Semantic Assisted, Multiresolution Image Retrieval in 3D Brain MR Volumes

by

Azhar Quddus

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2010

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Content Based Image Retrieval (CBIR) is an important research area in the field of multimedia information retrieval. The application of CBIR in the medical domain has been attempted before, however the use of CBIR in medical diagnostics is a daunting task. The goal of diagnostic medical image retrieval is to provide diagnostic support by displaying relevant past cases, along with proven pathologies as ground truths. Moreover, medical image retrieval can be extremely useful as a training tool for medical students and residents, follow-up studies, and for research purposes.

Despite the presence of an impressive amount of research in the area of CBIR, its acceptance for mainstream and practical applications is quite limited. The research in CBIR has mostly been conducted as an academic pursuit, rather than for providing the solution to a need. For example, many researchers proposed CBIR systems where the image database consists of images belonging to a heterogeneous mixture of man-made objects and natural scenes while ignoring the practical uses of such systems. Furthermore, the intended use of CBIR systems is important in addressing the problem of "Semantic Gap". Indeed, the requirements for the semantics in an image retrieval system for pathological applications are quite different from those intended for training and education. Moreover, many researchers have underestimated the level of accuracy required for a useful and practical image retrieval system. The human eye is extremely dexterous and efficient in visual information processing; consequently, CBIR systems should be highly precise in image retrieval so as to be useful to human users. Unsurprisingly, due to these and other reasons, most of the proposed systems have not found useful real world applications.

In this dissertation, an attempt is made to address the challenging problem of developing a retrieval system for medical diagnostics applications. More specifically, a system for

semantic retrieval of Magnetic Resonance (MR) images in 3D brain volumes is proposed. The proposed retrieval system has a potential to be useful for clinical experts where the human eye may fail. Previously proposed systems used imprecise segmentation and feature extraction techniques, which are not suitable for precise matching requirements of the image retrieval in this application domain. This dissertation uses multiscale representation for image retrieval, which is robust against noise and MR inhomogeneity. In order to achieve a higher degree of accuracy in the presence of misalignments, an image registration based retrieval framework is developed. Additionally, to speed-up the retrieval system, a fast discrete wavelet based feature space is proposed. Further improvement in speed is achieved by semantically classifying of the human brain into various "Semantic Regions", using an SVM based machine learning approach.

A novel and fast identification system is proposed for identifying a 3D volume given a 2D image slice. To this end, we used SVM output probabilities for ranking and identification of patient volumes. The proposed retrieval systems are tested not only for noise conditions but also for healthy and abnormal cases, resulting in promising retrieval performance with respect to multi-modality, accuracy, speed and robustness.

This dissertation furnishes medical practitioners with a valuable set of tools for semantic retrieval of 2D images, where the human eye may fail. Specifically, the proposed retrieval algorithms provide medical practitioners with the ability to retrieve 2D MR brain images accurately and monitor the disease progression in various lobes of the human brain, with the capability to monitor the disease progression in multiple patients simultaneously. Additionally, the proposed semantic classification scheme can be extremely useful for semantic based categorization, clustering and annotation of images in MR brain databases. This research framework may evolve in a natural progression towards developing more powerful

and robust retrieval systems. It also provides a foundation to researchers in semantic based retrieval systems on how to expand existing toolsets for solving retrieval problems.

Acknowledgments

I would like to take this opportunity to thank those who have contributed to my thesis and education. My deep appreciation goes to my thesis supervisor Prof. Otman A. Basir, for his constant help, guidance and the countless hours of attention he devoted throughout the course of this research work. His incessant guidance and encouragements helped me in making valuable contributions. Despite his hectic schedules, he always had time for me. I am grateful to Dr. Basir for so many other things including enlightening discussions, financial and moral support, and advices in my career development. His level of enthusiasm fascinates me and often inspires and challenges me to put forth my best efforts.

Special thanks to Dr. Mohamed Kamel, Dr. M. Yahiya Dabbagh, Dr. Olga Vechtomova and Dr. Zhou Wang for devoting their time out of their busy schedules. Dr. Kamel, Dr. Dabbagh, and Dr. Vechtomova, served on the committee for my comprehensive examination and contributed valuable insights and ideas from their own areas of specialization.

Special thanks are due to my colleagues and friends including Dr. Abdul-Rahim Ahmad, Mr. Syed Anas Vaqar, and Mr. Syed Hasan Riyaz, for their valuable help and encouragement. They have made my stay at the University of Waterloo (UW) an exceedingly pleasant and unforgettable experience.

Special thanks to Mrs. Ferah Cagoglu and Dr. Abdul-Rahim Ahmad for helping me in the proofreading of this dissertation, as well as to Mr. Syed Anas Vaqar for helping me with reviewing and finalizing the presentation for the thesis defense.

Appreciation is due to the computing facilities at Pattern Analysis and Machine Intelligence (PAMI) Lab. I would like to acknowledge the department of Electrical and Computer Engineering for its support during the course of my studies.

Acknowledgment is also due to Dr. S. E. Black at Sunnybrook and Womens' College Health Sciences Center in Toronto for providing some data. Similarly, acknowledgment is due to Montreal Neurological Institute (MNI) and CASILab at the University of North Carolina at Chapel Hill for the Brainweb and the MIDAS datasets, respectively.

Finally, I would like to express my deepest gratitude to my parents for their continuous encouragement, to my wife, Shamroz, for enduring me all these years. Shamroz, you have been an incredible source of motivation with your never-ending patience and support. Words cannot express enough how grateful I am, and always will be, to you.

Dedicated to

Shamroz, my wonderful wife,

whose supplication, patience and encouragement

led to this accomplishment

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Information retrieval is a multi-billion dollar industry, if companies like Google™and Yahoo™are of any indication. While the text-based retrieval systems have matured well in recent years, multimedia retrieval systems, such as those based on images and videos, remain in infancy. Consequently, multimedia information retrieval systems have become subject of extensive research efforts.

Investment in the use of Information Technology (IT) in the healthcare sector has grown rapidly during the last decade. The use of IT in the healthcare domain has potential to efficiently maintain patient records, reduce patient waiting times, improve emergency response times, improve drug deliveries, as well as processing of insurance claims. The benefits of the use of IT is not limited to these aspects only. In fact, information technology has the capability to revolutionize the way we perceive healthcare services. Some of the future applications of IT in healthcare include delivery of the healthcare records on mobile devices, remote monitoring of critically ill patients, and live medical expert advices during surgical procedures.

Governments all over the world have invested billions of dollars to modernize and to replace the old paper-based records with new computer-based systems. For instance, *Canada Health Infoway* is an independent, federally-funded, not-for-profit organization tasked with accelerating the development of *Electronic Health Records* (EHR) across Canada. As a strategic investor, this organization works with Canadian provinces and territories with the goal of creating electronic health records for Canadians. Infoway's members are Canada's 14 federal, provincial and territorial Deputy Ministers of Health. Infoway has recently developed a pan-Canadian electronic health information standard, leveraging on the existing HL7v3 standards. By the end of 2008, there were 276 EHR projects under way in Canadian hospitals, other health-care facilities, pharmacies and laboratories, with an investment value of $1.5-billion from Canada Health Infoway.

Similarly, *eHealth Ontario* is tasked with facilitating the development of provincial public Electronic Health Record system. *eHealth Ontario* was created in September 2008 through a merger of the Ontario Ministry of Health's electronic health program and the *Smart Systems for Health Agency* (SSHA). The mandate of *eHealth Ontario* is to create electronic health records for all the patients in the province by 2015. The Government of Ontario is investing the capital as well as the political will, in making this strategy a success. In May 2008, cabinet approved the funds from fiscal year 2008-09 through to 2011-12. The total cost of Ontario eHealth strategy over the three years is $2.133 billion [2].

It is note worthy that the *Picture Archiving and Communication Systems* (PACS) are an integral part of such electronic health records. PACS, are software components for storing and accessing large amounts of visual data used in medical departments. Therefore, robust and efficient image retrieval systems are extremely useful in PACS.

This chapter introduces the research in medical image retrieval and also highlights the problems associated with it. It is organized in various sections, where the first section introduces a general framework of content based image retrieval; the second section discusses the challenge of semantic gap; the third section discusses the lack of practical uses of *Content Based Image Retrieval* CBIR systems; the fourth section discusses the CBIR in medical domain which is followed by the section on motivations; the sixth section presents the thesis objectives and the final section provides the organization of this dissertation.

1.1 Content Based Image Retrieval

Due to the prevalence of digital images in recent times, the usefulness of *Content Based Image Retrieval* (CBIR) systems in various important applications cannot be overemphasized. The notion of CBIR is to find similar images in a large image database through some key attributes or through some inherent features in those images. There are many application areas for CBIR, including healthcare, face recognition, remote sensing, astronomy, paleontology and document retrieval.

A typical CBIR system is shown in Figure 1.1, using a basic, simplified block schematic. Here, a feature database is populated offline from the collections of the images. In the normal online mode, a human operator provides a query image to the system and the system is supposed to find the most similar images from the database. This is done by matching the feature vectors of the query image to that of the feature vectors of the images in the database. Numerous variations to this basic architecture have been proposed in the literature.

The challenge in the development of a CBIR system is due to two important and ill-

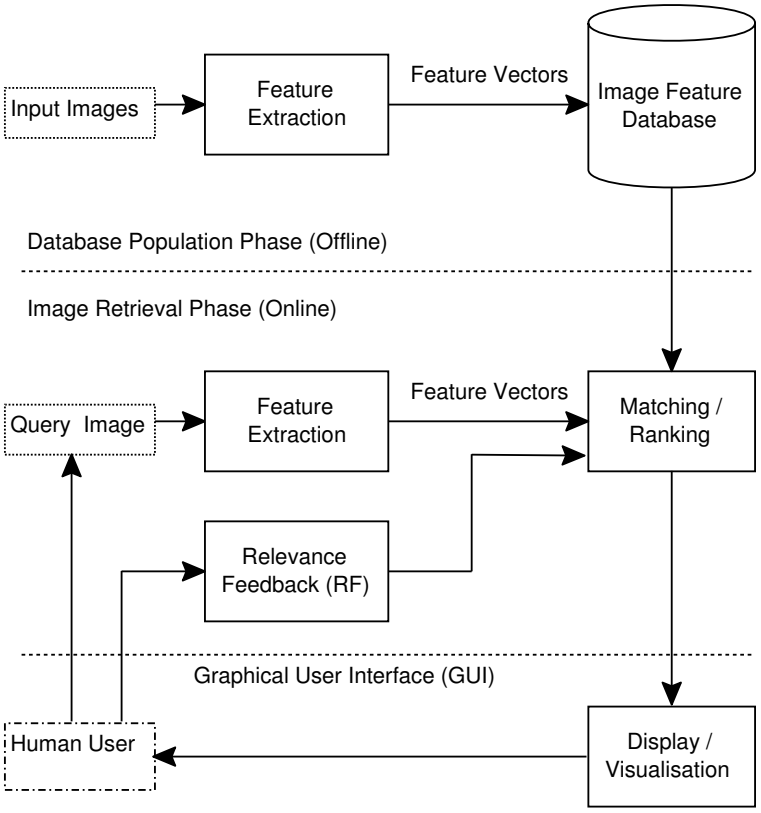

Figure 1.1: A typical and simplified block schematic of a CBIR system

posed problems. The first problem is the "Sensory Gap", which is the gap between the object in the world and the information in a computational description derived from a recording of the scene [3]. The second problem is the "Semantic Gap", which is the lack of coincidence between the information that one can extract from the visual data and the interpretation of the same data by a user in a given situation [3]. These gaps are shown in Figure 1.2, where the dashed lines indicate various modes of the semantic gap and dotted line indicates the sensory gap. The main components of a CBIR system are explained further in the following subsections.

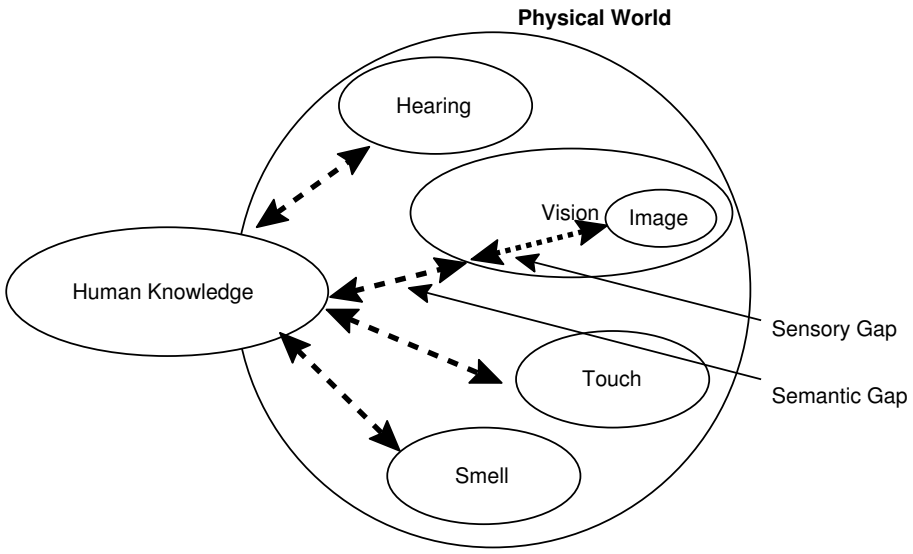

Figure 1.2: Sensory and semantic gaps in human knowledge accumulation

1.1.1 Feature Extraction

Feature extraction is a process to compute a set of inherent features in images, such as color, texture and shape etc. The selection of a proper image feature set is one of the most critical aspect of a CBIR system design. The object or an area of interest in images, may

appear differently due to change in color, shape, size, orientation or motion. Therefore, the problem of feature extraction has been extensively studied and explored in the areas of robotics and computer vision. Consequently, there are thousands of approaches presented and discussed by the research community. The problem of the feature extraction becomes extremely difficult when the object under consideration appears to be changing its color and shape due to motion, such as an octopus moving in water.

Some of the popular image feature sets include color, shape, texture, region and transform. In color-based techniques, feature sets include color histogram. In shape-based techniques, these sets normally include edges, corners, curvature scale space and chain codes. In texture-based techniques, these feature sets normally include co-occurrence matrices and Gabor filters, region based approaches use various kinds of segmentation schemes. In transform-based techniques, the feature sets normally include Fourier and wavelet transforms.

The type of images often dictate the type of features to be extracted, for example, color-based features would be preferred for color pictures. While feature sets are selected on the basis of some dominant property of the image, it is a normal practice to use different types of feature sets in a practical system.

The RGB color representation is a good choice for color-based systems, as it tries to model the input channels of the human eye. It is particularly useful when the color variance is minimal, as in case of images of paintings, photographs and trademarks taken in standard conditions. HSV representation has nice invariance properties under different object orientations [4], whereas the CIE-Lab representation is designed so that the Euclidean distance between two colors representations, models the human perception of color differences [3]. The texture in an image is another important feature type for both natural and

medical images. Textures are defined in terms of small repeated structures and randomness in the image. The existing literature in computer vision is rich in this regard.

The shape of an object is an important feature for any object-based retrieval systems. Here, object-based implies that images contain one or more specific objects, such as faces, flowers, plants, mountains, automobiles, trademarks and man-made objects. It is a common knowledge that people perceive and process knowledge mostly by recognizing objects and by associating relationships among these objects.

Natural objects tend to be more varied in shapes and sizes than man-made and industrial objects. Hence, most of the research in computer vision and pattern recognition has been done using the object-based paradigm. A review of the shape-based techniques can be found in [5]. In this regard, there are two broad categories of algorithms. In one category of techniques, an image is segmented based on color, grey-level or texture. Whereas in the other category of techniques, gradient based approaches are used where the objective is to extract the boundary of an object, assuming that the shape of the boundary can uniquely discriminate among various objects.

Some researchers in [6, 7, 8, 9, 10] have used machine learning based approaches to account for semantics. Multi-level image features have been suggested in [9, 10] and are found to be robust. In [11] image segmentation using statistical means is proposed.

1.1.2 Similarity Matching

Similarity matching involves comparing features of the query image and database images. An extensive amount of literature is available in the area of similarity matching. Researchers have used various kinds of distance measures to compare feature vectors. The

most common distance measure is the Euclidean distance which is used in many early works, including in [12]. A review of more complex distance metrics can be found in [13].

Similarity matching for the object based features requires shape matching algorithms. A review of such shape matching techniques can be found in [14]. In practice, shapes can be measured by various methods, such as moments, deformations and multiresolution-based approaches [15, 16]. For instance, in [17] shapes are matched using point correspondence. A recent work in [18] utilizes Triangle Area Representation (TAR) of each object stored at the corresponding node in the Curvature Tree (CT). The similarity between two multi-object images is measured based on the Maximum Similarity Subtree Isomorphism (MSSI) between their CTs. A recursive algorithm to solve the MSSI problem is proposed and an effective dynamic programming algorithm is developed to measure the similarity between the attributed nodes.

Recently, researchers have put efforts to map the similarity in semantic domains to low-level visual features through learning approaches, such as those proposed by [19] and [20]. In [19], Bhattacharya distance for statistical similarity is used where concept probabilities are determined for the combined color and texture feature vector by using multi-class SVM. In [20], the feature samples in the same category are trained as "similar" class and the feature samples in different categories as "dissimilar" class, and the training is achieved through a boosting framework. In [21], both supervised and unsupervised learning techniques are investigated to associate the low-level global image features (e.g., color, texture, and edge) in the projected PCA-based eigenspace to link high-level semantics with visual categories.

1.1.3 Query Formulation

Query formulation is an important subsystem from the user interaction perspective. Often query formulation dictates the significant portion of the design paradigm of a CBIR system. In content-based image retrieval, interaction is a complex interplay between the user, images, and semantic interpretations of these images [3]. Earlier systems used the simplest approach, where the user picks one of the database images as a query image and performs the retrieval. Consequently, most of the CBIR systems support "query-by-example"; some of them give additional facility to the user by providing "query-by-sketch" [22, 23] and keyword support. In order to improve the retrieval results, some systems, such as in [24, 25, 26], require the user to identify relevant regions.

1.2 The Challenge of the Semantic Gap

As already mentioned, the lack of acceptance of CBIR in mainstream and practical applications has largely been attributed to the so called "semantic gap" [3]. In other words, systems should be able to capture the intention of the user to efficiently provide better results. Generally, CBIR systems require the user to provide a query image to initiate the query. The issue of semantics also brings human subjectivity in the discussion because an image may contain more than one semantic aspects. Moreover, a user may interpret different semantics in an image based on his needs and requirements from time to time. There are many proposed techniques to address this problem. These techniques can be divided into two broad categories. The first category deals with the semantics in the query formulation while the second category involves semantics in the matching stage. These categories are further discussed in the following subsections.

1.2.1 Semantics in the Query Formulation

In this category, researchers have proposed additional facilities for the user to provide more accurate queries. Some of the systems require the user to identify relevant regions [24, 25, 26] to improve the results. Some researchers supplemented the query image with text fields to handle semantics in the retrieval process [27, 28, 29, 30, 31, 32]. In [33], both images and associated texts are indexed using medical concepts from the Unified Medical Language System (UMLS) meta-thesaurus in order to facilitate the automatic indexing and retrieval of large medical-image databases.

1.2.2 Semantics in Matching Subsystems

In this category, research activities have been focused on "Relevance Feedback" (RF). The main idea behind RF is to refine the retrieval results by allowing the user to teach the system by indicating good and bad results. The system is expected to modify its similarity matching criterion to improve the retrieval results. Variety of learning approaches have been presented in the literature to perform RF. While earlier techniques used incremental learning approaches presented in [34, 35], recently *Support Vector Machines* (SVM) have been popular [36] for this task.

The problem in RF is mainly due to the sparse training data and the non-availability and un-willingness of human operator to teach a system involving thousands of images. In order to circumvent some of these problems, some researcher have proposed to use the user log [37]. Moreover, human judgment is subjective and changes considerably from one person to another and hence reliability of the RF system is difficult to achieve in general sense.

1.3 Lack of Practical Uses of CBIR

Given the impressive amount of research efforts, the acceptance of CBIR for mainstream and practical applications is quite limited. This is due to the fact that the research on CBIR has largely been initiated as an academic pursuit rather than providing a solution to a human need. For example, many researchers proposed CBIR systems with databases consisting of heterogeneous mixture of variety of man-made objects and natural scene images without looking into the possibility of some practical uses of such systems.

Another reason for the lack of practical uses of CBIR technology is the semantic gap. Semantic gap in the abstract sense is difficult to implement because it would require a knowledge base of the semantics in the whole world. Therefore, the intended use of the system is important to address this gap. For example, semantics for the pathological system will be quite different than the one intended to be used for the training purposes. The semantics for a retrieval system for pathological use will include types and extents of diseases related to specific anatomical areas. On the other hand, the semantics for the retrieval system for educational purposes will require broad labeling of all the anatomical regions of healthy and abnormal conditions. Accordingly, meaningful semantics will be different for each class of the retrieval systems.

The researchers in the area of CBIR initially underestimated the level of accuracy required for a useful system. The human eye is extremely dexterous in visual information processing; consequently, CBIR systems have to be precise in order to be useful for users. As such most of the proposed systems have not been found useful in real world applications.

1.4 Medical Image Retrieval

The use of CBIR in the medical domain is the most important application area because digital images are produced in an ever-increasing quantities for the use in diagnostics and therapies. For example, the Radiology Department of the University Hospital of Geneva alone produced more than 12,000 images a day in the year 2002 [1]. In the same year, the total amount of cardiological image data produced in the Geneva University Hospital was around 1 TB.

The goal of medical information systems have often been defined as to deliver the right information at the right time, at the right place, to the right persons in order to improve the efficiency of the care processes [38]. Hence, such a goal would require more than a query by patient name, series ID or study ID for images.

Due to the advent of "Digital Imaging and Communications in Medicine" (DICOM), a standard for image communication, patient information can be stored along with the actual images. In many academic articles, the content-based retrieval of medical images for supporting clinical decision making has been proposed, which would improve the management of the clinical data. This also provides scenarios for the integration of content-based image retrieval methods into *Picture Archiving and Communication Systems* (PACS).

Clinical decision support techniques, such as case-based reasoning [39] and evidence-based medicine [40], can benefit greatly from efficient CBIR systems. Decision support systems in radiology and computer-aided diagnostics for radiological practice, as demonstrated at the Radiological Society of North America (RSNA), are on the rise and create a need for powerful data, meta-data management and retrieval systems [1]. The general clinical benefit of imaging systems has been demonstrated in [41].

In the medical application domain, many systems have been proposed where the image database consist of images of various anatomical regions with a variety of image modalities. One such database is ImageCLEFmed [42], which is developed to compare and evaluate general image retrieval algorithms. The image retrieval systems developed for these databases are more useful for applications in training and education rather than in diagnostic and decision support.

1.5 Motivations

The use of CBIR in medical diagnostics is the hardest but it is the most important application for image retrieval in the medical domain [1]. To be used as a diagnostic aid, image retrieval systems need to prove their performance to get accepted by clinicians. For application domains where evidence-based medicine or case-based reasoning is required, it is important to supply relevant and similar cases for comparison. Such retrieval will require visual features, which accurately model the visual detection of an MD using as much domain knowledge as possible.

There are two main ideas for supporting the clinical decision-making process. The first one is to supply the medical doctor with cases that offer a similar visual appearance. Results of the retrieval may supply a second opinion to the MD and support reasoning based on the various cases that are supplied by the system and the data available on the current patient. Another idea is the creation of databases containing normal (non-pathologic) cases and the comparison of the distance of a new case with the existing cases, thus performing dissimilarity retrieval as opposed to the similarity retrieval. This is even more natural compared to the normal workflow in medicine, where the first requirement is

to find out whether the case is pathologic or not [1].

One of the important application domains in medical imaging is the MR brain imaging. Given the fast aging population of baby boomers in North America and Europe, coupled with the prevalence of chronic illnesses such as Multiple Sclerosis (MS), Alzheimer (Dementia), and Stroke in this age-group, the usefulness of the MR brain imaging cannot be overemphasized. Figure 1.3 shows query T1 image (top-left) and three consecutive slices from the PD sequence all with $1mm^3$ voxels. It is clear from this figure that it is extremely difficult to retrieve the correct PD slice by the human eye.

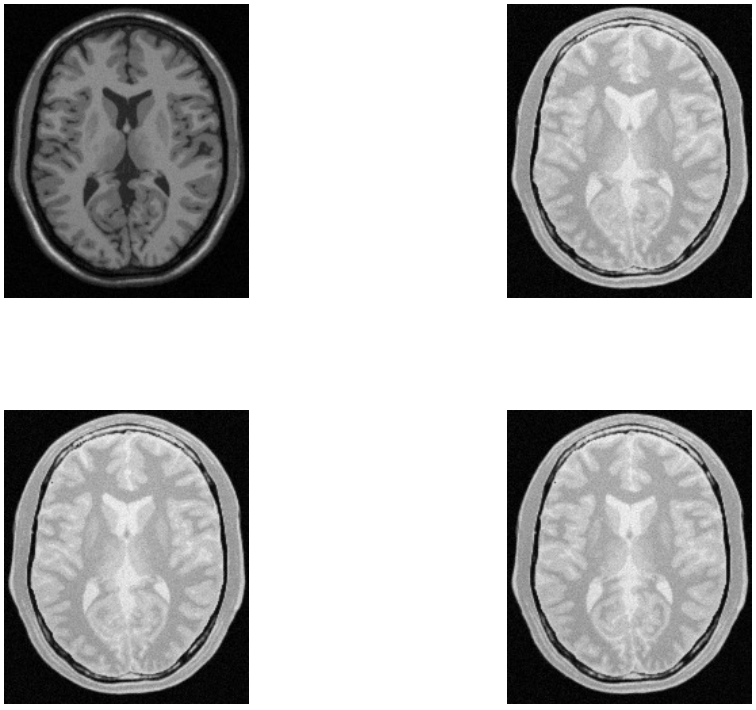

Figure 1.3: T1 query slice#82 (top left), PD slice#81 (top right), PD slice#82 (bottom left), PD slice#83 (bottom right)

Identifying a specific semantic area (such as "brain lobe") associated with the query slice would be extremely useful for quickly retrieving the results. This will also help

in quantifying, localizing and tracking of disease progression among various brain lobes. Identification of the 3D MR volume of a subject, among the volumes of a large number of subjects, given the query slice will help a great deal in monitoring and tracking of the disease progression among a group of patients (subjects). Therefore, implementing these ideas will help in achieving some of the goals of the medical information systems.

1.6 Thesis Objectives

Based on observations made in the previous section, this dissertation is targeted in addressing the challenging problem of developing an image retrieval system for medical diagnostics. More specifically, this dissertation proposes a system for semantic retrieval of MR images in 3D MR of brain volumes. Following are the thesis objectives in specific terms:

1. Develop a semantic based framework for retrieval of MR brain images for MR diagnostics applications.

2. Develop a fast retrieval application for multimodal, 2D brain MR image retrieval in 3D volumes.

3. Develop a robust system against variety of imaging related abnormalities such as MR inhomogeneity, noise and acquisition misalignments etc.

Previously, the proposed systems use imprecise segmentation and feature extraction techniques, which are not suitable for precise matching required for the retrieval of MR images. Several challenging problems need to be addressed in order to achieve these objectives:

1. The accuracy of the system needs to be extremely high.

2. The retrieval speed of the system needs to be acceptable.

3. The system needs to be robust against variety of imaging related abnormalities such as MR inhomogeneity, noise and acquisition misalignments.

In order to satisfy these requirements, a multiscale representation is used for image retrieval, which is robust against noise and MR inhomogeneity. An image registration based framework is developed to achieve high accuracy in the presence of mis-alignments. To speed-up the retrieval system, a fast discrete wavelet based feature space is used. Further improvement in speed is achieved by using semantic classification of the human brain into various semantic lobes. A novel fast retrieval system is proposed to identify a 3D volume in a database given a 2D query slice.

1.7 Organization of the Thesis

Chapter 1 provides a detailed introduction to CBIR systems. The problem of semantic gap has also been discussed with various mitigation techniques, such as relevance feedback (RF). The reasons for the lack of acceptance of image retrieval for general usage have been outlined. Motivations and objectives for this dissertation has been provided.

Chapter 2 provides an extensive literature review on medical image retrieval systems and discusses challenges associated with 2D image retrieval in 3D brain volumes.

Chapter 3 deals with the 2D rigid registration, where a multiscale wavelet based registration scheme is proposed. A multiscale greedy steepest gradient registration technique is

proposed for efficient implementation. The registration results with a greedy optimization technique are also compared against other two known optimization techniques. The results demonstrate the high efficacy of the approach under multimodal and noisy environment.

Chapter 4 deals with the intersubject retrieval part, which introduces two novel MI based schemes. The first retrieval technique utilizes full registration at each level of the retrieval, leveraging on all decomposition levels. The second retrieval technique starts with coarse registration using the coarse level decomposition and iteratively utilizes finer registration for the retrieval leveraging on finer levels of decompositions. It has been found that the second technique is generally five times faster than first approach while maintaining similar error rates. The proposed retrieval algorithms are tested under multi-modal and noisy conditions. Experiments show promising results with respect to multi-modality, accuracy, speed and robustness.

Chapter 5 introduces a novel, semantic assisted, intersubject retrieval technique to further speedup the retrieval system. A fast and simple feature extraction scheme is developed to extract features in the wavelet domain for the semantic classification. The feature set is compared against a well known feature extraction technique called *SIFT* and found be more effective in terms of speed and accuracy. In the first stage, a novel, SVM based, semantic classification scheme is proposed for classifying an incoming 2D query image into one of the semantic regions. These semantic regions are inspired by the four lobes of the human brain. In the second stage, registration based retrieval technique, proposed in chapter 4, is engaged in that specific semantic region, improving the speed by 50%. The classification errors are found to be small. The proposed retrieval algorithms are tested under multi-modal and noise conditions. The proposed retrieval algorithms are tested with healthy and MS images of the same patient. Experiments show promising results with respect to

multi-modality, accuracy, speed and robustness.

Chapter 6 presents a novel retrieval technique in a multisubject framework. In the first stage, a novel SVM based identification scheme is presented for identifying an incoming 2D query image and relating to a few highly likely subjects. This is achieved using a modified multiclass SVM with probabilistic outputs. In the second stage, an SVM based semantic classification scheme is presented for classifying an incoming 2D query image into one of the semantic regions. Finally, the registration based retrieval technique, proposed in chapter 4, is engaged in a specific semantic region of volumes belonging to the selected individuals. Simulations show promising results with respect to multi-modality, accuracy, speed and robustness.

Finally, in chapter 7, this dissertation is concluded and a few limitations of the proposed approach is mentioned. Various ways to extend this work in the future are also outlined.

Chapter 2

Background and Literature Review

2.1 Introduction

Content based Image retrieval has been an active research area over the last decade, however, first review articles on methods in image databases appeared as early as in 1981 [43]. In 1995, Enser [44] provided an extensive description of image archives, various indexing methods and common searching tasks, using mostly text-based searches on annotated images. In [45], an overview of the research domain is given and "Virage" system was introduced. The most complete overview of technologies was given by Smeulders et al. [3]. This article describes common problems such as the semantic and sensory gaps and provides links to a large number of articles describing the various techniques used in this problem domain. For an even deeper introduction into the domain, several books are available, such as [46, 47].

2.2 CBIR Systems

Although early systems existed already in the beginning of the 1980s (such as [48]), the majority would recall IBMs *Query By Image Content* (QBIC) as the start of content-based image retrieval [49]. The commercial QBIC system is definitely the most well-known system. Another commercial system for image and video retrieval is Virage [45] that has well known commercial customers, such as CNN. Most of the proposed systems are from academia. Some well-known examples include Candid [50], Photobook [51] and Netra [52] that use simple color and texture characteristics to describe the image content. The use of higher level information, such as segmented parts of the image for queries, was introduced by the Blobworld system [53]. On the other hand, PicHunter [54] is an image browser that helps the user to find an exact image in the database by showing to the user images on screen that maximize the information gain in each feedback step. Some systems are available as demonstration versions on the web such as Viper [55], WIPE7 or Compass [56]. In 2004, Mller et. al. [57] proposed image retrieval system for medical application known as "medGift". Table 2.1 provides a more complete list of the major CBIR systems along with the citations. A detailed comparison for some of these systems can be found in [58].

The proposed applications of CBIR systems include face recognition [11], healthcare [34, 73], remote sensing [74], astronomy [75], paleontology [76], document retrieval [77], textiles [78], robotics [79], airport management [80], trademarks and patents [81]. An other important application of this technology is to automatically annotate the images in a database [9, 82].

Table 2.1: CBIR Systems

Name	**Citations**
Blobworld	[53]
QBIC	[49]
PicHunter	[54]
Virage	[45]
PicToSeek	[59]
SIMPLIcity	[60, 61]
NeTra	[52]
MARS	[62]
WIBIIS	[63]
COMPASS	[56]
PICASSO	[64]
MediaNet	[65]
MUVIS	[66, 67]
PicSOM	[68]
WALRUS	[69]
Cortina	[30]
Viper	[55]
UCID	[70]
I-Browse	[71]
Kingfisher	[72]
medGIFT	[57]

2.3 Visual Image Features

There are many popular visual features such as color, texture and shape. Most of the available systems only use these primitive features unless manual annotation is coupled with the visual features. Even systems using segments and local features such as Blobworld [53] are still far away from identifying objects reliably. No system offers interpretation of images or even medium level concepts as they can easily be captured using text. This loss of information from an image to a representation by features is called the semantic gap [3]. The more a retrieval application is specialized for a certain, limited domain, the smaller the gap can be made by using domain knowledge.

It is now established that one kind of low-level feature is not sufficient for a complex task of image retrieval. Hence, a set of different features is normally used to achieve the desired performance.

2.3.1 Color

Color has been the most effective feature for the retrieval for general and stock photography images. Although, most of the images are in the red, green, blue (RGB) color space, this space is rarely used for indexing and querying because it does not correspond well to the human color perception. Other color spaces such as hue, saturation, value (HSV) [55, 53] or the CIE Lab [49] and LUV [83] spaces are much better with respect to the human perception and are more frequently used. In other words, the differences in the color space are similar to the differences between colors that humans perceive. Some effort has also been spent on creating color spaces that are invariant to shades and other influences such as viewing position [84, 85]. These approaches allow to identify colors even under varying

conditions but the information about the absolute colors is lost in the process. In medical domain, absolute color or gray level features are often of limited expressive power [1].

The edge histogram descriptor (EHD) is found to be quite effective for representing natural images [86]. This feature captures the spatial distribution of edges, similar to the color layout descriptor. To compute the EHD, a given image is first sub-divided into 4×4 sub-images, and local edge histograms for each of these sub-images is computed. Edges are broadly grouped into five categories: vertical, horizontal, 45^o, 135^o and neutral. Thus, each local histogram has five bins corresponding to the above five categories. The image partitioned into 16 sub-images results in 80 bins. These bins are non-uniformly quantized using 3 bits/bin, resulting in a descriptor of size 240 bits. The EHD is quite sensitive to objects or scene distortions [87].

