MEDICAL IMAGE MODALITY CLASSIFICATION USING FEATURE WEIGHTED CLUSTERING APPROACH

BHAVIK ANIL CHANDRA

UNIVERSITI SAINS MALAYSIA

2010

MEDICAL IMAGE MODALITY CLASSIFICATION USING FEATURE WEIGHTED CLUSTERING APPROACH

by

BHAVIK ANIL CHANDRA

**Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science**

December 2010

DECLARATION

Name : **Bhavik Anil Chandra**

Matric Number : **P-COM 0011/09**

School : **School of Computer Sciences**

Thesis Title :

Medical Image Modality Classification using Feature Weighted Clustering Approach

I hereby declare that this thesis I have submitted to **SCHOOL OF COMPUTER SCIENCES** on **DECEMBER 2010** is my own work. I have stated all references used for the completion of my thesis.

I agree to prepare electronic copies of the said thesis to the external examiner or internal examiner for the determination of amount of words used or to check on plagiarism should a request be made.

I make this declaration with the believe that what is stated in this declaration is true and the thesis as forwarded is free from plagiarism as provided under Rule 6 of the Universities and University Colleges (Amendment) Act 2008, University Science Malaysia Rules (Student Discipline) 1999.

I conscientiously believe and agree that the University can take disciplinary actions against me under Rule 48 of the Act should my thesis be found to be the work or ideas of other persons.

Students Signature: Date: **February 17, 2011**

Acknowledgement of receipt by: Date: **February 17, 2011**

ACKNOWLEDGEMENTS

I would like to first extend my gratitude and acknowledgements to my supervisor, Assoc. Prof. Dr Mandava Rajeswari for her timely support, the helpful advices and her tireless guidance throughout the duration of my dissertation.

Not forgetting my sincere appreciation to both of my parents, for without their sacrifice and support I would not have dreamt of realizing my goals and aspirations.

Acknowledgements also go to my friends from the CVRG lab, Henry Lee Song Yeow, Ong Kok Haur, Lim Khai Yin, Saravanesh Subramaniam, Alfian, Ehsan Abbasnejad and Anusha Acuthan for their useful advice and their willingness to lend a helping hand. I am also obliged to thank my batch mates, Mogana Vadiveloo, Vaithegy Doraisamy and Nurul Syafika Nadiah for their invaluable assistance and abundant support.

I would also like to express my gratitude to the Malaysian Government for sponsoring my studies and also to the School of Computer Sciences, Universiti Sains Malaysia for providing me with all the necessary essence to complete this dissertation.

Last but not least to all those who have been directly or indirectly indispensable to the successful completion of this dissertation, I genuinely thank you from the bottom of my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MRI	Magnetic Resonance Imaging
CT	Computed Tomography
GX	Graphics
MR	Magnetic Resonance
NM	Nuclear Medicine
PET	Positron Emission Tomography
PX	Optical Imaging
US	Ultrasound
XR	X-Ray
2,3,4,5D	2,3,4,5 Dimensional
JPEG2000	Joint Photographic Experts Group 2000
DICOM	Digital Imaging and Communications in Medicine
PACS	Picture Archiving and Communication System
CBIR	Content-Based Image Retrieval System
IMAGEClef	Image retrieval in Cross Language Evaluation Forum (CLEF)
SIFT	Scale-invariant feature transform
SURF	Speeded Up Robust Features
DoG	Difference of Gaussian

MPEG	Motion Picture Expert Group
SCAD	Simultaneous Clustering and Attribute Discrimination
FCM	Fuzzy C-Means
EHD	Edge Histogram Descriptor
HTD	Homogeneous Texture Descriptor
CLD	Color Layout Descriptor
DCD	Dominant Color Descriptor
TBD	Texture Browsing Descriptor
LBP	Local Binary Pattern
DCT	Discrete Cosine Transform
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
***k*-NN**	*k*- Nearest Neighbour
SVM	Support Vector Machine

PENGELASAN MODALITI IMEJ PERUBATAN MENGGUNAKAN KAEDAH PENGELOMPOKAN PEMBERAT CIRI

ABSTRAK

Sistem Dapat Semula Imej Perubatan merupakan satu bidang yang amat penting bagi pembekal penjagaan kesihatan. Dengan kepelbagaian modaliti imej perubatan yang digunakan, adalah penting agar sistem dapat semula imej perubatan mampu menapis hasil dapat semula berdasarkan modaliti. Kaedah terkini yang digunapakai untuk mengenalpasti modaliti adalah tidak tepat dan selalunya modaliti yang salah diberikan kepada imej perubatan. Matlamat kajian ini adalah untuk mencadangkan suatu kaedah untuk mengelaskan modaliti imej perubatan dengan menyiasat ciri-ciri visual imej tersebut dan mengurangkan intervensi manusia. Tesis ini mencadangkan kaedah pengelompokan imej perubatan menggunakan Pemerihal Visual MPEG-7, pengelompokan SCAD dengan diskriminasi ciri serentak dan menggunakan teknik pengelasan "*k*-Nearest Neighbour". Menggunakan kaedah yang dicadangkan, ciri-ciri visual yang relevan kepada kelompok akan diberikan nilai pemberat yang lebih tinggi yang akan menentukan ciri-ciri visual yang mewakili kelompok tersebut. Ciri-ciri visual yang tidak relevan tidak akan diabaikan sepenuhnya, sebaliknya akan diberikan nilai pemberat yang lebih rendah. Menggunakan kaedah ini, imej perubatan akan dikelompok dan dikategorikan kepada modaliti imej perubatan berdasarkan ciri-ciri visualnya. Dalam tesis ini, tumpuan diberikan kepada Pemerihal Histogram Pinggir (Edge Histogram Descriptor), Pemerihal Layout Warna (Color Layout Descriptor) dan Pemerihal Tekstur Homogen (Homogeneous Texture Descriptor) MPEG-7. Kaedah yang dicadangkan akan dibandingkan dengan kaedah pengelompokan

