University of Huddersfield Repository

Aboaisha, Hosain

The Optimisation of Elementary and Integrative Content-Based Image Retrieval Techniques

Original Citation

Aboaisha, Hosain (2015) The Optimisation of Elementary and Integrative Content-Based Image Retrieval Techniques. Doctoral thesis, University of Huddersfield.

This version is available at http://eprints.hud.ac.uk/id/eprint/26164/

THE OPTIMISATION OF ELEMENTARY AND INTEGRATIVE CONTENT-BASED IMAGE RETRIEVAL TECHNIQUES

HOSAIN ABOAISHA

A thesis submitted to the University of Huddersfield

in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

School of Computing and Engineering

University of Huddersfield

March 2015

Copyright Statement

I. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the "copyright") and he has given the University of Huddersfield the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.

II. Copies of this thesis, either in full or in extracts, may be made only in accordance with regulations of the University Library. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.

III. The ownership of patents, designs, trademarks and any and all other intellectual property rights except for the Copyright works, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.

Acknowledgements

Dedication…

I started this work before the revolt of the Libyan people against their tyrant.

Now the revolution is over, I would like to dedicate this work to the souls of the

brave martyrs who have sacrificed their lives for their beloved country so that

the word of Allah will always be up above.

Abstract

Image retrieval plays a major role in many image processing applications. However, a number of factors (e.g. rotation, non-uniform illumination, noise and lack of spatial information) can disrupt the outputs of image retrieval systems such that they cannot produce the desired results. In recent years, many researchers have introduced different approaches to overcome this problem. Colour-based CBIR (content-based image retrieval) and shape-based CBIR were the most commonly used techniques for obtaining image signatures. Although the colour histogram and shape descriptor have produced satisfactory results for certain applications, they still suffer many theoretical and practical problems. A prominent one among them is the well-known "*curse of dimensionality* ".

In this research, a new Fuzzy Fusion-based Colour and Shape Signature (FFCSS) approach for integrating colour-only and shape-only features has been investigated to produce an effective image feature vector for database retrieval. The proposed technique is based on an optimised fuzzy colour scheme and robust shape descriptors.

Experimental tests were carried out to check the behaviour of the FFCSS-based system, including sensitivity and robustness of the proposed signature of the sampled images, especially under varied conditions of, rotation, scaling, noise and light intensity. To further improve retrieval efficiency of the devised signature model, the target image repositories were clustered into several groups using the k-means clustering algorithm at system runtime, where the search begins at the centres of each cluster. The FFCSS-based approach has proven superior to other benchmarked classic CBIR methods, hence this research makes a substantial contribution towards corresponding theoretical and practical fronts.

List of Publications

- Aboaisha, Hosain, Xu, Zhijie and El-Feghi, Idris (2012); An investigation on efficient feature extraction approaches for Arabic letter recognition. In: Proc. Queen's Diamond Jubilee Computing and Engineering Annual Researchers' Conference 2012: CEARC'12. University of Huddersfield, Huddersfield, pp. 80-85. ISBN 978-1-86218-106-9.

- Aboaisha, H., El-Feghi, I., Tahar, A., and Zhijie Xu (March 2011); Efficient features extraction for fingerprint classification with multilayer perceptron neural network, 8th Int. Multi-Conference on Systems, Signals and Devices, 2011, pp. 22-25.

- Aboaisha, Hosain, Xu, Zhijie and El-Feghi, Idris (2010); Fuzzy Fusion of Colour and Shape Features for Efficient Image Retrieval. In: Future Technologies in Computing and Engineering: Proc. Computing and Engineering Annual Researchers' Conference 2010: CEARC'10. University of Huddersfield, Huddersfield, pp. 31-36. ISBN 9781862180932.

- El-Feghi, I.; Aboasha, H.; Sid-Ahmed, M.A.; Ahmadi, M. (Oct. 2010) "Content-Based Image Retrieval based on efficient fuzzy colour signature, IEEE Int. Con. on Systems, Man and Cybernetics, pp.1118-1124.

List of Abbreviations and Notations

AF   Average Feature

AR   Aspect Ratio

CFSD   Colour Frequency Sequence Difference

CBIR   Content-Based Image Retrieval

CCH   Conventional Colour Histogram

CSS   Curvature Scale Space

DFT   Discrete Fourier Transform

DHMM   Discrete Hidden Markov Model

DIP   Digital Image Processing

FCH   Fuzzy Colour Histogram

FFCSS   Fuzzy Fusion of Colour and Shape Signature

FDs   Fourier Descriptors

LM   Legendre Moments

OCR   Optical Character Recognition

OGs   Orthogonal Moments

PCA   Principal Component Analysis

PZMs   Pseudo-Zernike Moments

SAD   Sum-of-Absolute Difference method

SPCA   Shift-Invariant Principal Component Analysis

SGDs   Simple Global Descriptors

ZMs   Zernike Moments

SVM   Support Vector Machine

TM   Template Modification

List of Figures

List of Tables

Table of Contents

Chapter 1. Research Background

The continually increasing demands for multimedia storage and retrieval have promoted research into and development of various rapid image retrieval systems. Many applications such as anti-terrorism, policing, medical image databases, security data management systems are faced with having to acquire, store and access an ever growing number of captured digital images and video recordings. Research is needed to produce ever faster and more efficient processes and procedures.

The term information retrieval was first devised by Calvin Moores in 1951 based on (Gupta and Jain 1997). Generally, information retrieval is the description of a particular process by which a prospective user of information can process a request for information into the useful collection of query "hints and clues" for data.

Generally, there are two kinds of image retrieval systems: First, text-based systems which were introduced in the 1970s. These systems use keywords to describe each image in a database of collected images, which often suffer from limitations such as: the subjectivity of the user, and the need for manual annotation. They also require significant amount of human labour to maintain the systems and the work is often tedious and painstakingly slow. This text-based approach is usually valid only for a single language (Yong, Huang et al. 1998). The second are the so-called content-based retrieval systems which are multimedia-based search engines used to retrieve desired images, audios, and even videos from large databases containing collections of higher dimensional

data of varied formats. In this research the "content" is limited to images and their related characteristics hence the name "content-based image retrieval" (CBIR). The CBIR systems extract visual features based on such considerations as a study to image texture, colour, and shape patterns (El-Feghi, Aboasha et al. 2007).

Even though CBIR was first introduced in the 1980s it is still an active field in computer vision research and over the past two decades has been the one of the most active research areas in digital imaging (Yasmin and Mohsin 2012). CBIR is a technique which relies on the visual content features extracted from a query image such as texture, shape and colour to retrieve target images in terms of feature similarities from the image databases. The potential of CBIR was recognised after a number of successful applications such as facial recognition (Belhumeur, Hespanha et al. 1997; Gutta and Wechsler 1998) being published, and research into CBIR soon became widespread.

A group of researchers claimed that the concept of Query By Image Content (QBIC) proposed in the 1990s was the real start of modern CBIR systems (Flickner, Sawhney et al. 1995). One of the early QBIC systems was devised by researchers at IBM to interrogate large image databases, and the underlying algorithms used enabled the system to locate images within the database which have similarities with the sample images in the form of sketches, drawings, and colour palette. Virage is another outstanding commercial system for image retrieval (Bach, Fuller et al. 1996) and is capable of applying visual content features as primitives for face and character recognition.

The key in any effective image retrieval system is the feature representation scheme. Significant work has been done to identify visual features and their extraction methods (Cheng, Chen et al. 1998) (Laaksonen, Oja et al. 2000); (Jing, Mingjing et al. 2005). Most current CBIR systems engage three key processing stages as shown in Figure 1-1.
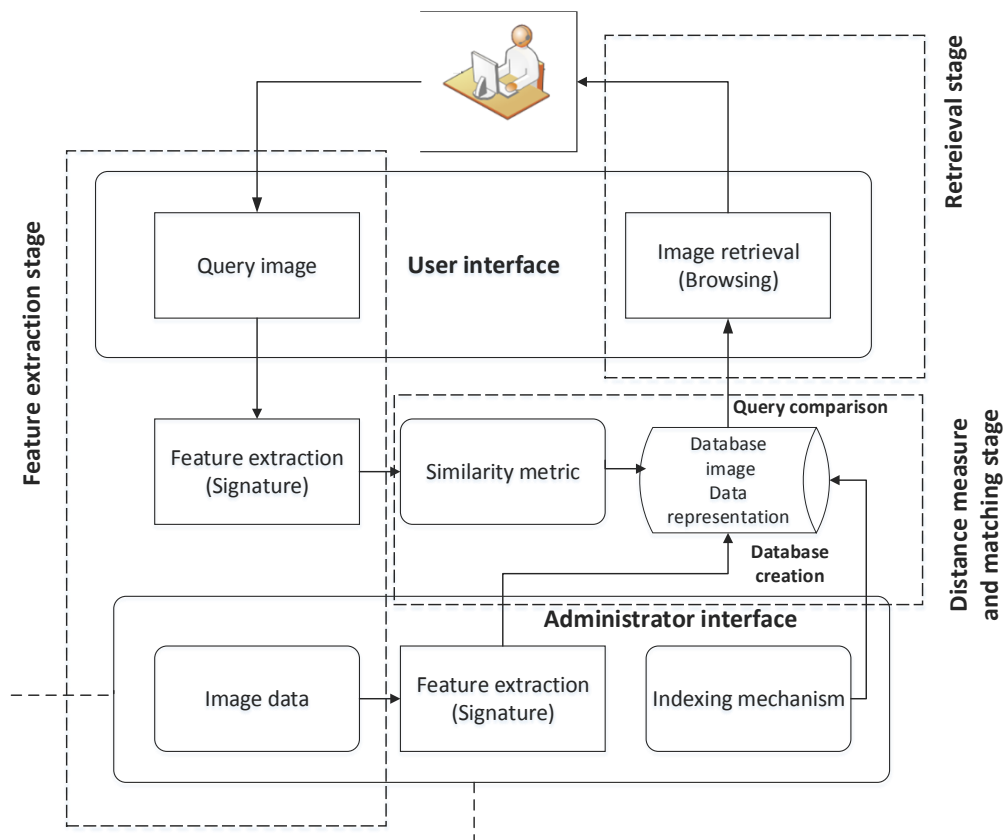
Figure 1-1 General Composition of CBIR Systems

The most challenging problem facing CBIR systems is the so called semantic gap: "*the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same date have for a user in a given situation*" (Smeulders, Worring et al. 2000). That is the retrieval is of an image represented by low level visual data and without any high-level semantic interpretation. A set of low level visual features cannot always precisely represent high-level semantic features in the human perception. The essential issue in CBIR

is that the user searches for semantic similarity, but the database can only reach homogeneity by data processing.

There are many conventional CBIR systems which have been widely used for general applications in image retrieval. Although, they have successfully produced good retrieval results in many applications, they still have major drawbacks (Qi and Han 2005):

1. They are too sensitive to visual signal distortion.

2. They have struggled to bridge the gap between low level features and the user's high level query semantics.

3. They are limited due to the lack of information about the spatial domain feature distribution.

In shape-based CBIR, discrimination power is required for a precise description, but when low level features are extracted these features usually lack the discrimination power required for accurate retrieval and this leads to inefficient retrieval performance (Kiranyaz, Pulkkinen et al. 2011).

There are five major approaches used to reduce the 'semantic gap' problem: Ontology-based techniques which rely on qualitative definitions of key semantic concepts and g are suitable for relative simple semantic features. Machine learning is capable of learning more complex semantic characteristics and is relatively easy to compute if the application problem can be well modelled. The Relevance feedback techniques are powerful tools to refine query results through modifying existing query samples till the users are satisfied. In order to improve the retrieval accuracy of CBIR techniques, this project has been focusing on reducing the semantic gap by using the Relevance feedback approach. The

FFCSS devised in this research bridges the gap between low level visual feature and high level semantic meaning through PZM iterations and the changing of moments parameters to satisfy users need. In the meantime, this research also focuses on the colour part of the object ontology through implementing the FCH method. Because the fuzzy membership function for weighting the colour features is more efficient than conventional "precise" methods. The FFCSS combines the advantages from both the Relevance feedback and object ontology for colour distribution, which leads to improved retrieval accuracy and speed. A new development in the field is called the Web Fusing that is considered as one of the state-of-the-art approaches in high image semantic level and its advantage stemmed from the vast knowledge pool on the Internet (Liu, Zhang et al. 2007).

CBIR techniques can be based on a single type of image features such as colours, shapes, or textures. Feature extraction using a single type of features is often inadequate. (Mianshu, Ping et al. 2010).

For bridging the gap between low level and high level concepts, advanced approaches are required and the techniques proposed in this research depend on the combination of different feature genres. Describing an image by combining multi-features is expected to give better results through enhancing the discrimination power of visual features to better interpret queries.

1.1 Motivation

CBIR is an attractive area of research because it is an active element of many important systems. In medical diagnosis imaging systems, where the medical

database is growing extremely fast, the effective use of such an immense system needs efficient CBIR-based indexing and retrieval technologies.

CBIR in also needed to ease access to the huge databases of digital images available on the internet, as well as supporting the enormous volumes yielded by digital cameras and scanning machines, where text retrieval and manual indexing become tedious and time consuming, if at all possible.

Traditional single feature CBIR techniques are relative simple to implement. However, the conventional colour-oriented and shape-oriented CBIR standalone features struggled to bridge the gap between the pixel values and the meaningful interpretation of an image. For example the colour histograms of some images look the same statistically but are completely irrelevant semantically.

This research studies difficulties that occur when using just individual features and to demonstrate how integrating these features can result a more efficient search clause for CBIR.

1.2 Aims and Objectives

The main aims of this research can be summarised as follows:

- To develop an efficient CBIR approach through the integration of fuzzy fusion of colour and shape features to produce superior performance on accuracy and speed over other conventional CBIR approaches.
- To design a new optimised fuzzy colour histogram-based technique for extracting representative colour feature vectors (signature) in high performance searching.

- To harness the power of shape feature moments for retrieval robustness in the presence of noise and variations.

1.3 Research Methodology

The general goal of this research is to investigate solutions to current CBIR problems through objectively and systematically analyse elementary and integrative CBIR techniques. The methodology follows in this context:

The problem identification process for this project starts with studying the challenges faces the CBIR application domain, including problem definition such as the problem of semantic gaps and curse-of-dimensionality. Then, the investigation moves on to how the proposed system would tackle the identified problems. The new methods are anticipated to add novel contributions to existing knowledge.

The research started by designing the first component of FFCSS, which is the FCH. By using fuzzy colour the so called curse of dimensionality can be avoided because the signature is compact by design.

The next stage of composition of the system was extracting the PZM descriptor feature and the orthogonal moments. PZM is used in this research because it has been successful applied to computer vision and pattern recognition.

The final stage was to merge FCH and PZM and link them together to define a strong and unified feature vector. There were many research methods used during the testing of the prototype system.

1.4 Thesis Structure

This dissertation is composed of seven chapters arranged in the following order:

Chapter 1- Research Background: introduces a brief of research background of CBIR and provides a summary of the proposed research contributions.

Chapter 2- Literature Review: reviews CBIR and current state-of–the-art techniques. An investigation of colour-based CBIR, shape-based CBIR and integration-based CBIR techniques are provided, and their advantages and limitations are discussed.

Chapter 3- Colour-based CBIR: provides an overview of colour-based CBIR concepts such as colour space, colour conversion and colour fuzzy techniques. A novel algorithm for computing the fuzzy fusion-based colour bins is presented, which relies on the fuzzy colour histogram.

Chapter 4- Shape-based CBIR: describes shape feature extraction, analysis, classification and segmentation. Several types of shape descriptor techniques are described. The pseudo-Zernike moments descriptor (PZM) which is the other vital component to build the proposed system (FFCSS) is introduced.

Chapter 5- Fuzzy Fusion of Colour and Shape Signature (FFCSS): introduces the prototype pipeline, design of FFCSS algorithms, and evaluation databases.

Chapter 6- Experimental Results and Evaluation: presents the evaluation of the results for FCH component, PZM component alone, and the final fusion FFCSS prototype. To examine the correctness and robustness of the proposed system, the FCH and PZM and the FFCSS systems are compared and how the FFCSS outperforms the FCH and PZM is described.

Chapter 7- Conclusions and Future Work: summarises the dissertation with a discussion of proposed algorithms and framework. The possibility of extension work is also discussed.

Appendix A- Representation of Pseudo-Zernike Moments: illustrates the computation of PZM in different levels.

Appendix B- Fuzzy Colour Histogram Algorithm and Results: shows the different query images for FCH and their retrieval results.

Chapter 2. Literature Review of Content-Based Image Retrieval

2.1 Introduction

With the rapid growth of digital devices for capturing and storing multimedia data, multimedia information retrieval has become a major research topic with image retrieval as one of the key challenges. In digital image processing and image retrieval systems, CBIR is an area of interest and has been applied widely in many computerised image applications (Hoi, Lyu et al. 2006). CBIR was first developed in the early 1990s to overcome the problems of the time-consuming manual-based image annotation approach. The image annotation was used to describe images in words and through searching process the user search word to bring similar text and corresponding image to that description. Although it is easy to build, it faces many challenges and it will discuss in Section 2.2.

The continuing rapid growth and enormous volume of the image collection databases in media technology demands more accurate search and retrieval approaches, since conventional database searches based on textual queries can, at best, provide only a partial solution to the problem. Database images are often not annotated with textual descriptions, and the vocabulary needed to describe the user's concept is not known to the user or may not exist. Moreover, a particular image can rarely be defined by a unique description. Thus, recently there has been immense activity in building direct content-based image search engines.

In CBIR, the image can be represented using visual features such as colour, texture, and shape or by combining different features. The features of all images in the database are extracted and compared with query image (Aggarwal, Ashwin et al. 2002). For any problem in CBIR, the solution starts from the image analysis and feature definition, the goal is to minimise the data and information which are not important, so that the redundant information can be neglected. The first task in any image analysis process is to select the criteria for identifying key information.

2.2 Image Annotation

Image annotation is defined as describing an image using a text format. Automatic image annotation or image classification is an important area in the field of machine learning and pattern recognition. Retrieval systems have traditionally used manual image annotation for indexing and responding to a query by retrieval from the image collection. These image collections are groupings of items, often documents or images. In image digital libraries, this designates all the works included, usually selected based on a collection management plan.

Manual image annotation suffers from several drawbacks. It is invariably tedious work, especially in large databases. It is also an expensive and labour intensive procedure, and it is limited to one language and is subjective to the user (Rahman, Desai et al. 2006).

2.3 CBIR Systems and Techniques

The term "content based image retrieval" (CBIR) was first used by (Kato 1992) to describe how to retrieve images automatically, based on the features of their contents. During the last two decades, valuable progress has been achieved

through research into both the theoretical and practical aspects of CBIR, and the literature shows a variety of approaches to describing images based on their content.

CBIR is considered an image search mechanism which can retrieve desired images relevant to the user's query from a large collection of images in a database. CBIR search techniques are sometimes denoted as query-by-image-content (QBIC) and the best known commercial CBIR approach was proposed and prototyped by IBM (Flickner, Sawhney et al. 1995), where a number of algorithms are deployed to allow users to form query clauses by combining multiple features such as colour, textures, and shapes.

CBIR operates on a different principle, retrieving stored images from a collection by comparing features automatically extracted from the query image with the targeted image sets. The commonest features used are statistical measures of colour, texture or shape. CBIR processes can be divided into five stages as illustrated in Figure 2-1.

The fundamentals stages of CBIR processes are:

- A first step which often requires segmentation, removal of image noise, and the conversion of the images into appropriate colour models.
- The second step is feature extraction, where the visual signals are often transformed into one or two dimensional vectors. In most cases, those features are coded texture, colour, and shape descriptors (El-Feghi, Aboasha et al. 2007).

- The third step is the actual retrieval of images. This consists of template matching where chosen features are used as matching tools and criteria for weighting the most similar images.

- The aforementioned processes in general explore the algorithmic ability to compute dissimilarities - distance measures - between the query image and the images in the database.

There is a large quantity of research and discussion concerning this development. Generally speaking, the signatures of an image should have two significant characteristics; it must be as representative of the image as possible; and it must be of reasonable dimensions. These two characteristics are essential to an accurate retrieval system in order to avoid the so-called a curse of dimensionality that incurs excessive computational cost (Smeulders, Worring et al. 2000).

Yong, et al. (1997) proposed a technique for converting image signatures in the image processing domain to a weighted signature in the information retrieval domain. This technique is unlike previous CBIR approaches, which were based only on image processing, but here the system, named the Multimedia Analysis and Retrieval System (MARS) was based on exploring the approach in both image processing and information retrieval. Yong, et al. also applied the relevance feedback technique from the information retrieval domain to assess retrieval results. MARS attempted to close the gap between high-level and low-level visual features and to reduce the subjectivity of the user through refining the user's query automatically at a feedback stage. It is considered as one of the first image retrieval approaches to apply a relevance feedback.
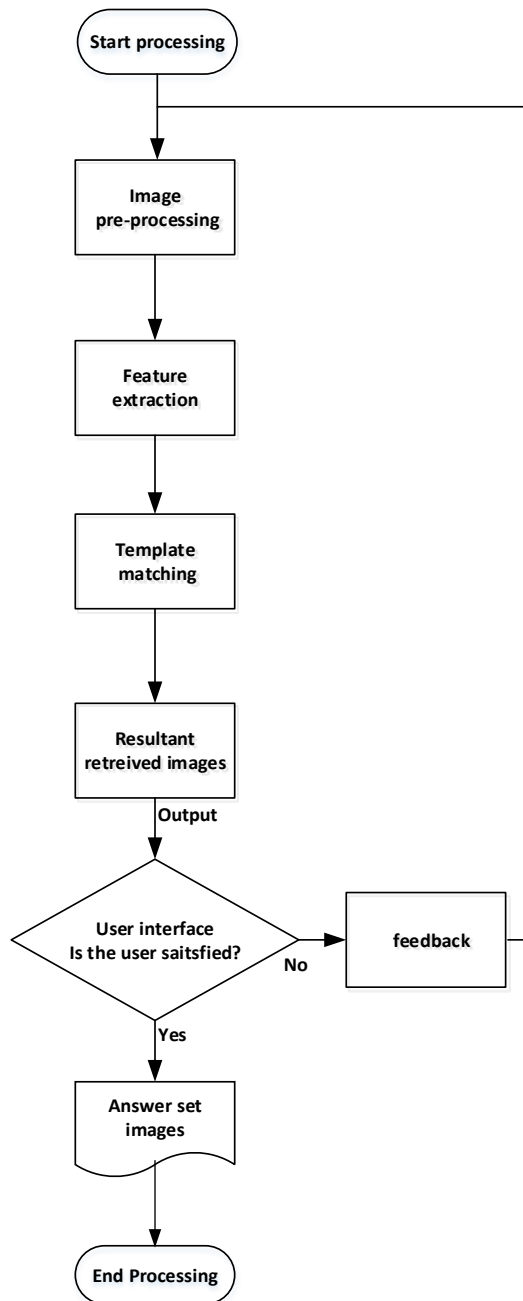
Figure 2-1   CBIR Processes

It also integrated two techniques to improve the retrieval process:

Firstly, obtain the features vector, convert the feature vector to a weighted vector and then use relevance feedback to evaluate the retrieved results. To extract a feature, Wavelet representation was used. The system receives a given image and a wavelet filter transferred that image into co-correlated sub-region. The feature extraction of orientation and scale of the original image is taken sub-region

by sub-region. Then the co-occurrence matrix representation approach used to extract texture features. Next the wavelet and co-occurrence feature vectors are combined to produce multiple components of both features as feature vectors.

2.3.1 Texture Content-Based Image Retrieval

The ability to retrieve images on the basis of texture similarity may not seem very useful, but can often be important in distinguishing between areas of images with similar colour histograms (such as sky and sea, or leaves and grass). A variety of techniques have been used for measuring texture similarity. The most established ones rely on comparing values of what are known as second-order statistics calculated from the query and stored images.

Essentially texture measures calculate the relative brightness of selected pairs of pixels from each image. From these it is possible to calculate measures of image texture such as the degree of contrast, coarseness, directionality and regularity (Tamura, Mori et al. 1978), or periodicity, directionality and randomness (Liu and Picard, 1996). Alternative methods of texture analysis for retrieval include the use of Gabor filters (Manjunath and Ma 1996). Texture queries can be formulated in a similar manner to colour queries, by selecting examples of desired textures from a palette, or by supplying an example query image. The system then retrieves images with texture measures most similar in value to the query. A recent extension of the technique is the texture thesaurus developed by Manjunath and Ma (1996). This method retrieves texture regions in images on the basis of similarity to an automatically-derived "codebook" representing important classes of texture within the collection.

The local pattern is considered to be one of texture as obtained by the CBIR technique, and this is considered similar to human perception. The well-known method used to describe local patterns is called a texture spectrum (histogram of all local patterns) and was first presented by He and Wang. This method is based on the idea of reducing the grey scales into three classes and counting all possible intensity patterns in a 3x3 window frame. The grey level value of the central pixel is compared with each of its eight neighbours. Each of the eight pixels is assigned the value 0 if its grey level value is less than the central value, 1 if the value is greater than central pixel and 2 if the pixels are of are equal value. The central pixel itself is not given any value and by using this method the number of grey levels is minimised to 3. The probability of possible combination is $3^8 = 6561$, and each pattern takes a value between 0 to 6561. Figure 2-2(a) shows the central pixel with all surrounding pixels brighter, whereas figure 2-2(b) illustrates the situation where the central and eight surrounding pixels are of equal brightness, and figure 2-2(c) depicts the patterns when all eight surrounding pixels are darker than the central pixel.

2	2	2
2		2
2	2	2

(a)

1	1	1
1		1
1	1	1

(b)

0	0	0
0		0
0	0	0

(c)
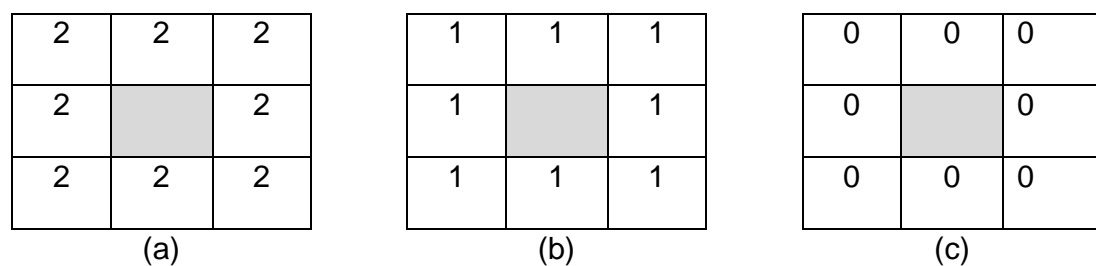
Figure2-2 The Central Pixel with Surrounding Pixels (a) Brighter, (b) Equally Bright or (c) Darker

Local contrast variations can be obtained using horizontal and vertical stripes instead of only light and dark spots which are compared with the central pixel. Jalaja and colleagues described several useful features such as surrounding contrast (SC) to describe any set of stripe patterns around the central pixel. These researchers used an alternating contrast (AC) texture feature to calculate the

occurrence frequency of any local pattern. Then they measured the frequency of occurrence by using local horizontal contrast and vertical contrast. The results showed that the contrast texture feature has benefits over other texture CBIRs. The pattern combinations were extended to a maximum possible of 3^{24}, with each pattern taking a value of between 0 to 3^{24}.

2.3.2 Colour Content-Based Image Retrieval

Several methods for retrieving images on the basis of colour similarity have been described in the literature (Yabuki, Matsuda et al. 1999; Seaborn, Hepplewhite et al. 2005; Falomir, Martí et al. 2010) but often are variations of the same principle. Each image added to a collection is analysed and a colour histogram computed which shows the proportion of pixels of each colour within the image. The matching technique most commonly used, histogram intersection, was first developed by (Swain and Ballard 1990).

The colour histogram for each image is stored in the database. At search time, the user can either specify the desired proportion of each colour (75% olive green and 25% red, for example), or submit an example image from which a colour histogram is calculated. Either way, the matching process retrieves those images whose colour histograms match those of the query to within specified limits. Variations of this technique are now used in a high proportion of current CBIR systems. Methods of improving on Swain and Ballard's original technique include the use of cumulative colour histograms (Pass, Zabih et al. 1997) combining histogram intersection with some element of spatial matching (Lazebnik, Schmid et al. 2006), and the use of region-based colour querying (Carson, Belongie et al. 1997).

The colour histogram serves as an effective representation of the colour content of an image. The colour histogram is easy to compute and effective in characterising both the global and local distribution of colours in an image. In addition, it is robust to translation and rotation about the view axis and changes only slowly with scale, occlusion and viewing angle. This is a very effective method if the colour pattern is unique compared with the rest of the data set. Any pixel in the image can be described by three components in a given colour space (for instance, red, green, and blue components in RGB space, or hue, saturation, and value in HSV space). The distribution of the number of pixels for each quantised bin, can be defined for each component and a corresponding histogram produced. Clearly, the more bins a colour histogram contains, the stronger discrimination power it has. However, a histogram with a large number of bins will not only increase the computational cost, but will also be inappropriate for building efficient indexes for image databases (Jung Uk, Seung-Hun et al. 2007).

Zhenhua and colleagues (2009) proposed a new method of colour feature extraction that depends on colour frequency. They used the HSV colour model instead of the RGB model because the RGB is suitable for display but is not appropriate for human perception (Zhang and Lu 2004). Thus the first step in their process is to change from RGB to HSV and complete the representation phase. Next is the colour quantisation process to reduce the number of distinct colours and this should be completed before feature extraction. The colour frequency sequence difference (CFSD) was proposed to solve problems associated with the

colour histogram such as high dimensionality of signature which leads to high computation requirements. These researchers used scalars to describe the colour feature of an image. Then the CFSD technique is integrated with

information entropy. Every image has a specific value of entropy but one value of entropy can apply to more than one image. Their experimental results showed outstanding accuracy of retrieval. The colour histogram is easy to calculate but faces several challenges such as the curse of dimensionality even with quantisation of the colour space. By using the CFSD method, the curse of dimensionality can be alleviated.

2.3.3 Shape Content Based Image Retrieval

In CBIR applications, shape features highlight local and global spatial distributions of the image patterns. Those shapes are defined by 2-D regions obtained from low-level pixel colour and distribution features, which are groups of connected image pixels sharing similar colours or textures. Generally speaking, the idea of image shapes is based on images appearing to share the same properties in the real world image scene defined by human vision systems, which is judged by human brains as geometric/affine invariant, noise/occlusion resistant and motion independent (Yang, Kpalma et al. 2008) .

Unlike texture, shape is a fairly well-defined concept and there is considerable evidence that natural objects are primarily recognised by their shapes (Biederman 1987). A number of characteristics of object shape are computed for every identifiable "item" within each stored image. Queries are then activated by comparing the same set of features from the query image, and retrieving those stored images whose features most closely match those of the query.

Two main types of shape feature are commonly used – global features such as aspect ratio, circularity and moment invariants (Wei, Li et al. 2009). Niblack et al. considered local features such as sets of consecutive boundary segments

(Mehrotra and Gary 1995). Alternative methods proposed for shape matching have included comparison of directional histograms of edges extracted from the image (Jain and Vailaya 1998).

Queries to shape retrieval systems are formulated either by identifying an example image to act as the query, or a user-drawn sketch (Kato, Kurita et al. 1992). Shape matching of 3-D objects is a more challenging task particularly where only a single 2-D view of the object in question is available. While no general solution to this problem is possible, some useful attempts have been made into the problem of identifying at least some instances of a given object from different viewpoints. One approach has been to build up a set of plausible 3-D models from available 2-D images, and match them with models already in the database (Chen and Stockman 1990). Another is to generate a series of alternative 2-D views of each database object, each of which is matched with the query image as done by (Shokoufandeh, Dickinson et al. 2002). Related research issues in this area include defining 3-D shape similarity measures (Shum, Hebert et al. 1996).

To develop an improved shape feature-based method for querying image databases, the Pseudo Zernike Moment (PZM) was adopted as a shape feature vector. To characterise shape features there are many approaches such a chain code, invariant moments, Fourier descriptors and the Zernike moment. In addition, statistical techniques to extract shape features are commonly used in pattern recognition due to their precise nature in computation and their inclusion of global and local features of the image.

