldsymbol{\alpha}_{\mathfrak{N}_{\boldsymbol{\omega}_i}}, \sigma_{\mathfrak{N}_{\boldsymbol{\omega}_i}}(z)^2) = \frac{1}{(2.\pi.\sigma_{\mathfrak{N}_{\boldsymbol{\omega}_i}}(z)^2)^{K/2}}.\exp\left(-\frac{\sum\limits_{\forall v}(\alpha_{\boldsymbol{\omega}_i,v} - \alpha_{\mathfrak{N}_{\boldsymbol{\omega}_i},v})^2}{2.\sigma_{\mathfrak{N}_{\boldsymbol{\omega}_i}}(z)^2}\right), \quad (7.27)$$

which satisfies the form of the local mixing prior as used by equation 7.19. $\sigma_{\mathfrak{N}}(z)$ was found to be well modelled by a Beta density function, details of which can be seen in section 7.4.

The posterior density of $\boldsymbol{\alpha}$ can now be defined, taking the form from equation 7.19

$$p(\boldsymbol{\alpha}|g, z, \boldsymbol{\alpha}_{\mathfrak{N}}, \sigma_{\mathfrak{N}}(z)^2) = \frac{p(g|\boldsymbol{\alpha}).p(\boldsymbol{\alpha}|\boldsymbol{\alpha}_{\mathfrak{N}}, \sigma_{\mathfrak{N}}(z)^2)}{p(g)} \quad (7.28)$$

where

$$p(g|\boldsymbol{\alpha}) = \frac{1}{\sqrt{2.\pi.\sigma_{\boldsymbol{\alpha}}^2}}.\exp\left(-\frac{(g-\mu_{\boldsymbol{\alpha}})^2}{2.\sigma_{\boldsymbol{\alpha}}^2}\right), \quad (7.29)$$

and recall from chapter 5, sub-section 5.1.3, the mixture mean is given by $\mu_{\boldsymbol{\alpha}} = \sum\limits_{\forall v} \mu_v.\alpha_v$ and the mixture variance is given by $\sigma_{\boldsymbol{\alpha}}^2 = \sum\limits_{\forall v} \sigma_v^2.\alpha_v^2$, where v is indicative of a particular classification class. Thus, the optimal mixture vector (for a particular voxel, $\boldsymbol{\omega}_i$) can also be determined using the expected mixture values w.r.t. the posterior distribution given in equation 7.28:

$$\boldsymbol{\alpha}_{\boldsymbol{\omega}_i,\text{opt}} = \begin{pmatrix} \mathbf{E}\left[\alpha_{\boldsymbol{\omega}_i,1}|g_{\boldsymbol{\omega}_i}, z_{\boldsymbol{\omega}_i}, \alpha_{\mathfrak{N}_{\boldsymbol{\omega}_i}}\right] \\ \mathbf{E}\left[\alpha_{\boldsymbol{\omega}_i,2}|g_{\boldsymbol{\omega}_i}, z_{\boldsymbol{\omega}_i}, \alpha_{\mathfrak{N}_{\boldsymbol{\omega}_i}}\right] \\ \cdots \\ \mathbf{E}\left[\alpha_{\boldsymbol{\omega}_i,3}|g_{\boldsymbol{\omega}_i}, z_{\boldsymbol{\omega}_i}, \alpha_{\mathfrak{N}_{\boldsymbol{\omega}_i}}\right] \end{pmatrix},$$

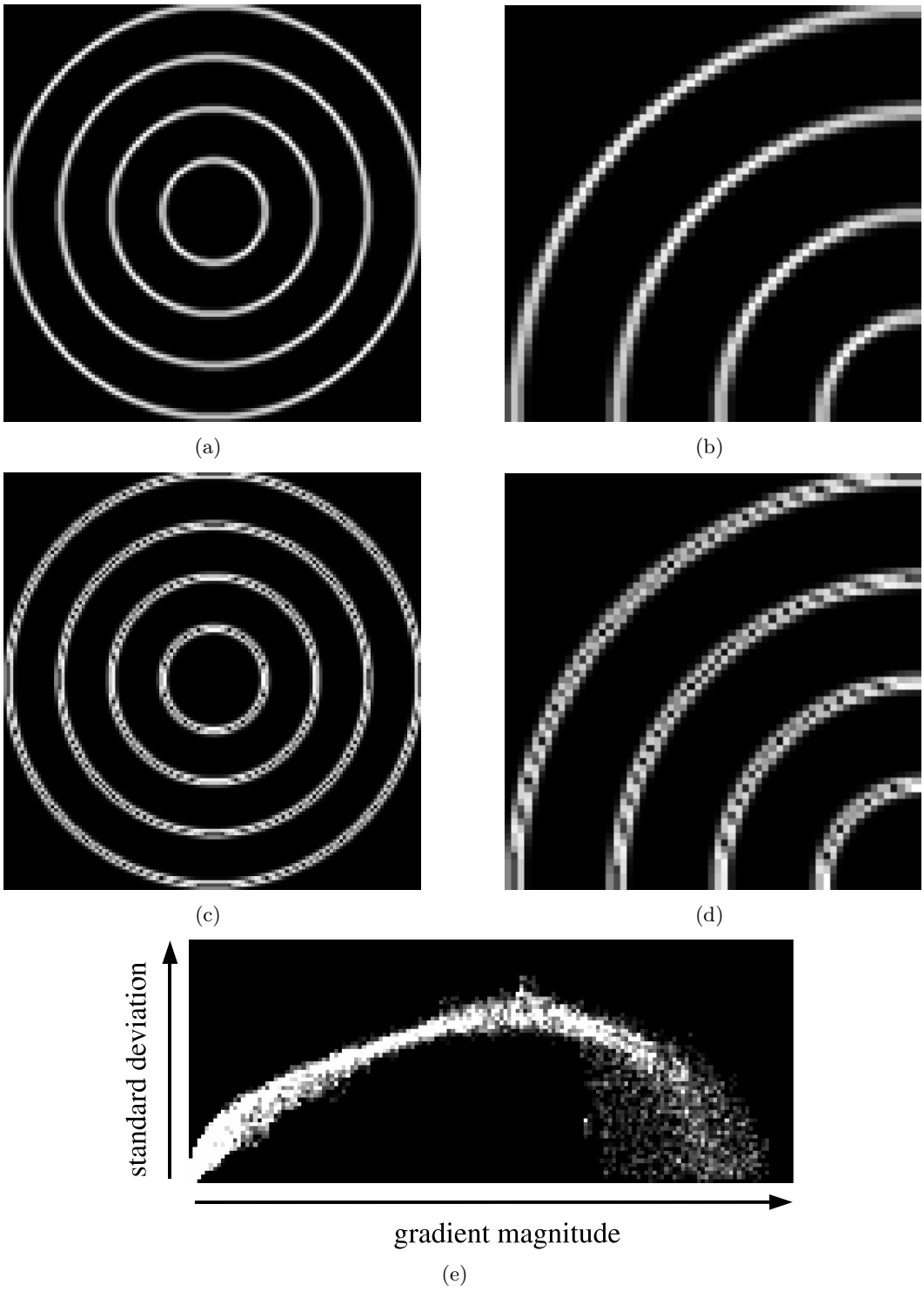


Figure 7.8: (a & b) are an exemplar slice and section of the 3-D gradient magnitude volume calculated from simulated PV data; (c & d) are an exemplar slice and section of the standard deviation data volume, where each pixel corresponds to a local value of $\sigma_{\mathfrak{N}}$ in equation 7.24; (e) is a gradient magnitude versus standard deviation scatter plot of the data shown in (a) and (c), illustrating that $\sigma_{\mathfrak{N}}$ varies as a function of z.

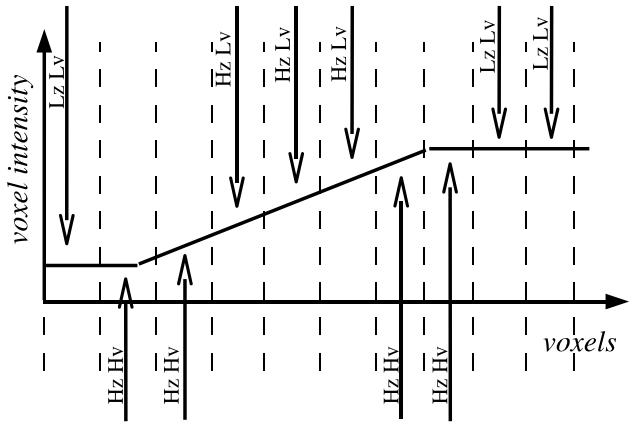


Figure 7.9: Illustration of the occurrence and location of high gradient (Hz) but low variance (Lv) voxels (Hz Lv). High Variation (Hv) voxels occur at the boundary of classification class components that are characterised by Low gradient (Lz) and Lv measurements. This figure helps to explain the gradient magnitude versus standard deviation scatter plot in figure 7.8 where for very high gradient magnitude measurements, many voxels tend to have low variation.

where the mixture content, $\alpha_{\boldsymbol{\omega}_i,v}$, for voxel $\boldsymbol{\omega}_i$ and classification class v is given by:

$$\mathbf{E}\left[\alpha_{\boldsymbol{\omega}_i,v}|g_{\boldsymbol{\omega}_i}, z_{\boldsymbol{\omega}_i}, \alpha_{\Re_{\boldsymbol{\omega}_i}}\right] = \int_0^1 \alpha_{\boldsymbol{\omega}_i,v}.p(\alpha_{\boldsymbol{\omega}_i,v}|g_{\boldsymbol{\omega}_i}, z_{\boldsymbol{\omega}_i}, \alpha_{\Re_{\boldsymbol{\omega}_i}}).d\alpha_{\boldsymbol{\omega}_i,v}. \tag{7.30}$$

7.3.2 Simulation of the Posterior Distribution

A method known as Riemann sums (see e.g. [113]) was utilized to evaluate the expectation integrals of the models in chapters 5 and 6. The method of Riemann sums divides the region of the integrand into a finite number of bins. For each bin, the function being integrated is evaluated and then used to approximate the integral. Riemann sums and other similar numerical integration type techniques can be slow and the results of which may be inaccurate, especially for integrations across many dimensions, as would be the case for multiple tissue or activity classes. Therefore an alternative technique is used that utilises Markov Chain Monte Carlo (MCMC) to simulate the posterior distribution from which expected values may be calculated numerically.

MCMC methods are different in that they rely on random samples being generated, so that many of the samples are generated in regions where the function being simulated is concentrated, therefore simulating a density of interest. This results in fewer samples having to be generated and can lead to more accurate integration results (when utilizing the simulation for integration) in comparison to the Riemann sums method which entails evaluating the function at regular intervals. MCMC methods are therefore suitable for high-dimensional sampling problems, such as is required for the evaluation of the conditional expectation in equation 7.30.

A type of particularly useful MCMC algorithm in Bayesian problems is known as the Independent Metropolis-Hastings (IMH) algorithm (see e.g. [41, 45, 113]). The algorithm can be stated as follows:

Algorithm 4 (Simulation via Independent Metropolis-Hastings (IMH))

1. *Given the previous simulated value,* $x^{(t-1)}$
   2. *Generate a new sample from the proposal density,* $X^\star \sim g(x)$
   3. *Calculate the acceptance ratio,* $R(x^{(t-1)}, X^\star) = \frac{f(x=X^\star)).g(x=x^{(t-1)})}{f(x=x^{(t-1)}).g(x=X^\star)}$
   4. *If the acceptance ratio,* $R > 1$ *then let* $R = 1$
   5. *Generate an instance of a uniform random variable, i.e.* $U \sim Uniform(0, 1)$
   6. *If* $R < U$ *then accept the new sample,* $x^{(t)} = X^\star$*, otherwise* $x^{(t)} = x^{(t-1)}$
   7. *Set* $t = t + 1$
8. *Return to step 1.*

The algorithm samples from a density known as the proposal (or instrumental) density, $g(x)$. These samples, $X^\star$ have to then be evaluated as possibly coming from the target density that is being simulated, $f(x)$. The evaluation is performed with an acceptance ratio, $R(,)$. If the acceptance ratio produces a value greater than 1, then the new sample is deemed to be a valid sample from the target density. If its value is less than 1 then a uniform random number, U is generated and this is used to determine whether the sample is accepted or not. The accepted samples form what is known as a Markov Chain (MC) and converge to the target density. This algorithm can be understood intuitively from the idea that a sample is generated from a proposal density that resembles the target density. If that sample has a higher probability than the earlier accepted sample then it is highly likely that it could equally have been directly sampled from the target density. For a more theoretical coverage of MCMC algorithms, see texts such as [113].

7.3.3 Choice of Proposal Distribution

The proposal distribution should reflect the shape of the posterior distribution that is being simulated. As already mentioned this enables samples to be concentrated in regions where the density is most likely to possess higher probabilities.

Therefore recall that the global two class mixture prior probability was found to be well described by equation 7.8, repeated here for convenience,

$$p_{Gauss}(\alpha_{a,b}) = C.\exp\left(\text{erf}^{-1}\left(2.\alpha_{a,b} - 1\right)^2\right),$$

where C is a normalizing constant. A number of observations and suggestions are now made.

- The Beta density for particular parameter values can provide a reasonable fit to the global mixture prior density (equation 7.8), as has previously been found by other authors for mixture distributions in non-medical images [56, 66, 156]. See figure 7.10 for an illustrative fit of a Beta density to the analytically derived prior density given by equation 7.8;

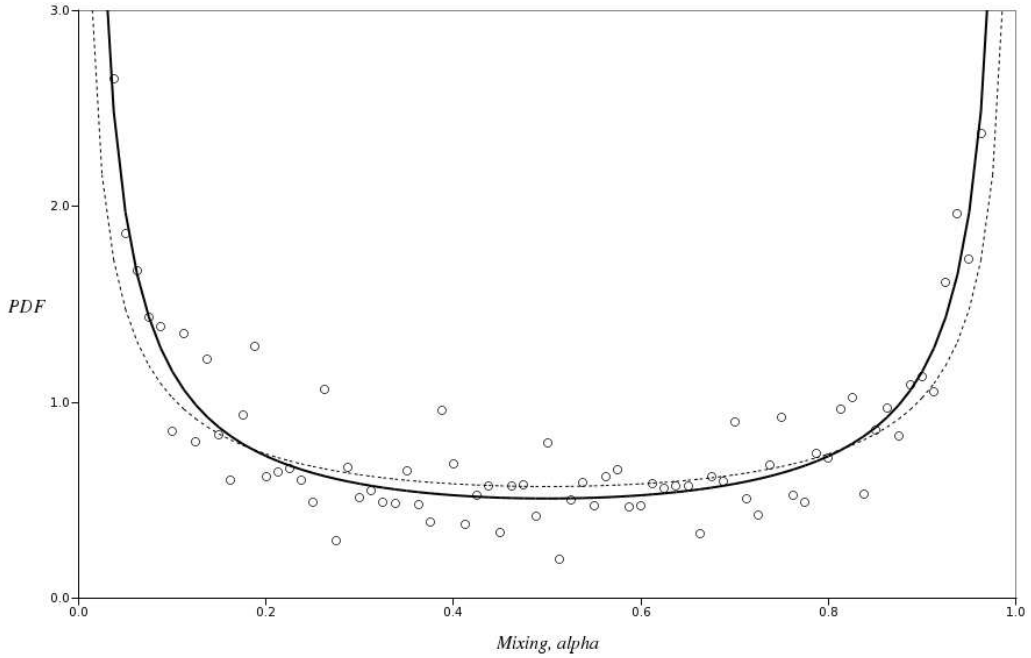


Figure 7.10: Illustration of fitting a Beta density (light dashed line) to the analytically derived mixture density of equation 7.8 (heavy line). Also illustrated are some exemplar data points (circles) from ground truth simulated PV data.

- The Beta density possesses a multi-variate extension, known as the Dirichlet density, thus providing a means for modelling more than two component mixtures, as was also used for non-medical images in [56];
- Computationally efficient methods exist that enable sampling from the Dirichlet density. These therefore enable more advanced numerical Bayesian techniques, such as MCMC algorithms to be implemented.

The Beta and Dirichlet densities also offer the advantage of possessing a property that constrains the random variable to pre-defined intervals. This is a requirement of the PV mixing variables, where the interval is $[0, 1]$ for all classification class mixing variables.

7.4 Model G: Implementation

7.4.1 The Form of the Neighbourhood Variance as a Function of Gradient Magnitude

Recall that the neighbourhood variance or standard deviation of equation 7.27 is well modelled by a Beta density function. The exact form and selection of parameters is now discussed. By calculating the mean neighbourhood standard deviation for a given value of gradient magnitude, data points such as the ones illustrated in figure 7.11 were obtained. With these data points a Beta density function was fitted using the Nelder-Mead Simplex algorithm, [92], (described in appendix E). The Beta density function

is defined as (also illustrated in figure 7.11)

$$\sigma_{\mathfrak{N}}(z) = \frac{\gamma_5.\Gamma(\gamma_1+\gamma_2)}{\Gamma(\gamma_1).\Gamma(\gamma_2).(\gamma_4-\gamma_3)^{\gamma_1+\gamma_2-1}}(z-\gamma_3)^{\gamma_1-1}(\gamma_4-z)^{\gamma_2-1} \tag{7.31}$$

where γ_4 and γ_3 define the extent of the function, so $\gamma_3 = 0$ and γ_4 is given by the maximum gradient magnitude. γ_5 is a scaling parameter and can be calculated from the area under the data points. The only remaining parameters to be estimated were γ_1 and γ_2 which define the *shape* of the Beta density (see figure 4.3 for a set of exemplar possible Beta density *shapes*). Using the Nelder-Mead Simplex optimisation algorithm it was found $\gamma_1 = 2.4$ and $\gamma_2 = 1.5$, so that equation 7.31 reduces to

$$\sigma_{\mathfrak{N}}(z) = \frac{4.814.\gamma_5}{\gamma_4^{2.9}} z^{1.4}.\sqrt{\gamma_4 - z} \tag{7.32}$$

where the scale, γ_5 and extent, γ_4 are approximated from the maximum ideal gradient magnitude in the data and the distance between the two most populous means in a region of interest surrounding the voxel of interest. In practise, for reasons of stability, the gradient magnitude has to be calculated from the noisy data. It was found that calculating the gradient magnitude from the estimated mixture values prevented the algorithm from converging due to the way in which the gradient magnitude was used to regulate the amount of spatial smoothness. Therefore z calculated from the noisy data may become greater than γ_4. To avoid this the actual equation used is given by:

$$\sigma_{\mathfrak{N}}(z) = \frac{4.814.\gamma_5}{\gamma_4^{2.9}} z^{1.4}.\sqrt{||\gamma_4 - z||}, \tag{7.33}$$

where $||x||$ represents the absolute value of the arbitrary variable x. Improved performance was also found with a Sobel gradient magnitude kernel due to the additional smoothing provided by the Sobel kernel, (see e.g. [94]).

7.4.2 Classifier Algorithm

The classification procedure is outlined in figure 7.12. The algorithm continues until nearly as few changes in the mixture means occur from one iteration to the next. This was measured by a RMS measure

$$\mathfrak{D}^{(a)} = \sqrt{\frac{1}{N}\sum_{\forall \boldsymbol{\omega}_i}\left(\mu^{(a)}_{\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}} - \mu^{(a-1)}_{\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}}\right)^2} \tag{7.34}$$

where N is the number of voxels being classified and $\mu^{(a)}_{\boldsymbol{\alpha}_{\boldsymbol{\omega}_i}}$ is the mixture mean given by $\sum_{\forall v} \mu_v.\alpha_v$, (see page 171), at iteration (a) and voxel $\boldsymbol{\omega}_i$. The iterations cease at an iteration (a') when $\left(\mathfrak{D}^{(a')} - \mathfrak{D}^{(a'-1)}\right) < \epsilon$ where ϵ was set to 1, although other more rigorously defined values may also be selected.

For each voxel, the mixture value is calculated over 60 successful (accepted) samples. At the time of implementation, this condition was found to be programmatically simpler instead of the usual MCMC approach where the length of the chain is defined on the number of accepted and unaccepted iterations.

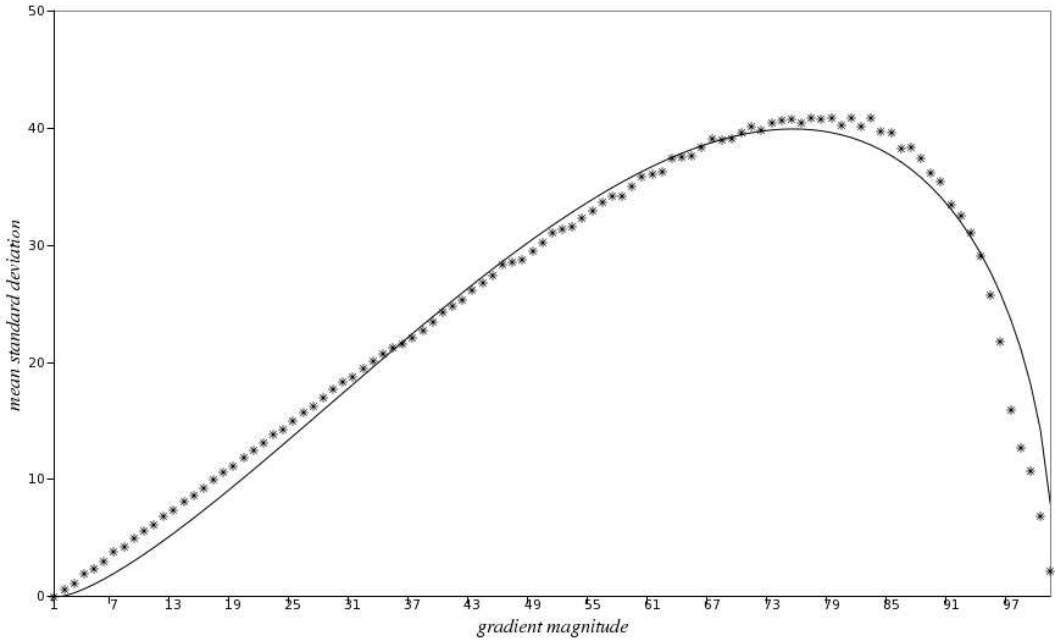


Figure 7.11: Exemplar illustration of mean standard deviation as a function of gradient magnitude data points ($*$) with a Beta density (equation 7.31) used to approximate the functional relationship between the mean standard deviation and the gradient magnitude (line). A Sobel gradient magnitude kernel was used to calculate the gradient magnitude values, (see e.g. [94]).

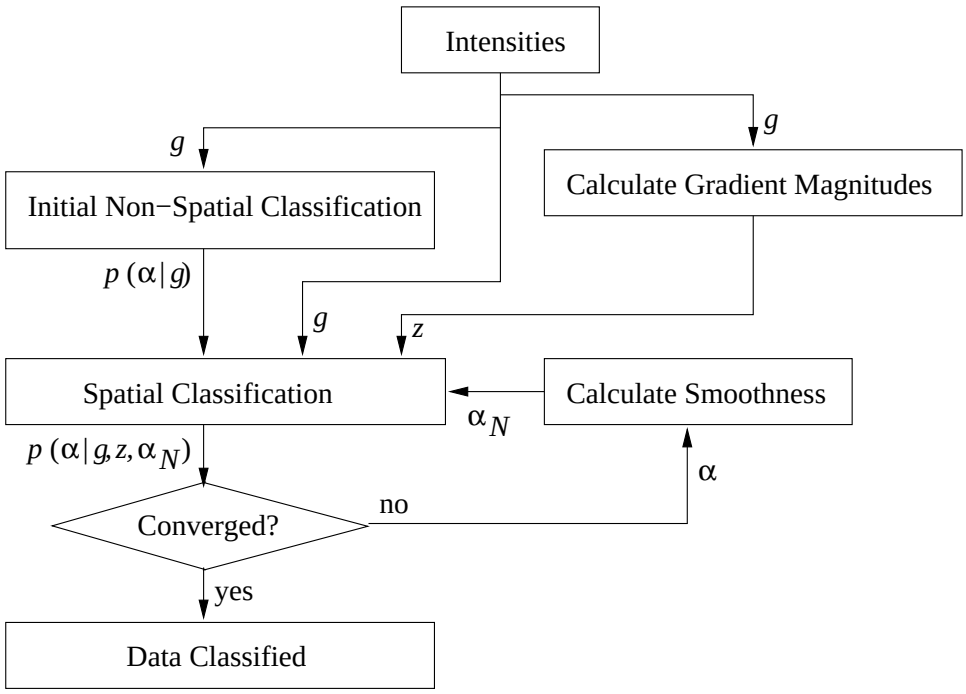


Figure 7.12: Illustration of the steps involved in the classification of data with classification class parameters known *a priori*. Convergence is tested with a RMS measure defined by equation 7.34.

7.5 Experiments, Results and Discussion

7.5.1 Classification of two Class Simulated PV Data

Experimental Procedure

Simulated PV data that was initially prepared for the classification experiments in chapter 5 is again used here to determine the classification performance of model G, of this chapter. As before, these simulations allow the classification performance of the PV classifier to be assessed under controlled conditions where the parameters are known *a priori*. Details of the parameters of the simulated PV data can be seen in table 5.1. The results of the classification of each data set by model G were then compared with the ground truth data (also prepared for chapter 5). This comparison enabled the RMS voxel error to be calculated, therefore providing a quantitative measure of the PV classification performance of model G which is compared with the PV models of the preceding chapters.

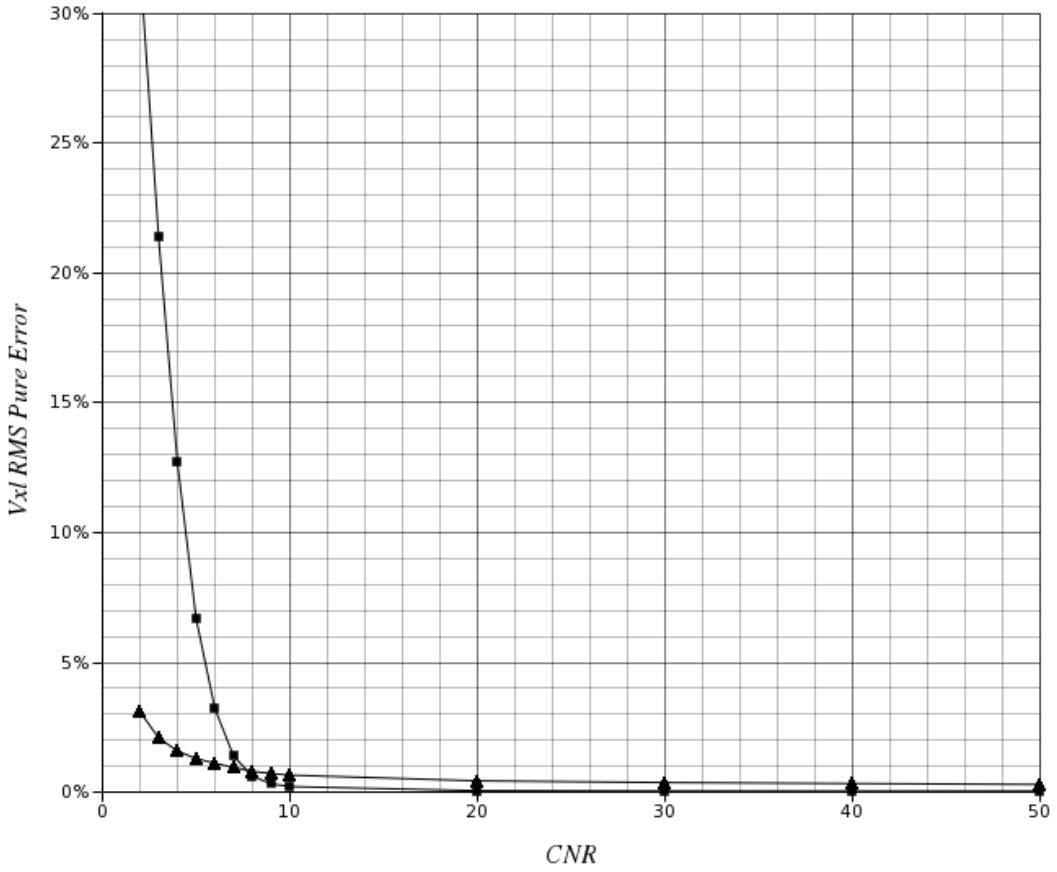
Results and Discussion

The voxel RMS error results of classifying the simulated PV data with parameters given in table 5.1 on page 103 with the PV classifier using model G can be seen in figures

7.13 and 7.14. For comparison, these figures also plot the results previously obtained for model F in chapter 6, the classifier using the reformulated 3-D gradient magnitude and intensity feature space.

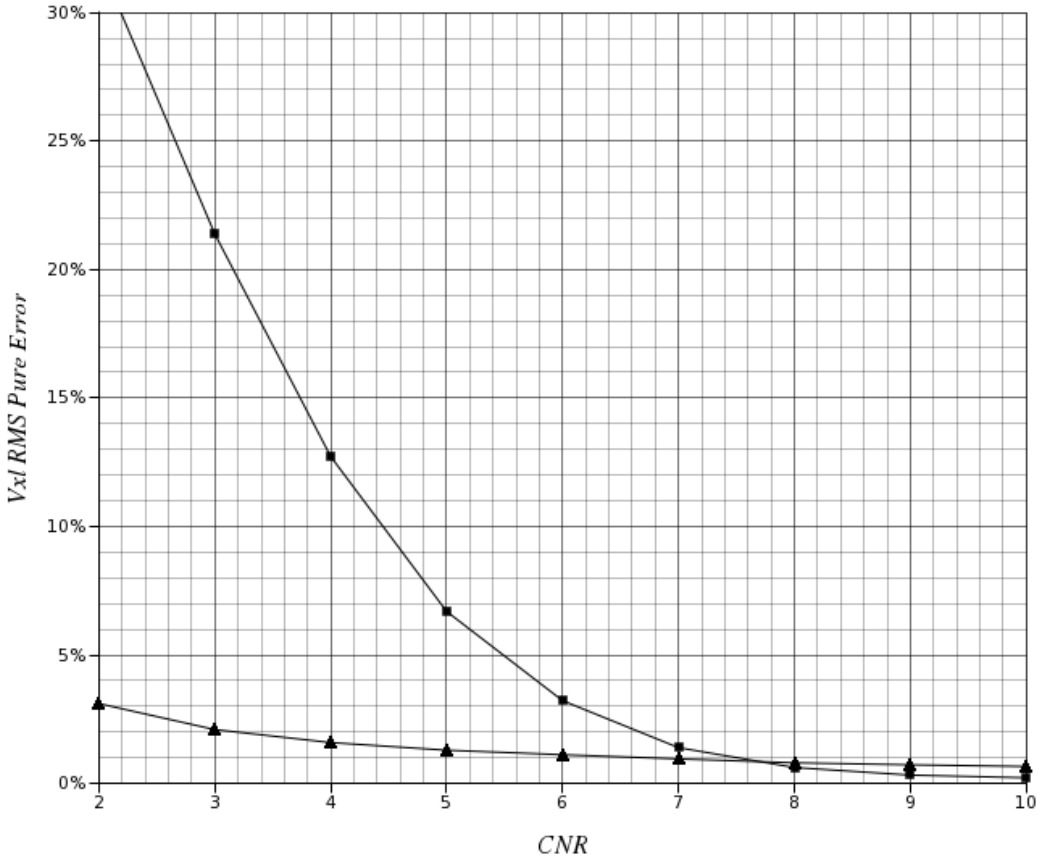
The errors illustrated in figures 7.13 and 7.14 demonstrate quite superior performance for model G over model F. The pure voxel errors obtained for model G are superior for most of the simulated PV data sets with CNR values below 10. The pure voxel RMS error for model G is never greater than 5.0% for any CNR equal or greater than 2. The PV voxel RMS errors for model G are similarly quite superior for CNR values below 9. However, model F does demonstrate slightly superior performance for CNR values greater than 6, although for most CNR values, the performance difference is not greater than 0.75%. This could be due to the limited benefit of including the gradient magnitude in the direct inference of the mixture values for model F. This is in contrast to the use of the gradient magnitude to control the amount of regulation imposed by model G. The use of the gradient magnitude for low CNR values in the direct mixture inference appears to offer no particular advantage.

The gradient magnitude in model F (as well as models D and E) is used as an immediate indication that a voxel is likely to contain a mixture of tissues or activities. Model G on the other hand utilizes the gradient magnitude to indicate the amount of spatial regularisation and this therefore does not impose the condition that a high gradient magnitude is highly likely to be caused by a PV voxel. It does however incorporate into the model the inclination that if there is a high gradient magnitude then less information can be drawn from the neighbouring voxels. These facets are of little importance when classifying these simple two class PV simulated data sets, but they are important when classifying PV data with more complex morphological structures such as might be found in the human brain. After a strategy for parameter estimation is considered in the next section of this chapter, the subsequent chapter presents the application of models F and G to emulated MR brain data that provides a more realistic simulation of the type of PV artefacts seen in clinical MRI data.



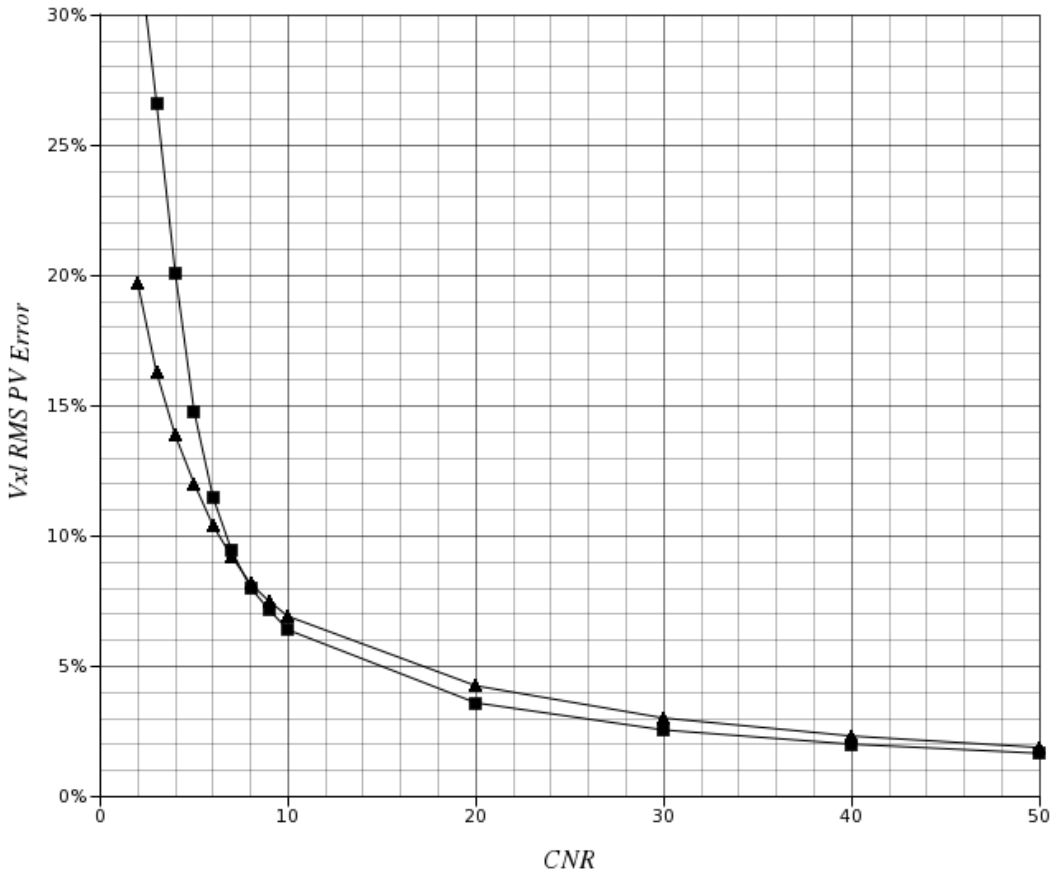
(a)

Figure 7.13: Pure voxel RMS errors obtained for models G (▲) and F (■). (a) is for CNR values from 2 to 50; (b) is for CNR values from 2 to 10. Continued on next page.



