Chapter 2

Background and Literature Review

2.1 Introduction

In this chapter a review of the background and previous work with respect to the research described in this thesis is presented. The chapter starts, Section 2.2, with a review of the MRI brain scan application domain. In Section 2.3 a brief review is then presented concerning image preprocessing, a necessary precursor to any image analysis process. Broadly the work described in this thesis falls within the domain of data mining, and more specifically image mining. Section 2.4 therefore introduced the concept of data mining and then goes on to consider, in more detail, the sub-domains of image mining and image classification. In particular the application of Case Based Reasoning (CBR) to the classification problem and frequent subgraph mining are discussed, because these are two of the techniques used (in part) to provide an answer to the research question central to this thesis (how best to process image collections so that efficient and effective classification, according to some ROI contained across the image set, can be achieved?).

One of the main issues associate with image classification is how best to represent the image input set so that an appropriate image mining technique can be applied. The nature of this representation is of course related to the nature of the image mining techniques to be adopted. A review of various representation techniques, that impact on the representations proposed in this thesis, is thus given in Section 2.5. A chapter summary is presented in Section 2.6.

2.2 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) came into prominence in the 1970s. MRI is similar to Computerized Topography (CT) in that cross-sectional images are produced of some object. MRI uses a strong magnetic field and radio waves to produce high quality and detailed computerized images of the inside of some object. MRI is based on the principle of Nuclear Magnetic Resonance (NMR), a spectroscopic technique used by scientists to obtain microscopic chemical and physical information about molecules. The first successful Nuclear Magnetic Resonance (NMR) experiments were conducted in 1946 independently by two scientists, Felix Bloch and Edward Purcell, both of whom were awarded the Nobel Prize in 1952. In 1977, Raymond Damadian discovered the basis for using MRI as a tool for medical diagnosis and demonstrated the first MRI examination of a human subject [32]. The technique was called MRI rather than Nuclear Magnetic Resonance Imaging (NMRI) because of the negative connotations associated with the word nuclear in the late 1970's. MRI started out as a tomographic imaging technique, that is it produced an image of the NMR signal in a thin slice through the human body. Since then MRI has advanced beyond a tomographic imaging technique to a 3D imaging technique.

MRI is commonly used to examine the spine, joints, abdomen, and pelvis. A special kind of MRI exam, called Magnetic Resonance Angiography (MRA) can be used to examine blood vessels. MRI is also used for brain diagnosis, for example to detect abnormal changes in different parts of the brain. A MRI of the brain produces a very detailed picture. An example brain scan image is given in Figure 2.1.

As noted in the previous chapter the focus of the work described in this thesis is directed at the classification/categorisation of MRI brain scans according to a particular feature (ROI) within these scans, namely the corpus callosum. MRI brain scans underpin the diagnosis and management of patients suffering from various neurological and psychiatric conditions. Analysis of MRI data relies on the expertise of specialists (radiologists) and is therefore subjective. More specifically the focus of the work is the classification of MRI brain scan data according to a feature called the corpus callosum. Figure 2.2 gives an example midsagittal slice of a MRI scan, the corpus callosum is located in the centre of

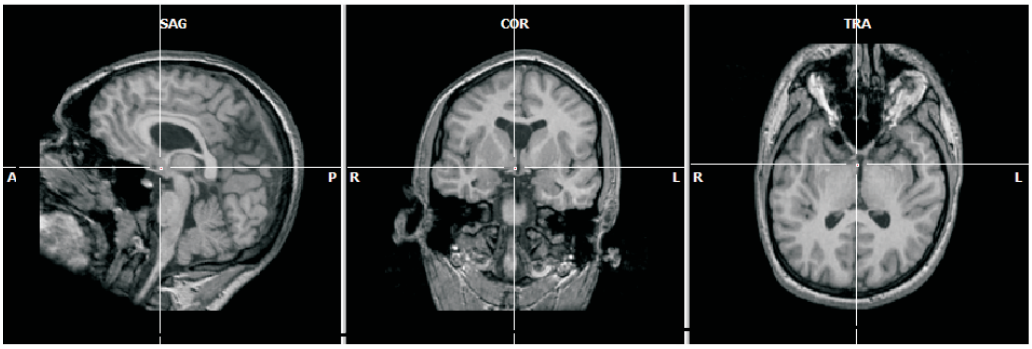

Figure 2.1: An example brain scan image. The three images show (from left to right) sagittal, coronal and axial planes. A common point is marked in each image.

the image. The *midsagittal slice* is the middle slice with a MRI scan, where a MRI scan comprised a sequence (or bundle) of "image slices".

The size and shape of the corpus callosum has been shown to be correlated to sex, age, neurodegenerative diseases (e.g. epilepsy, multiple sclerosis and schizophrenia) and various lateralized behaviour in people (such as handedness). It is also conjectured that the size and shape of the corpus callosum reflects certain human characteristics (such as a mathematical or musical ability). Within neuroimaging research considerable effort has been directed at quantifying parameters such as length, surface area and volume of structures in living adult brains, and investigating differences in these parameters between sample groups. Several studies indicate that the size and shape of the corpus callosum, in humans, is correlated to sex [4, 34, 128], age [128, 151], brain growth and degeneration [61, 98], handedness [31], epilepsy [27, 124, 149] and brain disfunction [38, 74]. It is worth noting that although the work described in this thesis is directed at MRI brain scan classification there are other features in MRI brain scans to which the techniques could be applied, such as the ventricles.

2.3 Image Preprocessing, Registration and Segmentation

The automated analysis of images, regardless of the nature of this analysis, typically requires some form of image pre-processing. This may include simple noise

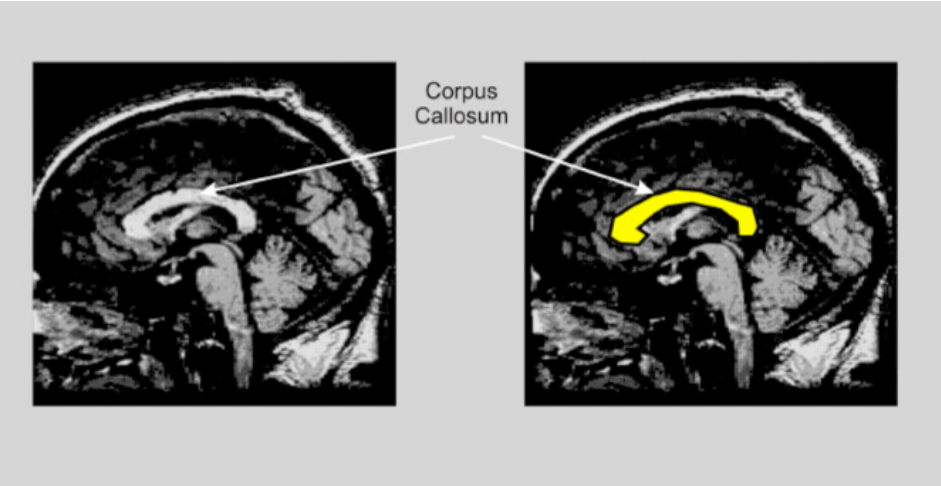

Figure 2.2: Midsagital MRI brain scan slice showing the corpus callosum (highlighted in the right-hand image).

removal and "deblurring" of objects. More advanced image preprocessing may require operations such as registration and segmentation.

Registration is the process where by one or more images are aligned with a reference image. This is important so that a collection of images can be effectively compared. Image registration is used in a range of application domains, such as medical image analysis (e.g. diagnosis), neuroscience (e.g. brain mapping), computer vision (e.g. stereo image matching for shape recovery), astrophysics (e.g. the alignment of images from different frequencies), etc. Image registration in the medical domain is particularly difficult when complex (e.g. nonlinear) geometric transformations are required to relate the images, e.g. when registering images of different human brains (multi-subject registration). Registration of medical image data sets can be defined as the problem of identifying a set of geometric transformations which map the coordinate system of one data set to that of the others. In the context of the work described in this thesis, image registration was undertaken to ensure that all brain MRI scans in a given collection conformed to a single set of coordinate axes. Therefore, all scans in the study have been aligned manually by trained physicians.

Image segmentation is defined as the partitioning of an image into non-

overlapping meaningful regions that are homogeneous with respect to some characteristic such as intensity, colour or texture or by extracting one or more specific objects in images such as medical structures. Ideally a segmentation method finds the group of pixels that correspond to anatomical structures or regions of interest in the image, usually this is done by distinguish objects or regions of interest (the "foreground") from everything else (the "background"). In the simplest cases there are only two pixel types (foreground and background), consequently the segmentation results in a binary image. Segmentation is often the first stage in an image analysis process. Once the regions of interest are isolated from the rest of the image, certain characterizing measurements can be made and these can be used to (say) classify the regions.

Image segmentation has gone through many stages of development and refinement over the last few decades [108]. However, any individual image segmentation technique is not likely to achieve reliable results under all circumstances; there is no all purpose "best" segmentation technique. Medical image segmentation is a key task in many medical applications such as surgical planning, post-surgical assessment and abnormality detection. Segmentation is used, for example: (i) in the detection of organs such as the brain, heart, lungs or liver in MRI scans; and (ii) to distinguish pathological tissues, such as tumours, from normal tissue. The most basic attributes used to identify the ROI are image grey scale level or brightness, but other properties such as texture can be used. Medical images mostly contain complicated structures and precise segmentation is often deemed necessary for clinical diagnosis. A lot of methods for automatic and semi-automatic image segmentation fail to partition medical images precisely due to poor image contrast, inhomogeneity and weak boundaries between regions. This includes brain image segmentation. The segmentation of brain MRI scans is a challenging problem that has received much attention [118, 154].

A variety of artifacts (noise) may appear in MRI data. Since the artifacts change the appearance of the image they may also affect the performance of a segmentation algorithm. The most significant artifacts in image segmentation are intensity inhomogeneities and the partial volume effect. Intensity inhomogeneities are not always visible to the human eye, but can nonetheless have negative influence on automatic segmentation. This may manifest itself by making intensities

in one part of the image brighter or darker than another part. It is often caused by the Radio Frequency (RF) coils used in the MRI acquisition process. The partial volume effect occurs when a pixel can not be accurately assigned to one tissue type. This is because the intensity of the pixel originates from more than one tissue. It occurs when one pixel covers a number of different tissue types and the collective signal emitted from these tissue cells makes up the detected intensity of the pixel. The partial volume effect is most apparent at edges between different tissues (regions).

Many image segmentation techniques are used in the context of brain MRI segmentation including pixel classification, deformable models, statistical models, fuzzy based approaches and graph based approaches. Each is discussed in the following sub-sections.

2.3.1 MRI Segmentation Using Pixel Classification

Pixel classification methods segment brain structures pixel by pixel based on spatial and appearance information, which are represented by pixel features. These pixel features can include pixel coordinates describing the location, as well as intensity values of filtered versions of MRI data that describe the appearance of the region of interest. In classification-based segmentation image pixels are represented as points in a high-dimensional space, in which the coordinate axes are defined by the feature values associated with the pixel (for example colour intensity). To segment an unlabelled target image using supervised classification techniques, first manually labelled example images are used to train a pixel classifier. This is done by first sampling pixels from the training images and mapping them to the feature space. In the feature space a decision boundary is then found that best separates the groups of pixels labelled as regions, from the pixels that were labelled as background in the examples. Different types of classifiers use different methods to derive this boundary from the training samples. After training, the classifier is applied to the unlabelled target pixels by mapping them into the feature space, and labelling them according to the decision boundary. The more pixel features that are used the better the classifier's ability to model the region's appearance and location. However, it also tends to increase the complexity of the decision boundary which increases the risk of over-training (where the

classifier is over tuned to a particular training set) causing errors when classifying images that were not used for training. This risk can be decreased by increasing the number of examples, constraining the complexity of the decision boundary, or decreasing the number of features by ignoring those that are not considered relevant to the classification accuracy. The most important differences between brain structure segmentation methods based on pixel classification are the number and type of pixel features used. In Powell et al. [117] up to 100 features were employed, including the intensities of pixels and their direct neighbours. Morra et al. [104] segmented the hippocampus using thousands of features and the AdaBoost ensemble classification method [53].