As an example of shape detection, (Yong-Xianga, Cheng-Minga et al. 2007) proposed a new technique for object contour tracking of images of fruit based on

the chain code descriptor. They used the chain code for feature extraction as one of the contour tracking methods because of its simplicity, effectiveness and accuracy, and less storage of data is needed. The properties of the applied chain code include circumference, graph perimeter height and width. At the beginning, the pre-processing of the image is as follows: first image enhancement is used to improve the quality of the grey-scale image to obtain the binary image. Next step in the segmentation relies on the grey level threshold. The small non-connected regions are deleted as noise. Now the target binary image is ready for contour extraction. After the contours of the image of the fruit have been extracted using a graph contour tracking method, the next step is for the chain code to compute relevant characteristics and features of the fruit.

Shape feature extraction is considered as most powerful feature in CBIR with which to extract meaningful information (Choras 2007). Although colour and texture features are used in analysing the fundamental apparent-based features images, the shape feature is more effective when dealing with detecting objects in binary images. There are many approaches used to extract shape features. The chain code technique is simple to implement and requires less storage.

Another application associated with shape CBIR, is a new multi-view based ear signature extraction technique developed by (Heng and Jingqi 2007; Liu, Zhang et al. 2007). The ear is rich in geometric features and using just a single frontal view of an ear image has proved adequate for most recognition tasks but there is still scope for improvement. Using a multi-view image-based reconstruction method that uses side, front and rear views of the ear makes the discrimination power much stronger. The first component in Liu's technique is a sampling system for capturing multi-view images of an ear in a dark room illuminated with fixed

lighting. The idea behind this is to avoid the effect of affine transformation on the ear shape. The second stage uses Tchebicef radial polynomials to obtain high-order invariant moments of the ear. These geometric shape features are extracted from both front and rear views of the ear including length, area and width. Liu and his colleaugues have used a neural network for the Principle Component Analysis (PCA). PCA is a statistical algorithm for converting a set of features of correlated variables into a set of linearly independent vectors, or uncorrelated parameters and these parameters are called the principal components. The uncorrelated variables should be fewer than or equal to the original correlated variables hence reduce the feature vector dimensions.

The coming steps of the PCA algorithm can be summarised as (Moore 1981);

- Extract the feature matrix from the data and the feature vector is represented by the columns of the matrix.
- Compute the covariance for whole matrix to obtain linear independence between the properties.
- Obtain the eigenvalues by calculating the characteristic determination.
- Obtain the converting feature for orthogonal linear transformation.

Orthogonal moments are commonly used for pattern recognition. The most famous of the orthogonal moments are the Zernike moment (ZM) and Legendre Moment (LM). Although ZM and LM are extensively used for pattern recognition, they belong to the continuous moments category and this may cause error and loss of precision when the order of the moments increases and moment transformations required. An important advance in shape CBIR came in 2006 when Wang and colleagues successfully combined two types of orthogonal

moments: the Tchebichef and the Krawtchouk (Xianmei, Yang et al. 2006). They used discrete time Hidden Markov Model (HMM) to combine these orthogonal moments.

In the training phase they applied pre-processing steps including noise removal, linearisation and boundary determination. The next stage was the feature extraction stage which is transferring of a 2-D image into a 1-D vector. Then the recognition stage which is measuring the distance between the 1-D vectors. Wang and his colleagues divided their research into two main parts. First, a theoretical investigation of the Ktawchouk and Tchebichef moments. In the second part, they concentrated on the integration of discrete orthogonal moments and features of DHMMs for recognition of off-line handwritten Chinese. The proposed technique has demonstrated outstanding accuracy in retrieval, better results than conventional HMM recognition, but processing speed is rather low.

2.3.4 Hybrid Content Based Image Retrieval

The idea behind combining image features is to overcome such problems as semantic gap, complexity of segmentation and when two different objects have the same distribution of colour. A single feature may not adequately describe an image, and this suggests that a series of features may be the best way to represent an image. For example, (Liu, Jia et al. 2008) presented a new CBIR method for integrating colour and texture. They demonstrated that using more than one single feature improved performance of the retrieval process. As signatures they extracted the HSV colour histogram, the co-occurrence matrix as texture feature, and the moment invariant as shape feature. Unlike the single

visual feature approach that can describe only parts of the image content, using a fusion of low-level features can improve retrieval.

Another example of feature integration-based CBIR by (Shahbaz Khan, Anwer et al. 2012) proposed the use of colour attributes to describe colour presentation and distributions for object recognition. Object recognition often engages the process of detecting an object and its location accurately to yield a robust object recognition system (Aslandogan and Yu 1999). Commonly shape features are used for object recognition but Aslandogan and Yu used colour as well because the use of colour attributes is more efficient and compact, and when integrated with the shape features increases the system reliable. Their algorithm was evaluated by using the Pascal Visual Object Classes (VOC) database (Everingham, Van Gool et al. 2010). This method is easy to implement and shows high performance.

Another popular CBIR system is Virage which was released by Alta Vista in the 1990s to facilitate image searching as its counterpart from Yahoo-Excalibur. The most prominent rivalry of these systems is VisualSEEK, which was developed by Columbia University and released in 1996, which introduced a novel approach for image searching by using visual features from collections of image databases (Smith and Chang 1996). It resulted in a computer program that combines spatial query techniques with image indexing. The regional colours and their distributions are located, after that the technique seeks the images that include the most similar sequences of similar regions. It is claimed that this process is similar to the way human's perceive an image.

Each image is divided into regions which have feature properties: firstly using a system of colour representation using colour sets, and secondly spatial features such as location and size. This is done for each separate region. Images are checked for similarity by comparing the relationship between their regions. The system assesses queries that provide colour, size and spatial layout and of arrangement of colour regions. (Chang, Chen et al. 1997) also considered several special issues about spatial query layout such as: capsulation and overlap of regions. They found VisualSEEK used a completely different approach to colour histograms and mean colour, because the approach used by VisualSEEK uses salient image regions and their associated colours and locations, to compare images. This combination of image contents and their spatial features leads to the benefit of using several colour and special queries and makes the searching query a powerful search.

Another popular system is Photobook (Pentland, Picard et al. 1996). Photobook directly searches the contents of the image through the use semantics-preserving image compression, with the advantage that it reduces the images to a set of perceptually significant coefficients. The semantics-preserving query searching depends on two important conditions: firstly, the search has to be efficient enough to be more interactive because the searching which is required long time to process just only one image is useless. Secondly, the ability to search should be a similarity metric for comparing image features or shapes. Pentland et al.,

introduced three types of photobook description: the search based on appearance, a second based on textural features and the third used 2-D shapes. These three approaches were designed by the Massachusetts Institute of Technology (MIT). The photobook is an interactive tool that allows the user to

browse and search the content of images in large databases. This technique uses the query tool to search for the content of images and does not rely on text image annotation. Pentland, et al. used regional colours and their relative locations for segmentation.

The MARS system (Yong, Huang et al. 1997) was also popular in the 90s because it has a wide variety of query methods based on integrations between the image regions, their colour and spatial information.

Aggarwal and Dubey (2000) proposed an efficient system called iPURE standing for Perceptual and User-friendly Retrieval. It attempted to include the subjectivity of the user during the retrieval process by utilising the then novel idea of intra-query modification and learning. The devised system yields an objective level view of the query image using an effective colour segmentation approach. The estimation of user perception is extracted from the user feedback. The feedback consists of successive iterations and this leads to improved precision in searching the database, hence reducing the overheads of searching and downloading. Comparing this system to others showed that iPURE outperformed most other CBIR systems. For example, in the VisualSEEK system, the user is restricted to manual manipulation of the global features of the image while the iPURE technique adjusts retrieval results automatically. In comparison, QBIC relies on user query assignment of different weighting properties, such as shape, colour and texture, where the assignment is not an easy task even for experts with significant knowledge of the features in the database. Figure 2-3 illustrates the structure of the iPURE CBIR system.
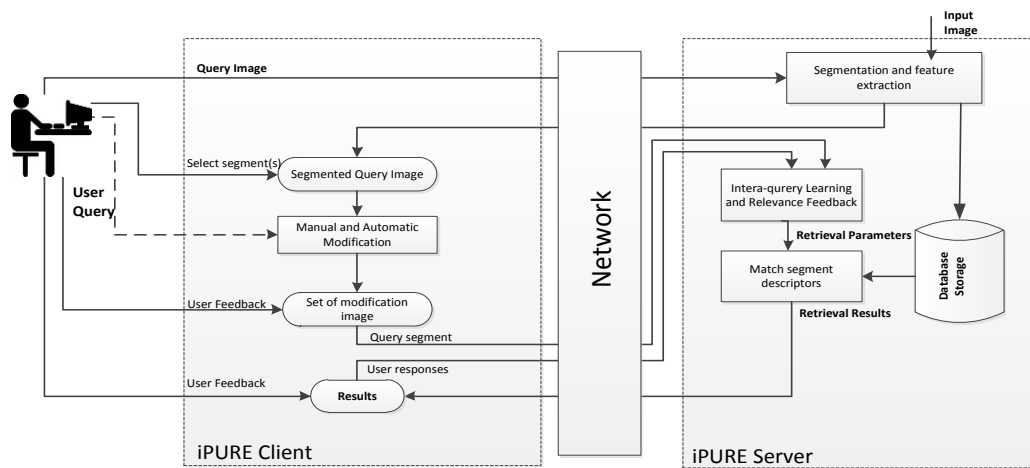
Figure 2-3   The Structure of iPure CBIR System (courtesy of Aggarwal and Dubey (2000))

Several studies in archaeology and history have focused on extracting character patterns from degraded historical documents. Many researchers use character pattern retrieval (CPR) to enhance the degraded historical documents. (Takakura, Kitadai et al. 2010) presented a similarity evaluation of Chinese character pattern to detect missing shape parts. They used historical documents to obtain 4911 Chinese character patterns called Kanji. They used feature extraction of patterns as the key and retrieved the most similar character from the archives as follows: The first step was to implement a feature matrix. In their approach they created a shape feature repository of the Chinese patterns using the chain code technique. The pixels were scanned to obtain non-liner characteristics. They applied a 3x3 pixel character pattern series for character recognition and acquired four different orientations of chain code features. Then, the next step was using the chain code to perform the contour tracking.

They then defined the 4 directional windows using an 8x8 grid Gaussian function for each direction. The distinctive features for a single Chinese character were recorded in a 64x4 matrix. However, there were still problems in the sampling

stage due to missing character sections, which was overcome by extracting the average colour from the "non-grey zones" and embedding the average colours into the missing sections. For further improvement, they proposed another method, called Template Modification (TM). This method does not create feature matrices of the template in the grey level. They used the binary image of the Chinese character pattern instead of using colour and grey scales because the historical documents (Mokkans) are too noisy to apply the gradient features of CPR. They improved the similarity evaluation method by proposing a feature extraction method based on simple presentation of gradient features, which had shown better repeatability than the chain code method on the trade-off of small degradations in retrieval accuracy.

In 2011, (Selvarajah and Kodithuwakku 2011) proposed a hybrid approach which combined texture and colour features for improving retrieval performance. They explored the colour histogram moment and used Discrete Wavelets Transform (DWT) to present a new approach that combined both colour and texture features. The DWT was selected because it can yield multi-resolution features of a discreet wavelet and simplifies the estimation of the edge of objects.

The process was; initially a 2-D DWT was computed through decomposing the given image into two sub-bands: the lower frequency band and the higher frequency band. The authors completed a three-level discrete wavelet transformations and explored 10 wavelet coefficients. They took 10 sets of measurements and then obtained the mean and standard deviation of the horizontal, vertical and diagonal data spreads. They then computed the histogram colour moments by using the mean and standard deviation of the horizontal, vertical and diagonal of the 10 coefficients. Then for the similarity measurement

they used the Sum-of-Absolute Differences (SAD) method for block matching of images. At system runtime, users submit a query image to the system for feature extraction, and then the feature set is compared to all image features in the database against a customisable threshold for determining the retrieval results. By using this fusion approach, more information on multi-resolution level can be encapsulated and utilised.

All approaches described above have limitations over sensitivity issues such as noise, image distortion, and struggled to bridge the gap between the user's high level semantic concepts and the low level image features.

As a consequence, a number of techniques such as spatial colour, colour correlogram, and colour coherence vector have been introduced in many CBIR systems to alleviate the shortcomings by using spatial information to find a special range of colour distribution in more details in spatial domain (Huang, Kumar et al. 1997; Rao, Srihari et al. 1999; Suryani and Guojun 2003). These approaches concentrate on the human perception and lead to the so-called semantic-based image retrieval. The conception is that the user defines the association of visual contents with semantic meanings of regions or objects.

2.4 Feature Extraction

Feature extraction is defined as the transformation of the content of an image into a suitable feature space. The process is one of obtaining the image's "signature vectors" that extract and preserve the measures of characteristics of interest such as texture. The features of an image can be categorised into two groups; general-purpose features and domain-specific features. The most used general-purpose features are texture, colour and shape. Examples of domain-specific features

often arise with special applications such as electronic pulses recorded from remote sensing devices, and volumetric scanning data from medical devices.

The two main requirements that are desired in any signature vector are discrimination power and the size of feature vector. The feature vectors should have the ability to represent all image details and the discrimination power is a way to uniquely describe an image and distinguish it from other visually similar images. The main aim of feature extraction is to establish appropriate and suitable feature vectors (signatures) that encapsulate the most meaningful and important information about the examined objects. It is one of the core processes in all pattern recognition systems.

In digital image processing (DIP) systems, the feature selection rules need to be justified with great care (El-Feghi, Tahar et al. 2011). Any digital images can be represented by a 2D grid of pixels. Thus, for classifying and recognising the properties of individual or groups of pixels, feature "descriptors" are often used, which are usually just a set of numbers in predefined formats and ranges. From this point of view, the studied subject patterns can be compared and recognised by matching the descriptors of an object against the extracted descriptors from unknown objects. For evaluating the retrieved results, effective descriptors should have four properties:

- They should be able to determine a complete feature set; thus two objects or regions should have identical descriptors if they are of the same shape.
- They should be based predominately on invariant properties to ensure robustness. For instance, rotation-invariant features.

- They should be of moderate dimension. As such, the descriptors will only take into account unique and meaningful feature information for distinguishing different objects. Any redundant information should be removed to make the feature descriptor as compact as possible, and to avoid the curse-of-dimensionality.

- They should be descriptive or congruent. That is, similar objects can be recognised if, and only if they have similar descriptors.

Based on the literature review carried out in this project, most of today's recognition systems only take into account subsets of these four criteria (Nixon and Aguado 2008). There are two types of feature extraction methods that have been deployed in these developed systems, the low-level-based and high-level feature-based strategies. Low-level features are well studied and can be extracted automatically from an image through statistical means such as a colour histogram, lines and corners. High-level feature methods are concerned with finding "shapes" in spatial terms that are often heavily influenced by pattern sizes and orientations (Liu, Zhang et al. 2007).

There are many semantic features such as colour layout and textures which are based on the basic characteristics of the images. Features in general can be categorised in three classifications, depending on their description methods.

- Low-level features: are the visual features such as colour or texture that can be obtained using computer vision approaches.

- Intermediate level features: obtained when segmentation techniques are applied to extract interest regions and can be fully automated by computer programs.

- High-level level features: need interpretation, and sometimes human intervention, for assigning semantic meanings to an image or image contents. The process often involves classifications of object types. The extraction of high level features is not an easy task without human assistance.

2.4.1 Texture Feature Extraction

Texture is another important property of images, which takes into account features such as roughness, and other regular or irregular patterns. It is a powerful feature in describing high-level feature semantics. Various texture representation schemes have been investigated for use in pattern recognition and computer vision. Basically, texture representation methods can be classified into two categories: structural and statistical. Structural methods, including morphological operator and adjacency graphs, describe texture by identifying structural primitives and their placement rules. They tend to be most effective when applied to textures that are very regular in appearance. Statistical methods which characterize texture by the statistical distribution of the image intensity include Fourier power spectra, co-occurrence matrices, shift-invariant principal component analysis (SPCA), Tamura feature, Wold decomposition, Markov random field, fractal model, and multi-resolution filtering techniques such as Gabor and wavelet transforms (Manjunath and Ma 1996). An example of texture feature extraction using wavelet transform is shown in Figure 2-4.

Original Image Transformation Wavelet Filter Wavelet Image
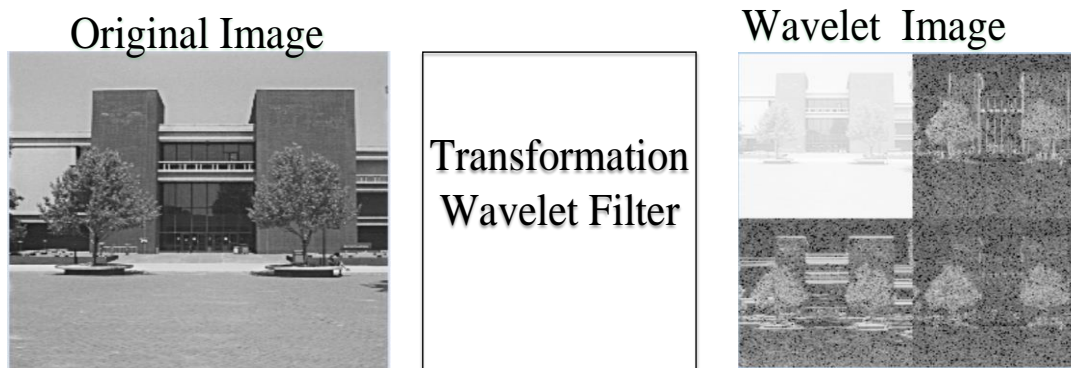
Breaking down of original image into frequency sub bands domain

Figure 2-4 Texture Features Extraction using Wavelet Transform

Texture is deemed a powerful approach to provide features in the CBIR domain. Texture has similarities with, but is more complex than, colour. It is a useful discrimination property for images. Feature extraction based on texture and shape are unlike colour since they describe regions or part of regions, while colour is defined by individual pixels. The main goal of texture representation is to extract some attributes or measurements which can be utilised later to specify a particular texture. Texture features are usually integrated with colour features to enhance retrieval accuracy. There are many types of measures for extracting texture from an image, the most common ones are Wavelets filters and Gabor filters. Co-occurrence matrices are also used for classifying textures (Mallat 1989; Manjunath and Ma 1996).

2.4.2 Colour Feature Extraction

The grey-level histogram is a function showing, for each intensity level, the number of pixels in the image having that grey value (Swain and Ballard 1990).

A Global Colour Histogram (GCH) is the most traditional way of describing the colour property of an image (Huang, Kumar et al. 1997). The GCH for an image is constructed by computing the normalized percentage of colour pixels in an image's n-dimensional feature vector (h_1, h_2, ... h_j ....h_n), where h_j usually represents the normalized percentage of colour pixels in an image corresponding to each colour element. More formally, the component h_j can be defined as a unique combination of the three values - red, green, and blue (obtained after normalization). The histogram by itself does not include any spatial information about an image. For example, the difference between a large region in red colour and a large number of scattered red pixels would not be captured by the colour histogram. In this context, the retrieval of similar images is based on the similarity between their respective GCHs. Since the similarity between two images is measured by the distance between their histograms, a very important feature of a colour signature is its strong discrimination ability. A common similarity metric is based on the city block distance between the abstracted feature vectors that

CITY BLOCK DISTANCE $$d(Q, I) = \sum \left| \left(h_j^Q - h_j^I \right) \right|$$ **Equation 2-1**

represent two images, and it is defined as:

Where Q and I represent the query image and one of the images in the image set, h_j^Q and h_j^I represent the coordinate values of the feature vectors of these images respectively. Note that a smaller distance reflects a closer similarity. The above inference stems from the fact that colour histograms are mapped onto points on an *n*-dimensional space, and similar images should therefore appear relatively close to each other (Huang, Kumar et al. 1997).

- **Primitive Features**

Colour is the most extensively used visual content for image retrieval. Its three-dimensional values make its discrimination potential superior to the single dimensional grey values of images. Before selecting an appropriate colour description, colour space must be determined. Although most of the images are in the RGB (Red, Green, and Blue) colour space, this space alone is rarely used in CBIR techniques. Other spaces such as HSV (Hue, Saturation, Value) (Cucchiara, Grana et al. 2001) or the CIE Laboratory (Ferschin, Tastl et al. 1994) and Luv (Tkalcic and Tasic 2003) are more frequently used as they correspond more closely to human perception.

- **Colour Moments**

Colour moments have been successfully used in many retrieval systems, especially when the image contains separate objects. The first order (mean), the second (variance) and the third colour moment (skewness) have been proved to be efficient and effective in representing colour distributions of images (stricker and Orengo 1995).Mathematically, the first three moments are defined, in order, as:

MEAN $$\mu_i = \frac{1}{N}\sum_{j=1}^{N} f_{ij}$$ **Equation 2-2**

VARIANCE $$\sigma_i = (\frac{1}{N}\sum_{j=1}^{N}(f_{ij} - \mu_i)^2$$ **Equation 2-3**

$$\textbf{SKEWNESS} \qquad\qquad s_i = \frac{1}{N} \sum_{j=1}^{N} \sqrt[3]{(f_{ij} - \mu_i)^3} \qquad\qquad \textbf{Equation 2-4}$$

Where f_{ij} is the value of the image pixel located at i and j special coordinates. N is the total number of pixels in the image. Using the skewness improves the overall retrieval performance compared to using only the first and second order moments. However, skewness sometimes makes the feature representation more sensitive to scene changes and thus may decrease the performance.

Hence only nine moments are used to represent the colour content of each image. Colour moments are a very compact description compared to other colour features. However, this compactness may lower the discrimination power. Usually, colour moments can be used initially to decrease the search space before other colour features are applied to improve retrieval (Long, Zhang et al. 2003).

2.4.3 Shape Feature Extraction

The shape feature is a powerful feature in CBIR systems. The shape feature vectors represent the shape image in numerical values, and each image has its own signature which should be unique to be able to differantate between shapes. In the retrieval processing the shape CBIR will retrieve the image most similar to the query image in terms of invariance to translation, rotation, and scaling. To construct a shape feature-based CBIR system which is fast, efficient and accurate has posed major challenges to real-world application developers. Because of its inherent complexity in defining and representing shapes, the shape signature-based CBIR has not matured as much as those for colour and texture.

Shape features or shape signature can be extracted from the silhouette of the regions of interest, but to extract these features one must convert the original formats to binary images (Hong, Yifei et al. 1998). To represent shape feature there are two shape descriptors, namely, region-based and contour based. Chapter 4 presents more information about these two basic descriptors.

2.4.4 Domain Specific Features

These features are used for specific applications of image retrieval. The following sections explain the domain specific application features vector which can be found in finger prints and face recognition, and other object recognition applications.

- **Fingerprint Features**

Each fingerprint image is converted to black and white first then the lines that flow in various directions across the fingerprints (ridges) and the spaces between ridges (valleys) are extracted to constitute the features for that fingerprint image (Hong, Yifei et al. 1998), see Figure 2-5.

Figure 2-5   Representation of Fingerprint

- **Face Recognition Process**

 In face recognition, the region of interest is only the main part of the face (Phillips, Flynn et al. 2005). It is required that the hair and other parts that do not contribute to recognition are removed. Thus some initial steps are required before extracting the face features, see Figure 2-6:

- Segmentation of objects in a scene.

- Approximate geometric shapes such as triangles, rectangles, circles, ellipses, etc.

- Create a secondary image that consists only of regions of interest.

- Extract features such as distance between objects and orientation.
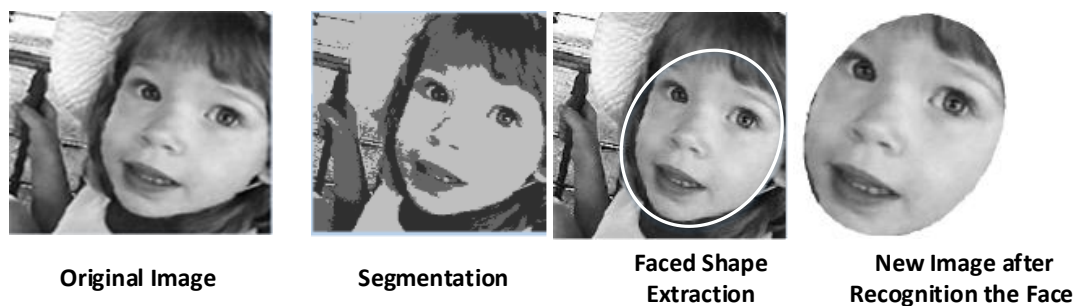

| Original Image | Segmentation | Faced Shape Extraction | New Image after Recognition the Face |

Figure 2-6   Some Steps Required before Extracting Face Features

The techniques used to detect faces in single image can be divided into four categories (Ming-Hsuan, Kriegman et al. 2002): knowledge-based techniques, face invariant feature methods, template matching approaches, and appearance based methods.

(Jaisakthi and Aravindan 2009) used a hybrid technique to combine Principle Component Analysis (PCA) and Lagrange Moment (LM) to detect faces. PCA is a method which often being used for larger data sets to reduce dimensionality. It has been successfully applied in many domains such as image compression, face

recognition and face detection. LM is considered as a powerful method in orthogonal moments approaches. Then the Support Vector machine (SVM) Segmentation method used to classify the faces after the shape features have been extracted by using LM. Although SVM is used as a two-class classifier, it can be expanded for multi-class uses.

The PCA and LM are often combined to form a hybrid approach because facial images are similar in nature but by using the LM, the variation in facial images can be highlighted. Also, both PCA and LM are robust, simple and their detection rate is high (Haddadnia, Faez et al. 2001) (Toygar and Acan 2003). Some external factors such as noise, light and facial expression can cause variations in the facial image but are ignored by most face detection techniques, while the weighted Eigenvalue technique can capture these variation and is very useful in the classification process.

 To extracting meaningful appearance based features of faces, pattern based LM moments are combined with the weighted Eigenvalue technique. The LM approach extracted moment based features from the input query image. Then, the SVM is used as a classifier to determine whether the input image is a face or not. The PCA and LM features are stored in a single signature.

To find a face when a query image is submitted to the system, the system first calculates the Eigenvalues and gives weights to all faces. The input face is represented by all the Eigenvalues. Next, the process extracts the LM for the given image. Finally, The SVM classifier receives the single vector which is the integration of the Eigenvalues and LM. The technique has been tested when

using PCA only and LM only. The results show that the hybrid technique outperforms the individual techniques.

- **Object Recognition**

Successful object detection and identification depends on object position, orientation, and size. Such a complex process requires an efficient algorithm with invariant object descriptors. Generally speaking, object descriptors should be robust, able to discriminate effectively, be independent and invariant.

In 2002, (Belongie, Malik et al. 2002) presented a method to measure similarity between shapes and utilised it for object detection. The basis of their algorithm was to estimate shape similarity and correspondences based on shape context. They tried to solve the correspondence problem between two shapes by calculating their correspondences and using this as an estimation of an aligning transform which is a measure of the distance between the two shapes. This measure relies on the computation of the sum of matching errors between corresponding points. The shape context approach attempts to find for each single point on the first shape the best corresponding point on the second shape. For matching they use local descriptors such as grey-scale instead of just using brightness at edge location.

- **Pattern Recognition**

Pattern recognition is defined as automated processing of tasks such as classification, recognition and description and these processes are important issues in many fields, especially in machine learning (Bishop 2006). There are many approaches used in pattern recognition and the most common is template

matching. This technique is applicable to detect the similarity between two shapes or curves, or even points.

One of the most famous pattern recognition applications is Optical Character Recognition (OCR), which is widely used in natural language processing, archiving and documentary management. In on-line recognition, handwriting is generated by means of an electronic pen or a mouse for acquiring or modelling time-dependent (Jain, Duin et al. 2000; Liwicki and Bunke 2007). In dealing with classification problems, a classifier is often estimated to support some desired invariant properties. For instance, the shifting invariance in character recognition is important so the changes of a character's location should not influence its clustering or classification results. A new approach has been introduced (Menasri, Vincent et al. 2007) based on the explicit grapheme segmentation method which incorporates the Hybrid Hidden Markov (HMM) model with a neural network-based recognition scheme. The advantages of this approach are rooted to its exploration of redundancy in shapes of, for example, Arabic letters.

2.5 Applications of CBIR

- **Medical Imaging and Diagnosis**

Content-based medical image retrieval is considered an important and prominent field. These databases contain enormous amount of digital medical images which can help thousands of doctors, on a daily basis, take the right decisions. CBIR in medical domain searches through huge collections of databases in an attempt to diagnose one case by using visual attributes. The main goal of medical imaging techniques is to store photos as a historic record in conditions such that, for

example, they will be readily accessible as an efficient tool to aid rapid diagnosis, assist in the training of doctors, and allow doctors to track the progress of a disease (Iakovidis, Pelekis et al. 2009).

The number of digitally produced medical images numbers tens of millions internationally and is rising rapidly. The management and access to these large image repositories becomes increasingly complex. Most access to these systems is based on the patient identification or study characteristics (modality, study description) (Müller, Michoux et al. 2004). This makes it difficult to retrieve images of the same characteristics for other (unknown) patients. To compare images of the same prosperities urgently requires CBIR technology advancements.

- **Crime Prevention**

Law enforcement agencies typically maintain large archives of visual evidence, including the facial photographs of past suspects (generally known as mugshots), fingerprints, and shoeprints. Whenever a serious crime is committed, they can compare evidence from the scene of the crime for its similarity to records in their archives. Strictly speaking, this is an example of identity rather than similarity matching, though since all such images vary naturally over time, the distinction is of little practical significance. More relevant is the distinction between systems designed for verifying the identity of a known individual (requiring matching against only a single stored record), and those capable of searching an entire database to find the closest matching records (Wayman, Jain et al. 2005).

- **Iris Recognition**

Iris recognition is a method of biometric authentication that uses pattern recognition techniques based on high-resolution images of the irides or irises of an individual's eyes. Not to be confused with another less prevalent ocular-based technology, retina scanning, iris recognition uses camera technology, and subtle IR illumination to reduce specular reflection from the convex cornea to create images of the detail-rich, intricate structures of the iris (Li, Tieniu et al. 2004). These unique structures are converted into digital templates, provide mathematical representations of the iris that yield unambiguous positive identification of an individual.

Iris recognition efficacy is rarely impeded by glasses or contact lenses. Iris technology has the smallest outlier (those who cannot use/enroll) group of all biometric technologies. The only biometric authentication technology designed for use in a one-to-many search environment, a key advantage of iris recognition is its stability, or template longevity as, barring trauma, a single enrollment can last a lifetime.

- **Digital Libraries**

Digital libraries are images of documents that contain mainly typed or typeset text. These are called textual images and are obtained by scanning the documents, usually for archiving purposes. They also use a procedure that makes the images and their associated indexing available to anyone through a computer network (Bin, Ramsey et al. 2000). The large portion of a textual image is text that can be used for retrieval purposes and efficient compression. In the case of textual images, further compression can be achieved by extracting the different symbols

or marks from images, building a library of symbols for them, and representing each image by a position in the library.

- **Intellectual Property Rights and Trademark Registration**

Trademark image registration is where a new candidate mark is compared with existing marks to ensure no risk of confusing property ownership. The trademark retrieval system analyses each image to characterize key shape components, grouping image regions into families that potentially mirror human image perception, and then derives characteristic indexing features from these families and from the image as a whole (Gudivada and Raghavan 1995).

- **Military Applications**

Military applications of imaging technology are probably the best-developed, though least publicised. Recognition of enemy aircraft from radar screens, identification of targets from satellite photographs, and provision of guidance systems for cruise missiles are known examples – though these almost certainly represent only the tip of the iceberg. Many of the techniques used in crime prevention are relevant to the military field (Gudivada and Raghavan 1995).

- **Web Searching**

Cutting across many of the above application areas is the need for effective location of both text and images on the Web, which has developed over the last five years into an indispensable source of both information and entertainment. Text-based search engines have grown rapidly in usage as the Web has expanded; the well-publicized difficulty of locating images on the Web indicates that there is a clear need for image search tools of similar power (Smith and

Chang 1997). There is also a need for software to prevent access to images which are deemed illegal or inappropriate. Alta Vista and Yahoo search engines now have CBIR facilities.

There are many challenges facing the concurrent CBIR techniques, namely a couple:

- Curse-of-dimensionality, which occurs when the size of a feature vector is too big that literally renders the practical computation and analysis impossible for high resolution images and massive scale image databases.
- Semantic gap, which often appears as the conflicts in between the interpretations of low-level visual feature and their high-level semantic and context-related meanings.