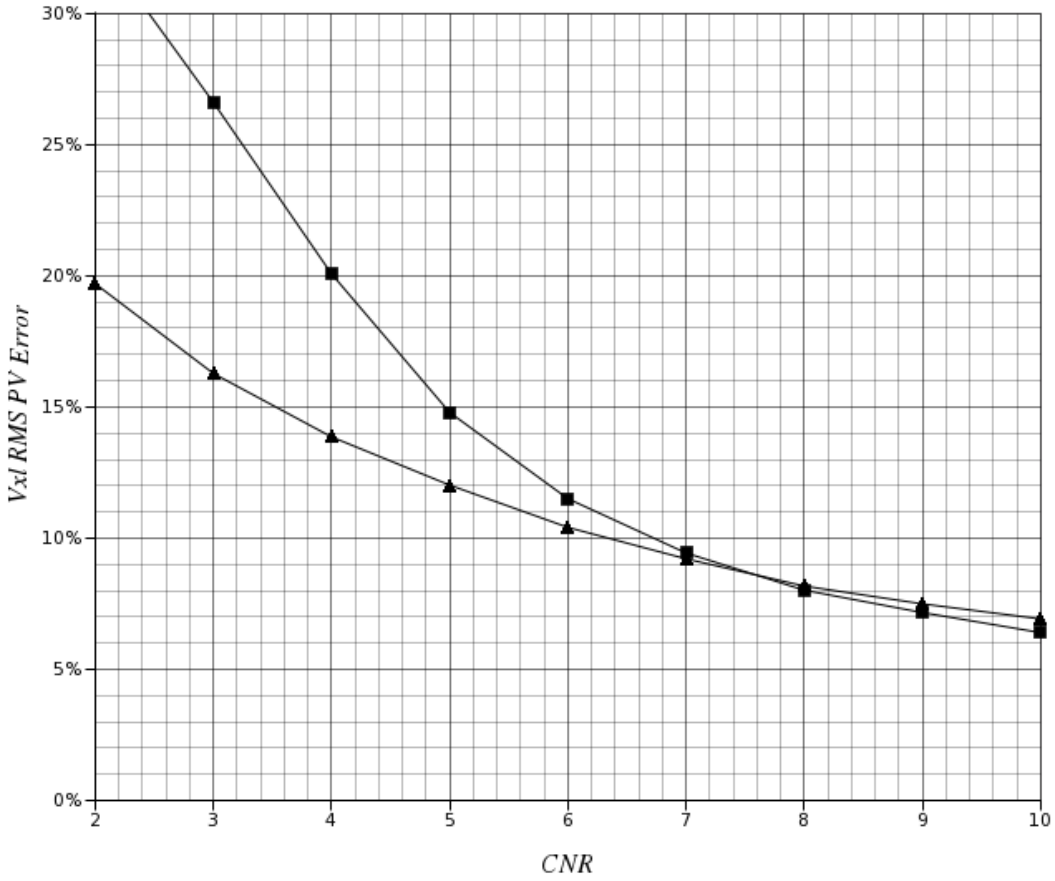
(b)

Figure 7.13: *continued*



(a)

Figure 7.14: PV voxel RMS errors obtained for models G (▲) and F (■). (a) is for CNR values from 2 to 50; (b) is for CNR values from 2 to 10. Continued on next page.



(b)

Figure 7.14: *continued*

7.6 Parameter Estimation

In all the preceding experiments, simulated PV data was utilized to determine the performance of the PV classifiers for a variety of different parameter values. In each of these experiments the exact parameter values used to simulate the data were also passed to the classifier in order to determine the intrinsic (best case) performance of the classifier under controlled conditions, without effects from incorrectly estimated parameters. Although this allows for systematic comparison of intrinsic performance, it is however, not a realistic scenario. Parameter values are not usually known *a priori* and have to be estimated during a classification process of biomedical data such as MRI or PET imaging data. The parameter estimation schema proposed here is based on Bayesian techniques, very similar to the technique described in section 7.3.2 to estimate the mixture values. Indeed, the class memberships of each pixel (voxel in this case) are often considered parameters in themselves (see e.g. [80, 160]).

7.6.1 Formalisation

Consider the parameters of the image data to be denoted by, $\theta = \{\boldsymbol{\mu}\ \boldsymbol{\sigma}\}$, where $\boldsymbol{\mu} = (\mu_1\ \mu_2\ ...\ \mu_n)^T$ is a vector of the N class means and similarly, $\boldsymbol{\sigma} = (\sigma_1\ \sigma_2\ ...\ \sigma_n)^T$ is the vector of the standard deviations. The parameters should be estimated over all the data. The imaging data consists of voxel intensities, $G = \{g_{\boldsymbol{\omega}_i} | \boldsymbol{\omega}_i \in \boldsymbol{\Omega}\}$, and voxel mixture values, $A = \{\boldsymbol{\alpha}_{\boldsymbol{\omega}_i} | \boldsymbol{\omega}_i \in \boldsymbol{\Omega}\}$ where $\boldsymbol{\Omega}$ is the set of all data points or voxels in the imaging data. Therefore, the probability of these parameters, θ, given the data, G and the mixture values, A can therefore be described by

$$p(\theta|G, A) = \frac{p(G|A, \theta).p(A, \theta)}{p(G, A)}. \tag{7.35}$$

Noting the independence A and θ, results in

$$p(\theta|G, A) = \frac{p(G|A, \theta).p(A).p(\theta)}{p(G|A).p(A)} = \frac{p(G|A, \theta).p(\theta)}{p(G|A)}. \tag{7.36}$$

The normalising term or marginal, $p(G|A)$, is difficult to realise. This is a common problem in many Bayesian problems, see e.g. [41]. Fortunately, the IMH algorithm described in algorithm 4, page 174, does not require realisation of the denominator. Therefore only the numerator has to be considered,

$$p(\theta|G, A) \propto p(G|A, \theta).p(\theta). \tag{7.37}$$

This leaves the two terms, $p(G|A, \theta)$, the joint likelihood over all the data and $p(\theta)$ the prior PDF for the parameters. The joint likelihood over all the data can be calculated as

$$p(G|A, \theta) = \prod_{\forall \boldsymbol{\omega}} p(g_{\boldsymbol{\omega}} | \boldsymbol{\alpha}_{\boldsymbol{\omega}}, \theta), \tag{7.38}$$

where it is assumed that individual intensity values are i.i.d. Spatial correlations of the voxel labels have already been considered through the voxel label estimation procedure

in the preceding section. Also, the voxel labels are the dependent terms not the variable being considered.

The prior PDF for the parameters are divided into two PDFs for the mean and standard deviation vectors as they are considered here to be independent, so that they are given the following form,

$$p(\theta) = p(\boldsymbol{\mu}).p(\boldsymbol{\sigma}). \tag{7.39}$$

The standard deviations of the classification class noise distributions, $\boldsymbol{\sigma}$, are unlikely to be very small or zero. The standard deviations are also less likely to be very much greater than the initial estimated values (to be discussed shortly). A right skewed density that is not defined below zero would fulfil these constraints. One such density is the Gamma distribution which is also easily sampled, thus making it a suitable choice for use in the IMH algorithm. Thus, the standard deviation prior distribution is given by a multi-variate Gamma distribution,

$$p(\boldsymbol{\sigma}) = \prod_{v=1}^{n} \frac{\sigma_v^{r-1}}{\Gamma(r).\lambda_v^r}. \exp\left(-\frac{\sigma_v}{\lambda_v}\right), \tag{7.40}$$

where the hyper-parameter λ_v controls the *shape* of the density and the hyper-parameter, r controls the *scale* of the density[1]. Exemplar plots of equation 7.40 for $n = 1$ and $r = 2$ and 3 can be seen in figure 7.15

No particular constraints should be placed on the mean vector prior distribution, $p(\boldsymbol{\mu})$, other than it being likely to have been generated from a symmetrically defined area surrounding the initialisation mean vector. Therefore the prior distribution for the mean vector is specified by a multivariate Gaussian distribution,

$$p(\boldsymbol{\mu}) = \prod_{i=1}^{N} \frac{1}{(2.\pi)^{N/2}.\sigma_i'} \exp\left(-\frac{(\mu_i - \mu_i')^2}{2.\sigma_i'^2}\right), \tag{7.41}$$

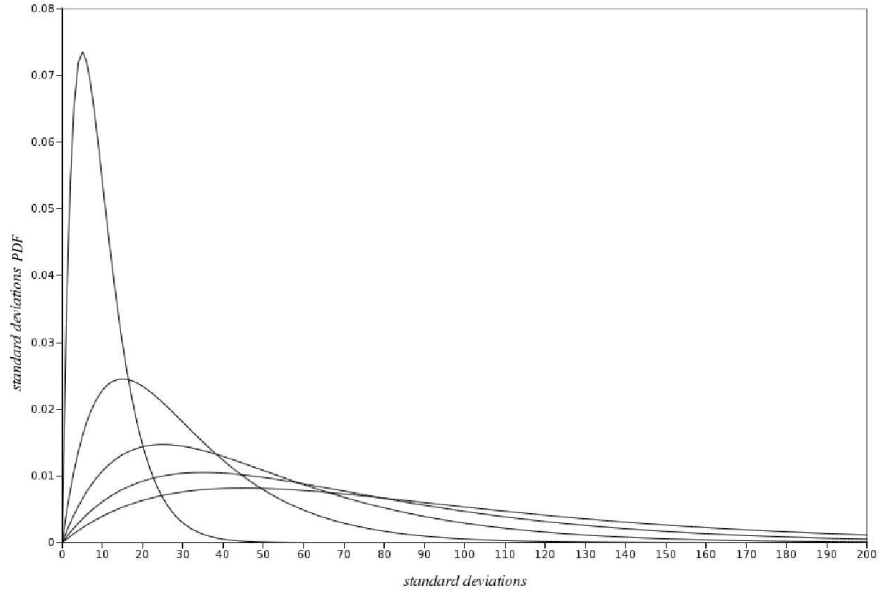
where μ_i' and σ_i' are the hyper-parameter mean and standard deviation for classification class i.

These two prior distributions can now be used in an IMH MCMC algorithm as described in algorithm 4, (page 174), where the prior densities are used as the proposal density to form possible samples. The samples are then used to calculate the acceptance ratio which determines whether the sample is likely to have come from the target density, i.e. the posterior, $p(\theta|G, A)$. Once chains of these samples have been built, they are then used to estimate the true values of the parameters, again via their expected values, $\mathbf{E}[\theta|G, A]$.

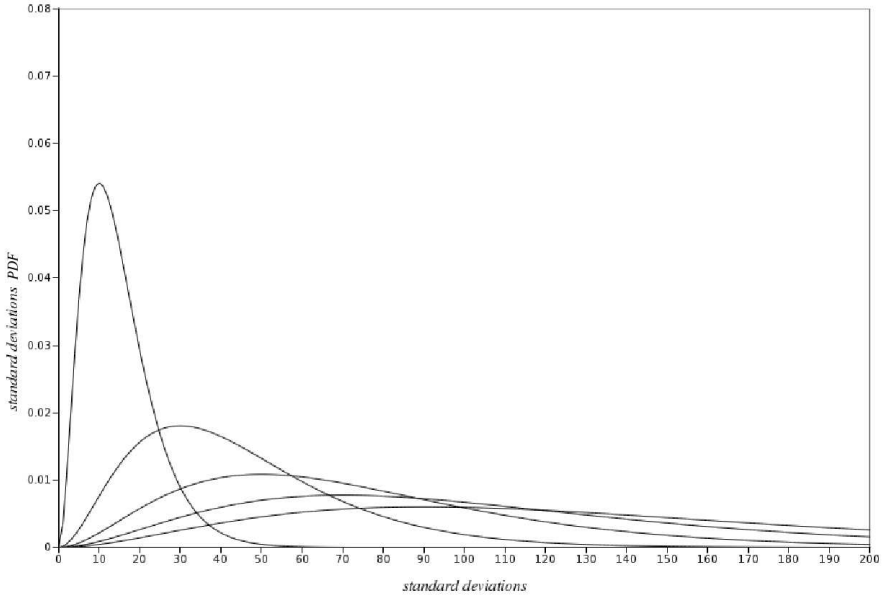
7.6.2 Parameter Initialisation

Initial parameter estimates are required to initiate the algorithm. A popular method of initialising parameters of a parametric unsupervised segmentation or classification

[1] A hyper-parameter is a type of parameter that is often associated with prior distributions in Bayesian problems where constraints can be placed on the hyper-parameters. Sometimes hyper-parameters possess prior distributions themselves, but in this work, they are assigned fixed values through empirical methods.



(a) $r = 2$



(b) $r = 3$

Figure 7.15: Illustration of the various shapes of the single variate Gamma density used as the prior density for the standard deviations (see equation 7.40). From left to right on both sub-figures, $\lambda = 5, 15, 25, 35, 45$, resulting in modes of $5, 15, 25, 35, 45$ for $r = 2$ and $10, 30, 50, 70, 90$ for $r = 3$.

procedure is with the use of a non-parametric technique such as one of the many clustering algorithms, e.g. k-means and fuzzy c-mean algorithms (see e.g. [133, 151]).

A simple minimum distance clustering algorithm which also includes spatial context is adopted here, details of which can be found in algorithm 5.

Algorithm 5 (Simple Clustering Algorithm)

1. *Given the imaging data and the number of desired data clusters, C,*
2. *Divide the CDF of the imaging intensities into C equally spaced values.*
3. *Let cluster intensities, μ_v equal the centre intensities of these units.*
4. *Perform an initial clustering by:*
   a. *For each voxel:*
      i. *Calculate distance between voxel's intensity, $g_{\boldsymbol{\omega}}$ and μ_v.*
   b. *Assign voxel $\boldsymbol{\omega}$ to the cluster with the smallest distance.*
   c. *Update cluster centre value with $g_{\boldsymbol{\omega}}$.*
5. *Perform smoothing cluster:*
   a. *Iterate through all voxels, $\boldsymbol{\omega}$.*
   b. *Calculate local voxel intensity mean, $\mu_{\boldsymbol{\omega}}$.*
   c. *Calculate distance between $\mu_{\boldsymbol{\omega}}$ and μ_v $\forall$ i.*
   d. *Re-assign voxel $\boldsymbol{\omega}$ to cluster with smallest distance.*
6. *Go back to (5) until no change can be seen in cluster intensities, μ_v.*
7. *Calculate standard deviations and priors for each cluster.*
8. *End.*

This algorithm is initialised by dividing the Cumulative Distribution Function (CDF) of the imaging data into C equal divisions. The centres of these divisions on the intensity axis are then used as the initial cluster intensities, or centroids. Clustering is then commenced. An initial clustering procedure is performed using just the voxels' intensities. After this the intensities of the neighbouring voxels to each voxel in the data volume are inspected. Each voxel is assigned to a particular classification class if the mean of the neighbouring voxels' intensities is the least distance away from the centroid of that classification class in relation to all the other centroid mean values. The centroid value is updated with the intensity of the voxel in question, not the mean of the neighbouring voxels' intensities. This process is repeated for every voxel in the data volume until no further changes occur in the cluster intensities. Using the neighbouring voxels' intensities in this way enables the algorithm to retain the spatial context that is inherent in the imaging data. Furthermore, only updating the centroid values with the individual voxel intensities reduces the smoothing effect that arises from the calculation of the mean values from the neighbouring voxel intensities.

7.6.3 Implementation

Some minor modifications were made to the overall classification algorithm described in section 7.4, (page 178), and illustrated in figure 7.12. These modifications included the simulation of the mean and standard deviation posterior distributions.

For each iteration of the classification algorithm, the parameters were estimated after the "Spatial Classification" stage illustrated in figure 7.12. At this point in the algorithm, the expected PV mixing values were known, so that equations 7.40 and 7.41 could be determined. Once the parameters were estimated, they were then used together with the expected mixture values to perform another iteration of the estimation step or were used to inform the algorithm user if enough iterations had executed. Five iterations in this case were found to be sufficient for the termination criteria to be fulfilled, but the algorithm was run for 10 iterations to confirm this was true by observing the values calculated by equation 7.34.

Each parameter estimation step generated 70 successful (accepted) samples before the expected parameter value was calculated. Similarly for each voxel, the mixture value was calculated over 60 successful (accepted) samples (as was previously used in the known parameter value experiments of the previous section). These conditions were found to be programmatically simpler at the time of implementation instead of the usual MCMC approach where the length of the chain is defined on the number of accepted and unaccepted iterations.

7.7 Further Experiments, Results and Discussion

7.7.1 Experimental Procedure

The simulated PV two class data used in the previous experiments is again used here (further details given in table 5.1). Instead of the known parameter values that were used to generate the simulated data, initial preliminary parameter values were estimated with the unsupervised clustering algorithm previously described in algorithm 5. These parameter estimates were then used to initialise the full PV classification and parameter estimation procedure.

7.7.2 Results and Discussion

The preliminary mean and standard deviation parameter estimates from algorithm 5, page 187, can be seen in tables 7.2 and 7.3 respectively.

Tables 7.2 and 7.3 illustrate that the PV parameter estimation algorithm is capable of improving the initial parameter estimates provided by the simpler non-parametric clustering algorithm. The absolute errors between the initial parameter estimates and the actual parameter values together with the PV estimated parameters are illustrated in figures 7.16 and 7.17. The errors for the mean parameters were normalised according to the distance between the two actual class means, in order to provide a fractional measure. The errors for the standard deviation errors were also normalised to the actual standard deviation values. These figures help to illustrate the benefit of incorporating knowledge about the PV effect into the parameter estimation schema, particularly for larger CNR values. When there are two classification classes, any non-PV parameter estimation schema will not be fully aware of the PV voxels with intensities between the two classes. This usually results in biased estimated mean values that are placed

Table 7.2: Estimates for the Mean Parameters. *First two columns represent initial estimates determined via the the simple clustering algorithm on page 187. Third and fourth columns are parameters obtained using the Bayes IMH simulation. Final two columns are the ground truth values that the algorithms were estimating.*

Clustering Estimates		**Bayes IMH Estimates**		**Ground Truth Actual Values**	
Class 1 μ_1	*Class 2* μ_2	*Class 1* μ_1	*Class 2* μ_2	*Class 1* μ_1	*Class 2* μ_2
512.0	1476.1	497.8	1501.7	500.0	1500.0
608.8	1381.9	597.5	1401.8	600.0	1400.0
707.4	1285.0	698.9	1301.7	700.0	1300.0
804.6	1191.4	798.7	1201.2	800.0	1200.0
901.6	1094.6	900.1	1100.9	900.0	1100.0
910.2	1085.5	908.7	1092.6	910.0	1090.0
919.9	1076.5	919.4	1081.5	920.0	1080.0
930.4	1067.3	929.9	1072.4	930.0	1070.0
941.1	1057.7	939.1	1062.6	940.0	1060.0
951.2	1047.8	949.5	1051.3	950.0	1050.0
961.0	1037.7	958.6	1041.1	960.0	1040.0
970.9	1028.4	968.9	1031.1	970.0	1030.0
981.6	1019.4	979.5	1022.6	980.0	1020.0

Table 7.3: Estimates for the Standard Deviation Parameters. *First two columns represent initial estimates determined via the the simple clustering algorithm on page 187. Third and fourth columns are parameters obtained using the Bayes IMH simulation. Final two columns are the ground truth values that the algorithms were estimating.*

Clustering Estimates		**Bayes IMH Estimates**		**Ground Truth Actual Values**	
Class 1 σ_1	*Class 2* σ_2	*Class 1* σ_1	*Class 2* σ_2	*Class 1* σ_1	*Class 2* σ_2
60.1	82.8	18.1	16.9	20.0	20.0
49.4	67.4	18.8	18.0	20.0	20.0
39.2	52.4	19.0	19.3	20.0	20.0
30.2	37.8	19.7	18.6	20.0	20.0
22.8	25.9	20.2	18.9	20.0	20.0
22.2	25.1	19.3	19.6	20.0	20.0
21.8	24.1	20.2	19.8	20.0	20.0
21.5	23.0	19.3	19.5	20.0	20.0
21.2	21.9	19.2	20.0	20.0	20.0
21.0	21.1	19.9	19.7	20.0	20.0
20.5	20.8	20.1	19.5	20.0	20.0
20.2	20.5	20.8	19.9	20.0	20.0
20.1	20.1	20.3	20.4	20.0	20.0

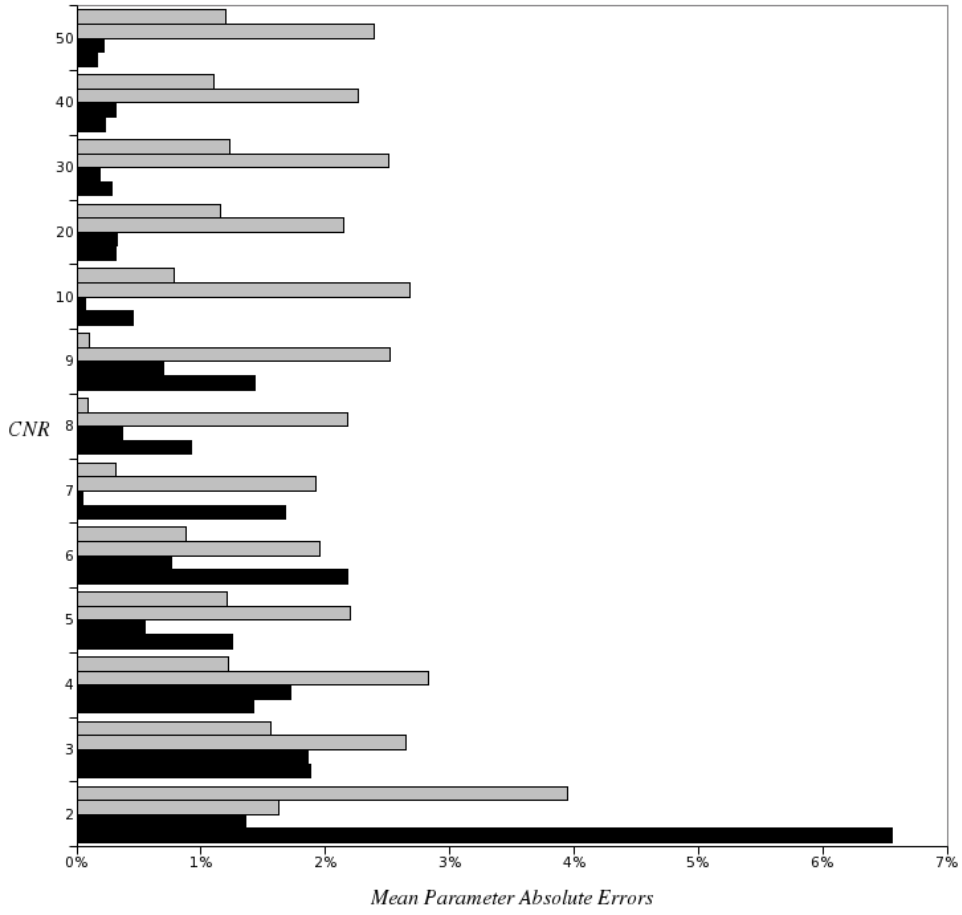


Figure 7.16: Absolute fractional errors of the estimates of the mean parameters with values given in 7.2, that include the initial estimates (*grey*) via the simple clustering algorithm and the PV estimates (*black*) via the expected value of the simulated posterior distribution. The errors are relative to the distance between the actual class means.

somewhere between the two class means. The effect on the variance is somewhat more significant, as can be seen from figure 7.17. The non-PV incorrectly estimated standard deviations are usually much greater than the true standard deviation values, partly due to the incorrectly estimated mean values *and* the larger apparent variation in per-class intensities.

As the CNR decreases, fewer intensities exist between the two classes, thereby reducing the error seen by the estimated non-PV standard deviations. This is in contrast to the PV aware estimation schema which provides similar errors for each standard deviation estimate for any value of CNR, confirming that the PV model possesses improved precision over the non-PV aware estimation schema. The fractional errors for the non-PV mean estimates remain very similar for most of the CNR values (see figure 7.16), increasing slightly for smaller CNR values. The PV aware mean estimates also increase for smaller CNR values, although it should be observed that the relative error of 6.5% of the smaller CNR values represents very few actual differences in intensities as can be seen from the values in table 7.2.

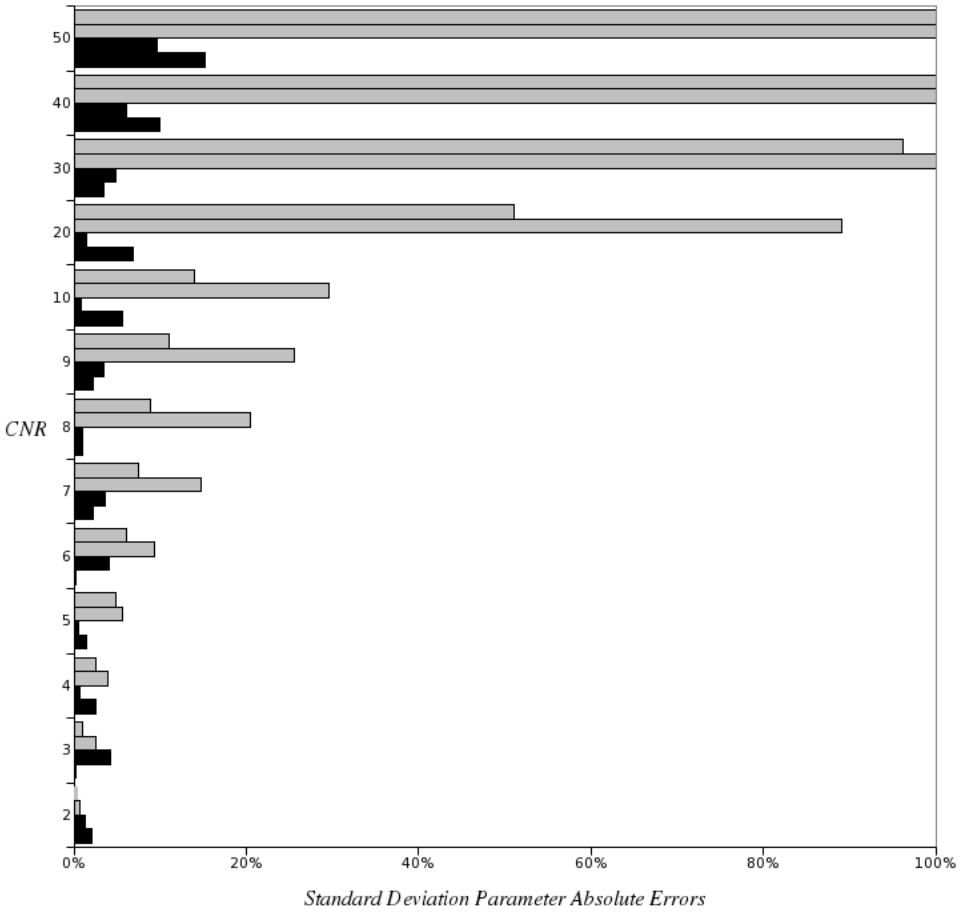


Figure 7.17: Absolute fractional errors of the estimates of the standard deviation parameters with values given in 7.3, that include the initial estimates (*grey*) via the simple clustering algorithm and the PV estimates (*black*) via the expected value of the simulated posterior distribution. The errors are relative to the actual class standard deviation values.

7.7.3 Classification Errors with Estimated Parameter Values

It might be expected that errors in the parameter values might affect the PV classification performance of the classifier. Therefore the pure and PV voxel RMS errors obtained for the classifications using the estimated parameter values together with the results obtained using the parameter values known *a priori* can be seen in figures 7.18 and 7.19 respectively. As might be expected, the pure voxel RMS errors are generally worse for the estimated parameter values (figure 7.18), but somewhat surprisingly the PV errors are improved with the estimated parameter values. The pure and PV errors also demonstrate somewhat erratic error values, this can be attributed to parameter values that have been incorrectly estimated. The pure results for data sets with CNR values of 3 and 4 are considerably better than the results for the data sets with CNR$= 2, 5$. This is because the distance between the two estimated mean values for CNR$= 3, 4$ are very close to the true distances between the mean values, (refer to the values in table 7.2). The data set with CNR$= 7$ has a mean value (929.9) that is very close to the true mean value (930), resulting in a smaller pure error (refer to figure 7.16). It is not clear as to why the data set with CNR$= 9$ also has reduced pure error, except perhaps a combination of these effects. The pure and PV errors appear to follow either of two trends, one of which is most likely the true trend (PV or pure) and the other is most likely a gross voxel RMS type error that is the pure and PV errors combined. This is now discussed.

The improvement in classification for the PV voxels is most likely due to the arbitrary division of voxels into pure and PV. When a significant number of pure voxels have been misclassified, then it is quite possible that the PV voxels will have improved classification due to the probabilistic nature of the problem. This is confirmed by observing the gross voxel RMS errors, as seen in figure 7.20. The gross voxel RMS error does not distinguish between a PV or a pure voxel, thereby removing this arbitrary division. The division is however useful when the true parameter values are known and the ratio of pure to PV voxels may change due to differing simulation parameters, such as anisotropy which was studied in chapter 6. Figure 7.20 illustrates that the performance of the classifier using estimated parameter values is in fact very similar to the classifier using actual simulation parameter values. The knowledge that the classifier can estimate parameters together with providing a classification performance that is very similar to one using actual parameter values is useful as it provides further validation to its applicability and inspires confidence for use with clinical data.

7.8 In Conclusion

The PV model developments described in this chapter, culminating in a seventh model of the PV effect, model G, have lead to superior classification performances over all the previous PV models (A to F). The first part of this chapter presented novel PV prior distributions. These included a PV prior distribution that was derived analytically via a physical model of the PV effect and the second using a phenomenological observation, relating a commonly found phenomenological law known as Benford's Law to the PV effect.

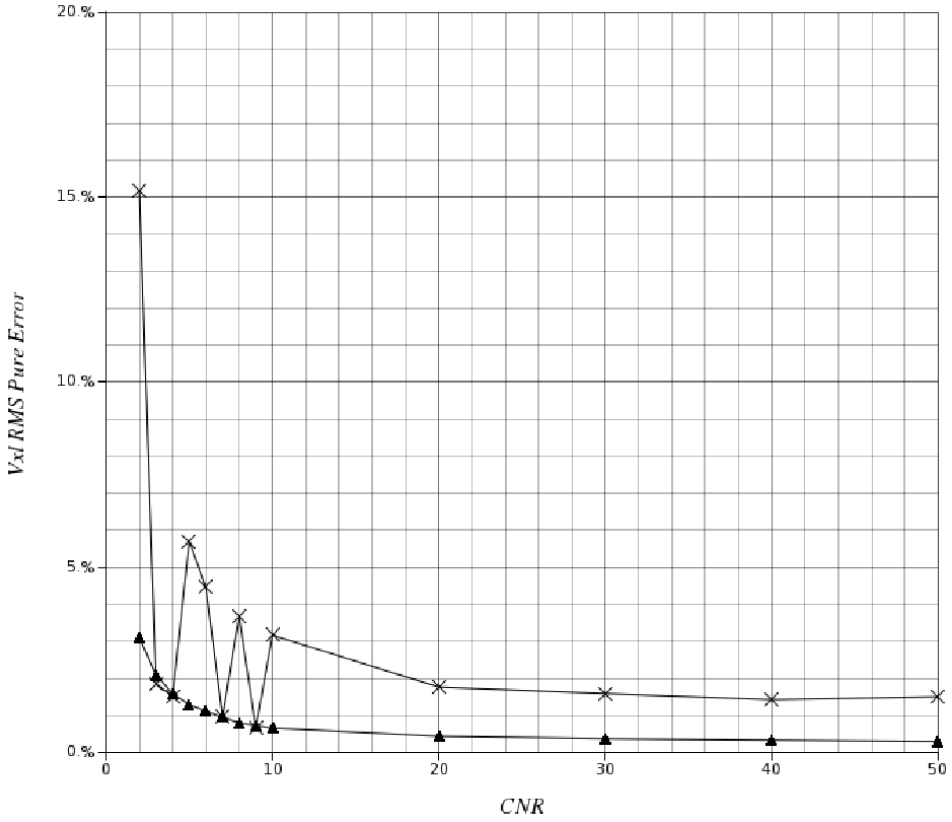


Figure 7.18: Pure voxel RMS errors obtained for model G using actual simulation parameters (▲) compared to model G using estimated parameter values (×).