2.3.2 Texture

Texture-based features try to capture the characteristics of the image or image parts with respect to changes in certain directions and the scale of the changes. This is most useful for regions or images with homogeneous texture. Due to the imprecise understanding and definition of texture, the research in texture-based features have larger variety than color-based features. However, texture provides important information in image classification as it describes the content of many real-world images such as fruit skin, clouds, trees, bricks, and fabric etc. Hence, texture is an important feature in defining high-level semantics for image retrieval purpose [87].

Texture features commonly used in image retrieval systems include spectral features, such as features obtained using Gabor filtering [88], and the statistical features such as the

six Tamura texture features [89]. Among the six Tamura features: coarseness, directionality, regularity, contrast, line-likeness, contrast and roughness, the first three were found to be more significant than the others. MPEG-7 has employed the regularity, directionality and coarseness as the texture browsing descriptor. The Wold features of periodicity, randomness and directionality have been proved to work well on Brodatz textures [90]. The limitation of Tamura features is its inability at multiple resolutions to account for scale. Wold feature is affected by image distortions such as scale and orientation variations due to perspective distortion [91]. These features are proved to be less effective when applied to natural scene image retrieval as texture regions in such images are not so structured and homogeneous [91].

Recent measures for capturing the texture in the images are Gabor filters [88, 92] and wavelets [93]. Perhaps, Gabor filters perform better as they correspond well to the properties of the human visual cortex [94]. Other traditional texture descriptors contain features derived from co-occurrence matrices [95], transform-based features [96, 97, 98].

2.3.3 Localized image features

Low-level features, such as color and texture, can be used on a global image level or on a local image level. The general approach is to use partitioning of the image for local feature extraction. Normally, image partitioning do not take into account any semantics of the image itself. When allowing the user to choose image regions of interest (ROI) [99], to delineate objects in the image [100] or when segmenting the image into areas with similar properties [24, 25], the locally extracted features contain more information about the underlying semantics.

2.3.4 Region and shape features

Automated segmentation of images into objects is an unsolved research problem. Even in fairly specialized domains, fully automated segmentation is often not easy to realize. In the area of image retrieval, several systems attempt to perform an automatic segmentation of the images before extracting image features [9, 34, 101]. To have an effective segmentation of images using varied image databases the segmentation process is normally done based on the color and texture properties of the image regions. Much has been written on medical image segmentation with respect to browsing image repositories [34].

MPEG-7 standard has included three shape descriptors for the object-based image retrieval [87]. First one is the 3-D shape descriptor, which is derived from 3-D meshes of the shape surface. The second descriptor is for the region-based shape, which is derived from Zernik moments. The third is for contour based shape, which is derived from Curvature Scale Space (CSS) [86]. Although the CSS descriptor is invariant to translation, scaling and rotation, it is sensitive to general distortions which can be resulted from objects taken from different point of view. Mokhtarian and Abbasi have extended the CSS descriptor to be robust to affine transform [15].

2.4 Similarity Metric

Previous techniques have used various distance measures to compare the feature vectors. The most simple distance measure is the Euclidean Distance which is used in several early works, including in [12]. A review of more complex ones can be found in [13]. Color based techniques resort to some type of histogram matching (such as [102, 103, 104]).

In case of region based features, image similarity is measured at two levels [87]. The first is the region-level, where the distance between two regions based on their low-level features is measured. The second is at the image-level, where the overall similarity of two images which might contain different number of regions is measured. Most researchers employ the Minkowski-type metric to define region distance. Suppose we have two regions represented by two p dimensional vectors $(x_1, x_2, ..., x_p)$, $(y_1, y_2, ..., y_p)$, respectively. The Minkowski metric is defined as

$$d(X, Y) = \left(\sum_{i=1}^{p} |x_i - y_i|^r \right)^{1/r} \tag{2.1}$$

When $r = 2$, it is the well-known Euclidean distance (L_2 distance). When r is 1, this metric becomes Manhattan distance (L_1 distance). Often weighted Minkowski distance function is used which introduces weighting to identify important features

$$d(X, Y) = \left(\sum_{i=1}^{p} w_i \, |x_i - y_i|^r \right)^{1/r} \tag{2.2}$$

where w_i , $i = 1, ..., p$ is the weight applied to different features.

Other distance types are also used in image retrieval, such as the Canberra distance, angular distance, Czekanowski coefficient [105], inner product, dice coefficient, cosine coefficient and Jaccard coefficient [106]. The overall similarity of two images is more difficult to measure and there are two ways to do this [87].

1. *One-One match*: Here, each region in the query image is only allowed to match one region in the target image and vice versa. Normally, each query region of the query image is associated to a single best matched region in the target image. Then the

overall similarity is defined as the weighted sum of the similarity between each query region in the query image and its best matched region in the target image.

2. *Many-Many match*: Here, each region in the query image is allowed to match more than one region in the target image and vise versa. A widely used method is the Earth Mover Distance (EMD), such as [107]. EMD is a general and flexible metric. It measures the minimal cost required to transform one distribution into another based on a traditional transportation problem from linear optimization, for which efficient algorithms are available. EMD matches perceptual similarity well and can be applied to variable-length representations of distributions.

Due to its simplicity, Minkowski metric is widely used in the previous systems to measure distance. However, extensive experiments have shown that it is effective in modeling perceptual similarity [108]. Measuring perceptual similarity is still a largely unanswered question because the processing of the human knowledge is not fully understood. Recently, researchers have suggested to use statistical learning-based approaches. The statistics (distribution) of the low-level features is largely unknown and often cannot be modeled hence non-parametric approaches are used. For example in [109], cosine similarity measure between feature vectors of query image and database image, for a particular feature input, is used which is based on probabilities of the multiclass SVM outputs.

Recently, a kind of statistical distance measure, namely the Bhattacharya distance [110], is applied for semantics based learning. The distance between a query image I_q and a database image I_j based on this representation [19] is defined as,

$$D_{S-correlation}(I_q, I_j) = \frac{1}{8}(\mu_q - \mu_j)^T \left[\frac{(\Sigma_q + \Sigma_j)}{2}\right]^{-1} (\mu_q - \mu_j) + \frac{1}{2} ln \frac{\left|\frac{(\Sigma_q+\Sigma_j)}{2}\right|}{\sqrt{|\Sigma_q|\,|\Sigma_j|}} \qquad (2.3)$$

where, μ_q and μ_j are the mean vectors of the local concept space and Σ_q and Σ_j are the covariance matrices of I_q and I_j, respectively. The Equation 2.3 is composed of two terms, the first one being the distance between the feature vectors, while the second term gives the class separability due to the difference between the covariance matrices.

2.5 Handling of Semantics

As mentioned in the previous chapter, reducing semantic gap is important in achieving desired performance from a CBIR system. All the visual features, and even features derived from segmented regions, are fairly low-level compared to high level concepts that are contained in the images. These visual features do not necessarily correspond to objects in the images or the semantic concepts or structures that a user might be interested in. Several articles speak of semantic or cognitive image retrieval (such as [60]). It is often required to connect visual low-level features with textual high level features and hence, the annotation of image collections for retrieval or for the combination with visual features for retrieval is another active research area [111, 9, 82, 112]. Many problems such as the subjectiveness of annotations need to be addressed even when working with restricted vocabularies.

According to Liu et. al. [87], the state-of-the-art techniques in reducing the semantic gap can be classified in five categories as follows.

1. Using object ontology to define high-level concepts [113, 114, 115, 116].
2. Using supervised or unsupervised learning methods to associate low-level features with query concepts [117, 9, 34, 111].

3. Introducing RF into retrieval loop for continuous learning of users intention [34, 118, 119, 111, 120].

4. Generating semantic template (ST) to support high-level image retrieval [121].

5. Making use of both the textual information obtained from the web and the visual content of images for Web image retrieval [112, 122].

2.5.1 Ontology

In many cases, semantics can be easily derived from any human language. For example, sky can be described as upper, uniform, and blue region. In systems using such simple semantics, initially different intervals are defined for the low level image features, with each interval corresponding to an intermediate-level descriptor of images, for example, light green, medium green, dark green. These descriptors form a simple vocabulary, the so-called object-ontology which provides a qualitative definition of high-level query concepts [87]. Database images can be classified into different categories by mapping such descriptors to high-level semantics (keywords) based on our knowledge [114, 115, 123, 124], for example, sky can be defined as region of light blue (color), uniform (texture), and upper (spatial location). A typical example of retrieval system using such an ontology-based system is presented in [113]. In this system, each region of an image is described by its average color in lab color space, and its position in vertical and horizontal axis, as well as its size and shape. Quantization of color and texture feature is the key in such systems. To support semantic-based image retrieval, a more effective and widely used method to quantize color information is by color naming.

Recently, a technique for automatically extracting the image content information into

MPEG-7 format and associating them to the existing domain ontologies is developed [116]. The validation experiments show that a high retrieval accuracy rate is obtained when all the image descriptors are combined with an ontology while building the semantic metadata for indexing images.

2.5.2 Supervised Learning

In order to handle semantics in the image retrieval, supervised learning schemes, such as support vector machine (SVM) [112, 125, 119, 19] and Bayesian classifier [54, 126, 127], are often used. Due to the strong foundation, SVM is used for object recognition and text classification. Moreover, it is also considered a good candidate for learning semantics in image retrieval systems. To learn multiple concepts, multi-class SVM is popular in recent works, such as [109], where both natural and medical images are semantically classified in a hierarchical framework. For the combined color and texture feature vector, the class or concept probabilities are determined by the prediction of the multi-class SVM. In [128], an SVM is employed for image annotation application, where binary SVM models are trained for each of the 23 selected semantics. In the testing stage, unlabeled regions are fed into each model. The concept from the model giving the highest positive result is associated with the region.

Another widely used learning method is Bayesian classification. In [129], high-level concepts of natural scenes are captured from low-level image features using binary Bayesian classifier. Database images are automatically classified in various types, such as indoor/outdoor. The outdoor images are further classified into various sub-classes, such as city/landscape. In [130], Bayesian network is used for indoor/outdoor image classification.

Recently, a boosting type framework is presented in [20] which uses images in the same category as "similar" class and images in different categories as "dissimilar" class. Here, two images in the training set are considered as similar if they match in semantic category or appear visually related. The goal is to simultaneously preserve both the semantic relevance as well as the visual similarity. For instance, two images could be defined to be similar only if they belonged to the same semantic category or similarity could be defined based on the images visual similarity according to the human perception.

Decision tree based techniques are also used to derive semantic features. Decision tree based methods, such as ID3, C4.5, and CART, build a tree structure by recursively partitioning the input attribute space into a set of non-overlapping spaces [131]. A set of decision rules is obtained by following paths from the root of the tree to leaves. In [132], the CART decision tree methodology is used to derive decision rules mapping between the global color distribution and the semantic keywords, such as Sunset, Marine, Arid images and Nocturne.

2.5.3 Unsupervised Learning

Unsupervised learning has no measurements of the outcome, rather the task is to find out how the input feature are organized or clustered. Image clustering is a typical unsupervised learning technique for retrieval purposes. It intends to group a set of image data in a way to maximize the similarity within clusters and minimize the similarity among different clusters. Each resulting cluster is associated with a class label and images within the same cluster are assumed to be semantically similar among each other. The traditional k-means clustering and its variations are often used for this task. In [133], k-means clustering is applied to the low-level color features belonging to the training images. The measurement

of variations within each cluster, is used to derive a set of mappings between the low-level features and the textual characterization (keywords) of the corresponding cluster. The derived mapping rules are used to index new untagged images added to the database.

In [19], self organizing maps (SOM) are used to represent images with automatically generated local visual concept. Here, visual concepts depict the perceptually distinguishable color or texture patches in the local image regions, which may not have clear semantic interpretation. For example, a predominant yellow color patch may be present in an image consisting of sun or a sunflower. In this framework, a visual concept vocabulary (codebook) is constructed by utilizing SOM. Subsequently, images are represented in a correlation and spatial relationship based concept feature spaces. This is achieved by exploiting the local neighborhood structure of the codebook, the local concept correlation statistics, and spatial relationships in the images. Finally, the features are used by a weighted similarity matching scheme, which is based on the relevance feedback framework. The feature weights are computed by considering both the precision and the rank order information of the top retrieved images of each representation, which adapts itself to individual searches to produce effective results.

2.5.4 Relevance Feedback (RF)

The main idea behind relevance feedback is to refine the retrieval results by allowing the user to teach the system by identifying good and bad results. In response, the system is expected to modify its similarity matching to improve the results. A large number of publications have appeared in the past to implement RF (such as [56, 54, 134, 135, 28, 136, 137, 138, 65, 139, 140, 118, 141]).

In the literature, machine learning techniques have been used in relevance feedback as well. SVM is often used to capture the concept of the query image by separating the relevant images from the irrelevant images using a hyper-plane in a projected space. The main advantage of SVM over other learning algorithms is its high generalization performance in the absence of a priori knowledge. The ability of SVM to work with small training sets is another advantage. In [142], *SVMactive* is proposed to use negative and non-labeled samples, and to learn a query concept more efficiently.

In most of the RF-based systems, the similarity measurement is fixed while the importance or weight of each descriptor is estimated through the RF from the user's feedback. In contrast to this, a generalized nonlinear RF algorithm for image retrieval has been proposed in [143]. In this approach, instead of adjusting the degree of importance of each descriptor, the similarity measure itself is estimated through an online learning mechanism. The method is based on the recursive optimal estimation of a nonlinear parametric relation of known functional components. However, due to the problem of optimization itself, the computation is expensive and algorithm may get trapped into a local minima.

Problem in RF is mainly due to the sparse training data and the non-availability or un-willingness of the human operator to teach a system involving thousands of images. In order to circumvent these problems, some researcher have proposed to use the user log [37, 144, 138]. However, the human judgment is subjective and changes significantly from one person to another.

2.6 Medical Image Retrieval

Medical domain is often cited as one of the principal application domain for CBIR technology in terms of potential impact [3, 145, 146]. The goals of medical information systems have often been defined to deliver the needed information at the right time, the right place to the right persons thereby improving the quality and efficiency of care processes [38].

Clinical decision support techniques such as case-based reasoning [39] and evidence-based medicine [40] may benefit greatly from efficient CBIR systems. Decision support systems in radiology and computer-aided diagnostics for radiological practice are on the rise which creates the need for powerful techniques for information retrieval [1]. The general clinical benefit of imaging systems is demonstrated in [41]. In the near future, purely visual image queries will not be able text-based methods, however, they have the potential to complement the text-based search. A hybrid retrieval system is proposed in [147].

For the clinical decision-making process, it is important to find similar images in various modalities acquired in various stages of the disease progression. This information is partly contained in the DICOM headers, however, DICOM headers have been found to contain a fairly high rate of errors, for instance, error rates in the range of 16% have been reported [148] for field anatomical regions.

Teaching and research in the healthcare domain may benefit significantly by the use of CBIR as visually interesting images are found in the existing large repositories. Visual features not only allow the retrieval of cases with similar pathologies but also help in the retrieval of visually similar cases but with different pathologies.

2.6.1 Uses in PACS and other Medical Databases

Implementation of CBIR systems in the medical domain is proposed in many papers [149]. Research articles describe the use of image retrieval within an image management framework, without stating what has actually been implemented and what is still in the status of ideas. Similarly, the integration into *Picture Archiving and Communication Systems* (PACS) or other medical image databases has been proposed often, but implementation details are generally rare [1].

Clinically relevant indexing and selective retrieval of biomedical images is explained in [150]. Some examples are given but no implementation details. It is proposed to make changes in the DICOM headers, which in principal ,is not allowed according to the standard. Many articles propose semantic retrieval based on images that are segmented automatically into objects and where diagnosis can be derived easily from the objects visual features.

PACS are the main software components to store and access the large amount of visual data used in medical departments. Often, layered architectures exist for quick short-term access and slow long-term retrieval. The general schema of a PACS system within the hospital is shown in Figure 2.1. The *Integrating the Healthcare Enterprise* (IHE) standard is aiming at the data integration in healthcare domain.

Indexing of the entire PACS causes problems due to the sheer amount of data that needs to be processed for allowing the image access by content. The issue of the amount of data that needs to be indexed is not discussed in any of the articles. In [151], an image classification system is integrated with PACS. Here, it is possible to search for certain anatomic regions, modalities or views of an image. A simple interface for coupling the PACS and the image retrieval system is stated as well. The search key is based on the

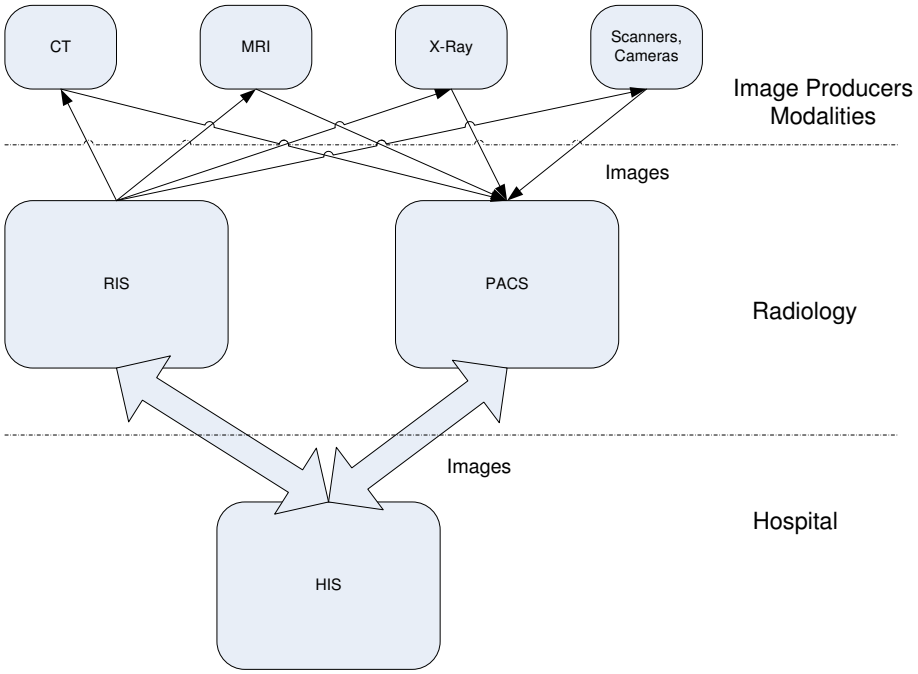

Figure 2.1: The basic position of a PACS within the information system environment in a hospital [1]

DICOM unique identifier (UIDs) of the images. There is a lack of publications describing the integration of image retrieval into the workflow in a medical institution. However, the use of content-based image retrieval is proposed in specialized collections. For example, in [152], CBIR is proposed in the context of a case database containing images and attached case descriptions.

2.6.2 Use in Various Medical Departments

The CBIR technology is proposed for various medical departments and specializations, however, most of the applications are centered around images produced in radiology departments [1]. A categorization of images used in various departments is described in [153]. The application of CBIR for CT brain scans is discussed in [154]. Mammographies are one of the frequent application areas for classification and content-based search [155, 119, 156].

CBIR for pathology images is often proposed where color and texture properties can easily be extracted. The tasks of a pathologist when searching for reference cases also supports the potential use of an image retrieval system. As early as 1986, the screening of cytological specimens is described in [157] and later extended towards CBIR in [158]. The use of CBIR in the retrieval of tuberculosis smears is described in [159]. An application of CBIR for histopathologic images is described in [160]. The retrieval of histologic images are proposed in [161]. Retrieval of cardiology MR images is proposed in [162]. Classification of dermatological images is explained in [163, 164]. Within the cardiology department, CBIR has been used to discover stenosis images [165].

The retrieval of High Resolution Computed Tomography (HRCT) scans of the lung is implemented in *ASSERT* project [166]. Here, the diagnostic quality assessed by using

the proposed retrieval system showed a significant improvement. The justification for the use of CBIR in this area is the hard decision-making task and the strong dependence of the diagnoses from texture properties. Descriptions of HRCT lung images, their visual features, and their pathologies are given in [167]. The retrieval of thorax radiographies is proposed in [168]. The retrieval of these images is more difficult because several layers are superposed and influence the visual content.

Many research papers propose the retrieval of medical images but the clinical evaluation is rarely done. For instance, in [153, 169], MR images of the brain are used to demonstrate the image search algorithms without mentioning the integration and evaluation. Table 2.2 lists several image retrieval systems proposed for various medical departments [1].

Table 2.2: Various image types and respective retrieval systems

Name	**Citations**
HRCTs of the lung	ASSERT
Functional PET	FICBDS
Spine X-rays	CBIR2, MIRS
Pathologic images	IDEM, I-Browse, PathFinder, PathMaster
Mammographies	APKS
Images from biology	BioImage, BIRN
Dermatology	MELDOQ, MEDS
Breast cancer biopsies	BASS
Varied images	I2C, IRMA, KMed, COBRA, MedGIFT, ImageEngine

2.6.3 Query Formulation

The query formulation based on visual features alone can be an issue as most systems in CBIR use the query by example (QBE) paradigm which needs an appropriate starting image for querying. This problem of missing starting image is known as the page zero

problem [1]. If text is attached to the images, which is normally the case in medical applications, then the text can be used as a starting point and once visually relevant images have been found, subsequent queries can be entirely visual. In the medical decision-making process, there are often images available for the current case and hence the problem of starting point does not arises.

In case of segmented images, the user may restrict the query to a certain *Region Of Interest* (ROI) [170], which may lead to more accurate retrieval results. The use of human sketches is also proposed in medical applications [147, 170, 171]. Considering the difficulty in exact drawing and the need for some artistic skills and time, this method will only be applicable for a small subset of queries, such as tumor shapes or spine X-rays. For general image retrieval tasks sketches are too time-consuming and often the retrieved results are not precise enough.

2.6.4 Use of Text in Medical Domain

In order to search image content, many systems propose to use text from the patient record [172] or studies [170]. Others researchers define a context-free grammar [165], a standardized vocabulary for image description [160]. In [173] researchers used text from radiology reports to transform it into concepts in the UMLS metathesaurus in order to retrieve the images. The use of text for queries is efficient, however it may not qualify to be termed as content-based as the text does not necessarily define an image content. The textual information in medical records rather puts images into the context they have been taken in. However, the combination of textual with visual features of the images does have the potential to produce excellent results [147].

2.6.5 Visual Features in Medical Domain

Images in the medical domain do not always contain color therefore the color properties are not important. However, some images, such as in dermatology, does have color information. Pathologic images are needed to be normalized as different staining methods can produce different colors. Within the radiology department, the normalization of gray levels between different modalities can cause problems when there is no exact reference point is available for the density of the tissue. A book chapter in [174] illustrates the dependency of intensity values of the brain tissues in various modalities. Here, the direct use of color and gray level features are not useful and texture and shape features gain importance. In mammography, denseness is used for finding small nodules [175]. It is interesting to compare various texture descriptors. Many of the texture descriptors model the same information and will most likely deliver similar results.

The shape features of image segments can be powerful depending on the accuracy of the image segmentation algorithm. Most common shape descriptors are variants of Fourier descriptors [99, 176] which allow to extract invariant descriptions. The use of image segmentation also permits to exploit spatial relationships within the image. Image segmentation based features are often proposed but consistent segmentation is difficult to realize. Nevertheless, many proposed techniques, such as in [177], automatic segmentation is proposed.

The use of eigenimages for the retrieval of medical images in analogy to eigenfaces for face recognition is proposed in [178, 179]. These features can be used for classification when a number of images for each class exist. Still, the features are purely statistical and it is hard to actually explain the similarity of two images based on these features which

can more easily be done using histogram intersection. In [170], signatures of the manually segmented objects of the images are proposed to reduce the list of resulting images. Other techniques include Tissue Time Activity Curve (TTAC), which is used in [180] for the retrieval of PET images. These are not really image features, rather 1D temporal signals that are compared.

Unfortunately, many articles those propose content based image queries do not explain reasons for the selection of a specific feature set. Sometimes, only a vague description such as general texture and color or gray level features are given, such as in [181, 153].

2.6.6 Recent Works

Recently, researchers have put effort to map the semantic domain to low-level visual features using machine learning approaches, such as in [21, 109, 19, 20]. In [19], Bhattacharya distance for statistical similarity is used. Here, for the combined color and texture feature vector, the class or concept probabilities are determined by using the prediction of a multi-class SVM classifier. In [20], authors used images in the same category as "similar" examples and images in different categories as "dissimilar" examples. These examples were learned through a boosting based binary classification framework. In [182], researchers map the low-level features to simple words (such as corner, texture etc) and subsequently used machine learning approach to map these simple words to high-level concepts.

Moreover, recent researchers, such as [21], have investigate both supervised and unsupervised learning techniques to associate low-level global image features (e.g., color, texture, and edge histogram) in the projected PCA-based eigenspace with their high-level semantic visual categories.

2.6.7 Performance Evaluation

Generally, it is difficult to compare two retrieval systems and it is true for medical image retrieval systems as well. However, there are several articles on the evaluation of imaging systems in medicine, such as [41]. Many researchers often present screenshots of retrieval results, which is not enough. Visual presentation of the result does not reveal a great deal about the real performance of the system, which is subjective as well.

Many system evaluations show measures with a limited power for comparison. In [183], the precision of the four highest ranked images is used without revealing the number of actually relevant items. In [179] researchers measure the number of times a differently scaled or rotated image retrieves the original which is not a fair performance evaluation. In medical statistics commonly used measurements are sensitivity and specificity which are defined as

$$sensitivity = \frac{positive\ items\ classified\ as\ pos.}{all\ positive\ items} \tag{2.4}$$

$$specificity = \frac{negative\ items\ classified\ as\ neg.}{all\ negative\ items} \tag{2.5}$$

Systems that use sensitivity and specificity include [184, 185]. These values can also be presented in the form of a ROC curve which contains much more information. As many of the presented systems use classifications of images, "accuracy" is often used to evaluate the system [186, 187, 176] which can be defined as,

$$accuracy = \frac{items\ classified\ correctly}{all\ items\ classified} \tag{2.6}$$

It is important to mention that CBIR systems not only find similar images but also classify the images. This is often more helpful as the practitioner must still judge the retrieved cases and the reasons for retrieving the images are often clearer, whereas, classification results are sometimes hard to detail and need to be explained. Precision and recall are well known measurements used in the domain of information retrieval [188]. These are also used in CBIR extensively [189] and defined as,

$$precision = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ items\ retrieved} \tag{2.7}$$

$$recall = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ relevant\ items} \tag{2.8}$$

Another rarely mentioned evaluation parameter is the retrieval speed which is important for an interactive system. In [190] it is only mentioned that the speed is reduced from hours to minutes for a set of 4000 images.

Performance evaluation metrics need to show the usefulness for application. Such an evaluation does not only contain the validation of a technology but also the inclusion of human factors into the process such as usability issues and acceptance of the technology [191], which can only be obtained through real user tests.

Finally, it will be interesting to evaluate the clinical impact of the application in real clinical practice [1]. Are these technologies able to reduce the length of stay of patients or do they manage to reduce the use of human resources for the patient care? Studies on clinical effects of image retrieval technologies might still be a distance away, however there are several necessities, such as the definition of standard and freely available databases, the definition of query topics for these databases and the definition of "gold standards".

2.7 Challenges in Medical Image Retrieval

Based on the discussion in the preceding sections, it is obvious that the image retrieval is a difficult problem. In this section, some of the major challenges in the area of medical image retrieval are outlined as follows:

1. Application of CBIR in medical domain is potentially useful.
2. Extraction of robust and precise visual features from medical images is a difficult problem.
3. The use of CBIR in medical diagnostics is important though it is difficult to realize.
4. To be used as a diagnostic tool, the CBIR systems need to prove their performance to be accepted by the clinicians.
5. In medical application domain many systems have been proposed where database consists of images of various anatomical regions with variety of image modalities (such as ImageCLEFmed database [42]). Such databases are useful as a benchmark to test various approaches in a general image retrieval framework, however these approaches are not useful for diagnostics support systems where high precision is required.
6. Useful semantics for medical image retrieval needs to be established.

2.8 Image Retrieval Application for Magnetic Resonance (MR) Brain Images

This dissertation attempts to address the need of image retrieval in medical diagnostics. The goal of diagnostic medical image retrieval is to provide diagnostic support by displaying relevant past cases, along with proven pathologies as ground truth [192]. Moreover, medical image retrieval may also be useful as a training tool for medical students and residents, follow-up studies, and for research purposes. Image registration is an important technique in the area of medical image analysis. Generally, it is needed for combining information from multiple imaging modalities, monitoring changes, image guided surgery or compare individuals anatomies to standard atlas. A majority of literature reviewed discusses medical image retrieval and registration as separate topics with respect to techniques and methodologies for a variety of image modalities.

Many of the proposed retrieval systems in the area of medical domain are adopted from general image retrieval schemes which perform satisfactorily with databases consisting of heterogeneous images of different modalities and anatomical regions. These systems use imprecise segmentation and feature extraction techniques which are not suitable for precise matching required for the retrieval of same 2D brain images (slices) in 3D volumes for diagnostic support. Only a couple of research papers ([193, 194]) have been reported to solve 2D slice retrieval problem. In [193] Karhunen-Loeve transform was used for the retrieval of relevant slices in the eigenimage domain. However, this technique requires a computationally expensive registration and intensity normalization step. Whereas, in [194], a computationally expensive preprocessing step is needed where brain tissue is removed from the skull and the mid-saggital plane is detected to compensate the alignment.

The system proposed in this dissertation exploits the use of 2D registration for retrieval purposes. Since mutual information provides high precision, it is used as a metric in the proposed research. Moreover, the use of image registration based framework provides desired properties of affine transform invariance and precise matching in an integrated fashion.

2.8.1 Image Registration for Retrieval Applications

Image registration is an important technique in the area of medical image analysis. Generally, it is needed for combining information from multiple imaging modalities, monitoring changes, guiding surgeries and comparing individuals anatomies to the standard atlas. Recently, it has been realized that both retrieval and registration techniques in medical image domain share some common steps and may complement each other [195, 196, 197].

In the context of fast and robust image retrieval, 2-D rigid registration, is preferred against 3D non-rigid registration for the following reasons:

- Medical images are multimodal and heterogeneous with temporal properties. Hence, multimodal image registration can be useful for medical image retrieval applications.
- Generally speaking, medical images are of high dimensionality, however these images are often organized and visualized as collections of 2D slices. Thus, 2D registration is a basic common denominator for image retrieval applications in the medical domain.
- The size of these images constitutes a challenge from the computational requirements point of view. Image retrieval applications involve large number of images to be processed.

- Image retrieval applications demand fast, robust and fully automatic registration techniques. Given the state-of-the-art, 2D rigid registration techniques appear to be more suitable as deformable registration techniques require some form of manual intervention [195].

- Deformable registration may easily fail in many situations due to image artifacts [198].

Image registration, in multi-modal setting, is a difficult problem. Mutual information (MI) based techniques have been quite successful in the area of medical image registration [199, 200]. The multiresolution and wavelet based approaches tend to be more robust [201, 202, 203, 204]. A wavelet based image retrieval is proposed for medical applications in [195]. In [201], multiresolution representation of the image is used for registration. There are several multiresolution based registration techniques proposed in the literature. In [202], Gabor Filters were employed for multiresolution decomposition. In [203], remote sensing images are registered using HL and LH coefficient of the wavelet decomposition. In [204], low-pass Haar wavelet coefficients along with MI and *Sum of Absolute Difference* (SAD) were used as registration metrics. Kullback-Leibler distance in multiresolution setting is employed in [205], where pre-aligned training images were used.

In multi-modal scenario, image structure is more important than absolute gray levels. Sometimes global count of MI can be misleading due to the absence of spatial information. Computation of a similarity index based on MI with spatial information is promising [199, 206]. Magnetic Resonance field inhomogeneity is a well known problem in MR images. Normally, literature on medical image registration do not discuss it because the same scanner is used to acquire images being registered. However, this assumption does not

hold in a general image retrieval scenario. Consequently, we propose to use multiscale wavelets with MI for fast, robust and automatic image registration.

It is tempting to "naively" argue that if registration is to be used in the retrieval process then it should be performed between 3D volumes. There are several problems with this approach:

1. Query information may contain only 2D slices and 3D volume may be not be available.

2. Even when query volume is available, it is not always possible to successfully register two volumes automatically. Since 3D volumes registration involves affine or deformable transformation, the registration will not be successful in case of large mis-alignment or occlusion.

3. Precise 3D deformable registration requires a few semi-automatic preprocessing steps, such as skull removal.

4. The registration between 3D volumes are computationally expensive.

Recent medical image registration techniques include multiscale edges [196], local Phase coherence [207, 208] and complex phase order (CPOL) [209]. While multiscale edge based technique use efficient Mallat's wavelet [210], phase based techniques utilize Kovesi's phase congruency computation [211]. It is worth noting that wavelet based techniques are multiscale in nature and hence expected to be more robust than phase congruency based techniques. Nevertheless, both the multiscale and the phase based techniques are attractive as they highlight the structures being registered rather than the absolute gray-level.

2.8.2 Slice Retrieval in 3D Brain Volumes

Many of the proposed retrieval systems in the area of medical domain are adopted from the general image retrieval schemes which perform satisfactorily with databases consisting of heterogeneous images belonging to different modalities and anatomical regions. Such systems use imprecise segmentation and feature extraction techniques which are not suitable for very precise matching required for the retrieval of 2D images in 3D MR brain volumes.

Only a couple of research papers [193, 194] have attempted to solve the 2D MR slice retrieval problem in the past. In [193], the Karhunen-Lo'eve transform is used for the retrieval of MR slices in the eigenimage domain. However, the proposed technique required a computationally extensive registration and intensity normalization steps. In [194], computationally extensive preprocessing steps are used to remove the brain tissue and to detect mid-saggital plane for compensating the alignment. These preprocessing steps work well in the central portion of the brain but may yield erroneous results in other areas. Moreover, these techniques are not tested in multimodal retrieval applications.

The author of this dissertation strongly feels that the 2D slice retrieval problem is important to address the needs of image retrieval in medical diagnostic. Specifically, a robust, multimodal and automatic retrieval system will be a significant contribution to the research community.

Chapter 3

Multiscale Image Registration for Retrieval Applications

3.1 Introduction

Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It geometrically aligns the reference and the sensed images. Differences between images arise due to different imaging conditions. Image registration is a crucial step in all image analysis tasks in which the final information is gained from the combination of various data sources, like in image fusion, change detection, and multichannel image restoration [212].

One of the most exciting and the fastest growing research areas in the field of medical imaging is image retrieval, with a particular interest in CBIR. Image retrieval involves finding similar images from a large image database with the help of some key attributes associated with images or features inherently contained in those images. There are few

medical CBIR systems reported in the literature, such as ASSERT [213], CBIR2 [214], IRMA [215], I-BROWSE [161] and Pathfinder [216].

