A HYBRID FEATURE EXTRACTION SCHEME FOR IMAGE INDEXING AND RETRIEVAL

Image Indexing and Retrieval based on Hybrid Feature Extraction and Classification Using Image Processing and Machine Learning Algorithms

Baharum Baharudin

Submitted for the degree of
Doctor of Philosophy

Department of Electronic Imaging and Media Communications

University of Bradford
2005

Abstract

A HYBRID FEATURE EXTRACTION SCHEME FOR IMAGE INDEXING AND RETRIEVAL

Image Indexing and Retrieval based on Hybrid Feature Extraction and Classification Using Image Processing and Machine Learning Algorithms

Baharum Baharudin

Keywords

Content-based image retrieval, classification, backpropagation, support vector machines.

Everyday more images are being created, stored and transmitted. However these three acts themselves do not really pose serious problems. The problem becomes apparent when the stored images need to be retrieved. Query using the traditional text-based approaches, though simple and easy to implement, are no longer sufficient when considering the large volume of images that have to be manually labelled. A logical solution to this problem is to search for images based on its content. Thus Content-Based Image Retrieval (CBIR) was born. Since then, many systems have been developed either commercially or in the form of research prototypes. The heart of any CBIR system is feature extraction. In other words features extracted from the images (usually in the form of a vector representation) becomes the index by which the images will be searched. In terms of the number of features used to represent images, it is generally accepted that the use of multiple image features is more desirable than using a single feature. This is evident judging from the major CBIR systems that have been developed.

In this thesis, a hybrid feature extraction scheme is proposed based on a combination of features derived from the compressed as well as the pixel domain. By using two well-known classifiers; the Backpropagation Neural Network and Support Vector Machines, the performance of the proposed hybrid feature approach is compared with that of the other feature based approaches which serve as benchmarks. From the results obtained it has been shown that the hybrid feature extraction approach outperforms all the other feature based methods used in the experiments.

ACKNOWLEDGEMENTS

I would like to express my thanks to the following people who have made this thesis possible.

First and foremost I would like to thank Professor Jianmin Jiang who has been instrumental in guiding me throughout my research studies. During these years he has helped and assisted me in many ways. His patience, enthusiasm and his never-ending stream of ideas have been absolutely essential for the results presented in this thesis. I am very grateful that he has spent so much time with me especially in my first year of study.

Next, I would also like to thank Dr Stan S Ipson, for his patience and willingness in reading through my writing (both conference papers as well as thesis). I certainly hope that for all the time I have spent with him, some of his writing flair would have rubbed itself on me. Needless to say, this thesis would not have been what it is, if not for him.

My sincere thanks to Dr Rami S Qahwaji, for his insights and helpful advice into the field of machine learning. Working with him on this subject matter is certainly wonderful and a "learning" experience, one that I will always cherish.

I would also like express my appreciation to Dr Dave Hobbs, the support staff of the EIMC department, led by Mrs Tracie Tighe and to Miss Rona Wilson for their kind assistance in office and administrative matters. Special thanks goes to Dr Miles Marks for giving me the "space" to do my research work.

My presence at the University of Bradford would not have been possible were it not for the vision of my sponsors, the PETRONAS University, under the leadership of Dr. Rosti Saruwono. Thank you.

My list would not be complete without mentioning these group of special people. To my wife, Puteri Norhashimah, a special thank you accompanying me in this "journey" and making it into a bearable and enjoyable experience. Also to my special and "LOTOT" children; Nadia, Nadim, Najib and Najwa – where would I be without you guys.

Last but not least this thesis is dedicated to my "emak" and my late father. If it were not for them, I would not be here today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Motivation

"...(T)he first microprocessor only had 22 hundred transistors. We are looking at something a million times that complex in the next generations—a billion transistors. What that gives us in the way of flexibility to design products is phenomenal."

— Gordon E. Moore

In 1965, Intel co-founder Gordon Moore came up with a prediction which later became popularly known as Moore's Law. In that prediction he stated that the number of transistors on a chip doubles about every two years. That law still holds today as seen by the evident increase in computing power. Instead of using computers only for number crunching activities, they are also being used to mimic the human senses namely the senses of hearing, touch, taste, smell and sight,. In the development of the sense of hearing, the majority of work that has been carried is for use in speech recognition systems (Kumar et al. 2003; Podder et al. 2003; Rigoll 1994). Tactile sensors for measuring the parameters of contact between a sensor and an object have been developed to provide object manipulators like robots with a sense of touch (Kageyama et al. 1999; Krishna et al. 2004; Mukai 2004; Voyles et al. 1996). Artificial electronic tongues (ET) have also been developed to imitate the sense of taste as mentioned in (Cole et al. 2004; Gardener 2005; Hauptmann et al. 2000; Lindquist et al. 2001; Lvova et al. 2004). Research done by (Brezmes et al. 2005; Castro et al. 2003; Daqi et al. 2004; Hauptmann et al. 2000; Kermani et al. 1999; Nagle et al. 1998) is related to the sense of smell, with the development of electric noses. Computer vision is

the area of research associated with sense of sight. It is one of the categories under computer imaging as shown in Figure. 1.1.

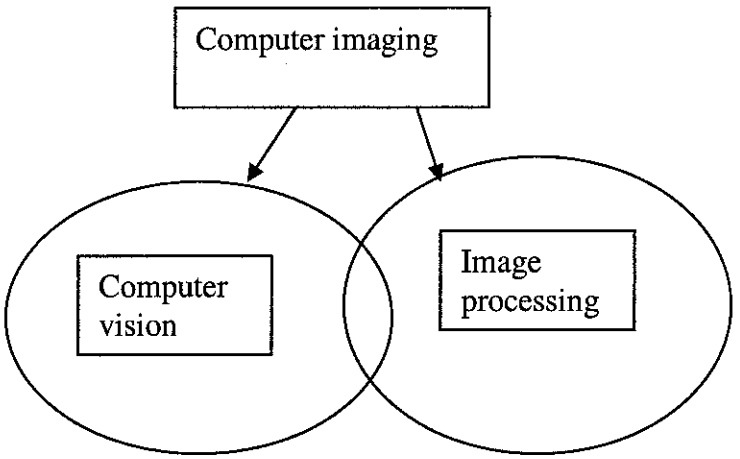


Figure 1. 1 Computer imaging separated into two overlapping areas

The main difference between the two is the way the output is being used. In the case of computer vision applications, the processed (output) images are for use by a computer, whereas in image processing applications, the output images are for human consumption.