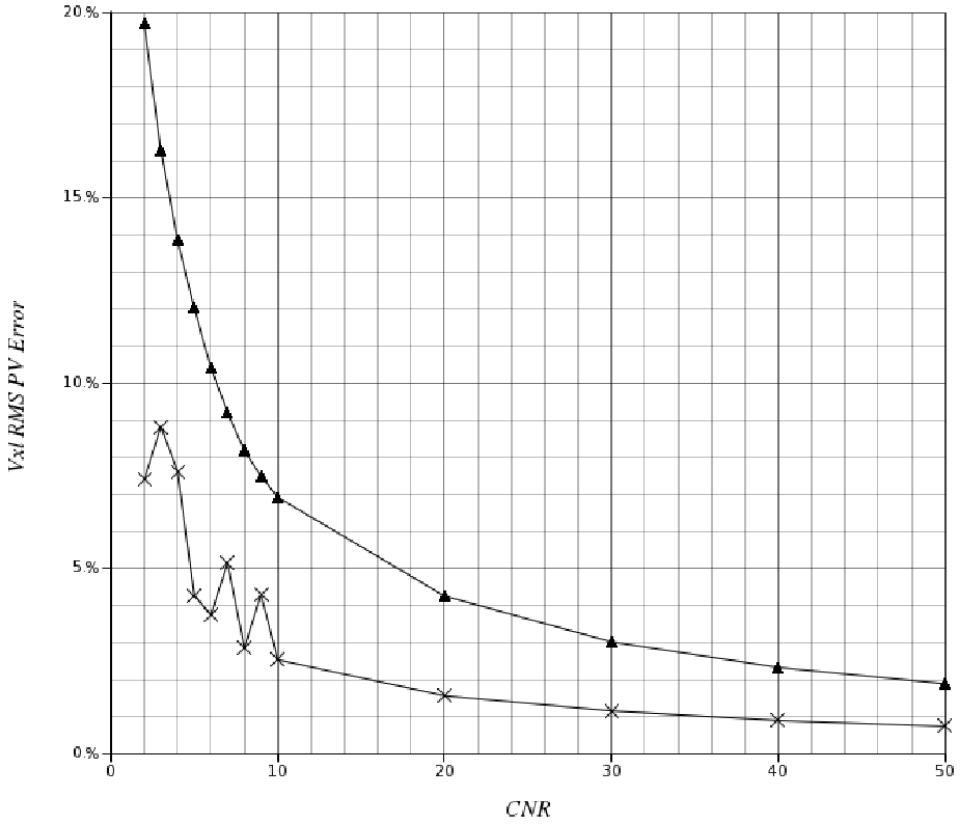


Figure 7.19: PV voxel RMS errors obtained for model G using actual simulation parameters (▲) compared to model G using estimated parameter values (×).

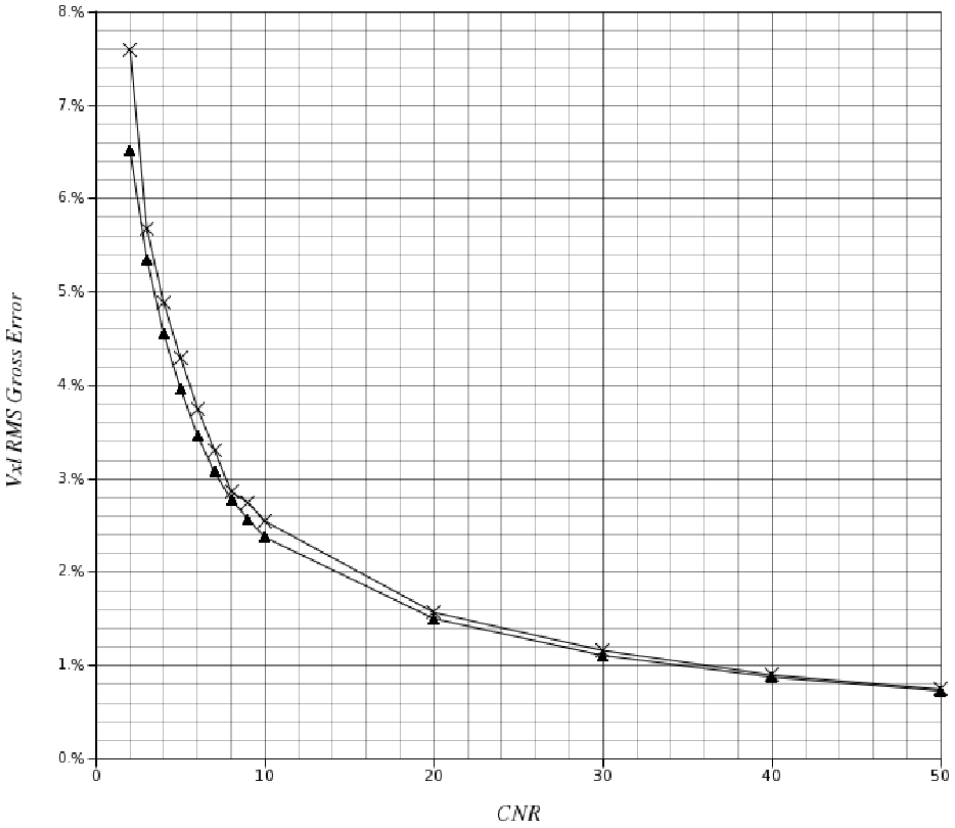


Figure 7.20: Gross voxel RMS errors obtained for model G using actual simulation parameters (▲) compared to model G using estimated parameter values (×).

The developments in this thesis prior to this chapter provided an insight into the nature of the PV effect, but neglected to include any form of spatial context except for the second order information provided by the gradient magnitude of models D-F in chapter 6. This gradient magnitude information was also included in the model of this chapter (G), but it was found that the inclusion of a spatial smoothness constraint removed the direct dependency on the gradient magnitude feature. The gradient magnitude was therefore re-introduced to control the amount of spatial regularisation. The use of gradient magnitude to control the spatial regularisation advances its use in PV classification by avoiding the deduction that only high gradient regions may result in PV voxels. This yields a more realistic model of possible PV instances that may occur together with extremely promising classification results for simulated PV data.

Chapter 8

Application to Biomedical Images

The preceding chapter (7), presented a method for estimating parameters of a PV model that included neighbourhood information in the form of a vectorial mean of the neighbouring mixture values. Improved classification performance was seen using this novel formulation over other PV classifiers in this thesis. However, the PV classifiers in this thesis, up until this point, have only been assessed in terms of their classification performance on simple two class simulated PV data consisting of bands of concentric spheroids. Other sources of data are available that can assist in the performance assessment of the PV classifiers under more realistic conditions. This is the topic of this chapter, where three classifiers have been selected and are now assessed utilising medical imaging data in the form of: simulated MR brain data sets containing variable amounts of noise, where PV ground truth templates are available for classification assessment; a PET phantom with a PV ground truth template derived from high resolution CT data; and 20 normal volunteer MRI brain data sets with expertly defined discrete ground truth maps. Each of these biomedical data sets are introduced below, in their respective sections.

A PV classifier from each chapter was selected based on their representative properties for each chapter. A PV classifier using the analogous Gaussian-Triangle model (B), from [148] was selected due to the approximate equivalence that was found with the explicit PV model (C); each both possessing superior PV classification performance over the simpler finite Gaussian mixture model (A). A PV classifier using the intensity and 3-D gradient magnitude reformulated (i.e. non-analogous) feature space (model F) was also selected based on the superior performance over the other intensity and gradient magnitude feature space classifiers from chapter 6. Also, the Gradient controlled Spatial Regularisation (GSR) PV classifier of chapter 7 was selected representing a PV model that includes PV neighbourhood contextual information. Therefore these three classifiers, using PV models B, F and G are utilised through out the work in this chapter.

8.1 Classification of McGill Simulated MRI Brain Data

The McGill simulated MRI brain phantom data [27, 28, 72, 73] is used here to assess the performance of the PV classifiers presented in this thesis. The simulated brain phantom is obtained from BrainWeb, provided by the Brain Imaging Center at the Montreal Neurological Institute, McGill University, Canada [12]. The simulated brain data consists of a morphological structure that is very similar to that found in actual MR images of the human brain. This data can therefore assess the performance of the PV classifiers under clinically realistic imaging conditions, where the PV artefact is affected by the complex brain morphology that are otherwise difficult to create with regular geometric structures. This is in comparison to the conditions provided by the simple two class simulated PV data that has so far been used. The simple two class simulated PV data provided a carefully controlled set of simulations that produced PV voxels and pure voxels with exact parameter values known *a priori* combined with a well-defined geometric structure. The simulated McGill brain data is also not created in the same manner, where parameter values of the PV models do not possess direct correspondence with the parameter values of the MR simulation process utilised to produce the simulated brain data.

8.1.1 Description of the Simulated MRI Brain Data

The process of creating the simulated MRI brain data is described in detail in [73], details of which are briefly summarised here.

Twenty-seven MRI T1 weighted scans of a single subject were acquired, registered and then averaged. This averaged data set was then classified with a fuzzy minimum distance classifier and manually corrected by a neuroanatomist. As noted by Kwan *et al.*, [73], (members of the Montreal Neurological Institute at McGill University), the most important factor of the result of this classification process is that it is representative of a *"plausible"* brain anatomy. The corrections applied by the neuroanatomist assure that this is the case.

Spin-models of the different tissues are then constructed that are combined with a particular pulse-sequence. This enables signal intensities for different tissues to be evaluated via equations such as the Spin-Echo imaging sequence signal equation (2.6), (see page 16). Kwan *et al.* also describe a simulation process in situations where convenient signal equations are not available. The signal intensities are then mapped to the 3-D anatomical templates derived from the classified averaged data volume. This anatomical template is then convolved in the spatial domain by an inter-slice 1-D PSF to simulate slice selection and then in the Fourier domain by a 2-D PSF to simulate the with-in slice PV effect. Zero mean additive Gaussian noise is then applied to the real and imaginary components to simulate the noise processes of MRI acquisition physics. Magnitude images are then calculated from these imaginary components which results in Rician distributed noise with pre-calculable parameters.

8.1.2 Experimental Methodology

A number of default MR T1 simulations together with their fuzzy tissue templates available from [12], were downloaded, details of which can be seen in table 8.1 and a set of single exemplar slices from each data set are illustrated in figure 8.1.

Table 8.1: Simulated Brain Data Set Details

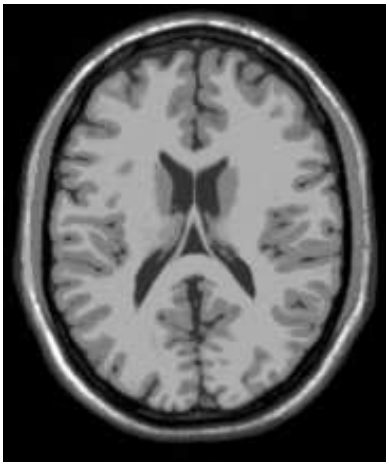
Noise	Voxel Dimensions	Inhomogeneity
1%	$1 \times 1 \times 1$ mm^3	0%
3%	$1 \times 1 \times 1$ mm^3	0%
5%	$1 \times 1 \times 1$ mm^3	0%
7%	$1 \times 1 \times 1$ mm^3	0%
9%	$1 \times 1 \times 1$ mm^3	0%

The simulated McGill brain data is also available with variable amounts of inhomogeneity artefact that is a common imaging artefact associated with the MRI acquisition process. This was previously discussed in chapter 2. However, the downloaded data sets were limited to 0% inhomogeneity to simplify performance assessment with this realistic simulated brain data because none of the classifiers in this thesis model the inhomogeneity artefact. The result of classifying data with the inhomogeneity artefact will be seen for real MR data in the last set of experiments in this chapter. Each simulated brain data set is described by a single percentage noise value, 1% to 9%. The percentage values represent the ratio of the standard deviation of the noise to the signal for a reference tissue, which for the pre-computed T1 simulations, (used here), is the WM tissue. Therefore a CNR value can be estimated using these noise values in conjunction with the noiseless signal levels of the 0% noise simulated MR brain data set. These calculated CNR values are given in table 8.2. These CNR values illustrate that the two-class simpler simulated PV data of the preceding chapters utilised a range of CNR values that are representative of the simulated brain data. These CNR values may therefore be considered as guides to classifier performance and may also be used to assess the quality of the estimated parameter values.

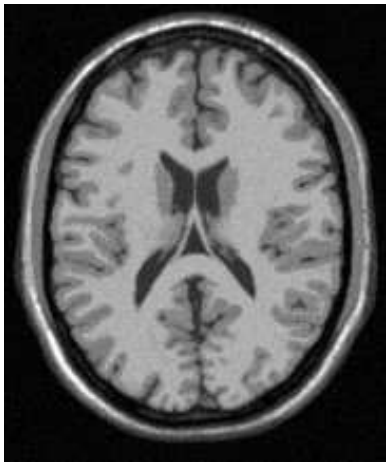
Table 8.2: Simulated Brain Data Set CNR Values

Noise	CSF-GM	GM-WM	CSF-WM
1%	44	26	70
3%	15	9	23
5%	9	5	14
7%	6	4	10
9%	5	3	8

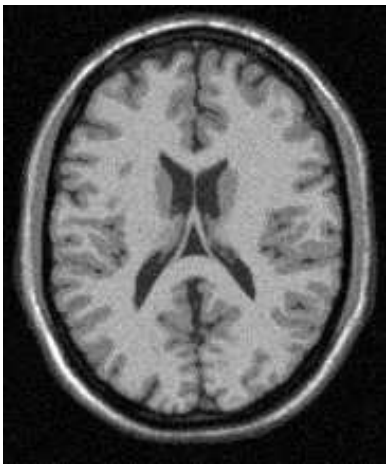
PV classifiers using models B, F and G were applied to the five brain web data sets. Model G together with the parameter estimation detailed in chapter 7 was used to estimate the parameters from which the classifiers using models B and F were able to classify the data as no parameter estimation technique has been investigated for these two models. Model G classifier utilise the same classification process as was followed in chapter 7 including the non-supervised spatially aware minimum distance classification



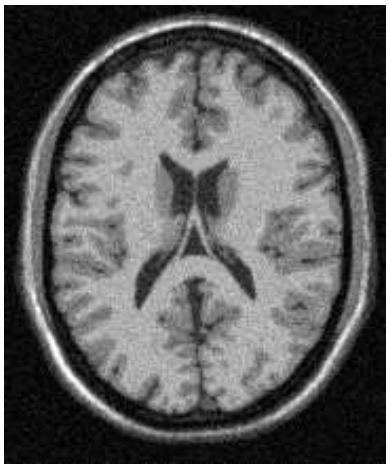
(a) 1% noise



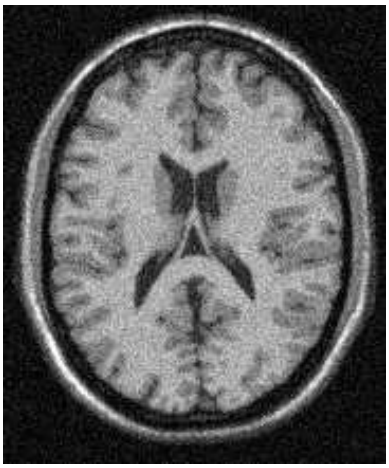
(b) 3% noise



(c) 5% noise



(d) 7% noise



(e) 9% noise

Figure 8.1: Exemplar image slices from the simulated brain data sets [73], details given in table 8.1. Notice the visual increase in the amount of noise as might be expected with the increase in the percentage of noise.

algorithm to initialise the process. The classifiers were set to classify for three classes, corresponding to the CSF, GM and WM CNS components. The remaining components were automatically excluded via a GM, WM and CSF brain mask created from the fuzzy GM, WM and CSF tissue templates, illustrated in figure 8.2. In practise these components could be excluded via a skull-stripping algorithm, e.g. see chapter 3.

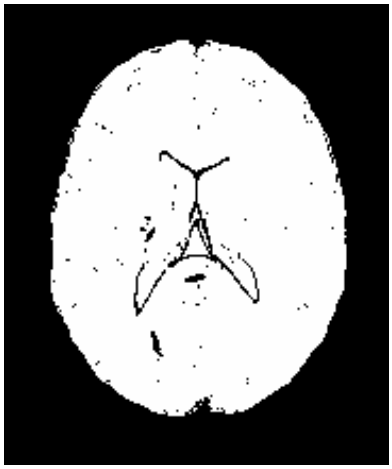


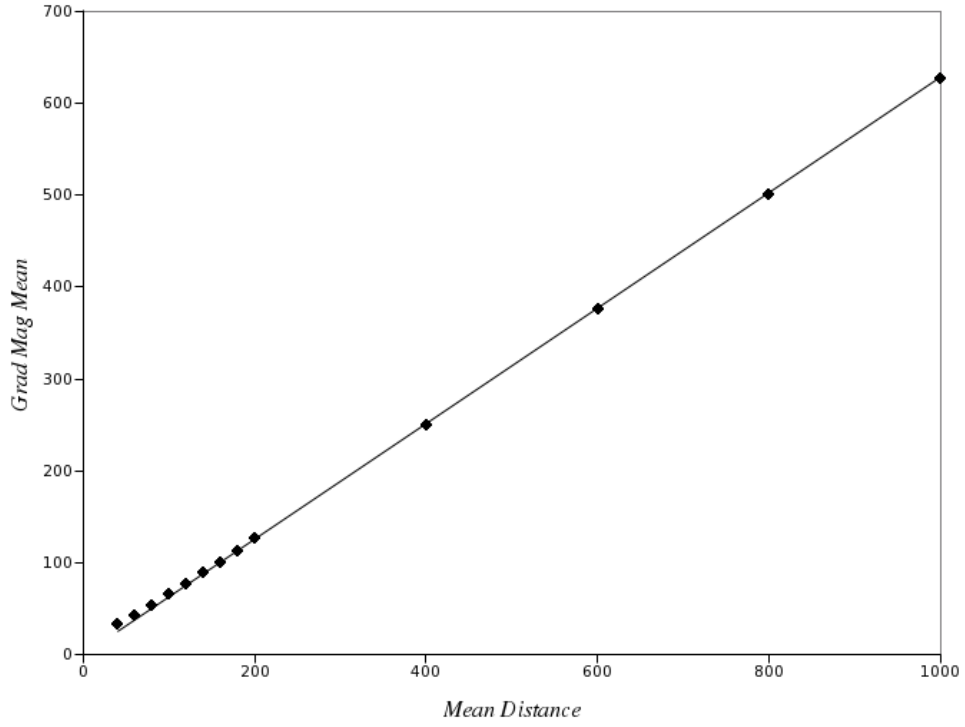
Figure 8.2: Exemplar slice from brain mask with voxels composed of more than 87% GM, WM and CSF components from the fuzzy tissue templates (illustrated shortly in figure 8.4). White represents voxel to be classified, black is not classified. Notice some holes many of which correspond to WM PV voxels with PV content below 87%.

The gradient magnitude feature space parameters for model F were inferred from a manual regression of the parameters found for the simple two class simulated PV data in the previous chapters. Previously these were calculated directly from the ground truth, but simple straight line relationships were found to exist between the maximum gradient magnitude mean for PV voxels, (maximum of $A(\boldsymbol{\alpha}|\tau_{v,a})$ – see page 130) and the distance between the pure class intensity means, (i.e. PV class event $\tau_{v,a}$ implies $||\mu_v - \mu_a||$). A similar relationship was also found between the width of the gradient magnitude distribution for PV voxels, (σ_z^2 – see page 130) and the distance between the pure class intensity means, (i.e. $||\mu_v - \mu_a||$). The results of these regressions can be seen in figure 8.3 for the gradient magnitude mean and width parameters.

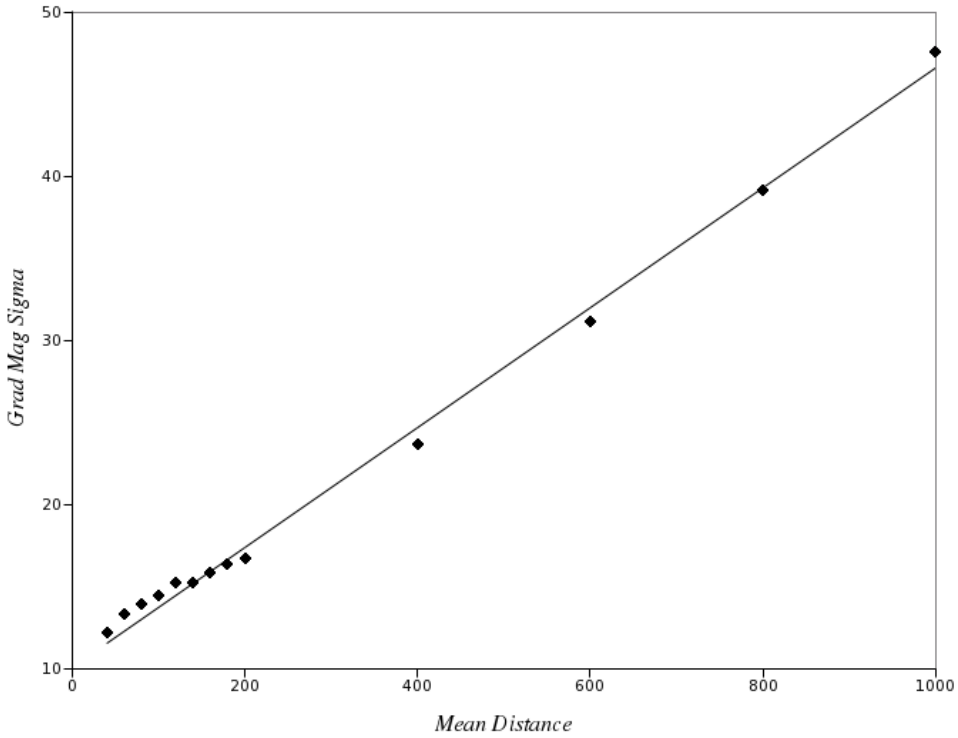
8.1.3 Results and Discussion

Exemplar image slices for the classification results for each classifier model B,F and G can be seen in figures 8.5, 8.6 and 8.7, respectively. Sub-figures 8.5(a-c), 8.6(a-c) and 8.7(a-c) repeat the corresponding fuzzy tissue templates (from figure 8.4) for visual comparison with the results in sub-figures 8.5(d-r), 8.6(d-r) and 8.7(d-r).

The exemplar image slices enable qualitative analysis of the results of the classification process. It is difficult to produce any conclusions with regard to the quality of the classification process except to say that the classifier utilizing model G has produced smoother looking results for the noisier data sets (5%,7% and 9% noise) in comparison to



(a) Gradient Magnitude Distribution Location



(b) Gradient Magnitude Distribution Width

Figure 8.3: Straight line approximations to the gradient magnitude feature parameters (squares), quantifying location (grad mag mean, (a), $A(\boldsymbol{\alpha}|\tau_{v,a}) = ||\mu_v - \mu_a|| \times 0.6283$) and width (grad mag sigma, (b), $\sigma_z = 0.0365 \times ||\mu_v - \mu_a|| + 10.1$) in relation to mean distance.

models B and F. This is despite the classifiers B and F utilizing parameters estimated from the parameter estimation of classifier G. The CNR values calculated from the estimated parameters can be seen in table 8.3, which help to illustrate the quality of the estimated parameters on this simulated MR brain data.

Table 8.3: Simulated Brain Data Set Estimated CNR Values

Noise	CSF-GM	GM-WM	CSF-WM
1%	22	17	56
3%	13	8	25
5%	9	5	17
7%	6	4	11
9%	5	3	8

The estimated CNR values appear to be very similar to the CNR values calculated from the percentage noise values, given earlier, in table 8.2, page 199. The 1% simulated brain data set possesses the greatest difference between actual and estimated CNR values. This is probably due, in part, to the way in which the CNR is calculated. Small differences in a small estimated standard deviation in relation to the true standard deviation will result in large differences in the estimated CNR value in relation to the actual CNR value.

The results of classifying the simple two class simulated PV data in chapters 5 and 6 together utilising similar CNR values as those in table 8.3 suggest that the classifiers utilizing models B and F might be expected to exhibit poor performance in relation to model G, particularly for the 7% and 9% noise data sets. This is confirmed by the RMS errors for these simulated brain data sets which can be seen in figure 8.8.

For the smaller CNR value (greater percentage of noise) simulated data sets, the errors increase quite considerably. Unsurprisingly, little improvement is seen for the classifier using model G over the classifiers using models B and F for the data sets with noise values smaller than 5%. This was also seen in the previous chapters, for simulated data sets with CNR values greater than 10 (see table 8.3). Model G offers the best performance improvement over the other models for the 7% and 9% simulated brain data sets. Model F, which utilized a gradient magnitude model to identify PV voxels does not appear to offer any classifier performance benefit over the other models. This is to be expected as the PV model used by the gradient magnitude models (D-F) of chapter 6 identify PV voxels as voxels with high gradient magnitude. This is not strictly valid for MR images of the brain, as was previously seen in chapter 2. The brain is composed of regions of PV voxels that are not necessarily associated with particularly high gradient magnitude values. This is illustrated in figure 8.9, where some PV voxel regions are associated with relatively low gradient magnitude in comparison to other PV voxels that arise from the edge regions. Whilst this effect may not fundamentally limit classifier performance, it does appear to be a contributory factor in the relatively poor performance of PV classifier for model F in relation to the performance of the intensity only classifier, model B. It should also be noted that qualitative comparison of the classifier output for models B and F (see figures 8.5 and 8.6) appear to illustrate fewer mis-classified voxels for model F (using the 3-D intensity and gradient magnitude feature space), particularly for large regions of contiguous WM.

(a) WM fuzzy template

(b) GM fuzzy template

(c) CSF fuzzy template

Figure 8.4: Exemplar image slices from fuzzy tissue templates used as the quantitative ground truth for performance assessment of the PV classifiers. White (voxel intensity= 255) represents 100% content and black (voxel intensity= 0) represents 0% content.

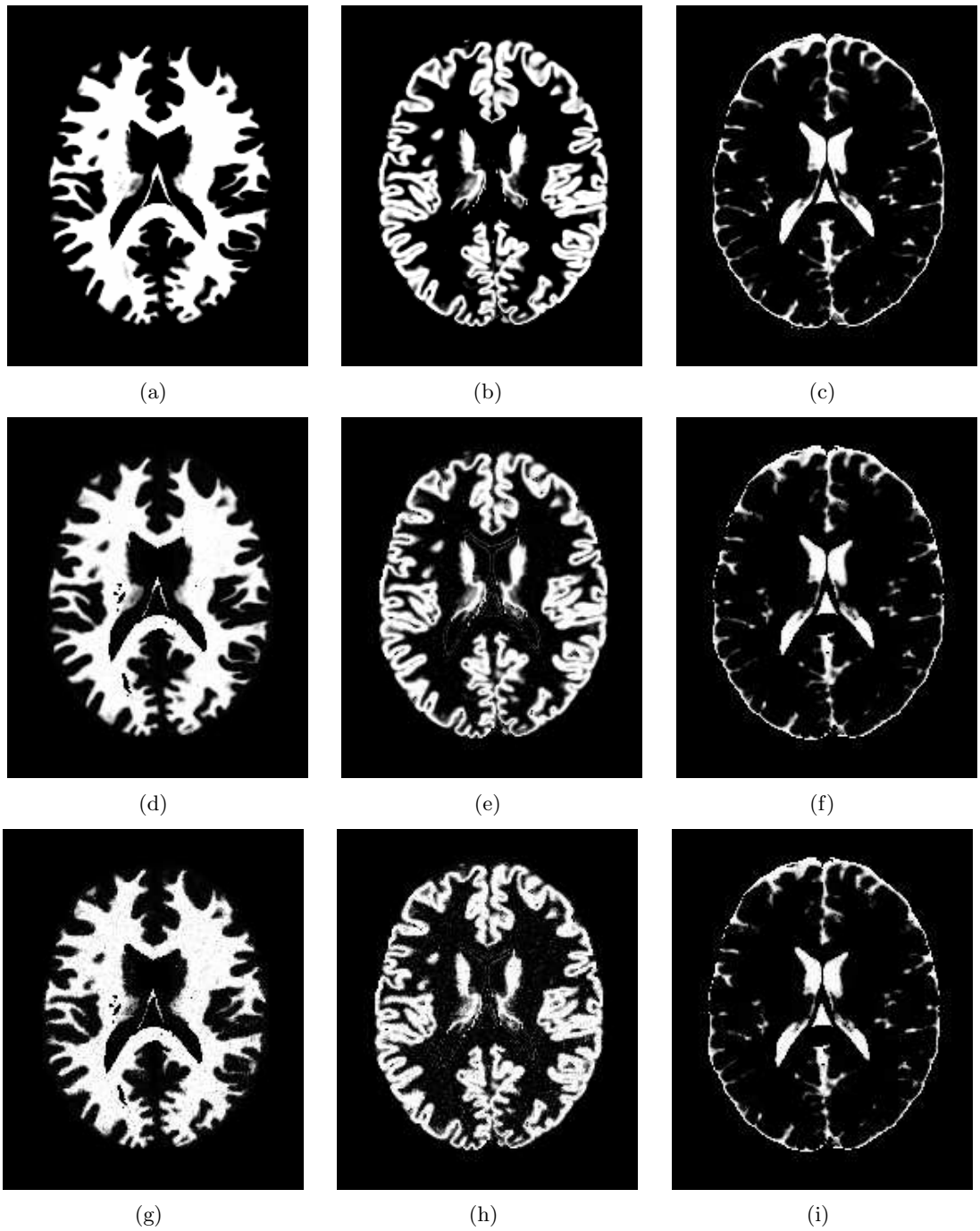
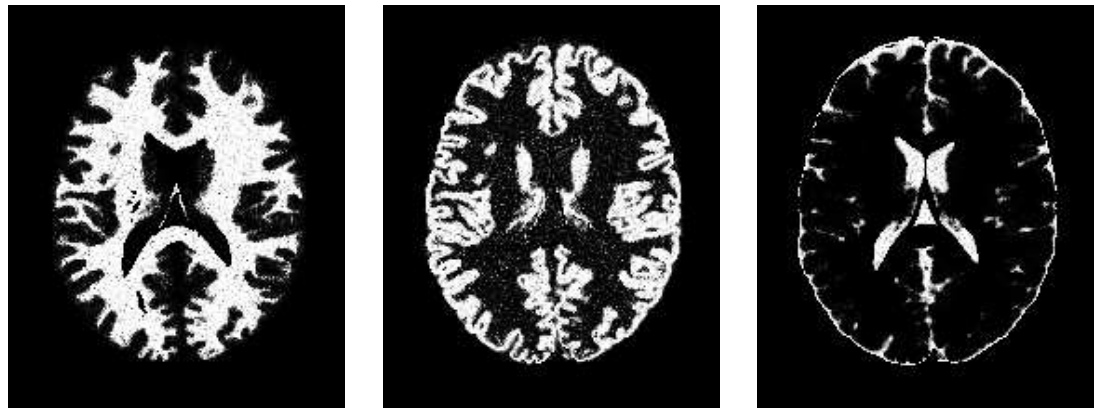
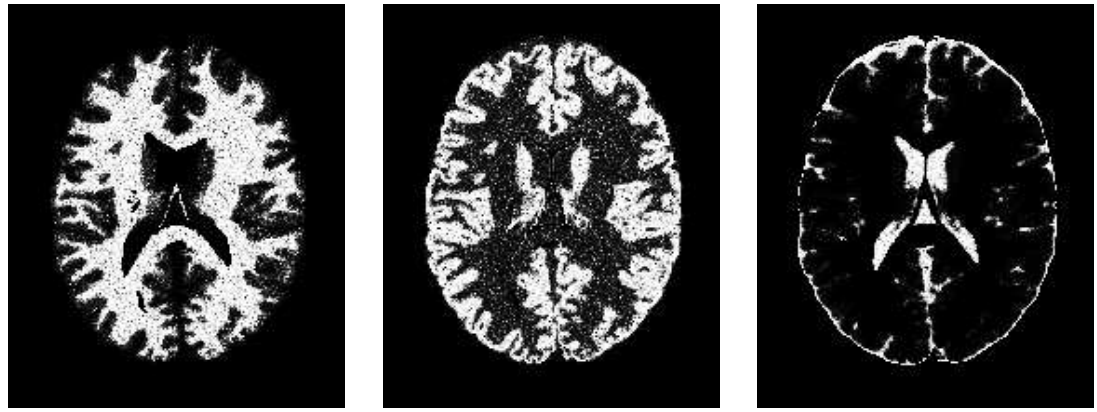


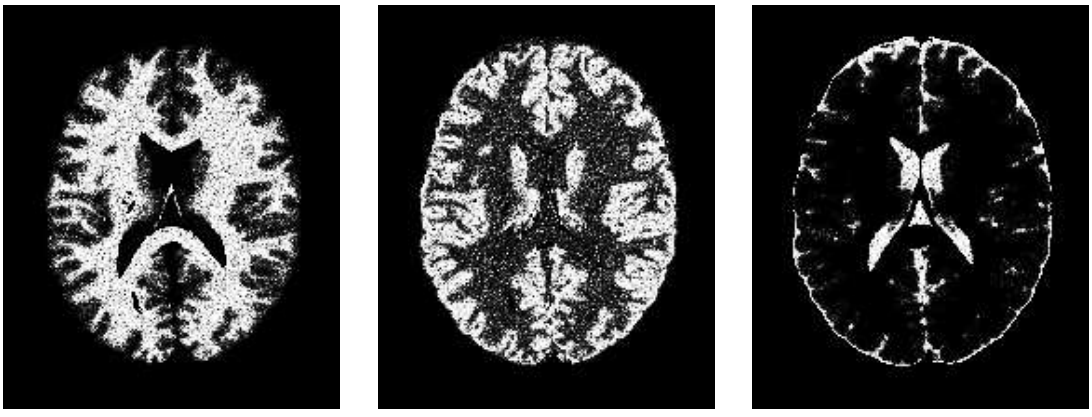
Figure 8.5: Exemplar slices from the WM, GM and CSF fuzzy tissue templates (a-c) and results obtained for classifier using model B (d-r), continued on next page. (d-f) WM, GM and CSF results for 1% noise; (g-i) WM, GM and CSF results for 3% noise; (j-l) WM, GM and CSF results for 5% noise; (m-o) WM, GM and CSF results for 7% noise; (p-r) WM, GM and CSF results for 9% noise. Note holes in WM classification are due to holes in the brain mask, illustrated in figure 8.2.