2.3.2 MRI Segmentation Using Deformable Models

Methods based on the *deformable model* paradigm, such as the *active contour model*, also called *snakes*, delineate brain regions of interest by fitting a boundary model to the image, which incorporates some form of global shape knowledge. Using Snake and Level Sets the shape information is relatively weak, as these techniques merely enforce smooth boundaries [11]. Statistical shape models enforce stronger shape constraints and are therefore better equipped to deal with low-contrast boundaries. These models are constructed by parametrizing the shapes of manually labelled examples, and learning their mean shape and typical variations. Examples of shape models used for brain segmentation have been reported in [71, 80]. The most important difference between these methods is the flexibility of the parametrization. As a general rule, models with many degrees of freedom can describe a complex boundary, but also require more labelled examples to represent the potential shape variation. The initialization of the boundary offers a way to incorporate spatial information in a deformable model. Deformable models operate by minimizing a cost function that generally has several local minima. This usually ensures that the final result is close to the initialization and can therefore be used as a de-facto constraint on the spatial domain of the segmentation.

2.3.3 MRI Segmentation Using Statistical Models

Methods based on the statistical model paradigm pay attention to spatially intrinsic characteristics. The most popular stochastic model is the Markov Random Field (MRF) model [14]. The MRF model, and its variants, have been successfully used for brain MRI segmentation [138]. Ruan et al. proposed a fuzzy Markovian method for brain tissue segmentation from magnetic resonance images that calculated a fuzzy membership value for each pixel to indicate the partial volume degree [125].

In unsupervised statistical segmentation techniques the number of class labels and the model parameters are assumed to be unknown. Hence, estimations of model parameters and image labels (where each group of pixels characterizing a region of interest can be represented as an image label) are required simultaneously. Since, the image label estimation depends upon the optimal set of parameters, the segmentation problem can be viewed as an incomplete data problem. To handle this problem, an iterative scheme, the Expectation-Maximization (EM) algorithm, has been proposed [103]. Zhang et al. proposed a Hidden Markov Random Field (HMRF) model to achieve brain MRI segmentation in the case of unsupervised statistical techniques [169]. A new Bayesian method for automatic segmentation of brain MRI was proposed in [99] where a variant of the EM algorithm was used so as to make the whole procedure more computationally efficient.

2.3.4 MRI Segmentation Using Fuzzy Based Approaches

Fuzzy based segmentation approaches are increasing popular because of recent developments in fuzzy set theory, the development of various fuzzy set based mathematical modelling techniques, and its successful application in computer vision systems [139]. Ahmed et al. proposed a bias correction Fuzzy C-Means (FCM) segmentation algorithm in which they incorporated a neighbourhood regularizer into the FCM objective function to allow labelling of a pixel to be influenced by the labels in its immediate neighbourhood [3]. The algorithm is realized by incorporating the spatial neighbourhood information into the standard FCM algorithm and modifying the membership weighting of each cluster. Siyal et al. presented a modified FCM algorithm formulated by modifying the objective func-

tion of the standard FCM and used a special spread method for the classification of tissues [137]. Wang et al. proposed a modified FCM algorithm, called mFCM, for brain MR image segmentation [148]. Aboulella et al. proposed a statistical feature extraction technique for diagnosis of breast cancer mammograms by combining fuzzy image processing with rough set theory [65].

2.3.5 MRI Segmentation Using Graph Based Approaches

Recently, automatic segmentation of MR images of the developing new born brain has been addressed [118] using graph clustering and parameter estimation for finding the initial intensity distributions. Cocosco et al. [24] used sample selection through minimum spanning trees for intensity-based classification. Shi and Malik [135] proposed the Normalized Cuts (NCut) algorithm for image segmentation problems, which is based on Graph Theory. The NCut algorithm treats an image pixel as a node of a graph, and considers segmentation in terms of a graph partitioning problem. A variation of this technique is used in the context of the work described later in this thesis. Chapter 3 gives more details of this image segmentation technique.

Note that with respect to the work described in this thesis a necessary pre-cursor for the desired classification is the identification (segmentation) of the corpus callosum, although the precise nature of the adopted segmentation algorithm is not significant as its is the relative performance of the proposed representations with respect the classification task that is central to the investigation described in this thesis.

2.4 Data Mining and Image Classification

Data Mining is broadly concerned with the identification of hidden patterns in data. Data mining is part of a super process called Knowledge Discovery in Databases (KDD) [47, 48, 64] and draws on the fields of statistics, machine learning, pattern recognition and database management. Data mining first came to prominence in the early 1990s when advances in computing technology meant that large quantities of data could be effectively collected and stored, and subsequently analysed. However, we can trace the origins of data mining to much earlier work on machine learning and statistics [18, 48, 49, 64]. Originally data

mining was directed at tabular data, for example the analysis of the contents of super market baskets [2]. Subsequently data mining techniques have been applied to many forms of data: examples include text [17], web usage logs [168], and graphs [72]. The work described in this thesis is directed at image mining.

Image mining deals with the extraction of implicit knowledge from image collections, for example image relationships or patterns, that are not explicitly stored in the image database. The main challenge of image mining is concerned with mechanisms to translate low-level pixel representations into high-level image objects that can be efficiently mined [69, 110, 114, 166, 167].

One branch of data mining, and that of interest with respect to the work described here, is classification (also sometimes referred to as categorisation). Classification is concerned with software systems to generate a "classifier" which can then be used to classify (categorise) new data. Typically a pre-labelled training set is used to build the classifier. Thus, in the case of the corpus callosum application; the training data will comprise a set of labelled images, (say) epilepsy and non-epilepsy, from which a classifier can be generated. This classifier can then be used to categorise new images. Thus the goal of image classification is to build a model that will be able to predict accurately the class of new, unseen images. There are various techniques that can be used to generate the desired classifier, examples include decision trees [122], Support Vector Machines (SVMs) [28], KNN [100] and Naive Bayes [95].

For the work described in this thesis, the popular C4.5 decision tree classifier [122] and SVMs [28] were adopted in Chapters (see 5 and 6). A C4.5 decision tree consists of internal nodes that specify tests on individual input variables or attributes that split the data into smaller subsets, and a series of leaf nodes assigning a class to each of the observations in the resulting segments. The C4.5 algorithm builds decision trees using the concept of information entropy [122]. The entropy of a sample S of classified observations is given by:

$$Entropy(S) = -p_1\, log_2(p1) - p_0\, log_2(p0), \tag{2.1}$$

where $p_1(p_0)$ are the proportions of the class values 1(0) in the sample S, respectively. C4.5 examines the normalised information gain (entropy difference) that results from choosing an attribute for splitting the data. The attribute with the

highest normalised information gain is the one used to make the decision. The algorithm then recurs on the smaller subsets.

Support Vector Machines (SVMs) have proved to be a useful technique for data classification. SVMs have provided improvements in the fields of handwritten digit recognition, object recognition and text categorization, among others. SVMs try to map the original training data into a higher dimensional space by a kernel function Φ. The aim is then to identify a linear separating hyper-plane, with the maximal margin between negative and positive samples, in this higher dimensional space. From the mathematical point of view, given a training set of instance label pairs (x_i, y_i); $i = 1 \cdots l$, where $x_i \in R^n$ and $y \in \{-1, 1\}^l$, SVMs search for a solution to the following optimization problem:

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \\ &subject\ to: \ y_i(w^T \Phi(x_i) + b) \geq 1 - \xi_i; \ \xi_i > 0 \end{aligned} \tag{2.2}$$

Here training vectors x_i are mapped into the higher dimensional space by the function Φ. $C > 0$ is the penalty parameter of the error term. Furthermore, $K(x_i, x_j) \equiv \Phi(x_i)^T \Phi(x_j)$ is the kernel function. There are four basic kernel functions: linear, polynomial, radial basis and sigmoid functions. SVMs are able to deal with two-class problems, but there exists many strategies to allow SVMs work with a larger number of categories.

The automatic categorization of medical images is a challenging task that can be of importance when managing real data collections. More specifically, because of the increasing amount of medical image data available, locating relevant information in an efficient way presents a particular challenge. Moreover, the extraction of appropriate visual descriptors to reduce the gap between the semantic interpretation and the visual content of medical images remains a research issue [106].

There has been a substantial amount of work on the application of classification techniques within the medical domain. For example, Antonie et al. [161] distinguished mammography tumours from the normal mammography using a medical image classification approach. Many successful techniques for automatic categorization of medical images rely on the same kind of feature space descriptors as more general methods for generic image classification. For example, Avni

et al. [7] have successfully applied a Bag-of-Features (BoF) approach to the classification of X-Ray images. Their method uses histograms of vector quantized scale invariant features for automatic organ recognition. Their Scale Invariant Feature Transform (SIFT) technique computes a 128-bin gradient histogram as a description vector for each salient region [96]. They have also applied their method to discriminate between healthy and pathologic cases on a set of chest radiography images. SIFT descriptors and the BoF representation have also been successfully applied to binary classification problems related to endomicroscopic images, thus enabling users to discriminate between neoplastic and benign cases with high accuracy [5]. Another tool for feature extraction in medical image retrieval and classification is the Local Binary Patterns (LBP) descriptor [109]. This descriptor is based on a very simple idea to efficiently describe the local texture pattern around a pixel. LBP comprises a binary code that is obtained by thresholding a neighbourhood according to the grey value of its centre. The effectiveness of LBP for radiographic images has been recently demonstrated by Jeanne et al. [78], who have found significant performance improvement over other common visual descriptors. In the context of MRI scan classification, Yang et al. [157] proposed an automatic categorization of brain MRI scans based on Independent Component Analysis (ICA) coupled with an SVM technique to discriminate MRI scans among Alzheimer patients, mild cognitive impairment, and control subjects using whole brain images. In [83], a hybrid approach combines an SVM technique and Genetic Algorithm (GA) to classify tumour and normal tissue of brain MRI scans based on wavelet descriptors.

2.4.1 Case Based Reasoning

Case-Based Reasoning (CBR) is a well established Artificial Intelligence (AI). CBR operates by using a Case Base (CB) of previous cases to solve a new case. Essentially the features of the new case are compared with the features of the cases in the CB and the identified most similar cases are used to formulate a solution to the new case. Thus the basic idea behind CBR is to solve a current problem by reusing solutions that have been applied to similar problems in the past. At the highest level of generality, the CBR cycle may be described by four tasks [1]: (i) retrieve the most similar case or cases, (ii) reuse the information and

knowledge in the retrieved cases to solve the problem, (iii) generate a solution, and (iv) retain in the CB the parts of this experience likely to be useful for future problem solving. In order to do the retrieval, case representation and case indexing techniques are required so that it is possible to retrieve cases using a similarity computation. The retrieved cases are then reused to provide a possible solution to the given problem and therefore, a case adaptation mechanism is also required.

CBR is not normally thought of as a data mining technique. However, if we think of a case base (CB) as comprising a set of examples each with an associated label then CBR can be used for classification purposes. CBR style systems have been applied successfully in the context of classification. For example Nearest-Neighbour Classification (NNC) systems [100] can be characterised as a simple CBR technique. As noted above the goal of classification is to predict the class membership of given entities. The basic idea of NNC is to use information about records (cases) for which the class membership is already known. In order to classify a new record (case), its description has to be compared to the descriptions of the known records. From an abstract point of view, each record can be characterised as a point in some problem space defined by the properties used to describe the record (see Figure 2.3). To predict the class of the new record its nearest neighbours within the problem space is identified using some distance metric. Finally, the information about the class membership of these nearest neighbours is used to predict the class the new record. Either the class of the actual nearest neighbour may be used or a weighted voting system may be adopted where k nearest neighbours have been identified. In the example shown in Figure 2.3 the entities belong either to the class "+" or "-". Using a 5-NNC the prediction for the shown query would be that it belongs to class "+", because 3 of the 5 nearest neighbours belong to this class. From the CBR point of view it is obvious that this approach requires no sophisticated adaptation methods as long as the number of cases exceeds the number of possible classes significantly. Excellent reviews of the foundations of CBR techniques, systems and tools can be found in [1] and [136].

In the context of CBR based image classification the main issue is the similarity matching technique to be adopted; given a case to be classified how best

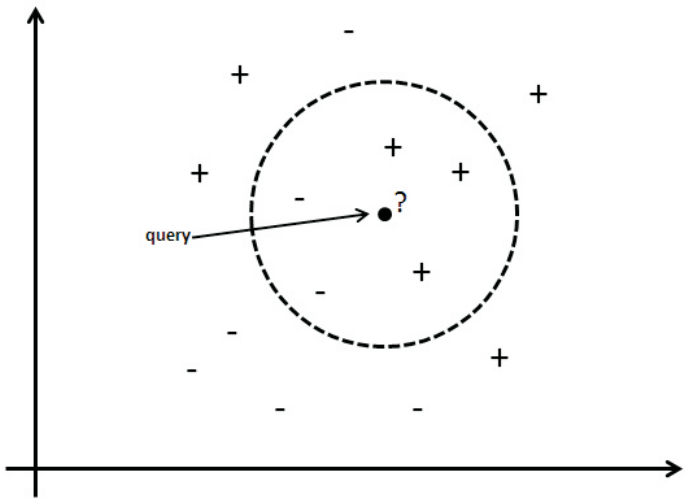

Figure 2.3: An example of 5-Nearest Neighbour Classification.

can this be matched with cases in the CB. The traditional measures used to compare cases in CBR are distance measures (assuming of course that we are only considering numerical attributes). Popular examples of such distance measures are: Minkowsky distance, Hausdroff Distance, Bottleneck Distance and Reflection Distance [143]. In the case of the work described here a Dynamic Time Warping (DTW) technique has been adopted, this is described further in Chapter 7.