C-Min Fuzzy. Tesis ini juga menyimpulkan bahawa pengelasan modaliti imej perubatan boleh dilaksanakan dengan jayanya dan bergantung kepada kekuatan pemerihal ciri yang digunakan untuk mengesan ciri-ciri imej perubatan tersebut.

MEDICAL IMAGE MODALITY CLASSIFICATION USING FEATURE WEIGHTED CLUSTERING APPROACH

ABSTRACT

Medical Image Retrieval System is an area of great importance to the healthcare providers. With a variety of modalities in use, it is important that retrieval systems can provide filtering of results based on modality. The current method of identifying modalities is not accurate and often wrong modality is assigned to the medical images. The objective of this work was to propose a methodology to classify existing medical images by inspecting their visual features and reducing human intervention. This thesis proposes the clustering of medical images using MPEG 7 Visual Descriptor, SCAD clustering with simultaneous feature discrimination and using *k*-Nearest Neighbour classification. Using this proposed work, the features that are relevant to the cluster are assigned higher weights, giving importance to a number of features that represent the clusters. The features that are not significant to the cluster are not totally ignored instead given lower weights. Using this method, the medical images will be clustered and categorized into their respective modality based on their visual features. In this thesis, emphasis is given to Edge Histogram Descriptor, Color Layout Descriptor and Homogeneous Texture Descriptor of MPEG-7. The proposed method is tested against Fuzzy C-Mean clustering. This thesis also concludes that medical image modality classification is possible depending on the strength of the feature descriptors to detect the characteristics of the medical images.

CHAPTER 1

INTRODUCTION

In this chapter, a general introduction to medical images, feature descriptors, clustering and classification is presented, particularly the relationship between feature descriptor, clustering and classification. The chapter explores further with explanations of the research problems, objectives, scope, contribution of the proposed work and finally the research methodology is presented.

1.1 Introduction

In recent years, medical imaging and its related technology in the areas of radiology has been a great driving force for fast and accurate diagnosis. Introduction of multiple modality in radiology, such as magnetic resonance imaging (MRI), positron emission tomography (PET), and computed tomography (CT) among others, has brought about great contribution in visualisation of the human anatomy. Doctors have been utilizing medical images produced from these modalities as an important tool to properly diagnose and assist them to save lives. Some of the medical images and their modalities are as follows (Zhou et al., 2008):

- 2D (or 2D-in-time) projected images or slices of the body, e.g., Chest X-ray; Echocardiogram;
- 3D volumetric images of the brain or viscera, e.g., brain Magnetic Resonance Imaging (MRI) or abdominal Computed Tomography (CT);
- 4D (3D-in-time or along different modalities), e.g., Cardiac CT/MR/Ultrasound; Fused

PET/CT, with the 4th dimension for anatomy versus function;

- 5D (3D-in-time and along modalities) images of a moving organ, e.g., respiratory-gated PET-CT study of the lung;

The flood of multiple modalities of medical images such as MRI, PET, CT and many more has required new techniques in computer vision to enable a computer to detect these images and its contents accurately. The detection of this image's content is important since all the modality have different texture and shape, which is a important factor that will allows modality classification to be carried out. Unfortunately, medical images lack a few characteristics of traditional images making existing feature detection techniques utilizing colour for example to be less beneficial. Texture and shape provides more information since medical images are usually large homogeneous regions with little color and intensity variation (Sargent et al., 2009). Figure 1.1 shows some of the medical image modality that are very common.

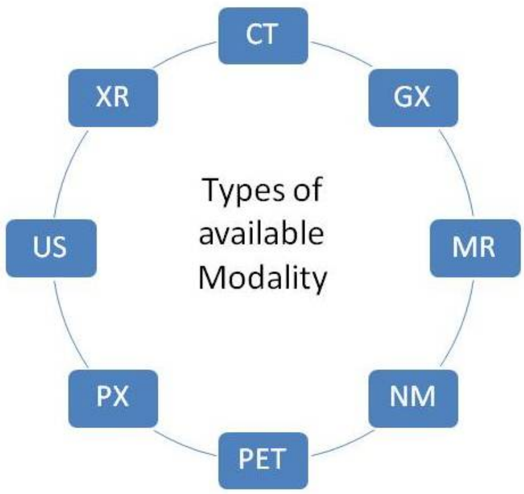

Figure 1.1: Types of modality in medical images

1.2 Background of the problem

Medical images are a set of 2 or 3 dimensional images of a human body. These images assist doctors to diagnose diseases or dysfunctions without expensive and time consuming medical procedures. Medical images are normally grey-scale and are in the format of JPEG2000 or DICOM, and are usually stored using PACS (Picture Archiving and Communication System). PACS provides a way of allowing patients' medical images to be stored securely at a centralized place, allowing a medical expert to access them whenever needed (Wong and Tjandra, 1999).