Content-based image retrieval (CBIR), also known as query by image content (QBIC), and content-based visual information retrieval (CBVIR) are applications of computer vision to the retrieval of images that is the problem of searching for digital images in large databases. "Content-based" means that the search makes use of the contents of the images themselves, rather than relying on human-inputted metadata such as captions or keywords. A *content-based image retrieval system* (CBIRS) is a software implementation of CBIR. CBIR is a well-known and active research area which started in the late 1970's. Currently it is carried out in two major domains; the pixel domain and the compressed domain. The major problem faced by CBIR is in finding a suitable

image representation for use in indexing and retrieval which is the main aim of this thesis.

1.2 The aims and objectives of the research

The Aim

This project aims to explore the feasibility of overcoming the shortcomings of current techniques that are being used for the automatic retrieval of images by its contents, and the limitations of the existing image retrieval systems.

The Objectives

More specifically, the main objectives of the project are as follows:

- To investigate current techniques in image indexing and retrieval and address related problems.
- To research suitable image features that can be used for representing image content
- To research into an effective image indexing and classification scheme for generating image indices automatically.
- To design an effective and suitable interface for interactive image retrieval
- To develop a prototype image retrieval system based on the proposed approach.
- To evaluate the performance of the above approach.

1.3 Thesis overview

The thesis is organized as follows.

Chapter 2:

In this chapter, a historical perspective of CBIR is given. This is followed by a description of some of the major elements of a CBIR system namely the feature

extraction approach, the similarity measures used and the performance evaluation. Some examples of existing major CBIR systems are also discussed here. Major problems associated with CBIR are also highlighted in this chapter.

Chapter 3:

The emphasis of this chapter is on the work that has been carried out in the compressed domain by the author. The fundamental principles of the JPEG compressed domain are introduced in this chapter as a basis for the description of the author's work. The first part of the work involves a progressive decoding scheme for JPEG. In the second part, based on the JPEG coefficient coding categories, a feature extraction scheme in the JPEG compressed domain is presented.

Chapter 4:

This chapter represents an extension of the work presented in chapter 3. Based on the feature extraction scheme, a region-based approach to image retrieval is carried out. Four algorithms are developed and evaluated against two benchmarked algorithms. Basically an image is segmented into regions and the regions become the index key of the image. Here we show that an image can be represented by a single key or by multiple keys. Hence the four algorithms involve different representations of image keys.

Chapter 5:

In this chapter a hybrid feature extraction scheme is introduced where a combination of image features are used. Here, an image feature, developed and mentioned in chapter 3 is combined with two well known image features. The hybrid

feature is then compared with other features assigned as benchmark features. All the image features are subjected to two machine learning algorithms, namely the Backpropagation Neural network and the Support Vector Machines.

Chapter 6:

This chapter provides the conclusions of the thesis. A summary of the thesis contributions and some scope for future work are also outlined.

Chapter 2 Content-based image retrieval

2.1 Introduction

A content-based image retrieval (CBIR) system is a system that has the ability to automatically index images based on their visual content. This is done by having predefined methods for extracting visual features from an image (known as signatures), replacing the image by a feature vector and employing rules based on these signatures during the query-retrieval process. In the retrieval process, the system accepts a query image, extracts the appropriate feature vector and then carries out a search process. The search process is carried out by computing a predefined "similarity measure" between the feature vector of the query and those of the images in the database. The retrieved images are presented to the user in the descending order of the similarity to the query.

The chapter is organised as follows. In Section 2.2 the development of CBIR systems is presented. The major components of a CBIR system are discussed in Section 2.3. Feature extraction, which is the component on which the rest of the CBIR system relies, is reviewed in Section 2.4 followed by a summary of similarity measures in Section 2.5. Section 2.6 highlights the important issues regarding the measurement of the retrieval performance of a CBIR system. Some major state-of-the-art CBIR systems are described in Section 2.7, followed by a summary of the chapter in Section 2.8.

2.2 Development of CBIR

The year 1979, can be regarded as a starting point for image retrieval with the organisation of a conference on Database Techniques for Pictorial Applications (Blaser 1979), held in Florence. At that time the techniques employed were generally based on textual annotation of images and not on the visual features. In other words the images were first annotated or labelled with text and then searched using a text-based approach adopted from traditional database management systems. The annotation was done by human indexers. Surveys of early text-based image retrieval methods can be found in (Chang et al. 1992; Tamura et al. 1984).

As computers became more powerful, the use of images became more widespread. CBIR began to find applications in areas such as digital libraries, medical diagnosis and records, intellectual property, art galleries and museum management, architectural and engineering design, interior design, remote sensing and earth resource management, geographic information systems, scientific database management, weather forecasting, retailing, fabric and fashion design, trademark and copyright database management, law enforcement and criminal investigation, picture archiving and communication systems.

Although there has been rapid technological advances in image data capture and storage, the expertise and techniques for effective image retrieval has not kept pace with the technology of image production (Mostafa et al. 1996). The human ability to retrieve images is a complex and multifaceted issue. While it is generally easier to state what an image consists of in terms of the objects it contains, one of the main difficulties arises from the subjective, individual interpretation of the 'non-verbal symbolism' of an image

(Hidderly et al. 1997). This subsequently leads to the variability problems faced by human indexers. Three other problems related to the annotation process are the overhead costs of employing personnel for manual annotations, the storage and the time required to carry out the annotation. Figure 2.1 below illustrates the problem involved in manual annotation. The picture is taken after a flash flood in a city. If these facts are not known the picture may very well be taken to just represent a picture of a river flowing in a city. Work by (Enser 1993; Keister 1994) highlighted problems pertaining to visual representation.