Our proposed system tackles the aforementioned problems by integrating optimised elementary and integrative visual features. For example, features defined by the fuzzy theory associated with colour histogram usages; and adaptive and robust invariant shape features based on the orthogonal moment's approaches.

Chapter 3. Colour-Based CBIR

This chapter introduces the concepts of Colour-based CBIR, fuzzy logic theory, membership functions and how colour representation can be calculated by using fuzzy colour distributions.

3.1 Introduction to Colour-Based CBIR

Colour-based CBIR is one of the most widely used features of CBIR. To improve data handling efficiency, colour has been used as a feature vector to represent the content characteristics within an image. Although colour can be an efficient representation for digital images, it carries little information about the spatial structures and shape features within the image. Because of this drawback, this project integrates different image features into the search and retrieval processes to make them more efficient and reliable. As highlighted by (Appass and Darwish 1999), the key to a good retrieval is the ability to use appropriate feature spaces to ensure the images are represented as precisely as possible.

A hybrid colour-shape scheme based on an adaptive fuzzy colour mechanism has been introduced in this research for constructing robust CBIR systems. The research has investigated the state-of-the-art of FCH CBIR implementation. Based on the review findings (see Chapter 2), development work has been carried out to integrate different feature spaces to improve the accuracy of image definition and the consequent retrieval operations. Artificial intelligence (AI) techniques, such as fuzzy logic systems, have been introduced into this

programme to improve the efficiency of the querying process by reducing redundant information and calculations.

3.2 Colour Space

The concept of colour space, also called "colour system" or "colour model", is used to describe the properties of colours according to image content and allocate these properties in a coordinate system and its subspaces, so that a single colour can be represented by a single point (Yining, Manjunath et al. 2001). Different colour spaces are available, appropriate for different applications such as RGB and HSV and brief explanations of these are provided below.

- **Representing Digital Images**

Given a digital image $f(x, y)$ sampled by M rows and N column of pixels. The resulting values of the coordinates can be transformed into discrete quantities. The coordinates (x, y) at the origin = (0, 0). The next values of coordinates in this array in this first row, see Figure 3-1, are $(x, y) = (0,1)$. The last coordinates along the first row are $(x, y) = (0, N - 1)$.

The notation of the full digital image, $M \times N$, is summarised in the following matrix:

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1)... & f(0, N - 1) \\ f(1,0) & f(1,1)... & f(1, N - 1) \\ \vdots & \vdots \quad \vdots & \vdots \\ f(M - 1,0) & f(M - 1,1)... & f(M - 1, N - 1) \end{bmatrix}$$

Thus the matrix, $f(x, y)$ describes the digital image. Each coordinate in the matrix is called a pixel, pel or an image element (Gonzalez and Woods 2002).
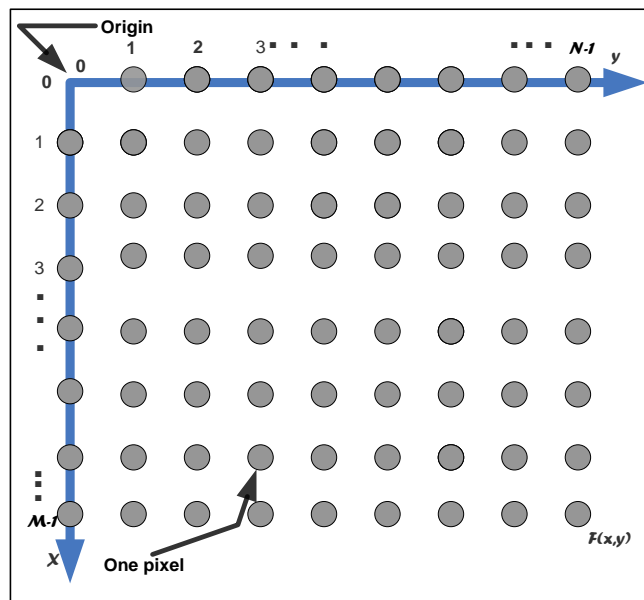
Figure 3-1 Representation of the Digital Image

In grey value images, the pixel values are between 0 and 2^L where L is the number of bits required to represent whole image. For example, when using 1 byte to represent each colour, the number of levels will be (2^8) =256. Generally, in colour images, each pixel can be represented by 3 bytes or 24 bits (Red, Green, and Blue) and each of the three colour components is represented by 256 levels.

- **RGB Colour Space**

Most digital image processing techniques treat the image as a fusion of pixels formed of red, green, and blue (RGB) values as shown in Figure 3-2, with red, green and blue forming the three Cartesian axes with values normalised to lie between [0, 1]. Each grey level along the diagonal and each component is categorised into 256 levels [0:255], the RGB colour model represents 8*3=24 bits. Then the total number of different colours that can be represented is equal to 2^{3*8}-1 = 2^{24}-1 = 16,777,216. RGB colour: quantize each component into 6 levels from 0 to 255 (Gonzalez and Woods 2002).
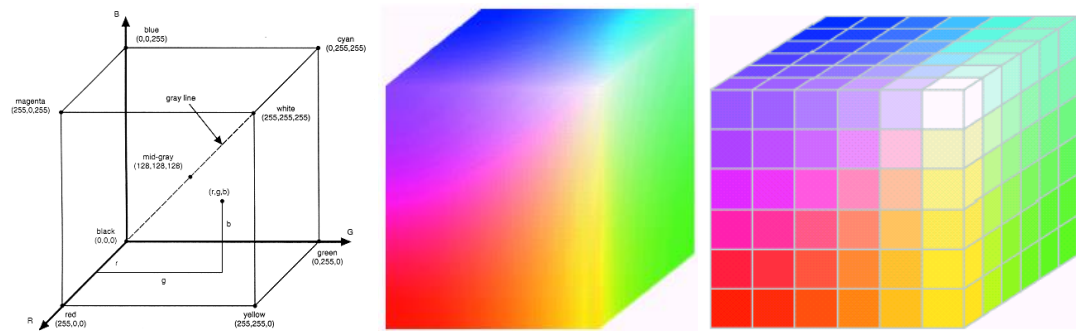
Figure 3-2 Representation of RGB Colour Space

The RGB colour space is very convenient for display purposes because computer monitors display the output colour by mixing different amounts values of red, green and blue.

- **HSV colour space**

However, the way computer monitors produce colours does not correspond to human perception. Humans perceive colours in terms of Hue (H), Saturation (S) and Value (V). HSV is usually referred to as Perceptual Colour Space (PCS) (Liu, Zhang et al. 2007). Thus the HSV is considered superior to RGB space because HSV represents colour in a way more closely corresponding to the human vision system. Although there are other perceptual colour systems such as HSL (Hue, Saturation, Lightness) (Tkalcic and Tasic 2003), our research focuses on HSV because it is commonly used. Figure 3-3 represents the HSV colour space as a cone with its apex pointing downward. Hue is considered as the angular value around the colour circle at the top of the cone. The zero angle line is assigned to be red. The saturation is defined as the radial distance from the origin (centre of the circle). The brightness is determined by the vertical position of the colours in the cone. The top of the cone is maximum brightness and the eye would see the

colour white. At the bottom of the cone, there is no brightness and the eye would see the colour as black.
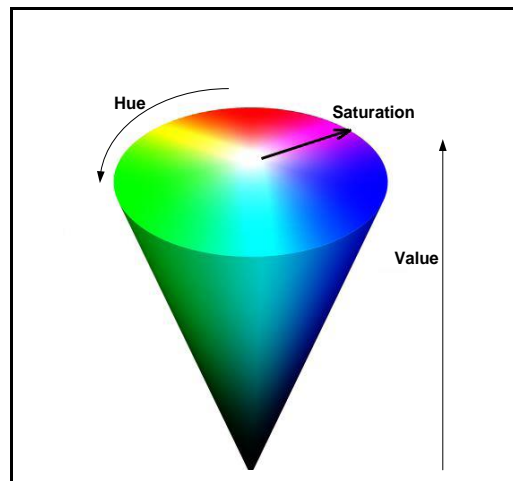

Figure 3-3   HSV Space

In this research, the Fuzzy Colour Histogram (FCH) technique is applied to the HSV colour model as the first component of FFCSS image retrieval. To overcome the problem when using histograms in colour-based CBIR which considers only global attributes and thus cannot effectively and efficiently represent an image, a fuzzy prototype has been developed to capture local attributes for more precise descriptions. The original colour image is divided into groups of sub images and colour histograms are constructed for every sub-image. This collection of colour histograms is combined into a multidimensional signature vector to search a collection of database for similar images.

- **Colour Conversion**

Given a colour represented by (R, G, B) where each of R, G, and B have a value in the range between 0 to 255, (and similarly for H, S, V), the colour can be defined as:

$$
\text{HUE} \quad H = \begin{cases} \left(\dfrac{G - B}{max - min}\right) * 60, & if\ R = max \\[2mm] \left(2 + \dfrac{B - R}{max - min}\right) * 60, & if\ G = max \\[2mm] \left(4 + \dfrac{R - G}{max - min}\right) * 60, & if\ B = max \\[2mm] Undefined, & if\ R = G = B \end{cases} \qquad \textbf{Equation 3-1}
$$

$$
\text{SATURATION} \quad S = \begin{cases} 0, & if\ max = 0 \\[2mm] \left(\dfrac{max - min}{max}\right) * 255, & otherwise \end{cases} \qquad \textbf{Equation 3-2}
$$

$$
\text{VALUE} \quad V = \{the\ maximum\ value\ of\ (R, G\ or\ B) \qquad \textbf{Equation 3-3}
$$

Let: min denote the minimum value of all the RGB pixels and max denote the maximum of these values (El-Feghi, Aboasha et al. 2007). Then the HSV colour model can be calculated as following formulas:

According to Eq. 3-1, it can be seen that the H value depends on the most dominant colour (R, G or B) which will be shown by the angle on the upper face of the cone shown in Figure 3-3. If Blue has the largest value, then the ratio of the difference between the red and the green divided by the difference between the min and the max values must lie in the range -1 to 1. This ratio is multiplied by 60 producing a value of the H between 60 and -60 (Ford and Roberts 1998). This is repeated for the other H values.

The S value ranges between 0 to 255 as shown in Eq. 3-2. If (R = G = B) then we have a grey value, if the colour is black, S = zero, and V is equal to the value of the max colour.

3.3 Conventional Colour Histogram (CCH)

Due to its relative simplicity, CCH, which captures global colour distribution in an image, is a popular representation that has been used in many image retrieval systems and is the most commonly used colour space in CBIR. CCHs are easy to calculate and implement, they provide large feature signatures but are not easy to index and have high search and retrieval cost and suffer from the "curse of dimensionality". Moreover, spatial information cannot be shown in a colour histogram. Thus a large blue colour on a red background will have the same colour histogram as an image containing the same number of randomly distributed blue and red pixels (Ju and Kai-Kuang 2002). CCH has been considered as one of the most basic techniques. To test performance improvements, the CBIR retrieval systems has been compared to a system using only colour histograms (Xiaoling and Hongyan 2009).

Image colour histograms have both advantages and drawbacks (Krishnan, Banu et al. 2007):

- They are easy to implement.
- They are fast, the histogram computation have $O(n^2)$ complexity for an image $(n * n)$ pixels, where n denotes the number of histogram bins for all different colours.
- It has low storage requirements. The size of the colour histogram is much smaller than the image size itself.
- But the drawback is that no spatial information can be represented because two completely different images may have equivalent histograms (El-Feghi, Aboasha et al. 2007).

3.4 Colour CBIR Component Based on Fuzzy Set Theory

Many decision-making and problem-solving tasks are difficult to analyse quantitatively. Acceptable solutions can often be found by using knowledge that is imprecise rather than precise. Fuzzy set theory resembles human reasoning in its use of approximate information and uncertainty to generate decisions. It has been specifically designed to mathematically represent uncertainty and provides formal tools for coping with the imprecision that exists in many real problems. Knowledge can be expressed in a more natural way by using fuzzy sets, and many engineering and decision problems, such as the scheduling of trains on the Sendai Subway in Japan where the use of fuzzy control enables the provision of the steadiest running subway system across the globe. The control of complex air conditioning systems to heat and cool large buildings in extreme weather conditions can be greatly simplified by using fuzzy logic (Dutta 1993). More formally, fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth - false-values between "completely true" and "completely false" (Klir 1995).

Fuzzy logic was known during the time of Plato, but it did not get much attention until 1965 when Zadeh published his work on fuzzy sets (Zadeh 1980), in which he described the mathematics of fuzzy set theory and, by extension, of fuzzy logic. This theory proposed making the membership function (or the values False and True) operate over the range of real numbers [0.0, 1.0]. The new operations have been recognised as a generalization of classic logic.

In classical set theory, datasets are represented by so-called crisp definitions. Any member either "belongs" or "does not belong" to a data set. In fuzzy set

theory, classes or groupings of data have boundaries that are not sharply defined (i.e., fuzzy). Any members in the fuzzy set can belong to the dataset with a certain value or membership. The process of finding the membership is called "fuzzification" which is a generalization of the concept of a crisp set to a fuzzy set with blurred boundaries. The benefit of extending fuzzy techniques and analysis methods to crisp theory is the increased power gained for solving real-world problems, which inevitably have imprecision and noise when measuring and using variables and parameters in decision making applications (Hong and Choi 2000). Accordingly, linguistic variables are a critical aspect of some fuzzy logic applications, where general terms such as "large", "medium" and "small" are used to capture a range of numerical values. While similar to conventional quantization, fuzzy logic allows these stratified sets to overlap (e.g., a 85 kilogram man may be classified in both the "large" and "medium" categories, with varying degrees of belonging or membership to each group). Fuzzy set theory encompasses fuzzy logic, fuzzy arithmetic, fuzzy mathematical programming, fuzzy topology, fuzzy graph theory, fuzzy data analysis and fuzzy neural networks, though the term fuzzy logic is often used to describe all of these.

The definition of a fuzzy set is (Zimmermann 2010), a fuzzy set A is a subset of the universe of discourse X that admits partial memberships. The fuzzy set A is defined as an ordered pair $A = \{x, \mu(x)\}$, where $x \in X$, and $0 \leq \mu_A(x) \leq 1\}$. The membership function $\mu_A(x)$ describes the degree to which the object x belongs to the set A, $\mu_A(x)=0$ represents no membership, and $\mu_A(x)=1$ represents full membership.

As an example, let X represent the age of all people. The subset A of X represents people who are young. As can be seen from Figure 3-4 all person aged 25 years

or younger have a membership of $\mu_A(25)=1$ while another person who is 50 years old has a membership of $\mu_A(50)=0.5$.
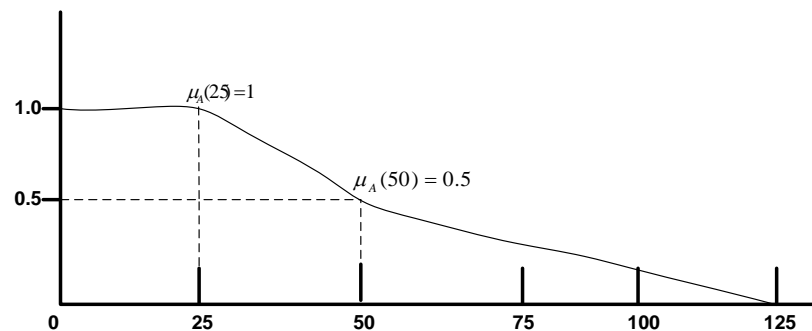
Figure 3-4   The Membership Function Describing the Relation between a Person's Age and the Degree to which that Person is Considered Young

Using crisp logic, a person is 25 years or younger is considered young and not young otherwise, which means that someone who is aged 24.99 years is considered young while someone who is aged 25.01 years is consider old. With fuzzy logic, on the other hand, a person who is aged 24.99 can be a member of the subset "young" with a given value and to subset "old" with another value; if these are the only two sets the sum of the values must be 1.0. For example if a person is a member of the "young" set with a value 0.9 s/he belongs to the "old" old with a value 0.1.

Another example, consider a fuzzy set *A* representing all numbers that are close to 6. Since "close to 6" is a fuzzy statement and there is no unique membership function representing the set A, the designer has to decide what the membership criterion should be. In crisp logic the membership can be either 1 or 0, but fuzzy logic allows flexible values to be assigned to represent "close to six".

Figure 3-5 shows two fuzzy ways of describing the membership function represented by the fuzzy statement "close to 6"
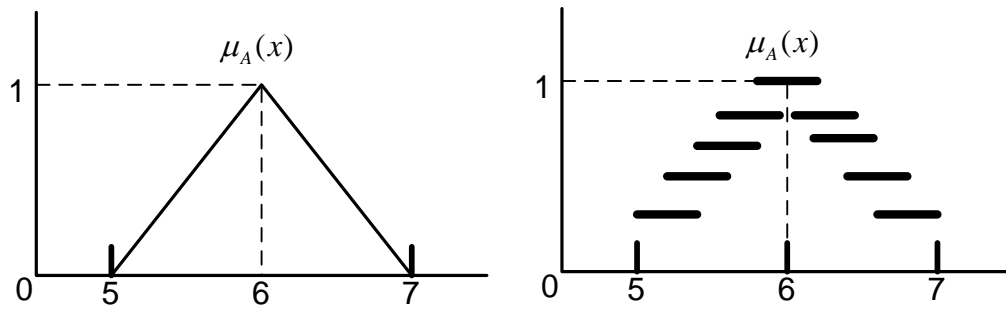
Figure 3-5   Two Representations of Membership Function of the Fuzzy Set that Represents "Real Numbers Close to 6"

- **Properties of Fuzzy Sets**

The operations on fuzzy sets are extension of those used with traditional sets. The common operations of fuzzy sets include *union, intersection, complement, comparison and containment.* Let X be the universe of discourse, $A \in X$ and $B \in X$ the following operations are defined (Zimmermann 2010):

Union (A $\cup$ B):	$\mu_{A \cup B}(x) = \max(\mu_A(x), \mu_B(x)) \ \forall x \in X$
Intersection (A $\cap$ B):	$\mu_{A \cap B}(x) = \min(\mu_A(x), \mu_B(x)) \ \forall x \in X$
Complement ($\overline{A}$):	$\mu_{\overline{A}}(x) = 1 - \mu_A(x) \ \forall x \in X$
Equivalence Relation:	$A = B \ \textit{iff} \ \mu_A(x) = \mu_B(x) \ \forall x \in X$
Containment:	$A \subset B \textit{iff} \ \mu_A(x) \prec \mu_B(x) \ \forall x \in X$

In addition to these operations, De Morgan's law, the distributive laws, algebraic operations such as addition and multiplication, and notation of convexity have fuzzy set equivalents. Table 3-1 presents properties of fuzzy sets.

Table 3-1 Properties of Fuzzy Sets

Commutative	$A \cup B = B \cup A, A \cap B = B \cap A$
Distributive	$A \cup (B \cup C) = (A \cup B) \cup C \quad A \cap (B \cap C) = (A \cap B) \cap C$
De Morgan's Laws	$\overline{(A \cup B)} = \bar{A} \cap \bar{B}, \overline{(A \cap B)} = \bar{A} \cup \bar{B}$
Algebraic Sum, $A + B$	$\mu_{A+B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x).\mu_B(x) \ \forall x \in X$
Algebraic Product, $A \otimes B$	$\mu_{A \otimes B}(x) = \mu_A(x).\mu_B(x) \ \forall x \in X$ $\mu_{A \otimes B}(x) = \mu_A(x).\mu_{AB}(x) \ \forall x \in X$
Bounded Sum, $A \oplus B$	$\mu_{A \oplus B} = min \ (1, \mu_A(x) + \mu_{AB}(x)) \ \forall x \in X$
Boundd Difference, $A \ominus B$	$\mu_{A \ominus B} = max \ (0, \mu_A(x) + \mu_B(x)) \ \forall x \in X$

3.4.1 Membership Function

A fuzzy set is completely characterized by its membership function. Since most of the fuzzy sets in use have a universe of discourse X, it is impossible to list all the pairs defining the membership function. A more convenient way is to express the membership function in a mathematical form. In this section we present the most popular mathematical forms for representing membership functions (Klir 1995):

Triangular Membership Functions: are defined as follows, see Figure 3-6, (Pedrycz and Gomide 2007):

TRIANGULAR MEMBERSHIP FUNCTION

$$f(x; a, b, c) = \begin{cases} 0, & if\ x \le a \\ \dfrac{a - x}{a - b}, & if\ a \ge x \ge b \\ \dfrac{x - c}{b - c}, & if\ b \ge x \ge c \\ 0, & if\ c \le x \end{cases}$$

Equation 3-4

Where the scalar parameters a and b represent the left and right boundaries of the set and c represents the value of the central parameter of the set as shown in Figure 3-6.
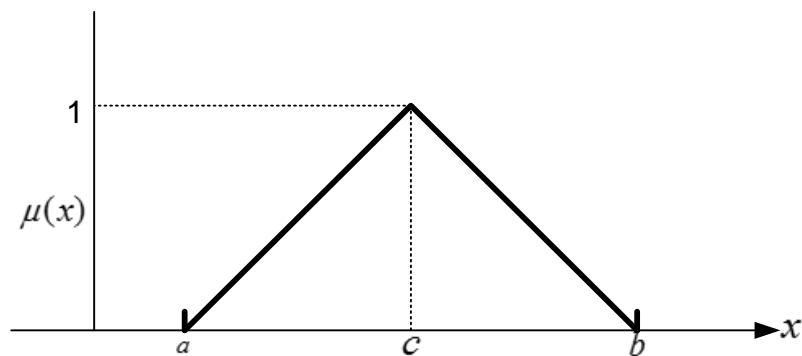
Figure 3-6   A Triangular Membership Function

Example, Triangular Membership $f(x, 3,6,8)$
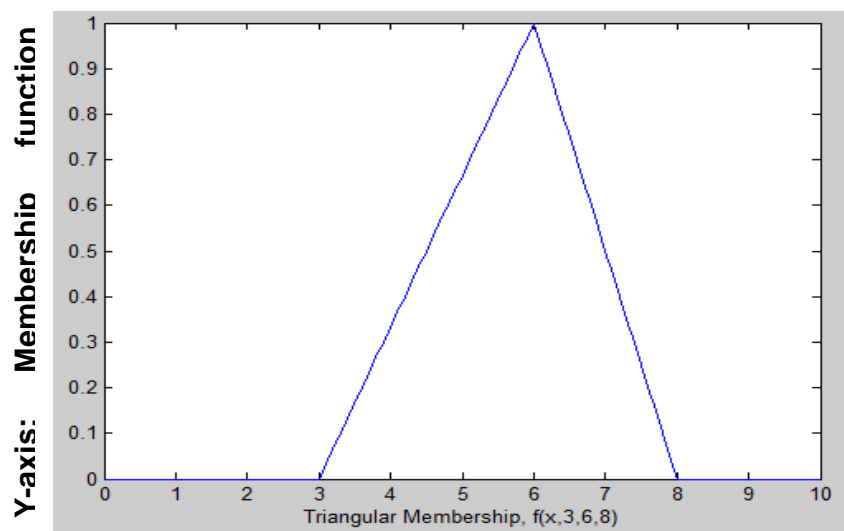
X-axis: Boundaries parameter and centre parameter
Figure 3-7   Triangular Membership Function $f(x, 3, 6, 8)$

Trapezoidal Membership Functions are defined as follows (Klir 1995):

TRAPEZOIDAL MEMBERSHIP FUNCTION

$$f(x, a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases} \qquad \textbf{Equation 3-5}$$

$f(x, a, b, c, d) = \max(\min(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}), 0).$

Example, Trapezoidal Membership $f(x, 1,5,7,8)$
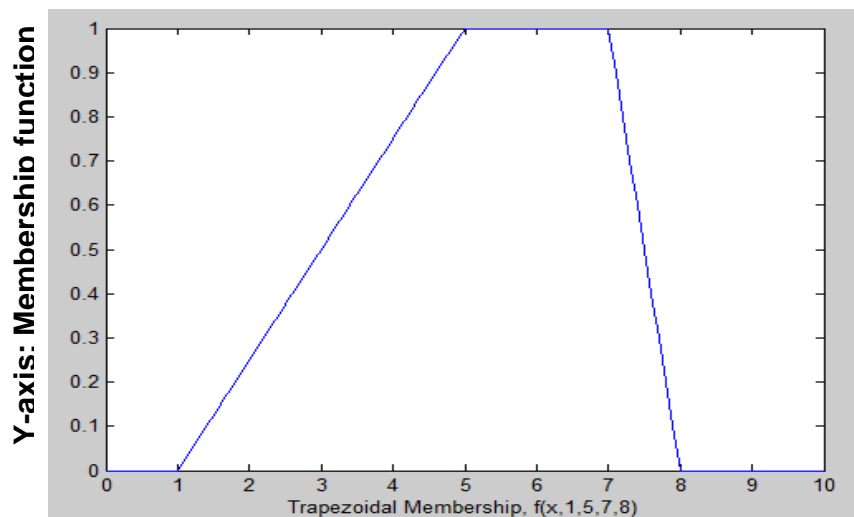
X-axis: Boundaries parameter and centre parameter
Figure 3-8   Trapezoidal Membership Function $f(x, 1, 5, 7, 8)$

- **Gaussian Membership Functions** are defined as follows (Pedrycz and Gomide 2007):

GAUSSIAN MEMBERSHIP FUNCTION

$$f(x, \sigma, c) = e^{\frac{-(x-c)^2}{2\sigma^2}} \qquad \textbf{Equation 3-6}$$

Example, Gaussian Membership $f(x, 2,5) = e^{\frac{-(x-c)^2}{2\sigma^2}}$ where c=2 and σ =5
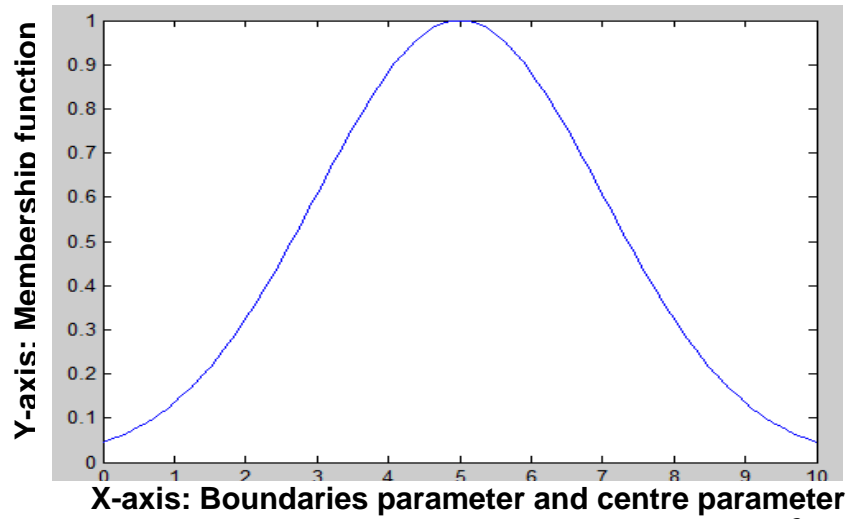
Figure 3-9   Gaussian Membership Function $e^{\frac{-(x-c)^2}{2\sigma^2}}$

- **Generalized Bell Membership Function**:

GENERALISED BELL MEMBERSHIP FUNCTION

$$f(x, a, b, c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}}$$ **Equation 3-7**

Where *a* denotes the width and *b* denotes the slope of the function, and *c* denotes the centre of the set.

Example of the Generalized Bell Membership Function with a=2, b=4 and c=6

Generalized Bell Membership, f(x,2,4,6)

X-axis: Membership function value
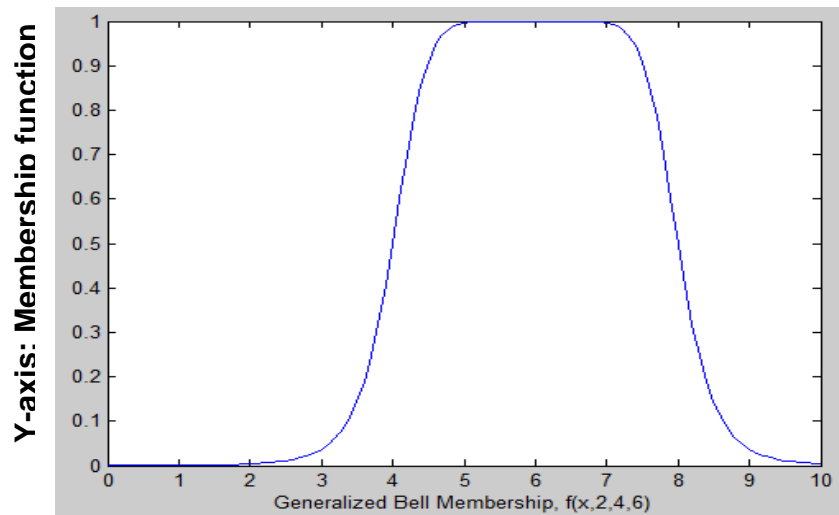
Figure 3-10   Generalized Bell Membership Function $f(x, 2, 4, 6) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}}$

3.5 Fuzzy Systems

Fuzzy rule-based systems are the most successful in decision-making systems, expert systems and control systems (Bellman and Zadeh 1970). A fuzzy rule-based system consists of 4 main parts: Fuzzifier, Knowledge Base, Inference Engine and Defuzzifier. The first part, the fuzzifier, converts input data from crisp values to fuzzy values based on membership of fuzzy sets. The knowledge base is implemented by fuzzy sets and fuzzy inference rules in the form of "If x is A then y is B". The inference engine implements a fuzzy inference algorithm that deduces results from inference rules and the given inputs (Pedrycz and Gomide 2007). The defuzzifier converts fuzzy data to crisp outputs. Fuzzy systems store

"rules" and estimate sample functions from linguistic inputs to generate linguistic outputs. Fuzzy methods are powerful in modelling human thinking and perception. They express human experience in a form that a machine can use to imitate human pattern recognition and judgment capabilities and convert the information into a form that people can understand.

3.5.1 Fuzzy Colour Histogram (FCH)

The CCH considers neither the colour similarity across different bins nor the colour dissimilarity in the same bin. Thus, it is very sensitive to noise interference such as illumination changes and quantization errors. Furthermore, the dimension of CCH makes the computation quite long. To overcome this problem, a colour histogram representation called FCH has been developed in this research. FCH considers the colour similarity of the colour of each pixel associated with every histogram bin through a fuzzy-set membership function.

In the system, each pixel value is converted to spectrum with 6(Hue) ×
3(Saturation) × 1(Value) bins. In addition 3(grey scale) bins have also been added
for the special case where Hue is undefined when we compute the signatures
from RGB histograms. A signature histogram is constructed by summing the
values of HSV for each of the 21 histogram bins. To overcome the colour
distribution problem mentioned in Section 3.3, and as shown in Figure 3-11, the
image is divided into 2 regions and two fuzzy colour histograms because like
Belgium and Germany flag have the same distribution of colour but in different
order and many systems can differentia between them when try to retrieve
wanted flag. Obliviously, from the figure the percentage of red colour, black and
yellow are the same in two pictures but the orientation is different. FCH signature
is obtained from those regions, which makes the dimension of the signature equal
to 2*21=42 bins. The first 21 bins represents the upper part of the image and
second 21 bins describes the lower part of the image. By dividing image into two
halves, the FCH can easily differentiate between the same images in a different
order.

**Figure 3-11   Two Different Images which have Same Colour Histogram
Distribution**

For example, people can be confused by the similarity between the flags of Chad and Romania, but our technique easily distinguishes the difference between the dark and light blues in the two flags, see Figure 3-12.

Romanian Flag   Chadian Flag

Figure 3-12   Proposed FCH Technique Recognises the Difference between Romanian Flag and Chadian Flag

The histogram (probability density function) of the values within an image (either colour or grey-scale) is largely used for content-based image retrieval. The retrieval means that the images within the databases are selected according to the resemblance of their histogram to the histogram of the query image.

FCH is constructed from fuzzy rules represented by IF - THEN constructions that have the general forms of "IF A THEN B ", where A and B are (collections of) propositions containing linguistic variables. A is called the premise and B is the consequence of the rule. The use of linguistic variables and fuzzy IF - THEN rules exploits the tolerance for imprecision and uncertainty (Krishnapuram, Medasani et al. 2004). We construct the FCH from the fuzzy rules as follows: for i=0,…,5, j=0,..2

$IF\ x\ is\ A_i\ and\ y\ is\ B_j\ THEN\ D_{i,j}$

Where the x and y, and A_i , B_j , and $D_{i,j}$ are linguistic variables.