Many clinicians have indicated that the ability to limit search to certain modalities in medical image retrieval systems has been a very important filter. The current retrieval system ((Society., 2010), (iVirtuoso, 2010)) allows this filtering but it is usually extracted from the caption and often proves to be wrong. Older images that are digitalised might not retain the modality information and this will be a time consuming process if the radiologist are to manually assign the modality to the images. DICOM images provide metadata with modality attributes but not all modality generating machines are able to take advantage of this, especially older machines that are still in active use.

The vast modalities available in the PACS system usually requires the doctor to manually specify what type of modalities used for older images, or images that are sourced from outside of PACS compliant modality generators. This is troublesome if the medical images come in very large quantities. There are few available options such as automated labelling and annotation based on the title of the image but this is not available for all images from different modalities. Another approach is by attaching textual annotation but again, this consumes time and requires active human intervention. The modalities classification by targeting the related visual features of the image promises a better solution to this problems.

1.3 Problem Statement

As PACS is becoming very common to medical practitioners, it is equally important to change the current records of the existing patients to digital format so that they can be referred immediately when the need arises. Converting from old patient information requires the existing medical images and films to be converted as well. When this conversion takes place, there is a very high probability of information loss. One example of such information loss is the modality of the source image, since during the conversion the modality information might not been saved. Furthermore, it might not be very practical to check every single image and assigning the modality manually. When this happens, there is no way for the image's original modality to be identified without visual inspection by a qualified radiologist. Manual inspection would consume very long hours and prone to human errors.

The current and existing approach tries to combine visual features with textual annotation to classify the modality. Although this approach guarantees better accuracy but it requires textual annotation to be prepared manually beforehand. Visual features of the images are sufficient to provide good insight of the modality of the image.

The research tries to highlight possible solution of classifying the related medical images into different modalities category based on the visual features of the medical images. The work proposed will be able to answer the following question:

How do we separate the different modalities of the images into its category effectively?

1.4 Research Objectives

The objectives of the proposed work are as stated:

- To determine the modality of medical images through inspection of visual features of the images;
- To propose a method to classify medical image modality by clustering using feature weighted visual features collected using MPEG-7 Visual Descriptors;
- To compare the performance of the SCAD and FCM clustering algorithm in modality classification.
- To evaluate accuracy of modalities classification for test dataset using table of confusion.

1.5 Scope of the Research

The scope of the proposed work is on medical images, either in JPEG2000 or DICOM format and includes all major modalities categories from radiography, magnetic resonance, ultrasound or optical sources such as CT, GX, MR, NM, PET, PX, US and XR.

1.6 Justification of the Research

The main motivation behind the proposed work is mainly to address the need of clinicians. User studies have shown that modality filtering is one of the most important filters that clinicians would like to have on medical image retrieval systems. This is important as there is currently no effective way to sort the medical images into the related modalities without any human intervention. In this research, the proposed work tries to put forth a method to classify the medical images into its related modality by looking at the visual features of the images, ridding the need of human intervention during this process.

Many clustering algorithms and techniques exist for various purposes. The clustering algorithm that is proposed to be used will be able to cluster the training images into its related cluster based on modality and will be utilizing feature weighting technique. All features that are related to the cluster will be given more weightage and the features that are not very relevant to the cluster will be given lower weightage. This is a simple feature discrimination/attribute selection technique integrated into the clustering mechanism.

1.7 Research Methodology

1.7.1 Research Procedures

This sub-section is about the steps and procedures of the research to be carried out. The entire flow of the research is presented. The starting point of the proposed research is by extracting the features from the medical images, so that we can use the feature vectors as data to be further used in the related implementation and testing phase. The second phase is to adapt the existing algorithm to make use of the feature vectors as an input to the clustering and classification of modalities. Finally, the final phase would be to produce an evaluation algorithm to compare the result of the classification and with the actual modality of the image. From the evaluation of the result, we will be able to know the accuracy of the proposed work. This steps is presented in figure 1.3.