In the medical domain, the goal of image retrieval is to provide diagnostic support by displaying the relevant past cases, along with the proven pathologies as ground truths [192]. Moreover, medical image retrieval may also be useful as a training tool for medical students and residents, follow-up studies, and for the research purposes. Image registration is an important technique in the area of medical image analysis. Generally, it is needed for combining information from multiple imaging modalities, monitoring changes, guiding surgeries and comparing an individual's anatomies to the standard atlas.

Recently, it has been realized that both retrieval and registration techniques in the medical domain share common image processing steps and require to be integrated in a larger system to complement each other [195]. There are two broad categories of image registration i.e. rigid and non-rigid (deformable) registration. Rigid registration only involves translations and rotations but affine and deformable registration involves many degrees of freedom such as scaling, shear and contour matching. In this chapter [1], a 2D rigid registration is preferred for fast and robust image retrieval applications due to the reasons mentioned in Chapter 2.

Image registration, particularly in the multi-modal setting, is a difficult problem. Mutual information (MI), as a similarity measure, has been quite successful in the area of medical image registration [199, 200]. Multiresolution and wavelet based approaches tend to be more robust, such as proposed in [201, 202, 203, 204]. In [201], multiresolution representation of the image is used for registration. Currently, there are several multireso-

[1] A part of this research has been presented at the International Conference on Bio-Medical Engineering and Informatics (BMEI 2008) in China.

lution based registration techniques in the literature. In [202], Gabor Filters are employed for multiresolution decomposition, while in [203], remote sensing data is registered using HL and LH coefficients of the wavelet representation. In [204], low-pass Haar wavelet coefficients along with MI and SAD (sum of absolute difference) are used for registration. Kullback-Leibler distance in multiresolution setting is employed in [205], where pre-aligned training images are used.

In the multi-modal scenario, the image structure is more important than the absolute gray levels. It has been shown that sometimes the global count of the mutual information can be misleading due to the absence of spatial information. The computation of a similarity index based on MI with spatial information is shown to be promising [199, 206]. The MR field inhomogeneity is a well know problem, however the literature on medical image registration do not generally discuss about it because the same scanner acquires images being registered. However, this assumption does not hold good in a general image retrieval scenario.

The proposed work differs significantly from the previous research work in several ways. First, it involves of the use of 2-D rigid registration for image retrieval applications in the medical domain, rather than the use of 3-D deformable registration. Second, MI in multiscale wavelet domain is employed to achieve fully automatic image registration. Third, this scheme does not inherit the problem of local maxima caused (in MI profile) by subsampling of the data at higher scales as discussed in [200, 201]. Fourth, we have suggested using variable bin sizes (at multiple scales) to speed-up the computation.

The rest of this chapter is organized as follows. The second section provides a brief overview of the multiscale wavelet decomposition; the third section discusses the use of mutual information (MI) in wavelet decomposition; the fourth section presents the image

registration based on greedy steepest gradient; the fifth section provides the results with 2-D images under rigid transformation. Finally, the sixth section concludes this chapter.

3.2 Multiscale Edge Representation and Decomposition

Multiscale edge representation of a 2-D signal provides characterization of singularity in an image, namely, Lipschitz exponents [217, 210]. This representation is efficiently computed at dyadic scales using separable low-pass and high-pass filters. In order to compute the decompositions at coarse scale, filters are upsampled instead of subsampling the image itself. Hence, this scheme does not inherit the problem of local maxima caused by subsampling of the data at higher scales. In [210], it has been shown that a close approximation of original signal can be reconstructed from its wavelet transform modulus maxima. In [218] this representation is used for the image registration using edge correlation as matching criterion.

This section introduces the multiscale edge detection and representation through dyadic wavelet transform. The same notations (as in the original work [210]) is used here for obvious reasons. In two dimensions, a multiscale edge detection can be formalized through a wavelet transform defined with respect to two wavelets $\varphi^1_{2^j}(x,y) = \frac{1}{2^{2^j}}\varphi^1\left(\frac{x}{2^j}, \frac{y}{2^j}\right)$ and $\varphi^2_{2^j}(x,y) = \frac{1}{2^{2^j}}\varphi^2\left(\frac{x}{2^j}, \frac{y}{2^j}\right)$. The wavelet transform of an image $f(x,y) \in L^2(R^2)$ at the scale has two components defined by, $W^1_{2^j}f(x,y) = f * \varphi^1_{2^j}(x,y)$ and $W^2_{2^j}f(x,y) = f * \varphi^2_{2^j}(x,y)$. The 2-D *dyadic wavelet transform* of $f(x,y)$ as the set of functions
$Wf = \left(W^1_{2^j}f(x,y), W^2_{2^j}f(x,y)\right)_{j\in Z}$.

The magnitude of this wavelet decomposition is given as,

$$M_{2^j} f(x,y) = \sqrt{\left|W^1_{2^j} f(x,y)\right|^2 + \left|W^2_{2^j} f(x,y)\right|^2} \tag{3.1}$$

If $\varphi^1_{2^j}(x,y)$ and $\varphi^2_{2^j}(x,y)$ are quadratic spline functions (derivative of a cubic spline function) then the wavelet transform can be implemented efficiently using simple separable filters [210]. The implementation details of this wavelet transform are provided in appendix A. Figure 3.1 shows a sample image and Figure 3.2 shows the multiscale wavelet decompositions ($W^1_{2^j} f(x,y)$,$W^1_{2^j} f(x,y)$ and $M_{2^j} f(x,y)$ column wise) at scales $j = 1...4$. Here, fine details are available at the lower scales while coarse details can be observed at the higher levels.

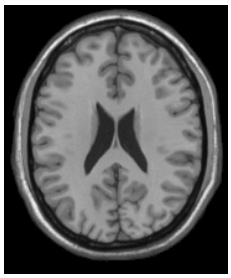

Figure 3.1: A sample image

3.3 Mutual Information in Multiscale Edge Decomposition

In the area of medical imaging, various image modalities provide complementary information. Here, images from different image modalities are often fused or combined to improve

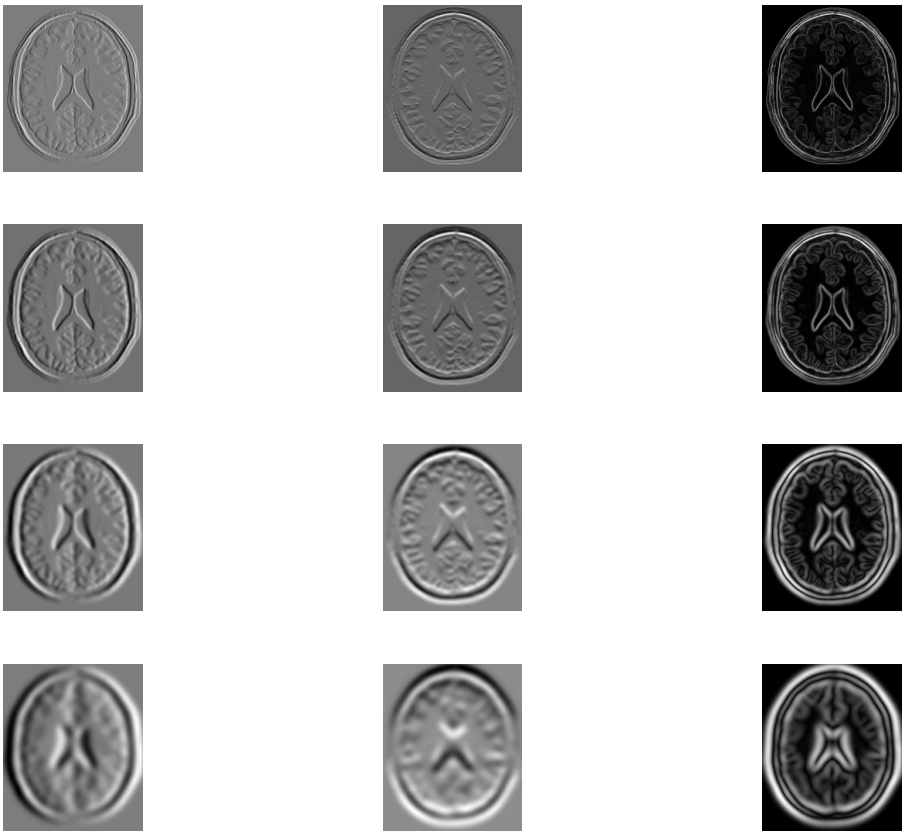

Figure 3.2: $W^1_{2^j} f(x,y)$,$W^1_{2^j} f(x,y)$ and $M_{2^j} f(x,y)$ (column wise) of the image in Figure 3.1, with $j = 1$ *to* 5 levels (top to bottom).

the diagnosis of the disease. Also different image modalities indicate same tissue with different gray levels. For example, a brain structure may appear as bright in MR but dark in SPECT scan [206]. Hence, matching the important structures in multiscale edges in the image produces a good registration.

The Mutual Information (MI) between two random variables A and B is given as [219],

$$I(A,B) = \sum_{a,b} P_{A,B}(a,b) log \frac{P_{A,B}(a,b)}{P_A(a)P_B(b)} \tag{3.2}$$

where $P_A(a)$ and $P_B(b)$ are marginal probability distributions and $P_{A,B}(a,b)$ is the joint probability distribution. MI is related to Shannon entropy as,

$$H(A,B) = H(A) + H(B) - H(A,B) \tag{3.3}$$

where $H(A) = -\sum_a P_A(a) log P_A(a)$ and $H(A,B) = -\sum_{a,b} P_{A,B}(a,b) log P_{A,B}(a,b)$

In practice, MI is computed from normalized joint histogram of the two images being registered [200]. The joint histogram is computed using 64 bins at each scale. It should be noted that using less number of bins also speeds up the computation of MI and hence the registration process.

Figure 3.3 shows the profile of MI with rotation at various scales where the top decomposition belongs to the finest resolution and the bottom decomposition belongs to the lowest resolution. The joint histogram is computed using 64 bins at each scale. Here, it can be observed that the capture range is larger at coarse scale while more precise localization of peak of MI can be obtained at the finest level. Moreover, the slope of MI profile is larger in the coarse scale. In retrieval applications, alignment may be required among the images

belonging to the different locations (for example 5 slices off). In such cases, MI profile at coarse scale may contain local maxima which may lead to the failure in registration, as shown in Figure 3.4. This affect may be reduced by using less number of bins in the coarse scale at the expense of smoothness of the peak, as shown in Figure 3.5.

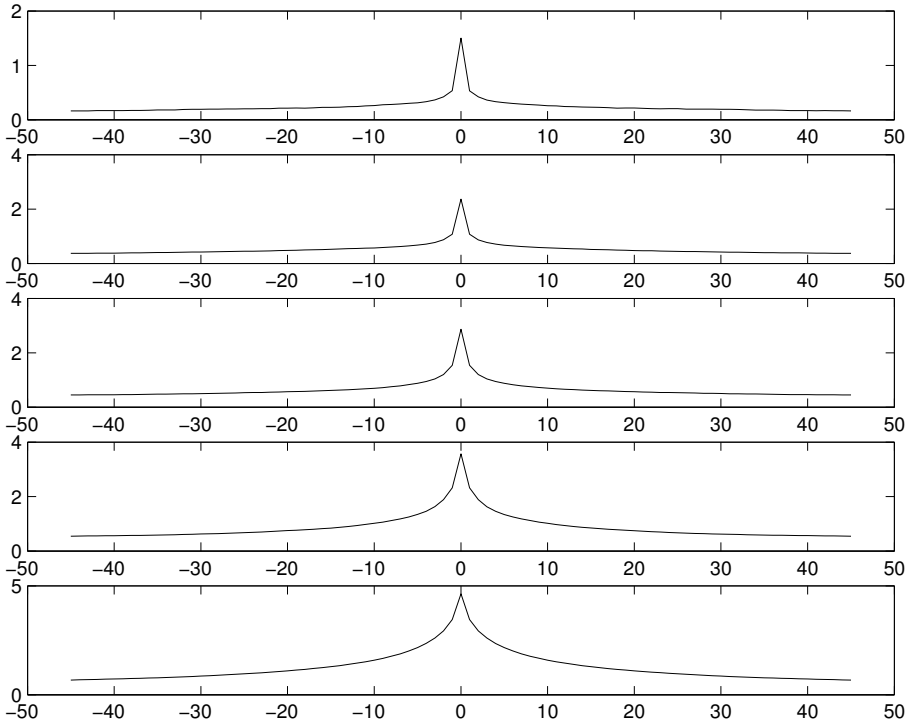

Figure 3.3: Profile of MI with rotation (degrees) at various scales (top: Finest and Bottom: Coarsest)

3.4 Multiscale Greedy Steepest Gradient Registration Technique

In this section, a novel multiscale optimization technique is proposed, which is suitable for achieving the robust image registration in multiscale edge representation. Traditionally

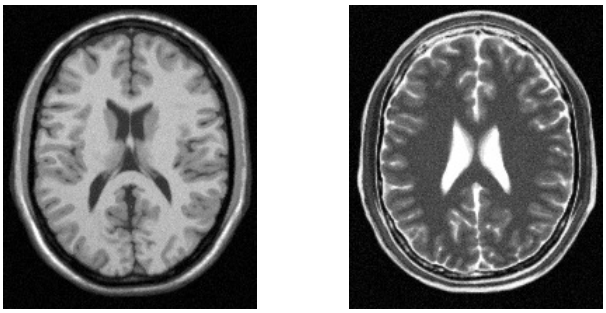

Figure 3.4: Images (T1 and T2) from different slices.

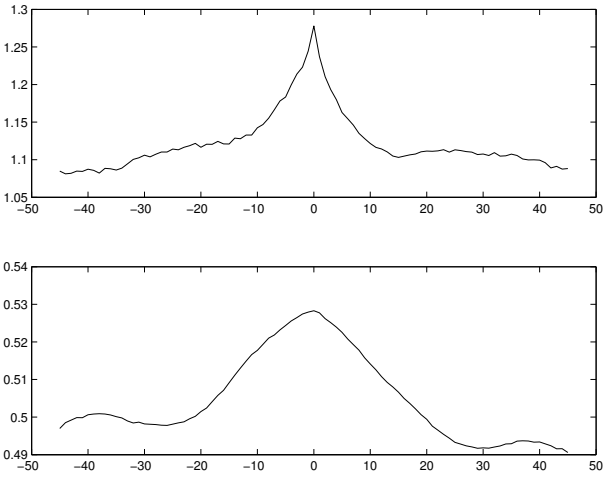

Figure 3.5: MI profile (between the images shown in Fig. 3.4 with rotations using 256 bins (top) and 4 bins (bottom) at coarsest scale.

Powell's optimization technique is used for MI based registration techniques [200]. However, we propose a variant of steepest gradient algorithm which is more robust and faster in multiscale domain. Our approach is similar to [203], where, approximate search for maximum correlation is done at coarse scale and refinement is done at finer scales. However, proposed approach is different than the previous techniques in several aspects. Firstly, we have proposed a dyadic search space (consisting of both rotation and translation) for each scale; and secondly, we have proposed a multiscale greedy steepest gradient optimization which does not require explicit gradient computation. Moreover, we compute MI using variable bin sizes for each level, thereby exploiting the multiscale wavelet decomposition.

Let Δ_D and Δ_R be the resolution in translation and rotation for the finest scale. For the level j we set resolution for translation $S_D^j = 2^{j-1}\Delta_D$ and $S_R^j = 2^{j-1}\Delta_R$ for rotation. If N_B is the number of bins to be used at finest scale then we set number of bins at level j as $S_B^j = N_B/2^{j-1}$. Let I_r^j be the magnitude of the wavelet decomposition of the reference image ($M_{2^j} f_r(x,y)$) at scale 2^j and similarly I_f^j be the magnitude of the wavelet decomposition of the floating image ($M_{2^j} f_f(x,y)$) at scale 2^j. If 2D rigid transformation of the floating image is denoted as $T(I_f^j, x, y, \theta)$ then mutual information between the two can be written as $H(I_r^j, T(I_f^j, x, y, \theta))$. Here x, y and θ are translations in X direction, Y directions and rotation applied to the rigid transformer, respectively. We have used nearest neighbor interpolation for speed considerations.

The optimization consists of two stages where first step is to search for the direction of the maximum difference in MI in the neighborhood. This difference can be written as,

$$\begin{aligned} \{x_m, y_m, \theta_m\} &= argmax\{H(I_r^j, T(I_f^j, \hat{x}, \hat{y}, \hat{\theta})) \\ &\quad -H(I_r^j, T(I_f^j, x, y, \theta))\} \end{aligned} \tag{3.4}$$

where $\hat{x} = \{x - S_D^j, x + S_D^j\}$, $\hat{y} = \{y - S_D^j, y + S_D^j\}$ and $\hat{\theta} = \{\theta - S_R^j, \theta + S_R^j\}$. The second stage is the greedy ascent in the direction of steepest (maximum) gradient without resorting to neighborhood search. If MI at the steepest direction $\{dx, dy, d\theta\}$ is $H(I_r^j, T(I_f^j, x_m, y_m, \theta_m))$ then evaluations of MI is done the same direction if $H(I_r^j, T(I_f^j, x_m + dx, y_m + dy, \theta_m + d\theta)) > H(I_r^j, T(I_f^j, x_m, y_m, \theta_m))$. In general, this difference can be written as,

$$\begin{aligned} \Delta H &= H(I_r^j, T(I_f^j, x + dx, y + dy, \theta + d\theta)) \\ &\quad - H(I_r^j, T(I_f^j, x, y, \theta)) \end{aligned} \tag{3.5}$$

This greedy movement in the steepest direction is continued until $\Delta H <= 0$. Subsequently, the search moves to the next level 2^{j-1} (utilizing I_r^{j-1} and I_f^{j-1}) till the neighborhood search reaches to a local maxima. The pseudo code is described below.

1. Initialization: Δ_D, Δ_R and N_B.
2. Compute $M_{2^j} f(x, y)$ for both reference image $f_R(x, y)$ and floating image $f_F(x, y)$ at each scale $j = 1...L$.
3. For each scale $j = L....1$,
   - Compute: S_D^j, S_R^j and S_B^j.
   - Set: $I_r^j = M_{2^j} f_r(x, y)$
   - Set: $I_f^j = M_{2^j} f_f(x, y)$
   - Do search according to (3.4).
   - Compute steepest direction $\{dx, dy, d\theta\}$.
   - Compute ΔH using (3.5).

- If $\Delta H <= 0$, Break.
- While $\Delta H > 0$,
  - Update $x = x + dx$, $y = y + dy$ and $\theta = \theta + d\theta$.
  - Compute ΔH using (3.5).
  - If $\Delta H <= 0$, Break.
- End While,

4. End For

The greedy step in (3.5) is not optimal, however it provides speed improvement. This algorithm works quite well due its multiscale nature. It should also be noted that the step sizes are precomputed for the different scales and are not changed during the course of optimization, which in turn provides simplicity for the implementation of the algorithm.

3.5 Registration Test Results

The test images are obtained from BrainWeb [220], a simulated MRI database at MNI, Montreal, Canada. This database provides a synthetic volume with several options of modalities, voxel sizes and noise levels. The size of the volumes are $181 \times 217 \times 181$ with voxel size of $1mm \times 1mm \times 1mm$. The multiscale decomposition is done at 5 levels ($L = 5$). The initial parameters are setup as $\Delta_D = 1$ pixel (both x, y directions), $\Delta_R = 0.5$ degrees and $N_B = 256$. The nearest neighborhood interpolation is used for the 2D rigid transformation for speed considerations.

The performance of the registration algorithm is tested on AMD64 Dual Core CPU with 2.8GHz and 3GB of RAM using MATLAB (R2008b) environment. Simulations are

done by rotating the floating image in the range of 0 to 20 degrees and shifting by 0 to 8 pixels in both x and y directions. The reason for small range in translation trails is due the fact that BrainWeb images does not have much empty spaces around. The performance of T1-PD registration under noisy conditions are shown in Table 3.1. Here, *Mean Absolute Error* (MAE) is used for the measurement of the registration accuracy.

Table 3.1: T1-PD Mean Absolute Error

%Noise	Translation		Rotation	
	Mean	Variance	Mean	Variance
%3	0.4703	0.4082	0.0456	0.0207
%5	0.7841	0.7127	0.3207	0.1610
%7	1.2457	2.2535	0.7693	1.7542
%9	1.8452	3.2284	0.9322	1.9132

Table 3.2: Comparison of Registration Accuracy

Initial parameters	Proposed	Nelder-Mead	Powell
0.1,0.1,0.1	0.1433	**9.4302**	0.2013
1,1,1	0.1333	**5.6246**	0.2554
5,6,2	0.1333	0.4745	0.2978
-5,-9,14	0.1333	0.2554	0.1993
-5.5,-0.5,-8.5	0.3667	**12.7806**	0.3002
7,-7,20	0.1333	0.3385	0.2373

Table 3.3: Comparison of Registration Speed (Seconds)

Initial parameters	Proposed	Nelder-Mead	Powell
0.1,0.1,0.1	6.75	**0.43**	9.72
1,1,1	5.7	**2.84**	9.0
5,6,2	4.48	2.98	13.93
-5,-9,14	5.43	5.47	8.30
-5.5,-0.5,-8.5	5.45	**1.88**	7.94
7,-7,20	5.43	5.11	10.6

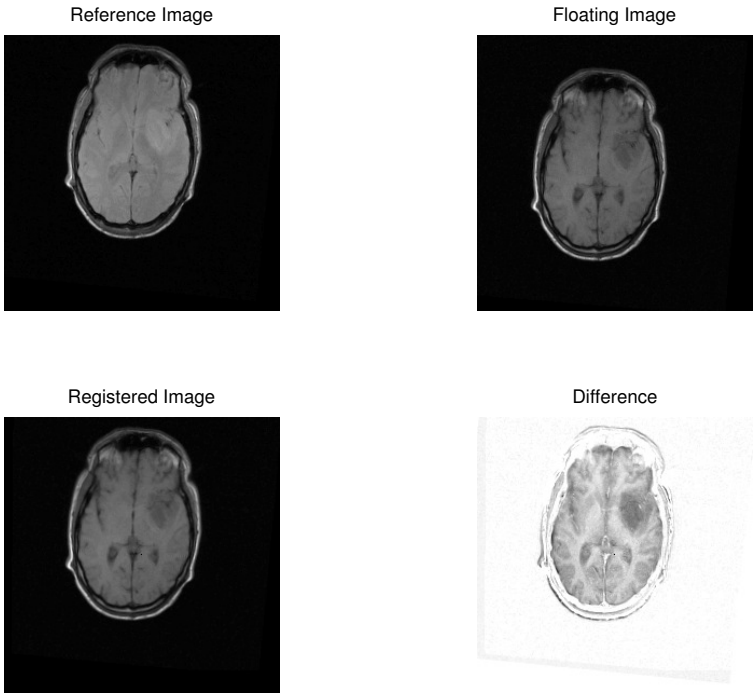

Figure 3.6: PD-T1 registration results.

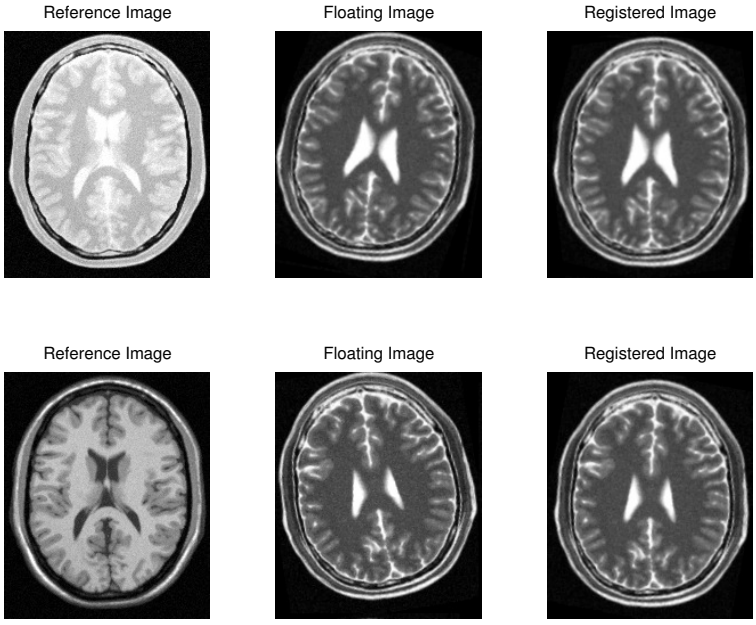

Figure 3.7: Multi-modal alignment results between different anatomical regions

In Figure 3.6, a T1 floating image against PD reference image is registered and the difference between the registered T1 image and the reference PD image is also shown. The accuracy and speed of the proposed "greedy" steepest ascent algorithm is compared against two well known optimization techniques, Nelder-Mead Simplex Method [221] and Powell's conjugate gradient method [222]. The same reference and floating image pair is used for this comparison using different initial search parameters and the accuracy and speed (seconds) are shown in the tables 3.2 and 3.3. The registration failures are indicated by entries with the bold font. In Table 3.2, it can be observed that Nelder-Mead method fails to register quite often and the accuracy of the proposed technique is consistently better than the other methods. In Table 3.3, it can be observed that the speed of the proposed technique is significantly better than Powell's method which is widely used for image registration tasks. It is worth mentioning that the Nelder-Mead and Powell methods are used to register images directly, without employing the wavelet decomposition.

Normally, affine and non-rigid (deformable) registrations techniques are more accurate than rigid registration techniques but are extremely slow. For example, demon-based affine registration technique, presented in [223], took around 7 seconds to register two images of 256×256 pixels. On the other hand, deformable registration proposed in [224], took 238 seconds to register two images of 256×256 pixels on the same platform.

Another feature of the proposed technique is the ability to align nearby areas, which is useful in image retrieval tasks. This requirement is not expected in the clinical image registration process. Multi-modal alignment results between different anatomical regions are shown in Figure 3.7.

3.6 Conclusions

In this chapter, a 2D rigid, registration scheme for image retrieval applications is presented. The main contribution in this chapter is the use of efficient multiscale representation in image registration. This representation is robust against noise and MR field inhomogeneity and facilitates efficient multimodal image registration. Another significant contribution is the development of multiscale registration scheme by exploiting the multiscale decomposition and suggesting appropriate step and bin sizes. We provided a multiscale greedy steepest gradient registration technique for the efficient implementation. The registration results using the proposed optimization technique are also compared against the two well-known optimization techniques. Simulation results demonstrate the high efficacy of the approach under multi-modal and noisy environments.

Chapter 4

Hierarchical Intersubject Multiscale Image Retrieval in 3D Brain Volumes

4.1 Introduction

Many of the proposed retrieval systems in the medical domain have been adopted from the techniques used in general image retrieval schemes. These techniques perform satisfactorily with the databases consisting of heterogeneous images of different modalities and anatomical regions. However, these systems are not suitable for the precise matching required in the retrieval of 2D images in 3D volumes. Figure 4.1 depicts this challenging problem, where even expert eye may fail. Only a couple of research papers [193, 194] have been reported regarding the 2D slice retrieval problem. In [193], Karhunen-Loeve transform is used for the retrieval of the relevant slice in the eigenimage domain. However, the technique required a computationally expensive registration and intensity normalization step. In [194], a computationally expensive preprocessing step is needed, where brain tis-

sues are removed from the skull and the mid-saggital plane is detected to compensate for the alignment. These preprocessing stages work well for the central portion of the brain, however, such preprocessing provides erroneous for other areas of the brain. Moreover, this technique has not been tested on multimodal scenarios. The proposed retrieval technique in this chapter employs an efficient registration technique, which does not requires any preprocessing stage [1].

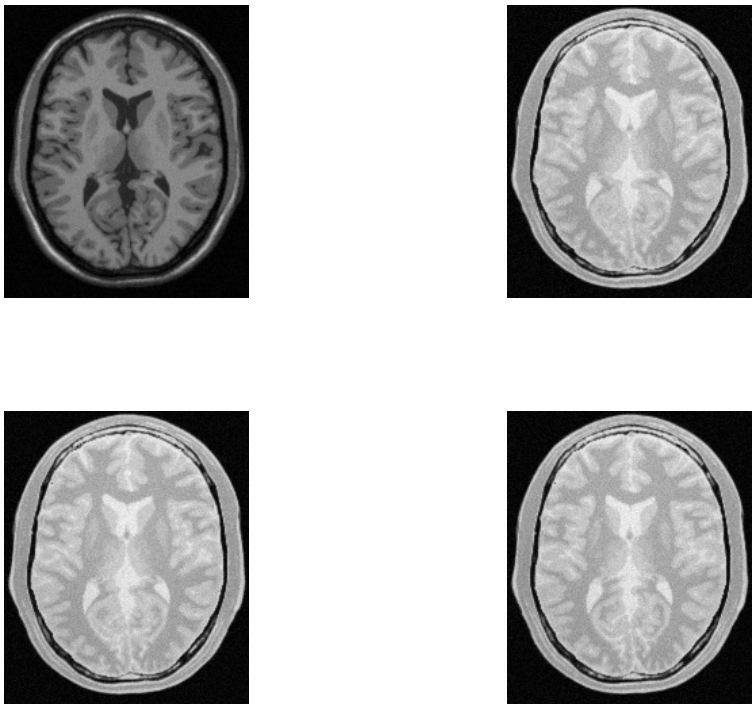

Figure 4.1: T1 query slice#82 (top left), PD slice#81 (top right), PD slice#82 (bottom left), PD slice#83 (bottom right)

In this chapter, a retrieval technique is proposed using fast 2D rigid registration, discussed in the previous chapter. The requirements for 2D slice retrieval can be listed as follows:

[1] A part of this work has been presented at the IEEE Toronto International Conference - Science and Technology for Humanity (TIC-STH 2009) in Toronto, Canada

- Medical images are multimodal and heterogeneous with temporal properties. Hence multimodal image registration is important for this retrieval application.
- The image registration should only allow 2D rigid (displacement and rotation) movements. The use of a shear or a scale component will not help, as consecutive slices are similar to each other.
- The registration technique should not be sensitive to signal variation across the image due to MR field inhomogeneity.
- This image retrieval application expects to align images from different, albeit nearby locations. This scenario is not expected in the general medical image registration.
- This image retrieval application demands a precise similarity metric.
- The retrieval time cannot be not more than a few minutes as the application will defeat its purpose.

This chapter is organized in three sections where the next section introduces two fast image retrieval techniques followed by results and conclusions.

4.2 Fast 2D Intersubject Slice Retrieval in 3D Volume

In the previous chapter, wavelet based 2D image registration was found to be robust and precise. In this chapter, a novel scheme is presented for intersubject image retrieval. Here, the term "intersubject" means that both the query image slice and the 3D volume belong to the same subject. In this section, mutual information (MI) is proposed as a metric for retrieving 2D slices from 3D volume in the multiscale domain. Here retrieval problem is

attempted by maximizing MI between query image and the candidate slices from target 3D volume. The profile of MI (after registration) between saggital query and saggital slices of 3D volume at various wavelet scales is shown in Figure 4.2. In this figure, query image is 100^{th} slice and the MI profile (in y-direction) is shown with slice index of the target 3D volume (in x-direction). Since 2D correlation is an established technique for pattern recognition, the profile of correlation between saggital query and saggital slices of 3D volume at various wavelet scales is shown in Figure 4.3. From this figure, it can be concluded that 2D correlation is not a suitable metric for the wavelet based multimodal image retrieval. It can be seen in Figure 4.2 that the MI maximizes near the true result.

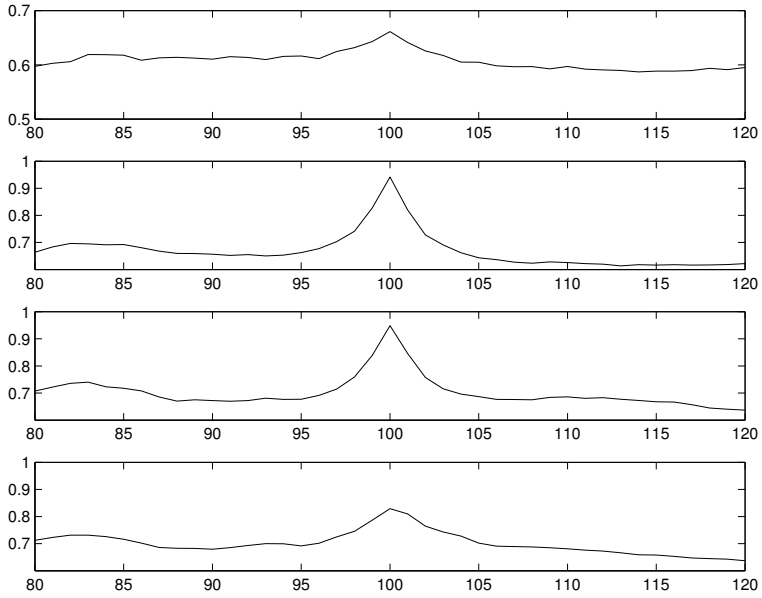

Figure 4.2: Profile of MI between saggital query (100^{th} slice) and all the slices of 3D volume at various wavelet scales

Here, a multiscale retrieval technique is proposed, which involves multiple 2D rigid registration attempts between the query slice and candidate slices in the target volume. In order to speed-up the retrieval process, the 2D rigid registration are attempted with the candidate slices at large slice interval in coarse scale. The candidate slice interval is

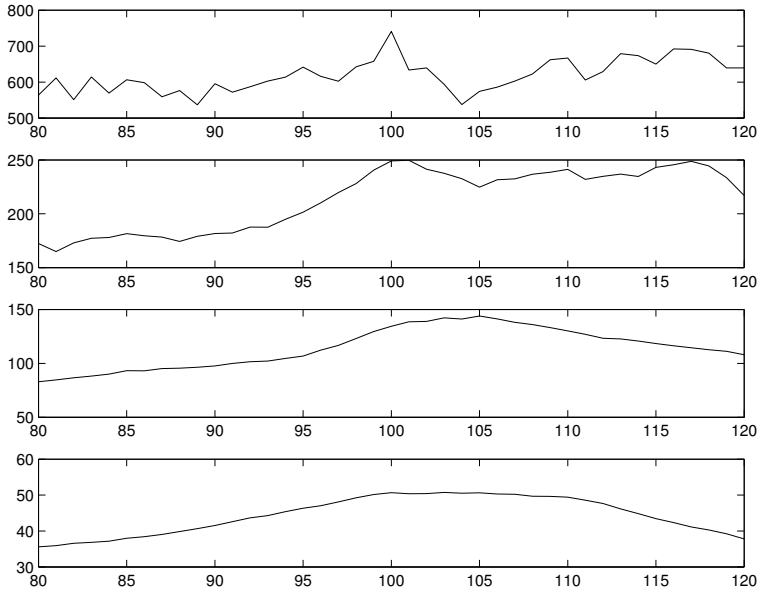

Figure 4.3: Profile of 2D correlation between saggital query and slices of 3D volume at various wavelet scales

reduced progressively as the retrieval process moves in finer scales. Let $\Delta_S = 1$ be the slice step size for the finest scale. For the level j, the step size is set at $R_S^j = 2^{j-1}\Delta_S$. The slice index search range at next fine scale j is set as $[R_S^{j+1} - R_S^j,\ R_S^{j+1} + R_S^j]$. Here, R_S^{j+1} is search results from previous scale. At coarsest scale, the algorithm does search until $(H_i - H_{max})/H_i > 0.4$ with slice index step size of R_S^j. Here H_i is the MI between the query image and the i^{th} candidate image slice in the target volume after the registration. The peak of MI during the retrieval process, in the current scale, is H_{max}. The slice index at which MI is maximum is used to start the search at the next finer scale.