3.5.2 Subsets Centres (FCH)

The output of the premise part is $\min(\mu(x,A_i), \mu(y, B_j))$ while the consequence part represents the increase of the FCH bin which corresponds to the If-Then rule. Every subset has a centre where its membership is equal to 1.

The HSV spectrums are divided into several centres as follows:

- **Hue Centre**

The Hue value is represented by 6 subsets as shown in Figure 3-13. The centres of the subsets are defined empirically by the following values: {0, 60.0, 120.0, 180.0, 240.0, 300.0 and 360.0}. Any hue value will activate only two subsets. The best results are given when 6 membership functions are selected. The idea behind this design comes from the Hue definition which is considered as the angular value around the colour circle at the top of the cone and each 60⁰ the colour change and there is new colour. Each value of the Hue is multiplied by the saturation and then multiplied by the brightness value.
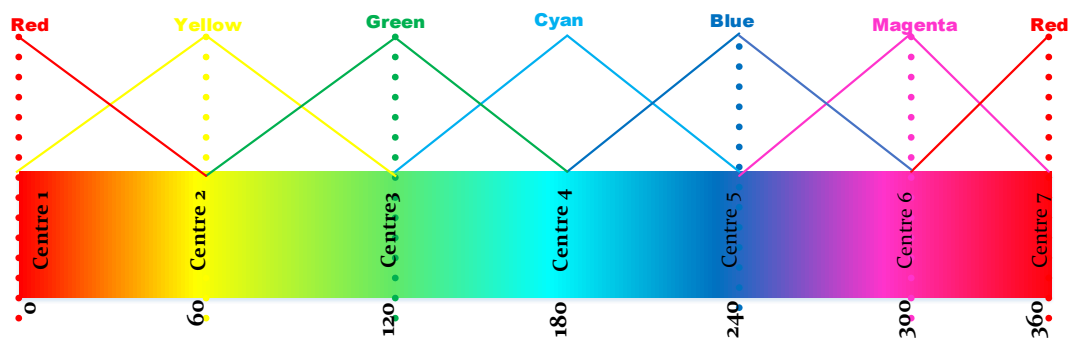
Figure 3-13   Hue Fuzzy Subset Centres

- **Saturation**

Saturation represents the strength of the colour. The highest saturation of 255 will correspond to membership of 1, while 0 saturation correspond to membership of

0. The higher the saturation, the higher the membership value. The first subset ranges from 0 to 128, the second is 0-255, and the third is 128-255. Each saturation value will activate two subsets. Figure 3-14 shows as an example the saturation of the colour red.
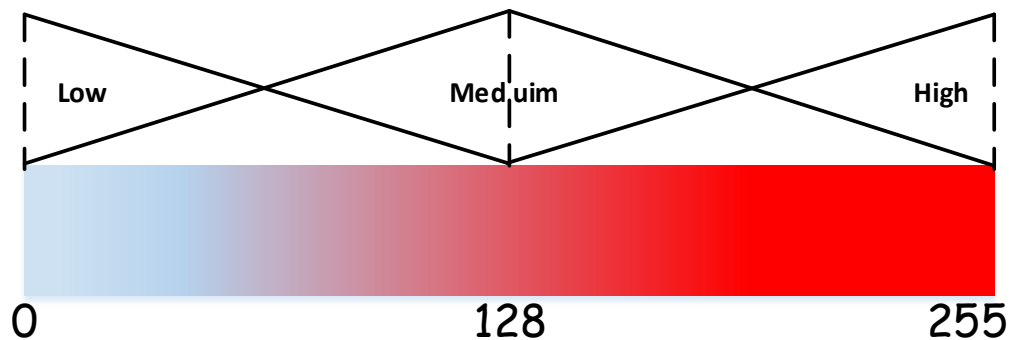
Figure 3-14   Saturation of RED Colour

Each Hue value is multiplied by the saturation. Values not activated will be assigned zero and will not be affected by the saturation.

- **Brightness Value**

It works in conjunction with saturation and describes the brightness or intensity of the colour from 0% to 100%.The brightness value is amount of brightness of one colour, see Figure 3-15.
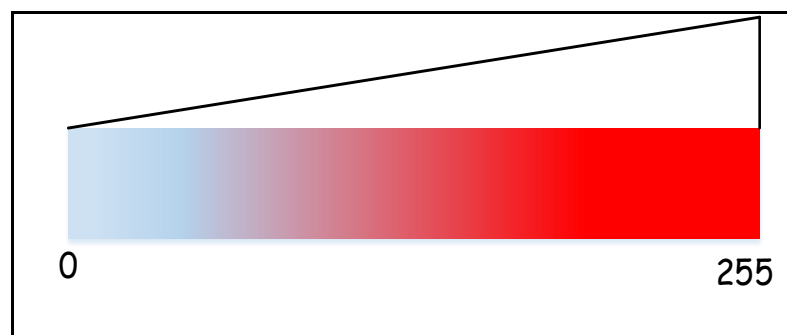
Figure 3-15   Brightness Value Fuzzy Subsets of RED Colour

- **Grey Scale**

The colours black, grey and white are represented by 3 separate bins since they are not included in the hue bins and the value of hue is undefined when colours are equal, see Figure 3-16.
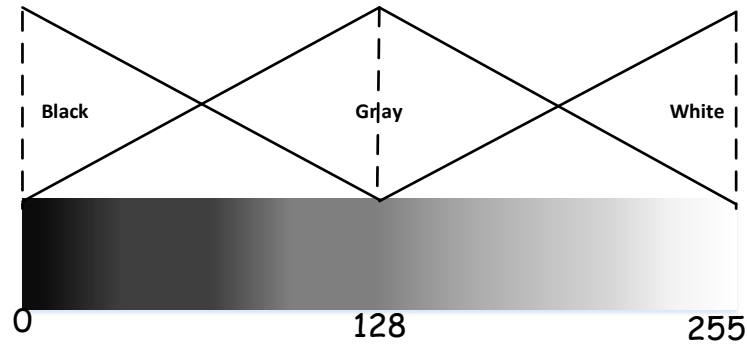
Figure 3-16   Representation Grey Level when R=G=B

3.5.3 Membership Function for FCH

Consider bin number $((i*6)*(j*3))$. HSV value will activate the triangular membership rule. For example, from, the value H=105 will belong to both the yellow and green sets. The membership of H in both yellow and green sets computed by Eq. 3-4. Green has its centre at 120U and yellow has its centre at 60, so μYellow(H)=0.25 and μGreen(H)=0.75. The saturation value is 169 which has membership in the medium to high range. Its membership is μmedium(S) =0.32 and μhigh(S) =0.68.

Six rules will be activated. If Hue is green and Saturation is medium, the bin representing green Hue and medium Saturation is incremented by an amount equal to:

$$\min\big(\mu_{Green\ Hue}(x), \mu_{Medium\ Saturation}(x)\big).$$

Since any value could activate up to 6 rules, all bins corresponding to activated rules are incremented the same way. If Hue is undefined, meaning R=G=B, the 3 bins of the FCH are incremented by the memberships of the grey values.

- **Distance Measures**

Pixels p, q and z have coordinates $(x, y), (s, t)$, and (v, w), respectively. D is a distance function or metric:

$(a)\ D(p, q) \geq 0$ $\qquad\qquad\qquad\qquad (D(p, q) = 0\ if\ (p = q))$,

$(b)\ D(p, q) \leq\ D(p, q),\quad$ and

$(c)\ D(p, z) \leq\ D(p, q) + D(q, z)h.$

The Euclidean distance (Amato and Di Lecce 2008) between p and q is defined as Eq. 3-8:

EUCLIDEAN DISTANCE

$$D(p, q) = \sqrt{(x - s)^2 + (y - t)^2}$$ **Equation 3-8**

For this distance measure, the pixels having a distance less than or equal to some values r from (x, y) are the points contained in a disk of radius r centered at (x, y).

The D4 distance (also called city-block distance) between p and q is defined as

CITY BLOCK DISTANCE

$$D4(p, q) = |x - s| + |y - t|$$ **Equation 3-9**

In this case, the pixels having a D4 distance less than or equal to some value r, form a diamond shape centred at (x, y).

The pixels having a D8 distance (also called chessboard distance) between p and q is defined as:

CHESSBOARD DISTANCE

$$D8(p,q) = Max(|x-s|, |y-t|)$$

- **Discussion of FCH Results**

The FCH has been tested in many experiments (see Chapter 6). It is concluded that the retrieved results for the colour-based CBIR were improved by using the proposed colour-based fuzzy algorithm and by applying fuzzy logic theory. The processing time is reduced significantly and accuracy of retrieval can be improved. The results demonstrate the efficiency and effectiveness of the FCH.

Chapter 4.   Shape-Oriented CBIR

4.1 Introduction

For huge databases, colour searching is inadequate for the sequential matching of images with a query image, due to the lack of efficiency and effectiveness in retrieval. Retrieval should be fast and accurate as possible and by matching many similar features, efficient retrieval cannot be achieved (Tangelder and Veltkamp 2004).

- **Shape Matching Methods**

To compare two shapes represented by two set of points (p, q), the problem is to determine their similarity. The measure of similarity depends on distance measures between the shapes. The less distance between shapes, the more similar they are.

Shape matching can be classified into two types (Mori, Belongie et al. 2001): First, feature based techniques: these take into account only the geometry of the shape. Here the first task is to measure the features and compare them using discrimination power (Cheikh, Cramariuc et al. 2003) to distinguish between shapes. Second, graph based techniques capture geometric information from a 3D shape using a graph to show how the shapes are related to each other.

- **Points**

There are two techniques to detect feature points. Firstly, detecting the points using local search methods such as least squares. Secondly, to find features in all the images in a database and later a matching processing can be applied in respect of their local appearance.

- **Corner detection**

Corner detection is defined as a method using pattern recognition techniques to seek certain types of features within objects such as an interest point distinguished from its neighbours by the abrupt change in direction of the boundary. Corner detection is important type of segmentation due to its ability to distinguish two objects, or mapping an object's pixels, and to cope with the correspondence problem between two images. Many application can use the corner detection approach, including motion detection, medical image matching of evolving diseases and palm print recognition.

4.2 Shape Formation

4.2.1 Shape Representation

Shape is a silhouette of the object with respect to its TRS (translation, rotation and scaling). The challenge facing CBIR using shape as a main feature is inhomogeneity which means non-uniform changes of object shapes. Solving the problem of inhomogeneity requires segmentation. A query image of simple shape such as a square, circle or triangle is easy to describe by examples or drawn sketches. Any shape-based CBIR consists of three main phases: shape description, shape similarity and shape retrieval (Emmanuel, Babu et al. 2007).

There are two techniques used to classify shapes: boundary-based and region-based, and these will be discussed in more detail in Sections 4.6 and 4.7. The classification of shape techniques is show Figure 4-1.
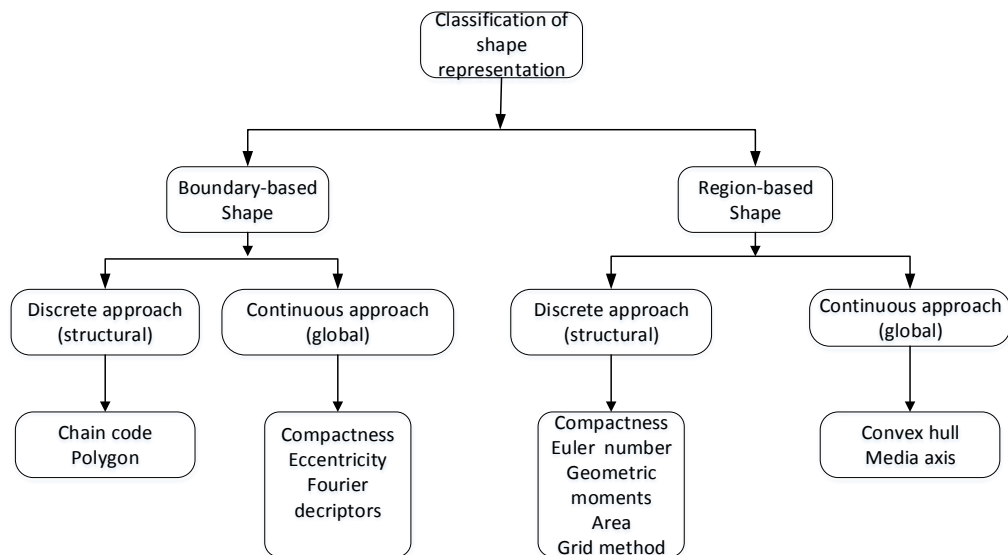
Figure 4-1 The Classification of Shape Techniques

For example, how to represent the shape of the original image in Figure 4-2– a butterfly in a green field – where our interest is solely on the main object?. The next image shows the border of the main object and a threshold process has been applied to reduce the grey value to just two colours either white or black. The image is completely transferred from its grey level to a binary image and the inner object is shown as white and the background of the object shown as black.

Original Image   Interested Object   Binary Image
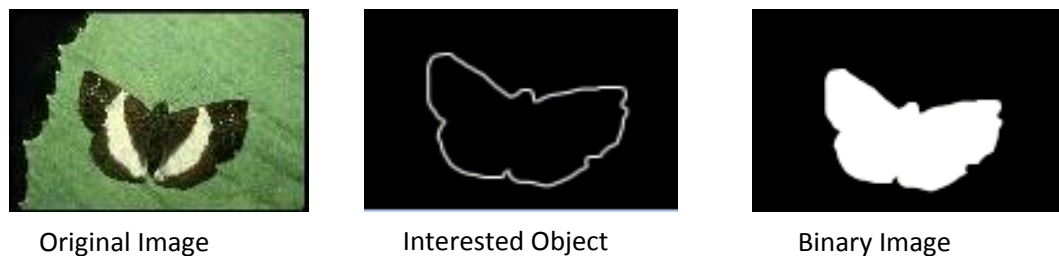
Figure 4-2 Example of Shape Detection by Converting an Original Image into Binary Image

- **Binary Image Analysis**

The binary image can describe the silhouette of an image or outline of an object of interest in terms of the values 1 or 0, where 1 denotes the object (white) and 0 describes its background (black).

To obtain these two values an image should be converted from grey levels into a binary image. This transform should be performed using a threshold processing method. A threshold operation assigns all the numbers above the specified average grey level the value 1, which means the maxima of the grey levels are represented by white pixels in binary images. All those values below the specified level are assigned 0 and represented by black pixels. The threshold approach is considered as the simplest operation to reduce the grey level values. Other methods to reduce the grey level values include half-toning, uniform bin with quantisation and false contouring. The binary image is useful in some applications such as object tracking or character recognition but it is limited by the small amount of information in the binary images when the original can have enormous content (Bin, Ramsey et al. 2000). This leads to possible ambiguity with shape-feature based CBIR which needs some pre-processing before transfer to binaries.

.

4.2.2 Shape Analysis

Shape analysis is the automatic analysis of shape geometries, for instance using a computer program to detect objects in a database and check for similarities between the shapes. To complete such analysis the objects must be described in digital forms. The boundary description or representation is the most common approach and is used to describe an object by its boundary. Then the data representing the object has to be simplified before matching process can be

accomplished to reduce the unnecessary information. The simplified description or shape descriptor is the process performed on the image content to retain the most important information, and this is the description stored ready to use for direct comparison with other shapes. Shape has certain advantages over colour and texture because shape is invariant to translation, rotation and scaling. The shape analysis process, see Figure 4-3, can be divided into three levels. First, shape pre-processing is applied to remove noise and unnecessary information. Second, data reduction reduces information by using a segmentation technique to determine regions which can describe the objects or the parts of interest. Third, signatures and features can be extracted for shape analysing and evaluation in real world applications. Shape analysis is a vital process in assessing the results. Feedback and repetition of previous stages is continued until acceptable results are obtained.
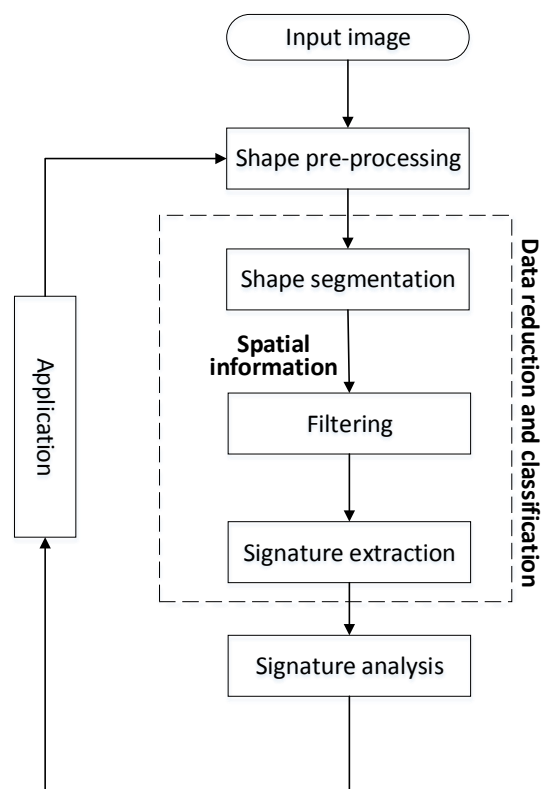
Figure 4-3   Shape Analysis Pipeline

4.3 Flexible Shape Extraction

4.3.1 Landmark Points

(Teh and Chin 1989) described landmarks as a finite set of values or points to define a landmarks scheme within the objects. The nodes of a polygon represents landmarks and has been used to find corresponding points on different images. After the matching is acheived, these corresponding landmark points are used to compute the probability of tranformation vectors (Umbaugh 2010).

4.3.2 Polygon Shape Descriptor

A polygon is basically a closed plane figure whose perimeter is a set of straight lines. Points where the straight lines meet are called corners.

The best way to represent the contour of the shape is called polygonal approximation approaches. The advantages of polygonal approximation are (Yang, Kpalma et al. 2008):

- To overcome the problem of noise.
- Polygonal approximation is preferable because the corner on a contour does not change its orientation after the polygonal approximation is finalised.

4.3.3 Dominant Points in Shape Description

Dominant points are defined as the points which retain important image information about the shape or objects of interest. To determine a dominant point requires finding the curve and to find the interest point which is called curvature.

An interest point is the point where the direction of the border of an object changes drastically. An interest point would normally be a corner or an intersection point between lines or edge segmentations. The importance of interest points, such as a corner, comes because it is relatively easy to distinguish between two images. Also interest points are robust because they do not suffer from aperture or ambiguity problems and it is relatively easy to find a unique match.

To detect the curvature of a curve there are two methods (Teh and Chin 1989):

- Direct detection and this is performed by detecting angles or corners.
- Determine polygonal approximation of the curve to detect a set of points on the query image. Then those detected points are compared to the intersections of the adjacent polygon.

4.3.4 Active Contour Model Approaches

(Kass, Witkin et al. 1988) presented the active contour or snake model for shape analysis. The active contour model is a means for detecting an object contour from a 2D image even in presence of noise. There are two components of snakes: the first component allows the snake to keep the original shape as is, to detect a corner. The second component determines where it goes on the image. This model enables the snake to minimise the energy related to the active contour as the fusion of an internal and external contour energy. When the position of the snake is at the boundary of an object, the external contour energy should be minimal.

- **Deformable Template**

The deformable template is one of the important techniques used in object estimation. This technique uses prior knowledge about the object. This prior knowledge can be described as a binary shape template or sketch. Then, later, prior information is encoded as edge information or parameter vector.

4.4 Segmentation

The purpose of segmentation is to divide a digital image and group the pixels into more meaningful information and this can be done by assigning all similar regions with the same label. Dividing an image into its uniform regions is an essential operation and there are many techniques for this. The most common technique is histogram and threshold. The histogram is based on the probability function to represent the signature of the image. For example, an image may have two levels of grey where the object is white and its background is black. To extract the object the image is divided into meaningful regions. The histogram will show a valley in between two peaks (Maria and Panagiota 1999). Thresholding is considered as one segmentation method. This method depends on the threshold values to convert a grey level image into a binary image. The selection of threshold values is crucial especially when there are multiple grey levels.

4.4.1 Concept of Segmentation

Image segmentation is the operation of seeking those regions that have the same kind of homogeneity. The purpose of image segmentation is to seek regions that describe objects or provide interested information about objects

- **Growing and Shrinking Regions**

Regions growing and shrinking is the most important approach to image segmentation. The technique depends on dividing the image into small regions operating on the pixels or the rows and columns considered to be the image space. The growth or shrinkage of a region is determined by the split and merge method. This method works by dividing the regions that do not pass the homogeneity criterion (neighbouring pixels which have similar attributes) and this leads to regions growing, and fusion of those regions. The homogeneity test is where neighbouring pixels are examined to determine whether they are similar to the initial pixel. If they are the region can be merged. However, if neighbouring pixels are different from the initial pixel then the region can be split.

4.4.2 Edge and Line Detection

The edges can be recognised as boundaries where there is a difference between regions of colours and/or textures. Although detecting edges can readily recognise regions and objects, it is a technically difficult task to segment an image into coherent regions. To determine the boundaries of an object in the image, the first step is to examine each pixel and compare it with its neighbours to determine whether this pixel value is located on a border of the object.

After this determination, the outline of the object will be found by the point values allocated to the borders named as edge points. The detection of the edge is performed by examining each pixel value and checking the transition of the grey level. If there is an abrupt change in the grey level, an edge can be defined.

- **First-order edge detection**

The first-order edge detection method is based on a differentiation operation that denotes changes in image intensity. It can highlight image variations and contrast, detecting boundaries of features within an image. First-order algorithms are often used in edge-detection including the Roberts cross, Sobel and Canny filters (Cherri and Karim 1989).

For example, the Roberts Cross filter seeks the boundary on the principle of gradient changes that can be obtained through any pair of orthogonal differences. This operator provides valuable results in vertical and horizontal positioning but is sensitive to noise.

The Sobel differential operator contains two masks to identify an edge in the vector form. Compared to other edge detecting operators, Sobel has two distinctive advantages (Wenshuo, Xiaoguang et al. 2010). Firstly, the extracted edges are explicit two columns/rows based differentiation; secondly, it has an inherent smoothing effect and is more robust to random noise.

The Canny algorithm is an optimal edge detector based on a specific mathematical model involving four essential processes (Umbaugh 2010). It starts with smoothing an input image by applying a Gaussian filter; then the filtering window scans direction and magnitude of the gradient. Non-maxima suppression is then applied to thin the marked edges before a threshold is applied to connect the pixels denoting edge points.

4.5 Shape Feature Extraction

The goal of shape feature extraction is to create feature vectors or signatures consisting of all meaninful information about the shape of an image. To achieve good results in any shape-based CBIR system, selecting the optimium signature of the feature vector is vital.

The shape techniques should be reliable and robust, we need to find shapes whether they are light or dark. The most important factor is to find the position of the object or detect its translation. Finally, the required parameter must find an object even when it is rotated (Nixon and Aguado 2008).

4.5.1 Introduction to Shape Descriptors

Although shape is one of the most powerful features in digital image processing and in CBIR systems, extracting the shape feature is considered the most difficult task because of their ambiguous geometrical and semantic descriptions.

The shape descriptor can be categorised into three parts

- Congruency: relies on invariance with transformation, i.e. shapes can be rotated, translated and scaled.

- Rinsic: is invariant with respect to isometric changes.

- Graph based: extracts geometric information and simplifies the shape but it is not easy to compare like shape vectors of shape signatures.

There are two basic processing steps in shape-based CBIR: feature extraction and similarity measurement. (Schlosser and Beichel 2009) have demonstrated that by introducing shape features, a powerful new feature dimension can be

introduced. Studies (El-ghazal, Basir et al. 2008) have confirmed that users prefer shape-oriented CBIR rather than texture or colour based CBIR, but that shape processing is often more difficult than colour or texture-based operation.

4.5.2 Shape Signatures

Shape signatures have been designed as a 1-dimensional model that represent 2-dimensional boundaries, often representing a unique object or shape and extracting the perceptual shape-features. They can be either real or complex.

4.6 Boundary-Based Shape Descriptors

Contour extraction is an important shape-detection approach (Celebi and Aslandogan 2005). Boundary (or contour) based shape descriptors ignore interior uniform characteristics and only focus on the boundary features. Fourier, Wavelet, and Curvature Scale Space descriptors are examples of the contour-based approaches. The computational complexity associated with these approaches is relatively low because only boundary pixels are required for computation; Figure 4-4 shows pixel-based outer and inner contour representations for an object boundary. As seen from the figure, contour may go through same pixel many times and from different orientation. The sequence means three circles has three neighbour contour and two circles has two. Even region which has just one pixel will also have contour.

4.6.1 Simple Global Descriptor (SGDs)

The shape boundary can be represented using scalar values that are based on their simple geometrical features.
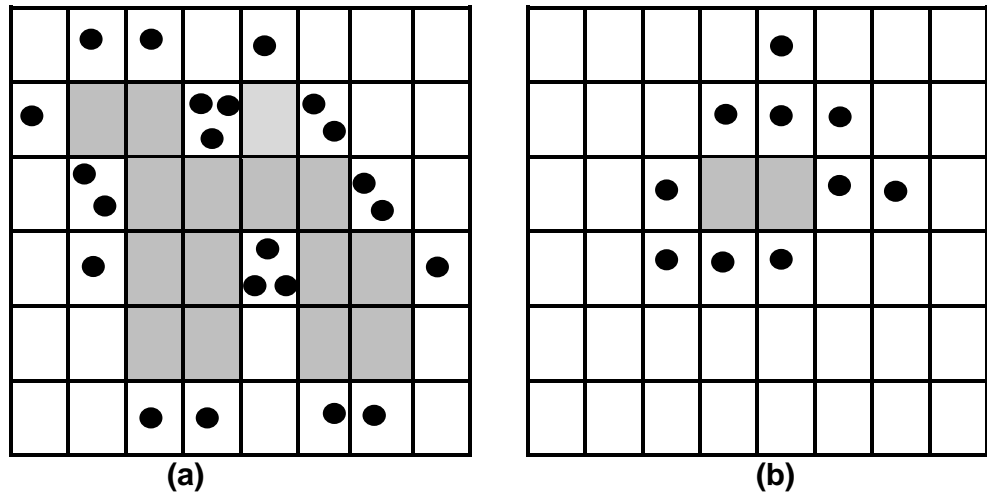
Figure 4-4 Pixel-based Boundary Representations a) Outer contour; b) Inner contour

- **Convexity (CX)**

Convexity is the ratio of perimeter of the convex hull to the original contour (p_0). Where the convex hull is the minimal convex covering of the object. Here it might be considered to be an elastic string stretched around the contour of the set of S points. The convex hull is the smallest convex set polygon that contains all S points. The convexity of a polygon containing S points is said to be convex if S is non-intersecting (Peura and Iivarinen 1997; Brlek, Lachaud et al. 2009). For example, two shapes used for defining the convexity and non-convexity are shown in Figure 4-5. The convexity is defined as (Jamil, Bakar et al. 2006).

CONVEXITY

$$CX = \frac{Convex\ Perimeter}{P_0}$$

Equation 4-1

Where the p_0 denotes polygonal original boundary
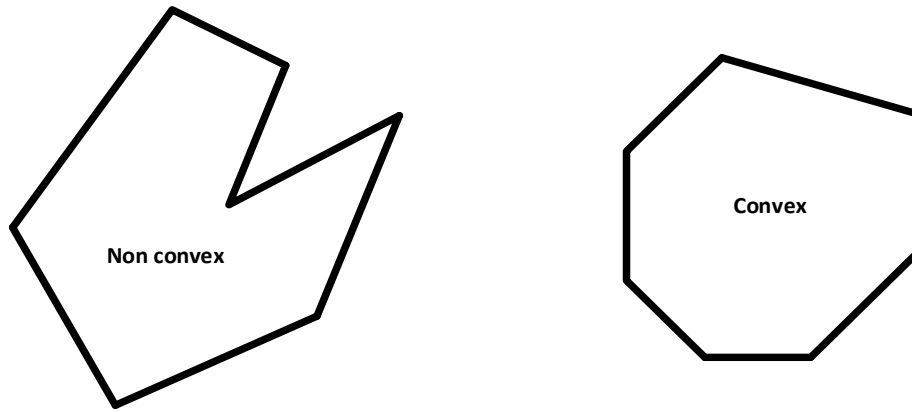
Figure 4-5   Examples of Convexity and Non-convexity

Examples of the convexity of four shapes are shown in Figure 4-6.

CX=0.8733   CX=1   CX=1   CX=1
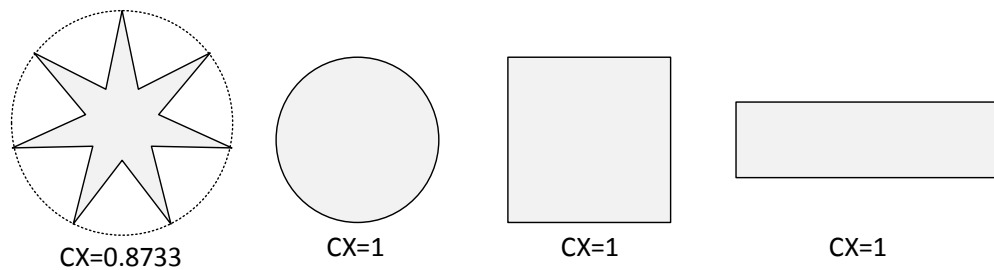
Figure 4-6   Examples of Shape Convexities

- **Bending Energy (BE)**

Bending energy is the quantity of energy necessarily to convert a closed boundary into a circle where the circle has the same perimeter as the original contour.

BENDING ENERGY

$$BE = \frac{1}{N} \sum_{n=0}^{N\_1} k(n)^2$$
                                                                Equation 4-2

Where N represents the number of points of the contour, and K denotes the contour curvature.

- **Aspect Ratio (AR)**

Aspect ratio is an important global descriptor and is used in conjunction with CSS descriptors (see Section 4.6.3) to define a closed contour of a two dimensional region in an image (Eakins, Riley et al. 2003). Aspect ratio is the height of the shape or the length along a major axis divided by its width along a minor axis, as:

ASPECT RATIO

$$AR = \frac{Height\ of\ the\ shape}{Width\ of\ the\ shape} \qquad \textbf{Equation 4-3}$$

4.6.2 Fourier Descriptor (FD)

The Fourier transform is considered as the most commonly used of all boundary object descriptors. Fourier descriptors can be defined as a set of complex numbers generated by a Fourier transformation of the contour function. The Fourier descriptors are formed for a complex contour using the discrete Fourier transform (DFT) (Kunttu, Lepisto et al. 2004). The discrete Fourier transform of a signature $z(k)$ is given as in Eq. 4-4.

FOURIER DESCRIPTOR

$$F(n) = \frac{1}{N} \sum_{k=0}^{N-1} z(k). e^{-j2\pi nk/N}, \quad n = 0,1,\ldots,N-1 \qquad \textbf{Equation 4-4}$$

4.6.3 Curvature Scale Space (CSS)

(Mokhtarian and Mackworth 1986) proposed the so-called curvature scale space (CSS), which defines a closed contour of a two dimensional region in an image. The representation of CSS is robust with regard to noise, scale and shift in orientation. A rotation of the image usually makes only a rotational change on its representation, which can be readily determined in the matching process.

Scaling hardly makes any change in representation due to length normalization. However, noise may cause some variations on the contours.

4.7 Region-Based Shape-Retrieval Descriptors

The region-based descriptors are another significant shape descriptor approach to represent shape features. Several studies such as (Lu and Sajjanhar 1999) (Kim and Kim 2000; Zhang and Lu 2004) have demonstrated that region-based descriptors are more effective than the contour based-descriptors. This is because it utilizes both the boundary and the interior pixels of regions to represent a shape and so is more robust to noise and distortions (Celebi and Aslandogan 2005). Empirically, region-based descriptor techniques outperform contour descriptors due to the fact that any scaling process effects only the local properties of a shape but maintain global visual features. Region shape descriptors comprise Geometrical moment, Legendre moment (LM), Zernike moment (ZM), and Pseudo-Zernike moment (PZM) that can be applied to generic shapes for describing their features. The following sections introduce commonly used region-based shape descriptors.