(j) (k) (l)



(m) (n) (o)



(p) (q) (r)

Figure 8.5: *continued.*

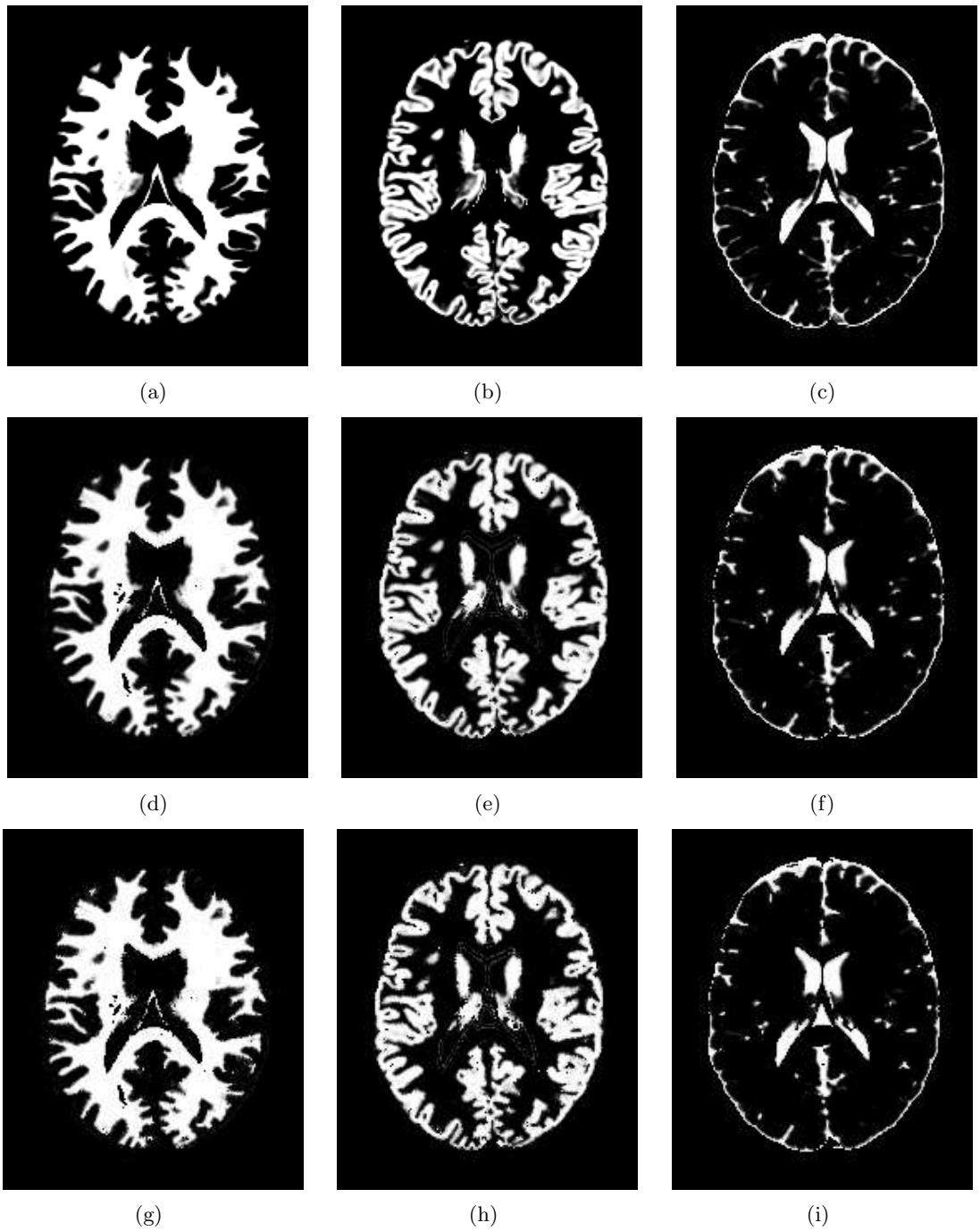
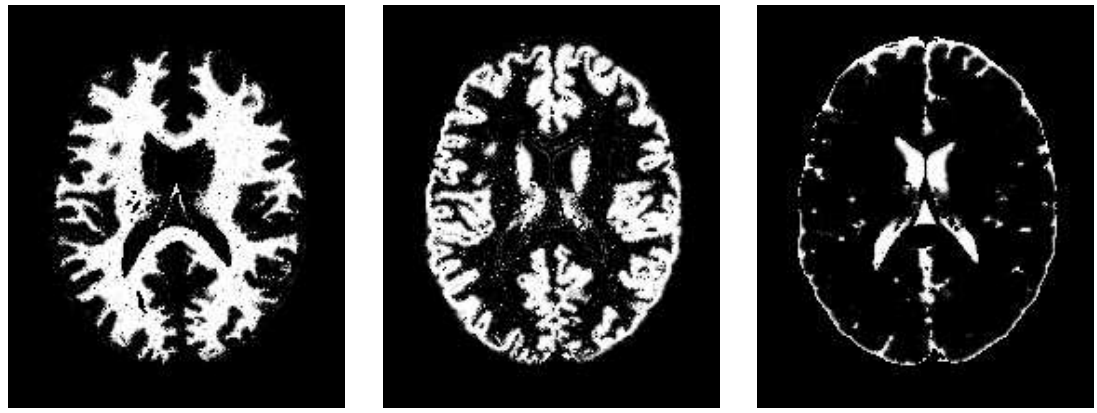
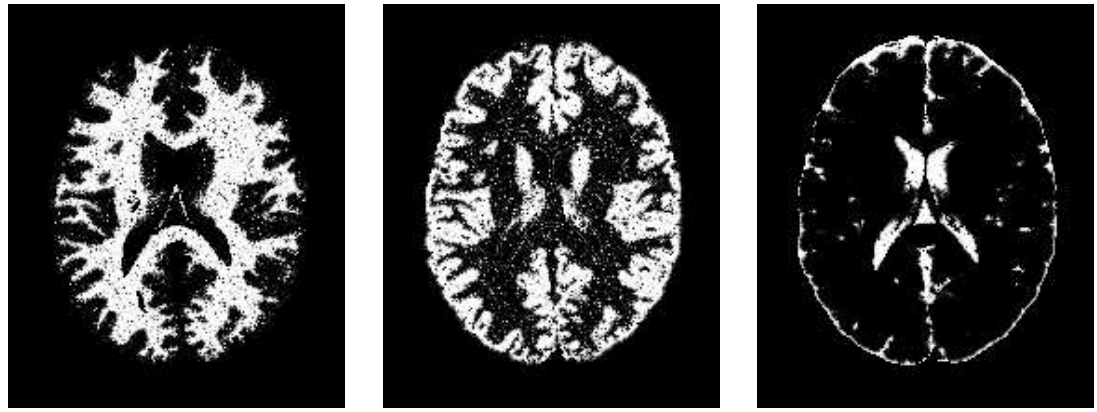


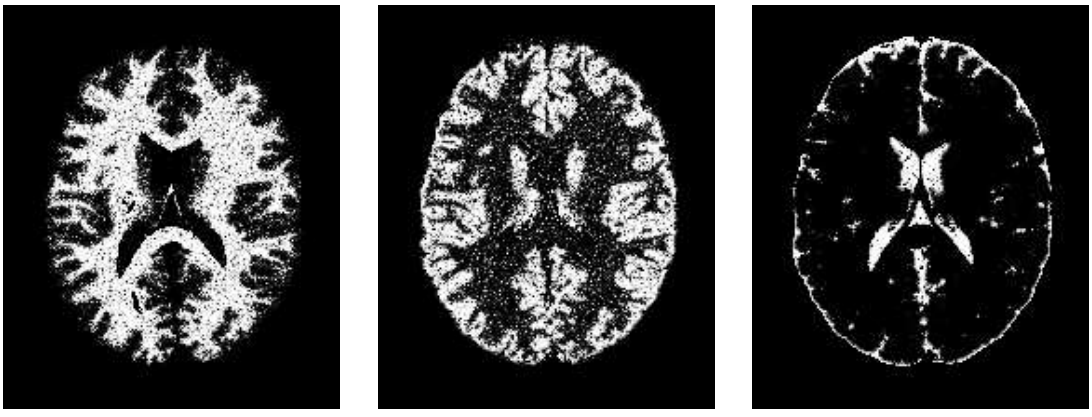
Figure 8.6: Exemplar slices from the WM, GM and CSF fuzzy tissue templates (a-c) and results obtained for classifier using model F (d-r), continued on next page. (d-f) WM, GM and CSF results for 1% noise; (g-i) WM, GM and CSF results for 3% noise; (j-l) WM, GM and CSF results for 5% noise; (m-o) WM, GM and CSF results for 7% noise; (p-r) WM, GM and CSF results for 9% noise. Note holes in WM classification are due to holes in the brain mask, illustrated in figure 8.2.



(j) (k) (l)



(m) (n) (o)



(p) (q) (r)

Figure 8.6: *continued.*

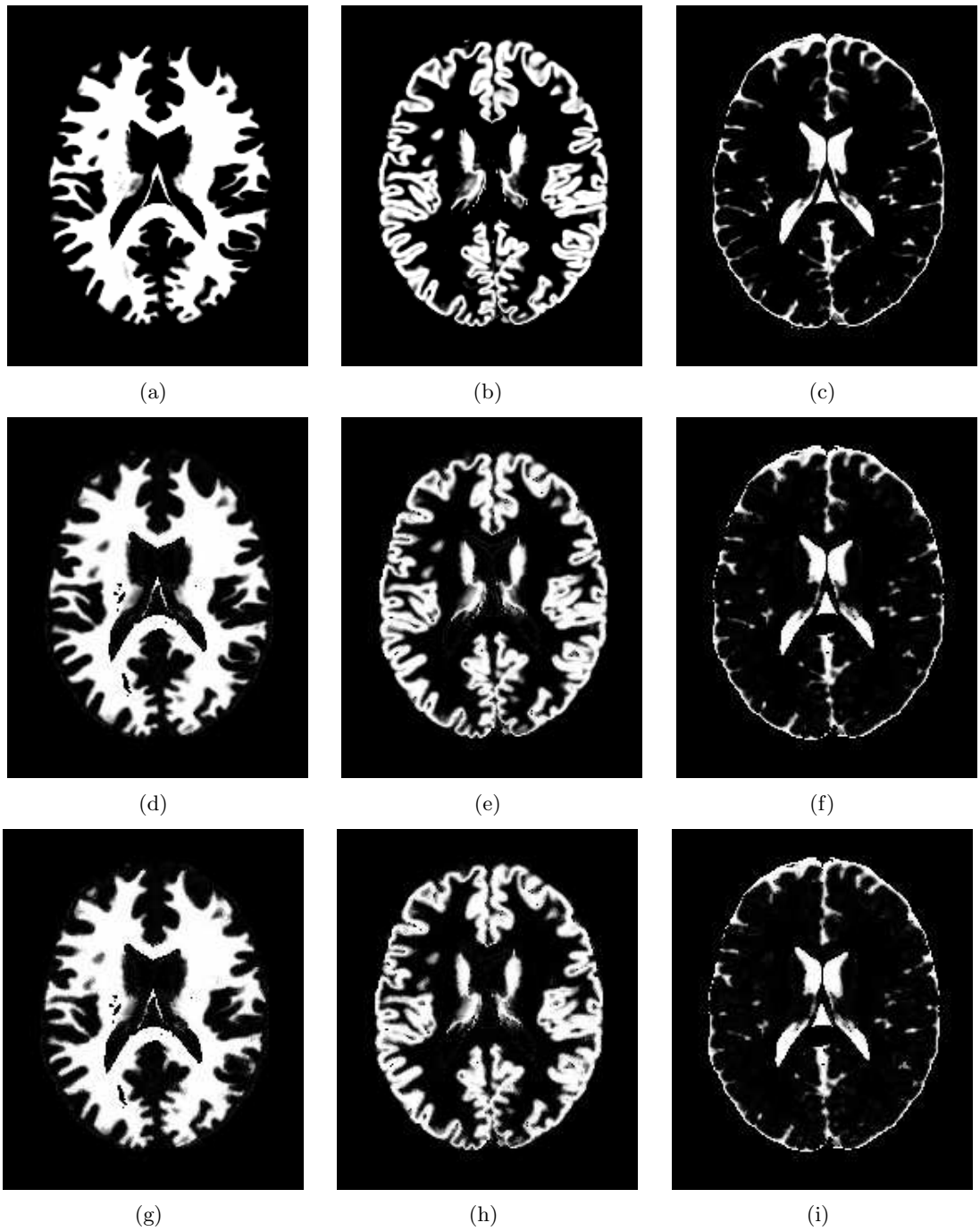
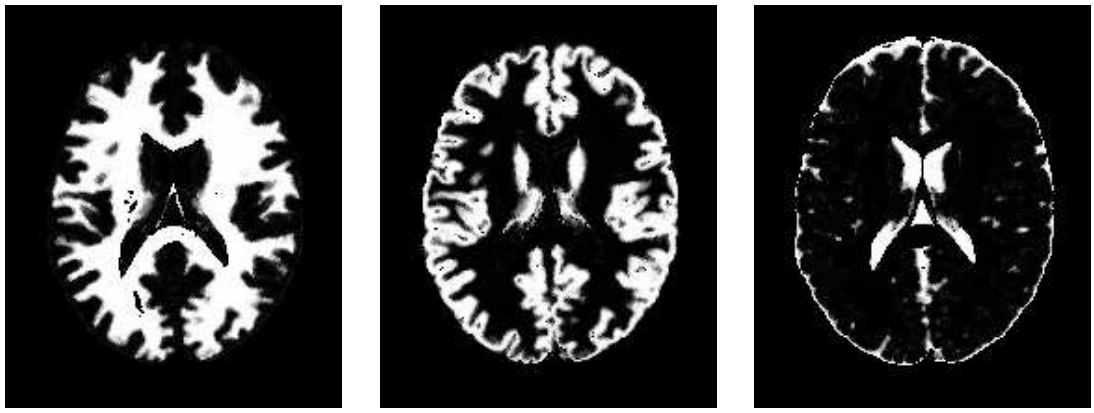
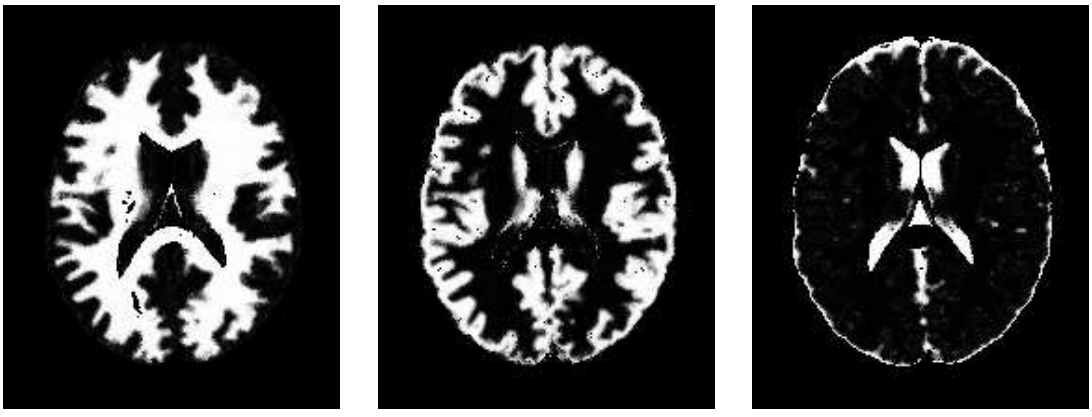


Figure 8.7: Exemplar slices from the WM, GM and CSF fuzzy tissue templates (a-c) and results obtained for classifier using model G (d-r), continued on next page. (d-f) WM, GM and CSF results for 1% noise; (g-i) WM, GM and CSF results for 3% noise; (j-l) WM, GM and CSF results for 5% noise; (m-o) WM, GM and CSF results for 7% noise; (p-r) WM, GM and CSF results for 9% noise. Note holes in WM classification are due to holes in the brain mask, illustrated in figure 8.2.



(j) (k) (l)

(m) (n) (o)



(p) (q) (r)

Figure 8.7: *continued.*

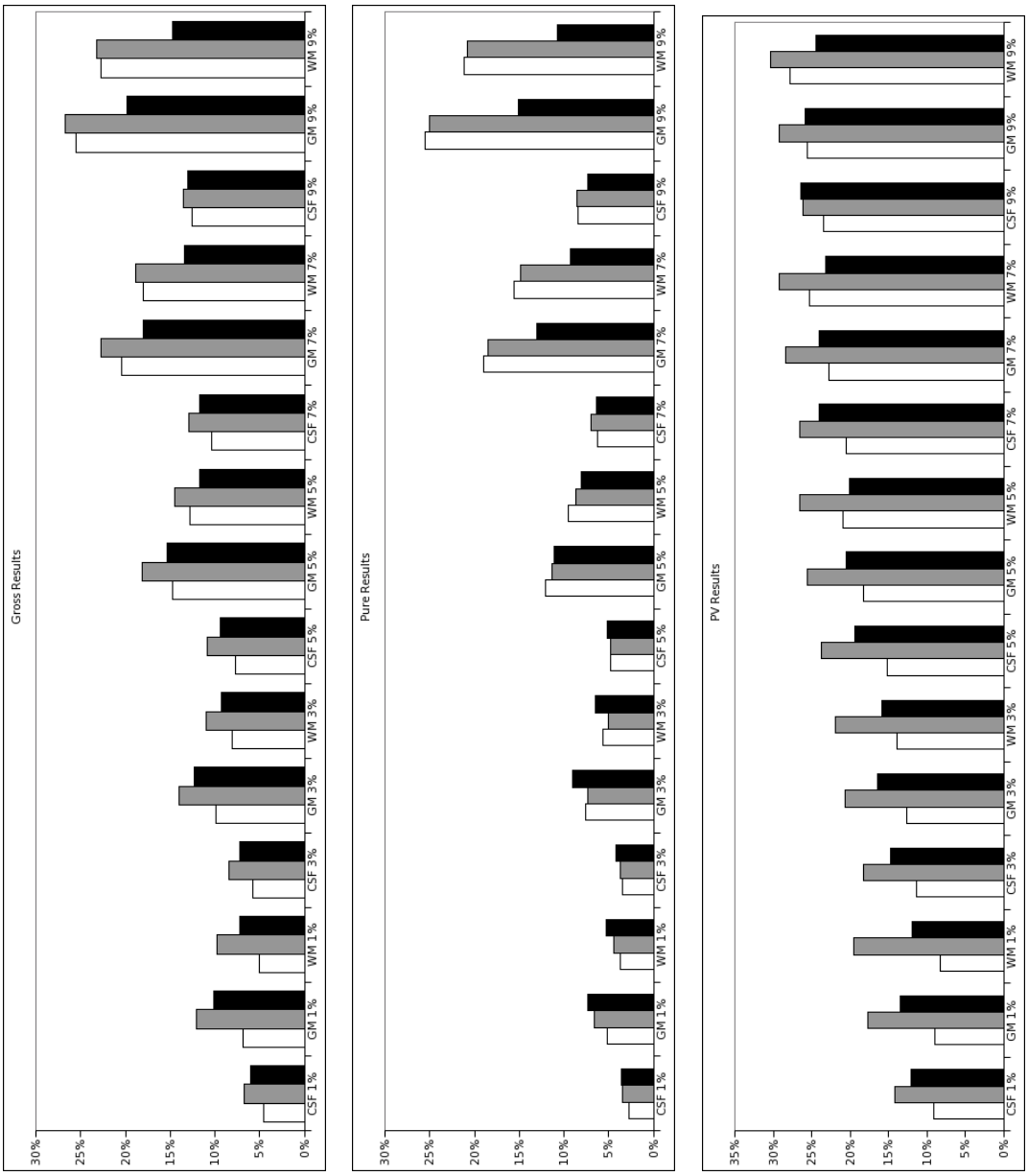


Figure 8.8: Gross, pure and PV voxel RMS error results for the simulated brain data sets, details of which can be seen in table 8.1. Voxel RMS errors are shown for the analogous intensity based PV model (B), in white; the intensity and 3-D gradient magnitude reformulated PV model (F), in grey; and the GSR PV model (G), in black. Note that for any model, GM appears to be the most challenging class for classification. This may be because it lies between the extremal WM and CSF intensity classes.

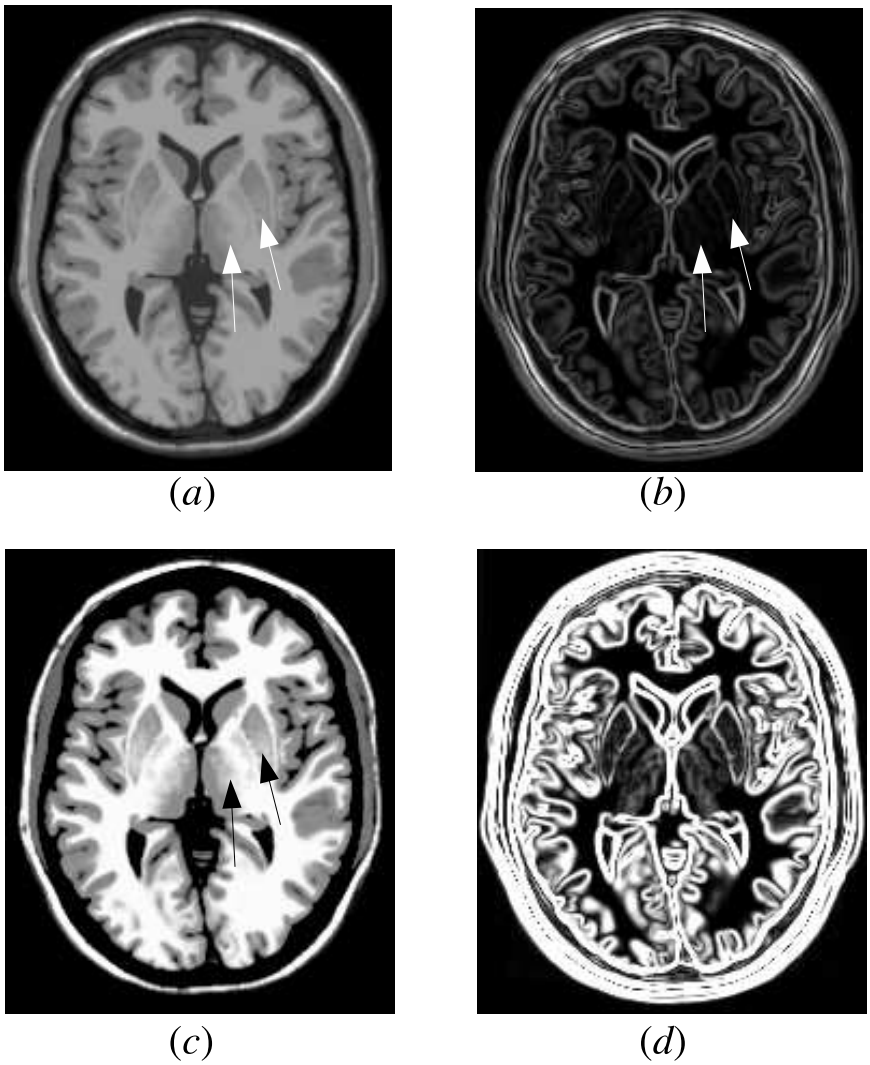


Figure 8.9: Illustration of relatively low gradient magnitude brain regions composed of a large population of PV voxels (marked by arrows) with an exemplar slice from the 0% noise simulated McGill brain data (a,c) and the result of a Sobel gradient magnitude convolution operation (b,d) both with different window widths and centres to emphasise the PV voxels. Window centres and widths are: (a) 1843.5, 3687; (b) 652.5, 1305; (c) 1709, 1400;(d) 133.5, 267.

Comparison with Further Classification Results

The comparison of results obtained with models B, F and G has helped to illustrate the different properties of these three classifiers and their suitability with regard to classifying MR neurological data. It is now instructive to compare the results of the best performing PV classifiers (B and G) with other classifiers that have previously been assessed using the McGill simulated brain data.

To aid identification the classifier utilizing model B is referred to here as TRI, short for Gaussian Triangle Convolution. Similarly, the classifier utilizing model G is referred to here as GSR, short for Gradient Spatial Regularisation.

Shattuck *et al.* in 2001, [128], tested brain segmentation techniques that modelled the pure tissue components as Gaussian distributions and the PV components as a convolution of these Gaussian distributions with a uniform distribution. Two techniques were tested, one in which no spatial prior was modelled and the other in which a discrete spatial prior was used. Both techniques relied upon the maximisation of the joint probability over all the voxels. These two techniques are referred to here as Shattuck Maximum A Posteriori (SMAP) and Shattuck Maximum Likelihood (SML) techniques. See table 8.4 for a summary.

Table 8.4: Classifiers previously used to classify simulated brain data

Acronym	*Refs*	*Note*
TRI	[148];[here,ch.5]	Gaussian Triangle Convolution
GSR	[here,ch.7]	Gradient Spatial Regularisation
FCM	[36, 101]	Fuzzy 'C' means
AGEM	[100, 101, 137]	Generalised Expectation Maximisation
SETS1	[101]	Statistical Estimation of Tissue Spread
SETS2	[101]	Statistical Estimation of Tissue Spread
SMAP	[128]	Shattuck Maximum A Posteriori
SML	[128]	Shattuck Maximum Likelihood

Pham and Prince in 2000, [101], published the results of testing a number of classification techniques applied to the simulated brain data. The primary models in their paper, referred to here as Statistical Estimation of Tissue Spread 1 and 2 (SETS1 and SETS2), relate the tissue proportion random variables to the spreading of the image data as a result of the image PSF. These *spread coefficients* are limited to two class mixtures (as was done for the models in chapters 5 and 6) but also incorporate a spatial prior which applies homogeneous spatial regularisation. Similar to [128], Pham and Prince also chose to maximise their posterior probabilities. SETS1 and SETS2 differ only by the value of a stationary prior used to manipulate likely pairs of tissue mixtures. They compare the results of applying SETS1 and SETS2 with two other previously published techniques: Fuzzy 'C' Means (FCM) which is a clustering algorithm that includes proportional class membership parameters, where a voxel may have multiple (in different proportions) class membership to model the variable voxel content [36]; and a Gaussian based Generalised Expectation Maximisation Algorithm (AGEM) which models only pure tissue types which are modelled as Gaussian distributions, but they

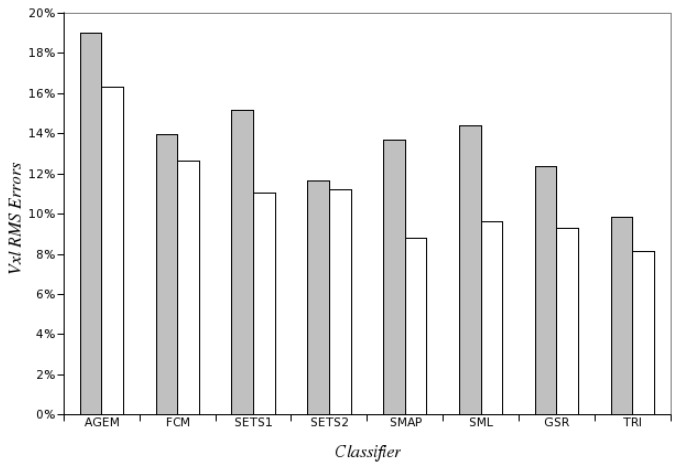


Figure 8.10: GM (grey) and WM (white) gross voxel RMS errors obtained for the PV classifiers on the simulated brain data set with 3% noise. See table 8.4 for a summary of the classifiers.

also incorporate a discrete based spatial smoothness constraint. These four methods are also summarised in table 8.4.

Pham and Prince limit their analyses to the 3% noise simulated brain data. Therefore, the first comparative assessment combines the results obtained by Pham and Prince, [101], and Shattuck *et al.* [128] for the simulated brain data with 3% noise, see figure 8.10 and table 8.5. The results illustrate that TRI has the best performance overall and GSR also has very good performance in comparison to the other classifiers. TRI has good performance because of the high contrast in the 3% simulated brain data set (see figure 8.1 and table 8.3). As was seen in chapter 7, little benefit is seen from incorporating spatial information at high CNR values, due to the intensity information in the voxel being more useful over the neighbourhood information at high CNR values. Most of the classifiers seem to have similar classification performances, except perhaps AGEM and FCM. AGEM does not even model the PV effect, so this result is not surprising. Similarly, FCM is based on proportional membership functions not a probabilistic formulation of the PV effect, although it does demonstrate somewhat better performance over AGEM. This is probably due to similarities between the continuous membership function of the FCM algorithm and the continuous nature of the PV effect.

Table 8.5: Gross voxel RMS errors obtained for the PV classifiers on the simulated brain data set with 3% noise, 0% inhomogeneity.

Tissue	**AGEM**	**FCM**	**SETS1**	**SETS2**	**SMAP**	**SML**	**GSR**	**TRI**
GM	19%	14%	15%	12%	14%	14%	12%	10%
WM	16%	13%	11%	11%	9%	10%	9%	8%

The classification performance of classifiers GSR and TRI were then compared with the remaining results of classifiers SMAP and SML for the simulated brain data sets

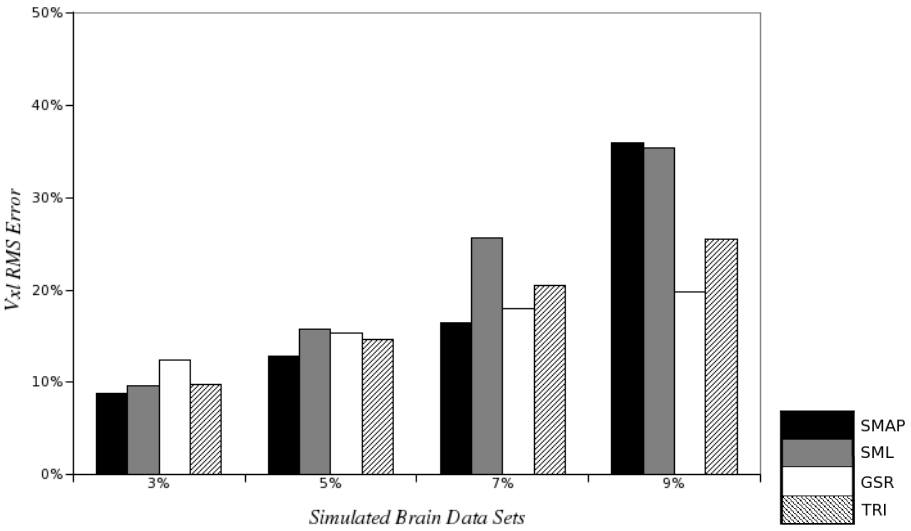
with 5%, 7% and 9% noise values, given in [128]. The results can be seen in figure 8.11 and table 8.6.

Notice the consistently poorer performance of SMAP and SML in relation to GSR and TRI for the WM in comparison to the GM results. The trend of the results for GSR and TRI appear to be consistent with the increased difficulty of correctly classifying GM voxels due to GM intensities taking values between CSF and WM intensities. It is not apparent as to why SMAP and SML perform worse for the WM voxels.

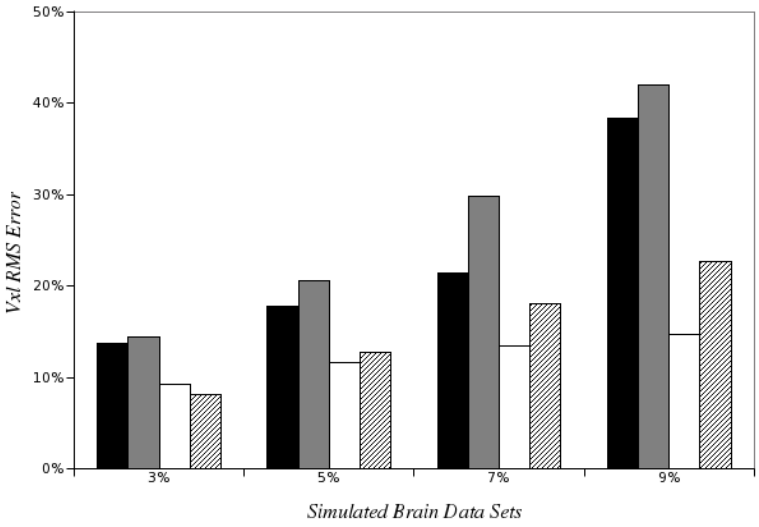
Table 8.6: Gross voxel RMS errors obtained for the PV classifiers on the simulated brain data sets with different noise values.

Data Set	*Tissue*	**SMAP**	**SML**	**GSR**	**TRI**
3%	*GM*	9%	10%	12%	10%
	WM	14%	14%	9%	8%
5%	*GM*	13%	16%	15%	15%
	WM	18%	21%	12%	13%
7%	*GM*	17%	26%	18%	21%
	WM	22%	30%	13%	18%
9%	*GM*	36%	35%	20%	26%
	WM	38%	42%	15%	23%

The comparison with SML and SMAP illustrate superior performance for GSR over SML and SMAP. SML does not incorporate any spatial information and therefore would not be expected to possess better performance over a model that does such as SMAP, particularly for data with small CNR values. This is illustrated by the somewhat least optimal performance of SML over all the other models. Quite surprisingly, TRI performs better than both SMAP and SML. TRI is similar to SML in that no spatial information is included in the model, but TRI does divide the PV distribution into two equal but opposite triangular distributions. This division enables individual class memberships to be determined and is in fact approximately equivalent to computing the expected tissue content per voxel, as demonstrated in chapter 5. In contrast, SMAP and SML treat all PV voxels as a single classification class, unlike GSR and TRI, which is probably another reason for their relatively poor performance. Other reasons for the poor performance of SMAP and SML might be due to the method in [128] incorporating an intensity inhomogeneity correction step. This step may in fact distort the true intensity values of this homogeneous data, thereby reducing the validity of their PV models. Similarly they utilized a brain mask that was derived from a separate skull-stripping procedure so that erroneous voxels were probably included and contributed toward a poor result for the final stage entailing PV classification. However, one can observe the differences between the exemplar classification result images presented in [128] and the results obtained for GSR illustrated in figure 8.7. There appears to be significant differences in classification quality that are difficult to attribute to a poor brain mask from the preceding skull-stripping algorithm.



(a) GM gross voxel RMS errors.



(c) WM gross voxel RMS errors.

Figure 8.11: Gross voxel RMS errors for (a) GM and (b) WM tissues for classifiers: SMAP (black, [128]), SML (grey, [128]), GSR (white, [here, ch. 7]), TRI (hatched, [148]).

8.2 Classification of PET Phantom Data

The McGill simulated MR brain phantom simulates the MR imaging process. As discussed in chapters 1 and 2, the PV effect is not limited to MRI data. Nuclear imaging data, such as PET imaging data, is also affected by the PV effect. In fact the relatively small sizes of voxels in functional PET data (e.g. $4\times4\times4\text{mm}^3$) in comparison to a structural modality such as MRI (e.g. $1\times1\times1\text{mm}^3$) produces proportionally more PV voxels.

The previous section utilized a MR simulation of the brain to assess the performance of the classifiers under more realistic conditions. This section takes a step further by utilising lower resolution real functional PET imaging data in combination with a noiseless ground truth derived from structural high resolution CT imaging data. CT also utilizes ionising radiation, in the form of X-rays rather than the gamma rays that are used in PET imaging. The X-rays do not originate from inside the patient, instead they are generated external to the patient and then the amount of X-ray attenuation is measured by sensors on the opposing side of the patient. The amount of attenuation is dependent on the variable density of the tissue found along vectors through the patient. The data collected by the CT scanner are then in a similar format to the PET data, in that the information are in the form of integrals over different vectors through the patient. This information is then reconstructed into the spatial domain to produce images that are comprehensible by a clinician. A CT scanner produces higher resolution images in comparison to PET and can therefore be used to determine the 'ground truth' image information at the relatively lower PET imaging resolution, particularly for a carefully designed phantom structure.