4.2.1 Retrieval using Full Registration at All Scales (FRAS)

In this technique, full registration (involving decompositions from all the scales) is done at larger interval of the slice indices at coarse scale and the slice interval is reduced as the

search progressed in the finer scales. Let $\Delta_S = 1$ be the slice step size for the finest scale. Then for the level j the step size is set at $R_S^j = 2^{j-1}\Delta_S$. The slice index search range at the next fine scale j is only $[k^{j+1} - R_S^j,\, k^{j+1} + R_S^j]$. Here, k^{j+1} is the search results from the previous scale. Let I_r^{ij} be the magnitude of the wavelet decomposition of the i^{th} candidate image among N, 2D slices ($M_{2^j} f_{ri}(x, y)$) at scale 2^j and similarly I_q^j be the magnitude of the wavelet decomposition of the query (floating) image ($M_{2^j} f_q(x, y)$) at scale 2^j. The registration process between I_q^j and I_r^{ij} can be described as,

$$< H_{ij}, t_{ij} >= R_{C_i}^{s_{ij}} \tag{4.1}$$

where $C_i = \{I_q^j, I_r^{ij}\}$ are the pairs used for image registration using s_{ij} as starting search parameter vector. Outputs of the registration process are MI (H_{ij}) and the solution vector (t_{ij}).

At coarsest scale the algorithm performs the search in the whole 3D volume with slice index step size of R_S^j, where registration is attempted between query image and candidate slices in the target 3D volume. Here, H_{iL} is the MI between the registered slices. Due to the symmetry between the right and the left brain hemispheres in saggital view, the retrieved result will contain slices from either of the two hemispheres. Consequently, the algorithm tries to detect two peaks in MI profile after the registration at coarsest scale. If the top two MI measurements after registration are H_{mL} and H_{nL} and corresponding slice indices are m and n then algorithm considers n as a valid index to be searched according to the following criterion.

$$DF > 22 \;\&\; \{(m < MD \;\&\; n > MD) \mid (m > MD \;\&\; n < MD)\} \tag{4.2}$$

where $DF = 100 \times (H_{mL} - H_{nL})/H_{mL}$ and MD is the index of the mid-slice of the volume. The value of 22 is selected after extensive simulations. If the equation 4.2 is satisfied then algorithm searches at slice index at n along with m at finer scales. In the following, pseudo code is provided.

1. Initialization: s_{1L}=[0,0,0], Δ_S, $MaxMI = 0$, $k2Flag = 0$.
2. Compute I_q^j for query image and all 2D slices of target volume I_r^{ij} at each i and j.
3. For $j = L....1$,
   - Compute: R_S^j.
   - If $j = L$,
     - Register I_q^j with I_r^{ij} slice with index in steps of R_S^j using $s_{1L} = [0, 0, 0]$ as initial search parameter.
     - $k^L = sort(< H_{iL} >,' descend')$
     - $m = k^L(1)$ and $n = k^L(2)$
     - Compute $DF = 100 \times (H_{mL} - H_{nL})/H_{mL}$
     - If Eqn. 4.2 is TRUE
     - Set $k2Flag = 1$
     - Endif
   - Else,
     - Register I_q^j with I_r^{ij} slice index m in step of R_S^j only in the range of $[k^{j+1} - R_S^j, k^{j+1} + R_S^j]$.
     - $k^j = argmax(< H_{ij} >)$

- Set $m = k^j$

- End if,

4. End For,

5. m is the first solution.

6. If $k2Flag == 1$,

   - For $j = L - 1....1$,

     - Compute: R_S^j.

       * Register I_q^j with I_r^{ij} slice index n in step of R_S^j only in the range of $[k^{j+1} - R_S^j,\ k^{j+1} + R_S^j]$.

       * $k^j = argmax(< H_{ij} >)$

       * Set $n = k^j$

     - End if,

   - End For,

   - n is the second solution.

7. End if,

A description of the retrieval scheme for $L = 3$ for two levels (3 and 2) is shown in Figure 4.4. In this figure, the WTM block computes magnitude of the wavelet transform for query image. The processing for the finest scale $j = 1$ is identical to the processing for scale $j = 2$ and hence not shown here. Here, it should be noted that index step size R_S^j is 4 and 2 for levels 3 and 2, respectively. In order to speedup the retrieval process, the

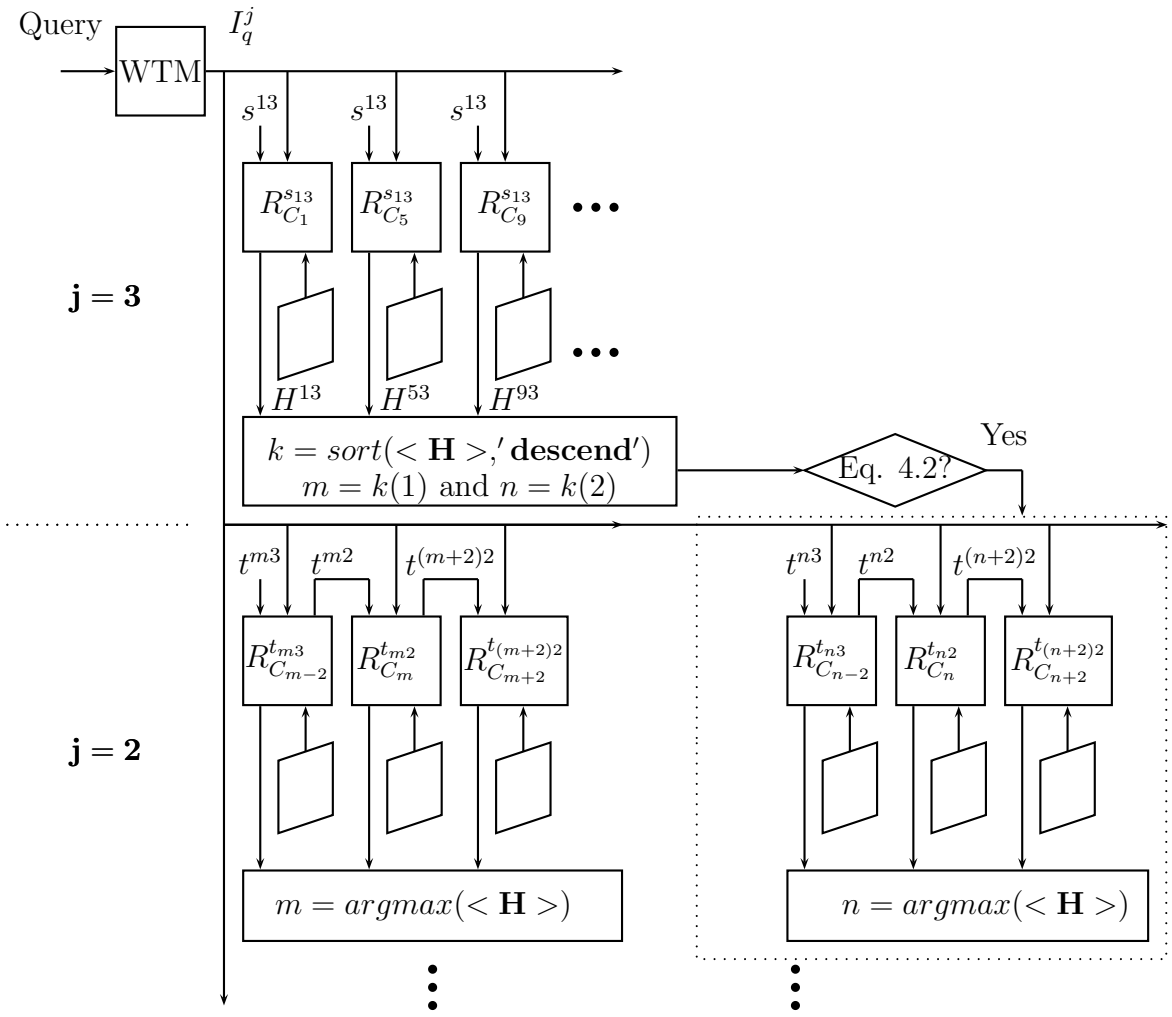

Figure 4.4: Description of $FRAS$ retrieval scheme for $L = 3$ for $j = \{3, 2\}$

initial transformation parameters for the registration among successive slices are set to the previous registration result.

Figure 4.5 shows a retrieval result using a 2D slice of MR T1 sequence as query image from BrainWeb [220] image dataset. Both query and 3D target volume are taken from 9% noise-level (noise relative to the brightest tissue). The retrieval is done on a MR Proton Density (PD) volume and the corresponding retrieved PD slice is also shown.

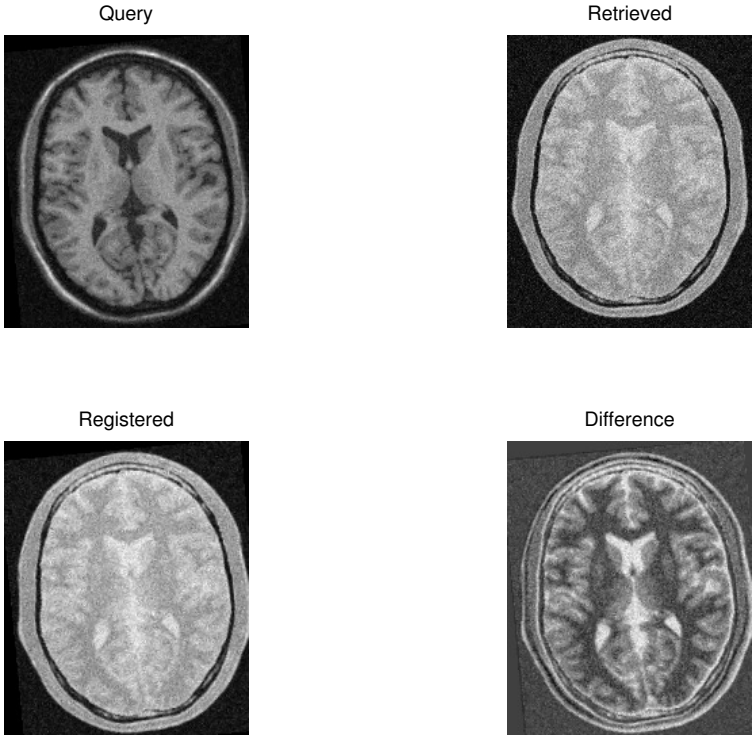

Figure 4.5: A T1 query slice, retrieved corresponding PD slice (from 3D, PD volume) with 9% noise-level. The bottom row shows the registration result.

In order to evaluate the retrieval performance, the *Mean Absolute Retrieval Error* (MAE) for N retrieval trails is defined as,

$$E_{MAE} = \frac{\sum_{i=1}^{N} min\{|S_m(i) - S_t(i)|, |S_n(i) - S_t(i)|\}}{N} \tag{4.3}$$

where S_m and S_n are the indices of the retrieved slice and S_t true (correct) retrieval slice.

Here, the MAE metric effectively computes the average retrieval error in the retrieved slice indices. Metrics based on squared error are not suitable because they cannot directly relate the error with slice index. Performance of the algorithm is tested on AMD64 Dual Core CPU with 2.8GHz and 3GB of RAM using MATLAB (R2008b) environment. In Table 4.1 the retrieval performance results are provided for retrieval of the T1 slices in PD 3D volume under various noise conditions. Since BrainWeb [220] images do not have much empty space, translations could not be employed. Consequently, all the query images are pre-rotated randomly in the range of 0 to 10 degrees. The retrieval is tested under various noise conditions using the 3D volumes available at BrainWeb [220] website. This database provides one synthetic volume with several options to choose regarding modalities, voxel sizes and noise levels. The size of the volumes are $181 \times 217 \times 181$ with voxel size of $1mm \times 1mm \times 1mm$. In order to limit the number trails, both the query and the target 3D volumes are used from same noise-level. In all the simulations, results show $E_{MAE} < 2$, indicating that the retrieval error, on the average, is limited to one slice. In Table 4.2 the retrieval timings are provided for the same set of simulations. In Table 4.3, performance is shown for multimodal scenario among the transverse slices with 9% noise-level.

Table 4.1: Mean Absolute Retrieval Error T1-PD under noise (FRAS)

Mode	$E_{MAE}(3\%)$	$E_{MAE}(5\%)$	$E_{MAE}(7\%)$	$E_{MAE}(9\%)$
Transverse	0.1294	0.5529	0.7529	0.9294
Saggital	0.0000	0.0993	0.3901	0.7234
Coronal	0.4241	0.6073	0.8901	1.1466

Table 4.2: Mean Retrieval Time (seconds) T1-PD under noise (FRAS)

Mode	$E_{MAE}(3\%)$	$E_{MAE}(5\%)$	$E_{MAE}(7\%)$	$E_{MAE}(9\%)$
Transverse	111	108	105	102
Saggital	127	121	122	124
Coronal	127	129	125	119

Table 4.3: Multimodal Performance with 9% noise (Transverse) (FRAS)

Mode	E_{MAE}	Mean Retrieval Time (Sec.)
PD - T1	1.2765	105
T1 - PD	0.9294	102
T2 - T1	0.7647	97
T1 - T2	1.9353	104
PD - T2	0.9529	101
T2 - PD	0.3588	96

4.2.2 Retrieval using Partial Registration at Each Scale (PRES)

This technique uses a registration, that requires coarse optimization step sizes and utilizes coarse level wavelet decomposition, while it iteratively uses finer registration for the retrieval in finer scales. This approach is extremely rapid as compared to the previous retrieval algorithm (FRAS). Let I_r^{ij} be the magnitude of the wavelet decomposition of the i^{th} candidate image among N, 2D slices ($M_{2^j} f_{ri}(x, y)$) at scale 2^j and similarly I_q^j be the magnitude of the wavelet decomposition of the query (floating) image ($M_{2^j} f_q(x, y)$) at scale 2^j. Then the partial registration at level 2^j between I_q^j and I_r^{ij} can be described as,

$$< H_{ij}, t_{ij} >= P_{jC_i}^{s_{ij}} \tag{4.4}$$

where $C_i = \{I_q^j, I_r^{ij}\}$ are the pairs used for registration using s_{ij} as starting search parameter vector. Output of the registration process are H_{ij} (MI) and solution vector t_{ij}. The partial registration P_j is achieved by registering wavelet decompositions from a given

level. The pseudo code of the registration algorithm can be described as follows.

1. Initialization: Δ_D, Δ_R, N_B and L.
2. Compute $M_{2^j} f(x, y)$ for both reference image $f_R(x, y)$ and floating image $f_F(x, y)$ only at scale $j = L$.
3. Set $j = L$,
   - Compute: S_D^j, S_R^j and S_B^j.
   - Set: $I_r^j = M_{2^j} f_r(x, y)$
   - Set: $I_f^j = M_{2^j} f_f(x, y)$
   - Do search according to (3.4).
   - Compute steepest direction $\{dx, dy, d\theta\}$.
   - Compute ΔH using (3.5).
   - If $\Delta H <= 0$, Break.
   - While $\Delta H > 0$,
     - Update $x = x + dx$, $y = y + dy$ and $\theta = \theta + d\theta$.
     - Compute ΔH using (3.5).
     - If $\Delta H <= 0$, Break.
   - End While,
4. Finish

Similar to equation 4.1, the partial registration process between I_q^j and I_r^{ij}, at level 2^j, can be described as,

$$< H_{ij}, t_{ij} >= R_{jC_i}^{s_{ij}} \tag{4.5}$$

where $C_i = \{I_q^j, I_r^{ij}\}$ are the pairs used for registration using s_{ij} as starting search parameter vector. Output of the registration process are H_{ij} (MI) and solution vector t_{ij}.

The retrieval algorithm starts with coarsest scale L (lowest resolution) utilizing coarse registration between the wavelet decompositions at level 2^L and subsequently moves towards the finer scale. At coarsest scale, the algorithm does search the whole 3D volume with slice index step size of R_{LS}^j, where partial registration is done between the wavelet decompositions of query image and the 2D slices in the target 3D volume. Here H_{iL} is the MI between the registered slices. Similar to the FRAS technique, the algorithm tries to detect two peaks in MI profile after the registration at the coarsest scale. If the top two MI measurements after registration are H_{mL} and H_{nL} and corresponding slice indices are m and n then algorithm considers n as valid index to be searched according to the criterion in 4.2. Since this approach is fast we employed tri-linear interpolation (as opposed to nearest neighborhood in FRAS) for the registration.

In Table 4.4, the retrieval performance is provided for the retrieval of T1 slices in PD volume under various noise conditions. Here, all the query images are pre-rotated by randomly in the range of 0 to 10 degrees. The simulation results show that $E_{MAE} < 2$, indicating that the retrieval error in the proposed technique, on the average, is limited to one slice. In Table 4.5, the retrieval timings are provided for the same experiment. In Table 4.6, performance are shown in multimodal (T1, T2 and PD) scenario with transverse slices under 9% noise-level. The retrieval accuracy is better than FRAS due to the use of tri-linear interpolation in the registration process. This technique is generally 5 times faster than the FRAS technique.

Table 4.4: Mean Absolute Retrieval Error T1-PD under noise (PRES)

Mode	$E_{MAE}(3\%)$	$E_{MAE}(5\%)$	$E_{MAE}(7\%)$	$E_{MAE}(9\%)$
Transverse	0.0294	0.3235	0.5235	0.7412
Saggital	0.0000	0.0284	0.3191	0.5532
Coronal	0.3403	0.5026	0.5079	1.0209

Table 4.5: Mean Retrieval Time (seconds) T1-PD under noise (PRES)

Mode	$E_{MAE}(3\%)$	$E_{MAE}(5\%)$	$E_{MAE}(7\%)$	$E_{MAE}(9\%)$
Transverse	22	22	20	20
Saggital	27	25	25	25
Coronal	25	23	22	21

Table 4.6: Multimodal Performance with 9% noise (Transverse) (PRES)

Mode	E_{MAE}	Mean Retrieval Time (Sec.)
PD - T1	0.8529	19
T1 - PD	0.7412	20
T2 - T1	0.7353	19
T1 - T2	1.6176	22
PD - T2	1.4882	20
T2 - PD	0.2412	19

4.3 Conclusions

In this chapter, a novel multimodal, multiscale, wavelet-based technique is proposed for 2D slice retrieval in 3D Magnetic Resonance (MR) brain volumes. Here the main contribution is the use of a multiscale registration scheme for retrieval of MR images. The target application domain for the proposed retrieval scheme is diagnostic and decision support, which is another contribution. For efficient retrieval of relevant 2D slices two novel MI based search schemes are presented. The first retrieval technique (FRAS) utilizes full registration, utilizing all decomposition levels, at each level of retrieval. Whereas, the second technique (PRES) starts with coarse registration using a coarse level decomposition and iteratively utilizes finer registration using finer levels of decompositions. We have found that the PRES technique is generally five times faster than the FRAS approach, while maintaining similar error rates. The proposed retrieval algorithms are tested under noisy conditions and the experiments show promising results.

Chapter 5

Semantic Assisted Intersubject Image Retrieval in 3D Brain Volumes

5.1 Introduction

In the previous chapter, a technique for wavelet based image retrieval in 3D MR brain volumes was presented. In this chapter, a novel semantic assisted scheme is presented for intersubject retrieval. Here, the term "intersubject" means that both the query image slice and the 3D volume belong to the same subject. This technique essentially shortens the retrieval time by first associating the incoming query image to a specific area of the brain and then executing the search in that limited area. This specific area of the brain is termed as "Semantic Region", a term inspired by the lobes of the human brain. The technique presented in this chapter associates the query image with one of the semantic regions, while speeding up the retrieval time simultaneously. Here, the first stage of associating the query image to a particular semantic region is termed as "classification" to distinguish it from

the retrieval part. The proposed technique not only retrieves 2D slice from 3D volumes but also classifies the 2D slice to a semantic region as well, which is the major contribution in this chapter.

The rest of this chapter is organized as follow. The second section introduces the semantic regions in the human cerebral cortex; the third section briefly introduces support vector machines (SVM); the fourth section provides the implementation details of the semantic classification in the retrieval scenario; the fifth section presents the overall strategy for the retrieval, which is supported by extensive simulation results for both healthy and MS patients. Finally, the sixth section concludes this chapter.

5.2 Semantic Regions in Human Cerebral Cortex

The anatomy of the human cerebral cortex is studied thoroughly by dividing it into four major lobes. Figure 5.1 depicts these lobes [225]. Here, it is obvious that these lobes are highly irregular and hence extremely difficult for automatic parcellation. However, there are some semi-automatic techniques developed for clinical research, such as in [226].

Figures 5.2, 5.3 and 5.4 show the percentage amount of brain matter belonging to various lobes in transverse, coronal and saggital views, respectively. The use of millimeters (instead of pixels) ensures consistency in measurements with different scanner resolutions.

Obviously, the brain lobes are highly irregular and overlap considerably and hence a 2D image slice in a specific view can be associated to more than one lobe simultaneously. Moreover, in the saggital view, various lobes fully overlap each other due the symmetry between left and right hemispheres of the brain. Consequently, the semantic classification of the 2D images slices in this view is not attempted.

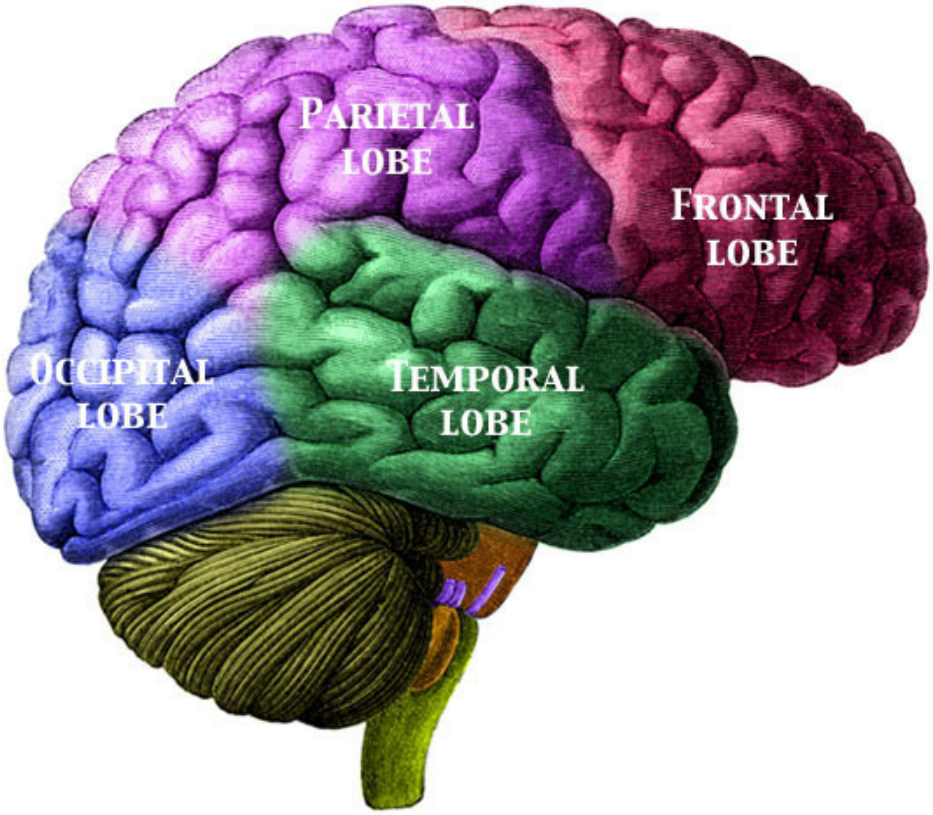

Figure 5.1: Cerebral cortex of the human brain showing the 4 major lobes

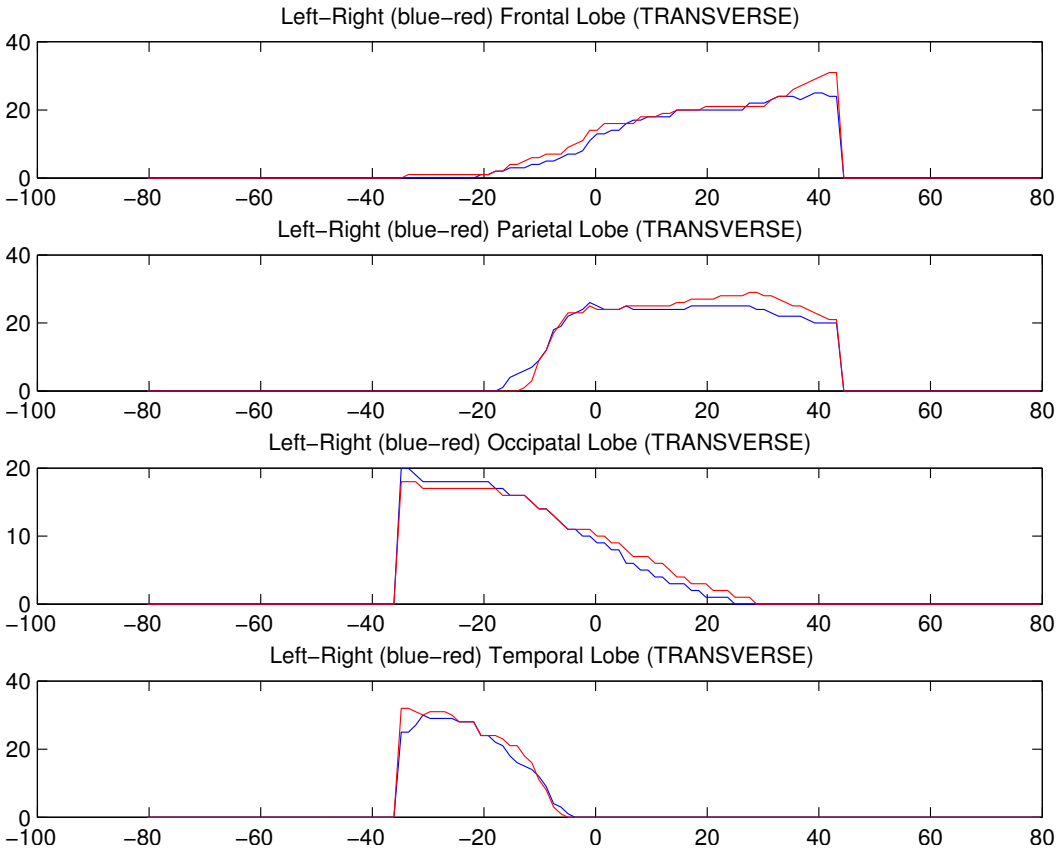

Figure 5.2: Percentage lobe volumes in Transverse view

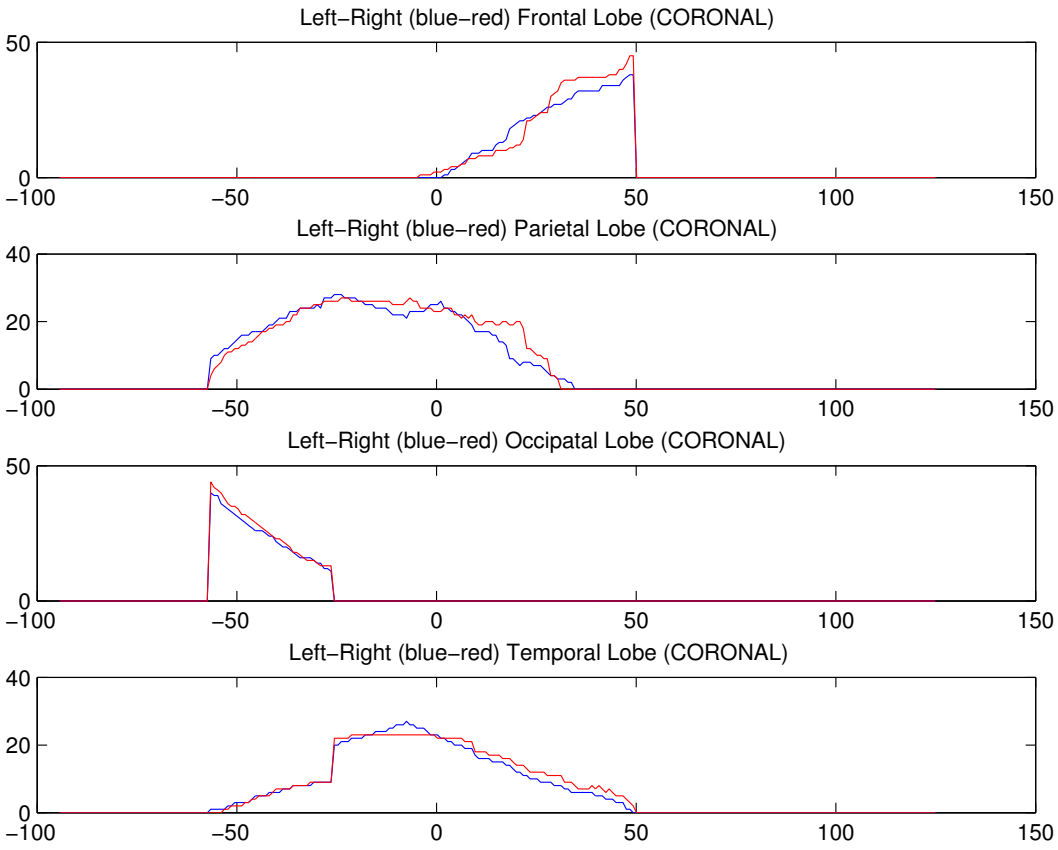

Figure 5.3: Percentage lobe volumes in Coronal view

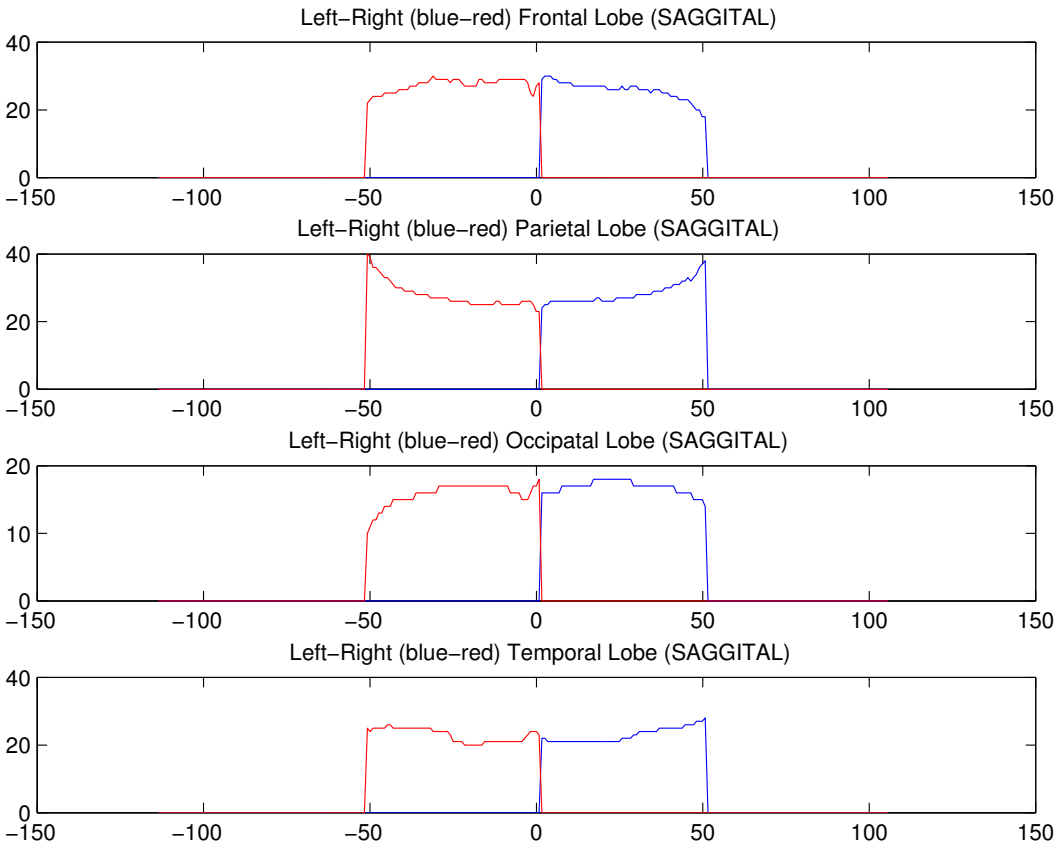

Figure 5.4: Percentage lobe volumes in Saggital view

5.3 Introduction to Support Vector Machines (SVM)

Support Vector Machine (SVM) is a popular technique among non-parametric learning techniques. This technique was first introduced by Boser [227] and Vapnik [228] in 1992 and 1998, respectively. Though, there are several known implementations of SVM available in the research community, in this work "LIBSVM" [229] is used exclusively. This library is the most popular implementation of SVM available in public domain.

A support vector machine constructs a hyperplane (or a set of hyperplanes) in a hyper-dimensional space, which is used for classification and regression tasks. A good separation is achieved by the hyperplane which has the largest distance to the nearest training data points of any given class because the larger the margin the lower the generalization error of the classifier. In the following subsections, an introductory treatment from [229] is provided for the sake of completeness and more in depth treatment can be found in [228].

5.3.1 C-Support Vector Classification (C-SVC)

Given the training vectors $x_i \in R^n, i = 1, 2, ..., l$, in two classes and a vector $y \in R^l$ such that $y_i \in \{1, -1\}$, C-SVC solves the following problem:

$$\begin{aligned} \underset{w,b,\xi}{min} \quad & \tfrac{1}{2} w^T w + C \textstyle\sum_{i=1}^{l} \xi_i \\ subject\ to \quad & y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \end{aligned} \tag{5.1}$$

Its dual is

$$\begin{aligned} \min_{\alpha} \quad & \tfrac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ \textit{subject to} \quad & y^T\alpha = 0 \\ & 0 \le \alpha_i \le C, \quad i = 1, ..., l \end{aligned} \tag{5.2}$$

where e is the vector of all ones, $C > 0$ is the upper bound, Q is an $l \times l$ positive semidefinite matrix, $Q \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is the kernel. Here training vectors x_i are mapped into a higher dimensional space by the function ϕ. The decision function is

$$sgn\left(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x_j) + b\right) \tag{5.3}$$

Traditionally, an SVM solves two class problem, however for this work multiclass SVM is needed as there are more than two semantic lobes in the brain. In the following subsection, an extension to the multiclass SVM is introduced.