Figure 2. 1 Flash flood in the city of Kuala Lumpur

In 1992, the National Science Foundation of the United States organised a workshop on visual information management systems (Jain 1992) to identify new directions for image database management systems. The principal new direction was to represent and index the visual images based on properties that are inherent in the images themselves. This signalled the birth of autonomous content-based image retrieval (CBIR). Subsequently much research work has been carried out as can be seen from several comprehensive surveys of the field (Furht et al. 1995; Mandal et al. 1999b; Rui et al. 1999; Smeulders et al. 2000)

2.3 Query types

A CBIR application is sometimes divided into three categories depending on the goal of the query:

- **Target search:** searching for a target image that is known to the user i.e. user knows what to look for.
- **Category search**: Searching for images from a particular category (e.g. a person wants to buy a shirt, but only has a general idea of the kind of texture and/or colour that he prefers).
- **General browsing**: searching for an item without really knowing what to look for, hence target can be very vague or even unknown.

2.4 Major components/functions of a CBIR system

The following are regarded as the basic components of a CBIR system

1. A user interface
2. A feature extraction scheme
3. A similarity measure

The user interface is the means by which a query is made, the display of the search results and for getting further input (in the case of relevance feedback). For example the query can be in the form of an example image, characteristics of an image or sketches. The second component, the feature extraction scheme, is the most important component as it is the one that generates the image signature on which the rest of the system depends for successful retrieval. The similarity measure is the next in importance. Both these components will be presented in more detail in the following subsections.

2.5 Feature Extraction

Image features (content) are the basis of CBIR as they provide a more useable representation of a particular image. In a broad sense, features may include both text-based features (key words, annotations) and visual features (colour, texture, shape). Within the visual feature scope, features can be further classified as broad or narrow (Smeulders et al. 1998). Broad or general features include colour, texture, and shape features while the narrow or domain specific features, are usually application-dependent and may include, for example, human faces and fingerprints. According to (Yang et al. 1999), due to image content varieties and diverse application subjectivity, there exists neither a universal feature for all images nor a single best representation for a given feature. In their work they concluded that the retrieval performance of features is task dependent whereby features like colour histograms and invariant feature histograms are suitable for databases of arbitrary colour photographs whereas for databases from a narrower domain, i.e. with clearly defined objects as content, the pixel values of the images in combination with a suitable distance measure are most important for good retrieval performance. The "best" features may change from image to image and from application to application.

Image indexing can be carried out in two well known domains; the pixel domain and the compressed domain. Some of the features used in the pixel domain are colour, texture, shape and sketch. The compressed domain can be further divided into two sub-domains; the transform domain and the spatial domain. Features used in the transform domain include the coefficients of the Discrete Fourier Transform (DFT), the Karhunen-Loeve Transform (KLT), the Discrete Cosine Transform (DCT) and Subbands/wavelets.

The Spatial domain uses Vector Quantization (VQ) and Fractals. A comprehensive survey on these domains can be found in (Mandal et al. 1999b).

2.6 Image features

The most commonly used image features are colour, shape and texture. The image features used can be either global or local in nature. A global descriptor uses visual features of the whole image, whereas a local descriptor uses the visual features of *regions* or *objects* to describe the image content.

2.6.1 Colour

Colour is the most extensively used image feature in image retrieval systems. Its three-dimensional values make its discrimination potentiality superior to the single dimensional grey values of images. Before selecting an appropriate colour description, the colour space should be determined. Some of the common colour spaces that are used are RGB, CIE Lab, CIE Luv and HSV. Some of the colour descriptors that have been implemented are colour histograms (Jeong et al. 2004; Nezamabadi-pour et al. 2004; Sawhney et al. 1994; Swain et al. 1991), colour coherence vectors (Pass et al. 1996), colour correlograms (Huang et al. 1997), and colour moments (Niblack et al. 1993).

2.6.2 Texture

Texture is another feature that has been widely investigated in pattern recognition and computer vision. Texture extraction techniques can be classified into two categories: *structural* and *statistical*. Structural methods, including *morphological operators* 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. Statistical methods, including *Fourier power spectra*, *co-occurrence*

matrices, *shift-invariant principal component analysis* (*SPCA*), *Tamura features*, *Wold decomposition*, *Markov random fields*, *fractal models*, and *multi-resolution filtering* techniques such as *Gabor and wavelet transforms*, characterize texture by their statistical distributions of image intensity. Some useful accounts of work on the use of texture can be found in (Bashar et al. 2003; Benazza-Benyahia et al. 2002; Chang et al. 1993; Chen et al. 1994; Lee et al. 2005).

2.6.3 Shape

Shape features of objects or regions have been used in many content-based image retrieval systems (Fudos et al. 2002; Gagaudakis et al. 2002; Guo et al. 2002a; Han et al. 2003; Kim et al. 2000). Shape feature extraction involves the segmentation of images into regions or objects. However, it should be noted that robust and accurate image segmentation is difficult to achieve, so much so that the use of shape features for image retrieval has been limited to special applications where objects or regions are readily available. A good shape representation feature for an object should be invariant to translation, rotation and scaling.

2.7 Similarity measures

Another important aspect of implementing image retrieval is determining the measure of similarity once a query has been submitted. The actual matching process can be seen as a search for images in the stored image set closest to the query specification. The similarity measure used depends on the types of features. In CBIR systems, image features are generally represented as n-dimensional feature vectors. Thus the query image and the database images can be compared by evaluating the distance between their corresponding feature vectors. Many kinds of distance functions (both general and specific) have been used in CBIR systems. The most common distance function is the

Euclidean distance. The Euclidean distance for two points $x = (x_1,...,x_n)$ and $y = (y_1,...,y_n)$ in Euclidean n-space is defined as

$$d(x,y) := \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)}^2$$

The cosine distance function is another type of measure that can be used for image retrieval. A property of the measure is that closer vectors give a higher value of the measure than more separated vectors. This is in contrast to the standard Euclidean distance where a higher value of the distance represents vectors that are further apart.