2.4.2 Graph Mining

Graph mining is the process of discovering hidden patterns (frequent subgraphs) within graph datasets. From the literature graph mining can be categorized in terms of transaction graph mining and single graph mining. In transaction graph mining the dataset to be mined comprises a collection of small graphs (transactions). The goal is to discover frequent recurring subgraphs across the dataset. In single graph mining the input of the mining task is one single large graph, and the objective is to find frequent subgraphs which occur within this single graph. Frequent Subgraph Mining (FSM) has demonstrated its advantages with respect to various tasks such as chemical compound analysis [73], document image clustering [13], graph indexing [156], etc.

The straightforward idea behind FSM is to "grow" candidate subgraphs in either a Breadth First Search (BFS) or Depth First Search (DFS) manner (can-

didate generation), and then determine if the identified candidate subgraphs occur frequently enough in the graph data for them to be considered interesting (support counting). The two main research issues in FSM are thus how to efficiently and effectively: (i) generate the candidate frequent subgraphs and (ii) determine the frequency of occurrence of the generated subgraphs. Effective candidate subgraph generation requires that the generation of duplicate or superfluous candidates is avoided. Occurrence counting requires repeated comparison of candidate subgraphs with subgraphs in the input data, a process known as subgraph isomorphism checking. FSM, in many respects, can be viewed as an extension of Frequent Itemset Mining (FIM) popularised in the context of Association Rule Mining (ARM). Consequently, many of the proposed solutions to addressing the main research issues effecting FSM are based on similar techniques found in the domain of FIM.

It is widely accepted that FSM techniques can be divided into two categories: (i) the Apriori-based approach (also called the BFS strategy based approach) and (ii) the pattern growth approach. These two categories are similar in spirit to counterparts found in ARM, namely the Apriori algorithm [2] and the FP-growth algorithm [63] respectively. The Apriori-based approach proceeds in a "generate-and-test" manner using a BFS strategy to explore the subgraph lattice of the given database. Therefore, before exploring any $(k+1)$-subgraphs, all the k-subgraphs should first be explored. For each discovered subgraph g, this approach extends g recursively until all the frequent supergraphs of g are discovered [62]. Pattern growth approaches can use both BFS and DFS strategies, but the latter is preferable to the former because it requires less memory usage.

One of the main challenges associate with FIM and FSM is the substantial number of patterns which can be mined from the underlying database. This problem is particularly important in the case of graphs since the size of the output can be extremely large.

The significance of FSM with respect to the work described here is that one of the techniques proposed uses this technique for the purpose of generating a feature space. The application of FSM algorithms to the datasets described in this work entails a significant computational overhead because of the great number of generated frequent subgraphs. To reduce this overhead a Weighted

FSM approach can be applied, the objective being to focus on the identification of those frequent subgraphs that are likely to be the most significant according to some weighting scheme.

2.5 Image Representation

As noted above, image representation is an important issue with respect to image classification. It is currently not computationally possible to present images to a classification algorithm in their entirety, consequently it is necessary to process the images in such a way that the classification process is tractable while at the same time minimising information loss. Thus, one of the fundamental challenges of image mining is to determine how the low-level pixel representation contained in a raw image or image sequence can be efficiently and effectively processed to identify high-level spatial ROIs and their relationships. Image classification techniques use visual contents such as colour, texture, and shape to represent and classify images. Colour and texture have been explored more thoroughly than shape. Because shape is a more intrinsic property of ROIs than colour or texture, and given the considerable evidence that natural ROIs are recognized based primarily on their shape [107], the increasing interest in using the shape features of ROIs for image classification is not surprising. The focus of this research is therefore on shape-based image classification. Since humans can often recognize the characteristics of ROIs using only their shapes, it is possible to expect shape-based techniques to be intuitive as a tool for classifying images. However, image classification by shape is still considered to be a more intrinsically difficult task compared to image classification based on other visual features [107]. In addition, the problem of shape-based image classification becomes more complex when the extracted ROIs are corrupted by occlusions or noise as a result of the image segmentation process.

Deriving shape descriptions is an important task in image classification. Once the ROI has been acquired in the image, a set of techniques can be applied in order to extract information from its shape, so that it can be analysed further. This process is called shape description and generally results in a feature vector (a shape descriptor). Shape description can be viewed as a mapping from a shape space to a vector space that satisfies the requirement that two similar

shapes will also have close-to-identical shape descriptors while still allowing for discrimination between different shapes. Note that it is not a requirement that the original shape can be reconstructed from the feature vector. In most studies, the terms shape representation and description are used interchangeably. Since some of the representation methods are used as shape descriptors, there is no well-defined separation between shape representation techniques and shape description techniques. However, shape representation and description methods are defined in [93] as follows. Shape representation results in non-numeric values of the original shape. Shape description results in numeric values and is a step subsequent to shape representation. For the sake of simplicity, we consider representation and description to be synonymous and, throughout the rest of this section, collectively refer to such techniques as shape description methods.

During the last decade, significant progress has been made with respect to both the theoretical and practical aspects of shape description, and the literature reports a variety of techniques directed at describing ROIs based on their shape. Two main approaches to deriving shape descriptors are region-based and boundary-based (also known as the contour-based approach). In the region-based approach, all the pixels defining a ROI are used in order to obtain the desired shape descriptors (vectors). On the other hand, the boundary-based descriptor approach uses only the boundary of a ROI to extract its shape descriptor. The two approaches are considered in more detail in subsection 2.5.1 and 2.5.2 below. The representations proposed in this thesis include both region based and boundary based descriptions.

The problem of shape analysis has been pursued by many authors, thus, resulting in a great amount of research. Recent review papers [93, 165] as well as books [29, 41] provide a good overviews of the subject.

2.5.1 Region Based Methods

Region based shape descriptors express the pixel distribution within a 2D ROI. They can be used to describe complex ROIs consisting of multiple disconnected regions as well as simple ROIs with or without holes. Since it is based on the regional property of a ROI, the descriptor is insensitive to noise that may be introduced inevitably in the process of segmentation. From the literature we can

identify a number of techniques for generating region based shape descriptors, examples include: (i) moments, (ii) simple scalar descriptors, (iii) angular radial transformation descriptors, (iv) generic Fourier descriptors, (v) grid descriptors, (vi) shape decomposition and (vii) medial axis transforms. A number of these techniques form the foundation for some of the representations formulated later in this thesis. Others have been included for consideration in this sub-section purely for comparison purposes or because they are of historical interest.

2.5.1.1 Moments

Moments are widely used as shape descriptors. These moments can be classified as being either orthogonal or non-orthogonal according to the basis used to derive them. The most popular type of moments are called geometric moments which are non-orthogonal moments that use a power basis $(x^p\, y^q)$. The geometric moment m_{pq} of order $p+q$ of an $(N \times M)$ digital image $f(x,y)$ is defined as:

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q f(x,y) \tag{2.3}$$

The use of moments for shape description was initiated in [70], where it was proved that moment-based shape description is information preserving. The zeroth order moment m_{00} is equal to the shape area assuming that $f(x,y)$ is a silhouette function with value one within the shape and zero outside the shape. First order moments can be used to compute the coordinates of the center of mass as $x_c = m_{10}/m_{00}$ and $y_c = m_{01}/m_{00}$. Based on the moments formulated in equation 2.3, a number of functions can be defined that are invariant under certain transformations such as translation, scaling and rotation. The moment invariants are then put into a feature vector. Global object features such as area, circularity, eccentricity, compactness, major axis orientation, Euler number and algebraic moments can all be used for shape description [120].

Central moments are constructed by subtracting the centroid from all the coordinates. These moments are represented as:

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - x_c)^p (y - y_c)^q f(x,y) \tag{2.4}$$

The disadvantage of using these moments is that the basis used to derive these moments is not orthogonal, so these moments suffer from a high degree of in-

formation redundancy. A generalization of moment transforms to other basis is also possible by replacing a conventional basis $x^p\, y^q$ by a form $P_p(x)\, P_q(y)$, for example orthogonal polynomials. In this case, the moments produce minimal information redundancy, which is important for optimal utilization of the information available in a given number of moments [93]. Some of the orthogonal polynomial systems include Legendre and Zernike polynomials [141]. Teague [141] adopt Zernike orthogonal polynomials to derive a set of invariant moments, called Zernike moments. Zernike moments are defined as the projections of $f(x, y)$ on complex Zernike polynomials which form a complete orthogonal set over only the interior of the unit circle; that is, $x^2 + y^2 \leq 1$. The function of complex Zernike moments with an order p and repetition q in polar coordinates is defined as:

$$ZM_{pq} = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 f(r,\theta).V_{pq}^*(r,\theta).\, r\, dr\, d\theta \tag{2.5}$$

$V_{pq}(r, \theta)$ is the Zernike polynomial, defined as:

$$V_{pq}(r,\theta) = R_{pq}(r).e^{iq\theta}, \quad i = \sqrt{-1} \tag{2.6}$$

where p is a positive integer or zero and q is a positive integer subject to the constraints that $p - |q|$ is even and $q \leq p$, r is the length of the vector from the origin to the (x, y) pixel; i.e.,$r = \sqrt{x^2 + y^2}$, θ is the angle between the vector r and the x axis in a counter-clockwise direction, the symbol $(*)$ denotes the complex conjugate, and $R_{pq}(r)$ is a real-valued Zernike-radial polynomial defined as follows:

$$R_{pq}(r) = \sum_{k=0}^{p-|q|/2} (-1)^k . \frac{(p-k)!}{k! \left(\frac{p+|q|}{2} - k\right)! \left(\frac{p-|q|}{2} - k\right)!} . r^{p-2k} \tag{2.7}$$

Many researchers have adopted Zernike moments for many applications, such as character recognition [15], face recognition [10], and shape retrieval [81]. Zernike moment invariants have been shown to outperform several other shape descriptors and to be highly effective in terms of image representation [85, 164]. Teh and Chin [123] evaluated the performance of several moments with respect to issues such as representation ability, noise sensitivity, and information redundancy. In terms of overall performance, Teh and Chin showed that Zernike moments outperform geometrical and Legendre moments. Further details about Zernike moments will

presented in Chapter 6 as this is the foundation of one of the representations considered in this thesis.

2.5.1.2 Simple Scalar Descriptors

A shape can be described using scalar measures based on its simple geometric characteristics. These simple descriptors of a shape can often discriminate only between shapes with large dissimilarities; therefore, they are usually used in the context of image classification to act as filters to eliminate false hits or are combined with other techniques to differentiate shapes. The advantage of these descriptors is that they have a physical meaning. A brief description of the most commonly used simple global descriptors is given below.

Eccentricity: Eccentricity (E) has been widely used as a scalar descriptor. It illustrates the way in which the pixels of a ROI are scattered around the centre of the ROI. E is defined as the ratio of the major axis of the ROI to the minor axis. It is calculated using central moments such that [153]:

$$E = \frac{\mu_{20} + \mu_{02} + \sqrt{\mu_{20}^2 + \mu_{02}^2 - 2\mu_{02}\mu_{20} + 4\mu_{11}^2}}{\mu_{20} + \mu_{02} - \sqrt{\mu_{20}^2 + \mu_{02}^2 - 2\mu_{02}\mu_{20} + 4\mu_{11}^2}} \tag{2.8}$$

where μ_{pq} are the central moments of order $(p + q)$.

Solidity: Solidity (S) is computed as the ratio between the ROI area A_0 and the area of the corresponding convex hull. Solidity is computed as follows [153]:

$$S = \frac{A_0}{Convex\ Area} \tag{2.9}$$

Extent (Rectangularity): Extent (EX) is a measure that reflects the rectangularity of a ROI [21], in other words, how much of a ROI fills its minimum enclosing rectangle (MER). Extent is defined as:

$$EX = \frac{A_0}{A_{MER}} \tag{2.10}$$

where A_{MER} is the area of the ROI's MER.

Circularity (Compactness): Circularity (C) is a measure of roundness and is defined as follows:

$$C = \frac{4\pi * A_0}{(P_0)^2} \tag{2.11}$$

where A_0 is the ROI area, and P_0 is the ROI perimeter.