4.7.1 Simple Global Descriptors (SGDs)

Scalar measures can be used to describe a particular region of a shape based on simple geometrical features. To distinguish between shapes, simple global descriptors can be used effectively only between shapes with large dissimilarities. The simple global descriptors are often applied in image retrieval systems as filters to avoid false hits or can be joined with other approaches to differentiate between shapes. The disadvantage of these descriptors is they are not appropriate as standalone shape features.

- **Eccentricity (E)**

Eccentricity (E) of an object can be described as the way in which the "points" on an object are spread around the centre of the image region. Eccentricity can be defined as the ratio of the major axis to the minor axis of the region. It is calculated using central moments, see Eq. 4-5:

ECCENTRICITY

$$E = \frac{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}{(\mu_{20} - \mu_{02})^2}$$

Equation 4-5

Where μ_{pq} are the central moments. Figure 4-7 shows several shapes with their eccentricities which can be used with represent OCR shape features. The Eccentricity is sensitive to elliptical shapes and good at representing moments based features.

| E=0.51 | E=1 | E=1 | E=1 |
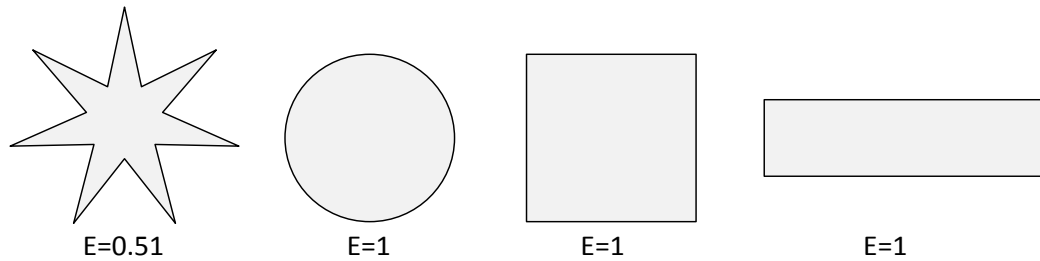
Figure 4-7 Examples of Shape Eccentricity

- **Solidity (S)**

Solidity can be defined as the ratio of the image object area, A_0, to the area of the convex hull (Jamil, Bakar et al. 2006), and calculated using Eq. 4-6:

SOLIDITY

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

Equation 4-6

Figure 4-8 shows various shapes and their solidities, compare with Eccentricity,Figure 4-7.

| S=0.7459 | S=1 | S=1 | S=1 |
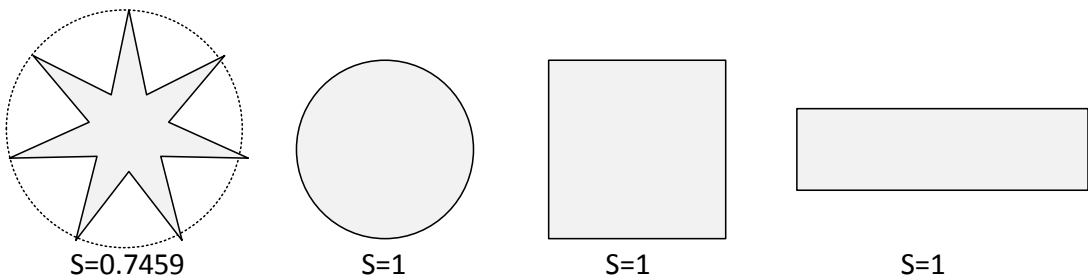
Figure 4-8 Examples of Solidity of Shapes

- **Rectangularity (Extent) (EX)**

The extent or rectangularity of an object can be described as how much of its shape fills its minimum enclosing rectangle (MER). Rectangularity thus has a value of 1 for a rectangular shape. Rectangularity is defined as:

RECTANGULARITY

$$EX = \frac{A_0}{A_{MER}}$$ **Equation 4-7**

Where A_0 is the area of the object, and A_{MER} is the area of the object's MER.

Figure 4-9 shows various shapes and their rectangularity.

| EX=0.51 | EX=0.785 | EX=1 | EX=1 |
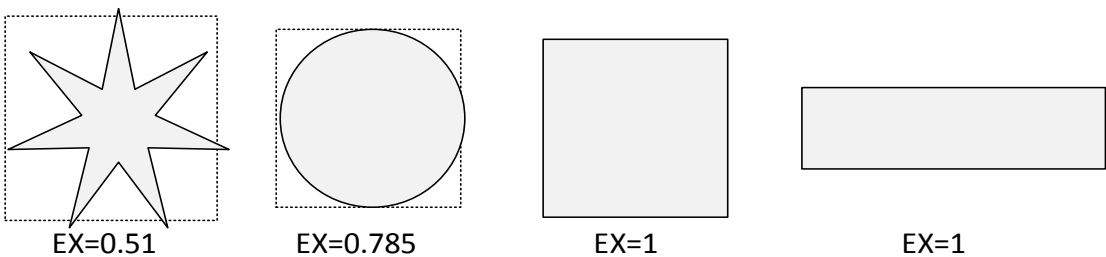
Figure 4-9 Examples of Rectangularity

4.7.2 Invariant Moments

Moments define the pixel arrangement or shape layout, and are considered as a global description of the shape. There are various popular moment techniques such as Zernike Moments (ZMs), Hu Moments and Legendre Moments (LMs). The invariant features describe the objects by a set of measurable scatter quantities. Invariance describes insensitivity to deformation and the capability to provide sufficient discrimination power to differentiate objects in different classes (Flusser, Suk et al. 2009).

4.7.3 Hu Moments

Hu derived seven common invariants of aspects (Ming-Kuei 1962). Shown below are the second-order moments $\emptyset_1$ and $\emptyset_2$. The remaining $\emptyset_3$, $\emptyset_4$, $\emptyset_5$ and $\emptyset_6$ are of third-order, where p and q provide the order of the moment. The last moment, $\emptyset_7$, is represented as a skew invariant implemented to distinguish mirror images, η_{pq} represents normalized central moments.

The well-known as Hu Moments are formed by the equations given below where η_{pq} were represents the normalised central moments. The normalisation process is done by $\eta_{pq\eta} = \frac{\mu_{pq}}{\mu_{00}{}^\lambda} \quad \lambda = 1 + \frac{p+q}{2}$.

HU MOMENTS

$$\emptyset_1 = \eta_{20+}\eta_{02}$$

$$\emptyset_2 = (\eta_{20-}\eta_{02})^2 + 4\eta_{11}^2$$

$$\emptyset_3 = (\eta_{30} - 3\,\eta_{12})^2 + (3\eta_{21-}\,\eta_{03})^2$$

$$\emptyset_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21+}\eta_{03})^2$$

$$\emptyset_5 = (\eta_{30} - 3\,\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21+}\eta_{03}$$

$$(3\eta_{21-}\eta_{03})(\eta_{21+}\eta_{03})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03}$$

Equation 4-8

$$\emptyset_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] +$$

$$4\eta_{11}^2\,(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\emptyset_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[\,(\eta_{30} + \eta_{12})^2$$

$$- 3(\eta_{21+}\eta_{03})^2\,] +$$

$$(3\eta_{21-}\eta_{03})(\eta_{21+}\eta_{03})[3(\eta_{30} + \eta_{12})^2 -$$

$$(\eta_{21} + \eta_{03})^2\,]$$

$$\mu_{pq} = \sum_x\sum_y(x - \bar{x})^p(y - \bar{y})^q\ f(x,y)\ \ p,q = 0,1,2,\dots \qquad \textbf{Equation 4-9}$$

Where $\bar{x} = m_{10}/m_{00}$ $\bar{y} = m_{01}/m_{00}$

m_{00} is the mass of the image or raw image moments include area of binary image or some grey level for grey level images. While m_{10} and m_{01} are the raw image moments with order 1,0 for m_{10} and order 0,1 for m_{01}.

$\bar{x}, \bar{y}$ define the components of centroid

$$m_{pq} = \sum_x\sum_y \psi_{pq}\,(x,y)f(x,y)\ \ p,q = 0,1,2,\dots$$

Equation 4-10

Where $\quad \psi_{pq} = x^p\,y^q$

4.7.4 Zernike Moments (ZMs)

Zernike Moment polynomials were first introduced in 1934 (Teague 1980) and have been considered as one of the most powerful techniques used in orthogonal

moments (Li, Moon-Chuen et al. 2009). The major justification for introducing orthogonal moments is the speed and stability of their numerical implementation. The ability of orthogonal moments to extract invariant features from scalar quantities (moments) depends on the theory of polynomials highlighted by Teague (Teague 1980). Image reconstruction from orthogonal moments can be achieved easily and the reconstruction is optimal because the orthogonal moments reduces the mean-square error.

ZERNIKE-RADIAL POLYNOMIAL, REAL VALUE

$$R_{nl}\left(r\right) = \sum_{s=0}^{(n-|l|)/2} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|l|}{2}-s\right)!\left(\frac{n-|l|}{2}-s\right)!} r^{n-2s} \qquad \text{Equation} \quad \textbf{4-11}$$

Where $R_{nl}\left(r\right)$ is a real-valued Zernike-radial polynomial and $n = 0,1,2,\dots, l = -n, -n+2, \dots, n$, the order of the polynomial is denoted by *n* and the repetition by *l*, the difference $n - |l|$ must be an even value.

For example if n =3 the repetition should be computed for three levels as follows:

n =3 repetition l= -3,-1,1,3

n =2 repetition l= -2,0,2

n =1 repetition l= -1,1

n =0 repetition repetition l= 0

The total of moments when n =3 is 10

The ZM polynomials are defined (Li, Moon-Chuen et al. 2009) as:

ZM POLYNOMIAL

Equation

4-12

$$Z_{nm} = \sum_{k=|l|,|l|+2,...}^{n} B_{nlk} \, r^k$$

Where the coefficients $B_{nlk} = \frac{(-1)^{(n-k)/2}((n+k)/2)!}{((n-k)\backslash 2)!((k+l)\backslash 2)!((k-l)\backslash 2)!} \, r^{n-2s}$

, and the l, k are the orders of repetition. Eq. 4-13 can be also defined as

ZM POLYNOMIAL

Equation

4-13

$$Z_{nm} = \frac{n+1}{\pi} \sum_{r} \sum_{\theta} f(r\cos\theta, r\sin\theta).R_{nm}(r).exp(jm\theta)$$

Where *n* is the radial magnitude, *m* is the radial direction and θ is the angle between the vector r and the x axis. There are many ways to convert from Cartesian coordinates to polar coordinates according to the particular application. Commonly, it is supposed that the centre of the rotation is at the centroid of the image. The r represents the length of the vector from the origin to the (x, y) pixel. If image has size Height (H) ×Width (w), r and θ can be defined as:

MAGNITUDE OF RADIAL VECTOR, r

Equation

4-14

$$r = \sqrt{(2*X - H + 1)^2 + (2*y - w + 1)^2} \, /H$$

ANGLE BETWEEN VECTOR AND X-AXIS

Equation

4-15

$$\theta = arctan\left(\frac{W - 1 - 2*y}{2*X - H + 1}\right)$$

4.7.5 Legendre Moments (LMs)

Legendre Moments belong to the class of orthogonal moments. The main advantage of the LM is that it retains orthogonality even on sampled images

(Flusser, Suk et al. 2009). The Legendre Moments of order (m, n), with image

intensity function $f(x, y)$, is defined as (Flusser, Suk et al. 2009) :

LEGENDRE MOMENT

$$L_{mn} = \frac{(2m+1)(2n+1)}{4MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} P_{m\left(-1+2\frac{x}{M-1}\right)} P_{n\left(-1+2\frac{y}{N-1}\right)}$$

Equation 4-16

Where x and y are the pixel locations. The moment order is denoted by m and n.

The Legendre polynomials are a complete orthogonal basis set defined over the

interval [-1,1]. For example: if the m=5, n=4 and we have image size $M \times N$=3×3.

Representation of the pixel of the image $\begin{bmatrix} f(0,0) & f(0,1) & f(0,2) \\ f(1,0) & f(1,1) & f(1,2) \\ f(2,0) & f(2,1) & f(2,2) \end{bmatrix}$ and to compute

$P_n(x)$ nth degree of Legendre polynomial or so-called Rodrigues formula is

defined by:

RODURIQUES FORMULA

$$P_n(x) = \sum_{p=0}^{n} a_{np} x^p$$

Equation 4-17

If the coefficients of the legend polynomials are denoted as a_{np}, then a_{np} can be

defined as:

LEGENDRE COEFFICIENTS

$$a_{np} = \begin{cases} (-1)^{(n-p)/2} \frac{1}{2^n} \frac{(n+p)!}{((n-p)/2)!((n+p)/2)!p!} & n-p \ even \\ 0 & n-p \ odd \end{cases}$$

Equation 4-18

4.7.6 Pseudo-Zernike Moments (PZMs)

PZMs were introduced by (Bhatia and Wolf 1954) and derived from the Zernike

Moments polynomial. PZMs take a form of a projection intensity function onto the

PZMs polynomial and those polynomials are described using a polar coordinate

of the image space. Thus, PZMs are commonly used in pattern recognition tasks that require rotational invariance.

PZMs were derived from the ZMs when the PZMs are released from the condition (n - 1) must be even. PZM has $(n+1)^2$ moments of order, whereas ZM has about half that number, only (n + 1)(n + 2)/2. Consequently, PZMs give more features than ZMs. PZMs and ZMs are both valuable moments for use in shape descriptors, but the PZM method outperforms ZM since the PZM has features such as rotation invariance and orthogonal moments (Hui, Zhifang et al. 2010).

4.7.7 PZM Descriptor Design

In general, moments are scalar values that represent a distribution of objects by associating those values to different powers. The number of moments will be computed using $(n + 1)^2$. The n is the order of PZMs polynomials. When n = 4, the number of PZM moments will be 25 and thus 25 bins represents the shape signature. Figure 4-10 shows the Pseudo-Zernike polynomial kernel for $n = 4$. The kernel can describe different levels of the shape and the 25 moments give feature vectors to represent the object as accurately as possible, for example the first icon *PZM₀₀* represents the conditions $n = 0$ and $m = 0$. The second row, commencing from left, represents the conditions the conditions $n = 1$ and $m = -1$, the centre icon $n = 1$ and $m = 0$ and icon on the right $n = 1$ and $m = 1$. Because $n = 4$ there are 4 levels plus *PZM₀₀*.
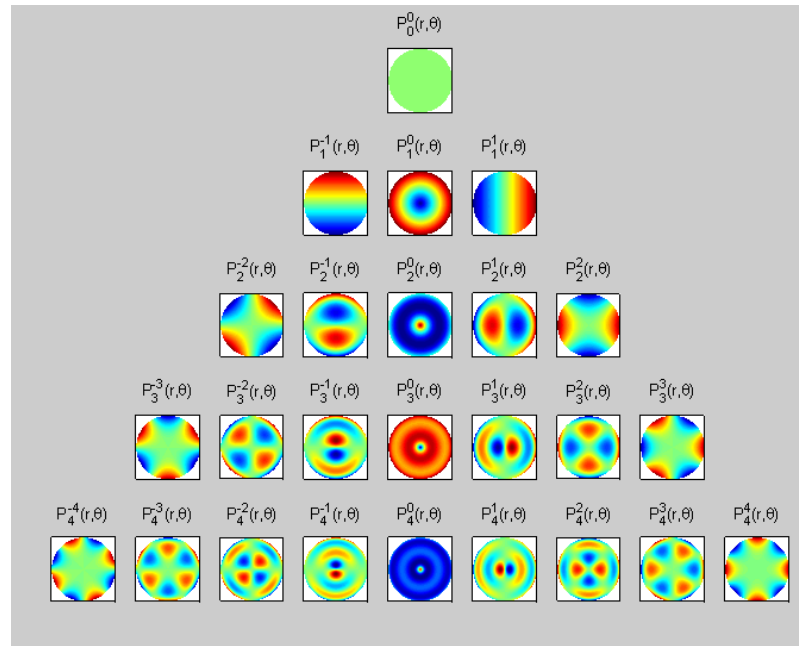
Figure 4-10   PZM Bases when n=4

Figure 4-11 shows the Pseudo-Zernike polynomial when $n = 8$ and the shape

signature is represented by 81 bins. The advantage gained when *n* is increased

is that there are more bins representing a greater number of possible curves. This

will allow a more detailed description and finer discrimination of two shapes.

However, the number of bins increases as the square of n and that can lead to a

huge number of moments and this will affect the time of processing the features.

Processing can be done adaptively by using relevance feedback. If the user is

not satisfied the value of n can be changed according the user's request.

For example, consider a shape illustrated as binary image as shown on the left in

Figure 4-12, extracted from the original colour image as shown on the right.Figure

4-13 illustrates the difference between the original image and the Pseudo-Zernike

Moments representation. The silhouette of the image is shown on the left of the

figure as a red shape on a blue background and it shows that the centre of the

image is the region of interest. The second image depicts the PZM polynomials

to represent the object in geometric aspect, to calculate angle and curve of the object borders.
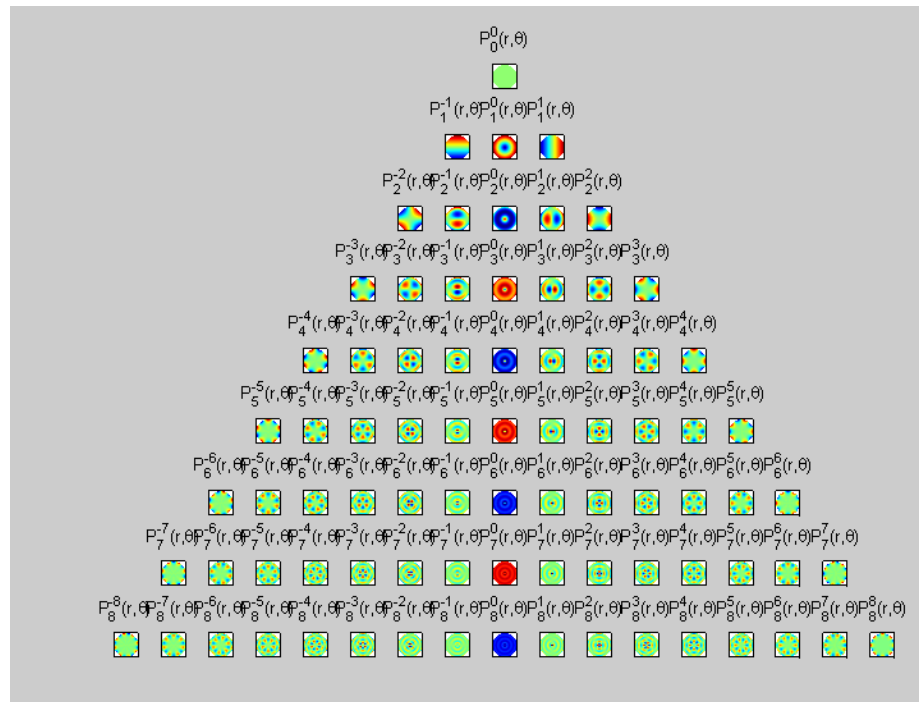
Figure 4-11   PZMs Bases when n=8

(a)   (b)

Figure 4-12   (a) Object binary image, (b) Original image as a colour image

The image on the right in Figure 4-13 shows the difference between the original image representation and the PZM representation.

The image on the right shows how the object is really represented in the good description and that clearly appears in experimental results. The yellow colour area represents the difference between the original image and PZM descriptor representation.
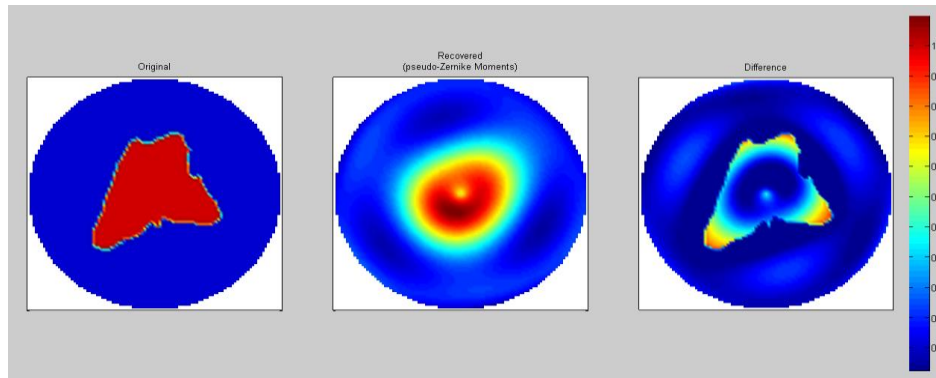
**Figure 4-13   Differences between Original Image Representation
and PZM Representation when n=4**

The experimental results collected and analysed in the research have demonstrated that the PZM is of superior performance to other geometric moments such as LM and ZM, especially when applied onto face recognition tasks due to its inherent merits when dealing with variant quality image databases (see section 6.3).

4.7.8 Moments-based Approaches and Their Pros-and-Cons

Moments represent a shape's layout.The main advantages of image moments are their abilities to identify the feature patterns and are robust to geometrical changes. The moments can be categorised into two main groups: statistical and mechanical moments. A mechanical moment represents the rate of change in momentum while the statistical moment represents the rate of change in shape feature characteristics.

There are various moment definitions such as Moment invariants, Geometrics Moments, Rotational Moments, Orthogonal Moments, Complex Moments, and Velocity moments. Moments Invariants are proposed to alleviate shortcoming of several geometric issues such as translation, rotation, scaling, affine changes and blurring. Due to these advantages invariants moments are widely used as

decimation property in pattern recognition applications. Describing images using moments instead of other more commonly used image features means that global properties of the image are used rather than local properties. The rotational moments rely on a polar coordinate description of the image, while orthogonal moments introduced by Teague describe the image from moments based on orthogonal polynomials. There are problems in the ZM identification, for example, geometric and approximation errors are more prominent (called finite precision errors). The velocity moments are often used in pattern recognition as tool to recognise moving objects. There are another type of moments called affine invariant moments. Their advantages are the robustness towards changes occurred in rotation, and scaling. The LM and Furier-melin have been widely studied to reduce the time needed to computer the polynomials basis. One drawback of moments is the so-called "overflow", which is a numerical instability problem appeared during computational stage

4.8 Evaluation of CBIR Based on Shape Features

The assessment and evaluation of shape-retrieval CBIR systems is a difficult task because of the lack of a standard database. The database most commonly used to evaluate the performance of shape-based CBIR retrieval is the MPEG-7 database (Manjunath, Salembier et al. 2002). This database has been designed to contain a variety of sets of shapes of a reasonable size and is commonly used as a database to check validation of shape CBIR systems.

MPEG-7 contains three categories, indexed by set A, set B and set C. Set A includes two subsets A1 and A2. Subset A1 consists of 420 shapes: 70 basic shapes and 5 shapes extracted from each basic shape by changing the scaling.

Subset A1 is used to test invariance to scaling. Figure 4-14 shows a sample image and its scaled image sequence.

Figure 4-14   Sample of Set A1 Used to Test Scaling

Subset A2 is applied to test rotation invariance. Set B is used to check retrieval performance. It consists of 1400 shapes classified into 70 categories each containing 20 similar shapes. Figure 4-15 shows examples of four image shapes, each with its own group.

Set C consists of 1100 images of sea fish and 200 more affine transformed. They are used to check robustness to distortion. Figure 4-16 shows samples of sea bream from group C. Figure 4-17 shows samples of sea marine fish from group C.
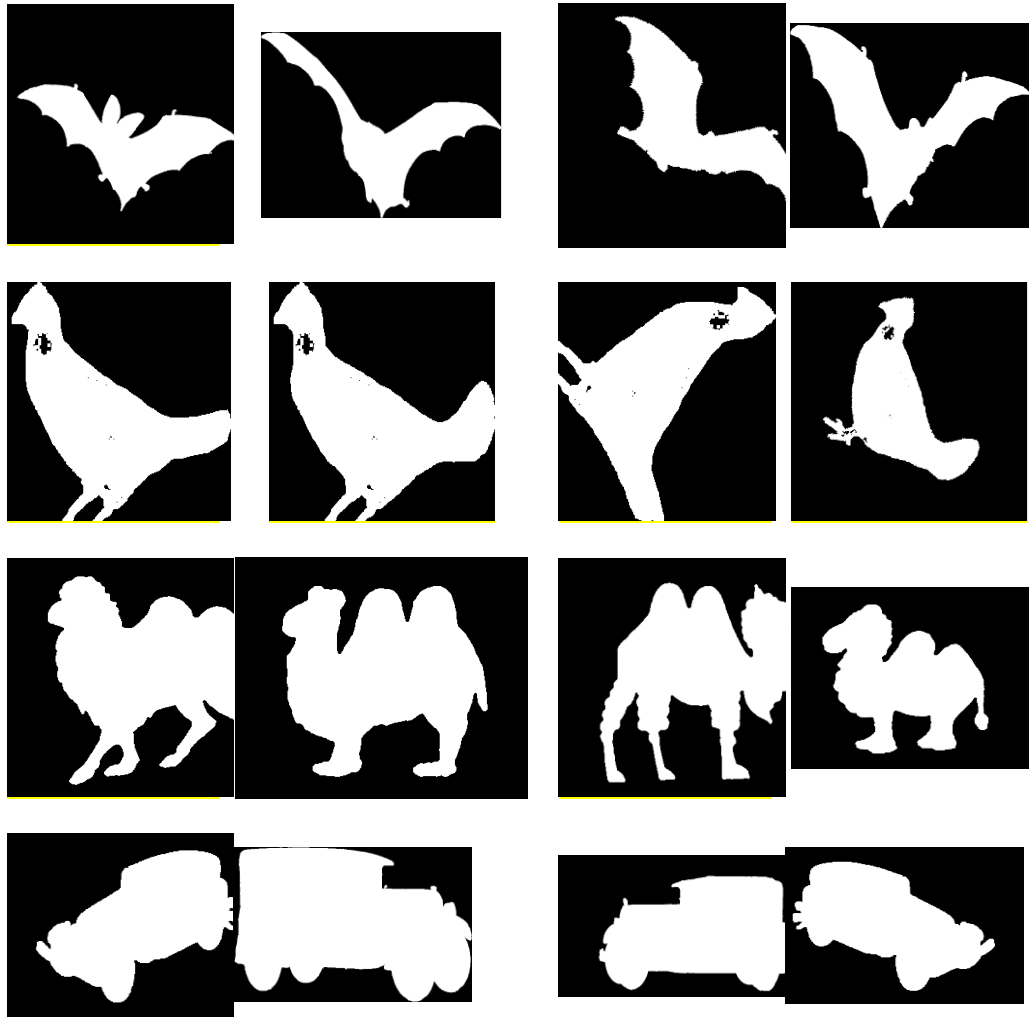
Figure 4-15   Sample Images from Set B of MPEG-7
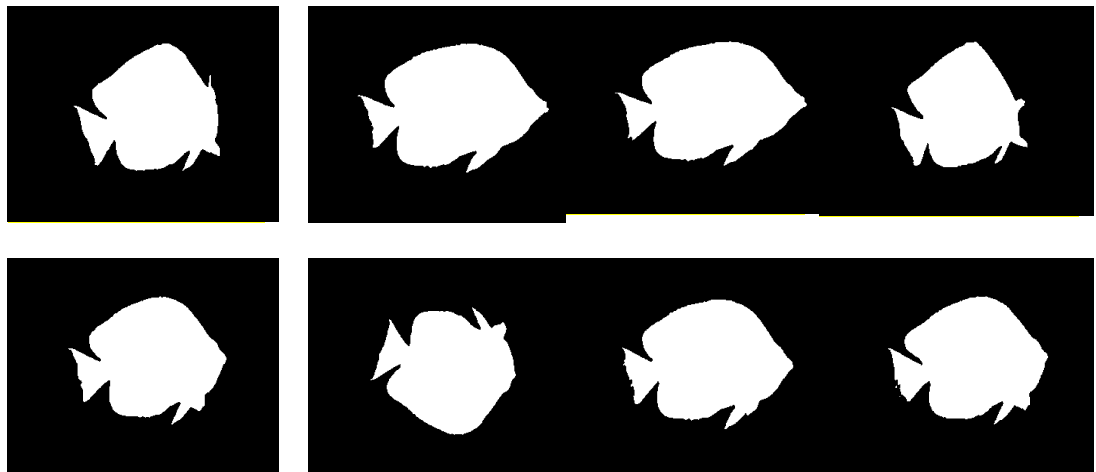
Figure 4-16   Samples of Sea Bream from Set C, First Group
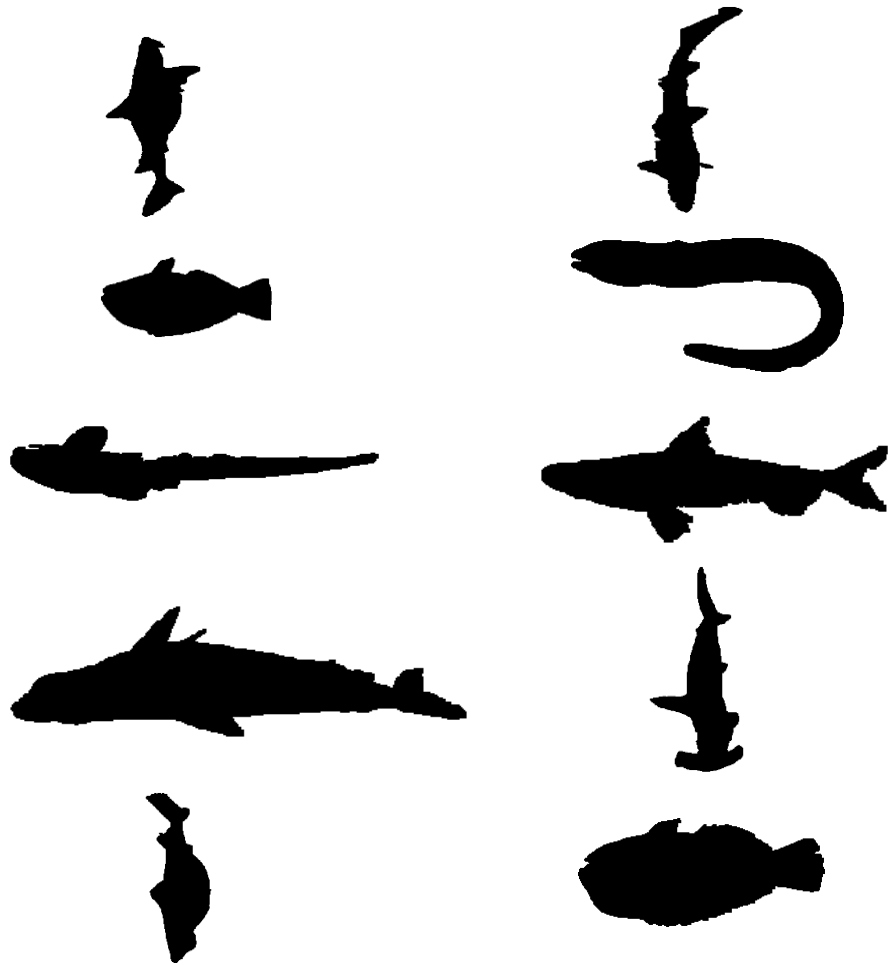
Figure 4-17   Samples of Sea Marine Fish from Set C, First Group

4.9 Image Processing for Local Shape

Local shape is the process of capturing the geometric properties of the image in detail. Local shape properties can be derived from colour derivatives (Smeulders, Worring et al. 2000). The local shape descriptor can be defined as the transformation from a small subset of the central surface at a point p to a vector, d. These vectors, or signatures, can then be matched and compared to decide how similar are the points which have been compared. Local shape descriptors and features are widely used specially in 3D shape matching to extract both distinctive points on the object's surface and then to link these points on diverse

models. They can also be applied to make the search process quicker to find matching objects by decreasing the model to a small vector of features which are readily compared. Good local descriptors should be invariant to geometric changes from rotation to translation and scaling. Consideration should be given to the local shape surface over a given point because the local shape descriptor can be used with 3D shapes to seek those unique points on the surface which match on different models. To speed up the search, the parameters within the models should be minimised.

Chapter 5. Fuzzy Fusion of Colour and Shape Signatures (FFCSS)

This chapter presents an evaluation of the main concepts and algorithms used in building the proposed FFCSS technique. FFCSS CBIR attempts to combine both colour- and shape-based techniques to achieve an efficient CBIR system to represent the image in a "high-tech" way so that the system includes all the main stages of image processing including query image, image representation and description, feature integration and weighting scheme. The goal behind using integration techniques is to overcome the limitations of a single feature type. The FFCSS method uses a combination of features which involve two different types of signatures and assign their weights automatically.The FFCSS method is a novel signature scheme which combines two different feature extractions: colour and shape. Combining these features makes for an extraction method more powerful than either separately.