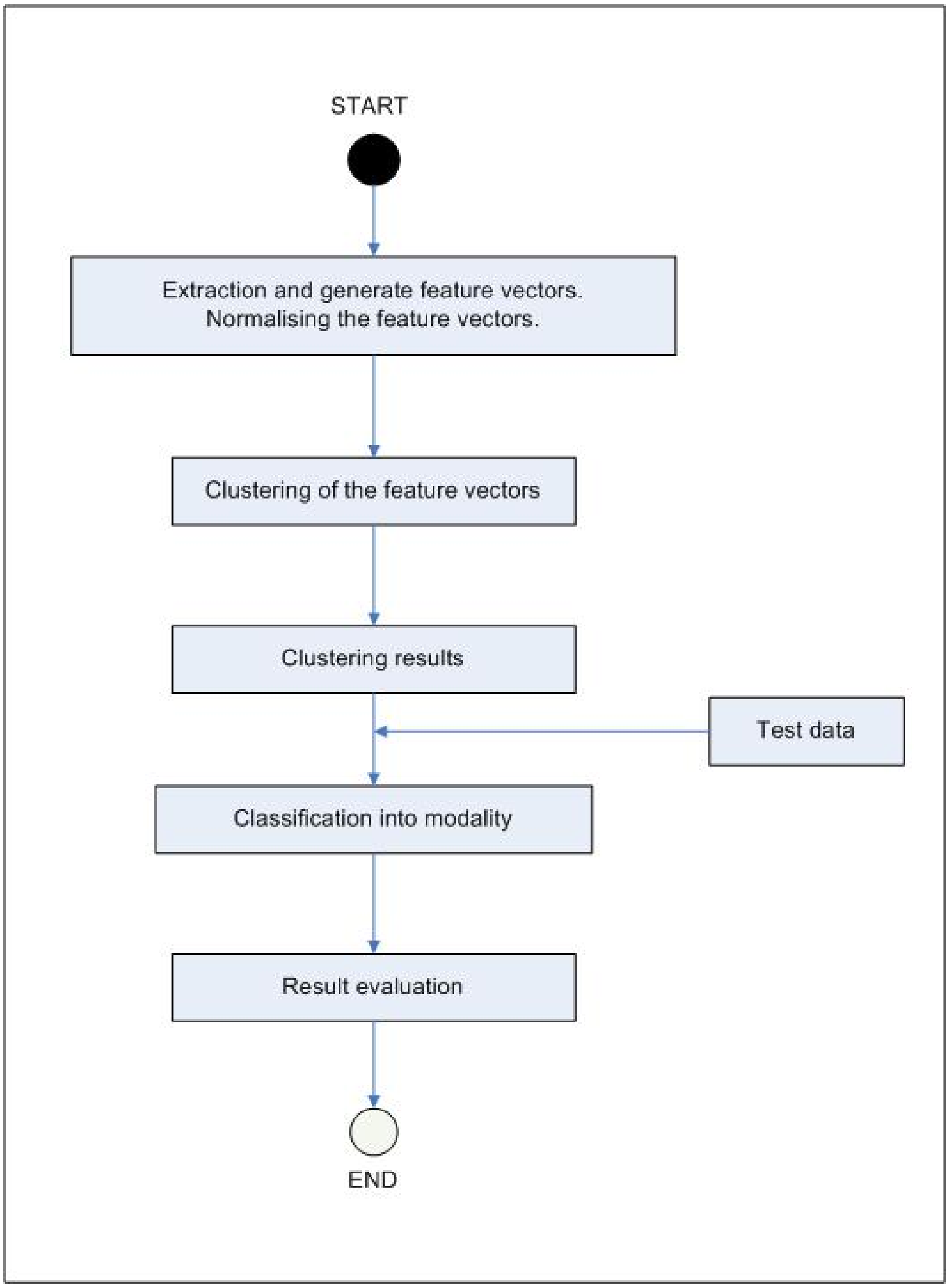

Figure 1.2: Flowchart of Research Procedures

1.7.2 Proposed Framework

The research proposes a method to cluster the visual features and further classify the medical images into its categories using SCAD (Simultaneous Clustering and Attribute Discrimination) clustering and *k*-Nearest Neighbour (*k*-NN) classification technique.

The clustering method is modified to accept a new type of input, that is, the feature vector generated using MPEG-7 visual feature descriptors. The feature vectors are then clustered. It is hypothesized that the clusters represent the image modality. The results of the clustering will be evaluated and will be compared with the manually labeled modality. This will allow us to compare the accuracy of the clustering. Evaluation and analysis of the results will be presented in Chapter A, where the hypothesis will be validated. Figure 1.3 provides the flow of the proposed work from input to the output. Our input to the proposed work is medical images with unknown modality and the output is assignment of a modality to the medical image. This is achieved with the proposed work, with steps listed in figure 1.2.