The imaging data used in this section to assess the performance of the PV classifiers was obtained from a combined PET/CT scanner that acquires the CT images simultaneously to the PET images. This is particularly useful as the issue of registering the PET images to the same origin as the CT images is already solved.

8.2.1 Experimental Methodology: Ground Truth Parameters

Experimental work was undertaken at the Royal Marsden Hospital, Sutton, UK using a Phillips Gemini PET/CT scanner to acquire registered PET/CT images of a phantom suitable for assessing the classification performance of PV classifiers. Some preliminary classifier assessment was undertaken in [5], but was limited to a classifier using model D in conjunction with non-uniform global PV prior distributions, such as those described in chapter 7, section 7.1. The classifier and assessment software used to carry out the performance assessment by Barry was created by the author. The analysis was also limited to two class manually defined Region of Interests (ROI)s.

The assessment here is developed further to compare the performance of PV classifiers using neighbourhood information, GSR (chapter 7, model G), the global PV prior intensity based analogous PV classifier (chapter 5, model B) and the reformulated 3-D gradient magnitude classifier (chapter 6). This selection of classifiers is the same as for the preceding set of experiments on the simulated brain data. Each classifier represents

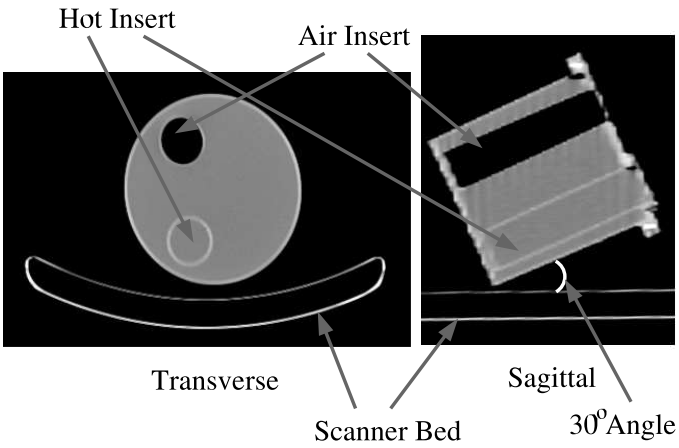


Figure 8.12: Illustration of PET phantom with exemplar transverse and sagittal image slices from the high resolution CT data.

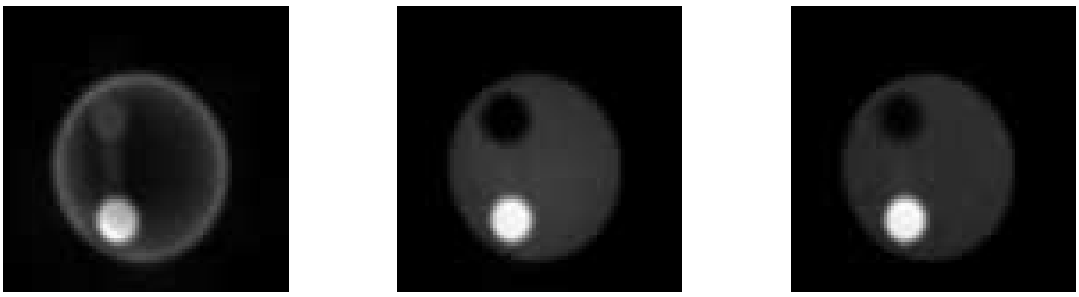
either the best performing or is representative for that type of classifier.The ROI is also extended to include three classes, excluding only the air based background that does not constitute part of the imaging phantom.

Data Acquisition

A National Electric Manufacturers Association (NEMA) phantom was used. This NEMA phantom is a plastic (polytetrafluoroethene-PTFE) cylinder (diameter 190mm x 200mm) containing three inserts (diameter 49mm) one of which was removed. One of the two remaining inserts contained air and the other insert together with the main body of the phantom were filled with different (but homogeneous) concentrations of Gallium-68. A 5:1 activity concentration was ensured between the main phantom body background medium (Bkgrnd) and the warmer insert by adjusting the concentration of Gallium-68 in relation to the amount of water. The NEMA phantom was placed at a 30^o angle on the scanner bed to produce a higher count of PV voxels. The use of Gallium-68, compared to the more common FDG (see chapter 2, section 2.2), was used to generate a larger proportion of PV voxels for the relatively large inserts used. Figure 8.12 illustrates the phantom geometry using a cross-section of the high resolution CT image data.

PET imaging data is usually processed prior to clinical presentation and quantitative analysis. Chapter 2 described various artefacts that affect PET imaging data and mentioned correction schemes employed to improve the visual and quantitative accuracy of the final image data. Therefore three different sets of data were produced, each utilising the Row-Action Maximum Likelihood Algorithm (RAMLA), [14], which is an iterative image reconstruction technique, previously described in chapter 2:

- Completely uncorrected RAMLA data (RAMLA no correction, RAMLAnc), illustrated in figure 8.13(a);
- CT attenuation corrected RAMLA data (RAMLActac), illustrated in figure 8.13(b);



(a) RAMLAnc (b) RAMLActac (c) RAMLAfc

Figure 8.13: Exemplar image slices from each of the error corrected data volumes. Notice the intensity variations in the uncorrected data volume, (a) RAMLAnc. These image slices were produced with the aid of MRIcro, [114], where the contrast auto-balance was used, which scales the highest and lowest 1% of image intensities to 255 and 0 grey levels respectively.

- CT attenuation corrected *and* Random and Scatter correction (RAMLA fully corrected, RAMLAfc), illustrated in figure 8.13(c).

It was thought that each correction step would change the classification performance of each of the PV classifiers.

Ground Truth Preparation

A ground truth representation of the imaging data was required in order to assess the quality of the output of the PV classifiers. As discussed earlier, the CT data is a higher resolution representation of the information represented by the PET imaging data[1]. A number of steps have to be carried out to prepare the information in the CT data so that it can be used as a ground truth to the PET imaging data which are described below and are also illustrated in figure 8.14.

Firstly, a simple threshold based manually initialised seeded region growing operation on the high resolution CT data was performed to define the space occupied by the cold insert. For the hot insert, the CT intensities between the hot insert contents and the surrounding warm background of the phantom were the same and the thin insert wall (1mm) did not present a great enough barrier to the growth of the region using intensities alone. Instead a seeded region growth on the hot insert wall was undertaken, where very little growth into the non-wall voxels occurred. Once the insert wall was defined, the inside to the insert was defined via a further region growing operation undertaken on each individual slice.

[1]It should be noted that the PET imaging data represents activity information where as the CT imaging represents structural information. This can be seen with the different appearance of the hot insert for the PET imaging data in comparison to the similar appearance of the hot insert in relation to the warm background for the CT imaging data. Nonetheless, for phantom data, there is a direct correspondence with the observed studies, used here for ground truth comparison. Such a correspondence can not always be assured for clinical applications.

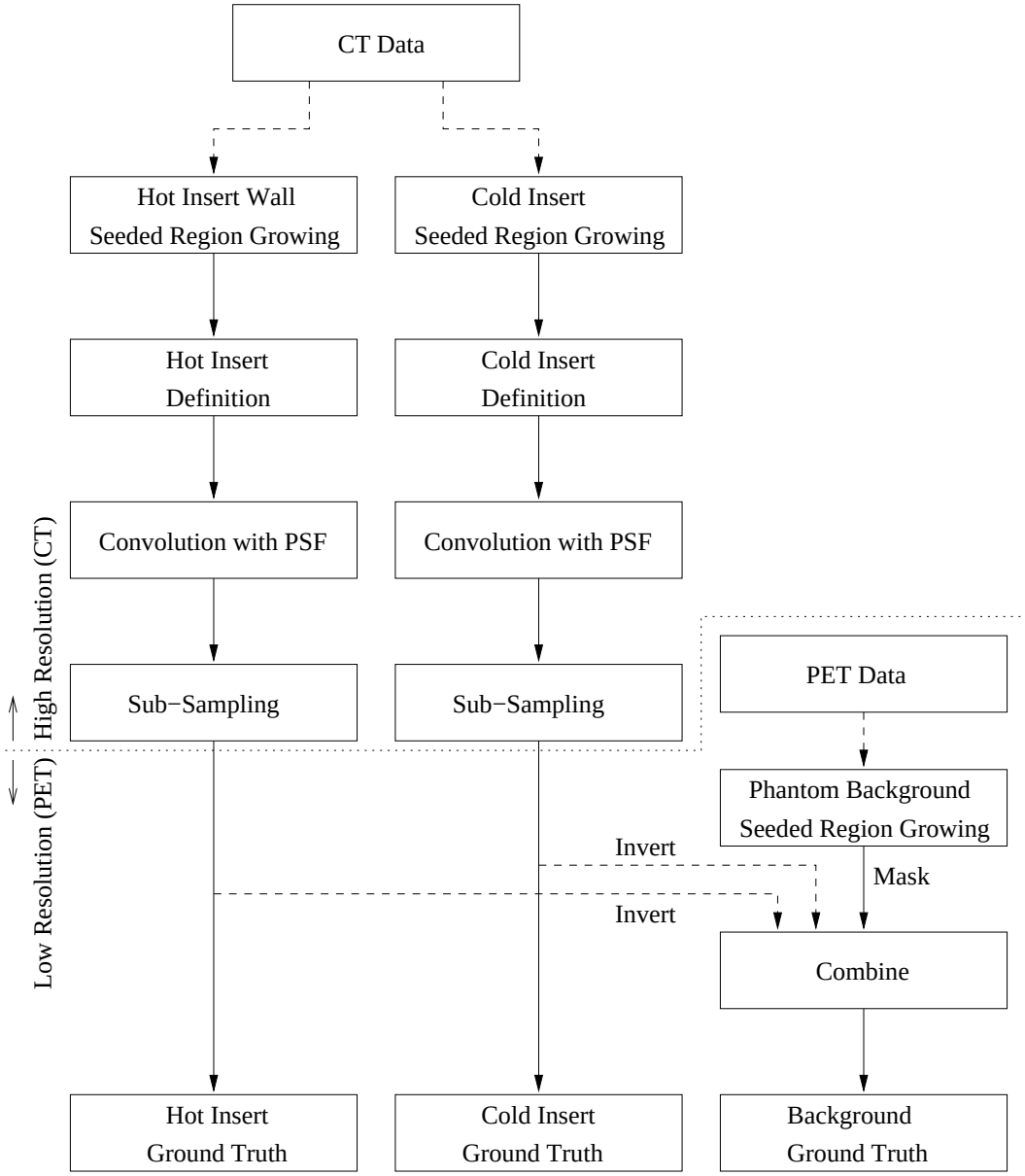


Figure 8.14: Ground truth creation process from the high resolution CT imaging data and PET PSF to the low resolution of the PET imaging data.

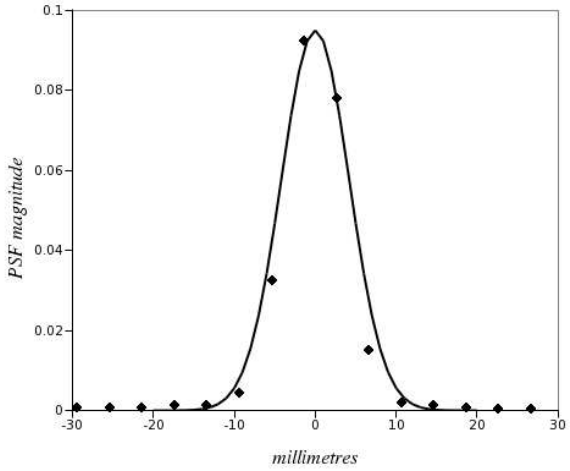


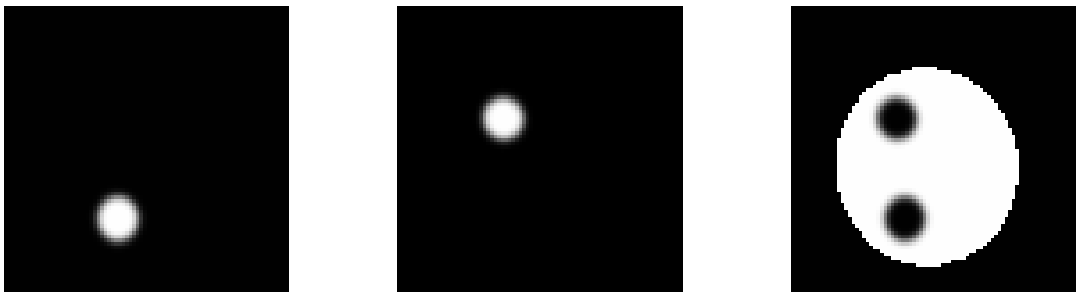
Figure 8.15: Illustration of Gaussian fitted (line) to experimental PSF measured points (squares)

After optimising both these strategies by manually adjusting the intensity range of the growing regions, the volumetric error of the segmented regions in relation to the volume calculated from the actual dimensions of the imaging phantom inserts were found to be 1.5% for the hot insert and 5.9% for the cold insert. The output of the region growing operations were then assigned to a 0 to 255 linear scale intensity range, where 0 represented an empty voxel and 255 100% voxel content. These were then convolved with a PSF that was measured from an experimental Ga-68 line spread function using the Phillips Gemini PET/CT imaging system for the PET modality. The PSF is illustrated in figure 8.15.

These PSF convolved noiseless data volumes can then be referred to as "ground truth" representations of the noisy PET data but sampled at the higher CT resolution. The CT data were acquired at resolution: $1.17 \times 1.17 \times 5.00\text{mm}^3$ in comparison to the PET data, acquired at $4.00 \times 4.00 \times 4.00\text{mm}^3$. The output of the classifiers would, however, be at the lower PET resolution. To enable comparison, the ground truth data were sub-sampled (using linear interpolation for non-integer sub-samples) to create the lower resolution ground truth data volumes. These are illustrated in figure 8.16.

The groundtruth for the background region of the image data was created by combining the two insert ground truths (hot and cold) into a single volume then inverting the intensities (i.e. $0, 255 \rightarrow 255, 0$) and limiting this inversion to the bounds of the phantom. The bounds of the phantom was also defined via a simple seeded region growing process but on the PET imaging data as the ground truths were, at this stage, at the lower PET resolution. The bounds of the phantom were eroded to limit any PV effect from the boundary between the air background and the phantom background. The ground truth to the phantom background can also be seen in figure 8.16.

The means, standard deviations and prior probabilities were then calculated from a combination of these ground truth volumes and the PET image data.



(a) Hot Insert (b) Cold Insert (c) Background

Figure 8.16: Exemplar image slices from each of the ground truth data volumes. White (voxel intensity= 255) represents 100% content and black (voxel intensity= 0) represents 0% content.

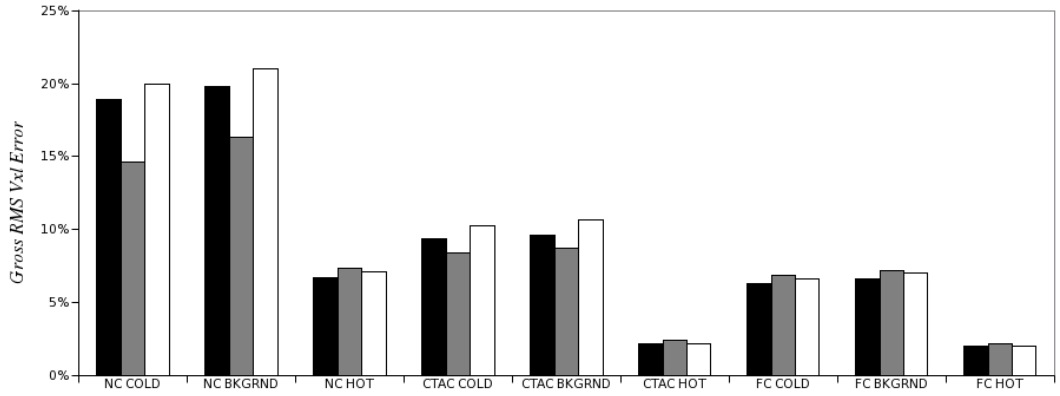
8.2.2 Results

The results of classifying the PET phantom data with the various levels of error correction can be seen in figure 8.17 for each of the three classifiers (PV models B,F and G), using the ground truth calculated parameter values. These result illustrate that the most significant reduction in error is seen between the PET data without any correction and with the CT attenuation correction. The correction for random and scatter errors also appears to produce improved PV classifier performances as well, although these latter correction schemes appear to have a less significant effect on classification.

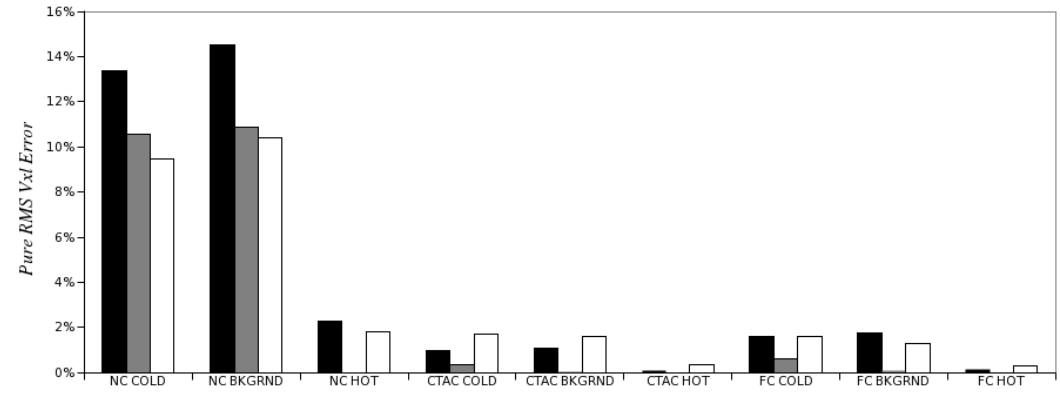
Model F, the classifier utilizing both intensity and localised image gradient magnitude information appears to possess better performance over models B and G for some of the PET phantom data. This was predicted by the simulated PV data results of chapter 6. This contrasts with the results obtained in the previous section of this chapter, where the performance of these classifiers were compared on simulated MR brain data. Simulated or real MR brain data is quite different from this PET imaging data in a number of ways. The simple geometry of the PET phantom is very similar to the simple geometry of the simulated PV data that was used in chapters 5,6 and 7. This is in contrast to the complex non-boundary type PV artefacts that are associated with MR images of the brain, which were discussed in chapter 2, section 2.4.

The gradient controlled spatial regulated PV classifier (model G) has not performed any better than models B or F. This can be understood from the fact that this PET phantom data possesses high CNR values for the hot and cold inserts in relation to the warm background of the phantom (see table 8.7). As was predicted by the experiments in chapter 7, model G did not offer any significant performance benefits for sufficiently high CNR values. Recall that this is probably due to the fact that neighbourhood information can not offer any additional useful information for high CNR data.

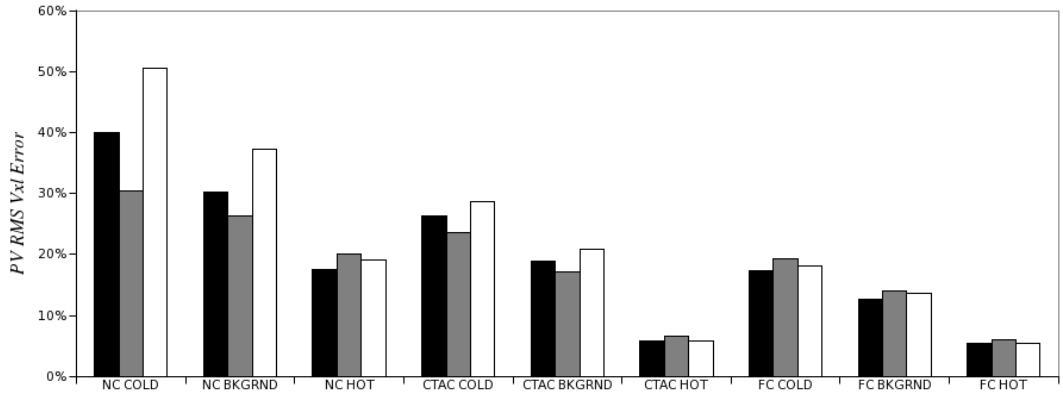
The PV voxel RMS error results illustrated in figure 8.17 are also higher than might be expected for the PV data possessing high CNR values. The pure voxel RMS errors appear to be quite acceptable. The larger PV errors are probably due to minor inaccuracies in the preparation of the ground truth data. In particular, the exact content of each of the PV voxels depends on a good approximation of the PSF of the scanner.



(a) Gross voxel RMS errors.



(b) Pure voxel RMS errors.

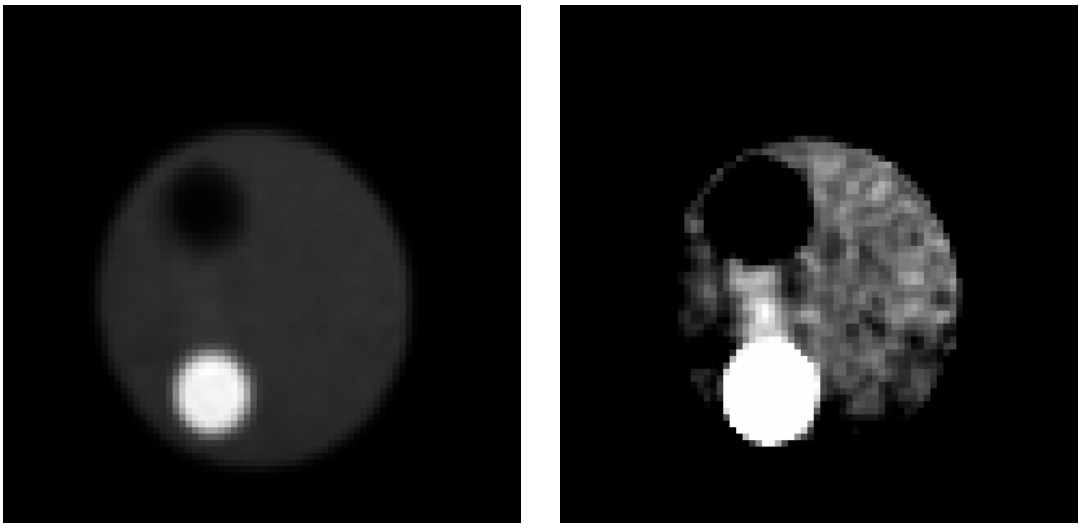


(c) PV voxel RMS errors.

Figure 8.17: Voxel RMS errors for classifiers using PV models B (black), F (grey) and G (white) using parameters calculated from the ground truth data. Key to abbreviations: NC, not corrected; COLD, cold insert; BKGRND, background; HOT, hot insert; CTAC, CT attenuation corrected; FC, fully corrected for attenuation and randoms and scatter.

Table 8.7: PET phantom data CNR values, calculated from the CT based ground truth data.

Data	*Hot/Bkgrnd*	*Cold/Bkgrnd*
RAMLAnc	11	2
RAMLActac	51	21
RAMLAfc	63	22

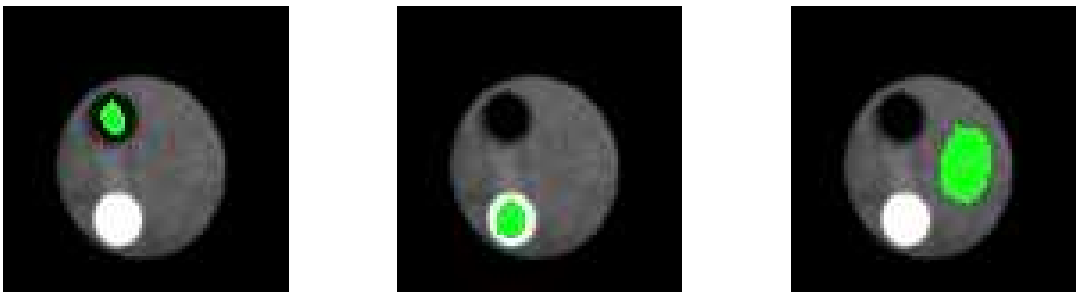


(a) RAMLAfc window centre,width=1397,2790 (b) RAMLAfc window centre,width=517,130

Figure 8.18: Illustration of inhomogeneities still present in the fully corrected PET imaging data. Notice in (b) light and dark regions in the warm background. Due to the high contrast between the hot insert and warm background these inhomogeneities are not apparent for a window centre/width automatically calculated with MRIcro (98% intensities scaled to 0 to 255). The inhomogeneities are small in relation to the CNR but they are not quite so small in relation to the SNR.

The volumetric error between the segmented inserts and the volume calculated from the actual insert dimensions were greater for the cold insert (5.9% in comparison to 1.5%). This difference in error helps, in part, to explain the greater errors seen for the classifications, where greater voxel RMS errors are consistently seen for the cold insert in comparison to the voxel RMS errors for the hot insert.

Also, the RAMLA constructed PET imaging data on close inspection appears to possess correlation of the noise and even for the fully corrected images (RAMLAfc), non-homogeneities exist between the two inserts (illustrated in figure 8.18). These imperfections that are not included in the PV models may also contribute to a degradation in the expected performance of the classifiers.



(a) Air insert initialisation ROI

(b) Hot insert initialisation ROI

(c) Warm background initialisation ROI

Figure 8.19: Illustration of ROIs used to initialise the clustering algorithms. Each image is the same exemplar transverse image slice through the RAMLAfc PET imaging data.

8.2.3 Experimental Methodology: Automatic Parameter Estimation

The classifier using PV model G in the previous section utilized parameters determined from the CT based ground truth. A framework has already been presented that allows classifier model G to estimate the parameters automatically (see chapter 7). Therefore it was interesting to see whether the parameter estimation scheme could operate successfully for the PET phantom data and also produce classified data volumes with similar results to those already presented.

Alternative Initialisation Previously the parameter estimation step described in chapter 7 was initialised via an unsupervised minimum distance spatially aware clustering algorithm, described in algorithm 5, on page 187. The clustering algorithm was itself initialised by dividing the CDF of the data into C equally spaced clusters. It was found for the PET phantom data of this set of experiments, that this initialisation step was not suitable. This was due to the relatively large distance between the hot insert and the warm background in comparison to the smaller distance between the warm background and the cold air insert. Therefore ROIs were manually defined to select realistic initialisation values for each of the inserts and the warm background. Three ROIs are illustrated in figure 8.19.

The mean intensities from these ROIs were calculated and used in place of steps 2 and 3 in the clustering algorithm, algorithm 5, (page 187).

Parameter estimated classification results The voxel RMS errors for the estimated parameter classification together with the ground truth estimated parameter classification are illustrated for comparison in figure 8.20. Tables containing numerical values of these errors together with the results obtained for models B and F of the previous section can be seen in tables 8.8, 8.9 and 8.10.

The CNR values calculated using the parameter estimation scheme of chapter 7 can be seen in table 8.11. These CNR values appear to be quite similar to the CNR values calculated from the CT based ground truth, (table 8.7).

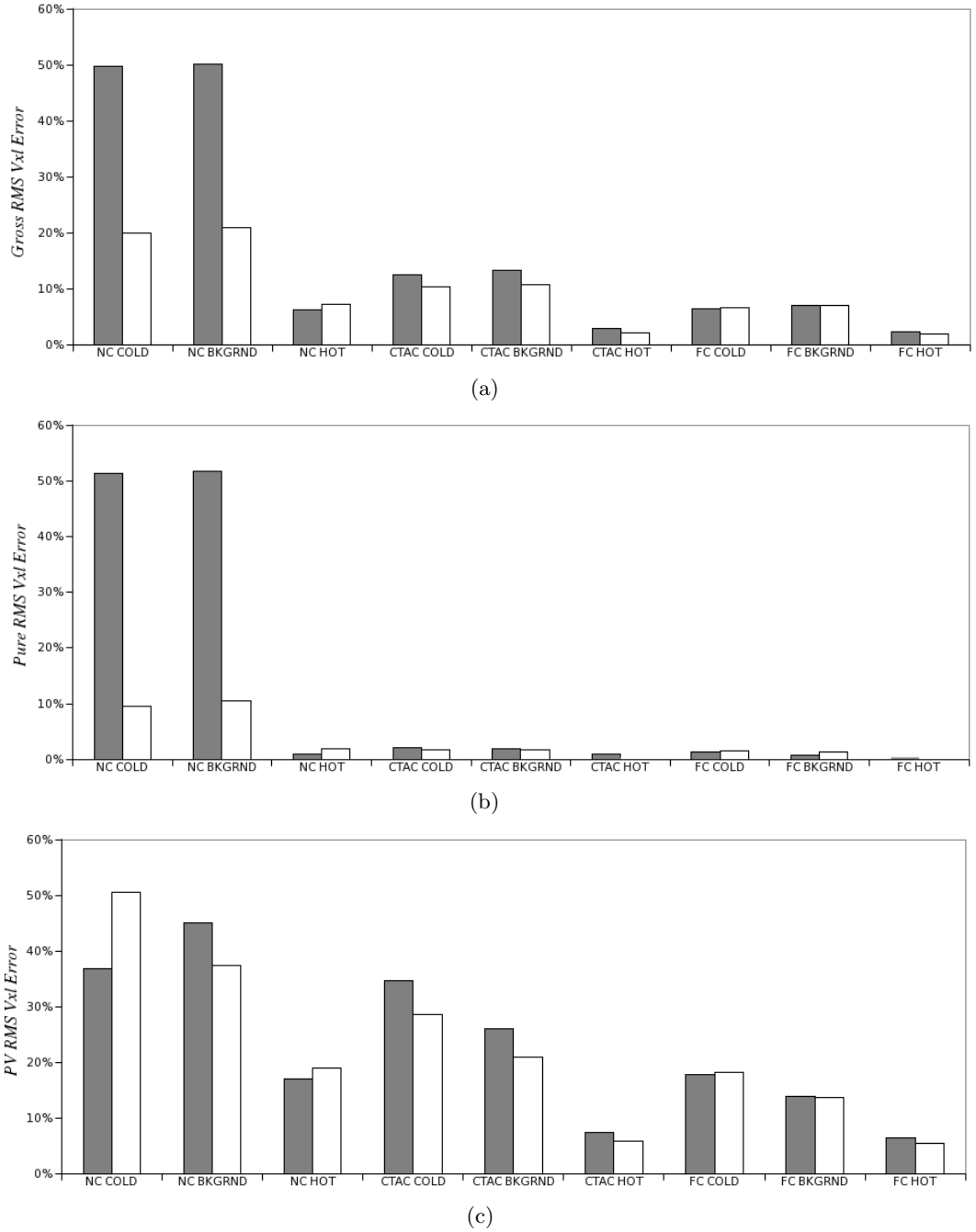


Figure 8.20: Voxel RMS errors comparing the performance of GSR PV classifier (model G), using ground truth estimated parameters (white) and automatically estimated parameters (grey). Key to abbreviations: NC, not corrected; COLD, cold insert; BKGRND, background; HOT, hot insert; CTAC, CT attenuation corrected; FC, fully corrected for attenuation and randoms and scatter.

Table 8.8: PET Results, gross Vxl RMS Errors. Model G(i) ground truth parameters; model G(ii) automatically estimated parameters.

		Classifier			
Data	*Region*	**B**	**F**	**G**(i)	**G**(ii)
RAMLAnc	Cold	19%	15%	20%	50%
	Bkgrnd	20%	16%	21%	50%
	Hot	7%	7%	7%	6%
RAMLActac	Cold	9%	8%	10%	12%
	Bkgrnd	10%	9%	11%	13%
	Hot	2%	2%	2%	3%
RAMLAfc	Cold	6%	7%	7%	6%
	Bkgrnd	7%	7%	7%	7%
	Hot	2%	2%	2%	2%

Table 8.9: PET Results, PV Vxl RMS Errors. Model G(i) ground truth parameters; model G(ii) automatically estimated parameters.

		Classifier			
Data	*Region*	**B**	**F**	**G**(i)	**G**(ii)
RAMLAnc	Cold	40%	30%	51%	37%
	Bkgrnd	30%	26%	37%	45%
	Hot	17%	20%	19%	17%
RAMLActac	Cold	26%	24%	29%	35%
	Bkgrnd	19%	17%	21%	26%
	Hot	6%	7%	6%	7%
RAMLAfc	Cold	17%	19%	18%	18%
	Bkgrnd	13%	14%	14%	14%
	Hot	5%	6%	5%	6%

Table 8.10: PET Results, pure Vxl RMS Errors. Model G(i) ground truth parameters; model G(ii) automatically estimated parameters.

		Classifier			
Data	*Region*	**B**	**F**	**G**(i)	**G**(ii)
RAMLAnc	Cold	13%	11%	9%	51%
	Bkgrnd	15%	11%	10%	52%
	Hot	2%	0%	2%	1%
RAMLActac	Cold	1%	0%	2%	2%
	Bkgrnd	1%	0%	2%	2%
	Hot	0%	0%	0%	1%
RAMLAfc	Cold	2%	1%	2%	1%
	Bkgrnd	2%	0%	1%	1%
	Hot	0%	0%	0%	0%

The results in figure 8.20 appear to illustrate reasonable agreement between the classifier performances of model G with and without the automatic parameter estimation. The difference between the two is greatest for the uncorrected data volume, (RAMLAnc). This is probably due to the iterative nature of the estimation procedure which attempts to adjust the parameters and voxel content depending on the amount of regularisation in the vicinity of a voxel. The quite significant inhomogeneities, (see fig. 8.13(a)), in this uncorrected data due to the attenuation of the photons presents data points that do not conform to the PV model. This is not the case for the corrected data volumes that appear to be implicitly well-modelled by model G.

Table 8.11: PET phantom data automatically estimated CNR values, calculated using the parameter estimation described in chapter 7.

Data	*Hot/Bkgrnd*	*Cold/Bkgrnd*
RAMLAnc	12	7
RAMLActac	31	18
RAMLAfc	70	31

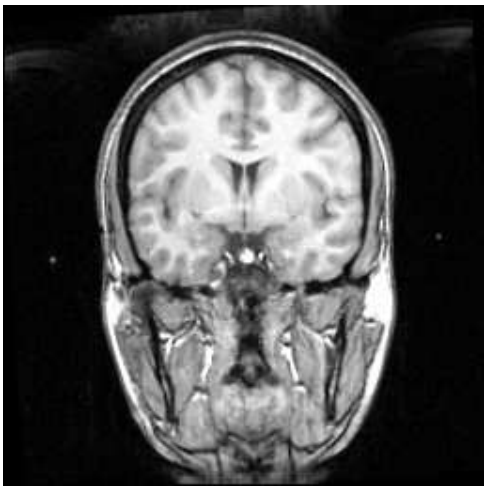
8.3 Classification of Neurological MRI Data

The classification of PV data has been limited to the concentric spheroids in chapters 5, 6 and 7; the simulated MR brain data in section 8.1; and a PET phantom in section 8.2. This section attempts to assess the performance of the gradient spatially regulated (GSR), model G, PV classifier on twenty neurological MR data sets obtained from a publicly accessible resource made available by the Center for Morphometric Analysis at Massachusetts General Hospital, USA [38]. Ground truth is available in the form of manual segmentations. The human operators who manually segmented the data sets could not (for obvious reasons) attempt to estimate the PV content for every voxel. They therefore assigned the voxels in a discrete manner, i.e. into either GM, WM or CSF. This therefore precludes any PV assessment. It does not, however, prevent some form of assessment based on greatest voxel quantity of the output of the PV classifier, in common with other PV classifiers, see e.g. [60, 107, 128].