It is sometimes used as inverse formula $C = \frac{(P_0)^2}{4\pi * A_0}$.

Euler's Number: Euler's Number is the difference between the number of contiguous parts and the number of holes in the ROI.

Elongatedness: Elongatedness is the ratio between the length and width of the region bounding rectangle.

2.5.1.3 Angular Radial Transform Descriptors

For region-based shape descriptions, the MPEG-7 working group has selected angular radial transform descriptors [19] as MPEG-7 region based shape descriptor. Angular radial transform (ART) descriptors are similar to Zernike moments descriptors. The main difference is that ART descriptors are based on cosine functions rather than on Zernike polynomials. The ART basis functions are separable along the angular and radial directions and are defined on the unit circle as follows [19]:

$$V_{pq}(x, y) = V_{pq}(r, \theta) = \frac{1}{2\pi} R_p(r) e^{iq\theta} \tag{2.12}$$

The complex ART descriptors of order p with repetition q for a continuous function $f(x, y)$ are defined in polar coordinates as:

$$ART_{pq} = \frac{1}{2\pi} \int_0^{\pi} \int_0^1 f(r, \theta).R_p(r).e^{-iq\theta}\, r\, dr\, d\theta \tag{2.13}$$

where

$$R_p(r) = \begin{cases} 1 & p = 0 \\ 2cos(p\pi r) & p \neq 0 \end{cases} \tag{2.14}$$

For a digital image, the integrals are replaced by summations to obtain:

$$ART_{pq} = \frac{1}{2\pi} \sum_r \sum_\theta f(rcos\theta, rsin\theta).R_p(r).e^{-iq\theta}\, r\, dr\, d\theta \quad , \quad r \leq 1 \tag{2.15}$$

One of the advantages of the ART method is the compactness and efficiency in describing the properties of multiple disjoint regions, simultaneously. It describes the ROIs that may be split into disconnected sub-regions in the segmentation process. It has been reported that Zernike moments outperform angular radial transform descriptors in classifying ROIs that can be classified based on their

contours, but that ART descriptors perform better in the case of complex ROIs, i.e., ROIs that consist of multiple contours, which can be classified only by region-based techniques [101].

2.5.1.4 Generic Fourier Descriptors

A one-dimensional Fourier transform has been used successfully in image analysis to derive shape descriptors from shape contours. Zhang and Lu [163] adopted the 2D Fourier transform to derive a set of invariant descriptors called Generic Fourier descriptors (GFD). In this method, the 2D ROI is transformed to the polar coordinates, and the 2D Fourier transform is then applied in order to derive shape descriptors. The rotation of the original ROI corresponds to a shift in the normal Cartesian coordinates. This property makes it possible to derive rotation invariant features by applying the 2D Fourier transform on the polar image and then taking the magnitude while ignoring the phase. The 2D Fourier transform of a polar image is give by [163]:

$$GFD(\rho,\phi) = \sum_{r}\sum_{i} f(r,\theta_i) e^{-j2\pi\left(\frac{r}{R}\rho + \frac{2\pi i}{T}\phi\right)} \tag{2.16}$$

where $0 \leq r = \sqrt{(x-x_c)^2 + (y-y_c)^2} < R$ and $\theta_i = i(2\pi/T), (0 \leq i < T)$; (x_c, y_c) is the center of mass of the ROI; $0 \leq \rho < R, 0 \leq \phi < T$. R and T are the radial and angular resolutions.

The experiments conducted by Zhang and Lu [163] showed that GFD outperformed the use of ZM in retrieving ROIs based on the content of a region. Li and Lee [91] proposed a technique which utilizes the magnitude of the Fourier transform as well as phase coefficients in order to derive a set of invariant features.

2.5.1.5 Fourier-Mellin Descriptors

The Fourier-Mellin transform takes advantage of the properties of the Fourier and Mellin transforms in order to define a new set of image invariants called Fourier-Mellin Descriptors (FMDs). The Fourier-Mellin transform for $f(r,\theta)$ is defined as follows [133]:

$$FM_{sl} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{0}^{\infty} r^{s-1}.f(r,\theta).e^{-jl\theta}\, r\, dr\, d\theta \tag{2.17}$$

This transformation shows a radial-Mellin transform with parameter s and an explicit circular-Fourier transform with parameter l [133]. The rotation invariant can be achieved by taking the magnitude of the transform. Translation invariance is conventionally achieved by placing the origin of the coordinates at the image centroid.

Sheng and Duvernoy [133] employed the Mellin transform and the angular Fourier complex componential factor to drive invariant Fourier-Mellin descriptors. The problem associated with Fourier-Mellin descriptors is that the kernel function of the Mellin transform is not orthogonal; hence, these invariants suffer from a high degree of information redundancy. To overcome this problem, Sheng and Shen [134] applied the Gram-Schmidt orthogonalization process to orthogonalize the kernel function of the Mellin transform and used the resultant orthogonal polynomial to derive a set of orthogonal Fourier-Mellin moments. Their experiments demonstrated that, for small images, the description produced by orthogonal Fourier-Mellin moments is better than that produced by Zernike moments in terms of image-reconstruction errors and signal-to-noise ratio.

2.5.1.6 Grid Descriptors

The grid-based method has attracted interest because of its simplicity with respect to representations [97]. In this technique, the given ROI is overlaid by a grid space of a fixed size. A value of one is assigned to cells if they are at least 15% covered, and a zero is assigned to each of the other cells [97]. Then the 1s and 0s are scanned from left to right and top to bottom to obtain a binary sequence for the ROI. It is also evident that the smaller the cell size, the more accurate the representation of the ROI, but the greater the storage and computation requirements. It is evident that the representation is translation invariant, but it is not invariant to scale and rotation. Rotation and scale normalization are thus necessary when grid descriptors are used to compare two ROIs. To achieve scale normalization, all ROIs are scaled proportionally so that the major axes have the same fixed length. For rotation invariance, the given ROI is rotated so that its major axis is in parallel with the x-axis. The key problem associated with grid descriptors is the problematic major axis normalization. The major axis is sensitive to noise and can be unreliable even in the absence of noise effects. Moreover,

online classification usually involves extremely complex computations due to the high degree of dimensionality of the feature vectors.

2.5.1.7 Shape Decomposition

This approach assumes that shape description is a hierarchical process and decomposes a ROI into a union of primitives that comprise the simplest elements which form the region. Given the decomposition, a graph representation is constructed. Syntactic pattern recognition techniques are, then, applied. Shape decomposition methods generally use mathematical morphology. A morphological shape decomposition technique is proposed in [116] where the binary ROIs are decomposed into a union of simple regions. A structuring element (disk) is defined as the simplest ROI component and the image is analysed as a union of the disks. The representation is shown to be unique and invariant under rotation, translation and scaling. Another morphological decomposition method, which is called Morphological Signature Transform (MST) was developed in [94]. The MST method utilizes multiresolution morphological image processing by multiple structuring elements. The idea of this approach is to process decomposed, multiple shapes instead of the original one. The decomposed shapes, called signature shapes, contain substantial information about the whole shape. The method calculates the area of the shape signatures obtained from multiple structuring elements and multiple ROI scales to generate shape descriptors. Multiple structuring elements are obtained by rotating single or multiple structuring elements. One of the earliest types of shape decomposition of image data is the quad-tree [92]. It is a tree-like representation of image data, where the image is recursively divided into smaller regions. A quad-tree is a tree data structure in which each internal node has exactly four children. The quad-tree decomposition is based on a recursive subdivision of the image block into four quadrants according to the complexity of the block. If a subimage is not a homogeneous block, it is subdivided into four equal sized subimages again until all the subimages are homogeneous blocks. A subimage is called a homogeneous block if the grey-level of each of the pixels in the block is the same with respect to some specified constant. More details concerning the quad-tree representation coupled with weighted frequent subgraph mining is given in Chapter 5.

One of the problems in shape decomposition is the definition of the components. The result of the decomposition does not always correspond to a human intuitive shape representation. Furthermore, the results are not always unique. The decomposition of similar shapes may result in different elements. In many cases, the derivation of the decomposition is rather tedious and the computational complexity is relatively high compared to the other region based shape description methods such as ART and moments.

2.5.1.8 Medial Axis Transform

The idea of the Medial Axis Transform (MAT) is to represent the shape by a graph in which the important shape features are preserved [35]. The shape graph is based on the region skeleton and the first step is the skeleton construction. The skeleton is the set of all region points, which have the same minimum distance from the region boundary for at least two separate boundary points. Generally, MAT techniques use the region of the shape skeleton in order to derive shape descriptors. The idea behind obtaining the skeleton of the region of a shape is to preserve the topological information of the shape and to eliminate redundant information. In other words, the skeletonization extracts a region-based shape feature that represents the general form of a ROI. The shape's skeleton can be obtained by several techniques such as thinning algorithms, mathematical morphologic-based algorithms, and distance map-based algorithms [59]. After skeletonization, the skeleton is decomposed into parts and represented as a graph. Matching shapes then becomes a graph-matching problem. The difficulty with skeleton-based techniques is that a small amount of noise or a variation in the boundary often generates redundant skeleton branches that may seriously disturb the topology of the skeleton's graph [9]. Therefore, one of the weaknesses in the medial axis transform is its sensitivity to noise in the shape boundary.

2.5.2 Boundary Based Methods

Boundary based shape description methods exploit only shape boundary information. The shape properties of a ROI boundary are crucial to human perception in judging shape similarity and recognition. Many authors who have studied the human visual perception systems, agree on the significance of high curvature points

of the shape boundary in visual perception [6]. In the psychological experiments, it is suggested that corners have high information content and, for the purpose of shape description, corners are used as points of high curvature. Therefore, the shape boundary contains more information than the shape interior, in terms of perception.

The previous sections have provided a review of several descriptors that are based on the internal content of the shape. However, in many applications, the internal content of the shape is not as important as its boundary. Boundary-based techniques tend to be more efficient for handling shapes that are described according to their contours [101]. From the literature we can identify a number of techniques for generating boundary based shape descriptors, examples include: (i) polygonal approximation, (ii) Hough transform, (iii) stochastic representation, (iv) boundary approximation, (v) Fourier descriptors, (vi) coding, (vii) Curvature Scale Space Descriptors and (viii) simple boundary descriptors. Each is discussed in further detail below.

2.5.2.1 Polygon Approximation

Polygon approximation is one of the most popular shape representation methods, where the continuous shape is represented by a set of vertices of a polygon [93]. The main idea is to eliminate the redundant and insignificant shape features by reducing the number of boundary points. This is performed by searching the break points on the boundary, based on an approximation criteria, such as minimal error, minimal polygon perimeter or maximal internal polygon area. One of the most widely used polygonal approximation algorithm is a heuristic method called the Douglas-Peucker algorithm [37]. This iterative procedure repeatedly splits the curve into smaller and smaller curves until the maximum of the perpendicular distances of the points on the curve from the line segment is smaller than the error tolerance ϵ. In the approaches which use the split and merge algorithm [113], where the curve is split into segments until some acceptable error is reached, the split segments are merged if the resulting segment approximates the curve within some maximum error. In [88] curve evolution is used as a polygonal approximation method. Curve evolution reduces the influence of noise and simplifies the shapes by removing the irrelevant features, keeping relevant ones. In

this method, a digital curve is regarded as a polygon, where each boundary point is assumed to be a vertex, at the beginning of the evolution. In every evolution step, a pair of consecutive line segments is substituted with a line segment, joining the end points. The evolution process halts when the difference between the evolved polygon and the original curve is higher than a given threshold. Curve evolution methods assume that the shape boundary consists of various sources of distortion, which should be removed by smoothing. The process of smoothing depends on the predefined threshold value, which halts the evolution.
Corner detection is another approach for polygonal approximation. Popular techniques for corner detection use Wavelet Transform Modulus Maxima (WTMM) [121]. First, the shape boundary is represented by a 1D signal such as the contour's orientation profile. The singularities on that signal are then detected by determining the local maxima of the wavelet transform of the signal.

2.5.2.2 Hough Transform

The Hough Transform (HT) is a widespread technique in image analysis. Its main idea is to transform the image to a parameter space where clusters or particular configurations identify instances of a shape under detection. HT based techniques are used for shape detection, either parametrized or generalized. The HT was first introduced by Paul Hough in 1962 [67] with the aim of detecting alignments in T.V. lines. It became later the basis of a great number of image analysis applications. The HT is mainly used to detect parametric shapes in images. It was first used to detect straight lines and later extended to other parametric models such as circles or ellipses, being finally generalized to any parametric shape [12].