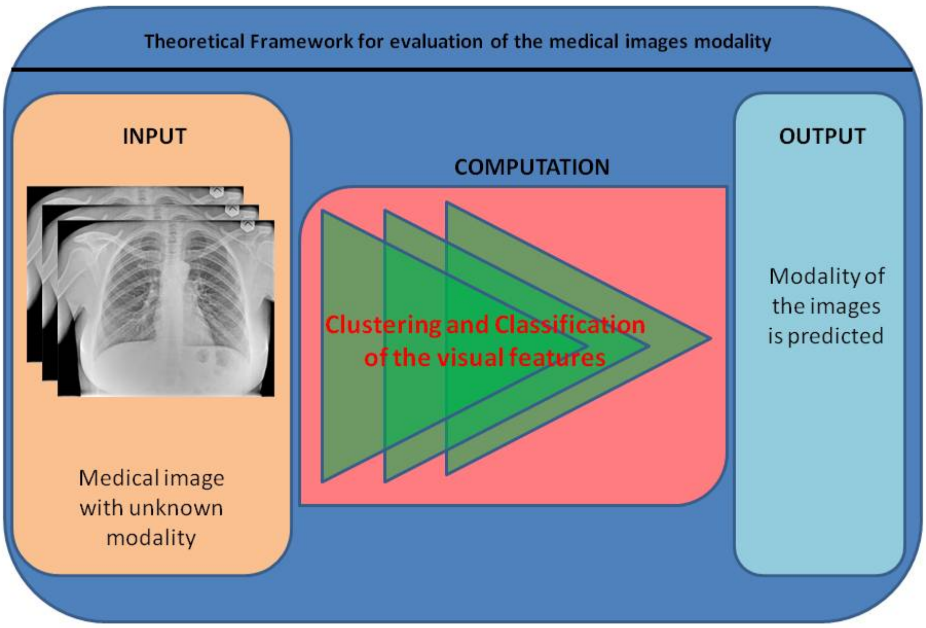

Figure 1.3: Theoretical framework for medical image modality classification

1.7.3 Research Design

This research will be conducted based on empirical evaluation technique whereby the proposed work will be evaluated in term of the accuracy of the modality predicted. Hypothesis set for this research are "Using visual features of medical images, classification of medical image modalities can be done accurately" and "SCAD clustering performs better than Fuzzy C-Means (FCM) clustering in modality classification". The aforementioned hypothesis will be tested and proven by means of experimentation, reinforcing the purpose of study stated earlier. The study setting for this proposed work is set to lab experiment. Time horizon is set to longitudinal study as this research will use multiple set of medical images to be tested, as a comparison of performance and reinforcement of the results.

1.8 Research Phases

Four steps or phase have been identified in this research, starting with extracting and generating the feature vectors, clustering of the features followed by classification to classify the clustered features into their modalities, and finally result evaluation to compare the accuracy of the proposed work.

1.8.1 Extracting, Generating and Normalizing the features vectors

In this proposed work, the input data is in the form of visual features of the medical images. The medical images used are extracted from IMAGEClef (Image retrieval in Cross Language Evaluation Forum (CLEF)) dataset of 8 different modalities. The dataset contains medical image in the format of JPEG.

The dataset is divided into 2 groups, one training group and another testing group. There are about 2380 medical images in training group divided into 8 modalities. The modality is categorized as follows:

- CT : Computerized tomography
- GX : Graphics, typically drawing and graphs
- MR : Magnetic resonance imaging
- NM : Nuclear Medicine
- PET : Positron emission tomography including PET/CT
- PX : Optical imaging including photographs, micrographs and gross pathology
- US : Ultrasound including (color) Doppler
- XR : X-ray including x-ray angiography

The training group consist of around 2600 medical images from all modality categories. Ground truth for the testing group is provided. More details on this dataset is discussed in Chapter 3.

Visual features of all these images are generated using MPEG-7 Visual features. The MPEG-7 Visual Descriptor is a series of descriptors that are designed as a part of a multimedia retrieval system. The visual component of MPEG-7 consists of descriptors in the following categories:

- Color
- Texture
- Shape
- Motion
- Localization
- Face recognition

For this research, focus is set on using Edge Histogram Descriptors (EHD) of MPEG-7 Visual. In additional experiments, Heterogeneous Texture Descriptor (HTD) and also Color Layout Descriptor (CLD) will also be used to compare their performances. The feature values are normalized using statistical normalization. The formula of statistical normalization used is

$$Z = (X - u) \tag{1.1}$$

where X represents the feature value, u represents mean value of the data. After the normalisation, the standard deviation of the data will be 1.

1.8.2 Clustering of the features

At this level, medical images will be clustered using SCAD clustering into their related clusters using training data. Clustering is required because the different types of medical images will be separated into the their group or cluster with this step.

In total, a group of 8 clusters will be created, referring to each different modality available. Each centroid of the clusters is saved for the next step of the process. For the purpose of comparison, the result and discussion of the experimental result will also include 4 and 6 cluster's performance.

1.8.3 Classification of modality

After identifying the different clusters, the next step is to proceed to further classify the test features or the unlabeled data so that it will be organized into different modalities based on corresponding clusters. The purpose of the classification is to find the suitable modality that matches the characteristics of the test data. Therefore, the proposed work will be utilizing *k*-

Nearest Neighbour (k-NN) classification technique to find the nearest cluster to the test features, assigns the modality of the cluster to the test feature. This will be checked against the manually assigned modality.

1.8.4 Result Evaluation

Finally, once the results has been obtained, we can then compare the accuracy of the classification by comparing the results of all the dataset classification done manually (ground truth) with the one done by the algorithm (proposed work). This way both results can be compared and the accuracy of the classification will be represented in table of confusion for further analysis and comparison with other algorithms.

1.9 Research Contribution

The main contributions from this research would be:

- Proposing a methodology to cluster medical images into their related modality category based on visual features of the images. Visual features are extracted using relevant MPEG-7 Visual descriptors.
- The proposal of clustering visual features using SCAD (a simultaneous clustering and feature weighting technique) whereby the irrelevant features are not totally neglected, instead the features are carefully selected for each cluster.

1.10 Organization of Thesis

This chapter describes the terms used in this thesis and also some background on the proposed work. It also defines the research problem, objectives and also justifications of the proposed work. The scope of the work has also been aforementioned as the research has a main concern on medical images and justifications of the research problems. The research methodology and contributions of the proposed work were also described in this chapter.