5.3.2 Multiclass SVM

SVM was originally designed for binary classification problems. However, when dealing with several classes one needs an appropriate multi-class method. As two-class or binary classification problems are much easier to solve, a number of methods have been proposed for its extension to multi-class problems [230, 231].

Here aim is to assign labels to the instances by using support vector machines, where the labels are drawn from a finite set of elements. One way to achieve this is to reduce the single multiclass problem into multiple binary classification problems. Each of these

problems produces a binary classifier, which is supposed to produce an output function that gives relatively higher values for samples from the positive class and relatively small values for the samples belonging to the negative class. There are generally two methods to build such binary classifiers, where each classifier discriminates (i) between every pair of classes (one-versus-one) or (ii) one of the labels to the rest (one-versus-all). The first method uses $L(L-1)/2$ binary classifiers for L number of classes, each of which provides a partial decision for classifying a data point. During the testing of a feature x, each of the $L(L-1)/2$ classifiers votes for one class. The winning class is the one with the largest number of accumulated votes. On the other hand, one against the others method compares a given class with all the others put together. This basically constructs L hyperplanes where each hyperplane separates one class from the other classes. In this way, it generates L decision functions and an observation x is mapped to a class with the largest decision function. In [230], it is shown that the one-versus-one strategy is more suitable for practical reasons.

In LIBSVM, one-versus-one approach is adopted which is proposed by Knerr et. al. [232]. For training data from the ith and jth classes, two-class classification problem is solved as follows,

$$\begin{aligned} \underset{w^{ij},b^{ij},\xi^{ij}}{min} \quad & \tfrac{1}{2}\left(w^{ij}\right)^T w^{ij} + C\sum_{i=1}^{l}\left(\xi^{ij}\right)_t \\ subject\ to \quad & \left(w^{ij}\right)^T \phi(x_t) + b^{ij}) \geq 1 - \xi_t^{ij},\ if\ x_t\ in\ the\ i^{th}\ class, \\ & \left(w^{ij}\right)^T \phi(x_t) + b^{ij}) \leq -1 + \xi_t^{ij},\ if\ x_t\ in\ the\ j^{th}\ class, \\ & \xi_t^{ij} \geq 0. \end{aligned}$$

A voting strategy is used in the classification process, where each binary classification is

considered to be a voting process where votes can be cast for all the data points $\mathbf{x}$. Finally, a data point x_i is designated to be in a class with maximum number of votes. When two classes have equal amount of votes, the smallest index is selected. A detailed comparative study on the selection criterion can be found in [230].

5.4 Semantic Classification of Brain Lobes

In this section, a novel technique is introduced for the classification of 2D images into various semantic lobes. This technique involves three steps as follows:

1. Feature Extraction: A translation and rotation invariant feature set is extracted in the wavelet domain.

2. Training: A multi-class SVM based training is used to produce a model.

3. Classification: A multi-class SVM model, produced in the previous stage, is used for associating the incoming image slice to a specific semantic region.

Justifications for the use of SVM in this work is provided in appendix C.

5.4.1 Feature Extraction in Multiscale domain

In the absence of any pre-registration, it is imperative to have translation and rotation invariant feature set. There are lots of feature extraction techniques reported in the literature, however only a few can be applied in wavelet based mutliscale domain. Most of the feature extraction techniques produce a large feature set, which are cumbersome for training and matching. Many feature extraction techniques require computationally expensive

preprocessing steps. In this work, a fast, novel and simple feature extraction technique in wavelet domain is proposed.

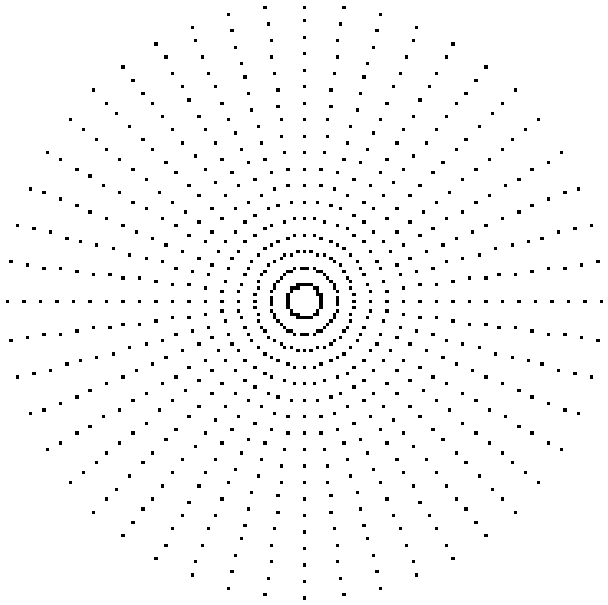

Figure 5.5: Feature map

Figure 5.5 depicts a pre-defined feature map where each point indicates the sample location of pixel in $M_{2^j}f(x,y)$ (defined in eqn. 3.1). The feature set consists of grey scale values of $M_{2^j}f(x,y)$, sampled circularly starting from the outer most circle. This scheme samples the wavelet transform magnitude more densely near the center which is a desirable property for MR brain images. This kind of sampling has origin in log-polar transformation used in RF coil design for Nuclear Magnetic Resonance (NMR) [233]. The Fourier transform of this sampling scheme contains Bessel functions of the order of samples taken in the N angular directions as follows,

$$\begin{aligned} S_N(X,Y) = \quad & \int_0^{r_{max}} \sum_{n=-\infty}^{\infty} i^n J_n(2\pi r R) e^{in\Theta} r dr, \\ & n = 0, \pm 2N, \pm 4N...., \end{aligned} \tag{5.4}$$

Derivation of this equation is quite involved and is presented in appendix B. The presence of Bessel functions of various order captures a range of frequencies by the proposed feature map. There are two components of this feature map, viz. angular interval and radial interval. The angular interval is at $2\pi/48$ radians and radial interval in 5 mm. Finally, each image produces a feature vector of $48 \times 18 = 864$ length.

There is one condition in the proposed feature map for achieving translation and rotation invariance. A "true" translation and rotation invariance can be achieved only if the center of the feature map coincides with the center of object in the image. This is achieved by automatically detecting the center of the object by thresholding the incoming image and averaging the rows and columns positions separately. In practice, the feature extraction technique is not sensitive to this threshold and in all the simulations a threshold value of 25 is used.

The proposed feature extraction scheme is compared against one of the best known feature extraction technique, namely "SIFT" [234]. To compare the proposed feature extraction scheme with "SIFT", a 2D query image from PD sequence is used to retrieve similar slice in a 3D volume of T2 sequence. The SIFT technique is applied without any preprocessing of the images and without any modifications to the original algorithm. Matching is done using the compiled executable from the original author. The proposed feature extraction technique is applied to the $M_{2^4} f(x, y)$, i.e. the coarsest scale. For the proposed scheme matching is done using simple Euclidean distance metric. Figure 5.6 shows the retrieval performance between the proposed extraction technique (top) and the SIFT (bottom) feature extraction.

Obviously, performance is much better in case of proposed algorithm. Moreover, retrieval time (feature extraction and matching) in case of proposed algorithm is 350 times

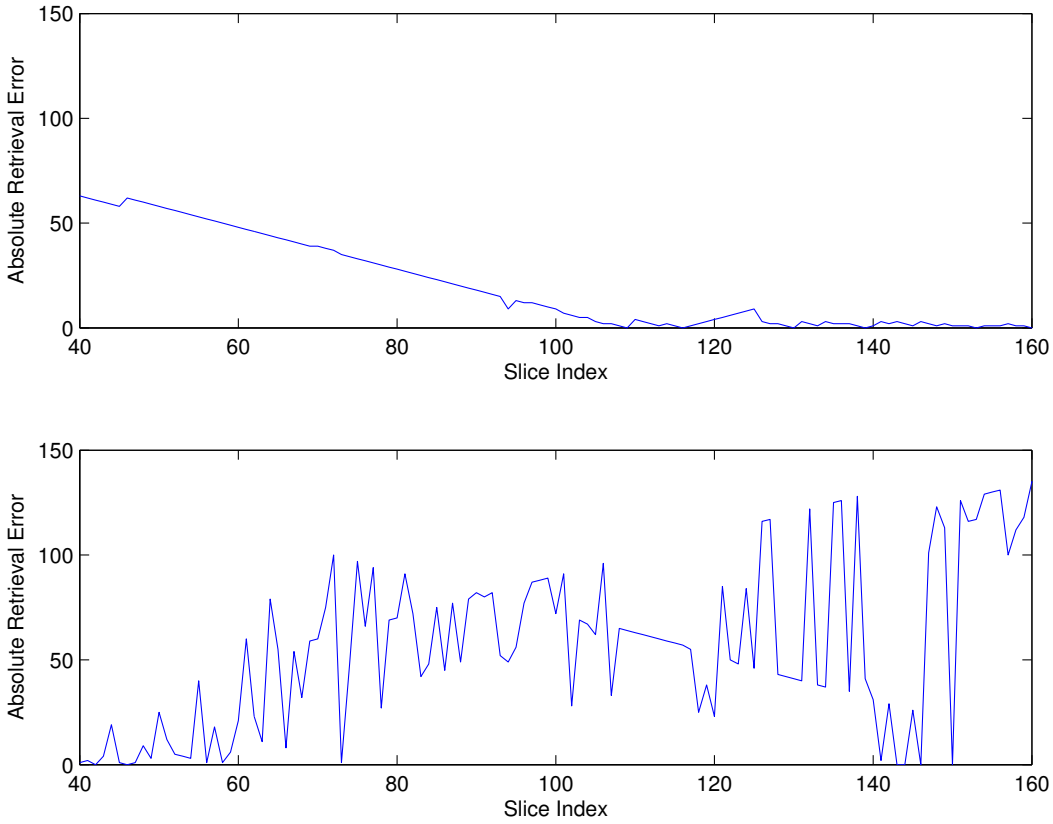

Figure 5.6: Absolute Error comparison between proposed (top) and the SIFT (bottom) feature extraction scheme.

faster than SIFT algorithm. This clearly shows that the proposed feature extraction algorithm is suitable for the retrieval of multi-modal images in the wavelet domain.

5.4.2 SVM based Training of Semantic Classes

Here, the main idea is to divide the 3D MR brain volume into semantically distinct regions. The brain lobes have irregular shapes and overlap significantly in transverse and coronal perspectives. Since SVM based classification requires distinct class labels, one feature vector cannot be associated with more than one class. Consequently, a new term *Semantic Region* is introduced to differentiate it from brain lobe. Though, *Semantic Region* is inspired from brain lobe, it is a non-overlapping and well defined region when viewed from a single perspective.

Based on Figure 5.2, the three semantic regions are shown in the Figure 5.7 and are defined in Table 5.1 for transverse view. Similarly, Figure 5.8 shows the semantic regions in coronal view and the Table 5.2 defines the training classes for coronal view based on Figure 5.3. In these tables the reference (origin) is a the center of 3D brain volume. The center of the 3D volume is found by averaging the 3D coordinates of voxels belonging to the gray-level greater than a threshold. In all the simulations a threshold value of 25 is used.

Table 5.1: Semantic Regions in transverse view

Semantic Region	Start (mm)	End (mm)
Temporal	first	-5
Occipatal	-6	24
Frontal	25	last

It can be observed from Figure 5.4 that all the lobes overlap between right and left

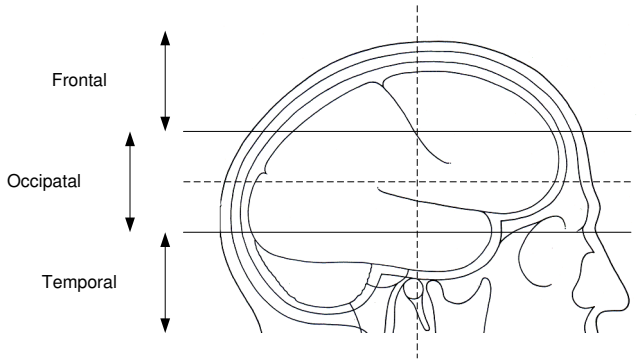

Figure 5.7: Semantic Regions in transverse view

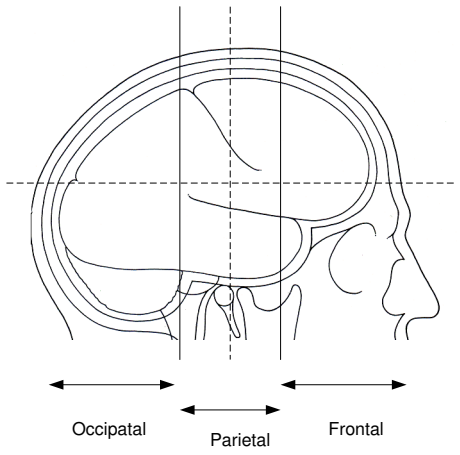

Figure 5.8: Semantic Regions in coronal view

Table 5.2: Semantic Regions in coronal view

Semantic Region	Start (mm)	End (mm)
Occipatal	first	-26
Parietal	-27	2
Frontal	3	last

hemisphere in saggital view and hence it is not possible to train the classifier across right and left hemisphere. Consequently, retrieval based on "Semantic Region" is not attempted in the saggital view. Nevertheless, the retrieval of 2D slices from saggital view can be accomplished using $PRES$ technique alone, presented in the last chapter.

Separate training process is adopted for each pair of modality. For example, for the training of modality pair PD-T1, the training images are taken from both the PD and T1 volumes. Duplicate training is avoided by using the same PD-T1 trained model for both PD-T1 and T1-PD classification. The training data is obtained by subsampling proximately 20% feature vectors from the whole feature set belonging to both modalities. Moreover, only volumes belonging to most noisy dataset (9% noise level) is used in the training set and the same model is used in the classification attempts for other noise-levels.

The LIBSVM library is used to train the SVM models using C-SVC scheme discussed earlier. Since the dimension of the feature set is large, linear kernel ($K(x_i, x_j) \equiv x_i^T x_j$) is found to be suitable, both in terms of accuracy and speed.

5.4.3 Classification in Semantic Regions

First, 2D wavelet transform magnitude is computed which is followed by the feature extraction. The classification is achieved according to the following equation,

$$sgn\left(\sum_{i=1}^{l} y_i K(x_i, x_j) + b\right) \tag{5.5}$$

which is the simplified version of equation 5.3 given the kernel ϕ is linear. The next section describes the full retrieval scheme.

5.5 Semantic Image Retrieval in 3D Volumes

This section provides the details of the overall retrieval process. Here, it is important to note that the semantic classification process, consisting of feature extraction, training and classification, utilizes wavelet transform magnitude only from the coarsest scale, i.e. $M_{2^4} f(x, y)$.

Initially, the wavelet transform magnitude of the incoming query image with 4 levels of decompositions is computed, which is followed by the feature extraction process. Next, the semantic classification is performed using a trained SVM model as shown in Figure 5.9. Finally, based on the result of the classification process, the retrieval using $PRES$ technique is conducted only in a specific semantic region. The $PRES$ technique is modified as shown in Figure 5.10, where the detection of two peaks based on equation 4.2 is avoided.

For the sake of clarity, feature extraction and semantic classification processes will jointly be termed as *Classification* and the PRES retrieval technique will be termed as

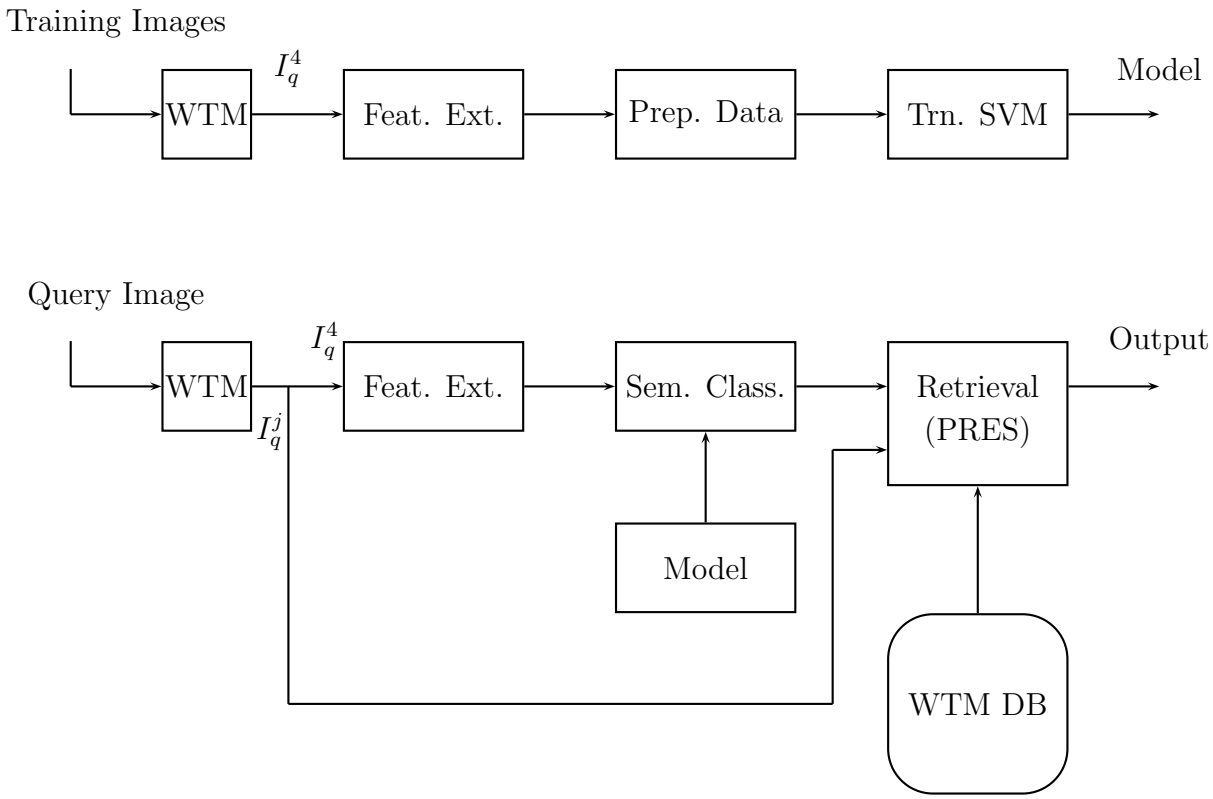

Figure 5.9: Description of SVM training (top) and classification with retrieval (bottom)

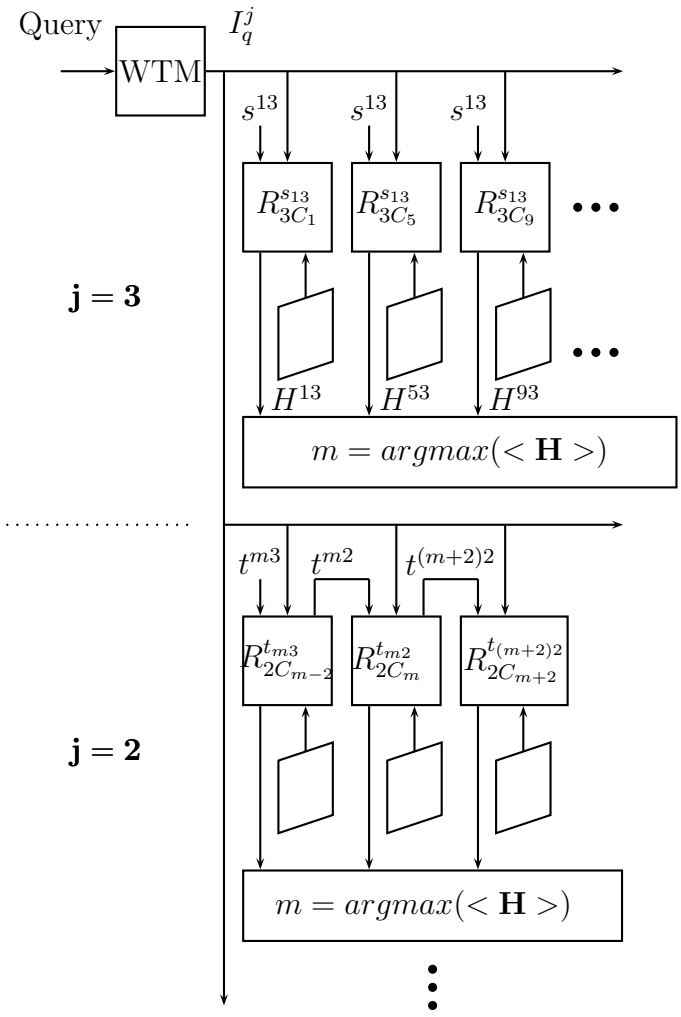

Figure 5.10: Description of modified $PRES$ retrieval scheme for $L = 3$ for $j = \{3, 2\}$

Retrieval. Also, the translation and rotation process in the retrieval process will collectively be termed as *Transformation*.

Though, the classification results are accurate, an error during the classification process may lead the retrieval process in the wrong semantic region leading to a large retrieval error. It should be noted here that an error in the classification near the transition between the classes may not necessarily result in retrieval error because PRES technique may find the correct slice while searching at finer scales of wavelet domain. However, there are chances of classification errors because the incoming query image is not pre-registered. In order to mitigate the effect of classification error, a novel two phase retrieval scheme is developed, which is shown in 5.11. In the first phase, the semantic classification is performed which is followed by the retrieval process proposed in the previous chapter. The retrieval result contains the best matching slice index along with the transformation parameter. In the second phase, the wavelet transform magnitude is transformed according to the transformation parameter and classification is attempted again. If the result of the second classification attempt is different than the first one, the retrieval processes is repeated. On the other hand, if the result of the second classification attempt is same as the first one then the result of the first retrieval process is accepted and the second retrieval attempt is avoided. Figure 5.11 shows the the flowchart of the proposed two phase retrieval scheme.

The performance of the proposed scheme is tested on AMD64 Dual Core CPU with 2.8GHz and 3GB of RAM using MATLAB (R2008b) environment. Since BrainWeb [220] images do not have much empty space, translations could not be employed, and consequently all the query images are randomly pre-rotated. The retrieval is tested under various noise conditions using the 3D brain volumes available at BrainWeb [220]. Here,

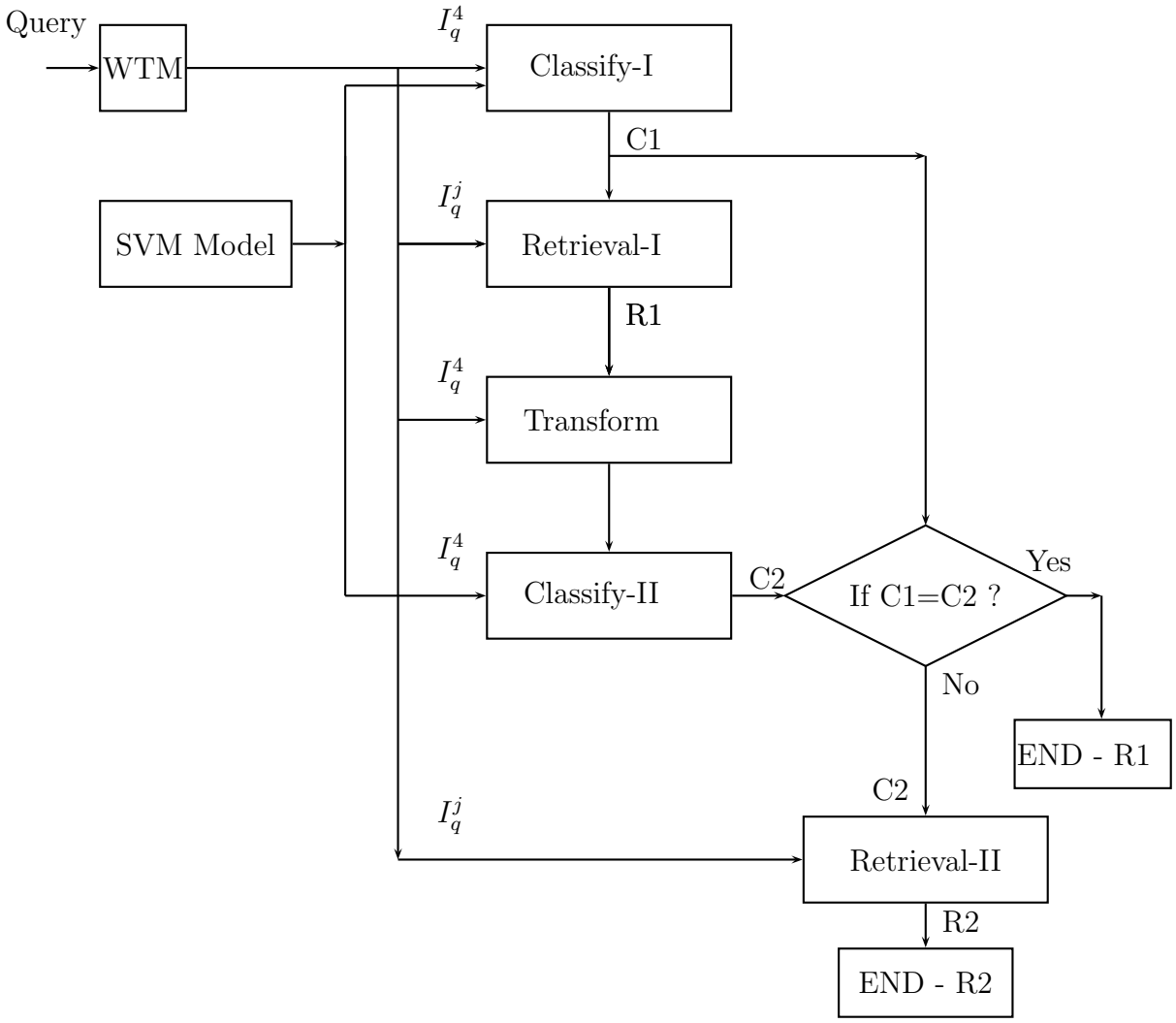

Figure 5.11: Two phase semantic assisted retrieval scheme

the noise is relative to the brightest tissue in the volume. This database provides one synthetic volume with several options to choose such as modalities, voxel sizes and noise levels etc. The size of the volume is $181 \times 217 \times 181$ with voxel size of $1mm \times 1mm \times 1mm$. In order to limit the number of simulations, both the query and the target 3D volumes belong to the same noise-level.

5.5.1 Retrieval among Healthy (Normal) Subjects

An extensive semantic retrieval simulation among healthy (normal) subjects is conducted. The tables 5.3, 5.4 and 5.5 show the classification errors, mean absolute errors and mean retrieval time in transverse view, respectively. Similarly, the tables 5.6, 5.7 and 5.8 in coronal view, respectively. The number of trails in the transverse and the coronal views are 180 and 217, respectively.

Table 5.3: Number of Classification Errors in Semantic Regions under noise (Transverse View)

	noise(3%)			noise(5%)			noise(7%)			noise(9%)		
Mode	T	O	F	T	O	F	T	O	F	T	O	F
PD-T1	0	2	2	0	2	2	0	2	2	0	2	3
T1-PD	0	4	0	0	4	0	0	4	0	0	4	0
T1-T2	0	2	3	0	2	4	0	2	4	0	2	3
T2-PD	0	3	2	0	4	3	0	3	3	0	5	0
T2-T1	0	5	0	0	5	0	0	6	0	0	6	0

5.5.2 Retrieval Between Normal and Abnormal(MS) Subjects

In this sub-section, retrieval simulations between normal and abnormal subjects with *Multiple Sclerosis* (MS) lesions, are presented. Table 5.9 shows the mean absolute errors and

Table 5.4: Mean Absolute Error in Semantic Regions under noise (Transverse View)

	noise(3%)			noise(5%)			noise(7%)			noise(9%)		
Mode	T	O	F	T	O	F	T	O	F	T	O	F
PD-T1	0.17	0.03	0.13	0.65	0.33	0.36	0.90	0.63	0.67	0.95	0.90	0.88
T1-PD	0.04	0.10	0.06	0.44	0.47	0.34	0.66	0.67	0.61	0.76	0.97	1.0
T1-T2	0.00	0.03	0.03	0.05	0.16	0.62	0.26	0.27	0.59	0.47	0.73	0.37
T2-PD	0.01	0.03	0.01	0.12	0.9	0.03	0.22	0.27	0.10	0.35	0.57	0.10
T2-T1	0.38	0.03	0.04	0.09	0.50	0.22	0.37	0.90	0.37	0.94	1.5	0.57

Table 5.5: Mean Retrieval Time (seconds) in Semantic Regions under noise (Transverse View)

	noise(3%)			noise(5%)			noise(7%)			noise(9%)		
Mode	T	O	F	T	O	F	T	O	F	T	O	F
PD-T1	12.3	9.1	12.1	12.2	8.8	11.2	12.8	9.9	10.7	12.9	9.7	10.9
T1-PD	13.4	11.4	12.3	13.9	11.4	11.3	12.9	11.0	11.6	13.2	11.2	11.3
T1-T2	13.1	10.7	12.5	12.7	10.4	12.4	13.6	11.3	13.1	12.9	11.3	12.0
T2-PD	12.1	10.5	12.3	12.7	10.1	11.9	12.2	10.9	11.8	11.9	14.4	12.5
T2-T1	12.1	10.4	10.9	12.2	11.5	10.4	12.4	12.2	9.9	12.5	11.4	10.3

Table 5.6: Classification Errors in Semantic Regions under noise (Coronal View)

	noise(3%)			noise(5%)			noise(7%)			noise(9%)		
Mode	T	O	F	T	O	F	T	O	F	T	O	F
PD-T1	1	1	0	1	1	0	1	1	0	1	0	1
T1-PD	1	0	1	1	0	2	1	0	1	1	0	1
T1-T2	0	1	1	0	1	1	0	1	1	0	1	1
T2-PD	1	2	0	1	2	0	1	2	0	1	2	0
T2-T1	1	2	0	1	2	0	1	2	0	0	1	1

Table 5.7: Mean Absolute Error in Semantic Regions under noise (Coronal View)

	noise(3%)			noise(5%)			noise(7%)			noise(9%)		
Mode	T	O	F	T	O	F	T	O	F	T	O	F
PD-T1	0.03	0.02	0.05	0.52	0.25	0.60	0.64	0.69	0.98	0.78	0.75	1.2
T1-PD	0.0	0.03	1.0	0.22	0.28	0.86	0.47	0.59	0.97	0.64	0.79	1.0
T1-T2	0.0	0.0	2.4	0.02	0.0	3.1	0.07	0.14	2.9	0.25	0.51	1.7
T2-PD	0.0	0.0	0.0	0.02	0.03	0.1	0.05	0.14	0.19	0.19	0.45	0.33
T2-T1	0.07	0.0	2.5	0.15	0.10	2.8	0.37	0.28	3.5	0.25	0.51	1.7

Table 5.8: Mean Retrieval Time (seconds) in Semantic Regions under noise (Coronal View)

	noise(3%)			noise(5%)			noise(7%)			noise(9%)		
Mode	T	O	F	T	O	F	T	O	F	T	O	F
PD-T1	13.9	9.8	12.9	13.2	9.5	12.8	13.6	10.2	12.8	12.5	9.8	12.2
T1-PD	9.3	9.1	9.4	9.1	8.2	9.3	9.1	7.5	8.5	9.2	8.5	9.0
T1-T2	9.5	9.5	6.3	9.5	9.5	5.5	9.5	9.5	3.9	12.0	9.4	19.8
T2-PD	9.3	8.5	8.5	9.33	8.5	8.2	8.9	8.5	7.2	8.5	7.5	6.7
T2-T1	8.1	8.5	7.3	7.7	8.2	7.1	7.2	7.5	6.4	11.9	9.3	19.7

mean retrieval time in transverse view with five different modality pairs. Similarly, Table 5.10 provides simulation results in coronal view. It is important to mention that the trained SVM models of healthy subject are used and the new set of training with MS cases is avoided, which lead to the performance degradation. The number of trails in the transverse and the coronal views are 180 and 217, respectively.

Table 5.9: Mean Absolute Error in Semantic Regions with MS Lesions (Transverse View)

	Mean Absolute Error			Mean Retrieval Time (seconds)		
Mode	T	O	F	T	O	F
PD(MS)-T1	0.8140	0.7000	0.7910	14.3683	9.6563	12.5198
PD(MS)-T2	2.0000	0.4333	0.2687	12.8349	12.9553	11.9003
T2(MS)-T1	1.1395	2.1667	0.6119	13.0562	14.3154	11.6275
T2-PD(MS)	0.2093	1.4000	0.1791	12.0148	11.8645	12.4509
T2(MS)-PD	0.1512	1.7333	0.1194	12.5790	14.9376	13.1212

Table 5.10: Mean Absolute Error in Semantic Regions with MS Lesions (Coronal View)

	Mean Absolute Error			Mean Retrieval Time (seconds)		
Mode	T	O	F	T	O	F
PD(MS)-T1	0.7966	0.9592	1.2333	12.7128	10.1579	13.3967
PD-T1(MS)	0.1695	1.5510	0.6333	14.9693	11.0515	13.7697
T1(MS)-PD	0.2034	1.7755	0.5833	15.4213	10.8440	15.4994
T1-PD(MS)	0.5763	3.3469	1.1833	12.9809	11.1807	15.4256
T2(MS)-PD	0.0508	0.1429	0.0833	15.6196	9.3268	15.0574

5.6 Conclusions

In this chapter, a novel, SVM based, semantic classification scheme is presented for classifying the incoming 2D query image into one of the semantic regions. This approach not only retrieves 2D slice in 3D volumes but also classifies them to semantic regions as well, which is a major contribution of this research work. The proposed semantic classification scheme can be extremely useful for semantic based categorization, clustering and annotation of images in databases. Here, the semantic regions are inspired by the lobes of the human brain. The classification is done by a multiclass SVM and the errors are found to be small. Another contribution is the development of a fast feature extraction scheme using polar, non-uniform, sampling grid, which is designed for extracting features in wavelet domain for the purpose of semantic classification. The feature set is compared against a well known feature extraction technique, called *SIFT*, and is found more effective. The proposed retrieval algorithms are tested under noisy conditions among the volumes of healthy subjects. Finally, retrieval results are provided between the volumes of healthy and abnormal scans of the same subject. Simulations show promising results with respect to multi-modality, accuracy, speed and robustness.

Chapter 6

Semantic Assisted Multisubject Image Retrieval in 3D Brain Volumes

6.1 Introduction

In the previous chapter, a semantic assisted image retrieval in 3D MR volumes was presented for intersubject scenario. In this chapter, a novel semantic assisted scheme is presented for multisubject retrieval. Here, the term "multisubject" means that the retrieval is done in a database of 3D volumes belonging to many subjects. Thus, the technique proposed in this chapter is more general. The overall retrieval technique essentially works in three stages by first associating the in coming slice to a few 3D brain volumes among many subjects; then associating the incoming slice to a specific semantic region; and finally executing the PRES based retrieval scheme within the identified volume(s) and in a specific semantic region. In this chapter, an SVM based scheme is presented for identifying the 3D volume of the subject using incoming 2D query image, which is the major contribution.