Colour histograms are typically used for representing colour distribution. Given a colour space defined by a number (usually three) of axes, the colour histogram is obtained by discretising the image colours and counting the frequency of each discrete colour that occurs in the image. Thus, the colours in the image are mapped onto a discrete colour space containing n colours. A colour histogram of image I is an n-dimensional vector, $H_j(I)$, where each element represents the frequency of colour j in image I. The histograms themselves are the feature vectors used as image indices. Statistically, it denotes the joint probability of the intensities of the three colour channels. Histograms are invariant to image rotation, translation and viewing axis (Swain 1993). The drawback in using a histogram is that it does not consider the perceptual similarity between the different bins. Apart from that the colour histogram also requires additional storage space and a large amount of processing. The computational complexity, can be decreased by reducing the number of bins. Other

variations of the colour histogram have also been developed. A histogram intersection, an L1 metric was proposed in (Swain et al. 1991), as the similarity measure for the colour histogram. (Niblack et al. 1993) introduced an L2-related metric in comparing the histograms. (Stricker et al. 1995) proposed the use of the cumulative colour histogram whereas the use of local histograms have been proposed by (Gong et al. 1994).

Another type of measure is the Hausdorff distance, attributed to Felix Hausdorff (1868 – 1942). It is a metric for two point sets that can be used for object detection as well as shape detection. Formally it is defined as follows.

For two sets of points $A = a_1, a_2, ...a_m$ and $B = b_1, b_2, ...b_n$

$$H(A,B) = \max(h(A,B), h(B,A))$$

where

$$h(A,B) = \max_{a \in A} \max_{b \in B} \| a - b \|$$

The Hausdorff distance has also been used to compare histograms as well as shapes in (Saber et al. 1997). The use of the Hausdorff distance measure can also be found in (Baudrier et al. 2004; Ko et al. 2002; Mukhopadhyay et al. 2004)

The Earth Mover Distance (EMD) (Rubner et al. 1997) is based on the old transportation problem (Hitchcock 1941). Suppose there are several *suppliers*, each with a given amount of goods. These suppliers are then required to supply several *consumers*, each with a given limited capacity. For each supplier-consumer pair, the cost of transporting a single unit of goods is given. The transportation problem is then to find the least-expensive flow of goods from suppliers to the consumers that satisfies the

consumers' demand. Similarly, signature matching can be regarded as a transportation problem by defining one signature as the supplier and the other as the consumer, and by setting the cost for a supplier-consumer pair equal to the ground distance between an element in the first signature and an element in the second. The solution to the signature matching problem then is to find the minimum distance between two elements, which can be further extended to finding the minimum distance between two sets or distribution. Research work involving the use of the EMD can be found in (Grauman et al. 2004; Greenspan et al. 2004; Surong et al. 2003; Yuan et al. 2003)

Another distance metric is the Mahalanobis distance developed by P. C. Mahalanobis in 1936. The basis of this measure is the correlations between variables by which different patterns can be identified and analysed. It is a useful way of determining *similarity* of an unknown sample set to a known one. Formally, the Mahalanobis distance from a group of values with mean $\mu = (\mu_1, \mu_2, ...\mu_p)$ and covariance matrix Σ for a multivariate vector $x = (x_1, x_2, ..., x_p)$ is defined as:

$$D(x) = \sqrt{(x-\mu)' \sum^{-1} (x-\mu)}$$

This measure has been used by (Schmid et al. 1997; Sebe et al. 2003) to measure the similarities between two invariant vectors. A survey of other types of distance measures can be found in (Puzicha et al. 1999) and (Kokare et al. 2003).

2.8 Measuring performance evaluation

A lot of effort has gone into CBIR research as can be seen by the many systems that have been developed so far. Unfortunately no way has been found to evaluate the effects of different techniques for image indexing, representation, and retrieval. In earlier CBIR systems, performances of a system were usually analysed in the form of one or more example queries. According to (Muller et al. 2001) this can be easily tailored to give a positive impression, since developers can choose queries which give the best results. Several review papers on the subject of performance evaluation have been written (Forsyth 2002; Gunther et al. 2001; Jermyn et al. 2002; Jorgensen 2001; Leung et al. 2000; Muller et al. 2001). Most, if not all, mentions the need for the following issues to be addressed. The first issue is that of a common database that can be used by all research groups. Ideally the images in these collections should be annotated for determining the relevancy of the top retrieved images. However, in order to compile a standard image collection the copyright issues involved in acquiring the image(s) have to be resolved. Image collections from Corel and Corbis, two major suppliers of digital image collection may be used, but the images are copyrighted and they are not free. Even then some of the groups that have access to the collection may use only a subset of the collection. The next issue is the use of a common performance measure. The most commonly accepted measure is precision and recall, a measure long used by the Information Retrieval (IR) community. Precision is defined as the ratio of the number of correct images retrieved to the number of images in the retrieved set, whereas recall is defined as the ratio of the number of correct images retrieved to the number of relevant images in the database. A recall of 1 implies that all images are retrieved.

Even then, using these values can lead to problems as described in (Forsyth 2002). A review of performance measures used for image retrieval can be found in (Muller et al. 2001). Currently there is an ongoing collaborative effort by the Benchathlon group (http://www.benchathlon.net) to develop a CBIR benchmarking environment.

2.9 CBIR systems

In this section some CBIR systems are to be discussed. The Query By Image Content (QBIC) System (Niblack et al. 1993) is an image retrieval system developed by IBM. It is regarded as the first commercial content-based image retrieval system and played an important part in the development of later systems. The QBIC system supports queries based on example images, user-constructed sketches and drawings, and selected colour and texture patterns, etc. The features used in this system are based on colour, texture and shape. A newer version of the system allows a text-based key word search which can be combined with a content-based similarity search.

MIT's Photobook (Pentland et al. 1994) is a set of interactive tools for browsing and searching images. The system consists of three subbooks namely; Appearance Photobook (face images), Texture Photobook and Shape Photobook from which face, texture and shape features are extracted. In its more recent version, relevance feedback is incorporated into the system. To perform a query, the user selects some images from a grid of still images displayed and/or enters an annotation filter. From the images displayed, the user can select another query images and repeat the search.

VisualSEEk (Smith et al. 1996b) is a visual feature search engine developed at Columbia University. The visual features used in their systems are a colour set and a

wavelet transform based texture feature. VisualSEEk supports queries based on both visual features and their spatial relationships. It also supports queries based on both keywords and visual content.