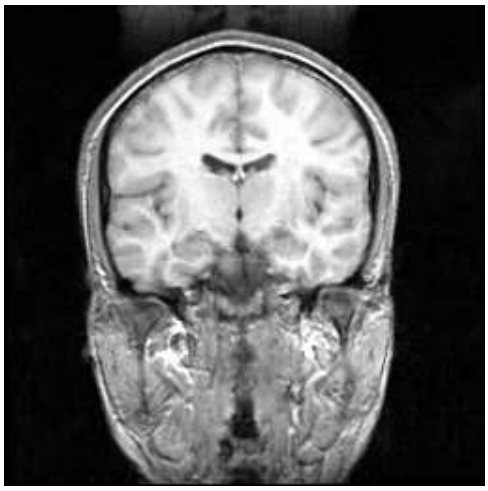
An advantage of a publicly available resource such as [38] is that detailed quantitative analysis of the segmentation results can be published. These results can be compared not only with the publicly available ground truth but also with other segmentation algorithms whose results have also been published.

8.3.1 Experimental Methodology

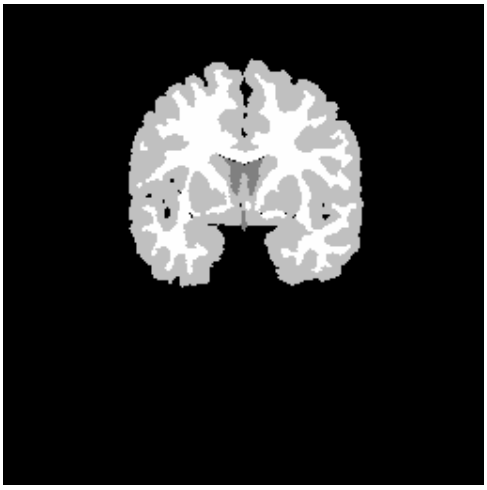
The 20 real MR data sets with ground truth were downloaded from [38], two of which are illustrated in figure 8.21 via exemplar image slices. The data sets were all acquired in the coronal axis with voxel dimensions of $1.0 \times 1.0 \times 3.0\text{mm}^3$. Linear interpolation was used to determine isotropic voxel values for the purposes of the spatial regularisation and gradient magnitude calculations.



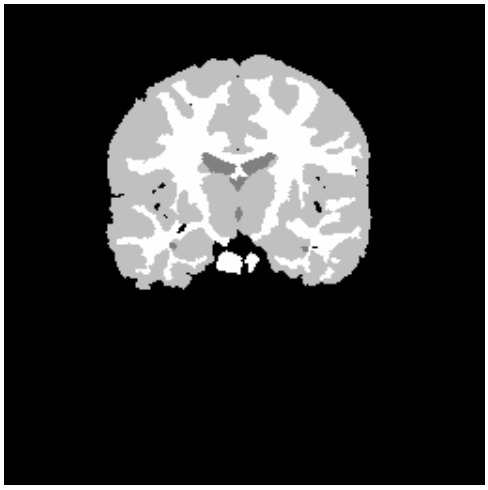
(a) 100-23



(b) 11-3



(c) 100-23 ground truth



(d) 11-3 ground truth

Figure 8.21: Exemplar coronal image slices from the 20 normal MR brain data sets together with the corresponding ground truth image slices for the same data sets. These data are available from [38].

The model G PV classifier, referred to here as GSR was then initialized and applied in the same manner as detailed in chapter 7. Initialisation was performed with the unsupervised minimum distance spatially aware classifier described in algorithm 5 on page 187.

The PV classifier was then initialised with the output of this algorithm. The PV classifier iterations were terminated with the criteria as described previously in section 7.4.2 on page 176.

The discrete voxel performance metric used by other authors to assess the performance of their algorithms on classifying the 20 normal MR brain data sets was the Jaccard similarity metric:

$$J(\Omega_{GT}|\Omega_{est}) = \frac{|\Omega_{GT} \cap \Omega_{est}|}{|\Omega_{GT} \cup \Omega_{est}|}, \tag{8.1}$$

where Ω_{GT} is the set of voxels identified by the ground truth data as belonging to a particular classification class; and Ω_{est} is the set of voxels identified as belonging to a particular classification class from the output of the discretization process post-PV classification. The numerator is the cardinality (number of set elements) of the intersection of these two sets and the denominator is the cardinality of the union of these two sets. This metric tends to 1.0 for perfect segmentations and 0.0 for imperfect segmentations. This was also adopted here for comparison purposes.

The discretized points were calculated by determining the largest individual tissue content for each voxel from the output of the PV classifier and assigning the voxel to that particular tissue class.

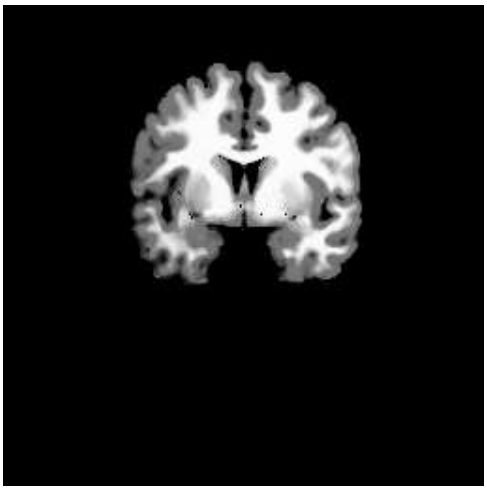
The combined GM,WM and CSF data volume were used as a brain mask, similar to the approach taken earlier in section 8.1. In practise these components could be excluded via a skull-stripping algorithm, see chapter 3.

8.3.2 Results and Discussion

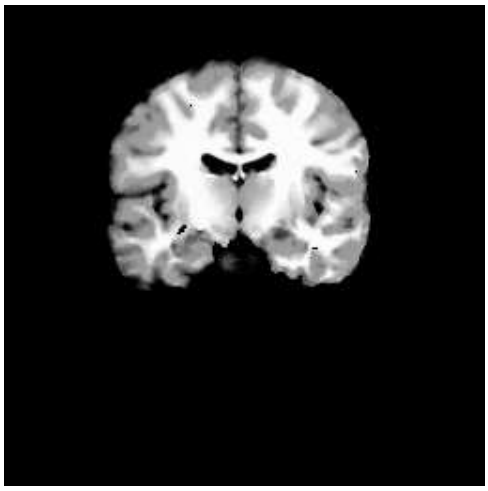
The GSR classifier and discretization process was applied to the data as described above. The PV and discretized results for the exemplar data sets previously illustrated in figure 8.21 can be seen in figure 8.22. An initial visual inspection of these exemplar slices appears to reveal moderately good agreement between the ground truth illustrated in sub-figures 8.21(c,d) and the results in figure 8.22. The discretization process appears to reduce the similarity of the classifier output when compared with what might be expected should there have been a PV ground truth for this real MR data.

The Jaccard similarity measure given by equation 8.1 was applied to the ground truth and discretized classifier output. The results of which can be seen in figure 8.23. The performance appears to be fairly consistent for 13 out of the 20 data sets, but then there are 7 other data sets that do not have such good performance. Visual inspection of these 7 particular data sets revealed that they were affected by severe inhomogeneity artefacts (see e.g. figure 8.24).

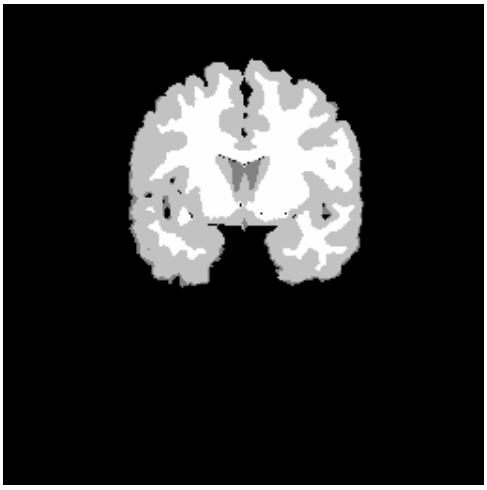
Due to the prevalence of the severe inhomogeneity artefacts in a number of the data sets the mean Jaccard similarity metric of the better classifier results was calculated. This



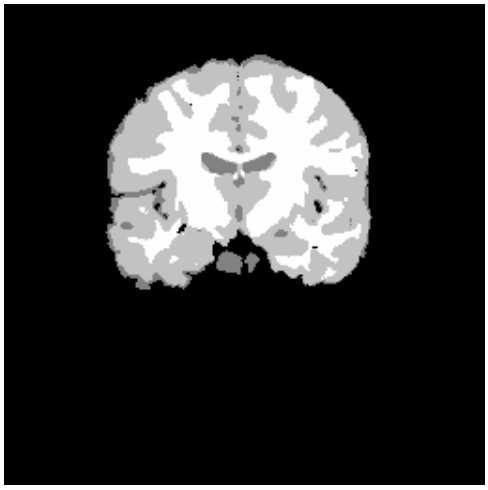
(a) 100-23 Classified PV result



(b) 11-3 Classified PV result



(c) 100-23 Classified discretized result



(d) 11-3 Classified discretized result

Figure 8.22: Exemplar coronal classification results for the MR data sets previously illustrated in figure 8.21.

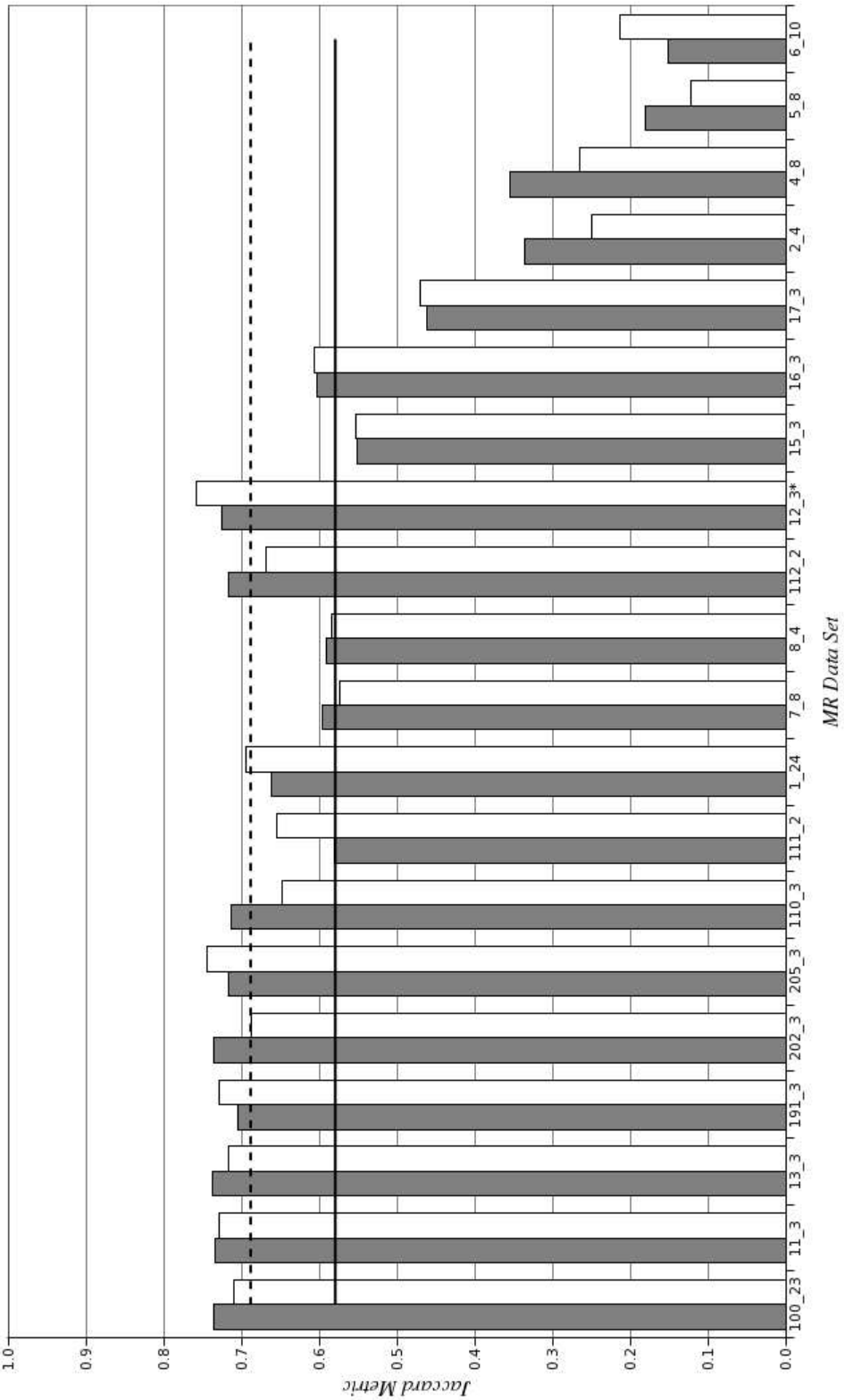
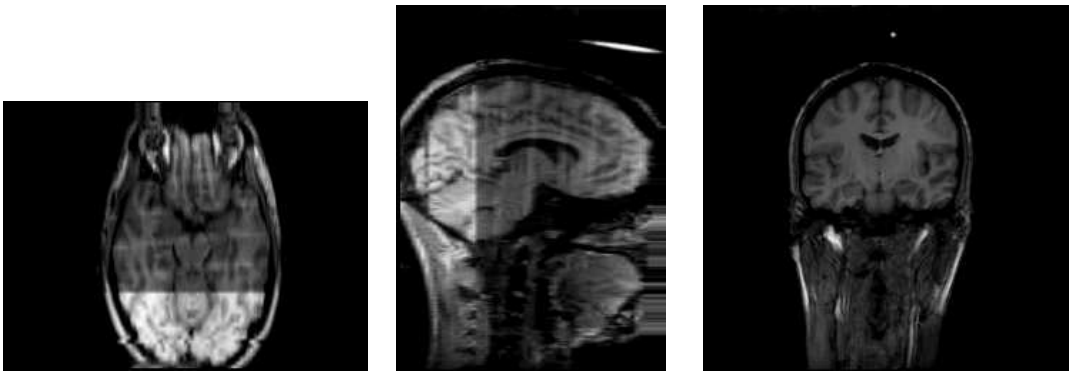


Figure 8.23: Jaccard similarity results for PV classifier GSR (model G) on GM (in grey) and WM (in white) classification of the 20 normal MR brain data sets. The overall mean Jaccard value is given by the solid line and the dotted line corresponds to the mean of the data volumes with less severe intensity inhomogeneities.



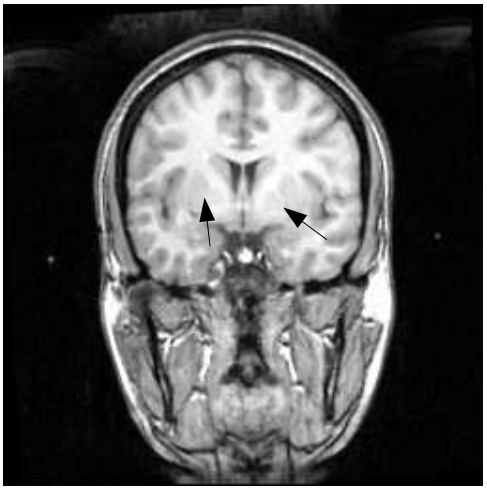
(a) Transverse mid-volume slice (b) Sagittal mid-volume slice (c) Coronal mid-volume slice

Figure 8.24: Illustration of one of the worst inhomogeneity artefacts in the 20 normal data sets [38]. This particular data set is the 6-10 data set. The artefact is also not only limited to the sudden change in intensity but also to more slowly varying inhomogeneities from slice to slice.

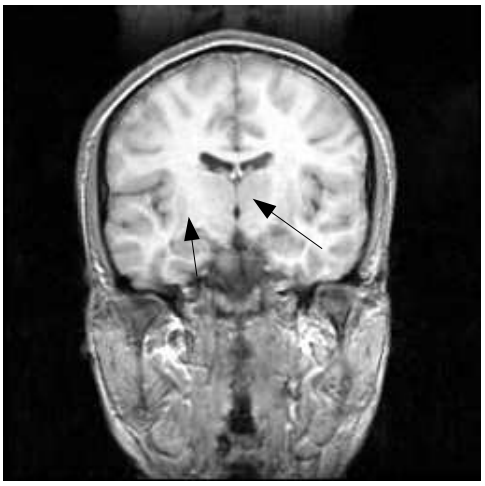
is also illustrated in figure 8.23, (dotted line). This illustrates a mean Jaccard similarity performance metric of 0.689 for GM and 0.684 for WM. This can be compared with the manual segmentation values of 0.876 for GM and 0.832 for WM. These manual segmentation values were determined by two human experts rated over segmenting the same (4) data sets and the similarity of their results was then calculated for these particular data sets. It is interesting to note that the GSR PV classifier does not model the intensity inhomogeneities usually associated with MR imaging data and even the better classified data sets still possess intensity inhomogeneities, but not as striking as those illustrated in figure 8.24. Intensity inhomogeneity correction algorithms exist and could be applied prior to classification to improve the classifier performance, see e.g. [141, 149, 129].

Peculiarities of PV Classification in MR Images of the Brain

Chapter 2, (page 34), introduced the idea that parts of the human brain were composed of regions that may not be conveniently identified as only GM *or* WM, such as the Basal Ganglia which includes the Globus Pallidus, Putamen and Caudate Nucleus, illustrated in figure 8.25. However, medically these regions are classed as GM. This is the case for the ground truth segmentations of the 20 normal brain data sets used to assess the PV classifier. This can be seen via the illustrative exemplar image slices in figure 8.21(a),(b), where despite the noise in the original images, the intensity of the Basal Ganglia GM is noticeably brighter than the GM in the surrounding Cerebral Cortex that envelopes the brain. Visual inspection of the PV classified data in figure 8.22(a),(b) illustrates that the PV classifier has mostly identified these regions as a mixture of both GM and WM. However, when the threshold of greatest PV component is applied to produce the discrete classified result, (fig. 8.22(c),(d)), the output is quite different from the discrete ground truth for the Basal Ganglia regions, (fig. 8.21(c),(d)). This, quite obviously, reduces the classification performance. This was not the case for the



(a) 100-23 data volume. Putamen (left) and Globus Pallidus (right) are annotated, (each part of the Basal Ganglia).



(b) 11-3 data volume. Globus Pallidus (left) and Caudate Nucleus (right) are annotated, (each part of the Basal Ganglia).

Figure 8.25: Annotated anatomical GM regions classed as PV voxels for the exemplar image slices previously illustrated in figure 8.21 with PV and discrete classification results in figure 8.22. These GM regions are characterised by a higher than usual myelin density for GM, (compared to Cerebral Cortex GM).

simulated PV brain data that was classified in section 8.1, where PV ground truth maps identify the Basal Ganglia as a mixture of GM and WM.

Comparison with other Classifiers

The availability of the results of other classifier and segmentation algorithms allows for a comparison of the performances of these algorithms with the results obtained for GSR. Table 8.12 summarises the algorithms to which GSR is compared. A brief discussion of these competing methodologies is now given.

Many of the methodologies were applied but not necessarily originated by Rajapakse and Kruggel in 1998 [107] to the 20 normal MR brain data sets. Rajapakse and Kruggel made their quantitative results available to the owners of the data sets so that they could be used for easy comparison, i.e. available for download on the website of [38]. Rajapkse and Kruggel's paper in 1998 was centred around two methods called AMAP and BMAP. These methods used additional classification classes to describe the PV components (as was initially proposed by Santago and Gage, 1995 in [123]) but also include a model for the intensity inhomogeneities and a MRF to improve spatial consistency. The differences between the BMAP and AMAP algorithms are that BMAP models the inhomogeneity as a multiplicative intensity where as the AMAP spatially varies the class parameters. Rajapakse and Kruggel explain that their methods would be equivalent to a MAP technique [42] without the inhomogeneity sections of their models. Similarly if AMAP/BMAP techniques did not also include the MRF then AMAP/BMAP would be equivalent to a ML technique, see e.g. [142]. They also include a comparison with

the fuzzy 'C' means clustering algorithm that includes proportional class membership parameters, where a voxel may have multiple (in different proportions) class membership to model the variable voxel content [36]. A further method which was also used to initialise all the algorithms was the tree-structure k-means algorithm (T_{sk}-means, [40]) which is a spatially aware k-means algorithm that utilizes a tree-structure, similar to other segmentation type techniques that have been designed specifically for image data, see e.g. [133].

Other notable methods include: a method by Noe and Gee in 2001, [95] that also utilizes additional classification classes to represent the PV content in combination with limiting their mixture model to only allow two tissue mixtures. Noe and Gee utilize an independent inhomogeneity correction step to reduce the effect of the intensity inhomogeneities, [129]. Ibrahim *et al.* in 2006, [60] utilize hidden Markov models (HMMs) to model the likely spatial configurations of tissues in the MRI data.

The results of applying the methods detailed above and summarized in table 8.12 can be seen in figures 8.26 and 8.27.

Table 8.12: Classifiers previously used to classify the 20 normals data sets. The legend column presents the symbols used to represent the results the plots in figures 8.26 and 8.27.

Acronym	*Refs*	*Note*	*Legend*
AMAP	[107, 38]	adaptive-MAP	$\Diamond$
BMAP	[106, 107, 38]	biased-MAP	$\triangle$
FCM	[36, 107, 38]	fuzzy 'C' means	$\bowtie$
MAP	[38, 107, 42]	*Maximum A Posterior*	$\lhd$
MLC	[38, 107]	Maximum Likelihood	$\rhd$
T_{sk}**-means**	[40, 38, 107]	tree structure means	$+$
NoeGee	[95]	mixture model clustering	$\triangledown$
Ib-cont	[60]	continuous HMM	$\square$
Ib-dist	[60]	discrete HMM	o
GSR	[here,ch.7]	Gradient Spatially Regulated	$*$

The results demonstrate that GSR performs moderately well in comparison to all the other techniques. The greatest difference can be seen for the data volumes with the most serious artefacts (to the left of the plots). Even so, GSR appears to perform closest to the HMM type classifiers for most of the data volumes. This is despite the HMM techniques (Ib-cont and Ib-dist) utilizing supervised training details of which were not given. Similarly, most of the techniques (including AMAP [107], BMAP [106], NoeGee [95], Ib-disc and Ib-cont [60]) all utilise pre-processing or incorporate explicit models of the intensity inhomogeneities prevalent in these test data sets.

These observations together with the results obtained for the simulated brain data indicate that GSR provides a suitable framework for further development. Further development that should improve the discrete classification performance on these 20 normal brain data sets might include extension of the model to include intensity inhomogeneities. A further important development might include prior knowledge of the peculiarities of the PV effect when considering regions such as the Basal Ganglia.

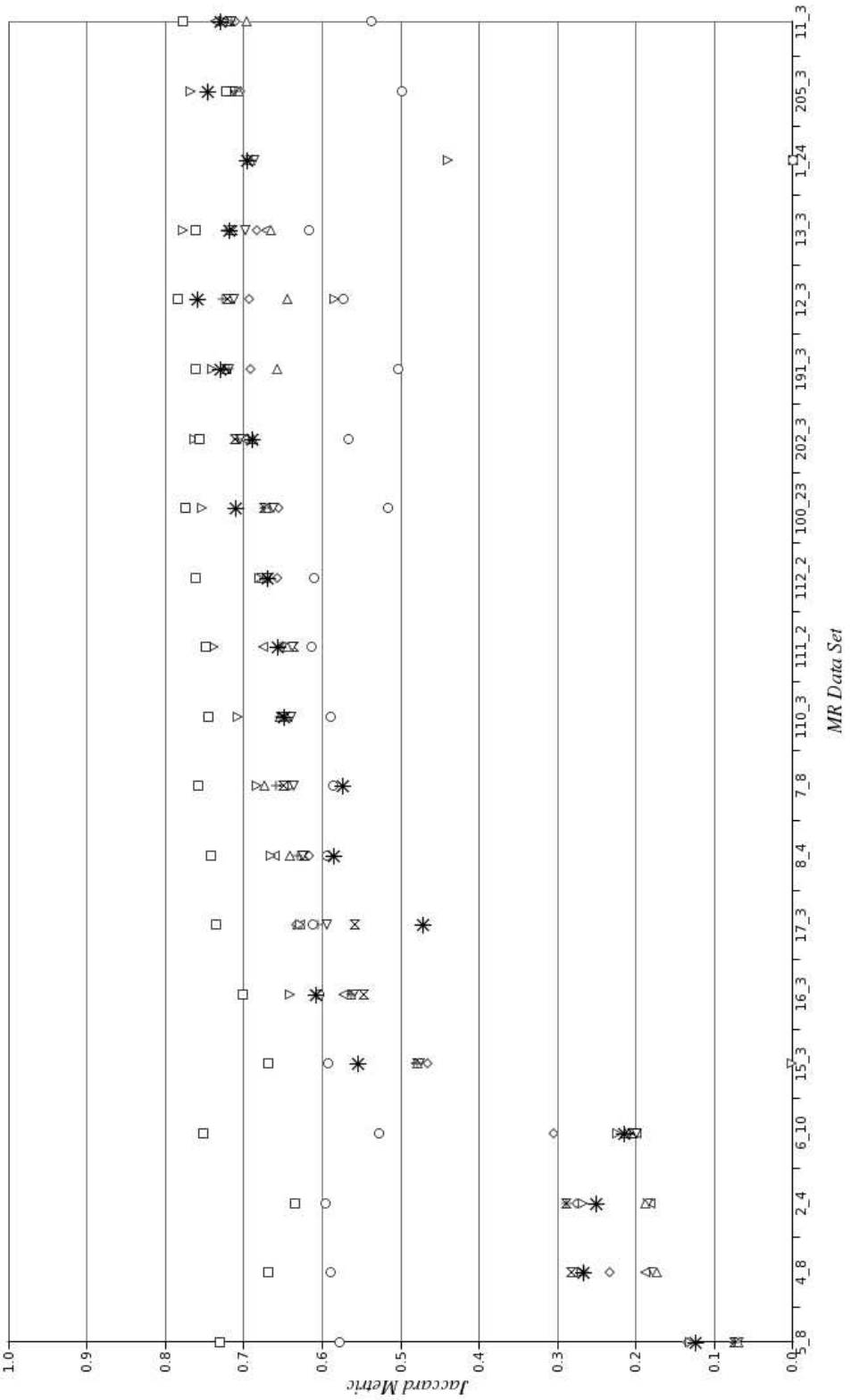


Figure 8.26: Jaccard WM similarity results for the 20 normals data sets using the classifiers summarised in table 8.12 (including the symbols). Classifier using PV model G, GSR, is shown by the line.

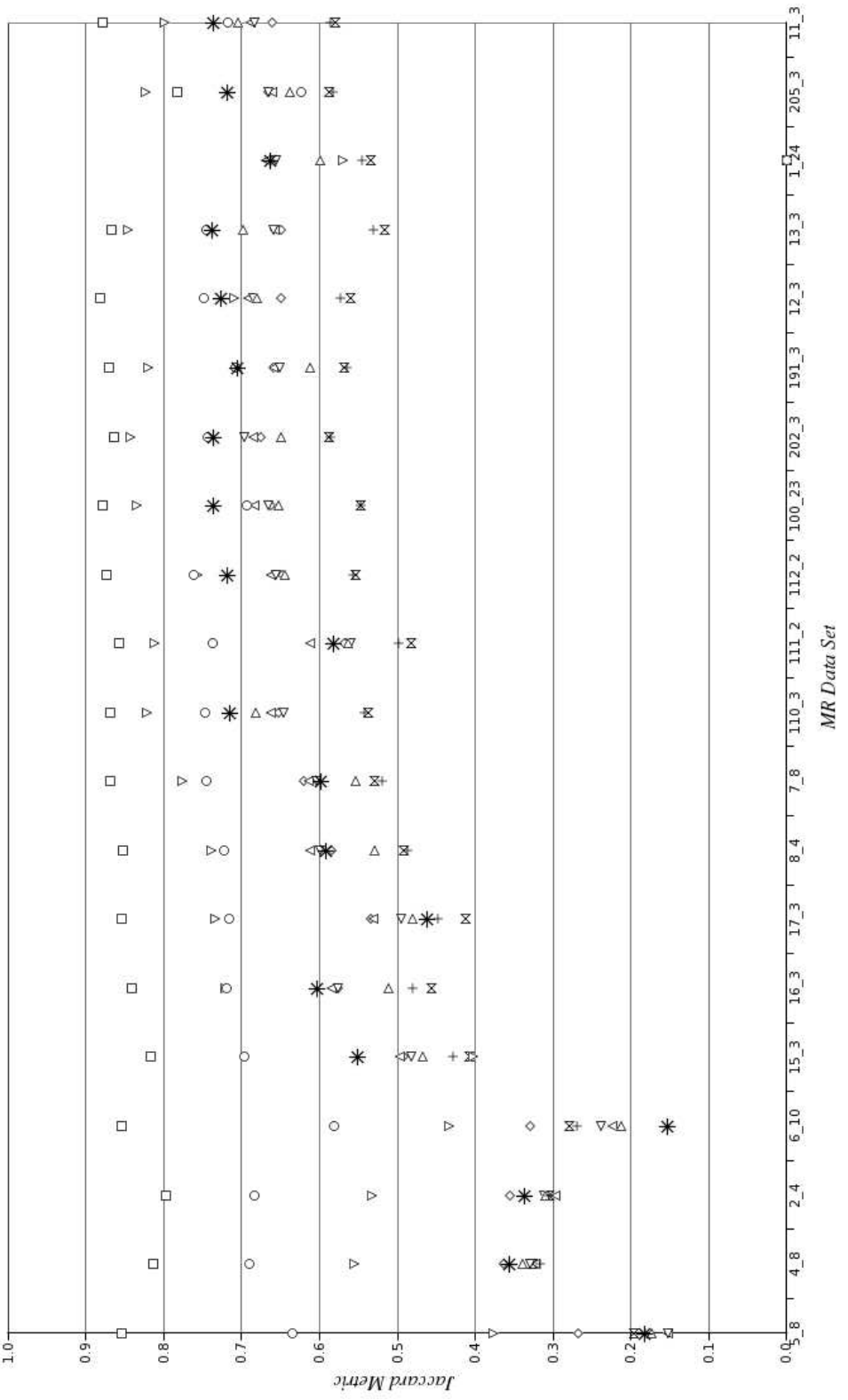


Figure 8.27: Jaccard GM similarity results for the 20 normals data sets using the classifiers summarised in table 8.12 (including the symbols). Classifier using PV model G, GSR, is shown by the line.

8.4 In Conclusion

The PV model developments of the global prior based intensity classification schemes in chapter 5, the global prior based intensity and gradient magnitude models of chapter 6 and the gradient magnitude spatially regulated intensity model of chapter 7 have now been compared on various types of PET and MRI data. The global prior based intensity models of chapter 5 appear to possess relatively good performance in comparison to other existing techniques when assessed using the simulated MR brain data from McGill University [12]. But overall the gradient magnitude spatially regulated intensity model appeared to perform the best on this intensity inhomogeneity free simulated data.

Application of the classifiers to the PET image phantom with hot and cold inserts illustrated the benefit of the gradient magnitude based global prior classifier (model F) of chapter 6. The relatively poor performance of GSR was probably due to the spatial regularisation at high CNR values providing relatively little useful additional information. This was found to be in agreement with the results obtained for the simple simulated PV data used to assess the development of the classifiers in the earlier chapters.

The final set of performance assessment was undertaken with the aid of 20 normal MR brain data sets provided by [38]. The ground truth of these real MRI data sets was provided by human experts consisting of discrete voxel allocations to individual tissues. The gradient magnitude spatially regulated PV classifier (GSR, model G) was applied to this data and the output of the classifier was discretized. The performance of GSR was then compared with other published results on these test data sets and found to have comparable and in some cases superior performance to many of the existing PV classifiers exhibiting high performance. This was made more remarkable as the classification process with GSR did not model or pre-process the image data to remove prevalent MR artefacts that manifest as severe intensity inhomogeneities particularly on some of the data sets. GSR also does not utilize supervised training, *ad hoc* rules or arbitrarily selected parameter values due to its probabilistic mathematically consistent formulation. The highly competitive performance, and the rigorous underlying theory therefore make it an attractive proposition for brain classification.

Chapter 9

Conclusions and Further Work

9.1 Summary of Results and Conclusions

The work in this thesis was primarily motivated by the need to improve models of the PV effect and, as a result, improved classification performance of PV affected images.

Chapters 1 and 2 introduced the concept of a medical image. These chapters also introduced the fact that medical images are imperfect pictorial representations of human anatomy or physiology. These chapters also defined the nature of one such artefact known as the PV effect. Exemplars from MRI and PET modalities were selected because these modalities represent the state of the art in imaging technology due to the excellent soft tissue delineating properties of MRI and the excellent sensitivity of PET to physiological activity.

The PV effect in neurological MRI and PET imaging data is of particular importance due to the need to accurately identify brain tissue quantities or activity concentrations for global (entire brain) and local (clusters of voxels) volumes. Pre-processing of neurological data is often required to remove extraneous non-brain voxels so that classification algorithms do not have to incorporate numerous additional non-CNS classification classes. Therefore chapter 3 introduced an improved method for skull-stripping neurological MRI data. This was also adapted to include CSF voxels in the final segmentation. This unique method automatically identifies transverse image slices in which the majority of the NMR voxel signal could be identified as corresponding to GM, WM and CSF. The technique was then assessed on adult and for the first time on infant neurological MR data, and found to be comparable and in some cases superior to the well known BET, [131].

A review of the existing PV literature was given in chapter 4, introducing the current state of the art.