5.1 Image Database

An image database is a collection of records containing images or a path indicating the location and the name of the required image. The size of an image database is measured by the number of records it contains. There are several important factors in image processing which effect the processing power requirements (Aggarwal, Ashwin et al. 2002). These include the image format and image size. The bigger the image size, the higher processing power required.

This research used several image sizes ranging from 128×85 to 128×96 pixels, it also used several types of images text files. One database contained flags of 224 different countries (Zagrouba, Ouni et al. 2007). Another database called Vary (El-Feghi, Aboasha et al. 2007) contains about 10,000 standard images of natural scenes. In addition, there was a third data base of images which were captured by the author. The format used here is Bitmap of 24-bits for all three databases. Our selection of this format is for several reasons, including its compatibility with most of image processing software, its simplicity and the fact that it is very widely used.

5.2 Prototype Pipeline

The FFCSS process starts with the submission of either the query image itself or sketch of the query image into the system. The system then extracts the colour features of the query image using the fuzzy logic scheme and extracts the shape signature by using orthogonal moments. Next the system integrates both colour and shape features into a unified vector. The next step, is to match similarities and compare from the extracted features of query image and the features of images in the database. Finally, the system retrieves the most similar images to the query image. The Prototype Pipline is illustrated by Figure 5-1.
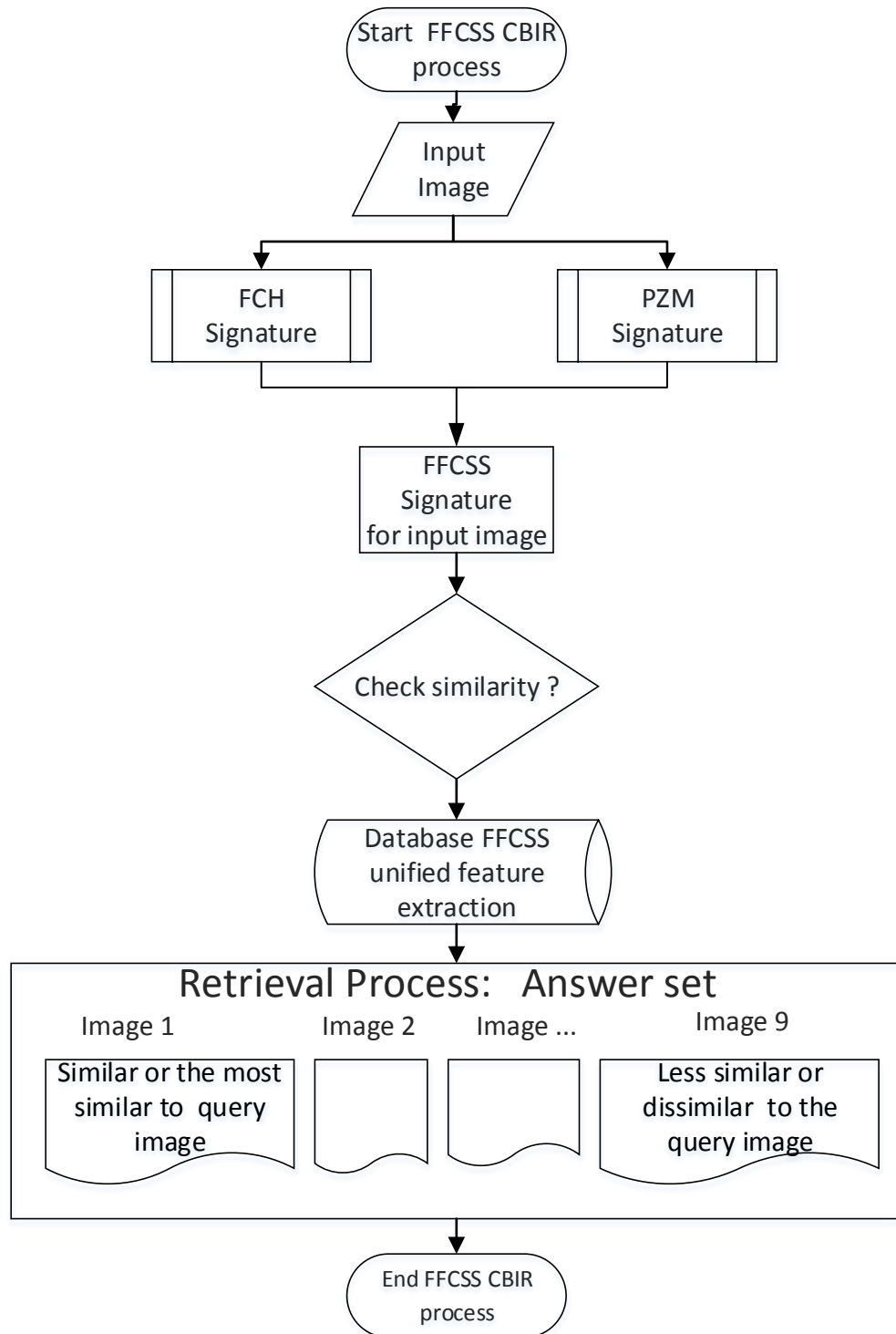
Figure 5-1 The Prototype Pipeline

5.3 Colour-Based CBIR Component

The image is partially represented by a vector of Part 1 of the image:

6 Hue × 3 Saturation × 1 Value + 3 x Grey scale= 21 bins, and partially by a vector of Part 2 of the image:

6 Hue × 3 Saturation × 1 Value + 3 Grey scale= 21 bins. The vector signature is thus a total of 42 bins when Parts 1 and 2 are combined, with each containing the accumulated values of the corresponding fuzzy rules. Figure 5-2 shows the signature store in the text file.
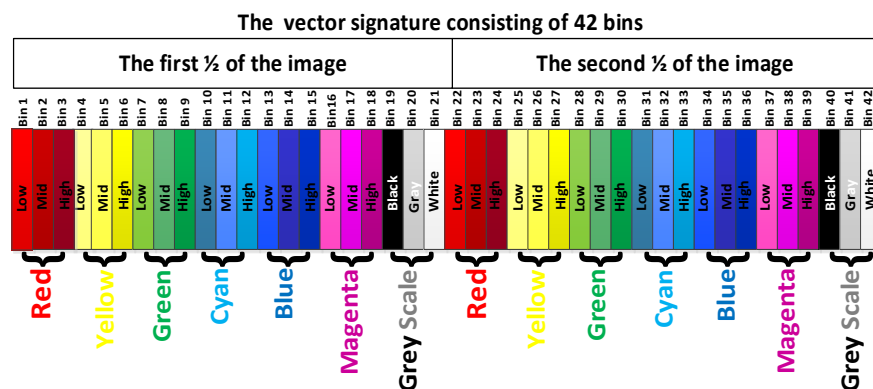
Figure 5-2   Representation of the FCH Signature

The size of the text files of a CCH for the Vary database, for all 10 000 images is about 20 Mb. However, the size of the text files for a FCH is only 5 Mb. Obviously, the total files for all images using CCH is four times more than for a FCH.

Table 5-1 shows the representation of the features of all 42 Bins.

Table 5-1 Representation the Features of all 42 Bins

This column represents the "first half" of the image	This column represents the "second half" of the image
Bin 1=Hue(**RED**)*Saturation(**LOW**)	**Bin 22**=Hue(**RED**)*Saturation(**LOW**)
Bin 2=Hue(**RED**)*Saturation(MID)	**Bin 23**=Hue(**RED**)*Saturation(MID)
Bin 3=Hue(**RED**)*Saturation(HIGH)	**Bin 24**=Hue(**RED**)*Saturation(HIGH)
Bin 4=Hue(**YELLOW**)*Saturation(LOW)	**Bin 25**=Hue(**YELLOW**)*Saturation(LOW)
Bin 5=Hue(**YELLOW**)*Saturation(MID)	**Bin 26**=Hue(**YELLOW**)*Saturation(LOW)
Bin 6=Hue(**YELLOW**)*Saturation(HIGH)	**Bin 27**=Hue(**YELLOW**)*Saturation(HIGH)
Bin 7=Hue(**GREEN**)*Saturation(LOW)	**Bin 28**=Hue(**GREEN**)*Saturation(LOW)
Bin 8=Hue(**GREEN**)*Saturation(MID)	**Bin 29**=Hue(**GREEN**)*Saturation(MID)
Bin 9=Hue(**GREEN**)*Saturation(HIGH)	**Bin 30**=Hue(**GREEN**)*Saturation(HIGH)
Bin 10=Hue(**CYAN**)*Saturation(LOW)	**Bin 31**=Hue(**CYAN**)*Saturation(LOW)
Bin 11=Hue(**CYAN**)*Saturation(MID)	**Bin 32**=Hue(**CYAN**)*Saturation(MID)
Bin 12=Hue(**CYAN**)*Saturation(HIGH)	**Bin 33**=Hue(**CYAN**)*Saturation(HIGH)
Bin 13=Hue(**BLUE**)*Saturation(LOW)	**Bin 34**=Hue(**BLUE**)*Saturation(LOW)
Bin 14=Hue(BLUE)*Saturation(MID)	**Bin 35**=Hue(**BLUE**)*Saturation(MID)
Bin 15=Hue(**BLUE**)*Saturation(HIGH)	**Bin 36**=Hue(**BLUE**)*Saturation(HIGH)
Bin16=Hue(MAGENTA)*Saturation(LOW)	**Bin 37**=Hue(**MAGENTA**)*Saturation(LOW)
Bin 17=Hue(**MAGENTA**)*Saturation(MID)	**Bin 38**=Hue(**MAGENTA**)*Saturation(MID)
Bin18=Hue(**MAGENTA**)*Saturation(HIGH)	**Bin39**=Hue(**MAGENTA**)* Saturation(HIGH)
Bin 19=RGB(**BLACK**)	Bin 40=RGB(BLACK)
Bin 20=RGB(GREY)	Bin 41=RGB(**GREY**)
Bin 21=RGB(WHITE)	Bin 42=RGB(**WHITE**)

5.4 Shape-Based CBIR Components

(Teh and Chin 1988) introduced a set of invariants for PZMs and presented a mathematical method for extracting the scale invariant function of PZMs based on use of its polynomials. They also showed that PZMs are less sensitive to image noise than the conventional ZMs, radial moments and geometric moments. However, the time complexity (processing time) of PZMs is greater than for ZMs (Chong, Raveendran et al. 2003). The PZMs are used throughout this project as a shape descriptor to extract the shape feature. The steps to extract the image features are as follows:

The complex moments for PZM descriptors of order n, with repetition l, for a continuous function in polar coordinates can be described (Hui, Zhifang et al. 2010) as:

COMPLEX MOMENT

$$PZM_{nl} = \frac{n+1}{\pi} \int_0^\pi \int_0^1 \widetilde{R_{nl}}.e^{-jq\theta}.f(r,\theta).drd\theta \qquad \textbf{Equation 5-1}$$

Where $\widetilde{R_{nl}}$ is the radial component.

RADIAL COMPONENT

$$\widetilde{R_{nl}}(x,y) = \sum_{s=0}^{n-|l|} S_{nls}(x^2+y^2)^{n-s/2} \qquad \textbf{Equation 5-2}$$

Where the S_{nls} is the coefficients and computed as follows:

The coefficients

$$S_{nls} = \frac{(-1)^{n-s}(n+s+1)!}{(n-s)!\,(s+l+1)!\,(s-l)!} \qquad \textbf{Equation 5-3}$$

To replace this function for a digital image, the integrals become summations to get:

COMPLEX MOMENT in DIGITAL IMAGE

$$PZM_{nl}$$

<div align="right">**Equation 5-4**</div>

$$= \frac{n+1}{\pi} \sum_r \sum_\theta . \widetilde{R_{nl}}(r). e^{-jq\theta}. f(r\cos\theta, r\sin\theta), r \leq 1$$

Where $n = 0,1,2,\ldots,\infty$ and l can be a positive or negative value subject to $|l| \leq n$.

The same algorithm can be applied to Eqs. 5-1, 5-2 and 5-3, to validate that the absolute value of the complex PZM of a rotated image is identical to original image before rotation. (Abu-mostafa and Psaltis 1985) proposed complex moments for feature extraction, with the aim of extracting a simpler moment descriptor while maintaining the property of invariance.

PZM polynomials are orthogonal sets of complex values of the nth order with repetition l and are defined as:

PZM POLYNOMIALS USED AS OUR PROBOSED SHAPE VECTOR

$$PZM_{nl} = \frac{n+1}{\pi} \sum_{\substack{s=0 \\ n-l-s \ even}}^{n-|l|} S_{nls} \sum_{j=0}^{\frac{n-|l|-s}{2}} \sum_{m=0}^{|l|} \binom{\frac{n-|l|-s}{2}}{j} .$$

<div align="right">**Equation 5-5**</div>

$$\binom{|l|}{m}(-i)^m \, CM_{n-s-2j-m,2j+m} \; +$$

$$\frac{n+1}{\pi} \sum_{\substack{s=0 \\ n-l-s\ odd}}^{n-|l|} S_{nls} \sum_{j=0}^{(n-|l|-s+1)/2} \sum_{m=0}^{|l|} \binom{\frac{n-|l|-s+1}{2}}{j} \cdot$$

$$\binom{|l|}{m} (-i)^m RM_{n-s+1-2j,2j+m}$$

Where $CM_{p,q}$ denotes the scale invariant central moment:

SCALE INVARIANT CENTRAL MOMENT

<div align="right">**Equation 5-6**</div>

$$CM_{p,q} = \mu_{pq} = \sum_x \sum_y (x - \overline{x})^p \cdot (y - \overline{y})^q\, f(x,y)$$

$$RM_{p,q} = \mu_{pq} = \sum_x \sum_y (x - \overline{x})^p \cdot (y - \overline{y})^q \sqrt{(x - \overline{x})^2 + (y - \overline{y})^2} \cdot f(x,y) \quad \text{**Equation 5-7**}$$

Where p and q are the orders, with values = 0,1,2, and f(x,y) is the grey value of the digital image at locations *x,y*.

To make PZM invariant with respect to scale and rotation requires the central moments be normalised by dividing by the zeroth moment (Chong, Raveendran et al. 2003):

The radial geometric moments, *RM*$_{p,q}$, are given by:

For invariant scale and rotaion, the normalised radial geometric moment should be divided by the zeroth moment according to:

SCALE INVARIANT CENTRAL MOMENT NORMALISED

$$CM_{p,q} = \frac{\sum_x \sum_y (x - \overline{x})^p \cdot (y - \overline{y})^q\, f(x,y)}{m_{00}^{\left[\frac{p+q}{2}\right]+1}} \quad \text{**Equation 5-8**}$$

NORMALISED RADIAL GEOMETRIC MOMENT

$RM_{p,q}$

$$= \frac{\sum_x \sum_y (x - \bar{x})^p \cdot (y - \bar{y})^q \sqrt{(x - \bar{x})^2 + (y - \bar{y})^2} \cdot f(x, y)}{m_{00}^{\left[\frac{p+q}{2}\right]+1}}$$ **Equation 5-9**

Where $\bar{x} = \frac{m_{10}}{m_{00}}$, $\bar{y} = \frac{m_{01}}{m_{00}}$.

Central moments are formed by extracting the centroid from all coordinates.

The normalisation process is performed as: $\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\lambda}}$ **Where** $\lambda = 1 + \frac{p+q}{2}$

To compute zeroth order (m_{00}) and first order (m_{10}), use Eq. 5-10:

RAW IMAGE MOMENTS

$$m_{pq} = \sum_x \sum_y x^p \cdot y^q \cdot f(x, y)$$ **Equation 5-10**

The rotation of the image does not affect the magnitude of the PZMs but only changes the phase. This operation maintains invariance under rotation of images.

5.5 Data Clustering and Indexing

Clustering is an operation of classifying a data set into a group of objects, where each group surrounds a particular target object. There are many clustering techniques such as supervised and unsupervised clustering. The support vector machines (SVM) is considered as one of the most popular supervised learning clustering method. It starts with assigning a set of training examples. A SVM training algorithm builds a model that marks new examples into one category, making it a non-probabilistic binary linear classifier. A SVM model denotes the examples as points in a feature space, and mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and can be predicted to belong

to a category based on which side of the gap they fall on. Moreover, besides linear classification, SVM can be efficiently performed for non-linear classification using what is called the kernel trick (Bishop 2006).

Most popular clustering techniques employ the sum-of-squares principles, which reduce the sum-of-squares error as in classic k-means, vector quantization, and Fuzzy k-means exercises.

Searching large databases can be very costly. All members of the database are examined and compared with the query image. This comparison requires calculations on the distances between the query image and all images in the database which requires a considerable amount of time and computation power, especially in large databases. To reduce the computation burden, we have introduced a novel approach that will reduce the required computations by an order of magnitude. The approach is based on using the k-means clustering algorithm (Kanungo, Mount et al. 2002) (Jain 2010) to divide the feature vectors of all images in the database into several groups based on the similarities of their signature and hence their contents. After dividing the databases into groups we compute the centre of each group which is the mean of the signatures of all images belonging to the group. We compute the similarity between the query image and the centres of all groups. The aim of this algorithm to divide the data into k clusters and due to this the sum of squares can be minimised.

After looking for the most similar centre to the query, we can then search that group for the closest distance between the query and the group images. To illustrate this concept, assume that we have a database of 10,000 images sub-divided into 50 groups each containing 200 images. We first find the group with

its signature closest to that of the query image. This will require 50 comparisons. After finding the closest centre, we search that group to find the image in the group closest to the query image, which will require a further 200 comparisons. Instead of making 10,000 distance calculations, we require only 250. Using this technique, we can reduce the calculation time by as much as 97.5%. However, we not only reduce the time requirement but also increase the search accuracy. The situation is illustrated by Figure 5-3. The A, B, C and D are groups where the signature of each member is similar to the centre of that group. The red dots represent the signatures in each group and the black squares represent the centre or centroid of each group. For efficient analysis the data must be clustered in such a way that the distance between any two images in the same group is smaller than between any two images in different groups. The k-mean clustering algorithm is an efficient choice for this type of comparisons.
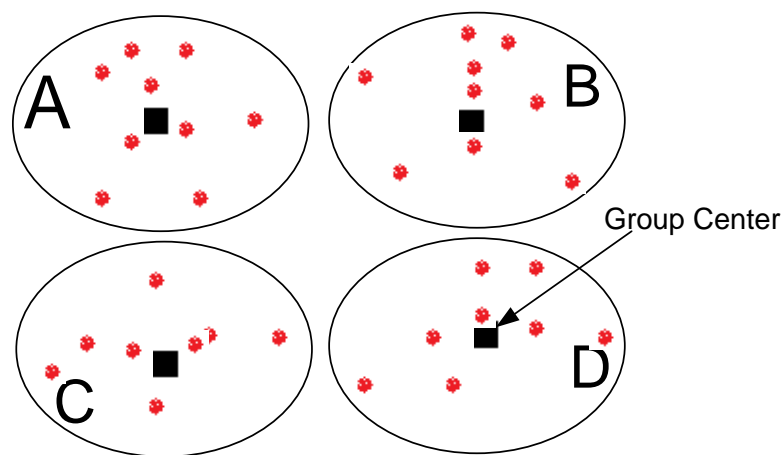
Figure 5-3   Clustering Groups

- **The K-Means Clustering Algorithm**

The k-means algorithm, partitions a collection of data containing n vectors x_j, where j=1,…,n into c-groups, G_i, i=1,…,c and finds a cluster centre in each group such that some objective function (or cost function) of distance is minimised (Hartigan and Wong 1979). When the Euclidean distance is chosen as a measure

of distance between vector x_k in group j and the corresponding cluster centre c_i, the objective function can be defined as:

OBJECTIVE FUNCTION OF CROUP G$_i$

$$J = \sum_{i=0}^{c} J_i = \left(\sum_{k, x_k \in G_i} \|x_{k-}c_i\|^2 \right)$$

Where J_i is the objective function within group i. Thus the value of J_i depends on the geometric properties of G_i and the location of c_i .

In general, the geometric distance function, defined as $d(x_k, c_i)$ can be applied for vector x_k in group i and the resulting overall objective function will be:

OVERALL OBJECTIVE FUNCTION

$$J = \sum_{i=0}^{c} J_i = \left(\sum_{k, x_k \in G_i} d(x_k, c_i) \right)$$

For simplicity this research project uses the "city block" distance (Yi, Jagadish et al. 1998) as a measure of dissimilarity. The partitioned groups are typically defined by a *c x n* binary membership matrix U, where the element u_{ij} is 1 if the jth data point x_j belongs to group I and 0 otherwise. Once the cluster centres c_i are fixed, u_{ij} can be the minimized as follows :

ELEMENT u$_{ij}$ OF MEMBERSHIP MATRIX U

$$u_{ij} = \begin{cases} 1 \ if \ \|x_j - c_i\|^2 \\ 0 \ otherwise \end{cases} \leq \|x_j - c_k\|^2 \ k \neq i$$

This can be stated as x_j belongs to group *i* if c_i is the closest centre amongst all centres. Since a given data point can only belong to one group, the membership matrix U has the following properties:

Equation

5-14

$$\sum_{i=1}^{c} u_{ij} = 1, \ \forall j = 1, \dots, n \ \ AND \ \ \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij} = n$$

On the other hand, if u_{ij} is fixed, the mean of all vectors in group i will be:

Where $|G_i|$ is the size of G_i or $|\boldsymbol{G_i}| = \sum_{j=1}^{n} \boldsymbol{u_{ij}}$

MEAN OF ALL VECTORS IN GROUP G$_i$ WITH u$_{ij}$ FIXED

$$c_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k \qquad \textbf{Equation 5-15}$$

When the K-means algorithm is presented with a data set $x_i, i = 1, \dots, n$; the algorithm determines the cluster centres c_i and the membership matrix U iteratively using the following steps (Kanungo, Mount et al. 2002)

- Step 1: Initialize the cluster c_i i=1,…,c. This can be achieved by selecting the point c randomly.

- Step 2: Determine the matrix U using Eq. 5-13.

- Step 3: Compute the cost function according to Eq. 5-11. Stop if it is either below a certain tolerance value or the improvement over the previous iteration is below a certain threshold.

- Step 4: Update the cluster centres according to Eq. 5-15.

The k-means algorithm has been used in this research as the main clustering technique for its simplicity amid its unsupervised learning manner. This learning method is efficient and works well facing enormous data set.

5.6 Integration Rules for Mixing Colour and Shape Features

INPUT

Query Image in Colour space

OUTPUT

Retrieved images according to their similarity to the query image

Algorithm

- **First component : Colour-Based CBIR**

Step1. Convert RGB colour space into HSV colour space;

Step2. Calculate fuzzy membership function sets from the HSV colour space;

Step3. Compute the fuzzy membership function by using the triangular function;

Step4. Compute the feature extractor using Fuzzy Colour Histogram (FCH) and place them into one vector which contains 42 bins.

- **Second component: Shape-Based CBIR**

Step 1. Convert RGB colour space into Grey Scale Space;

Step 2. Computer the real-valued polynomial or radial of Pseudo Zernike moment (PZM);

Step 3 Compute the r, the length of the vector from the origin to the image(x,y);

Step 4. Compute the PZM polynomials.

Step 5. Calculate PZM_{nl} polynomials vector which contains 25 bin if the n=4

- **FFCSS-Based CBIR**

Step1. Combine the 42 colour bins with the 25 shape bins to create a unified feature vector.

Step2. Retrieve the results in an iterative style if the user is not satisfied with the PZM parameter n.

5.7 FFCSS Feature Extraction

Design of the FFCSS system is as shown in Figure 5-4 is a combination of the fuzzy colour histogram (to get high performance of smaller feature of colour) with the PZMs signatures as a shape weighting scheme. Clearly, the use of a combination of colour and shape signatures will provide strong discrimination of most visual features.
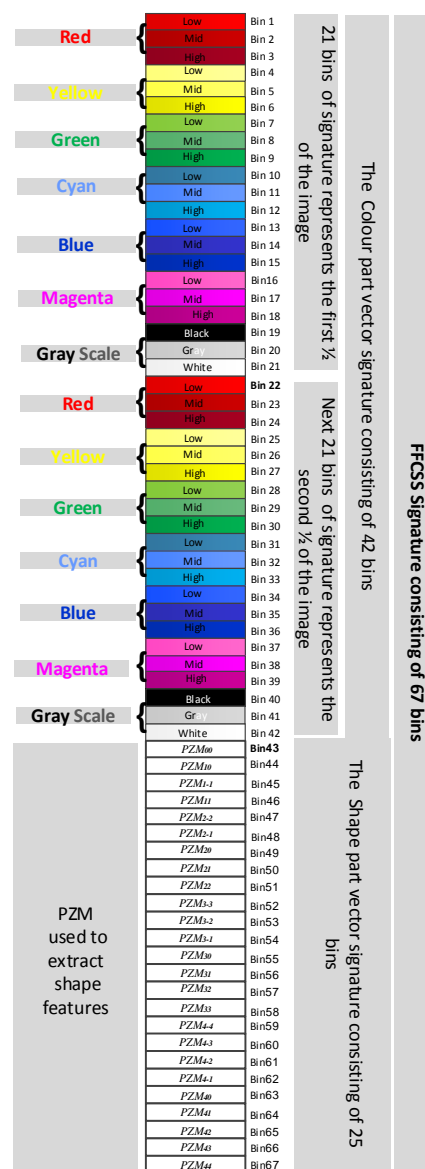
Figure 5-4   FFCSS Signature Design

- **Conclusions**

The design of the proposed system has carefully taken into account the concept that any reliable system should have at least the ability to represent a query by example. In this thesis, colour and shape fusion features have been used to improve the accuracy of image retrieval, first the fuzzy colour histogram (FCH) is extracted and the fuzzy colour histogram used to reduce the processing time.

Then, the shape Pseudo-Zernike Moments (PZM) feature is used to detect shapes and objects. Finally, both features are combined to get a unified feature, and the query retrieves the best match images for fusion of colour and shape. The system should be user friendly with a system interface which allows the user to browse and review the results through a relevant feedback stage.

Chapter 6. Experimental Results and Evaluation

This chapter reports the performance of the system prototype when verified and evaluated. The robustness of colour-based and shape-based CBIR and the integrated FFCSS scheme were assessed and it has been demonstrated through practical experiments that integration of colour-oriented and shape-oriented features produced a system which outperformed either feature on its own.

- **Similarity Measure**

Comparing two images is an essential step for visual feature identification and matching in CBIR approaches. In most cases to compare two images, checking their similarity by using one-to-one pixel comparison is not efficient. According to (Amato and Di Lecce 2008), a more practical approach would be:

- Exploring an adequate set of visual features such that these features can be used to extract priorities and then later used as a measure.
- Providing specific metric measure to feature space.

6.1 Performance Measures of Query Results of FCH

The most common measures of system performance are response time and space. Other factors for the evaluation of a retrieval system include ranking the relevant image set, effectiveness of indexing and operational efficiency (Müller, Müller et al. 2001). The type of evaluation to be considered depends on the

objectives of the retrieval system. However, it should be noted that obtaining a single satisfactory measure by using the features from both colours and shapes for evaluating a retrieval system is extremely difficult.

The effectiveness of an information retrieval system is a measure of the system's ability to satisfy the users in terms of the relevance of the retrieved information, and is often measured by using recall and precision (Wilbur 1992). Here, recall measures the ability of a system to retrieve the relevant documents, and precision measures the ability to reject irrelevant ones. There are alternative evaluation measures such as the E measure (Wilbur 1992), the harmonic mean (Xu 2009), satisfaction (Shou, Wang et.al, 2007) and frustration (Chance, Nioka et.al, 2001)

The following sub-section contains a more detailed explanation of recall and precision and details several experiments conducted by author to assess the effectiveness and efficiency of FFCSS methods.

6.1.1 Recall and Precision

Recall and precision are the most common tools used to measure and assess information retrieval systems. In general the relevant documents used in information systems but in image retrieval the relevant images specifically used. The recall function for such a representation is given by:

$$\text{RECALL} = \frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ relevant\ documents} = \frac{|Ra|}{|R|}$$

Equation 6-1

For example, if the database comprises of 100 relevant documents for a query, and the search procedure retrieved only 10 of these, then the recall of the system for this particular query would be 10%.

Conversely, precision is a measure of the information retrieval system to retrieve only the relevant documents (Turpin and Scholer 2006). For a given query document the precision is the ratio of the number of relevant documents to the total number of retrieved documents:

$$\frac{Number\ of\ relevant\ documents\ retrieved}{Total\ number\ of\ retrieved\ documents}$$

PRECISION $$= \frac{|Ra|}{|A|}$$ **Equation 6-2**

For example, if the number of retrieved documents is 100 and 20 of these are relevant, then the precision of the system for this particular query would be 20%. Both recall and precision have values in the interval [0, 1]. For an efficient retrieval system the values of both recall and precision should be close to 1.

Figure 6-1 shows some sample images selected to measure precision and recall.

The first experiment carried out was to test the performance of the colour component FCH relative to the CCH. The experiment was performed on three large heterogeneous databases. The VARY database which contains 10,000 images of natural scenes where the images range in size from 128 x 85 to 128 x 96 pixels. This database is internationally used as a benchmark for CBIR testing. The second database was a database of the flags of 224 different nations. This is a simple database used for primary testing.

The third database is the author's own and contains 1000 colour images taken under different conditions. These images are all about 1024 x 256 pixels. Figure 6-1 shows some image samples and the comparable recall and precision measurements for FCH and CCH. The results demonstrate that the FCH technique performed better than the CCH because the values of both precision and recall for FCH were higher than the values obtained with CCH.

For example suppose there are 6 images as expected answers and the set size is 16.