The rough outline of the remainder chapter of the thesis is organized as follow:

Chapter 2 will discuss the background of the work and also literature review of related work for the chosen methods and related topics of discussion. This chapter introduces the basics on feature descriptors,clustering and classification method considered and also some background studies of the chosen clustering and classification method.

Chapter 3 proposes Simultaneous Clustering and Attribute Discrimination (SCAD) clustering technique used for clustering the features and k-Nearest Neighbour classification to match the modalities to the features. This chapter also focuses to explain the mechanics of the algorithm and the research work. Evaluation of the results and experimental setup used is also explained.

Chapter 4 discusses the results of the proposed work and how it justifies the objectives of the research.

Chapter 5 concludes the findings of the proposed work and also discusses the future work that can be carried out to enhance the result of the proposed work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In this chapter, a more detailed exploration of the research area and its related components is presented. The chapter starts with introducing Content-based image retrieval (CBIR) systems, with specifics towards medical image retrieval. Next, Picture Archiving and Communication System (PACS) and medical images and its related modality is introduced and discussed. The components of the proposed work such as feature descriptors is then explained and discussed. Some current approaches in modality classification is then presented and followed by summary of the chapter. In this chapter, the focus is to justify the research methodology and the justification for the methods and technique used in order to achieved the objectives of the research.

2.2 Content-Based Image Retrieval systems

Content-based image retrieval (CBIR) systems are built with aims of finding similar images based on the query images or some relevant keywords. CBIR systems are developed to instantly search medical images based on the visual features (Arafa, 2010). An example of commonly used CBIR engine in medical domain is GNU Image Finding Tool (GIFT), made available at (FSF, 2010).

Most common CBIR systems receives input combination of visual and textual annotations in order to make the searching and retrieval to be more meaningful. CBIR performs very poorly when compared to text searching when it is used on large databases that have wide spectrum of modalities (Stougiannis et al., 2010). Visual features such as color is commonly used in other CBIR domains such as natural images or objects to retrieve similar images but it is not very commonly used in medical images since medical images are mostly monochromatic or gray-scale. This makes existing searching based on color descriptors to be less effective. Texture and shape based descriptors are more commonly used as input to the searching of images in CBIR systems in medical images domain and a trend of combining one or more descriptors is in the rise.

Introduction of PACS in medical healthcare has generated new ways of storing and retrieving medical images in an organized and reliable manner; patients now wait for shorter time for their diagnosis and healthcare delivery efficiency has vastly improved (Stoian et al., 2008). PACS have also made the usage of the medical images easily accessible to doctors; allowing them to do more with the newly acquired medical image, including 3D-visualizations of tumor structures or organs and using computer aided detection to allow detection of anomalies by the radiographers, to name a few. CBIR in medical images domain is very frequently combined with PACS in order to allow medical experts to search for related images, from different patients who might have been diagnosed or treated before. Combining PACS and CBIR allows experts to share and learn from previous cases or history of a patient.

Therefore, there is very instant need for accurate detection of similar images and also from different modality. Modality detection and searching is not currently available in many CBIR systems because this filtering is unfortunately until today is not accurate and consistent. Current systems that utilize this filtering most of the time gets the modality information from the image's file name, most probably assigned manually by the technicians in-charge of the digitalizing the images from analog source of the patient's record. Thus, no verification on the modality assigned to the medical images exist and therefore, this filtering is proven to be not accurate most of the time.

2.2.1 Medical Images and Modality Types

Below is some description of medical image modality types:

- COMPUTED TOMOGRAPHY (CT)

  Computed Tomography (CT), also commonly referred to as a CAT scan, is a medical imaging method that combines multiple X-ray images taken from different angles to produce detailed cross-sectional pictures of areas inside the body. The resulting images provide more information than regular X-rays, and allow doctors to look at individual slices within the 3D images.

  CT is often used to evaluate:

  - Organs in the pelvis, chest and abdomen.
  - Colon health (CT colongraphy).
  - Presence of tumors.
  - Pulmonary embolism (CT angiography).
  - Abdominal aortic aneurysms (CT angiography).
  - Spinal injuries.

- GRAPHICS (GX)

  Graphics (GX) are typically drawing and graphs generated by measuring equipments, reports and findings.

- MAGNETIC RESONANCE IMAGING (MR)

  Magnetic Resonance Imaging (MR) is a medical imaging technology that uses radio waves and a magnetic field to create detailed images of organs and tissues. MR has proven to be highly effective in diagnosing a number of conditions by showing the difference between normal and diseased soft tissues of the body.

MR is often used to evaluate:

- Blood vessels.
- Breasts.
- Organs in the pelvis, chest and abdomen (heart, liver, kidney and spleen).

- NUCLEAR MEDICINE (NM)

  Nuclear medicine (NM) is used mainly to allow visualization of organs and regions within organs that cannot be seen on conventional x-ray images. Space occupying lesions (injury or abnormality), especially tumors, may stand out on nuclear medicine images. Generally, these lesions are seen as areas of reduced radioactivity (called a "coldspot"); however, in some instances, like bone scanning, areas of increased activity (called a "hotspot") represent disease or injury (pathology).

  NM is often used to evaluate:

  - Bone scanning.
  - Heart Disease.