Netra (Ma et al. 1997) is a prototype image retrieval system developed by the University California, Santa Barbara (UCSB) Alexandria Digital Library (ADL) project. The system uses colour, texture, shape, and spatial location information in segmented image regions to search and retrieve similar regions from the database.

MARS (Rui et al. 1997) was developed at the University of Illinois at Urbana-Champaign (UIUC). The visual features used in this prototype are colour, texture and shape. It allows combined features queries using combinations of global or local image features with textual keywords associated with the images. It is an interactive system using relevance feedback

Virage is a content-based image search engine developed at Virage Inc (Bach et al. 1996). It supports visual queries based on colour, composition (colour layout), texture, and structure (object boundary information). A unique feature of the query system is that it allows combinations of the above four atomic queries. The users can adjust the weights associated with the atomic features to match their own emphasis.

WebSeek (Smith 1997) is a video and image cataloguing and retrieval system for the world-wide web developed at Columbia University. It automatically collects online visual material from the web and populates a database using an extendible subject taxonomy. Webseek uses text as well as colour features to index the visual

material. In carrying out a search query, users are allowed to manually modify an image colour histogram before repeating the search.

Blobworld (Carson et al. 2002) is an image retrieval system developed at UC Berkeley. It uses the Expectation Maximization (EM) algorithm to segment images into regions of uniform colour and texture (blobs). Image features used in this system are colour, texture and shape. In carrying out a query, the user first selects a category, thus limiting the search space. Given an initial image, the user selects a region (blob), and indicates the importance of the blob, the blob's color, texture, location, and shape More than one regions can be used for querying.

ImageRover (Scarlogg et al. 1997), is a system developed in Boston University. It combines textual and visual statistics into a single index for content-based search of a web image database. The visual statistics are based on colour and texture orientation histograms. To initiate a search of the ImageRover index, the user specifies a few keywords describing the desired images. Later the user can refine his query through relevance feedback.

PicSOM (Laaksonen et al. 2000; Laaksonen et al. 2002) is an image browsing system based on the Self-Organizing Map (SOM), developed at the Laboratory of Computer and Information Science at Helsinki University of Technology, Finland. The SOM is used to organize images into map units in a two-dimensional grid so that similar images are located near each other. The features are derived from colour, texture, shape and MPEG-7 features. By applying a tree-structured version of the SOM algorithm (Tree Structured Self-Organizing Map, (TS-SOM)) it creates a hierarchical

representation of the image database. During the queries, the TS-SOMs are used to retrieve images similar to a given set of reference images. Image retrieval with PicSOM is an iterative process utilizing the relevance feedback approach. A summary of the above-mentioned systems is provided in Table 2.1 below.

Table 2. 1 Summary of CBIR systems, image features and query methods

	CBIR	image features					query
	systems	colour	texture	shape	text	edge	features
1	QBIC	Yes	Yes	Yes	Yes	No	E,F,S,T
2	Photobook	Yes	Yes	Yes	No	No	RF
3	VisualSEEk	Yes	Yes	No	No	No	S,E,T
4	Netra	Yes	Yes	Yes	No	No	E,F
5	Mars	Yes	Yes	Yes	No	No	RF
6	Virage	Yes	Yes	Yes	Yes	No	E,F
7	WebSeek	Yes	Yes	No	Yes	No	E,F,RF
8	Blobworld	Yes	Yes	No	No	No	E,F
9	ImageRover	Yes	Yes	No	No	No	T,RF
10	PicSom	Yes	Yes	No	No	Yes	RF

E - Example
F – Selected features/ feature weights
S – Sketch
T – Text
RF – Relevance Feedback

2.10 Other approaches to image retrieval

Another approach to developing CBIR systems is by implementing an ontology-based approach. Though it is primarily used in text retrieval systems, its use can enhance the capabilities of a CBIR system as reported by (Chiang et al. 2001; Hollink et al. 2004; Pastra et al. 2003; Town et al. 2004; Zhuge 2004).

2.11 Summary

In this chapter the development of CBIR has been presented. The major components have also been outlined. In summary the problem faced in CBIR is a general one irrespective of the domain used. Finding the most suitable image

representation in the form of image features is of utmost importance in every system. Having selected a feature extraction scheme, the next issue is in the selection of similarity measure. Problems of comparing CBIR systems have also been discussed. The issues of having a common database and a common set of performance measures needs to be resolved before different CBIR systems can be compared. As mentioned earlier the domains and sub-domains in which the CBIR platform is developed is also wide in scope.

Chapter 3 Work done in the compressed domain

3.1 Introduction

As the use of images has become more widespread, storing images in the compressed domain has become very common. Apart from the savings in terms of storage space, it also reduces network costs in the transmission of files. However, the use of compressed files means that the files need to be decompressed before they can be viewed or processed further; in other words the data must be transformed back from the compressed domain to the pixel domain. Doing so, leads to a significant increase in the computing cost overhead as the process can be time consuming and it also increases the complexity of algorithm design and development. This negative aspect of compression usage has resulted in a new wave of research effort directed at techniques for working in the compressed domain. In the field of image retrieval, compressed domain indexing techniques can be broadly classified into two categories: transform domain techniques and spatial domain techniques (Mandal et al. 1999b). Transform domain techniques are generally based on the Discrete Fourier transform (DFT), Karhunen-Loeve transform (KLT), Discrete Cosine Transform (DCT) and Subbands/Wavelet transforms. Spatial domain techniques include vector quantisation (VQ) and fractals. The main focus of this chapter will be the DCT domain because it is used in JPEG compression.

The chapter is organised as follows. In Section 3.2 a basic introduction to JPEG compression is provided as JPEG compression is important to the whole work. Section

3.3 describes the joint work done with a member of my research group in developing a progressive decoding scheme (Guocan. et al. 2002). Section 3.4 describes a feature extraction scheme that has been carried out in the DCT domain. The first involves a decoding scheme for JPEG compressed images and the second involves an image feature extraction scheme. A summary of the chapter is provided in Section 3.5.

3.2 Basic JPEG Compression

JPEG is an image compression standard developed by the Joint Photographic Experts Group after which it is named. It can achieve compression ratios as high as 50 to 1 (Pennebaker et al. 1993). It features a simple lossy technique known as the Baseline method, which is currently the most widely implemented JPEG method. Figure 3.1 indicates the main processing steps involved in the baseline JPEG compression of an image.