One major advantage offered by the HT is that it is highly robust against image noise and degradation. The HT is used for extracting *shape signatures* which can be used as a feature vector in the classification process. The HT is suitable for this task because it maintains the spatial information associated with an image ROI (feature). The "classic" HT performs a mapping between the X-Y image space onto a $\rho - \theta$ space. The transformation is $\rho = xcos\theta + ysin\theta$, where (x, y) are the coordinates of the pixel to be transformed, and (ρ, θ) are the parameters of a corresponding line passing through the pixel. The parameter

space is suitably quantized into bins and represented by means of an accumulator matrix initially set to zeros. Each pixel (x, y) can be conceptualised as a sinusoidal (parametric) curve in the new $\rho - \theta$ space,where θ varies from the minimum to the maximum values, giving the corresponding ρ values. The corresponding sinusoidal positions can be stored in an accumulator matrix A, where each row corresponds to one value of ρ, and each column to one value of θ. The cells in A are then incremented by the parametric curve. The local maxima within A then correspond to the dominant boundary lines of the ROI. Further details about the HT are presented in Chapter 4.

2.5.2.3 Stochastic Representation

stochastic representation methods are based on the stochastic modelling of a one dimensional function obtained from the shape boundary [93]. The idea is to interpret the one dimensional function as a stochastic process realization and use the model parameters obtained by estimation, as shape descriptors. For example the Autoregressive (AR) model proposed in [82] to represent the closed shape boundary. The one dimensional representation is obtained using a centroid distance function. The AR model is, then, characterized by a set of unknown parameters and an independent noise sequence. Given the function $u(s)$, the stochastic process is defined by:

$$u(s) = \alpha + \sum_{j=1}^{m} \theta_j u(s-j) + \sqrt{\beta}\omega_s \tag{2.18}$$

where θ_j is the AR coefficients to be estimated, α is a constant to be estimated, $\sqrt{\beta}$ is the variance of prediction noise and ω_s is the independent random noise source. $\theta_1, ..., \theta_m, \alpha, \beta$ are called the model parameters. The shape descriptor vector is constructed using these parameters estimated by the Maximum Likelihood Method. The same AR model is used in [39] with additional methods for improving classification performance. In [33], the AR model is extended to the bivariate case and the classification results are further improved. The disadvantage of the AR model is the sensitivity to shape occlusion [66]. The reason for this drawback is that it models the whole shape with only one set of predictive parameters. If the shape contains a large number of sample points and the contour varies radically, the shape may seem unpredictable. Therefore, an AR model with a finite

number of parameters is not adequate for the whole shape. In order to overcome this drawback, the AR model is combined with the Hidden Markov Model in [66]. This approach segments closed shapes into segments, describes each segment by AR modeling and finally analyzes the resulting vectors using a Hidden Markov Model.

2.5.2.4 Boundary Approximation

Boundary approximation methods represent the curves by piecewise polynomial interpolation, which results in a set of smooth curves inflected at control points. Splines are used for the interpolation of functions to approximate shape segments. The power of splines comes from the approximation of a given function with a curve having the minimum average curvature. The main disadvantage of splines is that the local function value modification changes the complete spline representation [93]. For this reason, B-splines are constructed so that the local function value change does not spread to the rest of the intervals. Bsplines are piecewise polynomial curves, whose shape is closely related to their control polygon which is a chain of vertices giving a polynomial representation of a curve. If a control polygon vertex changes its position, a resulting change of the spline curve occur only in a small neighbourhood of that vertex. In [25], curve representation and matching is performed by B-splines. Another method, described in [26], use splines for ROI modelling and shape estimation. Shape preserving approximation based on splines is proposed in [68].

2.5.2.5 Fourier Descriptors

The Fourier representation decomposes a shape contour into its frequency components (Fourier descriptors) obtained via its Fourier transform. The Fourier transform is applied to the boundary function and the resulting coefficients are used for shape description. For a given closed curve, which in turn is represented by a one dimensional function $u(s)$, the discrete Fourier transform is defined by:

$$a_n = \frac{1}{N} \sum_{s=0}^{N-1} u(s) exp(-j2\pi ns/N) \tag{2.19}$$

The coefficients a_n, $n = 0, 1, ..., N-1$, are used to derive Fourier descriptors. In order to achieve translation and rotation invariance the phase information of

a_n is ignored and only the magnitude $|a_n|$ is used. The magnitudes are divided by the DC component $|a_0|$, to achieve the scale invariance. In order to obtain one dimensional function from the shape boundary, various methods have been suggested such as tangent angle [160] and centroid distance [162]. The "classical method" is to express the shape as a sequence of coordinates in the complex plane, namely $u(s) = x(s) + jy(s)$. Modified Fourier descriptors are used in [126] to efficiently compute the feature matching. The Fourier descriptors represent global information about the boundary. Therefore, local spatial information about the shape is not readily available and the level of shape detail can only be controlled on a global basis.

2.5.2.6 Coding

Coding methods describe a ROI by a sequence of unit size line segments with a given orientation. One of the most popular coding scheme is Freeman's chain code [51], which is essentially obtained by mapping the shape boundary into a 2D parameter space, which is made up of codes. The chain code is defined as the direction of the ROI's contour from a starting point. This representation is typically based on 4- or 8-connectivity, and the resulting chain code is a sequence of numbers. There are many variations of Freeman chain codes. A generalized chain code is proposed in [52], where the nodes surrounding a center node are enumerated counter-clockwise in ascending order from inside out.

2.5.2.7 Curvature Scale Space Descriptors

Scale space representations are based on tracking the position of inflection points in a boundary, filtered by low-pass Gaussian functions of variable widths [93]. The inflection points, remaining in the representation are expected to be significant object characteristics [93]. The Curvature Scale Space (CSS) approach proposed in [102] is the most popular one in this class of methods. It has been selected as the MPEG-7 contour based shape descriptor. Mathematically speaking, let Γ be a closed planar curve and u be the normalized arc length parameter on Γ:

$$\Gamma = \{(x(u), y(u)) \,|u \in [0, 1]\}\,. \tag{2.20}$$

In order to smooth the curve, functions x(u) and y(u) is convolved with a one dimensional Gaussian kernel at different levels:

$$X(u,\sigma) = x(u) * g(u,\sigma),$$
$$X(u,\sigma) = y(u) * g(u,\sigma), \tag{2.21}$$

$g(u,\sigma)$ is a Gaussian kernel of width σ. The curvature of the smoothed curve is then computed as:

$$k(u,\sigma) = \frac{\dot{X}(u,\sigma)\ddot{Y}(u,\sigma) - \ddot{X}(u,\sigma)\dot{Y}(u,\sigma)}{\left(\dot{X}^2(u,\sigma) + \dot{Y}^2(u,\sigma)\right)^{3/2}}, \tag{2.22}$$

where $\dot{X}$, $\dot{Y}$ and $\ddot{X}$, $\ddot{Y}$ are the first and second derivatives of $x(u,\sigma)$ and $y(u,\sigma)$. As σ increases, the shape of Γ_σ changes. This process of generating ordered sequences of curves is referred to as the evolution of Γ. During evolution, the locations of curvature zero crossings of every Γ_σ is determined. The points are then plotted in (u,σ) plane. The result of this process is represented as a binary image, called a CSS image of the curve. Finally, shapes are described by the positions of their CSS contour maxima. These positions projected onto the simplified object contours give the positions of the mid points of the maximal convex/concave arcs, obtained during the curve evolution. Since the small contours on the CSS image represent some information about the existing noise on the actual object, those maxima which are lower than a threshold are discarded. The shape similarity measure between two shapes is computed by relating the positions of the maxima of the corresponding CSSs. The CSS image representation is invariant to scale, rotation and translation. It is also robust to significant non-rigid deformations and perspective deformations. The basic drawback of this representation is the difficulty in the determination of a threshold value in order to remove small contours in CSS image. This process results in a decreased resolution, depending on the predefined threshold value. Therefore, it requires procedures, which find the necessary level of detail or use empirical parameters for sufficient resolution of the shape contour.

2.5.2.8 Simple Boundary Descriptors

Boundary descriptors are mostly based on the geometric properties of the boundary. Because of the discrete character of digital images, all of them are sensitive

to image resolution. In the following, popular geometric descriptors are provided for the sake of completeness:

Bending Energy: Given a plane curve $\Gamma = (x(u), y(u))$ and its curvature function $C(u)$ is given by:

$$C(u) = \frac{\dot{X}(u)\ddot{Y}(u) - \ddot{X}(u)\dot{Y}(u)}{\left(\dot{X}^2(u) + \dot{Y}^2(u)\right)^{3/2}}, \tag{2.23}$$

where $\dot{X}$, $\dot{Y}$ and $\ddot{X}$, $\ddot{Y}$ are the first and second derivatives of $x(u)$ and $y(u)$. The bending energy is defined as:

$$E = \frac{1}{N}\sum_{u=1}^{N} C(u)^2 \tag{2.24}$$

Centroid Distance: The centroid distance function is expressed by the distance of the boundary points from the centroid (x_c, y_c) of the ROI:

$$D_{centroid}(s) = \sqrt{(x(s) - x_c)^2 + (y(s) - y_c)^2} \tag{2.25}$$

Convexity(CX): A convex hull is the minimal covering of an ROI. A ROI convexity can be defined as the ratio of perimeters of the convex hull to that of the original contour (P_0). The convexity is represented as [115]:

$$CX = \frac{\text{Convex perimeter}}{P_0} \tag{2.26}$$

Aspect Ratio (AR): The aspect ratio is defined as the height of the ROI divided by its width and is expressed as:

$$AR = \frac{\text{Height of the ROI}}{\text{Width of the ROI}} \tag{2.27}$$

2.6 Summary

This chapter has presented the background and previous work with respect to the research described in this thesis. Image classification, as a sub-domain of image mining, and the challenges of medical image categorization were described. In the context of image pre-processing prior to the application of image classification, several image segmentation techniques that may be applied to medical

images were introduced. A review of the literature concerned with ROIBIC was presented. This review indicated that the number of types of shape descriptors for image classification has been rapidly increasing over the last decade; however, each technique has a number of shortcomings. In the next chapter the medical brain MRI scan datasets that were used for the performance evaluation of the proposed techniques are described. The process of preparing these images for image classification, including image registration and image segmentation, will also be presented.

Chapter 3

MRI Datasets: Preprocessing and Segmentation

3.1 Introduction

There have been major advances in the field of medical imaging over the past two decades. New medical imaging technologies have provided physicians with powerful, non-invasive techniques to analyse the structure, function, and pathology of the human body. In more recent years the improvement in (and the development of) many image acquisition techniques, the enhancement of the general quality of the acquired images, advances in image processing and the development of large computational capacities; have considerably enhanced the analysis task. The acquisition of medical images in 2D or 3D has become a routine task for clinical research applications. Image acquisition techniques include Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET) and functional MRI (fMRI). The corresponding increase in the quantity and detail of the available imagery has highlighted the need for more efficient and more effective computer automated, or semi-automated, techniques to support the interpretation of this data in order to provide better diagnosis and treatment options.

The analysis of the increasingly detailed amount of information available constitutes a great challenge for the medical imaging community, and requires significant innovations in all aspect of image processing. One particular challenge is related to the delineations of anatomical structures from medical images, this is seen as a critical step for many clinical and research applications. As noted in the introduction to this thesis the anatomical structure of interest is the corpus

callosum, a distinctive feature in MRI brain scans. Some background concerning this application domain is therefore presented in Section 3.2.

The automated recognition of meaningful image components, anatomical structures, and other regions of interest, such as the corpus callosum, is typically achieved using some kind of *registration* and *segmentation* techniques. Some background concerning the registration of brain MRI scans is therefore considered in Section 3.3 as well as mechanisms to align collections of MRI scans to some standard coordinates system. Some background concerning MRI brain scan segmentation techniques is then presented in Section 3.4. In this section a variation of the Normalized Cut segmentation algorithm is proposed, an additional contribution of this thesis, to identify the corpus callosum regions within brain MRI scans. The section commences with some details concerning the original "standard" normalized cuts algorithm, and the multiscale normalized cuts algorithm. The latter was introduced to overcome the computational complexity problem encountered when the standard normalized cuts algorithm is applied to large images. Although, the multiscale normalized cuts produces good results with respect to many image segmentation applications it was found not work efficiently with respect to the application domain considered in this thesis. Therefore, a variation of the multiscale normalized cuts was proposed to deal with this problem. This is fully described at the end of Section 3.4.

This chapter is concluded (Section 3.5) with some details concerning the image data sets used for evaluation purposes with respect to the work described in the reminder of this thesis.