The recall is computed as follows:

$$Recall = \frac{4\ images\ (\ number\ of\ relevant\ images\ retrieved)}{6\ images\ (Total\ number\ of\ relevant\ images\ )} = \frac{4}{6} = 66.7\ \%$$

And the precision is computed as follows:

$$Precision = \frac{4\ images\ (\ number\ of\ relevant\ images\ retrieved)}{16\ images\ (Total\ number\ of retreived\ images\ )} = \frac{4}{16} = 25\ \%$$

Query Image **Pumpkin** (size of expected answer: 6 images)

16 best matches by FCH – Recall: 4 /6 = 66.7 %, Precision: 4 /16=25%

16 best matches by CCH – Recall: 2 /6=33.3 %, Precision: 2 /16=12.5 %
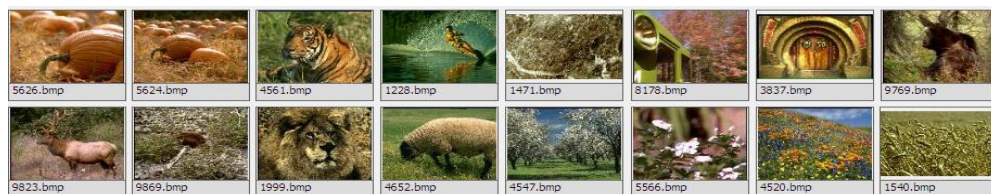
Query Image **Space Machine** (size of expected answer: 18 images)

16 best matches by FCH – Recall: 7/18=38.9%, Precision: 7 /16=43.8%
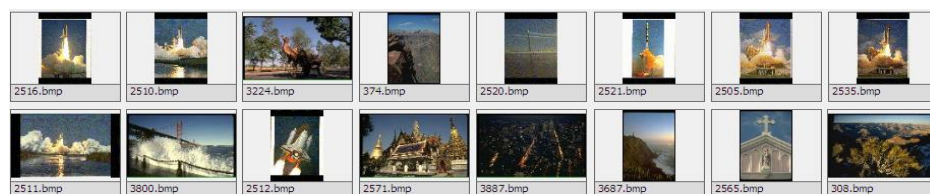
16 best matches by CCH – Recall: 6/18=33.3%, Precision: 6/16=37.5%

Query Image **Waves** (size of expected answer: 7 images)

1392.bmp

16 best matches by FCH – Recall: 5/7=71.4 %, Precision: 5/16=31.3%

1392.bmp | 1363.bmp | 1482.bmp | 1787.bmp | 2082.bmp | 1393.bmp | 2060.bmp | 2652.bmp

1366.bmp | 1391.bmp | 1670.bmp | 1396.bmp | 1419.bmp | 2304.bmp | 1784.bmp | 1778.bmp
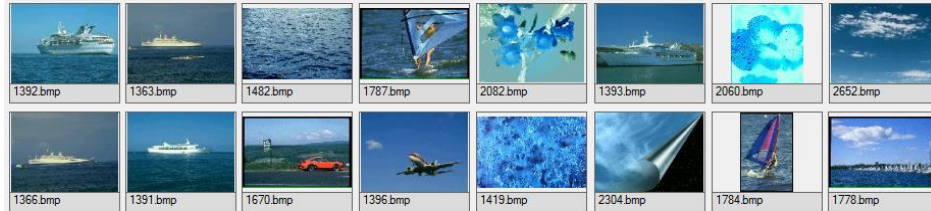
16 best matches by CCH – Recall: 1/7=14.3 %, Precision: 1/16=6.3 %

1392.bmp | 1787.bmp | 1713.bmp | 2206.bmp | 2421.bmp | 2393.bmp | 1699.bmp | 1745.bmp

1584.bmp | 2571.bmp | 1422.bmp | 1164.bmp | 1050.bmp | 1428.bmp | 1456.bmp | 1790.bmp

Query Image **Austria flag** (size of expected answer: 6 images)

Austria.bmp

16 best matches by FCH – Recall: 4/6=66.7 %, Precision: 4/16=25%

Austria.bmp | Peru.bmp | Poland.bmp | French_Polynesia.b | Malta.bmp | Lebanon.bmp | Chile.bmp | Tunisia.bmp

Hungary.bmp | Yemen.bmp | Trinidad_and_Tobag | Turkey.bmp | Indonesia.bmp | Russian_Federation. | Singapore.bmp | Hong_Kong.bmp

16 best matches by CCH – Recall: 2/6=33.3 %, Precision: 2/6= 12.5 %

Austria.bmp | Poland.bmp | Yemen.bmp | Tunisia.bmp | Iraq.bmp | Hungary.bmp | Madagascar.bmp | China.bmp

Italy.bmp | Laos.bmp | Trinidad_and_Tobag | Peru.bmp | Russian_Federation. | Slovakia.bmp | Lebanon.bmp | Egypt.bmp

Figure 6-1 Recall and Precision for FCH and CCH for Different Databases

6.1.2 Lighting Intensity Test

We can evaluate the performance of image retrieval according to the Normalised Rank Sum (NRS) (McCaffrey 2009). The database is divided into sets; each set contains a number of similar images. When a query image belonging to a set is selected to test the performance of the CBIR system, CBIR can respond by a list of images sorted according to the distance from the query image. The rank of each image corresponds to its order in the retrieved list. The normalized rank sum is defined as follows:

NORMALISED RANK SUM (NRS)

$$NRS = \frac{Sum\ of\ ranks\ in\ the\ set}{Sum\ of\ ranks\ retrieved\ from\ the\ set} = \frac{\sum_{i=0}^{n-1} ranks\ (I_{i \in S})}{\sum_{j=0}^{m-1} ranks\ (I_{j \in R})} \qquad \textbf{Equation 6-3}$$

Ideally, if all images were retrieved according to their expected order, then NRS would be equal 1 (the ideal result). The performance of the system was also evaluated under changes to the light intensity.

Ten images were chosen randomly and each image presented with thirteen levels of brightness compared to the original image, -35%, -30%,-25%,-20%, -15%,-10%, no change (using original image), +10%, +15%, +20%, +25%, +30%, and +35%.

Figure 6-2 Selected Images for Testing FCH and CCH with Change in Light Intensity

From each image in Figure 6-2, thirteen images were created, including the original image, and were added to the database. Each image was used as a single query. The retrieval results are presented in ascending order of light intensity, and the NRS was computed using Eq. 6-3. Results are presented in Table 6-1 and a comparison of the NRS values obtained using FCH and CCH shows clearly that the proposed FCH system can accomplish much better discrimination.

- **Computation of NRS**

The NRS is computed using Eq. 6-3. For example, the retrieval ranking of the thirteen images was 1,2,3,4,5,6,7,8,9,10,11,12,30. For example, the entry 30 means image was retrieved in position 30 in the answer set.

$$NRS = \frac{n(n+1)/2}{\sum_{j=0}^{m-1} ranks(I_{j\epsilon R})} = \frac{13(13+1)/2}{1+2+3+4+5+6+7+8+9+10+11+12+30} = \frac{91}{108} = 0.842$$

The accuracy of retrieval is 0.842*100=84.2%

Table 6-1 NRS Values Obtained for Ten Query Images with Thirteen Levels of Relative Brightness for FCH and CCH.

		Brightness (%)												
		-35	-30	-25	-20	-15	-10	0	10	15	20	25	30	35
Image 1	FCH	0.989	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	CCH	0.052	0.074	0.108	0.127	0.180	0.260	0.294	0.206	0.152	0.132	0.130	0.092	0.076
Image 2	FCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	CCH	0.105	0.136	0.197	0.247	0.332	0.431	0.500	0.508	0.416	0.324	0.341	0.262	0.218
Image 3	FCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	CCH	0.011	0.016	0.022	0.032	0.053	0.081	0.225	0.119	0.067	0.040	0.028	0.018	0.015
Image 4	FCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	CCH	0.106	0.138	0.188	0.223	0.306	0.337	0.401	0.374	0.298	0.232	0.209	0.164	0.113
Image 5	FCH	0.674	0.850	0.968	1.00	1.00	1.00	1.00	1.00	0.978	0.968	0.929	0.858	0.752
	CCH	0.011	0.014	0.017	0.024	0.035	0.051	0.095	0.057	0.036	0.023	0.016	0.012	0.009
Image 6	FCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.989
	CCH	0.018	0.023	0.027	0.037	0.054	0.067	0.116	0.084	0.064	0.044	0.031	0.023	0.017
Image 7	FCH	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	CCH	0.026	0.031	0.038	0.047	0.064	0.086	0.116	0.104	0.078	0.062	0.045	0.035	0.027
Image 8	FCH	0.919	0.948	0.968	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.978
	CCH	0.017	0.019	0.026	0.038	0.059	0.096	0.314	0.170	0.098	0.047	0.032	0.020	0.014
Image 9	FCH	0.948	0.958	0.968	0.989	1.00	1.00	1.00	1.00	1.00	1.00	0.989	0.938	0.978
	CCH	0.027	0.948	0.948	0.948	0.948	0.049	0.057	0.067	0.066	0.078	0.062	0.023	0.052
Image 10	FCH	0.968	0.968	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	CCH	0.883	0.910	0.948	0.958	0.978	0.989	1.00	0.989	0.948	0.901	0.867	0.827	0.771
FCH rate		89.68	100.0	100.0	100.0	99.78	99.68	99.18	97.96	89.68	88.34	97.24	99.04	99.89
CCH rate		13.13	24.47	31.18	26.79	22.24	18.83	17.60	14.75	13.13	12.57	23.08	25.19	26.82

Conventional colour-based methods (CCH) are sensitive to illumination variations due to the over-definition of the colour combinations. In this research, the proposed FCH utilises the local background information around each pixel to take into account just basic colour bins for computation.

6.1.3 Noise Test

The principal sources of noise in digital images arise during image acquisition or transmission. Salt and Pepper noise has been used to test the noise ability. This noise is given by Eq. 6-4, where z represents grey level. The noise is applied to the chosen image into database and searching by query image which is the original image before salt and pepper noise has been added.

$$\begin{cases} P_a & for\ z=a \\ P_b & for\ z=b \\ 0 & otherwise \end{cases}$$

Equation 6-4

SALT AND PEPPER NOISE *P(z)*

Where μ is the mean value of z, and σ^2 is its variance, P_a is the probability dinsty of pepper noise , P_b is the probability density of salt noise. $\mu = \frac{a}{b}, \quad \sigma^2 = \frac{b}{a^2}$

If *b* >*a*, grey level *b* will appear as a light dot in the image. Conversely, level *a* will show as a dark dot if either P_a or P_b is zero. Figure 6-3 shows the probability density function for Salt and Pepper noise.
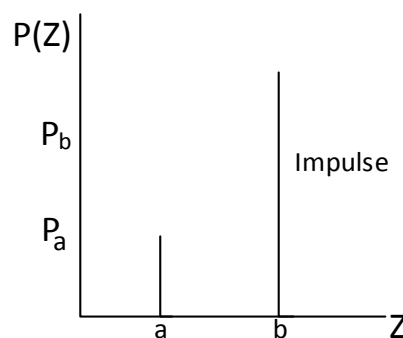
Figure 6-3   Probability Density Functions for Salt and Pepper Noise

The system has been tested with different mean values of P_a and P_b of 0.4, 0.5, 0.6, and 0.7 on the query image. The retrieval results were excellent until value

of the mean reached 0.8. Figure 6-4 shows the original image on the left and the

added noise with 0.5 of probability of P_a and P_b on the right.

Original Image   Added noise with same value (0.5) to
                        P_a and P_b

Figure 6-4   Probability Density with Mean Value 0.5 for both Salt and Pepper Noise

6.2 Results and Discussion for FCH

In this research, we have used several image databases for the tests which were

performed on a laptop computer running Windows 8 with Intel 1.6 GHz processor.

The method of testing is to provide a query image and its signature or feature

extraction and the system is asked to find the most similar images by measuring

distances between signatures. The result of the search is a list of images sorted

by their similarity to the query.

In Figure 6-5, the query image is shown on the left and the results shown on the

right. The answer set contains the closest 16 images obtained by the proposed

system (FCH) and another commonly used technique (CCH). The proposed FCH

technique was tested on the three databases described above.

Query Image	Using	Retrieval Results

Figure 6-5   Results Obtained Using VARY Database

Results obtained by using the VARY database are shown in Figure 6-5, retrieving results of Country Flags database are shown in Figure 6-6, and results obtained from the author's own database are shown in Figure 6-7. It has been proven in these tests that the proposed FCH system can perform much better than the more conventional CCH method.

The next test was retrieval of four query images (of given flags) from the database containing the images of 224 flags.

Figure 6-6 Retrieval Results Obtained Using FCH and CCH with Database of Flags of 224 Countries

The final test was to retrieve eight query images from the author's own database of **Aboaisha** images. Results are presented in Figure 6-7.

Query Image	<u>Using</u>	**Retrieval Results**
1216.bmp	FCH	1216.bmp 1214.bmp 1215.bmp 1211.bmp 1213.bmp 1217.bmp 1212.bmp 1221.bmp 1219.bmp 1218.bmp 1222.bmp 1220.bmp 1149.bmp 700.bmp 701.bmp 964.bmp
	CCH	1216.bmp 1212.bmp 1211.bmp 1214.bmp 1215.bmp 1213.bmp 1222.bmp 1207.bmp 754.bmp 722.bmp 1051.bmp 752.bmp 697.bmp 696.bmp 1203.bmp 1056.bmp
12.bmp	FCH	12.bmp 15.bmp 71.bmp 9.bmp 39.bmp 27.bmp 31.bmp 133.bmp 100.bmp 57.bmp 35.bmp 104.bmp 117.bmp 111.bmp 78.bmp 347.bmp
	CCH	12.bmp 15.bmp 39.bmp 57.bmp 547.bmp 560.bmp 71.bmp 559.bmp 394.bmp 549.bmp 303.bmp 104.bmp 301.bmp 302.bmp 306.bmp 307.bmp
800.bmp	FCH	800.bmp 801.bmp 802.bmp 867.bmp 804.bmp 899.bmp 806.bmp 1057.bmp 881.bmp 811.bmp 1064.bmp 809.bmp 900.bmp 1156.bmp 1063.bmp 883.bmp
	CCH	800.bmp 801.bmp 802.bmp 1151.bmp 789.bmp 858.bmp 896.bmp 214.bmp 213.bmp 877.bmp 895.bmp 920.bmp 221.bmp 210.bmp 209.bmp 754.bmp

.

Query Image	Using	Retrieval Results
512.bmp	FCH	512.bmp, 513.bmp, 511.bmp, 515.bmp, 601.bmp, 518.bmp, 516.bmp, 514.bmp, 523.bmp, 540.bmp, 600.bmp, 599.bmp, 597.bmp, 593.bmp, 517.bmp, 509.bmp
	CCH	512.bmp, 513.bmp, 511.bmp, 524.bmp, 523.bmp, 563.bmp, 543.bmp, 535.bmp, 526.bmp, 522.bmp, 510.bmp, 396.bmp, 506.bmp, 527.bmp, 397.bmp, 317.bmp
670.bmp	FCH	670.bmp, 678.bmp, 675.bmp, 673.bmp, 677.bmp, 671.bmp, 674.bmp, 669.bmp, 680.bmp, 676.bmp, 672.bmp, 681.bmp, 679.bmp, 682.bmp, 683.bmp, 298.bmp
	CCH	670.bmp, 671.bmp, 680.bmp, 676.bmp, 674.bmp, 678.bmp, 675.bmp, 677.bmp, 672.bmp, 673.bmp, 681.bmp, 669.bmp, 547.bmp, 303.bmp, 304.bmp, 301.bmp
343.bmp	FCH	343.bmp, 338.bmp, 340.bmp, 335.bmp, 344.bmp, 330.bmp, 341.bmp, 529.bmp, 331.bmp, 562.bmp, 561.bmp, 413.bmp, 336.bmp, 649.bmp, 579.bmp, 575.bmp
	CCH	343.bmp, 413.bmp, 405.bmp, 515.bmp, 333.bmp, 409.bmp, 689.bmp, 686.bmp, 516.bmp, 415.bmp, 688.bmp, 687.bmp, 411.bmp, 406.bmp, 410.bmp, 336.bmp

Query Image	Using	Retrieval Results
250.bmp	FCH	
	CCH	
2.bmp	FCH	
	CCH	

Figure 6-7   Retrieval Results Obtained Using FCH and CCH with the Author's Own Database of Aboaisha Images

6.3 PZM Descriptor Evaluation and Results

In this section, the results obtained by using the PZM technique depended on the orientation of the shape within the image. The results were as follows:

The MPEG-7 standard database was used to test the performance of the PZM approach and to evalute general retrieval. Subset B was used. The PZM technique gave good results when the query image was submitted. All images in the answer set related to the original query image and were retrieved in an efficient manner. A review of the answer set retrieved using the PZM approach, to the naked eye, shows the system to have high accuracy.

For example, a pigeon shape was chosen randomly from the datasets and used as a query image as shown in Figure 6-8.

The answer set was the retrieved images shown in Figure 6-9. The results are quite accurate and all the retrieved images are relevant to the query image.
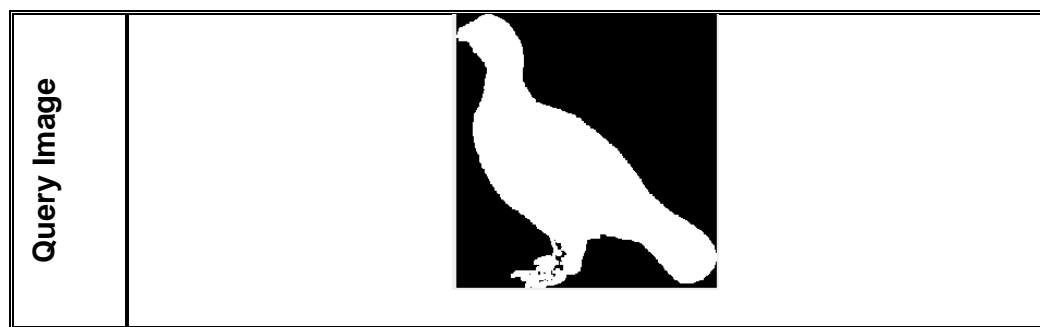
Figure 6-8   Query Image Used to Test Performance of the PZM Approach
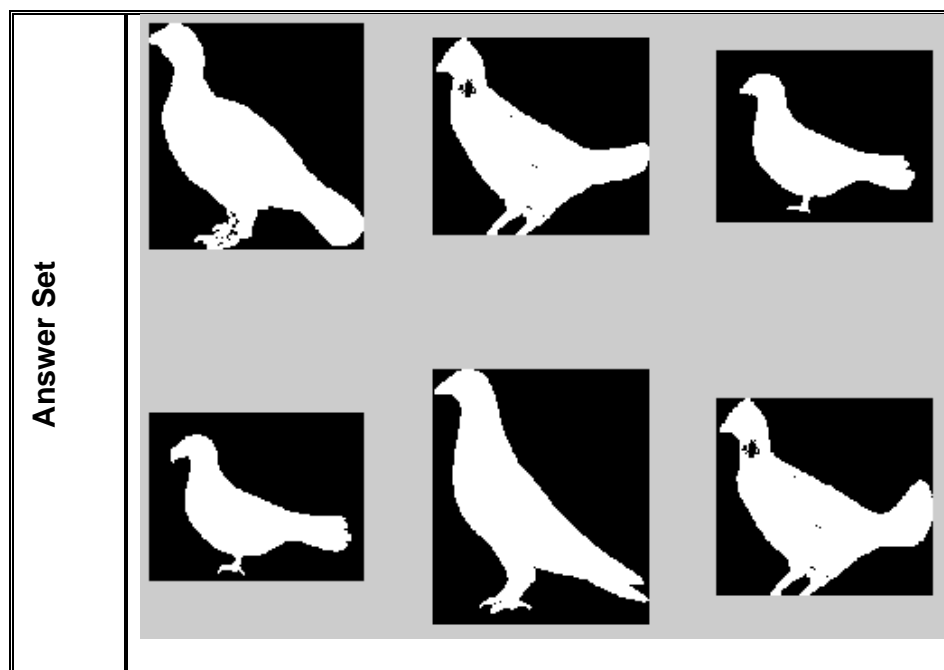
Figure 6-9   Retrieved Results using PZM Technique with database MPEG7-set B

To test the performance when scaling occurs, the MPEG7- subset A1 was used. The results obtained show the accuracy of retrieval of PZMs was superior to ZMs and LMs with a retrieval ratio greater than 99%.The system was also tested with images of different sizes.

6.4 FFCSS Prototype System

The first step in the fusion system is to compute the fuzzy colour histogram as the colour signature. For example, consider the query image shown in Figure 6-10, the bar chart in Figure 6-11 represents the distribution of FCH features.

The magnitudes of the first 21 bars in Figure 6-11 are the bin values 1 to 21 corresponding to Figure 6-10. Here we note that values are relative high for bins 11 and 12 which means that cyan is a dominant colour in the image, similarly, the high values for bins 13 and 14 show that the colour blue is also important.

The magnitudes of the second 21 bars in Figure 6-11 (bars 22 – 42) are the bin values 22 to 42 corresponding to Figure 6-10. It is worth noting that bins 26 and 27 which represent the colour yellow have the highest values. As presented in FCH signature in Chapter 5 (Section 5.2), it is clear that these peak values refer to cyan, blue and yellow which are dominate colours in the image. The retrieved results using the FCH alone are shown in Figure 6-12.

Figure 6-10   Query Image
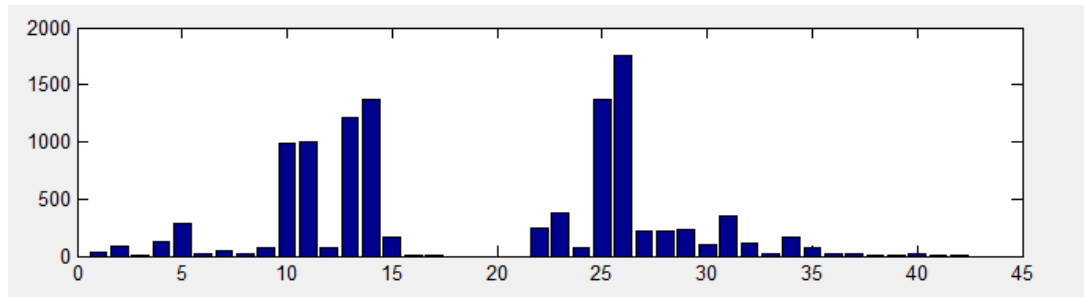
Figure 6-11   Presentation of the FCH Signature
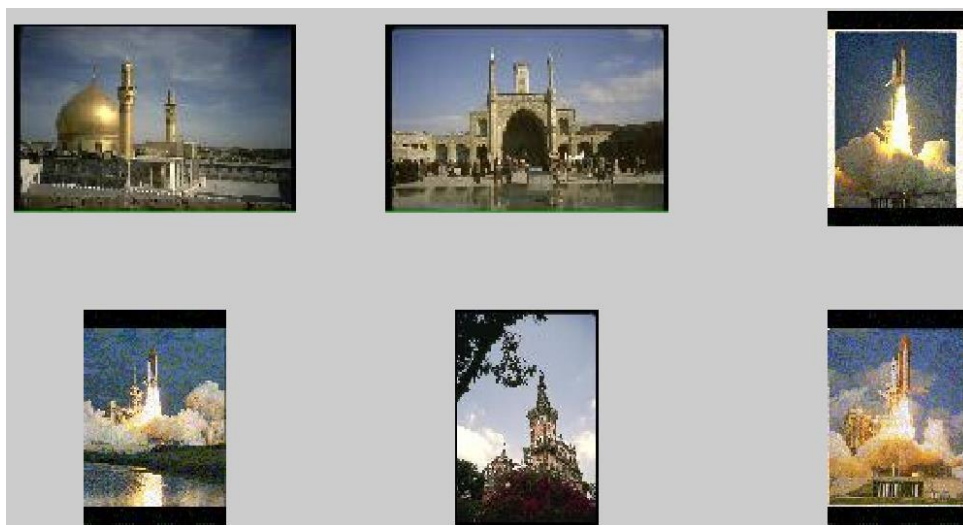
Figure 6-12   Images Retrieved Using FCH Based CBIR

The second step was to again use the image shown in Figure 6-10 as the query image, but the retrieved images were found by applying only the PZM descriptor features, see the bar chart in Figure 6-13. It is obvious that the values of the moments are normalised to fit between -1 to 1. Each bin represents one moment of the PZM and together make the PZM shape signature. Figure 6-14 shows the retrieved images for the PZM signature only (effectively bins 1-42 in Figure 6-13 are again omitted from consideration).
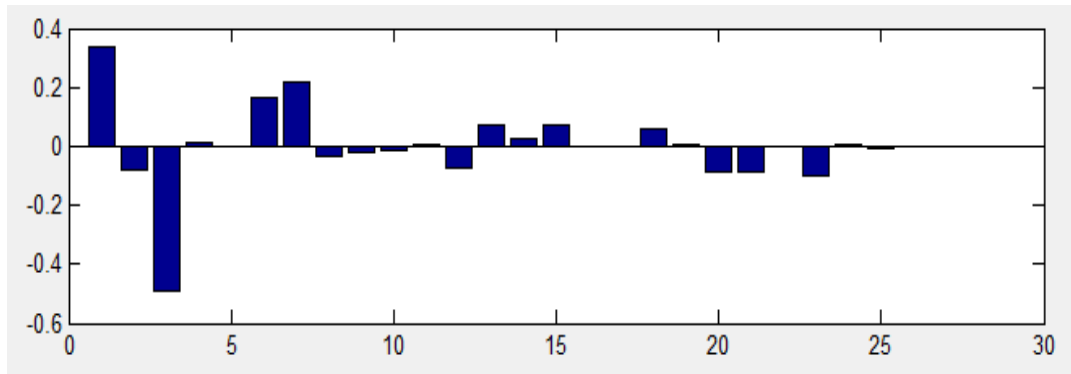
Figure 6-13   The Presentation of The PZM Signature
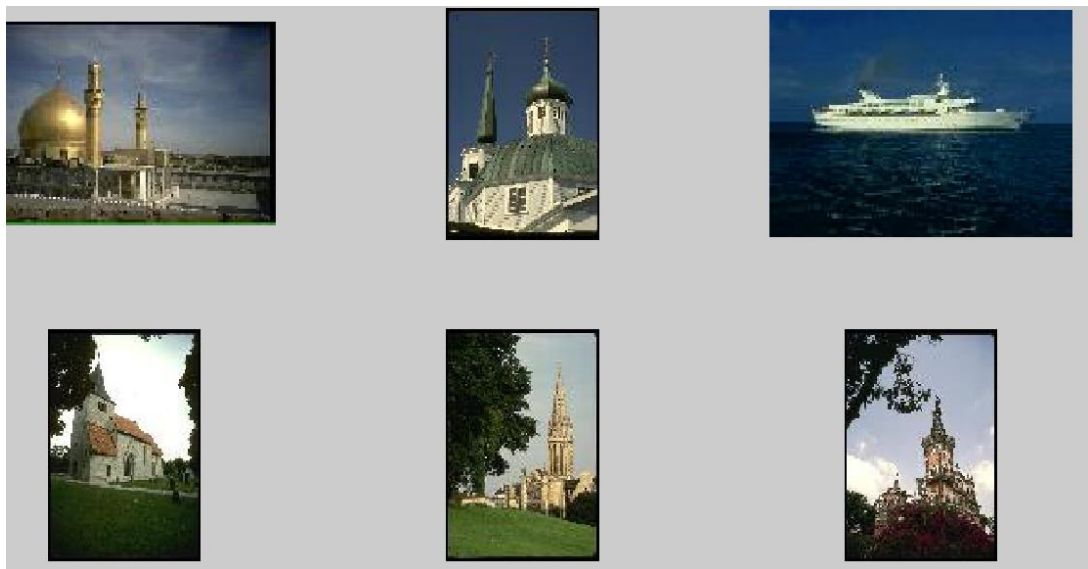
Figure 6-14   Images Retrieved Using PZM Descriptor

The second step in the application of the FFCSS system is to extract the shape signature using only the PZM polynomials which provide the remaining features in bins 43-67. The weighting scheme for the PZM descriptor depends on the n-value of moments and dictates the length of a PZM signature, see appendix A.

The final step in the FFCSS system is to link FCH and PZM to complete the integrated FFCSS signature, so that all bins (1 to 67) are used in the retrieval process, as discussed in Chapter 5. The FFCSS system gives a combined weighting scheme combining the colour vector signature obtained from FCH and shape vector signature obtained from PZM. Figure 6-15 shows the results obtained using FFCSS technique.
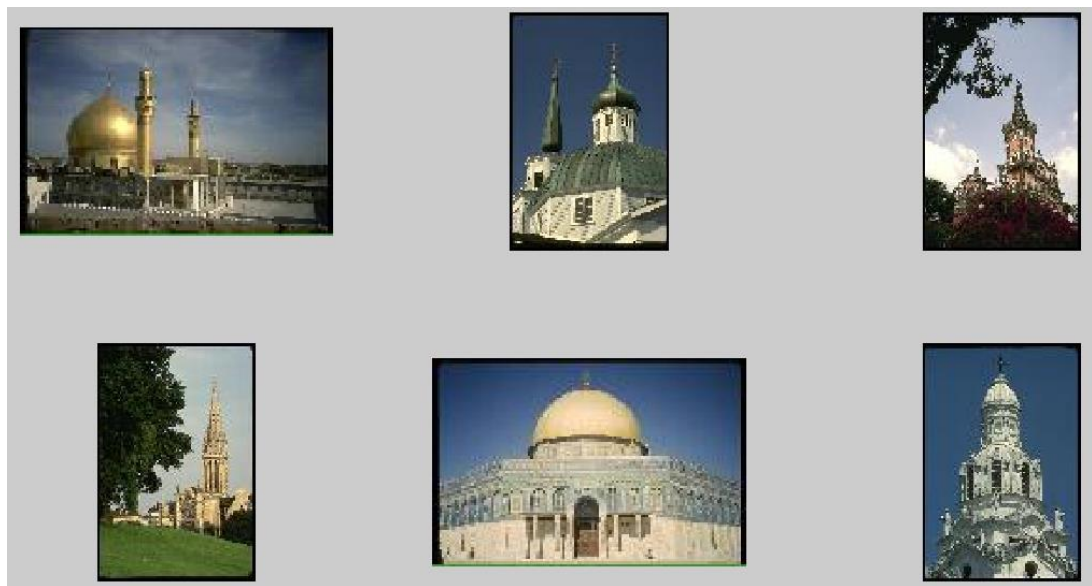
Figure 6-15   Images Retrieved Using the FFCSS Technique

6.5 Comparison of FFCSS with FCH and CCH

Image retrieval using FFCSS has achieved better results than colour or shape-only signatures. As discussed in Chapter 1, using just one single feature cannot be sufficient to fully or adequately define all details concerning an image content. Here the improvement comes from the integration of two single features: colour and shape.

The proposed system has the ability to combine the search for similar shapes and colours, to combine colour and shape in unified fusion-oriented features using a specific weighting scheme. Although FFCSS has achieved highly accurate results, the FCH or PZM are separate components and can be used to retrieve some special results by using relevance feedback. For instance, the user may desire to retrieve images where the dominant colour is blue, as with a sea or sky. FCH can do this perfectly.

A FFCSS system gives the user the choice when detecting every important corner or border on an image, through a shape component feature by choosing the number of moments. The greater the number of moments, the more details that can be obtained when the PZM shape feature is applied.

The FFCSS technique does not simply link the colours and shapes, it applies a selective weighting scheme. When determining all the colour ratio distributions using the fuzzy membership function, the colours are then extracted according to a fuzzy theory that models human perception. To avoid the curse of dimensionality, the selection of the signature feature vector is small to minimise time complexity. This leads to high performance and reduces processing time for the colour component.

When weighting with the PZMs features, especially when the number of small moments relies on the value of n which determines the number of moments, as discussed in Chapter 4, when the value of n is larger, more details are extracted. These two factors mean the FFCSS method is very effective and efficient.

6.6 FFCSS Results and Discussion

The shortcomings contained in early measurements of the CBIR systems meant they were limited to demonstrating just few sample of images results such as QBIC (Flickner, Sawhney et al. 1995). We believe that is why early CBIR systems were far from optimal.

The results obtained when using a fusion of shapes and colours are much better than the results obtained using colours or shapes only. When colours alone is used the focus is on the distribution of colours in the image and does not distinguish the distribution of shapes or objects, which can lead to serious omissions. Of course, one can retrieve all objects with a shape similar to the query image using shape-based CBIR and then complete a third stage to compare the two sets of results.

The advantage of the proposed system, is that we can use a powerful fusion of colour and shape to retrieve - in a single process - all objects which have same colour and shape. An experimental investigation has been carried out to investigate the performance of the proposed unified system, examine the robustness of the system and demonstrate its outstanding performance.

Chapter 7. Conclusions and Future Work

This research proposes a new method for image retrieval which combines the advantages of colour-based CBIR, relying on the fuzzy colour histogram, and shape-based CBIR relying on geometric moments. This integration leads to a powerful union of features for use in image retrieval.

7.1 Conclusions

This thesis investigated the problem of the sematic gap in current approaches to CBIR with the goal of improving image retrieval results by integrating more than one image feature into the retrieval process. This thesis began by investigating whether a single feature is adequate to describe the entire scene of the image or object on the image, and demonstrated that using just one single feature is not adequate for satisfactory retrieval performance.

The novel solution proposed for this problem, and the main object of this work, was to integrate more than one feature, combining them into a more powerful CBIR. Combining features provides greater discrimination power between images.

The main contribution of this thesis is the design of a new image extraction model which combines fuzzy colour and shape signatures to define image content. Chapter 3 describes a novel colour-based CBIR, the Fuzzy Colour Histogram (FCH).

The fuzzy logic has the ability to represent information in a manner suited to humans has been further validated in this research.

This research also added a new benefit to the CBIR system to increase retrieval speed by clustering all the images in the database into a number of groups. The preliminary search is limited to the centres of the clusters of images. After finding the closest centre to the query, only that cluster is searched. This step provides a drastic reduction in search duration, see Chapter 5.

Experiments were performed testing the new system on VARY, a heterogeneous database of images and the UCID standard. State-of-the-art fusion of colour and shape were investigated, and the colour-based CBIR and shape-based CBIR approaches mostly used for retrieving and indexing images were studied in Chapter 2.

The FCH introduced a novel structure for defining feature signatures used for image retrieval. This new signature has lower sensitivity to noise, greater robustness and enhanced flexibility. The flexibility allows the FCH to search different databases with high speed. The devised method in this research decreases the length of the signature of images to only 42 bins.

Chapter 4 introduced a shape-descriptor based on the PZM method. The PZM descriptor is used for abstracting shape information because of its intrinsic invariant discrimination power. The experiments were conducted on the MPEG-7 database to evaluate results under TRS rules.

The experiments clearly showed the high performance of the system and the robustness of PZM due to its geometric invariant characteristics.