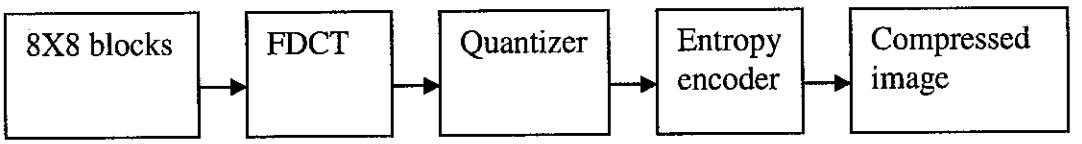


Figure 3.1 Baseline JPEG image compression

Assuming that the input is an 8-bit greyscale image, an image is first divided into non-overlapping blocks of 8×8 pixels. Since each block contains integer values in the range 0...255., each block then undergoes a normalization step by having the number 128 subtracted from each pixel, to bring the values into the range -128 to 127. The resulting block is then input to the Forward Discrete Cosine Transform (FDCT) producing a block of DCT coefficients. Within each DCT block, the coefficient at location (0,0) is called the DC coefficient and the other 63 coefficients are called AC coefficients. The DCT coefficients are fed into a Quantizer, in which each coefficient is divided by a constant value contained in a quantization table. The output from the quantizer is then

fed into the entropy encoder, where the quantized DC and AC coefficients are entropy coded separately. The DC coefficients, after differential pulse code modulation (DPCM), are encoded using Huffman coding. DPCM computes the difference between the two quantized DC coefficients in the current and previous DCT blocks. In the case of the AC coefficients, the two-dimensional array of AC coefficients is first transformed into a one-dimensional array using a zigzag scan ordering (see Figure 3.2). The resulting one-dimensional array of quantized AC coefficients are then converted into the "Run-Level" form by variable length coding (VLC). The run value is the number of zeros before a non-zero element in the one-dimensional array and the Level value is the value of a non-zero element. These results are then put through an AC Huffman encoder and output as a JPEG bit stream. A header is then combined with this stream to form a JPEG compressed file. The decoding process is just the inverse of the encoding procedure.

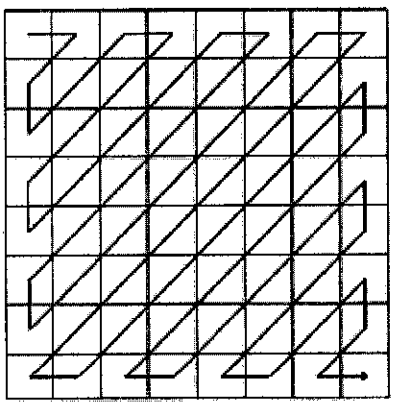


Figure 3.2 The JPEG zig-zag scan ordering sequence.

3.3 A progressive decoding scheme for JPEG compressed images

3.3.1 Introduction

The use of compressed images by internet applications is becoming widespread. Consequently, an important issue to pertaining to the proliferation of these compressed

images is to achieve the maximum efficiency in the transmitting and the decoding of these compressed images on the Internet (Armstrong et al. 2001; Helsingius et al. 2000). Traditional non-progressive transmission of images requires the complete set of data constituting an image to be transmitted before the viewer at the receiver's end can see the whole image. With a typical 28,800 bps modem connected to the internet, a large image can take up to several minutes to be displayed. Progressive image transmission (coding) can alleviate this problem by providing a coarse version of an image during the early stages of transmission, which is gradually refined by subsequent transmissions. Generally speaking, progressive coding is aimed at achieving (1) higher transmission rate or (2) higher speed decoding or both. For JPEG-compressed images, the conventional progressive decoding can be carried out using the methods of spectral selection or successive approximation as defined by the JPEG standard. They transmit spatial frequency information progressively so that an image grows clearer as more data is received. In more detail, contiguous coefficients in the zigzag sequence are grouped into bands, and each band is sent in a separate scan. In the early scan, only DC and a few lower frequency AC coefficients are transmitted, which provide a blurred but recognisable rendition of the image. However, conventional progressive coding involves the Inverse Discrete Cosine Transform (IDCT) which takes a lot of computational effort even at low bit rates. Though it can achieve good image quality, using it requires more computational cost than the JPEG baseline method at the receiving end, especially for the successive approximation approach. To overcome these shortcoming, we propose a comprehensive progressive decoding scheme based on applying the Taylor approximation of the IDCT function which will be described in Section 3.3.3.

3.3.2 Related work

Tong and Zhang (Tong et al. 1998) proposed an improved progressive transform scheme for decoding colour images based on the RGB colour space. In comparison with the traditional way of building up spatial resolution, their proposed method uses a different approach to construct colour information progressively, and to make the reconstruction of a sharp and clear gray scale image possible at an early stage of the transmission. However this scheme is only a framework which does not include the actual compression mechanism of the image data.

In the work done by (Manohar et al. 1999), the authors introduced a model-based vector quantization (MVQ), a variant of vector quantisation (VQ), into the compression of remotely sensed images. VQ is an asymmetric compression technique that has great potential for image data archival and distribution (Tilton et al. 1994). The advantage of using the MVQ is that it does not require codebook training as well as the storage and transmission of codebook as required in the use of conventional VQ. By making comparisons with other VQ techniques and with the JPEG/DCT approach, they declared their proposed coding scheme as an efficient and effective approach for disseminating image data across networks.

3.3.3 The progressive decoding design

As mentioned in Section 3.1, this was a joint work with a colleague in the research group. Specifically the author's contribution in this work are detailed as follows.

i. Carrying out Phase 2 of the experiment which involves assessing the quality of the decoded images from Phase 1. This is done by obtaining the Peak Signal-to-Noise Ratio (PNSR) values from the encoded image (Section 3.3.4.2)

ii. Carrying out Phase 3 of the experiment which involves finding the optimal number of coefficients. This is done by computing the PSNR values using coefficients varying in number from 1 to 64 along the zigzag route (Section 3.3.4.3).