3.2 Application Domain

MRI brain scans underpin the diagnosis and management of patients suffering from various neurological and psychiatric conditions. Analysis of MRI data relies on the expertise of specialists (radiologists) and is therefore subjective. The focus of the research work described in this thesis is brain MR images, and in particular a specific structure in these images called the corpus callosum which connects the two hemispheres of the brain. Figure 3.1 gives an example of a brain MRI scan, the corpus callosum is located in the centre of the image. The corpus callosum was selected to evaluate the work described because it is of interest to medical

researchers for a number of reasons. The size and shape of the corpus callosum has been shown to be correlated to sex, age, neurodegenerative diseases (e.g. epilepsy, multiple sclerosis and schizophrenia) and various lateralized behaviour in people (such as handedness). It is also conjectured that the size and shape of the corpus callosum reflects certain human characteristics (such as a mathematical or musical ability). Within neuroimaging research, considerable effort has been directed at quantifying parameters such as length, surface area and volume of structures in living adult brains, and investigating differences in these parameters between sample groups. Several studies indicate that the size and shape of the corpus callosum, in humans, is correlated to sex [4, 34, 128], age [128, 151], brain growth and degeneration [61, 98], handedness [31], epilepsy [27, 124, 149], special skills (e.g.musical ability) [111, 132] and brain disfunction [38, 74].

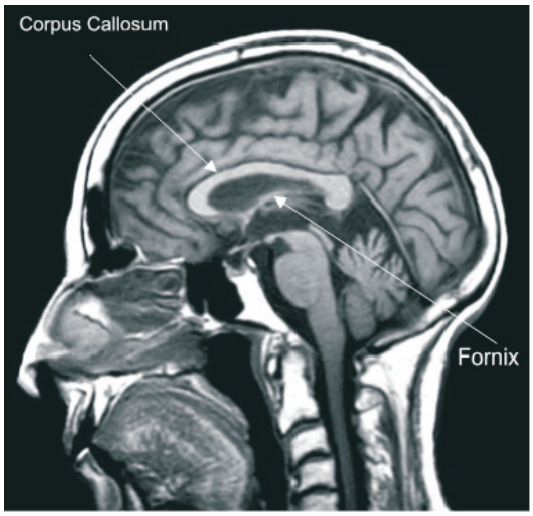

Figure 3.1: The location of the corpus callosum in a brain MR image.

3.3 Image Preprocessing and Registration

Each MRI scan comprised a sequence of "image slices", we refer to this as a *bundle*. The raw dataset used to evaluate the techniques described in this thesis consist of collections of MRI scan bundles. For our experiment we only required

the middle slice from each bundle. This is referred to as *midsagittal slice* and is the slice that separates the left and the right hemispheres of the brain.

It should be noted that as a part of the collection process, all slices in all bundles were aligned so that each bundle was centered on the same axes. The alignment (registration) was conducted manually by trained physicians using the Brain Voyager QX software package [56]. Figure 3.2 shows a typical MRI brain scan registered to a "standard" coordinate system using the Brain Voyager QX software package.

BrainVoyager uses several different coordinate systems: the internal axes, the standard Dicom and Talairach axes and the OpenGL axes. The Talairach transformation is at present the most widespread method for brain normalization and registration. The Talairach transformation is based on the 8 Talairach point landmarks that have to be specified: the anterior commissure (AC) and posterior commissure (PC) located on the midsagittal plane (MSP); and 6 cortical landmarks determining the extents of the brain in the anterior (A), posterior (P), left (L), right (R), superior (S), and inferior (I) directions. The AC is taken to be the origin of the coordinate system, the AC-PC line to be the y axis, the vertical line passing through the interhemispheric fissure to be the z-axis, and the line passing through the AC and at right angles to the y and z axes to be the x-axis. The three axes, along with a line parallel to the x-axis passing through the PC, divide the brain into 12 cubic rectangular regions. An appropriate translation parameters (in x, y and z directions) will be specified manually to determine these 8 points. The image bundle is then transformed using a trilinear interpolation. Once the image registration is completed, the midsagittal slice is extracted to delineate the ROI (the corpus callosum with respect to the work described here).

3.4 Brain MR Image Segmentation

As noted in the previous chapter the objective of image segmentation is to partition images into meaningful regions. The segmentation of medical MRI scans is a difficult task for a variety of reasons. Firstly segmentation algorithms tend to operate on the intensity or texture variations of a given image and are therefore sensitive to artifacts produced by the image acquisition process such as: image noise, image intensity inhomogeneity or non-uniformity, and partial volume av-

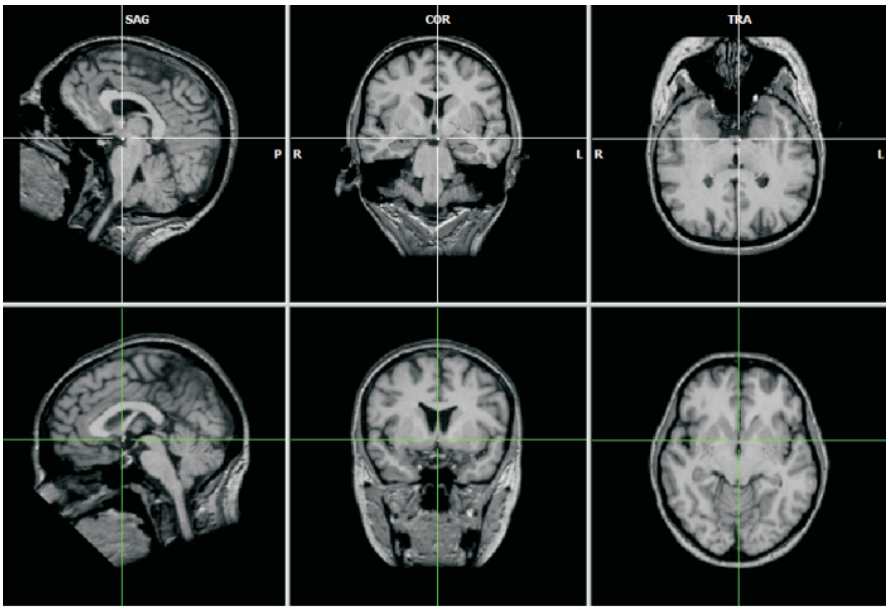

Figure 3.2: A typical brain MRI scan, (a) before the registration (first row scans), and (b) after the registration (second row scans).

eraging effect(as described in Chapter 2). Recall that if only a single tissue type is present in a voxel, the signal intensity will be characteristic of that tissue type. However, if more than one tissue type is present, the signal will be a combination of the contributions of the different tissues. This is known as the *partial volume effect*. Consequently, it blurs the intensity distinction between tissue classes at the border of the two tissues. In the case of brain MRI segmentation the partial volume effect makes it particularly difficult to accurately locate the boundaries of features such as the corpus callosum boundary. For example, if we consider the MRI brain scan given in Figure 3.1 the corpus callosum is located in the centre of the image. A related structure, the fornix, is also indicated. The fornix often blurs into the corpus callosum and thus presents a particular challenge in the context of the segmentation of these images so as to isolate the corpus callosum.

In this section, we discuss the segmentation of the midsagittal brain MRI slice to delineate the corpus callosum. To this end a variation of the Normalized Cuts (NCut) segmentation technique is proposed. NCut formulates segmentation as a graph-partitioning problem. The basic NCut algorithm, proposed by Shi and Malik [135], is presented in the Section 3.4.1. An illustration of the computation

of the optimal partitioning of a graph using NCuts, as proposed by Shi and Malik, is presented in Subsection 3.4.2. The basic NCuts algorithm does not operate well when applied to large images such as MRI brain scan images. An established enhancement to the basic NCuts algorithm, the multiscale normalized cuts algorithm proposed by Cour et al. [30], is therefore presented in Section 3.4.3. In the context of the corpus callosum application it was found that the multiscale normalized cuts algorithm could be improved upon so as to reduce the computational resource required to achieve the segmentation. A variation of the multiscale normalized cuts algorithm, proposed by the author of this thesis, is therefore presented in Section 3.4.4.

3.4.1 Normalized Cuts Criterion

In the image segmentation problem based on graph theory, an image is represented by an undirected graph $G = (V, E)$ where V is the set of nodes and E is the set of edges. A weighting matrix W is then defined to measures the similarity between nodes. Each node $v_i \in V$ corresponds to a locally extracted image pixel and each edge $e_i \in E$ connects pairs of nodes. The edges are determined by the proximity of pixels: each pair of pixels (nodes) is connected by an edge if they are located within a distance r from each other. An edge e_i has a weight proportional to the similarity of the properties of the connected nodes (e.g. pixel intensities). Given such a graph representation, image segmentation becomes equivalent to partitioning the nodes of the graph into disjoint sets. Formally, the graph $G = (V, E)$ can be partitioned into two disjoints sets of nodes V_1, V_2, where $V_1 \cup V_2 = V$ and $V_1 \cap V_2 = \phi$. The degree of dissimilarity between these two sets can be expressed as:

$$Cut(V_1, V_2) = \sum_{u \in V_1, v \in V_2} w(u, v). \tag{3.1}$$

where $w(u, v)$ is the similarity between nodes u and v. The weight of the edges that stay within the same set should be as high as possible and the weight of the edges that connect the two sets should be as low as possible. The optimal bipartitioning of a graph is one that minimizes this cut value. Finding the minimum cut is a well-studied problem and there exist efficient algorithms for solving it. Despite the merits of the graph cut formulation and its proposed algorithms, it is

biased towards producing cuts that contain a small number of nodes. Therefore Shi and Malik [135] proposed the normalized cuts algorithm (NCuts) as a new measure of disassociation, the normalized cut (NCut) is defined as follows:

$$NCut(V_1, V_2) = \frac{cut(V_1, V_2)}{SumCon(V_1, V)} + \frac{cut(V_1, V_2)}{SumCon(V_2, V)} \tag{3.2}$$

where $SumCon(V_1, V)$ is the total weights of the edges connecting nodes from a set V_1 to all nodes in the original set V and $SumCon(V_2, V)$ is the total weights of the edges connecting nodes from a set V_2 to all nodes in the original set V. The weight of the edges that stay within the same set are contained in $SumCon(V_1, V)$ and $SumCon(V_2, V)$, the weight of the edges that connect the two sets are contained in $cut(V_1, V_2)$. By minimizing this criterion the similarity across partitions is minimized and the similarity within partitions is maximized simultaneously. Unfortunately, minimizing this criterion constitutes an NP-complete problem but Shi and Malik showed that when the normalized cut problem is embedded in a real valued domain an approximate discrete solution can be found efficiently. The details of solving this minimization problem is described in next section.

3.4.2 Optimal Graph Partition Computation

Shi and Malik [135] developed an efficient computational technique based on a generalized eigenvalue problem to minimize the normalized cut. First, W is defined as the affinity matrix of the graph which determines the connectivity of the graph and defines the edges weights. Then the edges weights need to be determined based on the similarity and distance between pixels. This is done by applying the following formula for each edge:

$$w_{i,j} = \begin{cases} exp\frac{-\|I_i - I_j\|^2}{\sigma_I^2} \cdot exp\frac{-\|X_i - X_j\|^2}{\sigma_X^2} & if\ \|X_i - X_j\| \leq r \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

where: (i) $w_{i,j}$ is the weight of the edge that connects node i with node j, (ii) I_i is the intensity of pixel i, (iii) X_i is the spatial location of pixel i, (iv) I_j is the intensity of pixel j and (v) X_j the spatial location of pixel j. If there is no edge between node i and node j then $w_{i,j} = 0$ where the distance between node i and node j is greater than distance r. The effect of this procedure is that the pixels that are located very close to each other and that have a very small difference in

intensity get a high weight on the edge that connects them. The pixels that are located very far from each other and that have a very big difference in intensity get a low weight on the edge that connects them. All the weights $w_{i,j}$ are saved in a matrix W.

By using matrices Shi and Malik reformulated $NCut(V_1, V_2)$ to:

$$\frac{(\mathbf{1}+x)^T(\mathbf{D}-\mathbf{W})(\mathbf{1}+x)}{k\mathbf{1}^T\mathbf{D}\mathbf{1}} + \frac{(\mathbf{1}-x)^T(\mathbf{D}-\mathbf{W})(\mathbf{1}-x)}{(1-k)\mathbf{1}^T\mathbf{D}\mathbf{1}} \tag{3.4}$$

where x is an indicator vector, $x_i = 1$ if node i is in set V_1 and $x_i = -1$ if node i is in set V_2, $\mathbf{D}$ is a diagonal matrix with $d_i = \sum_j w_{i,j}$ on its diagonal and $k = \frac{\sum_{x_i=1} d_i}{\sum_i d_i}$. After some transformations, defining $b = k/(1-k)$ and setting $y = (\mathbf{1}+x) - b(\mathbf{1}-x)$, the normalized cut problem can be translated into the following formula:

$$min_x NCut(x) = min_y \frac{y^T(\mathbf{D}-\mathbf{W})y}{y^T\mathbf{D}y} \tag{3.5}$$

with the condition $y_i \in \{1, -b\}$ and $y^T\mathbf{D}\mathbf{1} = 0$. This formula is known as the Rayleigh quotient, if y is relaxed to take on real values, it can be minimized by solving the generalized eigenvalue problem:

$$(\mathbf{D}-\mathbf{W})y = \lambda\mathbf{D}y \tag{3.6}$$

The eigenvector corresponding to the second smallest eigenvalue is the real valued solution of the normalized cut problem. The approximate Lanczos method is often used to compute such an eigenvalue problem [58].