Here, a subset of the MIDAS dataset [235], consisting of approximately 100 MRI datasets of normal subjects, is used. The selected dataset comprises of magnetic resonance angiography (MRA), T1-Flash, T2 and T1-MPRage sequences.

Rest of the chapter is organized as follow. The second section describes the identification of 3D volumes from 2D slice images using SVM; the third section presents the multisubject semantic classification of lobes, which is the extension of the previous chapter; the fourth section describes the full retrieval scheme; the fifth section presents the experimental results and performance evaluation. Finally, the sixth section concludes this chapter.

6.2 3D Volume Identification Using SVM Probabilistic Outputs

This section discusses a novel approach for identifying a 3D volume given a query slice image. The feature extraction scheme in the wavelet domain is described in the previous chapter. Here, the multicalss SVM based machine learning technique is used for identifying similar 3D volumes from a set of volumes. Normally, a multiclass SVM only predict class outputs, however in the process of identification some measure of similarity is needed to rank the results. In the present work, a probability estimation approach, discussed in [236], is used for multi-class SVM classification.

6.2.1 SVM with Probabilistic Outputs

For the purpose of training, the input to the retrieval system is a set of feature vectors from the training images. Here, each image is manually annotated with a single semantic

label selected out of M labels or categories. In the present application, the class labels are identities of the subjects. A set of M labels are defined as $C_1, C_2, ..., C_M$, where each $C_i, i \in 1, ..., M$ characterizes the representative semantics of an individual volume. In the testing stage of the probabilistic classifier, each non-annotated or unknown image feature vector is classified against M identities. The output of the classification process produces a ranking on the M identities. Each identity would assign a probability (confidence) score to the image. Consequently, the confidence score represents the weight of a category label in the overall description of the incoming image. The probabilities or confidence scores of the identities form an M-dimensional vector for a feature x_m of image i as follows:

$$p_i(x_m) = [p_{i1}(x_m)...p_{ik}(x_m)...p_{iM}(x_m)]^T \tag{6.1}$$

Here, $p_{ik}(x_m), 1 \leq k \leq M$, denotes the posterior probability that an image i belongs to identity C_k in terms of input feature vector x_m. Finally, an image i belongs to an identity $C_l, l \in 1, ..., M$ using feature vector x_m where the identity label is determined by

$$l = argmax_k \left[p_{ik}(x_m)\right] \tag{6.2}$$

which is the label of the identity with the maximum probability score. In this context, given the feature vector x_m, the goal is to estimate

$$p_k = P(y = k|x_m), \qquad k = 1, ...M \tag{6.3}$$

the pairwise class probabilities r_{kj} are estimated as an approximation (as in [236]),

$$r_{kj} \approx P(y = k | y = k \text{ or } j, x_m) \approx \frac{1}{1 + e^{Af+B}} \tag{6.4}$$

where A and B are the parameters estimated by minimizing the negative log-likelihood function, and f are the decision values of the training data (fivefold cross-validation (CV) is used to form an unbiased training set). Finally, p_k is obtained from all these r_{kj}'s by solving the following optimization problem:

$$\min_{p} \frac{1}{2} \sum_{k=1}^{M} \sum_{j:j \neq k} \left(r_{j_k} p_k - r_{j_k} p_j \right)^2 \tag{6.5}$$

$$subject\ to \quad \sum_{k=1}^{M} p_k = 1, \quad p_k \geq 0 \ \forall k$$

where p(e.g., $p_i(x_m)$) is a M-dimensional vector of multi-class probability estimates. Similarly, the label vector of a query image q can be found online by applying its feature descriptors as inputs at different levels to the SVM classifiers as

$$p_q(x_m) = \left[p_{q_1}(x_m) ... p_{q_k}(x_m) ... p_{q_M}(x_m) \right]^T \tag{6.6}$$

6.2.2 3D Volume Identification Scheme

The length of the feature set is large, therefore, a linear kernel with $K(x_i, x_j) \equiv x_i^T x_j$ is found to be most useful for SVM training, both in terms of accuracy and speed. The first stage involves computing 2D wavelet transform magnitude, which is followed by the feature extraction. Given feature vector of the 2D image slice, SVM classification produces probabilities $p_q(x_m)$ for each class label. The process of data collection for training and

testing is depicted in Figure 6.1. The top ranked volumes with normalized probabilities $p_q(x_m)$ greater than 60% of the maximum is selected for further image retrieval process.

Figure 6.2 shows the SVM training and identification process. During the identification process, the SVM based classifier outputs a probability vector of size M, where M is the number of identities.

The accuracy of the SVM based identification process can be seen in Figures 6.3, 6.4, 6.5, 6.6, 6.7, 6.8 and 6.9 for different combinations of modalities between MRA-T1F, T1F-MRA, MRA-T1R, T1R-MRA, T1F-T1R, T1R-T1F and T2-T1R, respectively.

It can be observed from these figures that the SVM based identification scheme improves the performance significantly as compared to the Euclidean distance based ranking. It is worth mentioning that the SVM based identification works extremely fast; normally it takes only 0.2 seconds to do identification in $\approx$ 100 volumes. The LIBSVM library is employed to train the SVM models using C-SVC scheme discussed in the previous chapter.

6.3 Multisubject Semantic Classification of Lobes

A scheme for the semantic classification was proposed in the previous chapter to classify the incoming 2D query image into one of the semantic region of the brain. In multisubject scenario, this problem can be handled in two ways; (i) train SVM classifiers for every subject or (ii) train single SVM classifier with training data from all the subjects. Here, second approach is used due to the lower complexity, which is shown in Figure 6.10. Hence one SVM classifier is trained for each pair of modalities and for each view.

For transverse view, the three semantic regions are defined as shown in Table 6.1. In

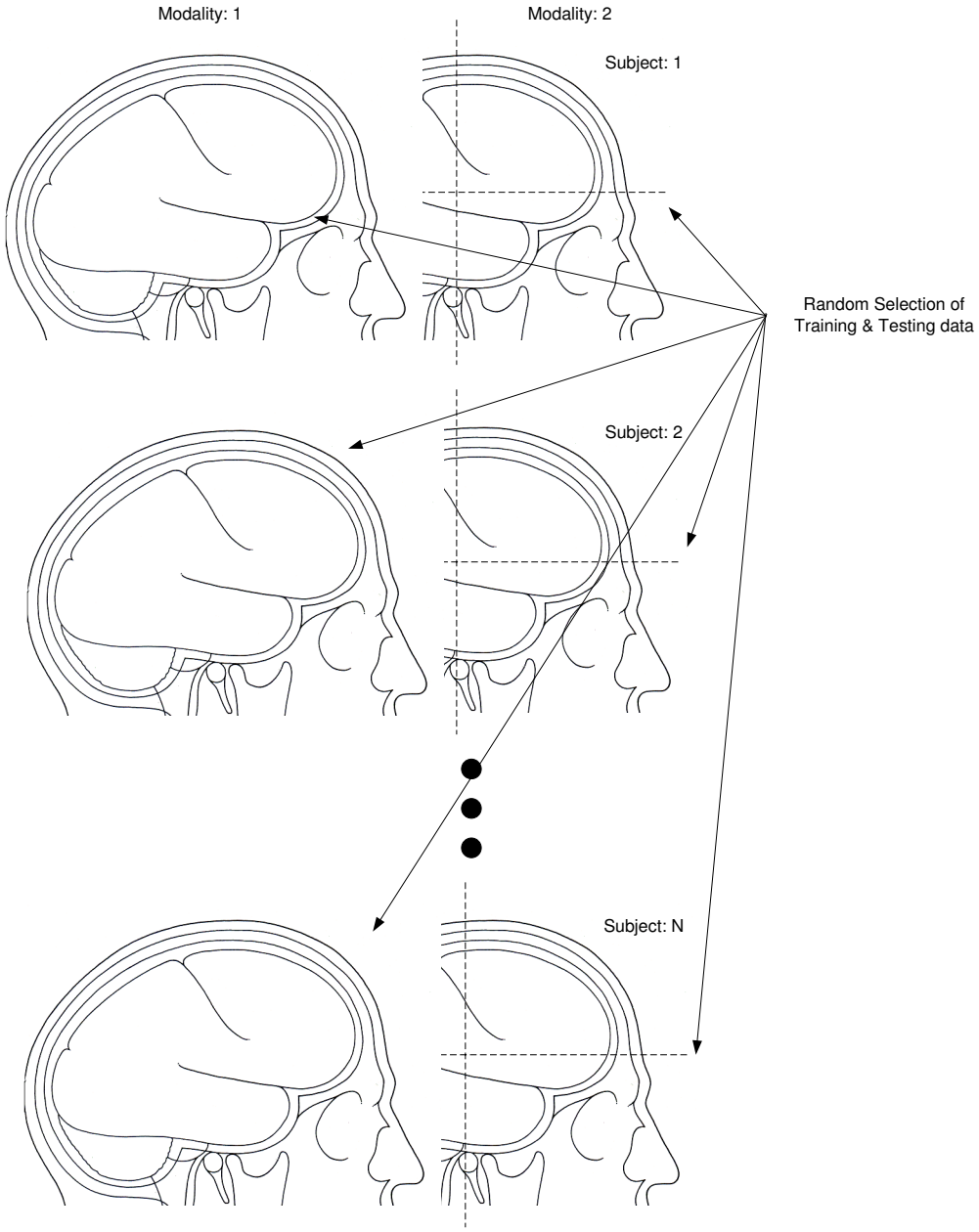

Figure 6.1: Training and Testing for Identification

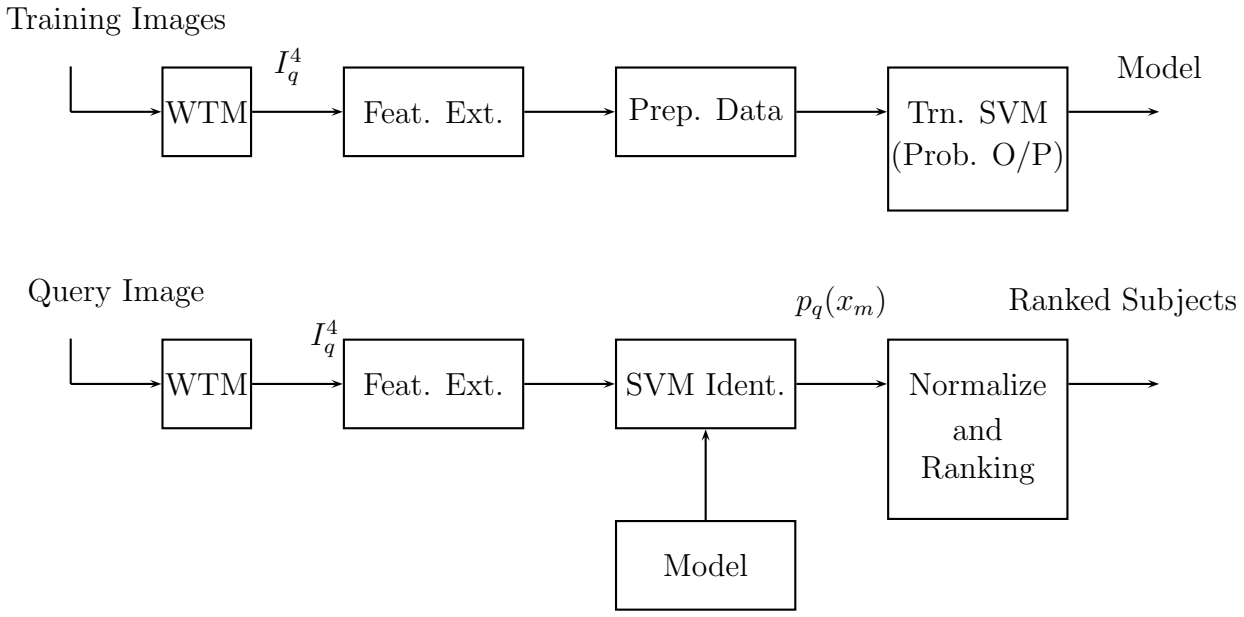

Figure 6.2: Description of SVM training (top) and identification (bottom)

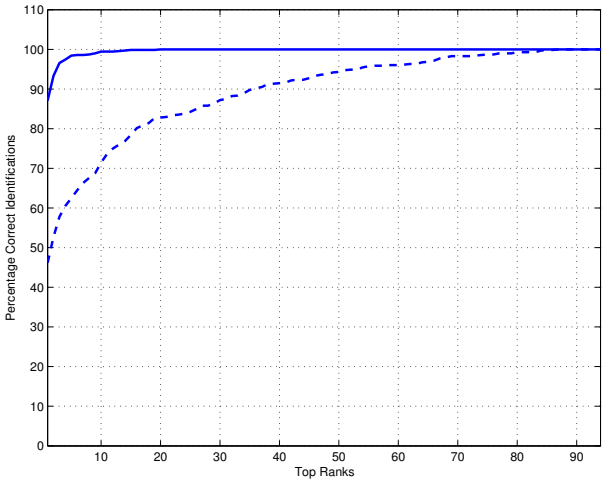

Figure 6.3: MRA-T1F: SVM based identification (solid) versus Euclidean distance (dotted)

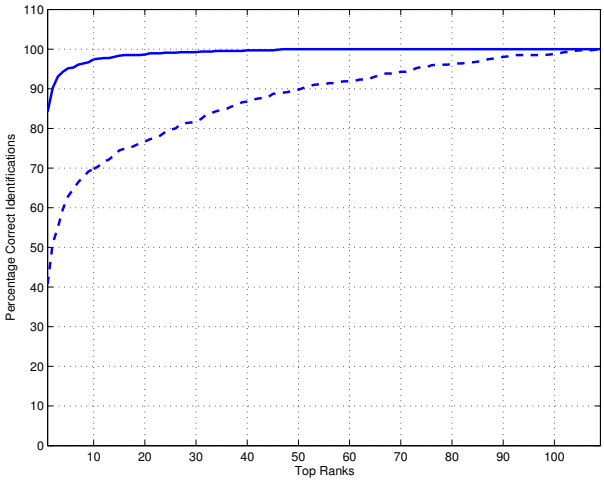

Figure 6.4: T1F-MRA: SVM based identification (solid) versus Euclidean distance (dotted)

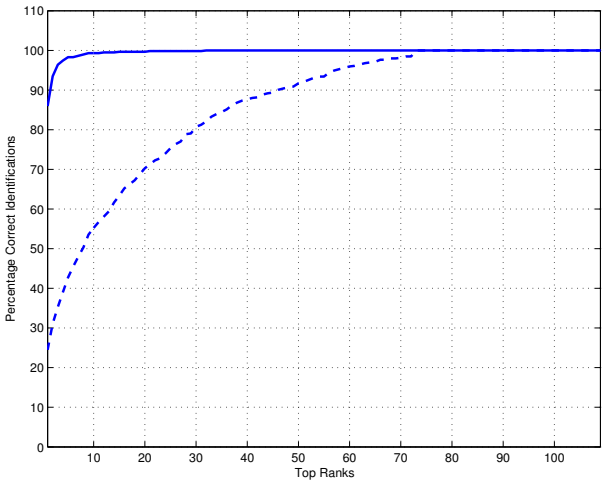

Figure 6.5: MRA-T1R: SVM based identification (solid) versus Euclidean distance (dotted)

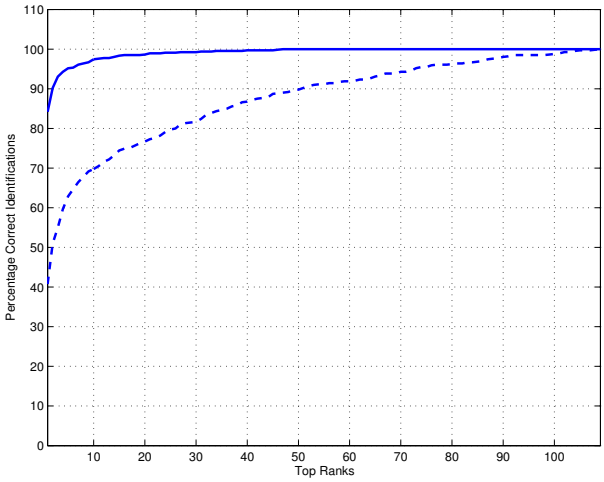

Figure 6.6: T1R-MRA: SVM based identification (solid) versus Euclidean distance (dotted)

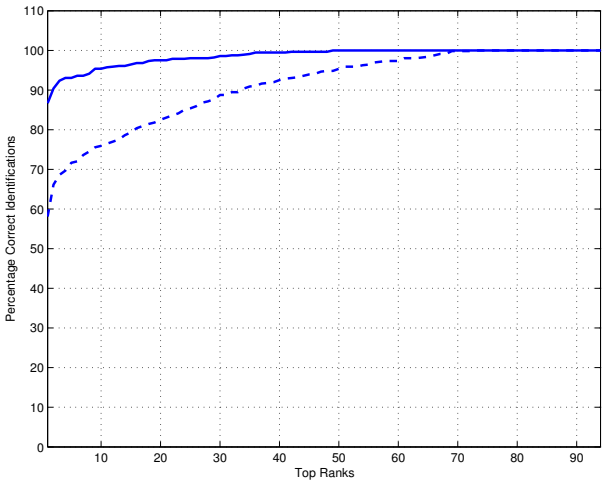

Figure 6.7: T1F-T1R: SVM based identification (solid) versus Euclidean distance (dotted)

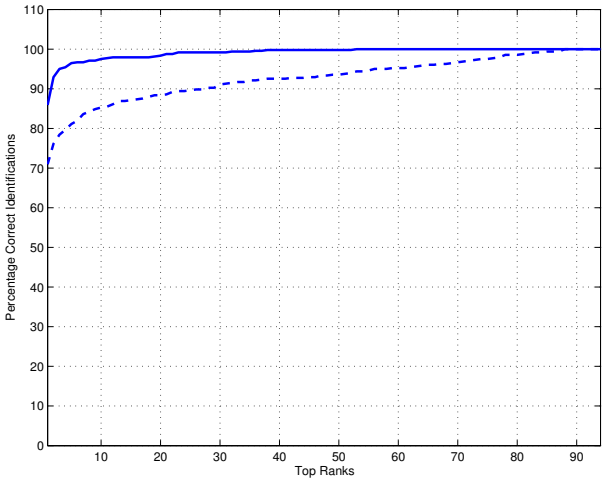

Figure 6.8: T1R-T1F: SVM based identification (solid) versus Euclidean distance (dotted)

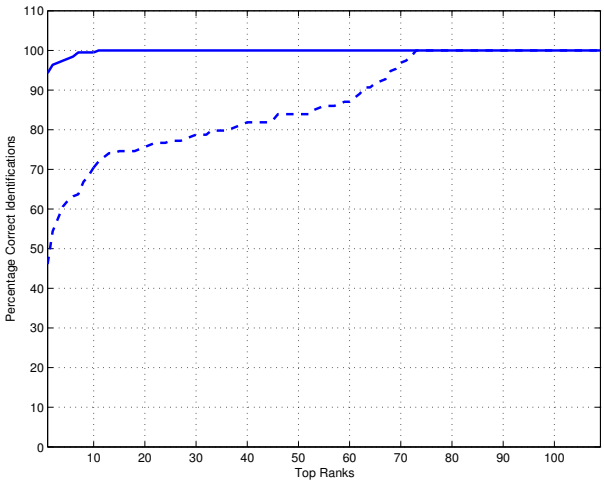

Figure 6.9: T2-T1R: SVM based identification (solid) versus Euclidean distance (dotted)

this table, the reference (origin) is a the center of 3D brain volume, which is found by averaging the 3D coordinates of voxels belonging to the non-zero grey-levels.

Table 6.1: Semantic Regions in transverse view

Semantic Region	Start (mm)	End (mm)
Temporal	first	-5
Occipatal	-6	24
Frontal	25	last

The training data is extracted by subsampling approximately 45% feature vectors from the whole feature set belonging to all the subjects and specific pair of modalities (such as MRA-T1F). Duplicate training is avoided by using same model for the pair of modalities of query and database images, regardless of which of them is used for query and retrieval. For example, one SVM classifier is trained for both MRA-T1F and T1F-MRA modality pairs.

The LIBSVM library is used to train the SVM models using C-SVC scheme discussed earlier. Since the dimension of the feature set quite large (864) for each 2D image slice, linear kernel, with $K(x_i, x_j) \equiv x_i^T x_j$, is found to be most useful, both in terms of accuracy and speed.

6.4 Multisubject Image Retrieval Scheme

In this section, full retrieval scheme is described to retrieve 2D image slice in multiple 3D volumes. It consists of three stages, viz. volume identification, semantic region classification and registration-based retrieval. Figure 6.11 shows the schematic diagram of the full retrieval scheme. In the first stage, given the query slice image, the identification of volume(s) is done using SVM based technique. The identification process identifies a subset of

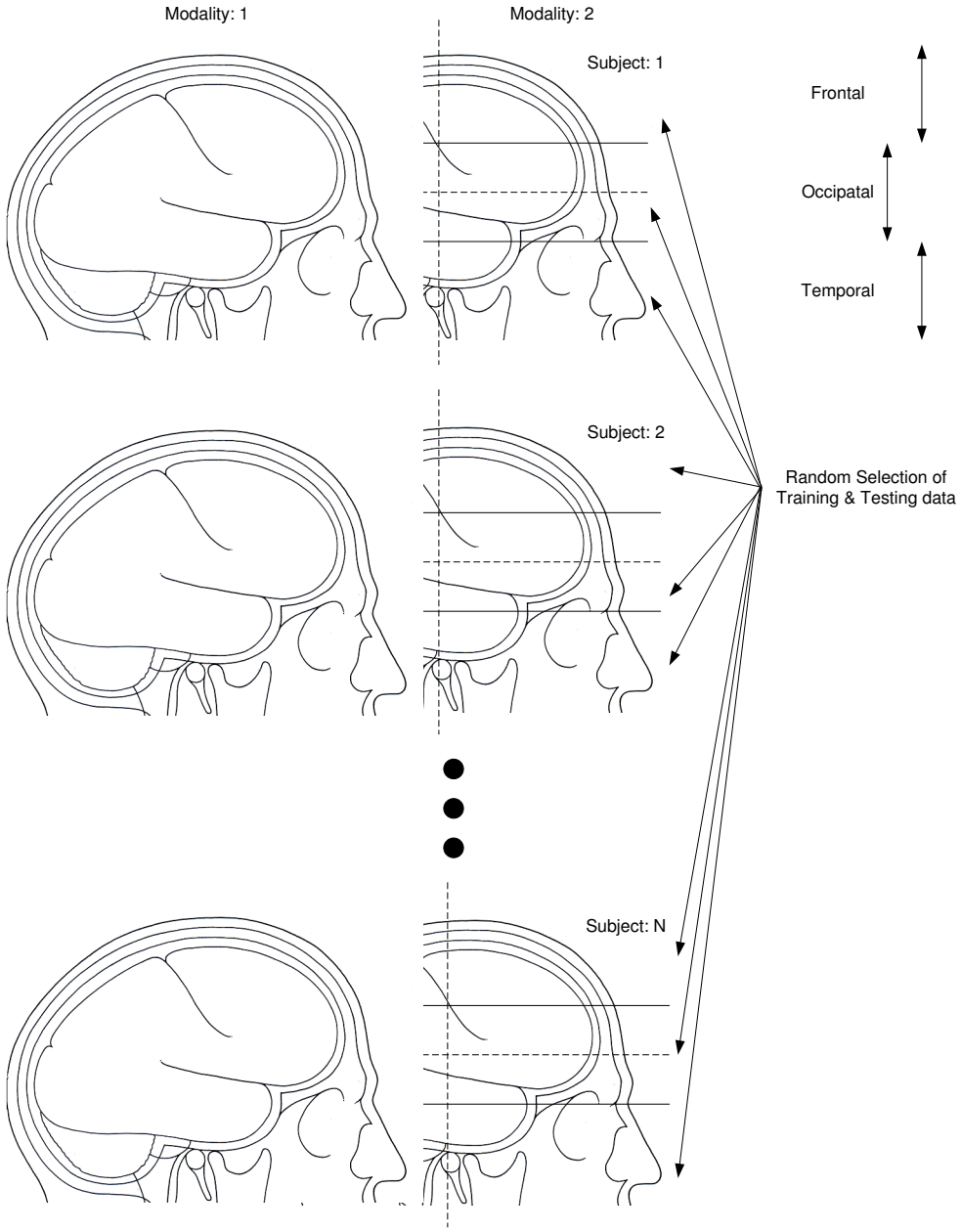

Figure 6.10: Training and Testing for multi subject classification

highly ranked subjects given a 2D query image. In the second stage, SVM based classifier associates an incoming 2D query image to a specific semantic region. Finally, based on the results of the first and second stage, registration based retrieval (PRES) is executed within the volumes of the identified subjects and in that specific semantic region.

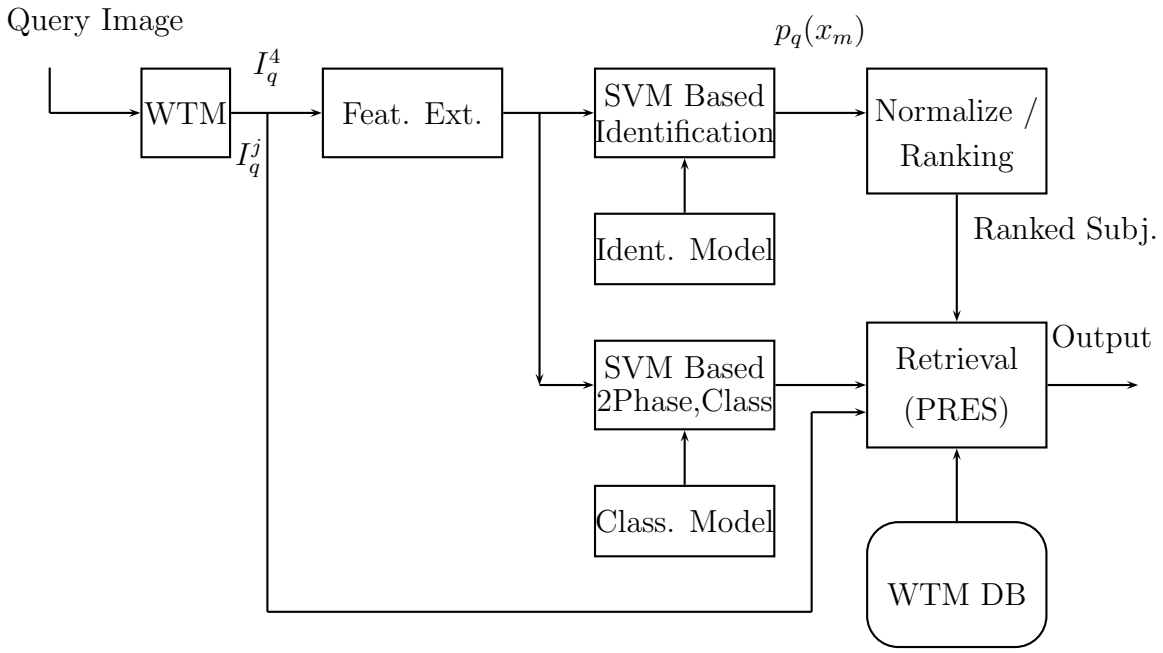

Figure 6.11: Description of full retrieval scheme

6.5 Experiments

In order to test the algorithms proposed in this chapter, MIDAS dataset [235] consisting of approximately hundred MRI datasets of normal subjects is used. The dataset comprises of magnetic resonance angiography (MRA), T1-Flash, T2 and T1-MPRage sequences. The dataset is quite realistic in the sense that it is captured on various MR scanners and at

different times. The details of the dataset is presented in Table 6.2, which shows that the MRA and T2 volumes are anisotropic while the volumes of T1-Flash and T1-MPRage images are isotropic. Consequently, MRA and T2 image volumes are converted to isotropic voxels using MRICRO tool, which leads to different image sizes as shown in Table 6.3.

Table 6.2: MIDAs Dataset - Original

Modality	Voxel size(mm)	image dimensions
MRA	$0.51 \times 0.51 \times 0.8$	$448 \times 448 \times 128$
T1-Flash	$1.0 \times 1.0 \times 1.0$	$176 \times 256 \times 176$
T1-MPRage	$1.0 \times 1.0 \times 1.0$	$208 \times 256 \times 128$
T2	$0.5 \times 0.5 \times 1.0$	$392 \times 512 \times 160$

Table 6.3: MIDAs Dataset - Modified

Modality	Voxel size(mm)	image dimensions
MRA	$1.0 \times 1.0 \times 1.0$	$228 \times 228 \times 101$
T1-Flash	$1.0 \times 1.0 \times 1.0$	$176 \times 256 \times 176$
T1-MPRage	$1.0 \times 1.0 \times 1.0$	$208 \times 256 \times 128$
T2	$1.0 \times 1.0 \times 1.0$	$196 \times 256 \times 160$

In order to properly register and display the images, all the 2D slice images are resized to larger image sizes by padding with dark noisy background. For example, the transverse (or axial) images are padded to form images of 256×256 pixels for all the modalities. Moreover, the MIDAS dataset is not a "closed set" among various modalities. Here, "closed set" means that there are a few individuals (subjects) who are present in one modality while absent in the other modality. Therefore a "closed set" is extracted and utilized in this research.

6.5.1 Training the SVM

As shown in Figure 6.11, an SVM based machine learning scheme is employed for both "identification" and "classification" stages. In order to train an SVM model for identification process, the training set is created by taking features from every 4th slice from all 3D volumes belonging to the pair of modalities. The remaining slices are used for testing purpose. In order to train the SVM for classification process, every 3rd slices from all 3D volumes belonging to the pair of modalities are used as training set. There are three semantic classes as discussed in the previous chapter.

6.5.2 Performance Evaluation

Here, *False Match Rate* (FMR) is used as a performance evaluation metric for the proposed retrieval system. This metric is extensively used in the evaluation of biometric systems. In the current retrieval framework, *FMR* can be defined as, *the probability that the system fails to find a correct match given the query slice in the database.* Basically, it measures the percent of valid queries which are incorrectly matched and can be computed as,

$$FMR = \frac{number\ of\ incorrect\ matches}{total\ number\ of\ attempts} \times 100 \tag{6.7}$$

This metric is justified here because a 2D query image matches correctly with only one image slice in the identified 3D volume. Traditionally, image retrieval systems use *Precision* and *Recall* as performance metrics which are adapted from the information retrieval framework. The reason for using these performance measures is the premise that a query image retrieves a set of images. In the simulations, approximately 500 random retrieval

trails are conducted for each modality pairs and the accuracy and mean retrieval time are presented in Table 6.4.

Table 6.4: Performance Evaluation - Transverse

Modality pairs	FMR (%)	Mean Retrieval time (seconds)
MRA:T1-Flash	8.4	18.70
MRA:T1-MPRage	9.8	17.65
T1-Flash:T1-MPRage	8.5	18.69
T2:MRA	9.5	20.65
T2:T1-MPRage	4.3	16.99

It is important to note that the identification stage does not increase the retrieval time significantly because most of the time it identifies the correct volume in the first rank. The errors are identified manually by visual inspections of all the retrieval results. Simulations of the algorithm is performed on AMD64 Dual Core CPU with 2.8GHz and 3GB of RAM using MATLAB (R2008b) environment. Figures 6.12 and 6.13 show the results with noisy MRA query images in T1-Flash database. Figure 6.14 shows the retrieval result in the presence of RF field inhomogeneity.

It is worth mentioning that the proposed image retrieval technique in coronal and saggital views is not attempted as images in these views are rectangular with high aspect ratio (as shown in Table 6.3), hence the circular feature map could not be applied for extracting image features consistently. Moreover, in these views, the center of the object is not closer to the center of the image.

6.6 Conclusions

In this chapter, a novel scheme for retrieving 2D image slices in multiple 3D brain volumes is presented. Here, an SVM based identification scheme for identifying the 3D volume of the

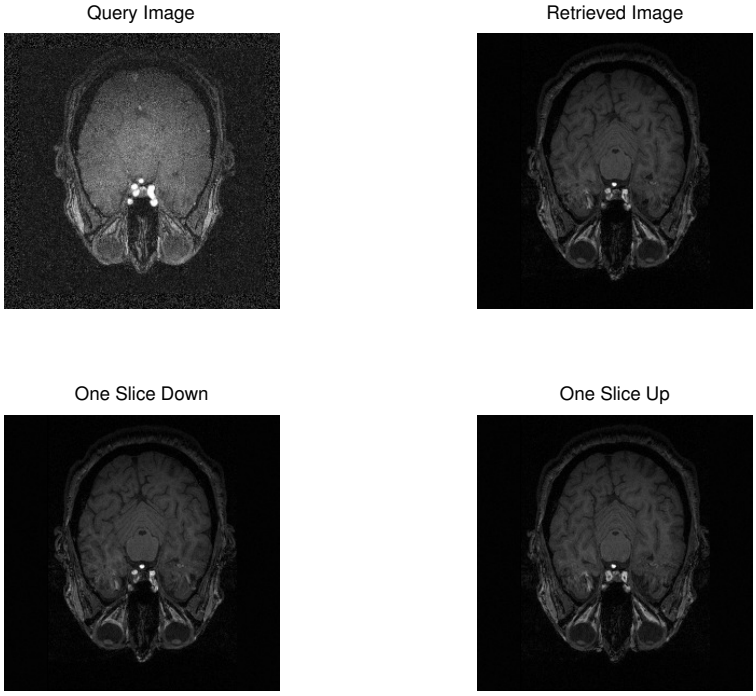

Figure 6.12: Retrieval Result (MRA vs T1-Flash)

subject using an incoming 2D query image is proposed, which is the major contribution of this chapter. In the next step, a semantic classification scheme for classifying the incoming 2D query image into one of the semantic regions is employed. This classification technique is similar to the one presented in the previous chapter. This is followed by the regular registration-based retrieval, as presented in the fourth chapter. The identification errors are found to be small resulting in a fast algorithm because in more than 95% of cases it finds the correct volume in the top most rank. Simulations show promising results with respect to multi-modality, accuracy, speed and robustness.