Before going into the proposed decoding scheme design, an introduction to the Taylor approximation of IDCT is presented. Specifically, given the DCT coefficients, $C(u)$ ($u = 0,1,\ldots, 7$), of an 8-point 1D signal $x(i)$, its IDCT transform can be defined as

$$x(i) = \frac{1}{2}\sum_{u=0}^{7} \alpha(u)C(u)\cos\left(\frac{(2i+1)u\pi}{16}\right) \quad i=0, 1, \ldots,7 \qquad (1)$$

where the normalization factor $\alpha(u)$ satisfies the equation

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{2}} & for\ \ u = 0. \\ 1 & otherwise \end{cases}$$

As i and u vary, the decoded signal will be determined by the resulting angles inside the cosine function together with the DCT coefficients. In order to simplify the IDCT operation, the Taylor series can be applied to expand the cosine function as a power series to required order. Hence, approximations of the IDCT can be achieved to any number of linear or non-linear terms depending on the users' requirement. To minimize the error of such an approximation, however, the angle inside the cosine function must be small, the smaller the better. To this end, the expression $(2i+1)u$ (i=0,1,2..., 7; u=1,2,...,7) in (1) can be rearranged as

$$(2i+1)u = 8(4k+l) + \beta_{i,u} \qquad (2)$$

where $\beta_{i,u} = (2i+1)u$ **mod** 8; $l = (((2i+1)u - \beta_{i,u})$ **mod** 32)/8; $k = [(2i+1)u /32]$ and the operators **mod** and /, are modulus and integer division respectively. This rearrangement can be confirmed point by point when both i and u vary within [0, 7].

From equation (2), it can be seen that: $0 \le \beta_{i,u} < 8$ and $0 \le l < 4$. Therefore, we have:

$$\cos\left(\frac{(2i+1)u\pi}{16}\right) = \cos\left(\frac{8(4k+l)\pi + \beta_{i,u}\pi}{16}\right) = \begin{cases} \cos\frac{\beta_{i,u}\pi}{16} = \cos\frac{\gamma_{i,u}\pi}{16} & r_{i,u} = \beta_{i,u}, l = 0 \\ -\cos\frac{(8-\beta_{i,u})\pi}{16} = -\cos\frac{\gamma_{i,u}\pi}{16} & r_{i,u} = 8-\beta_{i,u}, l = 1 \\ -\cos\frac{\beta_{i,u}\pi}{16} = -\cos\frac{\gamma_{i,u}\pi}{16} & r_{i,u} = \beta_{i,u}, l = 2 \\ \cos\frac{(8-\beta_{i,u})\pi}{16} = \cos\frac{\gamma_{i,u}\pi}{16} & r_{i,u} = 8-\beta_{i,u}, l = 3 \end{cases}, = (-1)^{\left[\frac{l+1}{2}\right]}\cos(\frac{\gamma_{i,u}\pi}{16}) \qquad (3)$$

This equation essentially transfers the angle $\frac{(2i+1)u\pi}{16}$ into the first quadrant, and $\frac{\gamma_{i,u}\pi}{16}$ is an acute angle. The corresponding values of $\gamma_{i,u}$ ($\gamma_{i,u} = 0,1,\ldots,7$) can be worked out as given in equation (4), and $S^0_{i,u}$, the sign of $\cos\frac{(2i+1)u\pi}{16}$, is determined in equation (5).

$$\gamma_{i,u} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 3 & 6 & 7 & 4 & 1 & 2 & 5 \\ 0 & 5 & 6 & 1 & 4 & 7 & 2 & 3 \\ 0 & 7 & 2 & 5 & 4 & 3 & 6 & 1 \\ 0 & 7 & 2 & 5 & 4 & 3 & 6 & 1 \\ 0 & 5 & 6 & 1 & 4 & 7 & 2 & 3 \\ 0 & 3 & 6 & 7 & 4 & 1 & 2 & 5 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{pmatrix} \qquad (4)$$

$$S^0_{i,u} = sign\left\{\cos\frac{(2i+1)u\pi}{16})\right\} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 \\ 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix} \qquad (5)$$

where $sign\{x\} = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$. By applying the Taylor series, the function, $\cos\frac{\gamma_{i,u}\pi}{16}$ ($\gamma_{i,u} \neq 0$), can be expanded into a series of items at the point of $\pi/4$, which is given below:

$$\cos(\frac{\gamma_{i,u}\pi}{16}) = \sum_{k=0}^{n} \cos^{(k)}(\frac{\pi}{4}) \frac{(\frac{\gamma_{i,u}\pi}{16} - \frac{\pi}{4})^n}{n!} + R_n(\frac{\gamma_{i,u}\pi}{16}) \quad (6)$$

where R_n is the residue, which satisfies

$$|R_n| < \frac{\left(\frac{\pi}{4}\right)^{n+1} \left|\frac{\gamma_{i,u} - 4}{4}\right|^{n+1}}{(n+1)!} \leq \frac{\left(\frac{3\pi}{16}\right)^{n+1}}{(n+1)!} \quad (7)$$

according to the property of Taylor Series. If we only consider those items up to the 2nd order and ignore the rest, the equation (6) can be rewritten as

$$\cos(\frac{\gamma_{i,u}\pi}{16}) = \frac{1}{\sqrt{2}}\left(1 + \frac{\pi}{16}(4-\gamma_{i,u}) - \frac{\pi^2}{512}(4-\gamma_{i,u})^2\right) + R_2 \quad (8)$$

The residue, R_2, is determined from (7) as being less than 0.03. This suggests that the IDCT given in (1) can be approximated by a linear transform with an error less than 0.03. In most practical cases, such a small error is acceptable, considering the lossy nature of JPEG compression. In other words, the error introduced can be easily outstripped by the information loss introduced by quantization. Hence, the order-2 approximation of the IDCT can be specified as follows:

$$\begin{aligned} x(i) &= \frac{\sqrt{2}}{4}C(0) + \frac{1}{2}\sum_{u=1}^{7} C(u) S_{i,u}^{0} \cos\left(\frac{\gamma_{i,u}\pi}{16}\right) \\ &\approx \frac{\sqrt{2}}{4}\left(C(0) + \sum_{u=1}^{7} C(u) S_{i,u}^{0}\left(1 + \frac{\pi}{16}(4-\gamma_{i,u}) - \frac{\pi^2}{512}(4-\gamma_{i,u})^2\right)\right) \\ &= \frac{\sqrt{2}}{4}\sum_{u=0}^{7} C(u) S_{i,u}^{0} + \frac{\sqrt{2}\pi}{64}\sum_{u=1}^{7} C(u) S_{i,u}^{0}(4-\gamma_{i,u}) - \frac{\sqrt{2}\pi^2}{2048}\sum_{u=1}^{7} C(u) S_{i,u}^{0}(4-\gamma_{i,u})^2 \\ &= x^0(i) + x^1(i) + x^2(i) \end{aligned} \quad (9)$$

where $i \in [0,7]$, and $x^0(i)$, $x^1(i)$ and $x^2(i)$ represent the approximation of IDCT to the 0th order, 1st order and 2nd order respectively.