Chapter 5 described in detail two existing intensity based PV models (A - a finite Gaussian mixture model and B, a PV finite mixture model, [148]) that use the posterior probabilities obtained from Bayes theorem as analogies to actual voxel content. A third model (C) was then defined that explicitly incorporated a voxel content (mixture) random vector. This novel third model was in fact found to be approximately equivalent

to the second analogous model via appropriate analysis of the underlying formulations. Simple two class simulated PV data sets were then generated with various CNR values. PV models B and C were found to possess similar PV classification performance, over and above model A, which did not include any attempt to correctly model the PV effect. The intrinsic performance of the global PV prior intensity based classifiers (models B and C) for PV RMS voxel error was found to be equal to the reciprocal of the CNR. It was therefore decided that further information should be included into the PV models in order to improve classification performance.

Therefore chapter 6 looked at PV models that incorporated locally calculated image gradient magnitude measures together with the raw image intensities. The first model (D) was based on a technique found in the literature [157] and utilized a parametric model for the 2-D gradient magnitude. This model built on the analogous concepts introduced by the intensity only models (A and B) of chapter 5. The novel intensity formulation of model C, where the PV mixing was modelled explicitly rather by analogy was then combined with a novel intensity and gradient magnitude likelihood (model E). This resulted in an alternative intensity and gradient magnitude PV feature space that was smoother in appearance. A further development was also provided that extended the parametric 2-D gradient magnitude description of the feature space to a 3-D formulation. This provided improved PV classifier performance when assessed with simulated PV data, reducing the PV voxel RMS error by up to 1.5% for CNR values between 4 and 10. This model therefore provided decreased classification error over all the preceding models, such as the intensity only based classifiers (models B and C), reducing the PV voxel RMS errors by up to 5% to 7% for simulated PV data with CNR values between 4 and 10. Chapter 6 helped to illustrate that improvements to the PV classifier performance were possible by improving and extending the PV models.The improvements described in chapter 6 still did not improve classifier performance sufficiently for the classification of low-CNR PV data (< 10 with voxel RMS classifier errors of over 25% for CNR values below 5).

The developments of chapters 5 and 6 did not explicitly incorporate spatial contextual information that many other PV models in the literature provide. It was expected that spatial information would be particularly useful for low CNR imaging data. Especially as medical image data quite often possess moderately low CNR values and the PV classifiers of chapters 5 and 6 possessed relatively high errors for low CNR values. Therefore chapter 7 reformulated the PV models to include spatial contextual information. Initially global prior mixing densities were proposed. The first was derived analytically and the second was derived from observation using a phenomenological law known as Benford's law. These two prior distributions were found to possess better χ^2 fits to simulated PV prior distributions in comparison to a mixing prior distribution proposed elsewhere.

Manipulation of the appropriate PDFs was then undertaken to achieve a realisable formulation that included spatial information, intensity and gradient magnitude random variables. It was found that the gradient magnitude PDF introduced in chapter 6 was cancelled out of the formulation due to the additional, newly included, spatial prior. However, a gradient magnitude random variable was still included in the form of a dependent term of the spatial prior. This spatial prior was then formulated so as to take

advantage of the gradient magnitude to control the amount of spatial regularisation, i.e. how much information should be drawn from the neighbouring voxels for particular values of gradient magnitude. This new model formulation was then implemented via simulation using MCMC. This was due to the fact that no closed form solutions of the integrals in the model could be found. The sampling density was conveniently based on the Dirichlet PDF due to its similarity to the analytically derived prior distribution that was determined in the first part of chapter 7. This sampling and proposal Dirichlet distribution also fully implements the PV random variable constraints, i.e. no voxel should have more than 100% of a single tissue or activity concentration.

Assessment with the simple two class PV simulated data found this gradient spatially regulated (GSR-model G) formulation to possess superior performance for low CNR value data.The principal improvement was due to improved classification performance of low gradient voxels which for this simple simulated PV data consisted of only pure voxels. Equally, contiguous regions of PV voxels found in some PV imaging data, such as the Basal Ganglia regions in MR neurological images would also benefit from the contextual classification process offered by this classifier.

The PV models of chapter 5, 6 and 7 were assessed by determining their performance on classifying simple two class simulated PV data consisting of concentric spheroids. Therefore the final chapter, 8 applied exemplar PV models B (the Gaussian-Triangle PV intensity based classifier), F (the best performing 3-D gradient magnitude and intensity based classifier) and G (the gradient control spatially regulated classifier-GSR) to more rigorous assessment. This included simulated MR brain data from McGill University, Canada; a PET phantom with ground truth derived from high resolution structural CT data from Royal Marsden Hospital, UK; and 20 normal MR brain data sets from Massachusetts General Hospital, USA. Each of these data sets had ground truth to which the output of the classifiers could be compared to provide objective quantitative measures of their performance.

The results of these analyses illustrated that some of the broad conclusions drawn from the experiments performed in the earlier chapters were equally demonstrable on these more realistic data sets. Most notably that GSR provided improved classification results for smaller CNR value data, whilst not utilizing any supervised training, *ad hoc* rules or arbitrarily defined parameter values. Unfortunately model F, utilising a combined gradient magnitude and intensity likelihood, did not provide improved classification performance (over model B) on the simulated brain data, in contrast to the improved performance demonstrated on the simpler two class simulated PV data. This was most likely due to the different types of PV effect that model F was designed to detect: the human brain consists of different types of PV effect, not only PV voxels arising from simple boundary configurations but also larger volumetric expanses of PV voxels consisting of both GM and WM, as illustrated in chapter 2.

9.2 Future Work

Most of the developments in PV modelling described in this thesis do not have analytical solutions and therefore rely on numerical integration or simulation (MCMC)

techniques to determine not only improved estimates of model parameter values but also to determine individual voxel constituents. This is in contrast to the simpler analogous models (A,B and D) which provided closed form solutions to the appropriate calculation of voxel constituents. The reliance on computationally expensive simulation techniques most likely reduces the appeal to a wider audience than one to which accurate estimates of PV content are the primary interest. Therefore future work could include attempts to provide analytic (possibly approximate) solutions to the developments proposed in this thesis. However, simulation techniques such as MCMC are, say computational scientists, convenient and simple to code. These are becoming ever more popular due to their power in realising multi-dimensional Bayesian type problems. This popularity and the rise again in popularity of the parallel computing architectures (*circa* 2006) that enable parallel implementation of these computationally intensive simulation algorithms may not diminish any potential advances in a computational science that requires such simulations.

The application of the models to real imaging data also illustrated that the PV effect is not the only imaging artefact that should be included in a comprehensive model of the imaging data. In particular, the application to the 20 normal brain data sets [38] demonstrated the performance limiting intensity inhomogeneities that are prevalent in the MR image acquisition process. These intensity inhomogeneities should be included in any imaging model to improve PV classification performance of MR data. The Bayesian approach is an ideal framework for such a task and direct extension of the developments in this thesis would be possible. These developments could include hyper priors (further hidden random variables) that model the variability or non-stationarity of the class parameters. These developments may be equally valid for other tomographic imaging modalities, such as the scatter imaging artefact in PET and SPECT (see chapter 2), although careful development of modality specific prior information would probably have to be taken.

Further *a priori* information could also be included to improve segmentation performance of the brain. As noted in chapter 2, the brain consists of large expanses of GM-WM PV voxels. Some of these regions are known anatomically as the Basal Ganglia of the human brain and are usually classed medically as GM. This is despite the Basal Ganglia regions possessing greater myelin density in comparison to other GM regions such as the surface of the cerebral cortex. This *a priori* information could be incorporated into the Bayesian PV model or perhaps as part of the discretization process using a HMM formulation.

It is thought that these two improvements alone would produce improved discrete classification performance that would be competitive with the supervised training utilized by other existing techniques [60].

Other open questions that this thesis identified included the validity of the assumption of linear mixing, especially when considering quantitative PV estimation of MR imaging data. The MR image acquisition process and post processing are potentially non-linear for some image acquisition sequences and tissue combinations. Therefore any algorithm that attempted to estimate the quantity of a particular tissue, in a particular voxel utilizing the linearity assumption (used here), may introduce systematic errors into the estimation process. Whether these errors would be greater than the lower bound error

of the classification process due to the intrinsic noise properties of the data is another question that would have to be answered in order to warrant further investigation.

The skull-stripping technique that was developed in chapter 3 provided good brain isolation performance in some cases, over another existing skull stripping technique known as BET. This complementary performance could be utilized by combining the output of the two algorithms to obtain the best performance of each. A supervised training approach could be used but this would require access to a large number of MR brain data sets. Alternative techniques could possibly utilize self-diagnostic criteria that would attempt to quantify the quality of the result based on some objective measures.

Appendix A

Abbreviations and Acronyms

- **2-D** two dimensional
- **3-D** three dimensional
- **AC** attenuation correction
- **a.k.a.** also known as
- **BET** brain extraction tool
- **Bkgrnd** background
- **CDF** cumulative distribution function
- **ch.** chapter
- **cm** centimetre
- **CNS** central nervous system
- **CSF** cerebrospinal fluid
- **CT** computed tomography
- **ECT** emission computed tomography
- **EM** expectation maximization
- **Err** error
- **FBP** filtered back projection
- **FGMM** finite gaussian mixture model
- **FID** free induction decay
- **FIR** finite impulse response
- **FM** finite mixture

- **FMM** finite mixture model
- **fMRI** functional magnetic resonance imaging
- **FOV** field of view
- **FWHM** full width at half maximum
- **GM** grey matter or gradient magnitude
- **GSR** gradient spatial regularisation
- **HDR** high density region
- **HMM** hidden markov model
- **HMRF** hidden markov random field
- **Hv** high variance
- **Hz** high gradient
- **ICG** inverse cumulative gaussian
- **ICM** iterated conditional modes
- **iid** independent and identically distributed
- **IMH** independent metropolis hastings
- **l.h.s.** left hand side
- **Lv** low variance
- **Lz** low gradient
- **MAP** maximum a posteriori
- **MC** markov chain
- **MCMC** markov chain monte carlo
- **ML** maximum likelihood
- **mm** millimetre
- **MR** magnetic resonance
- **MRF** markov random field
- **MRI** magnetic resonance imaging
- **NEMA** national electric manufacturers association
- **PACS** picture archival and communication system
- **PDF** probability density function

- **PET** positron emission tomography
- **PMF** probability mass function
- **PSF** point spread function
- **PTFE** polytetrafluoroethene
- **PV** partial volume
- **PVE** partial volume effect
- **RAMLA** iterative reconstruction technique
- **RAMLActac** RAMLA computed tomography attenuation correction
- **RAMLAfc** RAMLA fully corrected
- **RAMLAnc** RAMLA no correction
- **refs** references
- **RF** radiofrequency or random field
- **r.h.s.** right hand side
- **RMS** root mean square
- **ROI** region of interest
- **SMSS** statistical morphological skull stripper
- **SNR** signal to noise ratio
- **SPECT** single photon emission computed tomography
- **TRI** triangle
- **Vxl** voxel
- **WM** white matter
- **w.r.t.** with respect to

Appendix B

Gauss-Triangle Convolution Evaluation

[1] The Gaussian Triangle Convolution used as the Partial Volume (PV) distribution in the intensity PV model is derived in this appendix. A number of authors have suggested using a PV distribution to model the voxel intensities that are generated by a mixture of two or more pure classification classes. The form of this distribution is usually suggested to be of the form of a uniform distribution between the two pure class mean values, but also convolved with the pure class Probability Density Functions (PDFs) usually considered to be Gaussian distributed. Vokurka et al suggested an alternative form that consisted of two triangle distributions convolved with the pure class distributions so that mixtures arising from a single class could be considered independently [148].

B.1 Triangle Distribution

The equation for a triangle is the same as that for a straight line, except with constraints of a specific slope and two limits:

$$f_{triangle}(g|M,C) = \begin{cases} M.g + C & \text{for} \quad \mu_1 \leq g \leq \mu_2 \\ 0 & \text{for} \quad \text{elsewhere} \end{cases} \tag{B.1}$$

where M and C are the slope and intercept respectively. μ_1 and μ_2 are the limits of the triangle's extent. Suitable values for M and C can be derived using trigonometrical principles and the fact that when the triangular PDF is scaled, then any slope can be achieved. Therefore $M = +1$ or -1, depending on whether the PV distribution belongs to pure distributions with mean μ_1 ($M = -1$) or pure distribution with mean μ_2 ($M = +1$). A second observation can lead to values for C. The triangle can now be considered to be an equilateral triangle (i.e. a right angle triangle with slope $+/-1$). If $M = -1$ then from trigonometry the negative slope will result in a positive intercept

[1] A great deal of this derivation is due, in part, to Barbara Podda of the University of Cagliari, Sardinia who performed a version of the derivation when she was on placement at the University of Surrey in 2003.

so that $C = +\mu_2$. If $M = +1$ then from trigonometry the positive slope will result in a negative intercept, so that $C = -\mu_1$.

Hence, there are two forms of the triangle distribution, but for the purposes of the convolution (to be derived shortly), M and C will be retained in the equation so that only a single derivation has to be performed.

PDFs should be normalized, i.e.:

$$\int_{-\infty}^{+\infty} p_{triangle}(g).dg = 1 \tag{B.2}$$

This constraint can be determined by performing the integration:

$$\int_{\mu_1}^{\mu_2} \frac{1}{Q} \left(M.g + C\right).dg = 1 \tag{B.3}$$

As Q does not depend on g:

$$\int_{\mu_1}^{\mu_2} \left(M.g + C\right).dg = Q, \tag{B.4}$$

which results in:

$$\frac{M}{2}.[g^2]_{\mu_1}^{\mu_2} + C.[g]_{\mu_1}^{\mu_2} = \frac{M}{2}.(\mu_2^2 - \mu_1^2) + C.(\mu_2 - \mu_1) = Q. \tag{B.5}$$

where $\mu_2 > \mu_1$.

We therefore have a specific form for the PDF of the triangle distribution:

$$p_T(g|\tau_{a,b}) = \begin{cases} \frac{1}{Q}(M.g + C) & \text{for} \quad \mu_a \leq g \leq \mu_b \\ 0 & \text{for} \quad \text{elsewhere} \end{cases} \tag{B.6}$$

where $M = -1$, $C = \mu_b$ or

$$p_T(g|\tau_{b,a}) = \begin{cases} \frac{1}{Q}(M.g + C) & \text{for} \quad \mu_a \leq g \leq \mu_b \\ 0 & \text{for} \quad \text{elsewhere} \end{cases} \tag{B.7}$$

where $M = 1$, $C = -\mu_a$ and where $\tau_{a,b}$ are order dependent indicators for classification classes τ_a and τ_b.

B.2 Gaussian Distribution

The Gaussian distribution takes the form:

$$p_G(g|\tau_a) = \frac{1}{\sqrt{2.\pi.\sigma_a^2}}.\exp\left(-\frac{g^2}{2.\sigma_a^2}\right), \tag{B.8}$$

where σ_a is the standard deviation of the Gaussian distribution for class a.

B.3 The Convolution

The PV distribution for classes a and b is given by the convolution of the Gaussian distribution with the triangular distribution:

$$\begin{aligned} p_{GT}(g|\tau_{a,b}) &= p_G(g|\tau_a) * p_T(g|\tau_{a,b}) \\ &= \int_{-\infty}^{+\infty} p_G(g-\gamma|\tau_a).p_T(\gamma|\tau_{a,b}).d\gamma. \end{aligned} \tag{B.9}$$

Remembering that $p_T(g|\tau_{a,b}) = 0$ for $g > \tau_b$ and $g < \tau_a$:

$$\begin{aligned} p_{GT}(g|\tau_{a,b}) &= \int_{\mu_a}^{\mu_b} p_G(g-\gamma|\tau_a).p_T(\gamma|\tau_{a,b}).d\gamma \\ &= \frac{1}{Q.\sqrt{2.\pi.\sigma_a^2}}.\int_{\mu_a}^{\mu_b} \exp\left(-\frac{(g-\gamma)^2}{2.\sigma_a^2}\right).(M.\gamma + C).d\gamma \\ &= K_1.\int_{\mu_a}^{\mu_b} \exp\left(-\frac{(g-\gamma)^2}{2.\sigma_a^2}\right).(M.\gamma).d\gamma + \\ &K_1.\int_{\mu_a}^{\mu_b} \exp\left(-\frac{(g-\gamma)^2}{2.\sigma_a^2}\right).(C).d\gamma, \end{aligned} \tag{B.10}$$

where $K_1 = \frac{1}{Q.\sqrt{2.\pi.\sigma_a^2}}$. These two integrals can then be evaluated independently. Setting:

$$K_2(g) = M.\int_{\mu_a}^{\mu_b} \gamma.\exp\left(-\frac{(g-\gamma)^2}{2.\sigma_a^2}\right).d\gamma, \tag{B.11}$$

and

$$K_3(g) = C.\int_{\mu_a}^{\mu_b} \exp\left(-\frac{(g-\gamma)^2}{2.\sigma_a^2}\right).d\gamma. \tag{B.12}$$

Considering $K_3(g)$ first, if we change the variable being integrated to $\nu = \frac{g-\gamma}{\sqrt{2.\sigma_a^2}}$ so that $\frac{d\nu}{d\gamma} = -\frac{1}{\sqrt{2.\sigma_a^2}}$, $d\gamma = -d\nu.\sqrt{2.\sigma_a^2}$ and the limits of integration change to $\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}$ and $\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}$:

$$K_3(g) = -C.\sqrt{2.\sigma_a^2}.\int_{\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}}^{\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}} \exp\left(-\nu^2\right).d\nu. \tag{B.13}$$

Using the definition of the error function:

$$\mathrm{erf}(\mathbf{z}) = \frac{2}{\sqrt{\pi}} \int_0^{\mathbf{z}} \exp(-\mathbf{t}^2).d\mathbf{t}, \tag{B.14}$$

$K_3(g)$ becomes

$$K_3(g) = -C.\frac{\sqrt{\pi}}{2}.\sqrt{2.\sigma_a^2}.\left\{\mathrm{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \mathrm{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}. \tag{B.15}$$

Now to consider $K_2(g)$. Again letting $\nu = \frac{g-\gamma}{\sqrt{2.\sigma_a^2}}$ and therefore $\gamma = g - \nu.\sqrt{2.\sigma_a^2}$.

$$\begin{aligned}
K_2(g) &= M. \int_{\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}}^{\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}} (g - \nu.\sqrt{2.\sigma_a^2}).\exp\left(-\nu^2\right).(-d\nu.\sqrt{2.\sigma_a^2}) \\
&= -\sqrt{2.\sigma_a^2}.M. \int_{\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}}^{\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}} (g - \nu.\sqrt{2.\sigma_a^2}).\exp\left(-\nu^2\right).d\nu \\
&= -\sqrt{2.\sigma_a^2}.M.g. \int_{\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}}^{\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}} \exp\left(-\nu^2\right).d\nu \\
&\quad -\sqrt{2.\sigma_a^2}.M.(-\sqrt{2.\sigma_a^2}). \int_{\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}}^{\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}} \nu.\exp\left(-\nu^2\right).d\nu.
\end{aligned} \tag{B.16}$$

The first part of the equation can be deduced almost directly from the result obtained for $K_3(g)$, so that:

$$\begin{aligned}
&= -\sqrt{2.\sigma_a^2}.M.g.\frac{\sqrt{\pi}}{2}.\left\{\mathrm{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \mathrm{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\} \\
&\quad -\sqrt{2.\sigma_a^2}.M.(-\sqrt{2.\sigma_a^2}). \int_{\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}}^{\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}} \nu.\exp\left(-\nu^2\right).d\nu.
\end{aligned} \tag{B.17}$$

To consider the final integral, let:

$$K_4(g) = \int_{\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}}^{\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}} \nu.\exp\left(-\nu^2\right).d\nu \tag{B.18}$$

To evaluate, let $\iota = -\nu^2$, so that $d\nu = -\frac{d\iota}{2.\nu}$:

$$K_4(g) = \int_{-\frac{(g-\mu_a)^2}{2.\sigma_a^2}}^{-\frac{(g-\mu_b)^2}{2.\sigma_a^2}} \nu.\exp(\iota).\frac{d\iota}{-2.\nu} = -\frac{1}{2}\int_{-\frac{(g-\mu_a)^2}{2.\sigma_a^2}}^{-\frac{(g-\mu_b)^2}{2.\sigma_a^2}} \exp(\iota).d\iota. \tag{B.19}$$

The integral can then be evaluated resulting in:

$$K_4(g) = (-1).\frac{1}{2}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\} \tag{B.20}$$

Substituting $K_4(g)$ into $K_2(g)$:

$$\begin{aligned}
K_2(g) &= -\sqrt{2.\sigma_a^2}.M. \\
&\quad \left\{g.\frac{\sqrt{\pi}}{2}.\left\{\operatorname{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \operatorname{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}\right. \\
&\quad \left. + (-\sqrt{2.\sigma_a^2}).(-1).\frac{1}{2}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\}\right\} \\
&= -\sqrt{2.\sigma_a^2}.M. \\
&\quad \left\{g.\frac{\sqrt{\pi}}{2}.\left\{\operatorname{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \operatorname{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}\right. \\
&\quad \left. + \frac{\sqrt{\sigma_a^2}}{\sqrt{2}}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\}\right\}.
\end{aligned} \tag{B.21}$$

Then substituting the expressions for $K_2(g)$ and $K_3(g)$ into the Gaussian Traingle PDF:

$$\begin{aligned}
p_{GT}(g|\tau_{a,b}) &= K_1.\{K_2(g) + K_3(g)\} \\
&= K_1.\left\{-C.\frac{\sqrt{\pi}}{2}.\sqrt{2.\sigma_a^2}.\left\{\operatorname{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \operatorname{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\} + - \sqrt{2.\sigma_a^2}.M.\right. \\
&\quad \left\{g.\frac{\sqrt{\pi}}{2}.\left\{\operatorname{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \operatorname{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}\right. \\
&\quad \left.\left. + \frac{\sqrt{\sigma_a^2}}{\sqrt{2}}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\}\right\}\right.
\end{aligned}$$

$$= -K_1.\sqrt{2.\sigma_a^2}\left\{C.\frac{\sqrt{\pi}}{2}.\left\{\text{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \text{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\} + M.\right.$$
$$\left\{g.\frac{\sqrt{\pi}}{2}.\left\{\text{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \text{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}\right.$$
$$\left.\left. + \frac{\sqrt{\sigma_a^2}}{\sqrt{2}}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\}\right\}\right\}$$

$$= -K_1.\sqrt{2.\sigma_a^2}\left\{(M.g+C).\frac{\sqrt{\pi}}{2}.\left\{\text{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \text{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}\right.$$
$$\left. + M.\frac{\sqrt{\sigma_a^2}}{\sqrt{2}}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\}\right\}$$

$$= -\frac{1}{Q.\sqrt{2.\pi.\sigma_a^2}}.\sqrt{2.\sigma_a^2}\left\{(M.g+C).\frac{\sqrt{\pi}}{2}.\left\{\text{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \text{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}\right.$$
$$\left. + M.\frac{\sqrt{\sigma_a^2}}{\sqrt{2}}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\}\right\}$$

$$= -\frac{1}{Q.\sqrt{\pi}}.\left\{(M.g+C).\frac{\sqrt{\pi}}{2}.\left\{\text{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \text{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}\right.$$
$$\left. + M.\frac{\sqrt{\sigma_a^2}}{\sqrt{2}}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\}\right\}$$

Therefore the PDF for a PV distribution for a particular class, τ_a, sharing intensity values with τ_b is given by:

$$p_{GT}(g|\tau_{a,b}) = \frac{(M.g+C)}{-2.Q}\left\{\text{erf}\left(\frac{g-\mu_b}{\sqrt{2.\sigma_a^2}}\right) - \text{erf}\left(\frac{g-\mu_a}{\sqrt{2.\sigma_a^2}}\right)\right\}$$
$$+\frac{-M.\sigma_a}{Q.\sqrt{2.\pi}}\left\{\exp\left(-\frac{(g-\mu_b)^2}{2.\sigma_a^2}\right) - \exp\left(-\frac{(g-\mu_a)^2}{2.\sigma_a^2}\right)\right\} \tag{B.22}$$

Note This is now a PDF. A pair of these PV PDFs, defined by $p_{GT}(g|\tau_{a,b})$ and $p_{GT}(g|\tau_{b,a})$ should have equal prior values within a mixture model framework.

Appendix C

Sum of Gaussian Distributed Random Variables

This appendix demonstrates that the result of the sum of Q Gaussian distributed random variables is also governed by a Gaussian PDF. It also provides a means to calculate the parameters of the resulting Gaussian distributed random variable.

Consider Q random variables, X_q, that are normally distributed, i.e. $X_q : \mathcal{N}(\mu_q, \sigma_q^2)$ and the weighted sum of of these random variables is given by:

$$s = \sum_{q=1}^{Q} w_q.x_q, \tag{C.1}$$

where w_q is the weight of the q^{th} random variable instance, x_q which is an instantiation of the random variable, X_q, i.e. $X_q(\boldsymbol{\omega}) = x_q$, but to simplify presentation $\boldsymbol{\omega}$ will not be used.

The characteristic function[1] of X_q is given by [134]:

$$\boldsymbol{\Phi}_{X_q}(\psi) = E[e^{j.\psi.X_q}] = \int_{-\infty}^{+\infty} p_{X_q}(x_q).\exp(j.\psi.x_q).dx_q, \tag{C.2}$$

[1] A useful tool in the manipulation of functions of random variables is the characteristic function (see for example [96]). The result of a Fourier transform applied to a PDF, $p_{Z_a}(z_a)$, of an arbitrary random variable, Z_a, is known as the characteristic function, $\Phi_{Z_a}(\psi_a)$ of the random variable Z_a.

$$\Phi_{Z_a}(\psi_a) = \int_{-\infty}^{\infty} p_{Z_a}(z_a).\exp(j.\psi_a.z_a).dz_a,$$

where $j = \sqrt{-1}$. The characteristic function is often given as:

$$\Phi_{Z_a}(\psi_a) = \mathbf{E}\left[\exp(j.\psi_a.z_a)\right];$$

and the inverse Fourier transform of the characteristic function is given by:

$$p_{Z_a}(z_a) = \frac{1}{2.\pi} \int_{-\infty}^{\infty} \Phi_{Z_a}(\psi_a).\exp(-j.\psi_a.z_a).d\psi_a.$$

where $p_{X_q}(x_q)$ is the PDF of X_q. Scaling of a random variable can also be performed via a small modification to the above formula [146]:

$$\mathbf{\Phi}_{w_q.X_q}(\psi) = E[e^{j.\psi.w_q.X_q}] = \int_{-\infty}^{+\infty} p_{X_q}(x_q).\exp(j.\psi.x_q.w_q).dx_q, \tag{C.3}$$

It is well known that the summation of Q independently and identically distributed (i.i.d.) random variables can be performed via the product of their characteristic functions [134]. Hence, if $S = s$:

$$\mathbf{\Phi}_S(\psi) = \prod_{q=1}^{Q} \mathbf{\Phi}_{w_q.X_q}(\psi), \tag{C.4}$$

so that:

$$p_S(s) = \frac{1}{2.\pi} \int_{-\infty}^{+\infty} \mathbf{\Phi}_S(\psi).\exp(-j.\psi.s).d\psi, \tag{C.5}$$

where $p_S(s)$ is the PDF of the result of the weighted sum of Q Gaussian distributed random variables.

The individual characteristic functions of each scaled random variable are first evaluated:

$$\mathbf{\Phi}_{w_q.X_q}(\psi) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2.\pi.\sigma_q^2}}.\exp\left(-\frac{(g-\mu_q)^2}{2.\sigma_q^2}\right).exp\left(j.\psi.g.w_q\right).dg, \tag{C.6}$$

which results in:

$$\begin{aligned}
\mathbf{\Phi}_{w_q.X_q}(\psi) &= \frac{1}{\sqrt{2.\pi.\sigma_q^2}} \int_{-\infty}^{+\infty}.\exp\left(-\frac{(g-\mu_q)^2 - 2.g.j.w_q.\psi.\sigma_q^2}{2.\sigma_q^2}\right).dg, \\
&= \frac{1}{\sqrt{2.\pi.\sigma_q^2}} \int_{-\infty}^{+\infty}.\exp\left(-\frac{g^2 - 2.g.\mu_q + \mu_q^2 - 2.g.j.w_q.\psi.\sigma_q^2}{2.\sigma_q^2}\right).dg, \\
&= \frac{1}{\sqrt{2.\pi.\sigma_q^2}} \int_{-\infty}^{+\infty}.\exp\left(-\frac{g^2 - 2.g(\mu_q + j.w_q.\psi.\sigma_q^2) + \mu_q^2}{2.\sigma_q^2}\right).dg.
\end{aligned} \tag{C.7}$$

Completing the square:

$$
\begin{aligned}
&= \frac{1}{\sqrt{2.\pi.\sigma_q^2}}. \\
&\int_{-\infty}^{+\infty} \exp\left(-\frac{(g-(\mu_q+j.w_q.\psi.\sigma_q^2))^2-(\mu_q+j.w_q.\psi.\sigma_q^2)^2+\mu_q^2}{2.\sigma_q^2}\right).dg, \\
&= \frac{1}{\sqrt{2.\pi.\sigma_q^2}}.\exp\left(\frac{(\mu_q+j.w_q.\psi.\sigma_q^2)^2-\mu_q^2}{2.\sigma_q^2}\right). \\
&\int_{-\infty}^{+\infty} \exp\left(-\frac{(g-(\mu_q+j.w_q.\psi.\sigma_q^2))^2}{2.\sigma_q^2}\right).dg.
\end{aligned}
\tag{C.8}
$$

This can then be simplified by relying upon the symmetry of a Gaussian function integral, i.e.:

$$
\frac{1}{\sqrt{2.\pi.\sigma_q^2}}.\int_{-\infty}^{+\infty} \exp\left(-\frac{(g-(\mu_q+j.w_q.\psi.\sigma_q^2))^2}{2.\sigma_q^2}\right).dg = 1. \tag{C.9}
$$

So that:

$$
\begin{aligned}
\mathbf{\Phi}_{w_q.X_q}(\psi) &= \exp\left(\frac{(\mu_q+j.w_q.\psi.\sigma_q^2)^2-\mu_q^2}{2.\sigma_q^2}\right), \\
&= \exp\left(\frac{2.\mu_q.j.w_q.\psi.\sigma_q^2-\psi^2.\sigma_q^4}{2.\sigma_q^2}\right), \\
&= \exp\left(\mu_q.j.w_q.\psi\right).\exp\left(-\frac{\psi^2.w_q^2.\sigma_q^2}{2}\right).
\end{aligned}
\tag{C.10}
$$

Therefore the characteristic function for the sum of Q Gaussian distributed random variables, $\Phi_S(\psi)$, is given by the product of the individual characteristic functions:

$$
\mathbf{\Phi}_S(\psi) = \prod_{q=1}^{Q} \mathbf{\Phi}_{wX_q}(\psi) = \exp\left(j.\psi.\sum_{\forall q} w_q.\mu_q\right).\exp\left(-\frac{\psi^2}{2}.\sum_{\forall q} w_q^2.\sigma_q^2\right). \tag{C.11}
$$

The PDF of X_S can then be determined (letting $\mu_s = \sum_{\forall q} w_q.\mu_q$ and $\sigma_s^2 = \sum_{\forall q} w_q^2.\sigma_q^2$):

$$\begin{aligned} p_S(s) &= \frac{1}{2.\pi} \int_{-\infty}^{+\infty} \mathbf{\Phi}_S(\psi). \exp(-j.\psi.s).d\psi \\ &= \frac{1}{2.\pi} \int_{-\infty}^{+\infty} \exp\left(j.\psi.\mu_s\right). \exp\left(-\frac{\psi^2.\sigma_s^2}{2}\right). \exp(-j.\psi.s).d\psi \\ &= \frac{1}{2.\pi} \int_{-\infty}^{+\infty} \exp\left(\frac{2.j.\psi.(\mu_s - s) - \psi^2.\sigma_s^2}{2}\right).d\psi \\ &= \frac{1}{2.\pi} \int_{-\infty}^{+\infty} \exp\left(-\frac{\psi^2.\sigma_s^2 - 2.j.\psi.(s - \mu_s)}{2}\right).d\psi \end{aligned} \tag{C.12}$$

Completing the square:

$$\begin{aligned} &= \frac{1}{2.\pi} \int_{-\infty}^{+\infty} \exp\left(-\frac{(\psi - j.(s - \mu_s)/\sigma_s^2)^2 - (j.(s - \mu_s)/\sigma_s^2)^2}{2/\sigma_s^2}\right).d\psi \\ &= \frac{1}{2.\pi} \exp\left(-\frac{(s - \mu_s)^2}{2.\sigma_s^2}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{(\psi - j.(s - \mu_s)/\sigma_s^2)^2}{2/\sigma_s^2}\right).d\psi \end{aligned} \tag{C.13}$$

Similarly, the symmetry of the Gaussian can be used to simplify the expression, i.e.:

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(\psi - j.(s - \mu_s)/\sigma_s^2)^2}{2/\sigma_s^2}\right).d\psi = \frac{\sqrt{2.\pi}}{\sigma_s} \tag{C.14}$$

So that:

$$p_S(s) = \frac{1}{\sqrt{2.\pi.\sigma_s^2}}. \exp\left(-\frac{(s - \mu_s)^2}{2.\sigma_s^2}\right) \tag{C.15}$$

This shows that the result of the sum of Q weighted Gaussian distributed random variables also results in a Gaussian distributed random variable with parameters given by: $\mu_s = \sum_{\forall q} w_q.\mu_q$ and $\sigma_s^2 = \sum_{\forall q} w_q^2.\sigma_q^2$.