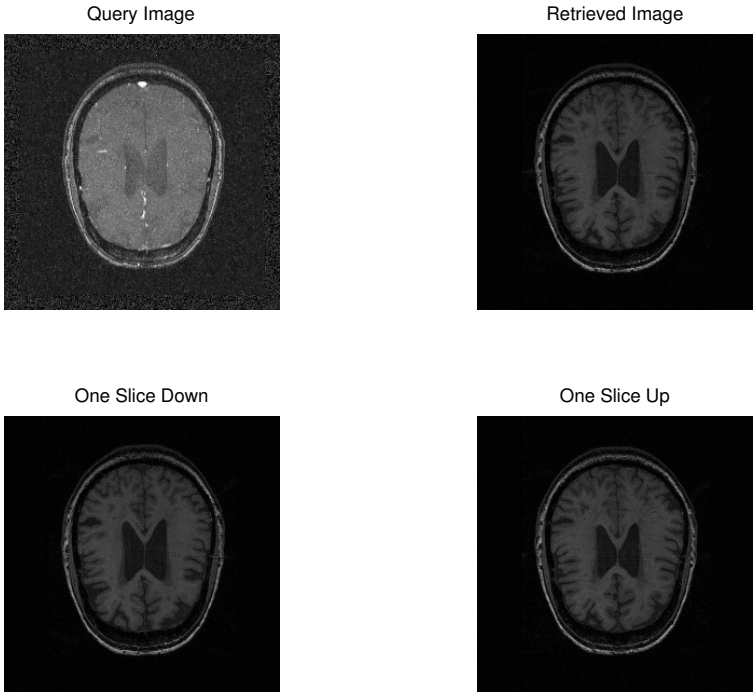

Figure 6.13: Retrieval Result (MRA vs T1-Flash)

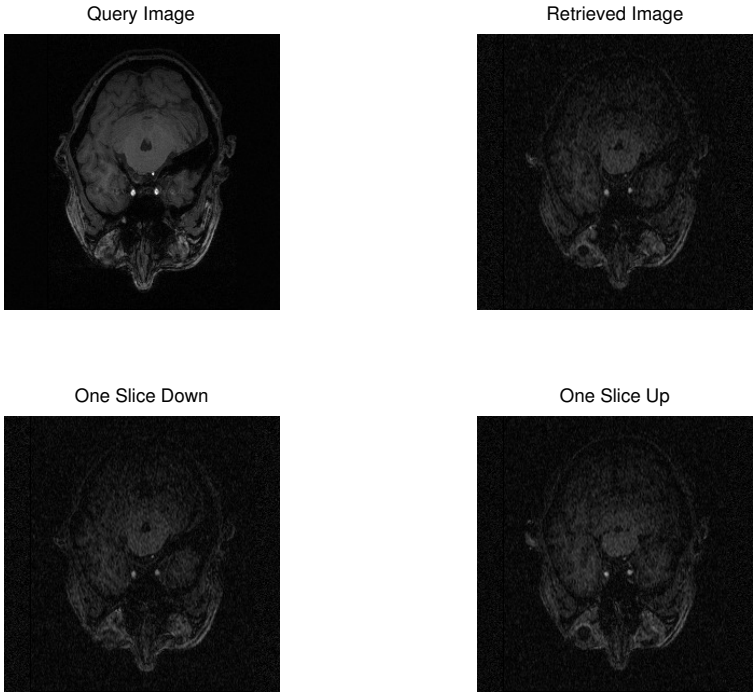

Figure 6.14: Retrieval Result (T1-Flash vs T1-MPRage)

Chapter 7

Conclusions and Future Work

This chapter provides some general comments and remarks on the insights that have been gained in this research in addition to the specific conclusions that are presented in the previous chapters. Furthermore, some promising future research directions are discussed.

The main objective of this thesis is to address the challenging problem of developing useful image retrieval techniques for medical diagnostics. Specifically, this dissertation proposed a system for semantic, robust and multimodal retrieval of magnetic resonance images in 3D MR brain volumes.

The work is presented in a bottom-up approach, where the core of the proposed retrieval system is based on the multiscale 2D rigid registration, as presented in chapter three. The proposed retrieval technique provides a precise algorithm for image retrieval within the 3D volume of a given subject, as discussed in detail in chapter four. The speed of the proposed retrieval technique has further improved with the introduction of a novel semantic assisted classification scheme, as discussed in chapter five. This classification technique uses a an efficient feature set in wavelet domain, which in turn is used for the training of Support

Vector Machines (SVM). Retrieval results are also provided between the volumes of healthy and abnormal scans of the same subject. A novel multimodal retrieval scheme for searching a 2D image slice in multiple 3D volumes is proposed, which is discussed in chapter six. Here, the rest of the retrieval stages uses a semantic classification and registration based framework.

7.1 Contributions

The highlights and findings of this research can be briefed as follows:

1. A precise, multimodal and fast retrieval 2D slice retrieval technique is proposed for the diagnostic applications in MR brain volumes.

2. The multiscale registration based framework works well for the retrieval task, using mutual information (MI) as a matching criterion. The proposed approach is robust against noise and RF inhomogeneity.

3. The SVM-based classification technique is found to be effective for learning semantic regions in the human brain based on various lobes. Even in the case of volumes of multiple subjects the classification accuracy is in the range of 95%.

4. The SVM-based identification technique is also found to be effective for identifying the subject given a 2D slice. The identification accuracy is in the range of 98%.

5. The overall accuracy in the multisubject retrieval technique is found to be greater than 90%. In intersubject (within the same subject) retrieval scenario, the *Mean Absolute Error* is less than 2, which indicates that on the average the error is within one slice.

7.2 Limitations

This dissertation achieves its objectives satisfactorily and is expected to contribute significantly to the research community. However, there are three limitations of the presented work:

1. Since proposed retrieval inherently does 2D image matching, a large rotation around the axis in the plane of the view would result in retrieval errors. To mitigate this problem, 3D volumes of both the query and the target should be crudely registered. However, this problem has not been encountered during this research which may indicate that the proposed retrieval approach is quite robust against typical misalignments.

2. Retrieval in saggital view is not possible with semantic regions because it is impossible to parcellate the brain lobes in this view. This is due to the symmetry of lobes between right and left brain hemispheres, therefore, *PRES* scheme should be used in this case.

3. Retrieval in coronal view could not be attempted with the multisubject scheme because the images in this view are rectangular with a high aspect ratio. Additionally, the circular feature map is not suitable for computing features consistently across various modalities. This problem is further aggravated due to the fact that the center of the object is far from the center of the image in coronal and saggital views. This can be handled by either embedding the image into a larger image or oversampling (reslicing) the volumes. However, these ideas were not explored in this work.

7.3 Future Directions

The proposed research work can be extended in many ways, as discussed below.

1. As already mentioned, the retrieval in saggital view is not possible with semantic regions due to the symmetry of lobes between the right and the left hemispheres. Therefore, some different kind of semantics may be explored.

2. The proposed semantic classification scheme is crude in the sense that it partitions the human brain into regions with 2D planes, which is quite different from the actual shape of the lobe. Moreover, these regions are of fixed sizes for all the scans (subjects). Therefore, a more accurate parcellation scheme may prove helpful.

3. Both the semantic "classification" and "identification" utilize the wavelet decomposition at the coarsest scale. Therefore, a multiscale extension of the classification and the identification scheme provide promising research direction. However, such multiscale extensions tend to increase the complexity of the system.

4. The SVM-based identification approach is found to be highly accurate with approximately 100 volumes available in the MIDAS database. However, there is a need for more thorough analysis with more subjects.

5. For interactive retrieval a faster registration scheme is needed. This can be achieved through various implementation strategies.

6. Integrating the proposed approach into an existing PACS system is highly desirable to asses its usefulness in a clinical setting. This may prove an interesting and a valuable future research direction.

7.4 Final Remarks

In this dissertation, a new research paradigm for image retrieval systems for diagnostic and decision support applications is proposed and implemented. More specifically, a system for semantic retrieval of MR images in 3D brain volumes is presented. This is an important application because it is expected to help experts where the human eye may fail. Previously proposed systems use imprecise segmentation and feature extraction techniques, which are not suitable for precise matchings required for the retrieval in such applications. Here, an efficient multiscale representation is used, which is robust against noise and MR inhomogeneity. In order to achieve high accuracy in the presence of misalignments, an image registration based framework is developed. To speed-up the retrieval system, a fast discrete wavelet based feature space is used. Further improvements in the retrieval speed is achieved by using the semantic classification of the human brain into various semantic regions. This classification is done by using an SVM based approach. A novel and fast 3D volume identification system is proposed for identifying a 3D volume given a 2D image slice. Here, the SVM output probabilities are used for ranking and identification of 3D volumes. The proposed retrieval systems are tested not only for noise conditions but also for healthy and abnormal cases, resulting in promising retrieval performance with respect to multi-modality, accuracy, speed and robustness.

This dissertation furnishes medical practitioners with a valuable set of tools for semantic retrieval of 2D images, where the human eye may fail. More specifically, the proposed retrieval algorithms provide medical practitioners ability in retrieving 2D MR brain images accurately and monitor disease progression in various lobes of the human brain, as well as ability to monitor disease progression in multiple patients simultaneously. Moreover,

the proposed semantic classification scheme can be extremely useful for semantic based categorization, clustering and annotation of images in MR brain databases. This research also provides a foundation to researchers in semantic based retrieval systems for expanding the existing toolsets in retrieval problems. The proposed research framework may evolve in a natural progression towards developing more powerful and robust retrieval systems.

Appendix A

Implementation of Fast 2D Dyadic Wavelet Transform

In this appendix, details of 2D dyadic wavelet transform is given for the sake of completeness, however, reader can get more details from original paper [210]. The two wavelets $\varphi^1(x,y)$ and $\varphi^2(x,y)$ can be written as separable products of functions of the x and y variables. Here $\varphi(x)$ belong to a class of 2π periodic functions $H(\omega)$, $G(\omega)$ and $K(\omega)$ that satisfy following constraints,

$$H(\omega) = e^{i\omega/2} \left(cos\left(\omega/2\right)\right)^{2n+1}, \tag{A.1}$$

$$G(\omega) = 4ie^{i\omega/2} cos\left(\omega/2\right), \tag{A.2}$$

$$K(\omega) = \frac{1 - |H(\omega)|^2}{G(\omega)}. \tag{A.3}$$

$$L(\omega) = \frac{1 + |H(\omega)|^2}{2}. \tag{A.4}$$

It can be proved (in [210]) that for $2n+1=3$, $\varphi(x,y)$ is a quadratic spline with compact support. These 2π periodic functions ($H(\omega)$, $G(\omega)$ and $K(\omega)$) can be viewed as transfer function of the discrete filters with finite impulse response which are presented in table A.1.

Table A.1: Finite Impulse Response of the Wavelet Filters

n	H	G	K	L
-3			0.0078125	0.0078125
-2			0.054685	0.046875
-1	0.125		0.171875	0.1171875
0	0.375	-2.0	-0.171875	0.65625
1	0.375	2.0	-0.054685	0.1171875
2	0.125		-0.0078125	0.046875
3				0.0078125

The two wavelets $\varphi^1(x,y)$ and $\varphi^2(x,y)$ are characterized by the discrete filters H, G, K and L. Moreover, L_p is the discrete filter obtained by putting 2^p-1 zeros between consecutive coefficients of the filter L. Let D be the Dirac filter whose impulse response is equal to 1 at 0 and 0 otherwise. Let $A*(H,L)$ be the separable convolutions of the rows and columns, respectively, of the image A with the 1-D filters H and L. Then the following algorithm computes the 2-D discrete wavelet transform of an image $S_1^d f$. At each scale 2^j, the algorithm decomposes $S_{2^j}^d f$ into $S_{2^{j+1}}^d f$, $W_{2^{j+1}}^{1,d} f$ and $W_{2^{j+1}}^{2,d} f$.

$j=0$

while $(j<J)$

$W_{2^{j+1}}^d f = \frac{1}{\lambda_j}.S_{2^j}^d f * G_j$

$S_{2^{j+1}}^d f = S_{2^j}^d f * H_j$

$j = j+1$

end of while

At each scale 2^j, $S_{2^j}^d f * G_j$ is divided by λ_j to obtain accurate measures of Lipschitz

exponents from the wavelet maxima. The values of λ_j are shown in table A.2.

Table A.2: Normalization coefficients λ_j

j	λ_j
1	1.50
2	1.12
3	1.03
4	1.01
5	1.00

The inverse wavelet transform algorithm reconstructs $S_1^d f$ from the discrete dyadic wavelet transform. At each scale 2^j, it reconstructs $S_{2^{j-1}}^d f$ from $S_{2^j}^d f$ and $W_{2^j}^d f$.

$j = J$

while $(j > 0)$

$$S_{2^{j-1}}^d f = \lambda_j . W_{2^j}^d f * K_{j-1} + S_{2^j}^d f * \widetilde{H}_{j-1}$$

$$j = j - 1$$

end of while

Here it is worth noting that at each scale, wavelet decompositions are of the same size as original image because filters are up-sampled. This produces convenient data which does not have artifacts due to downsampling in a traditional discrete wavelet domains.

Appendix B

Fourier Transform of Polar Sampled 2D Data

Here derivation is adapted from [233]. The Fourier transform of a continuous 2-D complex-valued function $f(x, y)$ is given as,

$$F(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2\pi i(xX+yY)} dxdy. \tag{B.1}$$

Substituting $r = \sqrt{x^2 + y^2}$ and $\theta = tan^{-1}y/x$, and changing the differentials $dxdy$ to $rdrd\theta$ equation B.1 can be written as,

$$F(X, Y) = \int_{0}^{2\pi} \int_{0}^{\infty} f(r, \theta) e^{-2\pi ir(Xcos\theta+Ysin\theta)} rdrd\theta, \tag{B.2}$$

Since angular sampling is done before radial sampling, equation B.2 can be re-arranged as

follows,

$$F(X,Y) = \int_0^{\infty}\int_0^{2\pi} f(r,\theta)e^{-2\pi ir(Xcos\theta+Ysin\theta)}rdrd\theta \quad (B.3)$$

$$= \int_0^{\infty}\int_0^{2\pi} f(r,\theta)e^{-2\pi ir(Rcos\Theta cos\theta+Rsin\Theta sin\theta)}rdrd\theta \quad (B.4)$$

$$= \int_0^{\infty}\int_0^{2\pi} f(r,\theta)e^{-2\pi irRcos(\theta-\Theta)} \quad (B.5)$$

where $R = \sqrt{X^2+Y^2}$ and $\Theta = tan^{-1}Y/X$. From [237] it is known that

$$e^{iucos\phi} = \sum_{n=-\infty}^{\infty} i^n J_n(u)e^{in\phi},$$

where J_n is the Bessel function of the order n, taking $u = 2\pi rR$ and $\phi = \theta - \Theta$, hence,

$$F(X,Y) = \int_0^{\infty}\int_0^{2\pi}\sum_{n=-\infty}^{\infty} f(r,\theta)i^n J_n(2\pi rR)e^{in(\Theta-\theta)}rdrd\theta, \quad (B.6)$$

The equation B.6 can be re-arranged as,

$$F(X,Y) = \int_0^{\infty}\sum_{n=-\infty}^{\infty}\left[\int_0^{2\pi} f(r,\theta)e^{-in\theta}d\theta\right] i^n J_n(2\pi rR)e^{in\Theta}rdr. \quad (B.7)$$

Here the integral in brackets is a 1-D Fourier transform with respect to θ. This transform over θ explicitly analyzes f into harmonic components of θ, indexed by their integer angular frequencies n. The amplitudes of the harmonic components in this transform provide weighting coefficients for the components used in the final synthesis of F, which vary with Θ as sinusoids and with R as Bessel functions of various orders.

Up to this point polar transformation in continuous domain is discussed, now the dis-

cretization of θ is discussed. With discrete values of θ this function in Cartesian and polar coordinates can be represented as,

$$s_N(x,y) = \sum_{i=1}^{N} \Delta(xsin\varphi_i + ycos\varphi_i), \tag{B.8}$$

$$s_N(r,\theta) = \sum_{i=1}^{N} \Delta(rcos\theta sin\varphi_i + rsin\theta cos\varphi_i), \tag{B.9}$$

where $\Delta(x)$ is defined as the 2-D, infinitely sharp line impulse at $x = 0$ and N is the number of angular samples from 0 to π. Hence FT can be written as,

$$S_N(X,Y) = \sum_{i=1}^{N} \int_{-\infty}^{\infty} \Delta(rcos\theta sin\varphi_i + rsin\theta cos\varphi_i) e^{-2\pi i r R'} |r| dr$$

This equation can be written in summation form as,

$$S_N(X,Y) = \sum_{i=1}^{N} S_{N,r}(R',\theta),$$

where $S_{N,r}$ can be written using convolution theorem,

$$S_{N,r}(R',\theta) = \Delta(R'cos\theta sin(\varphi_i + \pi/2) + R'sin\theta cos(\varphi_i + \pi/2)) * \int_{-\infty}^{\infty} |r| e^{2\pi i r R'} dr. \tag{B.10}$$

If r is limited to r_{max} beyond which $r = 0$ using a rectangular window then, integral in previous equation becomes,

$$\int_{-\infty}^{\infty} rect\left(\frac{r}{2r_{max}}\right) |r| e^{2\pi i r R'} dr = 2r_{max} sinc 2r_{max} R' - r_{max} sinc^2 r_{max} R'$$

and in the frequency domain,

$$S_{N,r'}(R',\theta) = \Delta\left(R'cos\theta sin(\varphi_i+\pi/2)+R'sin\theta cos(\varphi_i+\pi/2)\right) * (2r_{max}sinc2r_{max}R' - r_{max}sinc^2 r_{max}R') \quad \text{(B.11)}$$

If s is considered to be the integration for radii upto r_{max} of rings with radius r' each sampled N times over the range of 0 to π,

$$s_N(r,\theta) = \int_0^{r_{max}} \delta(r-r')(N/\pi)\aleph(\theta N/\pi)dr' \quad \text{(B.12)}$$

where $\aleph(x)$ is an infinite 1-D train of impulses spaced at integer increments in x is defined as,

$$\aleph(x) = \sum_{i=-\infty}^{\infty} \delta(x-i)$$

Alternatively,

$$s_{N,\theta}(r,n) = \int_0^{2\pi} (N/\pi)\aleph(\theta N/\pi)e^{-in\theta}d\theta = 2N\aleph(n/2N). \quad \text{(B.13)}$$

The factor $\aleph(n/2N)$ indicates that the $2N$ delta functions per ring of the sampling function can be synthesized from a series of waves around the ring, with frequencies that are integer multiples of the fundamental frequency $2N$. This can be substituted in equation B.7 to give,

$$S_N(X,Y) = \int_0^{r_{max}} \sum_{n=-\infty}^{\infty} i^n J_n(2\pi rR)e^{in\Theta}rdr, \quad n = 0, \pm 2N, \pm 4N...., \quad \text{(B.14)}$$

Up to this point r is considered as continuous now incorporating the discreteness in r in equation B.11 leads,

$$\begin{aligned} S_{N,r^d}(R',\theta) = \quad & \aleph\left(\tfrac{R'}{\Delta r} cos\theta sin(\varphi_i + \pi/2) + \tfrac{R'}{\Delta r} sin\theta cos(\varphi_i + \pi/2)\right) \\ & * (2r_{max} sinc 2r_{max} R' - r_{max} sinc^2 r_{max} R') \end{aligned} \tag{B.15}$$

Rather than a single ridge, the point response for each direction is now a series of ridges in the same direction $\theta_i + \pi/2$, replicated at the regular spacing $1/\Delta r$. This equation shows that the discrete polar FT is actually periodic in both r and θ.

Appendix C

Justifications for the Use of Support Vector Machines

There are two broad categories of the statistical classification schemes. First category is known as the Parametric, while the other is known as the Non-Parametric scheme. Inquisitive readers may find detailed treatments of this subject in the literature [238].

The parametric techniques, such as Gaussian Mixture Models (GMM), rely on the estimation of probability density functions (PDFs) for the training data. The estimation of the PDFs relies heavily on the nature and the size of the training set. The larger the dataset the better the results. As mentioned in Chapter 5, the vector length of feature set is 864 for each image with 217 slices in the transverse (axial) view. Obviously, the parametric techniques are less suitable in this application because of the large feature vector and a relatively small number of images in the volume.

There are several popular non-parametric classification techniques, viz. Bayesian, k-Nearest Neighbor (k-NN), Neural Networks and Support Vector Machines (SVM). These

non-parametric classification techniques have been used extensively in the area of machine intelligence, pattern analysis and recognition. The research reported in the literature is quite diverse, however there is a limited amount of research work on the justifications and comparison for these approaches. In [239], Linear Discrimination Analysis (LDA), k-NN, Bayes, and SVM based classifiers are compared and concluded that the SVM technique can find more complicated decision boundaries and the classification errors of k-NN and SVM are considerably lower than others. More recently, k-NN and SVM are directly compared in [240] on a variety of datasets, and concluded that the kNN is dominant on datasets with relatively low sparsity. On datasets with high to extremely high level of sparsity, k-NN starts performing poorly as it is unable to form reliable neighborhoods. Recently, neural networks and SVM are studied in text recognition framework in [241], where SVM has demonstrated superior classification accuracies as compared to the neural network based classifiers in many experiments.

In many image retrieval subsystems, such as Relevance Feedback (RF), SVM has been popular due to its ability to work well on sparse data. Besides, a recent research in [19] indicates SVM to be suitable for medical image retrieval.

Due to these observations, the use of SVM is justified for semantic classification and identification tasks, proposed in this dissertation.

References

[1] H. Mller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications–clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, pp. 1–23, 2004. xvii, 12, 13, 14, 23, 34, 35, 36, 37, 38, 39, 43

[2] *Ontario's eHealth Strategy 2009-2012*, eHealth Ontario. [Online]. Available: http://www.ehealthontario.on.ca/pdfs/About/eHealthStrategy.pdf 2

[3] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, December 2000. 5, 6, 9, 19, 22, 34

[4] A.Vadivel, S. Suralband, and A. K. Majumdar, "Human color perception in the hsv space and its application in histogramgeneration for image retrieval," in *Color Imaging X: Processing, Hardcopy, and Applications*, G. G. M. Reiner Eschbach, Ed., vol. 5667. SPIE, 2005, pp. 598–609. 6

[5] D. Zhang and G. Lu, "Review of shape representation andd escription techniques," *Pattern Recognition*, vol. 37, pp. 1–19, 2004. 7

[6] M. Christel, N. Moraveji, and C. Huang, "Evaluating content-based filters for image and video retrieval," in *SIGIR*. ACM, 2004, pp. 590–591. 7

[7] B. Ko and H. Byun, "Frip: A region-based image retrieval tool using automatic image segmentationand stepwise boolean and matching," *IEEE Transactions On Multimedia*, vol. 7, no. 1, pp. 105–113, February 2005. 7

[8] W. Jiang, K. L. Chan, M. Li, and H. Zhang, "Mappinglow-level features to high-level semantic concepts in region-basedimage retrieval," in *Conference on Computer Vision and Pattern Recognition*. IEEE, 2005, pp. 244–249. 7

[9] J. Fana, Y. Gaoa, H. Luo, and G. Xu, "Statistical modeling and conceptualization of natural images," *Pattern Recognition*, vol. 38, pp. 865–885, 2005. 7, 20, 25, 28

[10] S.-Y. Dai and Y.-J. Zhang, "Unbalanced region matching based on two-level description for image retrieval," *Pattern Recognition Letters*, vol. 26, pp. 565–580, 2005. 7

[11] R. Brown and B. Pham, "Image mining and retrieval using hierarchical support vector machines," in *Proceedings of the 11th International Multimedia Modelling Conference*. IEEE, 2005, pp. 446–451. 7, 20

[12] F. C. A.Quddus and M.Gabbouj, "Wavelet-based multi-level object retrieval in contour images," in *International Workshop on Very Low Bit Rate Video Coding (VLBV'99)*, Kyoto, Japan, 1999. 8, 25

[13] S. Santini and R. Jain, "Similarity measures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, pp. 871–883, 1999. 8, 25

[14] R. Veltkamp and M. Hagendoorn, *State-of-the-Art in Shape Matching*. Springer-Verlag, 2000, ch. Multimedia Search: State of the Art. 8

[15] F. Mokhtarian and S. Abbasi, "Shape similarity retrieval under afine transforms," *Pattern Recognition*, vol. 35, pp. 31–41, 2002. 8, 25

[16] A.Quddus, F. Cheikh, and M.Gabbouj, "Content-based object retrieval using maximum curvature points in contourimages," in *SPIE/EI'2000 Symposium, On Storage and Retrieval for Media Databases*. SPIE, 2000. 8

[17] L. Jia and L. Kitchen, "Object-based image similarity computation using inductive learning of contour-segmentrelations," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 80–87, 2000. 8

[18] N. Alajlan, M. S. Kamel, and G. H. Freeman, "Geometry-based image retrieval in binary image databases," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 30, pp. 1003–1013, 2008. 8

[19] M. M. Rahman, P. Bhattacharya, and B. C. Desai, "A unified image retrieval framework on local visual and semantic concept-based feature spaces," *J. Vis. Commun. Image R.*, vol. 20, pp. 450–462, 2009. 8, 27, 30, 32, 41, 141

[20] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S. C. Hoi, and M. Satyanarayanan, "A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 30–44, 2010. 8, 31, 41

[21] M. M. Rahman, P. Bhattacharya, and B. C. Desai, "A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback," *IEEE Transactions On Information Technology In Biomedicine*, vol. 11, pp. 58–69, 2007. 8, 41

[22] E. D. Sciascio, M. Mongiello, F. Donini, and L. Allegretti, "Retrieval by spatial similarity: an algorithm and a comparative evaluation," *Pattern Recognition Letters*, vol. 25, pp. 1633–1645, 2004. 9

[23] R. Krishnapuram, S. Medasani, S.-H. Jung, Y.-S. Choi, and R. Balasubramaniam, "Content-based image retrieval based on a fuzzy approach," *IEEE Transactions On Knowledge And Data Engineering*, vol. 16, no. 10, pp. 1185–1199, 2004. 9

[24] B. Moghaddam, H. Biermann, and D. Margaritis, "Image retrieval with local and spatial queries," in *International Conference on Image Processing*. IEEE, 2000, pp. 542–545. 9, 10, 24

[25] G. Aggarwal, T. V. Ashwin, and S. Ghosal, "An image retrieval system with automatic query modification," *IEEE Transactions On Multimedia*, vol. 4, pp. 201–214, 2002. 9, 10, 24

[26] K. Vu, K. A. Hua, and W. Tavanapong, "Image retrieval based on regions of interest," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 1045–1049, 2003. 9, 10

[27] K. Barnard and D. Forsyth, "Learning the semantics of words and pictures," in *Proc. International Conference on Computer Vision*, 2001, pp. 408–415. 10

[28] T. S. Huang, X. S. Zhou, M. Nakazato, Y. Wu, and I. Cohen, "Learning in content-based image retrieval," in *International Conference on Development and Learning*. IEEE, 2002, pp. –. 10, 32

[29] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 25, pp. 1075–1088, September 2003. 10

[30] T. Quack, U. Monich, L. Thiele, and B. Manjunath, "Cortina: A system for large-scale, content-based web image retrieval," in *International Conference on Multimedia*. ACM, 2004, pp. 508–511. 10, 21

[31] X.-J. Wang, W.-Y. Ma, G.-R. Xue, and X. Li, "Multi-model similarity propagation and its application for web image retrieval," in *International Conference on Multimedia*. ACM, 2004, pp. 944–951. 10

[32] F. Jing, M. Li, H.-J. Zhang, , and B. Zhang, "A unified framework for image retrieval using keyword and visual features," *IEEE Transactions On Image Processing*, vol. 14, no. 7, pp. 979–989, 2005. 10

[33] C. Lacoste, J.-H. Lim, J.-P. Chevallet, and D. T. H. Le, "Medical-image retrieval based on knowledge-assisted text and image indexing," *IEEE Transactions On Circuits And Systems For Video Technology*, vol. 17, pp. 889–900, 2007. 10

[34] S. Ghebreab, C. C. Jaffe, and A. W. M. Smeulders, "Population-based incremental interactive concept learning for image retrievalby stochastic string segmentations," *IEEE Transactions On Medical Imaging*, vol. 23, no. 6, pp. 676–689, June 2004. 10, 20, 25, 28, 29

[35] A. Grigorova and F. G. D. Natale, "Feature-adaptive relevance feedback fa-rf for content-based image retrieval," in *Storage and Retrieval Methods and Applications for Multimedia 2004*, C.-S. L. Minerva M. Yeung, Rainer W. Lienhart, Ed., vol. 5307. SPIE, 2004, pp. 20–28. 10

[36] M. Ferecatu, M. Crucianu, and N. Boujemaa, "Retrieval of difficult image classes using svm based relevance feedback," in *SIGMM MIR*. ACM, 2004, pp. 23–30. 10

[37] C. Hoi and M. R. Lyu, "A novel logbased relevance feedback technique in content-based image retrieval," in *International Conference on Multimedia*. ACM, 2004, pp. 24–31. 10, 33

[38] A. Winter and R. Haux, "A three-level graph-based model for the management of hospital information systems," *Methods Information Med.*, vol. 34, pp. 378–396, 1995. 12, 34

[39] C. LeBozec, M.-C. Jaulent, E. Zapletal, and P. Degoulet, "Unified modeling language and design of a case-based retrieval system in medical imaging," in *Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA)*, 1998. 12, 34

[40] J.-P. Boissel, M. Cucherat, E. Amsallem, P. Nony, M. Fardeheb, W. Manzi, and M. Haugh, "Getting evidence to prescribers and patients or how to make ebm a reality," in *Proceedings of the Medical Informatics Europe Conference (MIE)*, 2003. 12, 34

[41] B. Kaplan and H. Lundsgaarde, "Toward an evaluation of an integrated clinical imaging system: Identifying clinical benefits," *Methods Inform. Med.*, vol. 35, pp. 221–229, 1996. 12, 34, 42

[42] H. W, M. H, and K.-C. J., "The imageclefmed medical image retrieval task test collection," *J. Digit. Imaging.*, vol. 22, pp. 648–655, 2009. 13, 44

[43] S.-K. Chang and T. Kunii, "Pictorial data-base applications," *IEEE Computer*, vol. 14, pp. 13–21, 1981. 19

[44] P. Enser, "Pictorial information retrieval," *J. Document*, vol. 51, pp. 126–170, 1995. 19

[45] A. Gupta and R. Jain, "Visual information retrieval," *Commun. ACM*, vol. 40, no. 5, pp. 70–79, 1997. 19, 20, 21

[46] A. del Bimbo, *Visual Information Retrieval*. Academic Press, New York, 1999. 19

[47] S. Rahman, *Design & Management of Multimedia Information Systems: Opportunities & Challenges,*. Idea Group Publishing, London,, 2001. 19

[48] N.-S. Chang and K.-S. Fu, "Query-by-pictorial-example," *IEEE Trans. Software Eng.*, vol. 6, pp. 519–524, 1980. 20

[49] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. H. andB. Dom, M. Gorkani, J. Hafher, D. Lee, D. Petkovie, and D. S. P. Yanker, "Query by image and video content: The qbic system," *IEEE Computer Magazine*, pp. 23–32, September 1995. 20, 21, 22

[50] P. Kelly, M. Cannon, and D. Hush, "Query by image example: the candid approach," in *Proceedings of the of SPIE Conference on Storage and Retrieval for Image and Video Databases III*, W. Niblack and R. Jain, Eds., vol. 2420, 1995, pp. 238–248. 20

[51] A. Pentland, R. Picard, and S. Sclaro, "Photobook: Content-based manipulation for image databases," *J. Comput. Vis.*, vol. 18, pp. 233–254, 1996. 20

[52] W. Y. Ma and B. S. Manjunath, "Netra: A toolbox for navigating large image databases," in *International Conference on Image Processing.* IEEE, 1997, pp. 568–571. 20, 21

[53] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its applicationto image querying," *IEEE Transactions On Pattern Analysis And Machine Intelligence*, vol. 24, no. 8, pp. 1026–1038, August 2002. 20, 21, 22

[54] I. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N.Yianilos, "The bayesian image retrieval system, pichunter: Theory, implementation,and psychophysical experiments," *IEEE Transactions On Image Processing*, vol. 9, no. 9, pp. 20–37, JANUARY 2000. 20, 21, 30, 32

[55] D. M. Squire, W. Muller, H. Muller, and T. Pun, "Content-based query of image databases: inspirations from text retrieval," *Pattern Recognition Letters*, vol. 21, pp. 1193–1198, 2000. 20, 21, 22

[56] R. Brunelli and O. Mich, "Image retrieval by examples," *IEEE Transactions On Multimedia*, vol. 2, no. 3, pp. 164–171, 2000. 20, 21, 32

[57] H. Muller, A. Rosset, J.-P. Vallee, and A. Geissbuhler, "Comparing feature sets for content-based image retrieval in a medical casedatabase," in *Medical Imaging 2004:*

PACS and Imaging Informatics, H. K. H. Osman M. Ratib, Ed., vol. 5371. SPIE, 2004, pp. 99–109. 20, 21

[58] R. C. Veltkamp and M. Tanase, "Content-based image retrieval systems: A survey," Department of Computing Science, Utrecht University, Tech. Rep., 2002. 20

[59] T. Gevers and A. Smeulders, "Pictoseek: Combining color and shape invariant features for image retrieval," *IEEE Transactions On Image Processing*, vol. 9, no. 1, pp. 102–119, 2000. 21

[60] J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 947–963, 2001. 21, 28

[61] Y. Chen and J. Z. Wang, "A region-based fuzzy feature matching approach to content-based image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002. 21

[62] S. Mehrotra, Y. Rui, M. Ortega-Binderberger, and T. Huang, "Supporting content-based queries over images in mars," in *International Conference on Multimedia Computing and Systems*. IEEE, 1997, pp. 632–633. 21

[63] J. Wang, G. Wiederhold, O. Firschein, and X. Sha, "Content-based image indexing and searching using daubechies' wavelets," *International Journal of Digital Libraries*, vol. 1, no. 4, pp. 311–328, 1998. 21

[64] A. Bimbo, M. Mugnaini, P. Pala, and F. Turco, "Visual query for color perceptive regions," *Pattern Recognition*, vol. 31, pp. 1241–1253, 1998. 21

[65] M. Haas, J. Rijsdam, B. Thomee, and M. S. Lew, "Relevance feedback: Perceptual learning and retrieval in bio-computing,photos, and video," in *SIGMM MIR*. ACM, 2004, pp. 151–156. 21, 32

[66] F. Cheikh, B.Cramariuc, C.Reynaud, M.Quinghong, B.-A. B.Hnich, M.Gabbouj, P.Kerminen, T.Mäkinen, and H.Jaakkola, "Muvis: a system for content-based indexing and retrieval in large imagedatabases," in *SPIE/EI'99 Conference on Storage and Retrieval for Image and Video Databases*, San Jose, California, January 1999. 21

[67] F. A. Cheikh, "Muvis: A system for content-based image retrieval," Ph.D. dissertation, Tampere University of Technology,, Tampere, Finland, 2004. 21

[68] J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "Picsom - content-based image retrieval with self-organizing maps," *Pattern Recognition Letters*, vol. 21, pp. 1199–1207, 2000. 21

[69] A. Natsev, R. Rastogi, and K. Shim, "Walrus: A similarity retrieval algorithm for image databases," *IEEE Transactions On Knowledge And Data Engineering*, vol. 16, no. 3, pp. 301–316, 2004. 21

[70] G. Schaefer and M. Stich, "Ucid - an uncompressed colour image database," in *Storage and Retrieval Methods and Applications for Multimedia 2004*, M. M. Yeung, R. W. Lienhart, and C.-S. Li, Eds., vol. 5307. SPIE, 2004, pp. 472–480. 21