In the ideal case, the eigenvector components y_i corresponding to node i should only take on two discrete values. For example, if $y_i = 1$ or -1 then the signs tell us exactly how to partition the graph nodes V into V_1 and V_2 as follows: node i is in V_1 if the eigenvector components y_i, corresponding to node i, equals to 1 and node i is in V_2 if the eigenvector components y_i, corresponding to node i, equals to -1. i.e. the splitting point is 0 and the sign tells us exactly how to partition this graph ($V_1 = \{Vi|yi > 0\}, V_2 = \{Vi|yi <= 0\}$). However, eigenvector y is relaxed to take real values, therefore, a splitting point is needed to determine the two sets of nodes. To finally get the approximate discrete solution the eigenvector is split into two parts by mapping the real values to the discrete set of values over the interval $\{1, -b\}$. This mapping is done by sorting the real values of

this eigenvector and evaluating $NCut(V_1, V_2)$ for a few evenly spaced splitting points. The split point that produces the lowest $NCut(V_1, V_2)$ will then finally give us the approximate solution: the disjoint sets V_1 and V_2 together with their $NCut(V_1, V_2)$ value. The algorithm is recursively applied to every subgraph until the value of $NCut$ exceeds a certain threshold or the total number of nodes in the partition is smaller than a pre-specified threshold value.

From the above, the procedure for image segmentation based on normalized cuts can be summarized as follows:

1. Given an image I, construct an $N \times N$ symmetric similarity matrix $\mathbf{W}$ according to Equation 3.3 and $\mathbf{D}$ as a diagonal matrix with $d_i = \sum_j w_{i,j}$ on its diagonal.

2. Solve the generalized eigenvalue problem
   $(\mathbf{D} - \mathbf{W})\, y = \lambda \mathbf{D} y$
   and get the eigenvector with the second smallest eigenvalue.

3. Find the splitting point of eigenvector y, where $y_i \in \{1, -b\}$, that minimizes the $NCut$

4. Bipartition the graph nodes V into V_1 and V_2 according to this splitting point.

5. Repeat the bipartition recursively, stop if the $NCut$ value is larger than a pre-specified threshold value or the total number of nodes in the partition is smaller than a pre-specified threshold value.

The normalized cuts algorithm is computationally expensive when the dimension of the weight matrix is large because the pixel-based weight matrix required to compute the $N \times N$ weight matrix becomes very dense. Consequently the derivation of the eigenvalues becomes computationally expensive even though the approximate eigenvalue method and associated algorithm are designed to optimize this process.

3.4.3 Multiscale Normalized Cuts

Cour et al. [30] proposes a Normalized Cut adaptive technique, the Multiscale Normalized Cuts algorithm, that focuses on the computational problem created

by large graphs, which yields a better segmentation than the original Normalized Cuts. Cour et al. suggested the use of multiscale segmentations, decomposing a large graph into independent subgraphs. The main contribution of this technique is that larger images can be better segmented with linear complexity. In this technique, a graph representation of the image is again used as well as in the original normalized cuts segmentation approach described above. Basically this representation again comprises an undirected weighted graph $G = (V, E)$ where each node $v_i \in V$ corresponds to a locally extracted image pixels and the edges e_i in E connect pairs of nodes. The edges are determined by the proximity of pixels: each pair of pixels (nodes) is connected by an edge if they are located within a distance r from each other; also called the graph connection radius. A larger graph radius r generally makes segmentation better and facilitates the detection of objects described by faint contours against a cluttered background, but it is computationally expensive because the similarity matrix W becomes denser. In the multiscale normalised cuts algorithm, two cues are used to determine the connectivity of the graph; the intensity and position grouping cue and the intervening contours grouping cue. The edge weights can then be determined based on the similarity and distance between pixels. The intensity and position grouping cue of each edge is given by:

$$w_{IP}(i,j) = \begin{cases} exp\frac{-\|I_i - I_j\|^2}{\sigma_I^2} \cdot exp\frac{-\|X_i - X_j\|^2}{\sigma_X^2} & if\ \|X_i - X_j\| \leq r \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

where $w_{IP}(i,j)$ is the weight of the edge that connects node i with node j, I_i is the intensity of pixel i, X_i is the spatial location of pixel i, I_j is the intensity of pixel j and X_j is the spatial location of pixel j. If there is no edge between node i and node j then $w_{i,j} = 0$. This grouping cue, used separately; often gives bad segmentations because some natural images (including medical images) are affected by "texture clutter". The intervening contours grouping cue evaluates the affinity between two pixels by measuring the image edges between them. The measure of similarity regarding this grouping cue is given by[30]:

$$w_C(i,j) = exp - \left(\frac{max_{(x \in line(i,j))\varepsilon^2}}{\sigma_c} \right) \tag{3.8}$$

where $line(i,j)$ is a straight line joining pixels i and j and $\varepsilon = \|Edge(x)\|$ is the

image strength at location x. These two cues can be combined as shown by[30]:

$$w(i,j) = \sqrt{w_{IP}(i,j)\, w_C(i,j)} + w_C(i,j) \tag{3.9}$$

The multiscale normalized cut algorithm [30] works on multiple scales of the image so as to capture both coarse and fine levels of details. The construction of the image segmentation graph is given according to their spatial separation, as in:

$$W = W_1 + W_2 + ... + W_s, \tag{3.10}$$

where W represents the graph weights $w(i,j)$ and s the scale, i.e. W_s contains similarity measure between pixels with certain scale s. Two pixels i, j are connected only if $r_{s-1} < \|X_i - X_j\| \leq r_s$. The r value is a tradeoff between the computation cost and the segmentation result. The decomposition graph above can alleviate this situation. W_s can be compressed using recursive sub-sampling of the image pixels. Cour et al. determine representative pixels at each scale s as follows. First they sample representative pixels at $(2r+1)^{s-1}$ distance apart on the original image grid, and denote the representative pixels at each scale s by I_s and denote W_s^c as a compressed similarity matrix with connections between representative pixels in I_s, i.e. for the first graph scale, every pixel is a graph node and two nodes are connected with a graph edge if the two pixels are within r distance apart. They then construct W_1^c as a compressed similarity matrix with connections between representative pixels in I_1. For the second graph scale, they consider pixels at distance $2r+1$ apart in the original grid as representative nodes, I_2 and construct W_2^c as a compressed similarity matrix with connections between representative pixels in I_2. For the third graph scale, they consider pixels at distance $(2r+1)^2$ apart in the original grid as representative nodes, I_3 and construct W_3^c as a compressed similarity matrix with connections between representative pixels in I_3. In general, they consider pixels at distance $(2r+1)^{(s-1)}$ apart in the original grid as representative nodes, I_s. Then the full multiscale similirity matrix W can be expressed as:

$$W = \begin{pmatrix} W_1^c & & 0 \\ & \ddots & \\ 0 & & W_s^c \end{pmatrix} \tag{3.11}$$

Cour et al. reformulated the constrained multiscale normalized cut into the following formula:

$$\begin{aligned} maximize\ \varphi(X) &= \frac{1}{K}\sum_{l=1}^{K}\frac{X_l^T W X_l}{X_l^T D X_l} \\ subject\,to:\ CX &= 0 \end{aligned} \tag{3.12}$$

where X is an indicator vector s.t. $X \in \{0,1\}^{N^* \times K}$, $N^* = \sum_s N_s$, N_s is the number of representative pixels in I_s. The cross-scale constraint matrix C:

$$C = \begin{pmatrix} C_{1,2} & -I_2 & & 0 \\ & \ddots & \ddots & \\ 0 & & C_{S-1,S} & -I_S \end{pmatrix} \tag{3.13}$$

where $C_{s,s+1}(i,j) = \frac{1}{|N_i|} \quad if \quad j \in N_i$ and N_i is the neighbourhood which specifies the projection of $i \in I_{s+1}$ on the finer layer I_s.

The NP-complete constrained normalized cut is thus transformed into the following eigenvalue problem:

$$\bar{W}\bar{X} = \lambda\bar{X} \tag{3.14}$$

where:

$$\bar{W} = QD^{-1/2}WD^{-1/2}Q \tag{3.15}$$

and

$$Q = I - D^{-1/2}C^T(CD^{-1}CT)^{-1}CD^{-1/2} \tag{3.16}$$

The K first eigenvectors of $\bar{W}$ is used to find the indicator vectors X to segment the image into K partitions.

From the above, the image segmentation procedure based on multiscale normalized cuts can be summarized as follows:

1. Given an $m \times n$ image I, for $s = 1..S$ (S = number of scales):
   (a) Sample $\frac{m}{\rho} \times \frac{n}{\rho}$ pixels $i \in I_s$ from I_{s-1}, where ρ is the sampling factor.
   (b) Compute constraint $C_{s,s+1}(i,j) = \frac{1}{|N_i|},\ j \in N_i$ sampling neighbourhood of i.

(c) Compute the similarity matrix W_s^c for the representative pixels I_s.

2. Compute W, C from $(W_s, C_{s-1,s})$ as in (3.11),(3.13).

3. Solve the eigensystem $\bar{W}\bar{X} = \lambda\bar{X}$ as in (3.15), (3.16).

4. Use the first k eigenvectors of $\bar{W}$ to partition the image into k segments.

The multiscale normalized cuts algorithm has been shown to work well in many studies; however it was found that, in the context of the corpus callosum application, the algorithm did not work as well as expected with respect to all brain MRI scans in the test data sets. This was because: (i) MRI brain scans include a lot of noises, (ii) the intensity-level distributions between different soft tissues were not widely distributed and (iii) the complexity of tissue boundaries caused many pixels to contain mixtures of tissues. Although one can visually recognize the outline of the corpus callosum (Figure 3.1), portions of its boundary are indistinct, which can make it difficult to apply segmentation algorithms based on edge information alone. A further problem is that, quite often, intensity variations within the corpus callosum can be comparable or exceed the difference with the surrounding tissues. For these reasons it is desirable to enhance the contrast of the MR images. A variation of the multiscale normalized cuts algorithm was therefore developed. This is described in the following Subsection 3.4.4.

3.4.4 Proposed Approach

As noted above, the multiscale normalized cuts algorithm when applied to brain MRI scans to delineate the corpus callosum within these scans did not give satisfactory results with respect to all brain MRI scans in the test data. In the case of some scans the corpus callosum region can be clearly identified as shown in Figure 3.3 but blurred into other surrounding tissues in some other scans. Therefore a variation of the multiscale normalized cuts algorithm was developed that applied a threshold interval to extract objects with the same intensity values (such as the corpus callosum) during the application of the segmentation. This was found to give a much improved result.

The proposed variation of the multiscale normalized cuts algorithm was founded on the observation that the corpus callosum, which is located at the centre of the

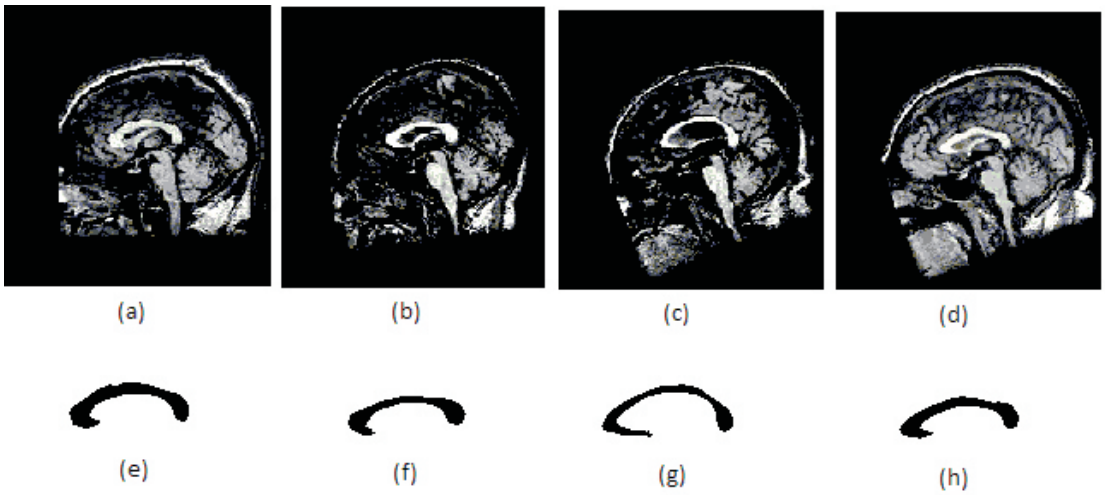

Figure 3.3: The results obtained by the multiscale normalized cuts algorithm, (a-d) the original brain MRI scans, (e-h) the corpus callosum obtained by the multiscale normalized cuts.