- POSITRON EMISSION TOMOGRAPHY (PET)

  Positron Emission Tomography (PET) is a type of nuclear medicine that provides physicians with information about how tissues and organs are functioning. PET, often used in combination with CT imaging, uses a scanner and a small amount of radio pharmaceuticals which is injected into a patient's vein to assist in making detailed, computerized pictures of areas inside the body.

  PET is often used to evaluate:

  - Neurological diseases such as Alzheimer's and Multiple Sclerosis.
  - Cancer.
  - Heart Disease.

- OPTICAL IMAGING (PX)

  Optical Imaging (PX) are photographs, micrographs and gross pathology that are taken to observe the conditions of the patient.

- ULTRASOUND (US)

  Diagnostic ultrasound, also known as medical sonography or ultrasonography, uses high frequency sound waves to create images of the inside of the body. The ultrasound machine sends sound waves into the body and is able to convert the returning sound echoes into a picture. Ultrasound technology can also produce audible sounds of blood flow, allowing medical professionals to use both sounds and visuals to assess a patient's health.

  Ultrasound is often used to evaluate:

  - Pregnancy.
  - Abnormalities in the heart and blood vessels.
  - Organs in the pelvis and abdomen.
  - Symptoms of pain, swelling and infection.

- X-RAY (XR)

  X-ray technology is the oldest and most commonly used form of medical imaging. X-rays use ionizing radiation to produce images of a person's internal structure by sending X-ray beams through the body, which are absorbed in different amounts depending on the density of the material.

  X-ray images are typically used to evaluate:

  - Broken bones.
  - Cavities.
  - Swallowed objects.
  - Lungs.

- Blood vessels.
- Breast (mammography).

The information gathered in this section is obtained at (Imaginis, 2010) and (MITA, 2010).

2.3 Feature Extraction and Descriptors

Feature descriptors are defined as a set of low level descriptors that describe different visual features such as colors, textures, shapes and motions as mentioned in (Wang et al., 2003). By using these descriptors, a computer can generate an idea on how a shape looks like in the actual medical image. Usually these feature descriptors will generate a feature vector that describes the feature, usually in binary or numerical form. The feature vector generated will differ in terms of dimensional length, as some feature descriptors are very detail in capturing the features, making the feature vectors to be very long. Some feature descriptors compress their feature vectors to produce shorter feature vectors in order to save time during similarity measurement and increases searching and retrieval performance.

Some authors do not agree that features especially visual features such as color, texture and shape would be enough when taken as feature descriptors. They recommend combining these features with semantic features, since this nature of image understanding is closer to that of human being (Shao et al., 2005). A few recent literatures heavily depends on combined descriptors for modality classification (Zhou et al., 2010) but this also increases time in generating the required features before the searching and retrieval task can be carried out.

Three main sources to get semantic features are also suggested further in the literature, mainly knowledge-based, human interaction and exterior information-based. A big gap, therefore, lies between low-level feature layout with high-level semantic concepts (Hiremath and

Pujari, 2007). Example of knowledge-based semantic features is prior knowledge of the color and texture distribution, human-interaction meaning human intervention is required to pass more information as a human will interpret it and exterior information-based semantic meaning another information gathered from a different source is needed to establish the high-level semantic concept. Generating these semantic concepts requires more processing steps, human intervention and is counter-productive as more time and resources needed to generate these semantic concepts.

In another literature, (Sargent et al., 2009) believe visual descriptors often fail to generate an acceptable number of features when applied on medical images, since medical images contain large homogeneous regions with little color, if any, and intensity variations. Therefore the algorithm to be chosen for our research purpose must be more robust in generating the feature vectors since medical images from few different modality have considerable intra-class heterogeneity and visual similarity. According to (Hiremath and Pujari, 2007), few algorithms make use of global color and texture features while some look closely on only local color and texture features as its main features. Global color and texture generates features by looking at the image as a whole, while local descriptors perform similar task by partitioning the images into smaller regions. Shape descriptors must be more robust to rotation invariant and small variations in shape. It is also also mentioned two most prominent, state-of-the-art discrete feature detection algorithms, the SIFT (Scale-Invariant Feature Transform) and the SURF (Speeded-up Robust Features). Both these algorithms have the ability to extract scale, rotation and contrast invariant feature descriptors. It is mentioned that SIFT is the most popular feature detector/descriptor currently in use in various computer vision applications. SURF meanwhile is based on SIFT, but makes use of many approximation techniques in order to increase its efficiency. MPEG-7 meanwhile promises a standard to allow many multimedia retrieval systems to be able to search much more larger and bigger number of databases.

The following section will review the different feature descriptors, compare their advantages and disadvantages and study their usage suitability in our research.

2.3.1 SIFT algorithm

SIFT algorithm, published by (Lowe, 2004), is an algorithm that transforms image data into scale-invariant coordinates relative to local features. Four important concepts in SIFT is scale-space extrema detection, keypoint localization, orientation assignment and keypoint descriptor. In scale-space extrema detection, points that have distinct surrounding patches are searched. An approximation to the scale-normalized Laplacian of Gaussian is presented as equation 2.1.

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \tag{2.1}$$

A Difference of Gaussian (DoG) approach is then combined with interpolation over the scale-space mentioned earlier to produce the locations of stable keypoints in the scale-space representation of the image. Next step after the localization is to assign an orientation to the key-points, which provides rotation invariance.