[71] H. L. Tang, R. Hanka, and H. H. S. Ip, "Histological image retrieval based on semantic content analysis," *IEEE Transactions On Information Technology In Biomedicine*, vol. 7, no. 1, pp. 26–36, 2003. 21

[72] M. Bruzzo, F. Giordano, and S. Dellepianea, "Kingfisher: a system for remote sensing image database management," in *Sensors, Systems, and Next-Generation Satellites VI*, H. Fujisada, J. B. Lurie, M. L. Aten, and K. Weber, Eds., vol. 4881. SPIE, 2003, pp. 669–676. 21

[73] T. Glatard, J. Montagnat, and I. Magnin, "Texture based medical image indexing and retrieval: Application to cardiacimaging," in *SIGMM MIR*. ACM, 2004, pp. 135–142. 20

[74] V. D. Lecce, A. Guerriero, I. Guarino, and P. DiBari, "A beowulf class parallel remote sensed image database retrieval system developedin assist environment," in *Storage and Retrieval Methods and Applications for Multimedia 2005*, N. B. Rainer W. Lienhart, Ed., vol. 5682. SPIE, 2005, pp. 1–9. 20

[75] A. Csillaghy, H. Hinterberger, and A. Benz, "Content-based image retrieval in astronomy," *Information Retrieval*, vol. 3, pp. 229–241, 2000. 20

[76] J. Landre' and F. Truchetet, "Multiresolution hierarchical content-based image retrieval of paleontologyimages," in *Wavelet Applications in Industrial Processing*, F. Truchetet, Ed., vol. 5266. SPIE, 2004. 20

[77] Y. Lu and C. L. Tan, "Document retrieval from compressed images," *Pattern Recognition*, vol. 36, pp. 987–996, 2003. 20

[78] X. Wang, K. Suzuki, H. Ikeda, and J. Suzuki, "Hd image retrieval utilizing networked image directory," *IEEE Transaction on Consumer Electronics*, vol. 48, no. 4, pp. 925–931, 2002. 20

[79] J. Wolf, W. Burgard, and H. Burkhardt, "Robust vision-based localization by combining an image-retrieval systemwith monte carlo localization," *IEEE Transactions on Robotics*, vol. 21, no. 2, pp. 208–216, 2005. 20

[80] B. Zhu, M. Ramsey, and H. Chen, "Creating a large-scale content-based airphoto image digital library," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 163–167, 2000. 20

[81] T. Zhao, L. H. Tang, H. H. Ip, and F. Qi, "On relevance feedback and similarity measure for image retrieval with synergeticneural nets," *Neurocomputing*, vol. 51, pp. 105–124, 2003. 20

[82] J. Fan, Y. Gao, H. Luo, and G. Xu, "Automatic image annotation by using concept-sensitive salient objects forimage content representation," in *SIGIR*. ACM, 2004, pp. 361–386. 20, 28

[83] S. Sclaroff, L. Taycher, and M. L. Cascia, "Imagerover: A content-based browser for the world wide web,," in *IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1997. 22

[84] T. Gevers and A. Smeulders, "Color-metric pattern-card matching for viewpoint invariant image retrieval," in *International Conference on Computer Vision and Pattern Recognition*. IEEE, 1996, pp. 3–7. 22

[85] J.-M. Geusebroek, R. van den Boogaard, A. Smeulders, and H. Geerts, "Color invariance," *IEEE Trans. PAMI*, vol. 23, pp. 1338–1350, 2001. 22

[86] B. S. Manjunath, P. Salembier, and T. Sikora, Eds., *Introduction to MPEG-7: Multimedia Content Description Interface,*. Wiley, New York, 2002. 23, 25

[87] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, pp. 262–282, 2007. 23, 25, 26, 28, 29

[88] A. C. Bovik, M. Clark, and W. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 55–73, 1990. 23, 24

[89] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception ." *IEEE Transactions on Systems Man and Cybernetics*, vol. 8, pp. 460–473, 1978. 24

[90] F. Liu and R. Picard, "Periodicity, directionality, and randomness: wold features for image modeling and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, pp. 722–733, 1996. 24

[91] W. Leow and S. Lai, *Scale and orientation-invariant texture matching for image retrieval, Texture Analysis in Machine Vision*, M. Pietikainen, Ed. World Scientific, Singapore, 2000. 24

[92] D. Dunn and W. Higgins, "Optimal gabor filters for texture segmentation," *IEEE Transactions on Image Processing*, vol. 4, pp. 947–964, 1995. 24

[93] P. Huang and S. Dai, "Image retrieval by texture similarity," *Pattern Recognition*, vol. 36, pp. 665–679, 2003. 24

[94] J. Daugman, "An information theoretic view of analog representation in striate cortex," *Computational Neuroscience*, pp. 9–18, 1990. 24

[95] J. Weszka, C. Dyer, and A. Rosenfeld, "A comparative study of texture measures for terrain classification," *IEEE Trans. Sys. Man Cybernetics*, vol. 6, pp. 269–285, 1976. 24

[96] S. Belkasim, X. Hong, and O. Basir, "Content based image retrieval using discrete wavelet transform," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, pp. 19–32, 2004. 24

[97] J. A. Lay and L. Guan, "Image retrieval based on energy histograms of the low frequency dctcoefficients," in *Proc. ICASSP*. IEEE, 1999, pp. 3009–3012. 24

[98] M.-C. Lee and C.-M. Pun, "Rotation and scale invariant wavelet feature for content-based texture imageretrieval," *Journal Of The American Society For Information Science And Technology*, vol. 54, no. 1, pp. 68–80, 2003. 24

[99] D. Comaniciu, P. Meer, D. Foran, and A. Medl, "Bimodal system for interactive indexing and retrieval of pathology images," in *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision (WACV98)*, 1998. 24, 40

[100] S. Perry and P. Lewis, "A novel image viewer providing fast object delineation for content based retrieval and navigation," in *Proceedings of the of SPIE Conference on Storage and Retrieval for Image and Video Databases VI*, I. Sethi and R. Jain, Eds., vol. 3312, 1997. 24

[101] Y. Gao and Y. Qi, "Robust visual similarity retrieval in single model face databases," *Pattern Recognition*, vol. 38, pp. 1009–1020, 2005. 25

[102] S.-H. Cha and S. N. Srihari, "On measuring the distance between histograms," *Pattern Recognition*, vol. 35, pp. 1355–1370, 2002. 25

[103] R. O. Stehling, M. A. Nascimento, and A. X. Falcao, "On "shapes" of colors for content-based image retrieval," in *International Workshop on Multimedia*. ACM, 2000, pp. 171–174. 25

[104] S. Krishnamachari and M. Abdel-Mottaleb, "Hierarchical clustering algorithm for fast image retrieval," in *SPIE Conference on Storage and Retrieval for Image and Video Databases*. SPIE, January 1999, pp. 427–435. 25

[105] D. Androutsos, K. Plataniotis, and A. Venetsanopoulos, "Distance measures for color image retrieval," in *Proceedings of the International Conference on Image Processing*, 1998. 26

[106] Z. Chen and B. Zhu, "Some formal analysis of rocchios similarity-based relevance feedback algorithm," *Inf. Retr.*, vol. 5, 2002. 26

[107] F. Jing, M. Li, H.-J. Zhang, and B. Zhang, "An efficient and effective region-based image retrieval framework," *IEEE Transactions On Image Processing*, vol. 13, pp. 699–709, May 2004. 27

[108] B. Li, E. Chang, and Y. Wu, "Discovery of a perceptual distance function for measuring image similarity," *Multimedia Systems*, vol. 8, pp. 512–522, 2003. 27

[109] M. M. Rahman, B. C. Desai, and P. Bhattacharya, "Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," *Computerized Medical Imaging and Graphics*, vol. 32, pp. 95–108, 2008. 27, 30, 41

[110] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. Academic Press, 1990. 27

[111] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, "A memory learning framework for effective image retrieval," *IEEE Transactions On Image Processing*, vol. 14, no. 4, pp. 511–524, April 2005. 28, 29

[112] H. Feng, R. Shi, and T.-S. Chua, "A bootstrapping framework for annotating and retrieving www images," in *International Conference on Multimedia*. ACM, 2004, pp. 960–967. 28, 29, 30

[113] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "An ontology approach to object-based image retrieval," in *International Conference on Image Processing*. IEEE, 2003, pp. 511–514. 28, 29

[114] N. Maillot, M. Thonnat, and C. Hudelot, "Ontology based object learning and recognition : Application to image retrieval," in *International Conference on Tools with Artificial Intelligence (ICTAI'04)*. IEEE, 2004, pp. 620 – 625. 28, 29

[115] S. Jiang, T. Huang, and W. Gao, "An ontology-based approach to retrieve digitized art images," in *International Conference on Web Intelligence (WI04)*. IEEE, 2004, pp. 131 – 137. 28, 29

[116] G. Allampalli-Nagaraj and I. Bichindaritz, "Automatic semantic indexing of medical images using a web ontology language for case-based image retrieval," *Engineering Applications of Artificial Intelligence*, vol. 22, pp. 18–25, 2009. 28, 30

[117] Y. Chen, J. Z. Wang, and R. Krovetz, "Clue: Cluster-based retrieval of images by unsupervised learning," *IEEE Transaction on Image Processing*, 2004. 28

[118] B. Li and S. Yuan, "A novel relevance feedback method in content-based image retrieval," in *International Conference on Information Technology: Coding and Computing(ITCC04)*. Los Alamitos, CA, USA: IEEE Computer Society, 2004, pp. 120–123. 29, 32

[119] I. El-Naqa, Y. Yang, N. P. Galatsanos, R. M. Nishikawa, and M. N.Wernick, "A similarity learning approach to content-based image retrieval: Applicationto digital mammography," *Ieee transactions on medical imaging*, vol. 23, no. 10, pp. 1233–1244, October 2004. 29, 30, 37

[120] P. H. Gosselin and M. Cord, "A comparison of active classification methods for content-based image retrieval," in *International Worshop on Computer Vision and Database*. ACM, 2004, pp. 51–58. 29

[121] Y. Zhuang and Y. P. X. Liu, "Apply semantic template to support content-based image retrieval," in *Proceedings of the SPIE, Storage and Retrieval for Media Databases*, vol. 3972, 1999, p. 442449. 29

[122] D. Cai, X. He, Z. Li, and W.-Y. Ma, "Hierachical clustering of wwwimage search results using visual, textual and link information 2004." in *Proceedings of the ACM International Conference on Multimedia*, 2004. 29

[123] T. Takagi, K. Kawase, K. Otsuka, and T. Yamaguchi, "Data retrieval using conceptual fuzzy sets," in *International Conference on Fuzzy Systems*. IEEE, 2000, pp. 94–99. 29

[124] A. Chavez-Aragon and O. Starostenko, "Ontological shape-description, a new method for visual information retrieval," in *International Conference on Electronics, Communications and Computers (CONIELECOMP04)*. IEEE, 2004, pp. 288 – 292. 29

[125] Y. Chen and J. Z. Wang, "Image categorization by learning and reasoning with regions," *Journal of Machine Learning Research*, vol. 5, pp. 913–939, 2004. 30

[126] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance feedback in content-based image retrieval: Bayesian framework,feature subspaces, and progressive learning," *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 924–937, 2003. 30

[127] J.-H. Lima and J. S. Jin, "Combining intra-image and inter-class semantics for consumer image retrieval," *Pattern Recognition*, vol. 38, pp. 847–864, 2005. 30

[128] R. Shi, H. Feng, T.-S. Chua, and C.-H. Lee, "An adaptive image content representation and segmentation approach to automatic image annotation," in *Proc. International Conference on Image and Video Retrieval (CIVR)*, 2004, pp. 545–554. 30

[129] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang, "Image classification for content-based indexing," *IEEE Trans. Image Processing*, vol. 10, no. 1, pp. 117–130, 2001. 30

[130] J. Luo and A. Savakis, "Indoor vs outdoor classification of consumer photographs using low-level and semantic features," in *Proc. International Conference on Image Processing (ICIP)*, vol. II, 2001, pp. 745–748. 30

[131] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, New York, 2001. 31

[132] I. Sethi and I. Coman, "Mining association rules between low-level image features and high-level concepts," in *Proceedings of the SPIE Data Mining and Knowledge Discovery*, vol. III, 2001, pp. 279–290. 31

[133] D. Stan and I. Sethi, "Mapping low-level image features to semantic concepts," in *Proceedings of the SPIE: Storage and Retrieval for Media Databases*, 2001, pp. 172–179. 31

[134] Y. Choi, D. Kim, and R. Krishnapuram, "Relevance feedback for content-based image retrieval using the choquet integral," in *International Conference on MultiMedia and Expo.* IEEE, 2000, pp. 1207–1210. 32

[135] Y. Zhuang, J. Yang, Q. Li, and Y. Pan, "A graphic-theoretic model for incremental relevance feedback in image retrieval," in *International Conference on Image Processing.* IEEE, 2002, pp. 413–416. 32

[136] F. Cheikh, B. Cramariuc, and M. Gabbouj, "Relevance feedback for shape query refinement," in *International Conference on Image Processing*. IEEE, 2003, pp. 9–12. 32

[137] R. Singh and R. Kothari, "Relevance feedback algorithm based on learning from labeled and unlabeled data," in *International Conference on Multimedia and Expo*. IEEE, 2003, pp. 433–436. 32

[138] X. Zhou, Q. Zhang, L. Liu, L. Zhang, and B. Shi, "An image retrieval method based on analysis of feedback sequence log," *Pattern Recognition Letters*, vol. 24, pp. 2499–2508, 2003. 32, 33

[139] X. He, "Incremental semi-supervised subspace learning for image retrieval," in *International Conference on Multimedia*. ACM, 2004, pp. 2–8. 32

[140] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Query feedback for interactive image retrieval," *IEEE Transactions On Circuits And Systems For Video Technology*, vol. 14, no. 5, May 2004. 32

[141] J. Luo and M. A. Nascimento, "Content-based sub-image retrieval using relevance feedback," in *International Workshop on Multimedia Databases*. ACM, 2004, pp. 2–9. 32

[142] E. Chang and S. Tong, "Svmactive-support vector machine active learning for image retrieval, , october 2001, pp. 107118." in *Proceedings of the ACM International Multimedia Conference*, 2001. 33

[143] A. Doulamis and N. Doulamis, "Generalized nonlinear relevance feedback for iterative content-based retrieval and organization," *IEEE Trans. CSVT*, vol. 14, pp. 656–671, 2004. 33

[144] H. Zhang, Z. Chen, M. Li, and Z. Su, "Relevance feedback and learning in content-based image search," *World Wide Web: Internet and Web Information Systems*, vol. 6, pp. 131–155, 2003. 33

[145] P. Aggarwal, H. Sardana, and G. Jindal, "Content based medical image retrieval: Theory, gaps and future directions," *ICGST-GVIP Journal*, vol. 9, pp. 27–37, 2009. 34

[146] D. Demner-Fushman, S. Antani, M. Simpson, and G. R. Thoma, "Annotation and retrieval of clinically relevant images," *International Journal of Medical Informatics*, vol. 78, p. e59e67, 2009. 34

[147] S. Antani, L. Long, and G. Thoma, "A biomedical information system for combined content-based retrieval of spine x-ray images and associated text information," in *Proceedings of the Third Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2002. 34, 39

[148] M. Gld, M. Kohnen, D. Keysers, H. Schubert, B. Wein, J. Bredno, and T. Lehmann, "Quality of dicom header information for image categorization," in *Proceedings of the International Symposium on Medical Imaging*, vol. 4685, 2002, pp. 280–287. 34

[149] H. Tagare, C. Jaffe, and J. Duncan, "Medical image databases: A content-based retrieval approach," *Journal of American Medical Informatics Association*, vol. 4, pp. 184–198, 1997. 35

[150] W. Bidgood, B. Bray, N. Brown, A. Mori, K. Spackman, A. Golichowsky, R. Jones, L. Korman, B. Dove, L. Hildebrand, and M. Berg, "Image acquisition context: procedure description attributes for clinically relevant indexing and selective retrieval of biomedical images," *Journal of American Medical Informatics Association*, vol. 6, pp. 61–75, 1999. 35

[151] T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, D. K. M. Kohnen, H. Schubert, and B. B. Wein, "Content-based image retrieval in medical applications for picture archivingand communication systems," in *Medical Imaging 2003: PACS and Integrated Medical Information Systems: Designand Evaluation*, O. M. R. H. K. Huang, Ed., vol. 5033. SPIE, 2003, pp. 109–117. 35

[152] A. Rosset, O. Ratib, A. Geissbuhler, and J.-P. Valle, "Integration of a multimedia teaching and reference database in a pacs environment, 22 (6) (2002)," *RadioGraphics*, vol. 22, pp. 1567–1577, 2002. 37

[153] A. Mojsilovic and J. Gomes, "Semantic based categorization, browsing and retrieval in medical image databases," in *International Conference on Image Processing*. IEEE, 2002, pp. 145–148. 37, 38, 41

[154] Y. Liu and F. Dellaert, "Classification-driven medical image retrieval," in *Proceedings of the ARPA Image Understanding Workshop*, 1997. 37

[155] A. Constantinidis, M. Fairhurst, and A. Rahman, "A new multi-expert decision combination algorithm and its application to the detection of circumscribed masses in digital mammograms," *Pattern Recognition*, vol. 34, pp. 1527–1537, 2001. 37

[156] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Wavelet optimization for content-based image retrieval in medical databases," *Medical Image Analysis*, vol. 14, p. 227241, 2010. 37

[157] F. Meyer, "Automatic screening of cytological specimens," *Comput. Vis. Graphics Image Proces.*, vol. 35, pp. 356–369, 1986. 37

[158] M. Mattie, L. Staib, E. Stratmann, H. Tagare, J. Duncan, and P. Miller, "Pathmaster: Content-based cell image retrieval using automated feature extraction," *J. Am. Med. Informatics Assoc.*, vol. 7, pp. 404–415, 2000. 37

[159] K. Veropoulos, C. Campbell, and G. Learnmonth, "Image processing and neural computing used in the diagnosis of tuberculosis," in *Proceedings of the Colloquium on Intelligent Methods in Healthcare and Medical Applications (IMHMA)*, 1998. 37

[160] M.-C. Jaulent, C. L. Bozec, Y. Cao, E. Zapletal, and P. Degoulet, "A property concept frame representation for exible image content retrieval in histopathology databases," in *Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA)*, 2000. 37, 39

[161] L. H. Tang, R. Hanka, H. H. S. Ip, and R. Lam, "Extraction of semantic features of histological images for content-basedretrieval of images," in *Proc. Symposium on Computer-Based Medical Systems*. IEEE, 2000, pp. –. 37, 51

[162] G. Robinson, H. Targare, J. Duncan, and C. Jaffe, "Medical image collection indexing: shape-based retrieval using kd-trees," *Comput. Vis. Graphics Image Proces.*, vol. 20, pp. 209–217, 1996. 37

[163] P. Schmidt-Saugeon, J. Guillod, and J.-P. Thiran, "Towards a computer-aided diagnosis system for pigmented skin lesions," *Comput. Med. Imag. Graphics*, vol. 27, pp. 65–78, 2003. 37

[164] A. Sbober, C. Eccher, E. Blanzieri, P. Bauer, M. Cristifolini, G. Zumiani, and S. Forti, "A multiple classifier system for early melanoma diagnosis," *Artificial Intel. Med.*, vol. 27, pp. 29–44, 2003. 37

[165] M. Ogiela and R. Tadeusiewicz, "Semantic-oriented syntactic algorithms for content recognition and understanding of images in medical databases," in *Proceedings of the second International Conference on Multimedia and Exposition (ICME)*. IEEE Computer Society, 2001, pp. 621–624. 37, 39

[166] A. M. Aisen, L. S. Broderick, H. Winter-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C.-R. Shyu, and A. Marchiori, "Automated storage and retrieval of thin-section ct images to assist diagnosis: System description and preliminary assessment," *Radiology*, vol. 228, pp. 265–270, 2003. 37

[167] D. Hansell, "High-resolution ct of diffuse lung disease," *Radiol. Clin. North Am.*, vol. 39, pp. 1091–1113, 2001. 38

[168] H. Abe, H. MacMahon, R. Engelmann, J. S. Q. Li, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. Metz, and K. Doi, "Computer-aided diagnosis in chest radiography: Results of large-scale observer tests at the 1996-2001 rsna scientific assemblies," *RadioGraphics*, vol. 23, pp. 255–265, 2003. 38

[169] C. Han, H. Chen, L. He, and W. Wee, "A web-based distributed image processing system," in *Proceedings of the SPIE Photonics West Conference on Internet Imaging IV*, S. Santini and R. Schettini, Eds., vol. 5018, 2003, pp. 111–122. 38

[170] E. A. El-Kwae and M. R. Kabuka, "Efficient content-based indexing of large image databases," *ACM Transactions on Information Systems*, vol. 18, pp. 171–210, 2000. 39, 41

[171] T. Ikeda and M. Hagiwara, "Content-based image retrieval system using neural networks," *International Journal of Neural Systems*, vol. 10, pp. 417–424, 2000. 39

[172] C. L. Bozec, E. Zapletal, M.-C. Jaulent, D. Heudes, and P. Degoulet, "Towards content-based image retrieval in his-integrated pacs," in *Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA)*, 2000, pp. 477–481. 39

[173] W. Hersh, M. Mailhot, C. Arnott-Smith, and H. Lowe, "Selective automated indexing of findings and diagnoses in radiology reports," *J. Biomed. Informatics*, vol. 34, pp. 262–273, 2001. 39

[174] A. Toga and P. Thompson, *An introduction to brain warping.* Brain Warping, Academic Press, 1998. 40

[175] S. Baeg and N. Kehtarnavaz, "Classification of breast mass abnormalities using denseness and architectural distorsion," *Electronic Lett. Comput. Vis. Image Anal.*, vol. 1, pp. 1–20, 2002. 40

[176] C.-T. Liu, P.-L. Tai, A.-J. Chen, C.-H. Peng, and J.-S. Wang, "A content-based medical teaching file assistant for ct lung image retrieval," in *Proceedings of the IEEE International Conference on Electronics, Circuits, Systems (ICECS)*, 2000. 40, 42

[177] E. Petrakis, "Content-based retrieval of medical images," *Int. J. Comput. Res.*, vol. 11, pp. 171–182, 2002. 40

[178] G. Bucci, S. Cagnoni, and R. D. Domicinis, "Integrating content-based retrieval in a medical image reference database," *Comput. Med. Imag. Graphics*, vol. 20, pp. 231–241, 1996. 40

[179] U. Sinha and H. Kangarloo, "Principal component analysis for content-based image retrieval," *RadioGraphics*, vol. 22, pp. 1271–1289, 2002. 40, 42

[180] W. Cai, D. Feng, and R. Fulton, "Content-based retrieval of dynamic pet functional images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 4, pp. 152–158, 2000. 41

[181] S.-K. Chang, "Active index for content-based medical image retrieval," *Comput. Med. Imag. Graphics*, vol. 20, pp. 219–229, 1996. 41

[182] C. G. Snoek and A. W. M. Smuelders, "Visual-concept search solved?" *IEEE Computer*, pp. 76–78, June 2010. 41

[183] C. Brodley, A. Kak, C. Shyu, J. Dy, L. Broderick, and A. Aisen, "Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision," in *Proceedings of the 10th National Conference on Artificial Intelligence*, 1999, pp. 760–767. 42

[184] W.-J. Kuo, R.-F. Chang, C. Lee, W. Moon, and D.-R. Chen, "Retrieval technique for the diagnosis of solid breast tumors on sonogram," *Ultrasound Med. Biol.*, vol. 28, pp. 903–909, 2002. 42

[185] K. Veropoulos, C. Campbell, and G. Learnmonth, "Image processing and neural computing used in the diagnosis of tuberculosis," in *Proceedings of the Colloquium on Intelligent Methods in Healthcare and Medical Applications (IMHMA)*, 1998. 42

[186] S. Beretti, A. D. Bimbo, and P. Pala, "Content-based retrieval of 3d cellular structures," in *Proceedings of the Second International Conference on Multimedia and Exposition (ICME)*, 2001, pp. 1096–1099. 42

[187] D. Keysers, J. Dahmen, H. Ney, B. Wein, and T. Lehmann, "A statistical framework for model-based image retrieval in medical applications," *J. Electronic Imaging*, vol. 12, pp. 59–68, 2003. 42

[188] G. Salton, "The evaluation of computer-based information retrieval systems," in *Proceedings of the 1965 Congress International Federation for Documentation (IFD 1965)*, 1965, pp. 125–133. 43

[189] H. Muller, W. Muller, D. M. Squire, and S. Marchand-Maillet, "Performance evaluation in content based image retrieval: overview and proposal," *Pettern Recognition Letters*, vol. 22, pp. 593–601, 2001. 43

[190] J. Bueno, F. Chino, A. Traina, C. Traina, and P.M.Azevedo-Marques, "How to add content-based image retrieval capacity into a pacs," in *Proceedings of the IEEE Symposium on Computer-Based Medical Systems (CBMS)*, 2002, pp. 321–326. 43

[191] T. Brkle, E. Ammenwerth, H.-U. Prokosch, and J. Dudeck, "Evaluation of clinical information systems. what can be evaluated and what cannot," *J. Evaluat. Clin. Practice*, vol. 7, pp. 373–385, 2001. 43

[192] H. D. Tagare, C. Jafe, and J. Duncan, "Medical image databases: A content-based retrieval approach," *Journal of the American Medical Informatics Association*, vol. 4, no. 3, pp. 184–198, 1997. 45, 51

[193] G. Bucci, S. Cagnoni, and R. D. Domenicis, "Integrating content-based retrieval in a medical image reference database," *Computerized Medical Imaging and Graphics*, vol. 20, p. 231241, 1996. 45, 49, 66

[194] D. Unay, A. Ekin, and R. Jasinschi, "Medical image search and retrieval using local binary patterns and klt feature points," in *Proceedings of International Conference on Image Processing (ICIP)*, 2008, pp. 997–1000. 45, 49, 66

[195] B. C. D. M. M. Rahman, T. Wang, "Medical image retrieval and registration: Towards computer assisted diagnosticapproach," in *IDEAS Workshop on Medical Information Systems: The Digital Hospital (IDEAS-DH04)*. IEEE, 2004, pp. 78– 89. 46, 47, 51

[196] A. Quddus and O. Basir, "Wavelet-based medical image registration for retrieval applications," in *International Conference on Bio-Medical Engineering and Informatics*, 2008, pp. 301–305. 46, 48

[197] A. Quddus, O. Basir, and F. Karray, "Wavelet-based image registration and alignment for image retrieval using mutual information and fuzzy logic," in *IASTED International Conference on Signal Processing, Pattern Recognition and Applications*, 2009, pp. 30–33. 46

[198] C. Jongena, J. P. Pluima, P. J. Nederkoornb, and M. A. V. W. J. Niessena, "Construction and evaluation of an average ct brain image for inter-subjectregistration," *Computers in Biology and Medicine*, vol. 34, pp. 647–662, 2004. 47

[199] M. Mellor and M. Brady, "Phase mutual information as a similarity measure for registration," *Medical Image Analysis*, vol. 9, pp. 330–343, 2005. 47, 51, 52

[200] J. Pluim, J. Maintz, and M.A.Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Transaction on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003. 47, 51, 52, 56, 59

[201] ——, "Mutual information matching in multiresolution contexts," *Image and Vision Computing*, vol. 19, pp. 45–52, 2001. 47, 51, 52

[202] J. Lui, B. Vemuri, and J. Marrouquin, "Local frequency representation for robust multimodal image registration," *IEEE Transaction on Medical Imaging*, vol. 21, pp. 462–469, 2002. 47, 51, 52

[203] J. Moigne, W. Campbell, and R. Cromp, "An automated parallel image registration technique based on the correlationof wavelet features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, pp. 1849–1864, 2002. 47, 51, 52, 59

[204] J. Wu and A. Chung, "Multimodal brain image registration based on wavelet transform using sadand mi," in *Proc. MIAR*, 2004, pp. 270–277. 47, 51, 52

[205] R. Gan, J. Wu, A. Chung, S. Yu, and W. Wells, "Multiresolution image registration based on kullback-leibler distance," in *Proc. MICCAI*, 2004, pp. 599–606. 47, 52

[206] J. Chua and P. Tischer, *A Similarity measure based on casual neighbors and mutual information.* IOS Press, Amsterdam, 2003, ch. Design and Applications of Hybrid Intelligent Systems, pp. 842–851. 47, 52, 56

[207] A. Wong and P. Fieguth, "Fast phase-based registration of multimodal image data," *Signal Processing*, vol. 89, pp. 724–737, 2009. 48

[208] R. Hassan, Z. Wang, and M. Salama, "Multi-sensor image registration based-on local phase coherence," in *Proc. ICIP 2009*, 2009. 48

[209] A. Wong, D. A. Clausi, and P. Fieguth, "Cpol: Complex phase order likelihood as a similarity measure for mr-ct registration," *Medical Image Analysis*, vol. 14, pp. 50–57, 2009. 48

[210] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 710–732, 1992. 48, 53, 54, 132, 133

[211] P. Kovesi, "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, vol. 1:3, pp. 1–26, 1999. 48

[212] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003. 50

[213] C. R. Shyu, A. Kak, C. Brodley, and L. Broderick, "Testing for human perceptual categories in a physician-in-the-loop cbirsystem for medical imagery," in *CBAIVL*. IEEE, 1999, pp. 111–132. 51

[214] G. Thoma, L. Long, and S. Antani, "Content-based image retrieval (cbir) of biomedical images," *Report to the NLM/LHC Board of Scientific Counselors*, pp. –, 2002. 51

[215] T. Lehmann, B. Wein, J. Dahmen, J. Bredno, and F. V. andM. Kohnen, "Content-based image retrieval in medical applications a novel multi-stepapproach," *Proc. SPIE*, vol. 3972, pp. 312–320, 2000. 51

[216] J. Z. Wang, "Pathfinder: Multiresolution region based searching of pathology images usingirm," in *Proc. AMIA Symposium*, 2000, pp. 883–887. 51

[217] S. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Transaction on Information Theory*, vol. 38, no. 2, pp. 617–643, 1989. 53

[218] C.-T. Hsu and R. Beuker, "Multiresolution feature-based image registration," in *Proc. VCIP*. SPIE, 2000, pp. 1490–1498. 53

[219] T. Cover and J. Thomas, *Elements of information theory*. John Wiley and Sons, New York, 1991. 56

[220] C. Cocosco, V. Kollokian, R.-S. Kwan, and A. Evans, "Brainweb: Online interface to a 3d mri simulated brain database," in *Proceedings of 3-rd International Conference on Functional Mapping of theHuman Brain, Copenhagen*, http://www.bic.mni.mcgill.ca/brainweb/, 1997. 61, 75, 76, 101

[221] J. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, pp. 112–147, 1998. 64

[222] M. J. D. Powell, "An eeficient method for finding the minimum of a function of several variables without calculating derivatives," *Computer Journal*, vol. 7, pp. 152–162, 1964. 64

[223] D.-J. Kroon and C. H. Slump, "Mri modalitiy transformation in demon registration," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009. (ISBI '09)*, 2009, pp. 963–966. 64

[224] S. Periaswamy and H. Farid, "Elastic registration in the presence of intensity variations," *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 865–874, July 2003. 64

[225] S. Camazine. (2008, October) Brain viewed from the right side showing the 4 major cerebral lobes which is a digitally enhanced version of an illustration from manuel de l'anatomiste, by charles morel and mathias duval, published in 1883. [Online]. Available: http://en.wikipedia.org/wiki/File:BrainLobesLabelled.jpg 83

[226] L. Dade, F.-Q. Gao, N. Kovacevic, P. Roy, C. Rockel, C. Otoole, A. Quddus, A. Feinstein, B. Levine, and S. Black, "Sabre: A time efficient semi-automated regional parcellation method for structural magnetic resonance brain images," in *MICCAI, Montreal.*, 2003. 83

[227] B. E. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. ACM Press, 1992, pp. 144–152. 88

[228] V. Vapnik, *Statistical Learning Theory*. Wiley, New York, 1998. 88

[229] C.-C. Chang and J. L. C. (2001) Libsvm: A library for support vector machines. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm/ 88

[230] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, pp. 415–425, 2002. 89, 90, 91

[231] E. Chang, G. Kingshy, G. Sychay, and W. Gang, "Cbsa: content-based soft annotation for multimodal image retrieval using bayes point machines," *IEEE Transactions on Circuits and Systems in Video Technology*, vol. 13, pp. 26–38, 2003. 89

[232] S. Knerr, L. Personnaz, and G. Dreyfus, *Single layer learning revisited: A stepwise procedure for building and training a neural network*, J. Fogelman, Ed. Springer-Verlag, 1990. 90

[233] B. E. Coggins and P. Zhou, "Polar fourier transform of radially sampled nmr data," *Journal of Magnetic Resonance*, vol. 182, pp. 84–95, 2006. 92, 135

[234] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004. 93

[235] E. Bullitt, D. Zeng, G. Gerig, S. Aylward, S. Joshi, J. K. Smith, W. Lin, and M. G. Ewend, "Vessel tortuosity and brain tumor malignancy: A blinded study," *Academic Radiology*, vol. 12, pp. 1232–1240, 2005. 108, 119

[236] T. F. Wu, C. J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004. 108, 110

[237] J. Waser, "Fourier transform and scattering intensities of tubular objects," *Acta Crystallographica*, vol. 8, pp. 142–150, 1955. 136

[238] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001. 140

[239] V. C. Chen, "Evaluation of bayes, ica, pca and svm methods for classification," in *RTO SET Symposium on Target Identification and Recognition Using RF Systems*, 2004. 141

[240] M. Grcar, B. Fortuna, and D. Mladenic, "knn versus svm in the collaborative filtering framework," in *Data Science and Classification*, ser. Studies in Classification, Data Analysis, and Knowledge Organization, V. Batagelj, H.-H. Bock, A. Ferligoj, and A. iberna, Eds., ACM. Springer Berlin Heidelberg, 2006, pp. 251–260. 141

[241] S. Arora, D. Bhattacharjee, M. Nasipuri, L. Malik, M. Kundu, and D. K. Basu, "Performance comparison of svm and ann for handwritten devnagari character recognition," *International Journal of Computer Science Issues*, vol. 7, no. 6, pp. 18–26, May 2010. 141

Glossary

Inhomogeneity Non-uniformity of MR radiation

Intersubject Within the same subject (patient)

Lobes Major anatomical areas of human brain

MR Magnetic Resonance

MS Multiple Sclerosis

Multimodal Involving several modes of MR sequences

Multiresolution Analyze images at multiple resolution

Multiscale Analyze images at multiple resolution

Multisubject Involving many subjects (patients)

PD MR Proton Density sequence

$T1$ weighted MR T1 sequence

$T2$ weighted MR T2 sequence

Volume 3D Data acquired from MR scanner