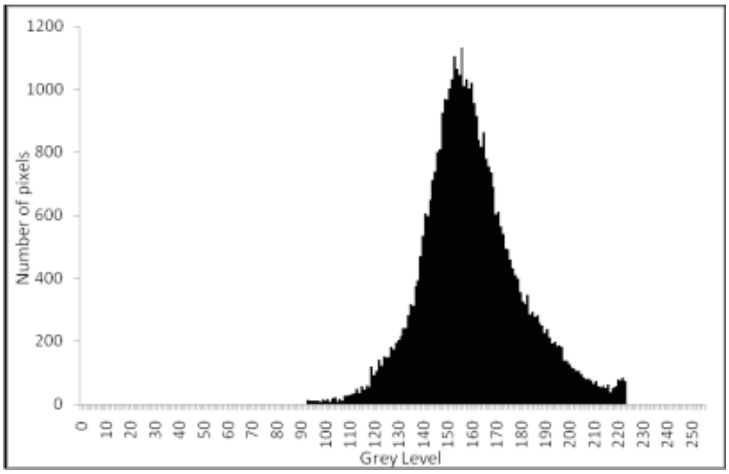

Figure 3.4: Histogram of the pixel greyscale values of the corpus callosum.

brain, comprises *white matter tissue* (i.e. the pixel represented corpus callosum have high intensity values). Figure 3.4 shows a pixel intensity value histogram of the corpus callosum derived from 30 selected MR images (256 grey levels were used) where the corpus callosum was very well defined and easy to detect using the multiscale normalized cuts algorithm. From the figure it can be seen that:

- The corpus callosum tends to have relatively high intensity values, and
- The distribution of intensity values seems to follow the normal distribution.

The latter is demonstrated in Figure 3.5 where it can be seen that the corpus callosum pixel values follow the normal distribution with mean $\bar{X} = 160$ and standard deviation $S = 20$. Figure 3.6a shows that with a threshold interval of $\bar{X} \pm S$, the corpus callosum can barely be recognized. With a threshold interval $\bar{X} \pm 2S$ a relatively distinct callosum shape is evident with a few other non-adjacent structures visible as shown in Figure 3.6b. With the threshold interval set at $\bar{X} \pm 3S$, the corpus callosum is clearly defined, although additional non-adjacent structures are also visible (Figure 3.6c). In Figure 3.6d, the corpus callosum starts to "blur" into the surrounding tissues using a threshold interval wider than $\bar{X} \pm 3S$. The significance here is that although the threshold values may differ depending on individual images, the high intensity property of the corpus callosum can be exploited to yield a segmentation algorithm that is both effective and efficient across the input image set. Therefore the interval $\bar{X} \pm 3S$ was chosen, so as to exclude intensity values outside the interval. This strategy was incorporated into multiscale normalized cuts algorithm and used to successfully extract the corpus callosum (and other incidental objects with the same intensity values).

The operation of the proposed image segmentation approach can be summarized as follows:

1. Apply the multiscale normalized cuts algorithm to the brain MRI scans datasets.
2. Calculate the sample mean $\bar{X}$ and standard deviation S corresponding to the fully segmented corpus callosum regions.

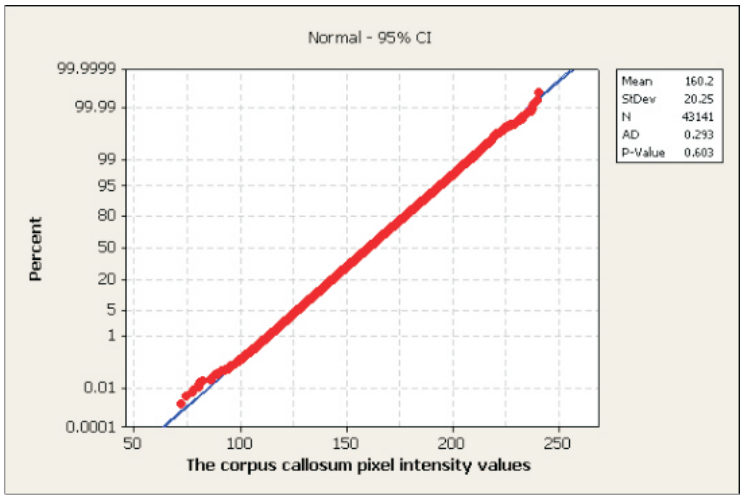

Figure 3.5: Probability plot of the corpus callosum pixel values.

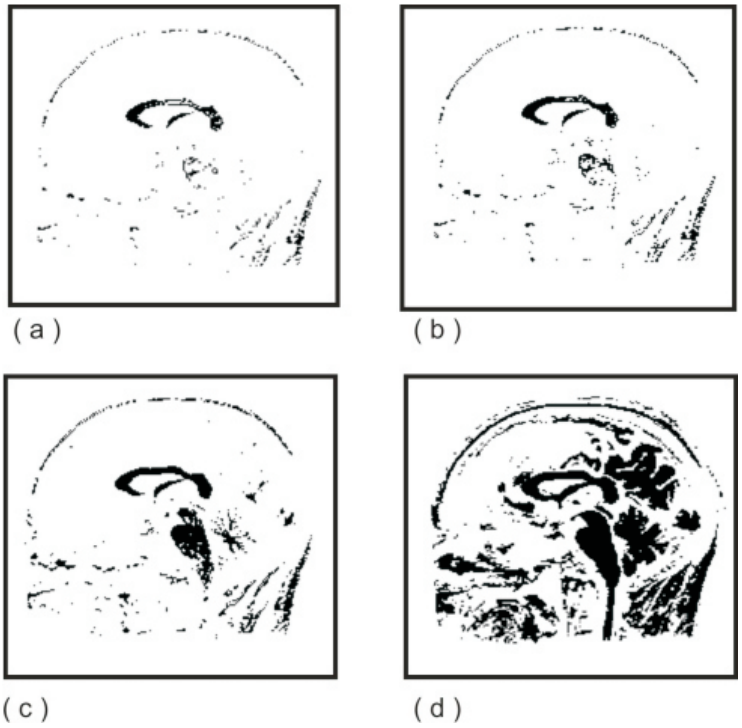

Figure 3.6: Thresholding with various threshold intervals.

3. Apply the threshold interval $\bar{X} \pm 3S$ to the reminder of the brain MRI scans where the corpus callosum is not delineated correctly.

4. Apply the multiscale normalized cut algorithm to the brain MRI scans resulting in step 3

For evaluation purposes the proposed approach was applied to brain MRI scans where the corpus callosum was not completely identified using the multiscale normalized cuts algorithm. Figure 3.7 presents a comparison between results obtained using the Multiscale Normalised Cuts algorithm (Figure 3.7(e) to (h)) and the proposed approach (Figure 3.7(i) to (j)). From the figure it can be seen that better results were obtained using the proposed approach.

Some *data cleaning* was also undertaken to remove the "incidental objects" discovered during segmentation. The heuristic used was that the object representing the corpus callosum can be identified using statistical measures (it is the largest object and is locate in roughly the centre of the brain). Having identified the corpus callosum object any other high intensity objects could be removed.

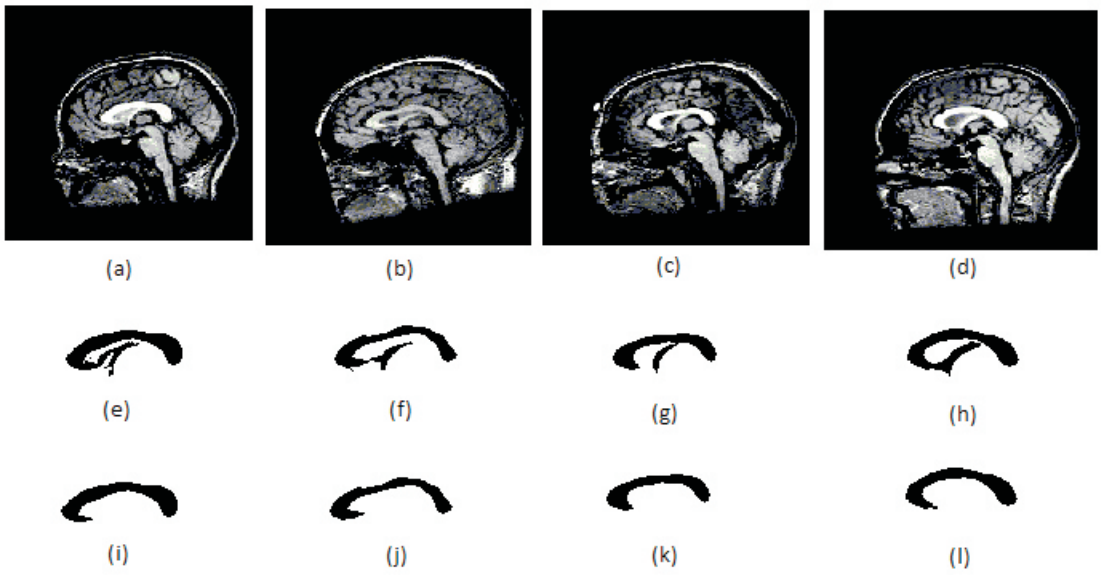

Figure 3.7: The results obtained by the multiscale normalized cuts algorithm and the proposed approach, (a-d) the original brain MRI scans, (e-h) the corpus callosum obtained by the multiscale normalized cuts, (i-l) the corpus callosum obtained by the proposed algorithm.

3.5 Medical Image Datasets

To evaluate the ROIBIC techniques described in this thesis to classify medical images according to a particular object that features across the data sets a number of data sets were used. As already noted the data sets were generated by extracting the *midsagittal slice* from MRI brain scan data volumes (bundles), one image per volume. Each data set thus comprised a number of brain MRI scans each measuring 256×256 pixels, with 256 greyscale levels and each describing a midsagittal slice. To support the evaluation the data sets were grouped as follows: (i) musicians, (ii) handedness and (iii) epilepsy. Each group is described in some further detail as follows:

Musicians datasets. For the musicians study the data set comprising 106 MRI scans, 53 representing musicians and 53 non-musicians (i.e. two equal classes). The study was of interest because of the conjecture that the size and shape of the corpus callosum reflects human characteristics (such as a musical or mathematical ability) [111, 132].

Handedness datasets. For the handedness study a data set comprising 82 MRI scans was used, 42 representing right handed individuals and 40 left handed individuals. The study was of interest because of the conjecture that the size and shape of the corpus callosum reflects certain human characteristics (such as handedness) [31].

Epilepsy datasets. For the epilepsy study three data sets (Ep_{106}, Ep_{159} and Ep_{212})were used. The first data set, Ep_{106},comprised the control group from the musicians study (the non-musicians) and 53 MRI scans from epilepsy patients so as to give a balanced data set. The second data set, Ep_{159}, used all 106 MRI scans from the musicians study and 53 epilepsy scans. The third data set, Ep_{212} was the same as the second but augmented with a further 53 epilepsy cases to create a balanced data set. The objective was to seek support for the conjecture that the shape and size of the corpus callosum is influence by conditions such as epilepsy [27, 124, 149].

All three brain MRI datasets were preprocessed using the variation of the multiscale normalized cuts algorithm described above (Subsection 3.4.4) to extract

the corpus callosum feature. On completion of data cleaning a "local" registration process was undertaken by fitting each identified corpus callosum into a Minimum Bounding Rectangle (MBR) so that each identified corpus callosum was founded upon the same origin.

3.6 Summary

This chapter has provided the necessary context to the preprocessing of the brain MRI scan data so as to detect and extract the desired ROI (the corpus callosum). A variation of the multiscale normalized cuts algorithm was proposed to achieve the desired segmentation. A description of the medical image datasets, used for evaluation purposes later in this thesis, was also given. In the following four chapters the four different techniques, whereby MRI brain scan data can be classified according to a particular ROI (namely the corpus callosum), considered in this thesis are described and evaluated. The first technique considered is the Hough Transform technique which is presented in the following chapter.