Appendix D

Alternative Derivation of Rician Distribution

In chapter 6, the Rician distribution was derived utilizing the parameters used by Rice in 1938, [111]. For the derivation in chapter 6, the gradient magnitude distribution was derived for $Z_x : \mathcal{N}(\mu_x, \sigma_z^2)$ and $Z_y : \mathcal{N}(\mu_y, \sigma_z^2)$. However, these parameter values do not realistically model the situation where the gradient kernels (M_x and M_y) will both be covering a heterogeneous image region (see figure 6.3). If both kernels are covering a heterogeneous image region, then $Z_x : \mathcal{N}(\mu_x, \sigma_z^2)$ and $Z_y : \mathcal{N}(\mu_y, \sigma_z^2)$, so that the CDF of the 2-D gradient magnitude becomes (compare to equation 6.11):

$$P_Z(z) = \iint\limits_{z \geq \sqrt{z_x^2+z_y^2}} \frac{1}{2.\pi.\sigma_z^2}. \exp\left(-\frac{1}{2.\sigma_z^2}((z_x-\mu_x)^2 + (z_y-\mu_y)^2)\right).dz_x.dz_y \tag{D.1}$$

$$= \iint\limits_{z \geq \sqrt{z_x^2+z_y^2}} \frac{1}{2.\pi.\sigma_z^2}. \exp\left(-\frac{1}{2.\sigma_z^2}(z_x^2 - 2.z_x.\mu_x + \mu_x^2 + z_y^2 - 2.\mu_y.z_y + \mu_y^2)\right).dz_x.dz_y$$

A change of variables is necessary to remove the double integrals on the single function. Taking advantage of the circular symmetry of the function, where $z^2 = z_x^2 + z_y^2$, $z_x = z.\sin(\theta)$, $z_y = z.\cos(\theta)$ and $dz_x.dz_y = z.d\theta.dz$, so that (also letting $U^2 = \mu_x^2 + \mu_y^2$):

$$P_Z(z) = \frac{1}{2.\pi.\sigma_z^2}. \int\limits_{-\infty}^{z} \int\limits_{0}^{2.\pi} \exp\left(-\frac{1}{2.\sigma_z^2}(z^2 - 2.z.(\mu_x.\sin(\theta) + \mu_y.\cos(\theta)) + U^2)\right).z.d\theta.dz, \tag{D.2}$$

$$= \frac{1}{2.\pi.\sigma_z^2}. \exp\left(-\frac{U^2}{2.\sigma_z^2}\right) \int\limits_{-\infty}^{z} \exp\left(-\frac{z^2}{2.\sigma_z^2}\right).z \int\limits_{0}^{2.\pi} \exp\left(\frac{z.(\mu_x.\sin(\theta) + \mu_y.\cos(\theta))}{\sigma_z^2}\right).d\theta.dz.$$

The trigonometric terms for the argument of the exponential in the inner integral, given by $(\mu_x.\sin(\theta) + \mu_y.\cos(\theta))$ maybe reexpressed as $U.\cos(\theta - \gamma)$ where γ is a phase shift,

so that:

$$P_Z(z) = \frac{1}{2.\pi.\sigma_z^2}.\exp\left(-\frac{U^2}{2.\sigma_z^2}\right)\int_{-\infty}^{z}\exp\left(-\frac{z^2}{2.\sigma_z^2}\right).z\int_{0}^{2.\pi}\exp\left(\frac{z.U.\cos(\theta-\gamma)}{\sigma_z^2}\right).d\theta.dz. \tag{D.3}$$

As the inner integral is of the same form as a special function, known as the modified Bessel function of the first kind, order zero, the inner integral can be replaced by using, [47], $\int_0^{2.\pi}\exp\left(\frac{z.\cos(\theta-\gamma).U}{\sigma_z^2}\right).d\theta = 2.\pi.I_0\left(\frac{z.U}{\sigma_z^2}\right)$, the CDF then becomes:

$$P_Z(z) = \frac{1}{\sigma_z^2}.\exp\left(-\frac{U^2}{2.\sigma_z^2}\right)\int_{-\infty}^{z}\exp\left(-\frac{z^2}{2.\sigma_z^2}\right).I_0\left(\frac{z.U}{\sigma_z^2}\right).z.dz. \tag{D.4}$$

This CDF, $P_Z(z)$ is now described in terms of the PDF of Z:

$$P_Z(z) = \int_{-\infty}^{z} p_Z(z).dz,$$

so the PDF of Z can be obtained from the CDF by removing the integral in equation D.4:

$$p_Z(z) = \frac{z}{\sigma_z^2}.\exp\left(-\frac{U^2+z^2}{2.\sigma_z^2}\right).I_0\left(\frac{z.U}{\sigma_z^2}\right). \tag{D.5}$$

Comparison of this equation with equation 6.18, suggests that the two equations are equivalent, where $U = \sqrt{\mu_x^2+\mu_y^2}$, so that if $\mu_x = 0$ then $U = \mu_y$ (where μ_y was given as μ_z in equation 6.18).

Appendix E

The Nelder-Mead Simplex Optimization Algorithm

Nelder and Mead in 1965 developed the Nelder-Mead Simplex optimization algorithm to minimize any arbitrary error function with respect to some data and a model describing the data [92]. A central concept to this simple optimization algorithm is a geometric figure known as a Simplex. A Simplex is defined by Borowski and Borwein in Collins Dictionary of Mathematics as [11]

> "...the most elementary geometric figure of a given dimension: the line in one dimension, the triangle in two dimensions, the tretrahedron in three, etc."

The Nelder-Mead Simplex algorithm utilizes a Simplex in the parameter-error space of the model that is being compared with some data. For every individual parameter of the data model there is a corresponding dimension in the parameter-error space. Therefore a Simplex is composed of $N + 1$ data points, where N corresponds to the number of dimensions or parameters for a particular data model. For every Simplex point, $\boldsymbol{y}_i$, the model, $f(\boldsymbol{y}_i)$ can be evaluated yielding a result that can be quantitatively compared with some corresponding data. The result is a set of points in the parameter-error space that the algorithm can determine error values. These error values can then be compared with each other and therefore determine which point (or set of particular parameter values) is least optimal.

The point with the greatest error between the model and the data for a particular set of parameter values (defining its location in the parameter-error space) is then reflected about the remaining points in the Simplex. A number of these reflections usually moves the Simplex to towards a minimal error for all of the points in the Simplex. If a new reflected point produces an error that is not lower than the previous position, then the reflection distance is reduced. This produces a focusing effect where the size of the Simplex reduces when it approaches a local minima.

A simple example is illustrated in figure E.1.

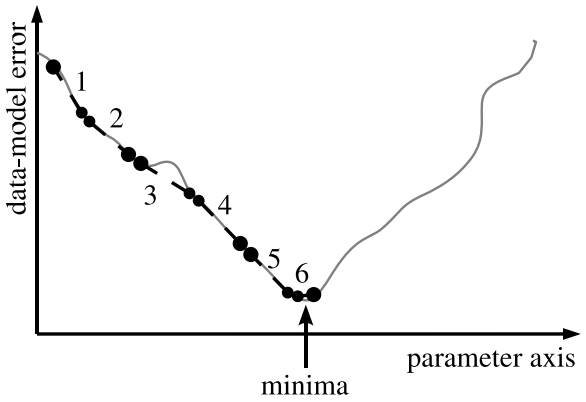


Figure E.1: An illustration of a simple example of the Simplex Nelder-Mead algorithm for a single parameter data model. Six Simplices are illustrated (numbers 1 to 6) where points for each Simplex move along the line of error for various values of the single parameter, until the final Simplex straddles the minima.

The algorithm may be terminated when the points of the Simplex are all approximately equal.

Bibliography

[1] A.H. Andersen and J.E. Kirsch. Analysis of noise in phase contrast MR imaging. *Med. Phys.*, 23:857–869, 1996.

[2] M.S. Atkins and B.T. Mackiewich. Fully automatic segmentation of the brain in MRI. *IEEE Trans. Med. Imaging*, 17:98–107, 1998.

[3] M.A. Gonzalez Ballester, A.P. Zisserman, and M. Brady. Estimation of the partial volume effect in MRI. *Medical Image Analysis*, 6:389–405, 2002.

[4] J. Barnes, R.I. Scahill, R.G. Boyes, C. Frost, E.B. Lewis, C.L. Rossor, M.N. Rossor, and N.C. Fox. Differentiating AD from aging using semiautomated measurement of hippocampal atrophy rates. *Neuroimage*, 23:574–581, 2004.

[5] M.A. Barry. Partial volume corrections in PET. Master's thesis, Dept. of Physics, School of Electronics and Physical Sciences, University of Surrey, UK, 2005.

[6] F. Bello, A.C.F. Colchester, and S.A. Roll. A geometry- and intensity-based partial volume correction for MRI volumetric studies. *Comput. Med. Imag. Graphics*, 22:123–132, 1998.

[7] F. Benford. The law of anomalous numbers. *Proc. Amer. Phil. Soc.*, 78:551–572, 1938.

[8] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Stat. Soc.*, 36:192–236, 1974.

[9] J.H. Bisese and A.M. Wang. *Pediatric Cranial MRI.* Springer-Verlag, 1994.

[10] M. Bomans, K.-H. Hohne, U. Tiede, and M. Riemer. 3-D segmentation of MR images of the head for 3-D display. *IEEE Trans. Med. Imaging*, 9:177–183, 1990.

[11] E.J. Borowski and J.M. Borwein. *Dictionary of Mathematics.* Collins, Glasgow, 1989.

[12] McGill University Brain Imaging Center, Montreal Neurological Institute. Simulated MRI brain data. "http://www.bic.mni.mcgill.ca/brainweb".

[13] P.A. Bromiley, M.L.J. Scott, M. Pokric, A.J. Lacey, and N.A. Thacker. Bayesian and non-Bayesian probabilistic models for magnetic resonance image analysis. *Image Vis. Comp.*, 21:851–64, 2003.

[14] J. Browne and A.R. De-Pierro. A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography. *IEEE Trans. Med. Imag.*, 15:687–699, 1996.

[15] M.E. Brummer, R.M. Mersereau, R.L. Eisner, and R.R.J. Lewine. Automatic detection of brain contours in MRI data sets. *IEEE Trans. Med. Imaging*, 12:153–166, 1993.

[16] L. Carlitz. The inverse of the error function. *Pacific J. Math*, 13:459–470, 1963.

[17] J. Chiverton, B. Podda, K. Wells, and D. Johnson. Fully automatic skull stripping of routine clinical neurological NMR data. In *Proc. IEEE Nuc. Sci. Symp. Med. Imag. Conf., Rome 2004*, volume 4, pages 2669–2673, Italy, October 2004.

[18] J. Chiverton and K. Wells. A partial volume 3-D gradient magnitude model. In *Conf. Proc. Med. Image Und. Anal.*, pages 49–52. BMVA, Imperial College, London, UK, 2004.

[19] J. Chiverton and K. Wells. Estimation of partial volume mixtures with a confidence measure combining intensity and gradient magnitude information. In *MIUA Conf. Rec.*, pages 227–230, Bristol, UK, 2005. British Machine Vision Association.

[20] J. Chiverton and K. Wells. A locally adaptive gradient controlled spatial regularization partial volume classifier. In *Proc. British Chapter of ISMRM*, 2006. Accepted for poster presentation.

[21] J. Chiverton and K. Wells. Mixture effects in FIR low-pass filtered signals. *IEEE Sig. Proc. Letters*, 13:369–372, 2006.

[22] J. Chiverton, K. Wells, E. Lewis, C. Chen, B. Podda, and D. Johnson. Statistical morphological skull stripping of adult and infant MRI data. *Comp. Biol. Med.*, 2007. Accepted for publication.

[23] J.P. Chiverton and K. Wells. Volumetric partial volume quantification via a statistical model of 3-D voxel gradient magnitude. In *Conf. Proc. IEEE Nuc.Sci.Symp. Med.Imag.Conf.*, volume 7, pages 4106–4110. IEEE, 2004.

[24] H.S. Choi, D.R. Haynor, and Y. Kim. Partial volume tissue classification of multichannel magnetic resonance images-a mixel model. *IEEE Trans. Med. Imag*, 10:395–407, 1991.

[25] S.M. Choi, J.E. Lee, and M.H. Kim. Volumetric object reconstruction using the 3D-MRF model-based segmentation. *IEEE Trans. Med. Imag*, 16:887–892, 1997.

[26] P. Clifford. Markov random fields in statistics. In G.R. Grimmett and D.J.A. Welsh, editors, *Disorder in physical systems*, pages 19–32, 1990.

[27] C.A. Cocosco, V. Kollokian, R. Kwan, and A.C. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5:S425, 1997.

[28] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes, and A.C. Evans. Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imag.*, 17:463–468, 1998.

[29] R.W. Cox. Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.*, 29:162–173, 1996.

[30] N. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1993.

[31] A.R. Crossman and D. Neary. *Neuroanatomy, An illustrated colour text*. Churchill Livingstone, 2000.

[32] A.M. Dale, B. Fischl, and M.I. Sereno. Cortical surface-based analysis. *Neuroimage*, 9:179–194, 1999.

[33] A.P. Dempster, N. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 42:1–38, 1977.

[34] P.A. Devijver and J. Kittler. *Pattern Recognition, A Statistical Approach*. Prentice-Hall, 1982.

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

[36] J.C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-seperated clusters. *J. Cybernetics*, 3:32–57, 1974.

[37] E.Ifeacher and B.Jervis. *Digital Signal Processing A Practical Approach*. Addison-Wesley, 1993.

[38] "Center for Morphometric Analysis at Massachusetts General Hospital". "20 normal MR brain data sets and their manual segmentations". "http://www.cma.mgh.harvard.edu/ibsr/".

[39] P.A. Freeborough, N.C. Fox, and R.I. Kitney. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Comput. Methods Programs Biomed.*, 53:15–25, 1997.

[40] J.K. Fwu and P.M. Djuric. Unsupervised vector image segmentation by a tree structure-ICM algorithm. *IEEE Trans. Med. Imag.*, 15:871–880, 1996.

[41] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian Data Analysis*. Chapman and Hall, CRC, 2004.

[42] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-6:721–224, 1984.

[43] L. Germond, M. Dojat, C. Taylor, and C. Garbay. A coorperative framework for segmentation of MRI brain scans. *Artif. Intell. Med.*, 20:77–93, 2000.

[44] J.W. Gibbs. *Elementary Principles in Statistical Mechanics*. Yale University Press, 1902.

[45] G.H. Givens and J.A. Hoeting. *Computational Statistics*. Wiley, 2005.

[46] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Addison-Wesley, 1992.

[47] I.S. Gradshteyn and I.M. Ryzhik. *Tables of Integrals, Series, and Products, 4th Corrected Enlarged Edition*. Academic Press, 1980.

[48] H. Gudbjartsson and S. Patz. The Rician distribution of noisy MRI data. *Magn. Reson. Med.*, 34:910–914, 1995.

[49] R.C. Gur, F.M. Gunning-Dixon, B.I. Turetsky, W.B. Bilker, and R.E Gur. Brain region and sex differences in age association with brain volume a quantitative MRI study of healthy young adults. *Am. J. Geriatr. Psychiatry*, 10:72–80, 2002.

[50] E.M. Haacke, R.W. Brown, M.R. Thompson, and R. Venkateson. *Magnetic Resonance Imaging Physical Principles and Sequence Design*. Wiley-Liss, New York, USA, 1999.

[51] H.K. Hahn, W.S. Millar, M.S. Durkin, O. Klinghammer, and H.O. Peitgen. Cerebral ventricular volumetry in pediatric neuroimaging. In *Bildverarbeitung fur die Medizin 2002, Algorithmen - Systeme - Anwendungen. Proceedings des Workshop vom 10.-12. Marz 2002 in Leipzig*. Springer Verlag, 2002.

[52] H.K. Hahn and H.-O. Peitgen. The skull stripping problem in MRI solved by a single 3D watershed transform. In *MICCAI 2000*, pages 134–143. Springer, 2000.

[53] J.M. Hammersley and P. Clifford. Markov field on finite graphs and lattices. *Unpublished*, 1972.

[54] E.R. Hancock and J. Kittler. Edge-labeling using dictionary-based relaxation. *IEEE Trans. Patt. Anal. Mach. Int.*, 12:165–181, 1990.

[55] R.H. Hashemi, W.G. Bradley, and C.J. Lisanti. *MRI The Basics*. Lippincott Williams and Wilkins, Philadelphia, USA, 2004.

[56] E. Hayman and J.O. Eklundh. Statistical background subtraction for a mobile observer. In *Proc. Ninth IEEE ICCV 2003*, volume 1, pages 67–74, 2003.

[57] K. Held, E. Rota Kops, B.J. Krause, W.M. Wells, R. Kikinis, and H.-W. Muller-Gartner. Markov random field segmentation of brain MR images. *IEEE Trans. Med. Imag*, 16:878–886, 1997.

[58] T. Hill. A statistical derivation of the significant-digit law. *Stat. Sci.*, 10:354–363, 1996.

[59] S. Huh, T.A. Ketter, K. Hoon Sohn, and C. Lee. Automated cerebrum segmentation from three-dimensional sagittal brain MR images. *Comput. Biol. Med.*, 32:311–328, 2002.

[60] M. Ibrahim, N. John, M. Kabuka, and A. Younis. Hidden Markov models-based 3D MRI brain segmentation. *Image Vis. Comp.*, 2006. accepted for publication.

[61] N. Iwasaki, K. Hamano, Y. Okada, Y.Horigome, J. Nakayama, T. Takeya, H. Takita, and T. Nose. Volumetric quantification of brain development using MRI. *Neuroradiology*, 39:841–846, 1997.

[62] N. Johnson, S. Kotz, and N. Balakrishnan. *Continuous Univariate Distributions, Volumes 1 and 2*. Wiley, Chichester, UK, 1999.

[63] E. Johnstone, M. Humphreys, F. Lang, S. Lawrie, and R. Sandler. *Schizophrenia Concepts and Clinical Management*. Cambridge, 1999.

[64] N.B. Joshi and M. Brady. Partial volume segmentation of MR images using non-parametric mixture model. In *Conf. Proc. Med. Image Und. Anal.* BMVA, University of Bristol, UK, 2005.

[65] T. Kapur, W.E.L. Grimson, W.M. Wells III, and R. Kikinis. Segmentation of brain tissue from magnetic resonance images. *Med. Image Anal.*, 1:109–127, 1996.

[66] A. Kitamoto and M.Takagi. The area proportion distribution-relationship with the internal structure of mixels and its application to image classification. *Sys. Comps. Japan*, 31:57–76, 2000.

[67] A. Kitamoto and M. Takagi. Image classification using probabilistic models that reflect the internal structure of mixels. *Patt. Anal. App.*, 2:31–43, 1999.

[68] J. Kittler and J. Foglein. On compatibility and support functions in probabilistic relaxation. *Comp. Vis. Graph. Image Proc.*, 34:257–267, 1986.

[69] J. Kittler and E.R. Hancock. Combining evidence in probabilistic relaxation. *Int. J. on Patt. Recog. Art. Int.*, 3:29–52, 1989.

[70] J. Kittler and J. Illingworth. Relaxation labelling algorithms - a review. *Image Vis. Comp.*, 2:206–216, 1985.

[71] F. Kuo and J.F. Kaiser, editors. *System Analysis by Digital Computer*, chapter Digital Filters. Wiley, 1966.

[72] R. Kwan, A. Evans, and G. Pike. An extensible MRI simulator for post-processing evaluation. In *Lecture Notes in Computer Science. Vol. 1131*, pages 135–140, Berlin, 1996. Springer-Verlag.

[73] R.K.-S. Kwan, A.C. Evans, and G.B. Pike. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans. Med. Imag.*, 18:1085–1097, 1999.

[74] D.H. Laidlaw, K.W. Fleischer, and A.H. Barr. Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Trans. Med. Imag.*, 17:74–86, 1998.

[75] C. Lee, S. Huh, T.A. Ketter, and M. Unser. Unsupervised connectivity-based thresholding segmentation of midsagittal brain MR images. *Comput. Biol. Med.*, 28:309–338, 1998.

[76] J.-M. Lee, U. Yoon, S. Hee Ham, J.H. Kim, I.-Y. Kim, and S.I. Kim. Evaluation of automated and semi-automated skull-stripping algorithms using similarity index and segmentation error. *Comput. Biol. Med.*, 33:495–507, 2003.

[77] P.M. Lee. *Bayesian Statistics An Introduction*. Arnold, London, UK, 1997.

[78] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans. Med. Imag*, 22:105–119, 2003.

[79] L. Lemieux, G. Hagemann, K. Krakow, and F.G. Woermann. Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. *Magn. Reson. Med.*, 42:127–135, 1999.

[80] S.Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer, 2001.

[81] Z. Liang, R.J. Jaszczak, and R.E. Coleman. Parameter estimation of finite mixtures using the EM algorithm and information criteria with application to medical image processing. *IEEE Trans. Nuc. Sci.*, 39:1126–1133, 1992.

[82] Z.P. Liang and P.C. Lauterbur. *Principles of Magnetic Resonance Imaging*. SPIE Press and IEEE Presss, Washington, USA and New York, USA, 2000.

[83] J.M. Links, L.S. Beach II, B. Subramaniam, M.A. Rubin, J.G. Hennessey, and A.L. Reiss. Edge complexity and partial volume effects. *J. Comput. Assist. Tomogr.*, 22:450–458, 1998.

[84] J.H. MacCabe, S.L. Rossell, X.A. Chitnis, A.S. David, and R.M. Murlay. Cognitive decline is associated with reduction in global white matter volume in schizophrenia. In *International Congress on Schizophrenia Research*, pages 144–145, 2003.

[85] D. MacDonald, N. Kabani, D. Avis, and A.C. Evans. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *Neuroimage*, 12:340–356, 2000.

[86] E.A. Martin. *Concise Colour Medical Dictionary*. Oxford, 1998.

[87] J. Matsuzawa, M. Matsui, T. Konishi, K. Noguchi, R.C. Gur, W. Bilker, and T. Miyawaki. Age-related volumetric changes of brain gray and white matter in healthy infants and children. *Cerebral Cortex*, 11:335–342, 2001.

[88] D.W. McRobbie, E.A. Moore, M.J. Graves, and M.R. Prince. *MRI from picture to proton*. Cambridge University Press, 2003.

[89] M.S. Mega, I.D. Dinov, J.C. Mazziotta, M. Manese, P.M. Thompson, C. Lindshield, J. Moussai, N. Tran, K. Olsen, C.I. Zoumalan, R.P. Woods, and A.W. Toga. Automated brain tissue assessment in the elderly and demented population: Construction and validation of a sub-volume probabilistic brain atlas. *Neuroimage*, 26:1009–1018, 2005.

[90] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Chem. Phys.*, 21:1087–1092, 1953.

[91] N. Miura, A. Taneda, K. Shida, R. Kawashima, Y. Kawazoe, H. Fukuda, and T. Shimizu. Automatic brain tissue extraction method using erosion-dilation treatment (BREED) from three-dimensional magnetic resonance imaging T1-weighted data. *J. Comput. Assist. Tomogr.*, 26:927–932, 2002.

[92] J.A. Nelder and R. Mead. A Simplex method for function minimization. *Comp. Journal*, 7:308–313, 1965.

[93] S. Newcomb. Note on the frequency of use of the different digits in natural numbers. *Amer. J. Math*, 4:S. 39–40, 1881.

[94] N. Nikolaidis and I. Pitas. *3-D Image Processing Algorithms*. Wiley, New York, USA, 2001.

[95] A. Noe and J.C. Gee. Partial volume segmentation of cerebral MRI scans with mixture model clustering. In *Conf. Proc. XVIIth Int Conf.Inf Proc Med. Imag*, page 423. Springer Verlag, 2001.

[96] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, USA, 1991.

[97] E. Parzen. On estimation of a probability density function and mode. *Annals Math Stat*, 33:1065–1076, 1962.

[98] J.S. Paulsen, V.A. Magnotta, A.E. Mikos, H.L. Paulson, E. Penziner, N.C. Andreasen, and P.C. Nopoulos. Brain structure in preclinical huntington's disease. *Biol. Psychiatry*, 59:57–63, 2006.

[99] A. Pfefferbaum, D. Mathalon, E. Sullivan, J. Rawles, R. Zipurksy, and K. Lim. A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. *Arch. Neuro.*, 51:874–887, 1994.

[100] D. Pham and J. Prince. A generalized EM algorithm for robust segmentation of magnetic resonance images. In *Proc. 33rd Ann. Conf. Inf. Sciences and Systems (CISS99)*, pages 558–563, 1999.

[101] D.L. Pham and J. Prince. Unsupervised partial volume estimation in single-channel image data. In *Proc. IEEE Workshop Mathematical Methods in Biomedical Image Analysis-MMBIA'00*, pages 170–177. IEEE, 2000.

[102] M. Pokric, N.A. Thacker, M.L.J. Scott, and A. Jackson. The importance of partial voluming in multi-dimensional medical image segmentation. In *Conf. Proc. MICCAI 2001*, pages 1293–1294. Springer, 2001.

[103] I. Poole. Optimal probabilistic relaxation labeling. In *Proc. BMVC*, pages 91–96, 1990.

[104] W. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 1992.

[105] D. Purves, G.J. Augustine, D. Fitzpatrick, L.C. Katz, A.S. LaMantia, J.O. McNamara, and S.M. Williams. *Neuroscience Second Edition*. Sinauer, 2001.

[106] J.C Rajapakse, J.N. Giedd, and J.L. Rapoport. Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Trans. Med. Imag.*, 16:176–186, 1997.

[107] J.C Rajapakse and F. Kruggel. Segmentation of MR images with intensity inhomogeneities. *Image Vis. Comp.*, 16:165–180, 1998.

[108] K. Rehm, K. Schaper, J. Anderson, R. Woods, S. Stoltzner, and D. Rottenberg. Putting our heads together: a consensus approach to brain/non-brain segmentation in T1-weighted MR volumes. *Neuroimage*, 22:1264–1270, 2004.

[109] A.L. Reiss, J.G. Hennessey, M. Rubin, L. Beach, M.T. Abrams, I.S. Warsofsky, A.M.C. Liu, and J.M. Links. Reliability and validity of an algorithm for fuzzy tissue segmentation of MRI. *J. Comput. Assist. Tomogr.*, 22:471–479, 1998.

[110] D.E. Rex, D.W. Shattuck, R.P. Woods, K.L. Narr, E. Luders, K. Rehm, S.E. Stolzner, D.A. Rottenberg, and A.W. Toga. A meta-algorithm for brain extraction in MRI. *Neuroimage*, 23:625–637, 2004.

[111] S.O. Rice. Mathematical analysis of random noise. *Bell Systems Tech J*, 23:1–162, 1938.

[112] R.A. Robb. *Biomedical Imaging, Visualization and Analysis*. Wiley-Liss, USA, 2000.

[113] C.P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

[114] C. Rorden and M. Brett. Stereotaxic display of brain lesions. *Behav. Neurol.*, 12:191–200, 2000.

[115] A. Rosenfeld, R.A. Hummel, and S.W. Zucker. Scene labelling by relaxation operators. *IEEE Trans. Sys. Man Cyb.*, 6:420–433, 1976.

[116] D.B. Rowe. *Multivariate Bayesian Statistics, Models for Source Separation and Signal Unmixing*. Chapman and Hall, 2003.

[117] S. Ruan, C. Jaggi, J. Xue, J. Fadili, and D. Bloyet. Brain tissue classification of magnetic resonance images using partial volume modeling. *IEEE Trans. Med. Imag*, 19:1179–1187, 2000.

[118] M. Rutherford. *MRI of the Neonatal Brain*. W.B. Saunders, 2002.

[119] G. Salamon, C. Raynaud, J. Regis, and C. Rumeau. *Magnetic Resonance Imaging of the Pediatric Brain*. Raven Press, New York, 1990.

[120] S. Sandor and R. Leahy. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Trans. Med. Imaging*, 16:41–54, 1997.

[121] S. Sanjay-Gopal and T.J. Hebert. Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. *IEEE Trans. Image Proc.*, 7:1014–1028, 1998.

[122] P. Santago and H.D. Gage. Quantification of MR brain images by mixture density and partial volume modeling. *IEEE Trans. Med. Imag*, 12:566–574, 1993.

[123] P. Santago and H.D. Gage. Statistical models of partial volume effect. *IEEE Trans. Imag. Proc*, 4:1531–1540, 1995.

[124] R.B. Sassi, M. Nicoletti, P. Brambilla, A.G. Mallinger, E. Frank, D.J. Kupfer, M.S. Keshavan, and J.C. Soares. Increased gray matter volume in lithium-treated bipolar disorder patients. *Neuroscience Letters*, 329:243–245, 2002.

[125] P. Schroeter, J.M. Vesin, T. Langenberger, and R. Meuli. Robust parameter estimation of intensity distributions for brain magnetic resonance images. *IEEE Trans. Med. Imag.*, 17:172–186, 1998.

[126] F. Segonne, A.M. Dale, E. Busa, M. Glessner, D. Salat, H.K. Hahn, and B. Fischl. A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22:1060–1075, 2004.

[127] Z.Y. Shan, G.H. Yue, and J.Z. Liu. Auomated histogram-based brain segmentation in T1-weighted three-dimensional magnetic resonance head images. *Neuroimage*, 17:1587–1598, 2002.

[128] D.W. Shattuck, S.R. Sandor-Leahy, K.A. Schaper, D.A. Rottenberg, and R.M. Leahy. Magnetic resonance image tissue classification using a partial volume model. *Neuroimage*, 13:856–876, 2001.

[129] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imag.*, 17:87–97, 1998.

[130] F.A. Smith. *A Primer in Applied Radiation Physics*. World Scientific, 2000.

[131] S. Smith. Fast robust automated brain extraction. *Hum. Brain Mapp.*, 17:143–155, 2002.

[132] H. Soltanian-Zadeh and J.P. Windham. A multiresolution approach for contour extraction from brain images. *Med. Phys.*, 24:1844–1853, 1997.

[133] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision Second Edition*. PWS, 1999.

[134] H. Stark and J.W. Woods. *Probability and Random Processes with Applications to Signal Processing Third Edition*. Prentice-Hall, USA, 2002.

[135] R. Stokking, K.L. Vincken, and M.A. Viergever. Automatic morphology-based brain segmentation (MBRASE) from MRI-T1 data. *Neuroimage*, 12:726–738, 2000.

[136] K.A. Stroud. *Further Engineering Mathematics*. Macmillan, London, 1996.

[137] M.A. Tanner. *Tools for Statistical Inference*. Springer, 1996.

[138] N.A. Thacker and M. Pokric. Noise filtering and testing for MR using a multi-dimensional partial volume model. In *Conf. Proc. BMVC 2004*. BMVA, UK, 2004.

[139] N.A. Thacker, M. Pokric, and D.C. Williamson. Multi-dimensional medical image segmentation with gradient analysis. In *Conf. Proc. Med. Image Und. Anal. , Bristol, UK, 2005*, pages 23–26. BMVA, UK, 2005.

[140] N.A. Thacker, D.C. Williamson, and M. Pokric. Voxel based analysis of tissue volume MRI data. Technical report, University of Manchester, 2003.

[141] N.A. Thacker, D.C. Williamson, and M. Pokric. Voxel based analysis of tissue volume from MRI data. *B. J. Rad.*, 77:S114–S125, 2004.

[142] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. John Wiley and Sons, Chichester, UK, 1992.

[143] A.W. Toga and P.M. Thompson. Image registration and the construction of multidimensional brain atlases. In *Handbook of Medical Imaging Processing and Analysis*, pages 449–463. Academic Press, 2000.

[144] J. Tohka, A. Zijdenbos, and A. Evans. Fast and robust parameter estimation for statistical partial volume models in brain MRI. *Neuroimage*, 23:84–97, 2004.

[145] N. Tzourio-Mazoyer, F. Crivello, M. Joliot, and B. Mazoyer. Biological underpinnings of anatomic consistency and variability in the human brain. In *Handbook of Medical Imaging Processing and Analysis*, pages 449–463. Academic Press, 2000.

[146] Rice University. Notes on characteristic functions. http://cnx.rice.edu/content/m11071/latest/.

[147] G. Villarreal, D.A. Hamilton, H. Petropoulos, I. Driscoll, L.M. Rowland, J.A. Griego, P.W. Kodituwakku, B.L. Hart, R. Escalona, and W.M. Brooks. Reduced hippocampal volume and total white matter volume in posttraumatic stress disorder. *Biol. Psychiatry*, 52:119–125, 2002.

[148] E.A. Vokurka, A. Herwadkar, and N.A. Thacker. Using Bayesian tissue classification to improve the accuracy of Vestibular Schwannoma volume and growth measurement. *Amer. J. Neurorad.*, 23:459–467, 2002.

[149] E.A. Vokurka, N.A. Thacker, and A. Jackson. A fast model independent method for automatic correction of intensity non-uniformity in MRI data. *J. Magn. Reson. Img.*, 10:550–562, 1999.

[150] Y. Wang, T. Adali, J. Xuan, and Z. Szabo. Magnetic resonance image analysis by information theoretic criteria and stochastic site models. *IEEE Trans. Inform Tech Biomed*, 5:150–158, 2001.

[151] A. Webb. *Statistical Pattern Recognition*. Arnold, 1999.

[152] K. Wells, J. Chiverton, M. Partridge, M. Barry, H. Kadhem, and R. Ott. The partial volume effect in PET/SPECT and Benford's Law. In *Proc. IEEE NSS and MIC, Puerto Rico*, pages 1804–1808, October 2005.

[153] K. Wells, J. Chiverton, M. Partridge, M. Barry, H. Kadhem, and R. Ott. Quantifying the partial volume effect in PET using Benford's Law. *Submitted to IEEE Trans. Nucl. Sci.*, 2007.

[154] W.M. Wells, W.E.L. Grimson, R. Kikinis, and F.A. Jolesz. Adaptive segmentation of MRI data. *IEEE Trans. Med. Imag*, 15:429–442, 1996.

[155] M. Wernick and J. Aarsvold. *Emission Tomography The Fundamentals of PET and SPECT*. Elsevier Academic Press, 2004.

[156] Y. Wexler, A. Fitzgibbon, and A. Zisserman. Bayesian estimation of layers from multiple images. In *Proc. ECCV, Copenhagen, Denmark, 2002*, pages 487–501, 2002.

[157] D.C. Williamson, N.A. Thacker, S.R. Williams, and M. Pokric. Partial volume tissue segmentation using grey-level gradient. In *Conf. Proc. Med. Image Und. Anal.*, pages 17–20. BMVA, UK, 2002.

[158] J.P. Windham, M.A. Abd-Allah, D.A. Reimann, J.W. Froelich, and A.M. Haggar. Eigenimage filtering in MR imaging. *J. Comp. Assis. Tomog.*, 12:1–9, 1988.

[159] J.P. Windham, A.M. Haggar, D.O. Hearshen, J.R. Roebuck, and D.A. Reimann. A novel method for volume determination using MR image sequences. In *Book of Abstracts: Society of Magnetic Resonance in Medicine, Seventh Annual Meeting, San Francisco, California, 22-26 August 1988.*, page 1081. Society of Magnetic Resonance in Medicine, 1988.

[160] G. Winkler. *Image Analysis, Random Fields, and Markov Chain Monte Carlo Methods Second Edition*. Springer, Germany, 2003.

[161] C.S. Won and R.M. Gray. *Stochastic Image Processing*. Kluwer Academic/ Plenum Publishers, 2004.

[162] M.W. Woolrich, T.E.J. Behrens, C.F. Beckmann, and S.M. Smith. Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Trans. Med. Imag*, 24:1–11, 2005.

[163] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imag*, 20:45–57, 2001.

[164] J. Zhong, A.M. Seifalian, G.E. Salerud, and G.E. Nilsson. The biological zero problem in laser doppler flowmetry. *IEEE Trans. Biom. Eng.*, 45:356–364, 1998.