The keypoint descriptors are calculated from the local gradient orientation and magnitudes in a neighborhood of the identified key-point. Finally, all the gradient orientations and magnitudes are combined and represented as a histogram. The descriptor outputs a normalized vector of 128 elements.

The main disadvantages of using SIFT is the long computational time needed to compute the keypoints. Although the number of keypoints detected is a lot more than those of SURF, it does not produce better result than that of SURF. In many cases, the SURF algorithm, that is also modified from SIFT, is far more robust and produces higher matching percentage than SIFT as mentioned in (Bauer, 2007). In term of scale invariance and blur images, SIFT performs better than SURF but doesn't perform better when there is illumination change (Juan and Gwun, 2009).

2.3.2 SURF algorithm

The SURF algorithm was designed to be a local invariant interest point detector and also a descriptor. For example, it provides benefits for finding correspondence between two images of similar scene or object. SURF is highly recognized in the areas of 3D reconstruction, image retrieval and object recognition. The disadvantage of using SIFT compared to SURF is the slowness of the SIFT algorithm itself. SURF produces 128 dimensional feature vectors with much faster processing time compared to SIFT.

In order to overcome the disadvantage of SIFT, SURF uses integral images for major speed up. Integral image also known as summed area tables, is actually an intermediate representation for the image and contains the sum of grey-scale pixel values of an image.

The detector used in SURF is based in Hessian matrix because it has better performance in computation time and accuracy (Bay et al., 2008). SURF describes images faster than SIFT by 3 times but does not perform as well as SIFT on invariance to illumination change and viewpoint change. The number of keypoints generated with SURF is much more lesser compared to SIFT. SURF is less superior than SIFT in detecting scale, rotation and blurring (Juan and Gwun, 2009).

2.3.3 MPEG-7 Visual Feature Descriptors

MPEG-7 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group), was formally named "Multimedia Content Description Interface". As the name might suggest, this is a standard for describing the multimedia content data that incorporates certain interpretation of the information, which can be passed to or accessed by a modern computer code or a device (Yanhong et al., 2009). Visual section in the MPEG-7 system is used to describe an image specifically. Visual low-level descriptors included in the visual part of MPEG-7 are color, texture, shape and motion descriptors which describe different features of a related visual content (Wang et al., 2003).

Based on (Spyrou et al., 2005), three MPEG-7 Visual descriptor that are relevant to our proposed work is Color Layout Descriptor (CLD), Edge-Histogram Descriptor (EHD) and Homogeneous Texture Descriptor (HTD). The reason Color Structure Descriptor (CSD) is not used, but replaced by HTD is because color histograms based descriptors such as CSD and Scalar Color Descriptor (SCD) is known to perform very badly on monochrome images.

In (Eidenberger and Horst, 2003), the authors tried to look at the retrieval performance using statistical point of view. Good descriptors will generate description with high variance, a well balanced cluster structure and high discrimination to be able to distinguish different media content. Statistical evidence would suggest the quality of the used description extraction algorithms. During the MPEG-7 design process, optimizing the retrieval was the major goal, but not how good the quality of recalled result is. The best descriptors are combination of CLD, Dominant Color Descriptor (DCD), EHD and Texture Browsing Descriptor (TBD). The other descriptors are highly dependent on these descriptors for their results.

TBD is better than HTD but TBD implementation of MPEG-7 is known to be unstable as it has a known bug in experimental software release, MPEG-7 XM (Tsapatsoulis and Theodosiou, 2009). Homogeneous texture is independent to media and it will be able to distinguish between strong feature and weak feature images but will not able to distinguish within the cluster of strong features (Xu and Zhang, 2006).

Shape descriptors such as EHD may perform well since edge histogram and region based shape descriptors are quite good with monochrome colors. Combination of several related descriptors will produce better retrieval accuracy than one, especially when working with domain color, texture histogram and shape descriptors (Wei et al., 2008) but it makes the feature vector to be longer, consuming more time to be generated. The memory usage to index, search and retrieve very long feature vectors will be larger and makes matching and retrieving process to be slower than compact feature vectors.

Color Layout Descriptor (CLD) is a compact and resolution-invariant MPEG-7 visual descriptor defined in the YCbCr color space and designed to capture the spatial distribution of color in an image or an arbitrary-shaped region. The feature extraction process consists of four stages. The input image is partitioned into $8x8 = 64$ blocks, each represented by its average color. A Discrete cosine transform (DCT) transformation is applied on the resulting $8x8$ partitioned image. The resulting coefficients are zig-zag-scanned and only 6 coefficients for luminance and 3 for each chrominance are kept, leading to a 12-dimensional vector. Finally, the remaining coefficients are nonlinearly quantized (Spyrou et al., 2005).

Edge Histogram Descriptor (EHD) captures the spatial distribution of edges. Four directions of edges (0°,45°,90°,135°) are detected in addition to non-directional ones. The input image is divided in 16 non-overlapping blocks and a block-based extraction scheme is applied to extract the five types of edges and calculate their relative populations, resulting in a 80-dimensional vector (Spyrou et al., 2005). Edge Histogram Descriptor is selected, because this descriptor represents the spatial distribution of edges, where this is an important feature to