A novel FFCSS multi-fusion signature combining the fuzzy-colour and PZM shape-descriptor advantages proven in the research has also been proposed (see Chapter 5) to alleviate some of the challenging problems occurred in single feature-based retrieval. The new multi-fusion features paradigm lead to superior accuracy in image retrieval which outperforms conventional CBIR techniques.

In interactive CBIR approaches, the developer or user should have the ability to interact with the system's interface in order to manipulate the results obtained from the query image, and this procedure can be performed several times to further refine the results until the user is satisfied. The FFCSS scheme has provided a viable medium for enabling and accelerating this process and in turn, narrowing the semantic gap between the meaningful information and their visual features.

Although there are still many obstacles to be investigated and overcome, the effort in fusing features in this research, as well as the devised techniques have achieved good results and made a significant step forward in helping to solve the challenges facing CBIR.

- **Limitations of FFCSS**

Although the FFCSS offers outstanding runtime results, there are still some drawbacks of its current form. For example, the speed for the extraction of the

PZM descriptor, especially when the n-value is too big, which does not affect off-line operations but causes negative impacts on online applications.

7.2 Future Work

Although the proposed methods and techniques in this research demonstrated valuable results, there are improvements that could further enhance future CBIR-related operations. In this study the FCH was fully automated and did not need any pre-processing, but the use of the shape component to determine objects or regions of interest in an image was done manually and required pre-processing.

Thus, an important extension of this research would be to improve shape extraction by segmenting images automatically. If this could be achieved the FFCSS algorithm would be substantially improved because the shape discrimination component would be made more powerful, improving the retrieval efficiency. This work confirms the views of many prominent computer vision and pattern recognition scientists that future research into effective methods of avoiding the "curse-of-dimensionality" would be of great significance if CBIR is ever to be widely applied rather than constrained to a few well-defined application domains. In addition, the marriage of research between algorithmic advances in image understanding and improvements in computer hardware such as consumable parallel processors (GPU, Cell-CPU) could derive semantically meaningful features in an effective manner (Jain and Vailaya 1998; Gupta and Jain 1997)

References

"http://www.free-country-flags.com/." (Coconuts Group).

Abu-mostafa, Y. S. and D. Psaltis (1985). "Image Normalization by Complex Moments." Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-7(1): 46-55.

Aggarwal, G., et al. (2002). "An Image Retrieval System with Automatic Query Modification." Multimedia, IEEE Transactions on 4(2): 201-214.

Aggarwal, G., and P. Dubey, (2000). "iPURE: Perceptual and User-Friendly Retrieval of Images". Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on.

Amato, A. and V. Di Lecce (2008). "A knowledge Based Approach for A Fast Image Retrieval System." Image and Vision Computing 26(11): 1466-1480.

Appass, A and A. Darwish (1999). "Image Indexing Using Composite Regional Color Channel Features" proc. SPIE, storage and retrieval for image and video Databases VII.

Aslandogan, Y. and C. Yu (1999). "Techniques and Systems for Image and Video Retrieval." Knowledge and Data Engineering, IEEE Transactions on 11(1): 56-63.

Bach, J., et al. (1996). "Virage Image Search Engine: An Open Framework for Image Management".

Belhumeur, P., et al. (1997). "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection." Pattern Analysis and Machine Intelligence, IEEE Transactions on 19(7): 711-720.

Bellman, R. and L. Zadeh (1970). "Decision-Making in A Fuzzy Environment." Management science 17(4): B-141-B-164.

Belongie, S., et al. (2002). "Shape Matching and Object Recognition Using Shape Contexts." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(4): 509-522.

Bhatia, A. and E. Wolf (1954). "On the Circle Polynomials of Zernike and Related Orthogonal Sets". Proc. Cambridge Philos. Soc, Cambridge Univ Press.

Biederman, I. (1987). "Recognition-By-Components: A Theory of Human Image Understanding." Psychological review 94(2): 115.

Bin, Z., et al. (2000). "Creating A large-Scale Content-Based Airphoto Image Digital Library." Image Processing, IEEE Transactions on 9(1): 163-167.

Bishop, C. M. (2006). "Pattern Recognition and Machine Learning ", springer New York.

Brlek, S., et al. (2009). "Lyndon+ Christoffel= Digitally Convex." Pattern Recognition 42(10): 2239-2246.

Carson, C., et al. (1997). "Region-Based Image Querying". Content-Based Access of Image and Video Libraries, 1997. Proceedings. IEEE Workshop on.

Celebi, M. E. and Y. A. Aslandogan (2005). "A comparative Study of Three Moment-Based Shape Descriptors". Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on.

Chance, B., Nioka, S., Li, C., & Nal, D. (2001, 2001). NIR imaging of prefrontal activation by anagram solutions. Paper presented at the Engineering in Medicine and

Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE.

Chang, S.-F., W. Chen, et al. (1997). "VideoQ: An Automated Content Based Video Search System Using Visual Cues". Proceedings of the fifth ACM international conference on Multimedia, ACM.

Cheikh, F. A., et al. (2003). "Relevance Feedback for Shape Query Refinement". Image Processing, 2003. ICIP 2003. Proceedings.

Chen, S. W. and G. Stockman (1990). "Wing Representation for Rigid 3D Objects". Pattern Recognition, 1990. Proceedings., 10th International Conference on.

Cheng, S., et al. (1998). "Semantic Visual Templates: Linking Visual Features to Semantics". Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on.

Cherri, A. and M. Karim (1989). "Optical Symbolic Substitution- Edge Detection Using Prewitt, Sobel, and Roberts Operators." Applied optics 28(21): 4644-4648.

Chong, C-W., et al. (2003). "The Scale Invariants of Pseudo-Zernike Moments." Pattern Analysis & Applications 6(3): 176-184.

Choras, R. S. (2007). "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems." International journal of biology and biomedical engineering 1(1): 6-16.

Cucchiara, R., et al. (2001). "Improving Shadow Suppression in Moving Object Detection with HSV Color Information". Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE.

Dutta, S. (1993). Fuzzy logic applications: Technological and strategic issues.Engineering Management, IEEE Transactions on, 40(3), 237-254.

Eakins, J. P., et al. (2003). "Shape Feature Matching for Trademark Image Retrieval". Image and Video Retrieval, Springer: 28-38.

El-Feghi, I., et al. (2007). "Content-Based Image Retrieval Based on Efficient Fuzzy Color Signature". Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on.

El-Feghi, I., et al. (2011). "Efficient Features Extraction for Fingerprint Classification with Multi Layer Perceptron Neural Network". Systems, Signals and Devices (SSD), 2011 8th International Multi-Conference on.

El-ghazal, A., et al. (2008). "A Context-Based Fusion Algorithm for Shape Retrieval". Information Fusion, 2008 11th International Conference on.

Emmanuel, M., et al. (2007). "Content-Based Medical Image Retrieval". Information and Communication Technology in Electrical Sciences (ICTES 2007), 2007. ICTES. IET-UK International Conference on.

Everingham, M., et al. (2010). "The Pascal Visual Object Classes (VOC) Challenge" International Journal of Computer Vision 88(2): 303-338.

Falomir, Z., et al. (2010). "A Pragmatic Approach for Qualitative Shape and Qualitative Colour Similarity Matching". CCIA.

Ferschin, P., et al. (1994). "A Comparison of Techniques for the Transformation of Radiosity Values to Monitor Colors". Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, IEEE.

Flickner, M., et al. (1995). "Query By Image and Video Content: the QBIC System." Computer 28(9): 23-32.

Flusser, J., et al. (2009). "Moments and Moment Invariants in Pattern Recongition". USA, Wiley.

Ford, A. and A. Roberts (1998). "Colour Space Conversions." Westminster University, London 1998: 1-31.

Gonzalez, R. and R. Woods, Eds. (2002). "Digital Image Processing"

Gudivada, V. N. and V. V. Raghavan (1995). "Content Based Image Retrieval Systems." Computer 28(9): 18-22.

Gupta, A. and R. Jain (1997). "Visual Informaion Retrieval." RCommun. ACM 40(5): 70-79.

Gutta, S. and H. Wechsler (1998). "Facial Image Retrieval Using Hybrid Classifiers". Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on.

Haddadnia, J., et al. (2001). "Neural Network Based Face Recognition with Moment Invariants". Image Processing, 2001. Proceedings. 2001 International Conference on.

Hartigan, J. A. and M. A. Wong (1979). "Algorithm AS 136: A k-means Clustering Algorithm." Applied statistics: 100-108.

Heng, L. and Y. Jingqi (2007). "Multi-view Ear Shape Feature Extraction and Reconstruction". Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on.

Hoi, S., et al. (2006). "A Unified Log-Based Relevance Feedback Scheme for Image Retrieval." Knowledge and Data Engineering, IEEE Transactions on 18(4): 509-524.

Hong, D. H. and C.-H. Choi (2000). "Multicriteria Fuzzy Decision-Making Problems Based on Vague Set Theory." Fuzzy Sets and Systems 114(1): 103-113.

Hong, L., et al. (1998). "Fingerprint Image Enhancement: Algorithm and Performance Evaluation." Pattern Analysis and Machine Intelligence, IEEE Transactions on 20(8): 777-789.

Huang, J., et al. (1997). "Image Indexing Using Color Correlograms". Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, IEEE.

Hui, Z., et al. (2010). "Object Recognition by A Complete Set of Pseudo-Zernike Moment Invariants". Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on.

Iakovidis, D. K., et al. (2009). "A Pattern Similarity Scheme for Medical Image Retrieval." Information Technology in Biomedicine, IEEE Transactions on 13(4): 442-450.

Jain, A. K. (2010). "Data Clustering: 50 Years Beyond K-means." Pattern recognition letters 31(8): 651-666.

Jain, A. K., et al. (2000). "Statistical Pattern Recognition: A Review." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(1): 4-37.

Jain, A. K. and A. Vailaya (1998). "Shape-Based Retrieval: A Case Study with Trademark Image Databases." Pattern Recognition 31(9): 1369-1390.

Jaisakthi, S. M. and C. Aravindan (2009). "Face Detection Based on Eigenfaces and Legendre Moments" Tencon IEEE Region 10 Conference.

Jamil, N., et al. (2006). "Image Retrieval of Songket Motifs Using Simple Shape Descriptors". Geometric Modeling and Imaging--New Trends, 2006.

Jing, F., et al. (2005). "A Unified Framework for Image Retrieval Using Keyword and Visual Features." Image Processing, IEEE Transactions on 14(7): 979-989.

Ju, H. and M. Kai-Kuang (2002). "Fuzzy Color Histogram and its Use in Color Image Retrieval." Image Processing, IEEE Transactions on 11(8): 944-952.

Jung Uk, Cho, et al. (2007). "A Real-time Color Feature Tracking System Using Color Histograms". Control, Automation and Systems, 2007. ICCAS '07. International Conference on.

Kanungo, T., et al. (2002). "An Efficient K-means Clustering Algorithm: Analysis and Implementation." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(7): 881-892.

Kass, M., et al. (1988). "Snakes: Active Contour Models." International journal of computer vision 1(4): 321-331.

Kato, T. (1992). "Database Architecture for Content-Based Image Retrieval".

Kato, T., et al. (1992). "A Sketch Retrieval Method for Full Color Image Database-Query by Visual Example". Pattern Recognition, 1992. Vol.I. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on.

Kim, H.-K. and J.-D. Kim (2000). "Region-Based Shape Descriptor Invariant to Rotation, Scale and Translation." Signal processing: Image communication 16(1–2): 87-93.

Kiranyaz, S., et al. (2011). "Multi-Dimensional Evolutionary Feature Synthesis for Content-Based Image Retrieval". Image Processing (ICIP), 2011 18th IEEE International Conference on.

Klir, G. J. (1995). "Fuzzy Logic." Potentials, IEEE 14(4): 10-15.

Krishnan, N., et al. (2007). "Content Based Image Retrieval Using Dominant Color Identification Based on Foreground Objects". Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on.

Krishnapuram, R., et al. (2004). "Content-Based Image Retrieval Based on A Fuzzy Approach." Knowledge and Data Engineering, IEEE Transactions on 16(10): 1185-1199.

Kunttu, I., et al. (2004). "Multiscale Fourier Descriptor for Shape-Based Image Retrieval". Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on.

Laaksonen, J., et al. (2000). "Analyzing Low-Level Visual Features Using Content-Based Image Retrieval". Proc int'l conf neural information processing, Taejon.

Lazebnik, S., et al. (2006). "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories". Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, IEEE.

Li, M., et al. (2004). "Efficient Iris Recognition by Characterizing Key Local Variations." Image Processing, IEEE Transactions on 13(6): 739-750.

Li, S., et al. (2009). "Complex Zernike Moments Features for Shape-Based Image Retrieval." Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 39(1): 227-237.

Liu, P., et al. (2008). "A New Image Retrieval Method Based on Combined Features and Feature Statistic". Image and Signal Processing, 2008. CISP '08. Congress on.

Liu, Y., et al. (2007). "A survey of Content-Based Image Retrieval with High-Level Semantics." Pattern Recognition 40(1): 262-282.

Liwicki, M. and H. Bunke (2007). "Combining On-line and Off-line Systems for Handwriting Recognition". Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on.

Long, F., et al. (2003). "Fundamentals of Content-Based Image Retrieval". Multimedia Information Retrieval and Management, Springer: 1-26.

Lu, G. and A. Sajjanhar (1999). "Region-Based Shape Representation and Similarity Measure Suitable for Content-Based Image Retrieval." Multimedia Systems 7(2): 165-174.

Mallat, S. G. (1989). "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." Pattern Analysis and Machine Intelligence, IEEE Transactions on 11(7): 674-693.

Manjunath, B. S. and W. Y. Ma (1996). "Browsing Large Satellite and Aerial Photographs". Image Processing, 1996. Proceedings., International Conference on.

Manjunath, B. S. and W. Y. Ma (1996). "Texture Features for Browsing and Retrieval of Image Data" Pattern Analysis and Machine Intelligence, IEEE Transactions on 18(8): 837-842.

Manjunath, B. S., et al. (2002). "Introduction to MPEG-7: Multimedia Content Description Interface ", John Wiley & Sons.

Maria, P. and B. Panagiota (1999). "Image Processing: The Fundamentals." England: John Wiley & Sons Ltd.

McCaffrey, J. D. (2009). "Using the Multi-Attribute Global Inference of Quality (MAGIQ) Technique for Software Testing". Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on.

Mehrotra, R. and J. E. Gary (1995). "Similar-Shape Retrieval in Shape Data Management" Computer 28(9): 57-62.

Menasri, F., et al. (2007). "Shape-Based Alphabet for Off-line Arabic Handwriting Recognition". Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on.

Mianshu, C., et al. (2010). "Image Retrieval Based on Multi-Feature Similarity Score Fusion Using Genetic Algorithm". Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on.

Ming-Hsuan, Y., et al. (2002). "Detecting Faces in Images: A Survey". Pattern Analysis and Machine Intelligence, IEEE Transactions on 24(1): 34-58.

Ming-Kuei, H. (1962). "Visual Pattern Recognition by Moment Invariants." Information Theory, IRE Transactions on 8(2): 179-187.

Mokhtarian, F. and A. Mackworth (1986). "Scale-Based Description and Recognition of Planar Curves and Two-dimensional Shapes." Pattern Analysis and Machine Intelligence, IEEE Transactions on PAMI-8(1): 34-43.

Moore, B. (1981). "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction." Automatic Control, IEEE Transactions on 26(1): 17-32.

Mori, G., et al. (2001). "Shape Contexts Enable Efficient Retrieval of Similar Shapes". Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on.

Müller, H., et al. (2004). "A Review of Content-Based Image Retrieval Systems in Medical Applications—Clinical Benefits and Future Directions." International Journal of Medical Informatics 73(1): 1-23.

Müller, H., et al. (2001). "Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals." Pattern recognition letters 22(5): 593-601.

Nixon, M. and A. Aguado (2008). "Feature Extraction & Image Processing ", Academic Press.

Pass, G., et al. (1997). "Comparing Images Using Color Coherence Vectors". Proceedings of the fourth ACM international conference on Multimedia, ACM.

Pedrycz, W. and F. Gomide (2007). "Fuzzy Systems and Computational Intelligence ", Willy.

Pentland, A., et al. (1996). "Photobook: Content-Based Manipulation of Image Databases." Int. J. Comput. Vision 18(3): 233-254.

Peura, M. and J. Iivarinen (1997). "Efficiency of Simple Shape Descriptors". Proceedings of the third international workshop on visual form, Citeseer.

Phillips, P. J., et al. (2005). "Overview of the Face Recognition Grand Challenge". Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on, IEEE.

Qi, X. and Y. Han (2005). "A novel Fusion Approach to Content-Based Image Retrieval." Pattern Recognition 38(12): 2449-2465.

Rahman, M., et al. (2006). "Supervised Machine Learning Based Medical Image Annotation and Retrieval" in CLEF2005. Accessing Multilingual Information Repositories, Springer: 692-701.

Rao, A., et al. (1999). "Spatial Color Histograms for Content-Based Image Retrieval". Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on, IEEE.

Schlosser, S. and R. Beichel (2009). "Fast Shape Retrieval Based on Shape Contexts". Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on.

Seaborn, M., et al. (2005). "Fuzzy Colour Category Map for the Measurement of Colour Similarity and Dissimilarity." Pattern Recognition 38(2): 165-177.

Selvarajah, S. and S. R. Kodithuwakku (2011). "Combined Feature Descriptor for Content based Image Retrieval". Industrial and Information Systems (ICIIS), 2011 6th IEEE International Conference on.

Shahbaz Khan, F., et al. (2012). "Color Attributes for Object Detection". Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on.

Shokoufandeh, A., et al. (2002). "On the Representation and Matching of Qualitative Shape at Multiple Scales". Computer Vision—ECCV 2002, Springer: 759-775.

Shou, Z.-g., Wang, F., & Jia, J.-m. (2007). A Cumulative Satisfaction Measure Model Based on Dynamic Customer Expectation. Paper presented at the Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on.

Shum, H.-Y., et al. (1996). "On 3D Shape Similarity". Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on, IEEE.

Smeulders, A., et al. (2000). "Content-Based Image Retrieval at the End of The Early Years." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22(12): 1349-1380.

Smith, J. R. and S.-F. Chang (1996). "VisualSEEk: A Fully Automated Content-Based Image Query System". Proceedings of the fourth ACM international conference on Multimedia. Boston, Massachusetts, USA, ACM: 87-98.

Smith, J. R. and S.-F. Chang (1997). "Visually Searching the Web for Content." IEEE multimedia 4(3): 12-20.

stricker, M. and M. Orengo (1995). "Similarity of Colour Images." SPIE The international scociety for Optical Engineering.

Suryani, L. and L. Guojun (2003). "Spatial Statistics for Content Based Image Retrieval". Information Technology: Coding and Computing [Computers and Communications], 2003. Proceedings. ITCC 2003. International Conference on.

Swain, M. J. and D. H. Ballard (1990). "Indexing via Color Histograms". Computer Vision, 1990. Proceedings, Third International Conference on.

Takakura, J., et al. (2010). "Techniques to Enhance Images for Mokkan Interpretation". Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on.

Tamura, H., et al. (1978). "Textural Features Corresponding to Visual Perception." Systems, Man and Cybernetics, IEEE Transactions on 8(6): 460-473.

Tangelder, J. W. H. and R. C. Veltkamp (2004). "A survey of Content Based 3D Shape Retrieval Methods". Shape Modeling Applications, 2004. Proceedings.

Teague, M. R. (1980). "Image Analysis Via the General Theory of Moments." JOSA 70(8): 920-930.

Teh, C. H. and R. T. Chin (1988). "On Image Analysis by the Methods of Moments." Pattern Analysis and Machine Intelligence, IEEE Transactions on 10(4): 496-513.

Teh, C. H. and R. T. Chin (1989). "On the Detection of Dominant Points on Digital Curves." Pattern Analysis and Machine Intelligence, IEEE Transactions on 11(8): 859-872.

Tkalcic, M. and J. F. Tasic (2003). "Colour Spaces: Perceptual, Historical and Applicational Background". Eurocon.

Toygar, Ö. and A. Acan (2003). "Face Recognition Using PCA, LDA and ICA Approaches on Colored Images." Journal Of Electrical & Electronics Engineering 3(1): 735-743.

Turpin, A. and F. Scholer (2006). "User Performance Versus Precision Measures for Simple Search Tasks". Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. Seattle, Washington, USA, ACM: 11-18.

Umbaugh, S. E. (2010). "Digital Image Processing and Analysis: Human and Computer Vision Applications with CVIP Tools ", CRC press.

Wayman, J. L., A. K. Jain, et al. (2005). "Biometric Systems: Technology, Design and Performance Evaluation ", Springer Science & Business Media.

Wei, C.-H., et al. (2009). "Trademark Image Retrieval Using Synthetic Features for Describing Global Shape and Interior Structure." Pattern Recognition 42(3): 386-394.

Wenshuo, G., et al. (2010). "An Improved Sobel Edge Detection". Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on.

Wilbur, W. J. (1992). "An Information Measure of Retrieval Performance." Information Systems 17(4): 283-298.

Wang, X., et al. (2006). "Combining Discrete Orthogonal Moments and DHMMs for Off-line Handwritten Chinese Character Recognition". Cognitive Informatics, 2006. ICCI 2006. 5th IEEE International Conference on.

Xiaoling, W. and M. Hongyan (2009). "Enhancing Color Histogram for Image Retrieval". International Workshop on Information Security and Application.

Xu, Z. (2009). Fuzzy harmonic mean operators. International Journal of Intelligent Systems, 24(2), 152-172.

Yabuki, N., et al. (1999). "Region Detection Using Color Similarity". Circuits and Systems, 1999. ISCAS '99. Proceedings of the 1999 IEEE International Symposium on.

Yang, M., et al. (2008). "A survey of Shape Feature Extraction Techniques." Pattern Recognition: 43-90.

Yasmin, M. and S. Mohsin (2012). "Image Retrieval by Shape and Color Contents and Relevance Feedback". Frontiers of Information Technology (FIT), 2012 10th International Conference on.

Yi, B.-K., et al. (1998). "Efficient Retrieval of Similar Time Sequences Under Time Warping". Data Engineering, 1998. Proceedings., 14th International Conference on, IEEE.

Yining, D., et al. (2001). "An Efficient Color Representation for Image Retrieval." Image Processing, IEEE Transactions on 10(1): 140-147.

Yong-Xianga, S., et al. (2007). "Shape Feature Extraction of Fruit Image Based on Chain Code". Wavelet Analysis and Pattern Recognition, 2007. ICWAPR '07. International Conference on.

Yong, R., et al. (1997). "Content-Based Image Retrieval with Relevance Feedback in MARS". Image Processing, 1997. Proceedings., International Conference on.

Yong, R., et al. (1998). "Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval." Circuits and Systems for Video Technology, IEEE Transactions on 8(5): 644-655.

Zadeh, L. A. (1980). "Fuzzy Sets Versus Probability." Proceedings of the IEEE 68(3): 421-421.

Zagrouba, E., et al. (2007). "A Reliable Image Retrieval System Based on Spatial Disposition Graph Matching." International Review on Computers and Software 2: 108-117.
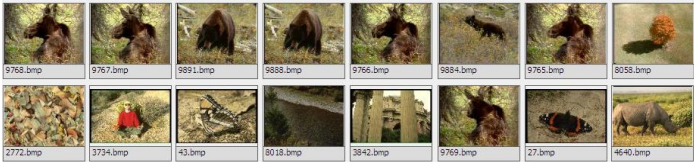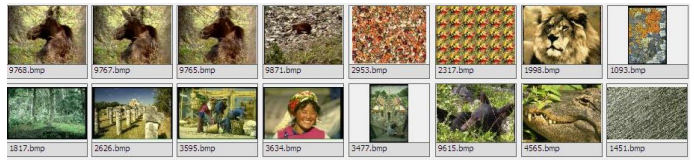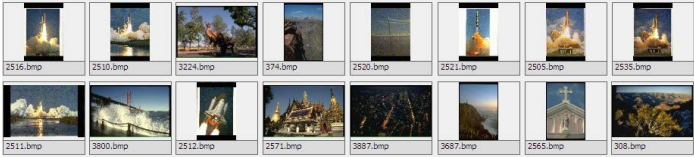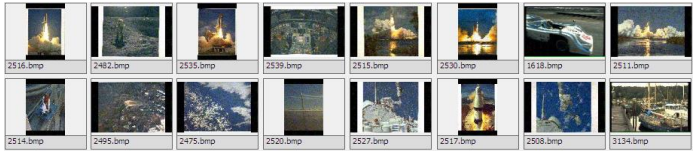
Zhang, D. and G. Lu (2004). "Review of Shape Representation and Description Techniques." Pattern Recognition 37(1): 1-19.

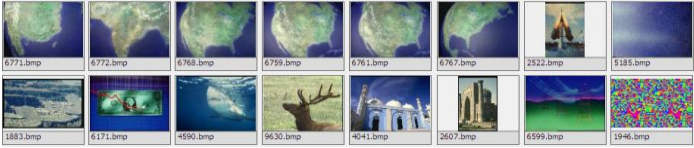Zimmermann, H. J. (2010). "Fuzzy Set Theory." Wiley Interdisciplinary Reviews: Computational Statistics 2(3): 317-332.
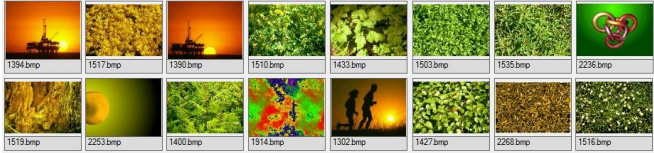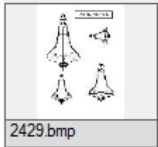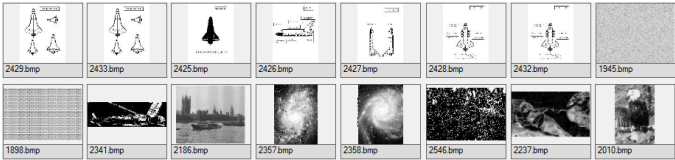
Appendix A: Representation of Pseudo-Zernike Moments (PZMs)

order (n)	Pseudo Zernike moment of order n with repetition l (PZM_{nl}) $l = -n, -n+1, -n+2, -n+3, \dots, n$	
0	PZM_{00}	Total moments = $(n+1)^2 = 101$
1	$PZM_{1-1}, PZM_{10}, PZM_{11}$	
2	$PZM_{2-2}, PZM_{2-1}, PZM_{20}, PZM_{21}, PZM_{22}$	
3	$PZM_{3-3}, PZM_{3-2}, PZM_{3-1}, PZM_{30}, PZM_{31}, PZM_{32}, PZM_{33}$	
4	$PZM_{4-4}, PZM_{4-3}, PZM_{4-2}, PZM_{4-1}, PZM_{40}, PZM_{41}, PZM_{42}, PZM_{43}, PZM_{44}$	
5	$PZM_{5-5}, PZM_{5-4}, PZM_{5-3}, PZM_{5-2}, PZM_{5-1}, PZM_{50}, PZM_{51}, PZM_{52}, PZM_{53}, PZM_{54}, PZM_{55}$	
6	$PZM_{6-6}, PZM_{6-5}, PZM_{6-4}, PZM_{6-3}, PZM_{6-2}, PZM_{6-1}, PZM_{60}, PZM_{61}, PZM_{62}, PZM_{63}, PZM_{64}, PZM_{65}, PZM_{66}$	
7	$PZM_{7-7}, PZM_{7-6}, PZM_{7-5}, PZM_{7-4}, PZM_{7-3}, PZM_{7-2}, PZM_{7-1}, PZM_{70}, PZM_{71}, PZM_{72}, PZM_{73}, PZM_{74}, PZM_{75}, PZM_{76}, PZM_{77}$	
8	$PZM_{8-8}, PZM_{8-7}, PZM_{8-6}, PZM_{8-5}, PZM_{8-4}, PZM_{8-3}, PZM_{8-2}, PZM_{8-1}, PZM_{80}, PZM_{81}, PZM_{82}, PZM_{83}, PZM_{84}, PZM_{85}, PZM_{86}, PZM_{87}, PZM_{88}$	
9	$PZM_{9-9}, PZM_{9-8}, PZM_{9-7}, PZM_{9-6}, PZM_{9-5}, PZM_{9-4}, PZM_{9-3}, PZM_{9-2}, PZM_{9-1}, PZM_{90}, PZM_{91}, PZM_{92}, PZM_{93}, PZM_{94}, PZM_{95}, PZM_{96}, PZM_{97}, PZM_{98}, PZM_{99}$	

Appendix B: FCH Query Images and their Retrieval Results Comparing to the CCH Results

Query Image	Using	Retrieval Results
	FCH	9768.bmp 9767.bmp 9891.bmp 9888.bmp 9766.bmp 9884.bmp 9765.bmp 8058.bmp 2772.bmp 3734.bmp 43.bmp 8018.bmp 3842.bmp 9769.bmp 27.bmp 4640.bmp
	CCH	9768.bmp 9767.bmp 9765.bmp 9871.bmp 2953.bmp 2317.bmp 1998.bmp 1093.bmp 1817.bmp 2626.bmp 3595.bmp 3634.bmp 3477.bmp 9615.bmp 4565.bmp 1451.bmp
	FCH	2516.bmp 2510.bmp 3224.bmp 374.bmp 2520.bmp 2521.bmp 2505.bmp 2535.bmp 2511.bmp 3800.bmp 2512.bmp 2571.bmp 3887.bmp 3687.bmp 2565.bmp 308.bmp
	CCH	2516.bmp 2482.bmp 2535.bmp 2539.bmp 2515.bmp 2530.bmp 1618.bmp 2511.bmp 2514.bmp 2495.bmp 2475.bmp 2520.bmp 2527.bmp 2517.bmp 2508.bmp 3134.bmp
	FCH	3060.bmp 3057.bmp 3099.bmp 3058.bmp 3059.bmp 3061.bmp 3054.bmp 3093.bmp 555.bmp 3084.bmp 3094.bmp 3053.bmp 3109.bmp 3039.bmp 3113.bmp 566.bmp
	CCH	3060.bmp 3057.bmp 3061.bmp 3211.bmp 968.bmp 1746.bmp 3386.bmp 3775.bmp 1883.bmp 3056.bmp 3091.bmp 512.bmp 1696.bmp 2638.bmp 8279.bmp 1086.bmp

Query Image	Using	Retrieval Results
6771.bmp	FCH	
	CCH	
	FCH	
	CCH	
	FCH	
	CCH	

Query Image	Using	Retrieval Results
1394.bmp (butterfly)	FCH	17.bmp, 6.bmp, 9.bmp, 13.bmp, 3.bmp, 19.bmp, 5033.bmp, 3859.bmp, 63.bmp, 91.bmp, 99.bmp, 1171.bmp, 83.bmp, 42.bmp, 328.bmp, 412.bmp
	CCH	17.bmp, 3871.bmp, 3921.bmp, 3816.bmp, 915.bmp, 204.bmp, 251.bmp, 2557.bmp, 3498.bmp, 3635.bmp, 1274.bmp, 74.bmp, 8709.bmp, 2590.bmp, 957.bmp, 1725.bmp
1394.bmp (oil rig sunset)	FCH	1394.bmp, 1390.bmp, 1513.bmp, 1965.bmp, 2314.bmp, 1395.bmp, 1330.bmp, 1855.bmp, 1532.bmp, 1927.bmp, 2046.bmp, 1185.bmp, 1383.bmp, 1384.bmp, 1382.bmp, 1517.bmp
	CCH	1394.bmp, 1517.bmp, 1390.bmp, 1510.bmp, 1433.bmp, 1503.bmp, 1535.bmp, 2236.bmp, 1519.bmp, 2253.bmp, 1400.bmp, 1914.bmp, 1302.bmp, 1427.bmp, 2268.bmp, 1516.bmp
2429.bmp (shuttle)	FCH	2429.bmp, 2433.bmp, 2425.bmp, 2426.bmp, 2427.bmp, 2428.bmp, 2432.bmp, 1945.bmp, 1898.bmp, 2341.bmp, 2186.bmp, 2357.bmp, 2358.bmp, 2546.bmp, 2237.bmp, 2010.bmp
	CCH	2429.bmp, 2433.bmp, 2426.bmp, 2428.bmp, 2432.bmp, 2425.bmp, 2427.bmp, 2127.bmp, 2116.bmp, 2123.bmp, 2047.bmp, 2341.bmp, 2532.bmp, 2060.bmp, 2303.bmp, 2404.bmp