From equation (9), it is clear that the original signal $x(i)$ can be successively approximated either according to the number of DCT coefficients, which is similar to the JPEG spectral selection scheme, or according to the order of the Taylor expansion. By controlling the order of the items inside Taylor series, such progressively reconstructed signals, $x(i)$, from IDCT can be designed to have any level of accuracy up to the exact decoding of the JPEG normal decompression mode. To reduce computing costs and achieve the best possible efficiency, however, we will only consider the 0th, 1st

and 2nd order of terms when various approximations of IDCT are discussed and analyzed in the following paragraphs. The error incurred in doing so can be estimated via (7), which proves to be trivial in practice.

The detailed representations of $x^0(i)$, $x^1(i)$ and $x^2(i)$ can be further simplified as follows:

$$x^0(i) = \frac{\sqrt{2}}{4}\sum_{u=0}^{7} C(u)s_{i,u}^0 \tag{10}$$

$$x^1(i) = \frac{\sqrt{2}\pi}{64}\sum_{u=1}^{7} C(u)S_{i,u}^0\ (4-\gamma_{i,u}) = \frac{\sqrt{2}\pi}{64}\sum_{u=1}^{7} C(u)S_{i,u}^1 \tag{11}$$

$$x^2(i) = -\frac{\sqrt{2}\pi^2}{2048}\sum_{u=1}^{7} C(u)S_{i,u}^0(4-\gamma_{ui})^2 = -\frac{\sqrt{2}\pi^2}{2048}\sum_{u=1}^{7} C(u)S_{i,u}^2 \tag{12}$$

At the 0th order, the approximation given in (10) is similar to the Walsh-Hadamard transform. Although the transform is not as efficient as the DCT in terms of energy compaction, it can outline the silhouette of the original signal, which will be shown by latter experiments presented in the next section.

By introducing matrix forms $\boldsymbol{x}^0$=[$x^0(0)$, $x^0(1)$, ..., $x^0(7)$]^T and $\boldsymbol{C}$=[$C(0)$,$C(1)$, ..., C (7)]^T for the input signal and its DCT coefficients respectively, the above approximation of the IDCT can be represented in matrix form as follows:

$$\mathbf{x} \approx \mathbf{x}^0 + \mathbf{x}^1 + \mathbf{x}^2 = \left(\frac{\sqrt{2}}{4}\mathbf{S}_{i,u}^0 + \frac{\sqrt{2}\pi}{64}\mathbf{S}_{i,u}^1 + \frac{\sqrt{2}\pi^2}{2048}\mathbf{S}_{i,u}^2\right)\mathbf{C} \tag{13}$$

where $s_{i,u}^1$ and $s_{i,u}^2$ can be worked out from (11-12) as:

$$S_{i,u}^1 = \begin{pmatrix} 0 & 3 & 2 & 1 & 0 & -1 & -2 & -3 \\ 0 & 1 & -2 & 3 & 0 & -3 & -2 & 1 \\ 0 & -1 & 2 & -3 & 0 & -3 & 2 & 1 \\ 0 & -3 & -2 & 1 & 0 & 1 & 2 & -3 \\ 0 & 3 & -2 & -1 & 0 & -1 & 2 & 3 \\ 0 & 1 & 2 & 3 & 0 & 3 & 2 & -1 \\ 0 & -1 & -2 & -3 & 0 & 3 & -2 & -1 \\ 0 & -3 & 2 & -1 & 0 & 1 & -2 & 3 \end{pmatrix} \tag{14}$$

$$S^2_{i,u} = \begin{pmatrix} 0 & -9 & -4 & -1 & 0 & -1 & -4 & -9 \\ 0 & -1 & -4 & 9 & 0 & 9 & 4 & 1 \\ 0 & -1 & 4 & 9 & 0 & -9 & -4 & -1 \\ 0 & -9 & 4 & 1 & 0 & -1 & 4 & 9 \\ 0 & 9 & 4 & -1 & 0 & 1 & 4 & -9 \\ 0 & 1 & 4 & -9 & 0 & 9 & -4 & 1 \\ 0 & 1 & -4 & -9 & 0 & -9 & 4 & -1 \\ 0 & 9 & -4 & 1 & 0 & 1 & -4 & 9 \end{pmatrix} \tag{15}$$

As can be seen in (5), all elements in $S^0_{i,u}$ take the value of 1, 0 or –1. This means that there exist only additions in the 0^{th} approximation. For higher order approximations, however, multiplication is involved since the elements of $S^1_{i,u}$ and $S^2_{i,u}$ are not limited to 1, 0, or –1. In order to reduce the number of multiplications in our progressive coding design, all the elements inside the matrix $S^1_{i,u}$ and $S^2_{i,u}$ are rearranged so their non-zero values are made to have either 1s or –1s. To this end, the matrix $S^1_{i,u}$ and $S^2_{i,u}$ are decomposed as follows:

$$S^1_{i,u} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & -1 & 0 & -1 & 0 & 1 & 0 & 1 \end{pmatrix} + 2 \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & -1 & -1 & 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & -1 & 1 \end{pmatrix} = S^{10}_{i,u} + 2S^{11}_{i,u} \tag{16}$$

$$S^2_{i,u} = \begin{pmatrix} 0 & -1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} + 4 \begin{pmatrix} 0 & -1 & -1 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & -1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & -1 & 1 \end{pmatrix} + 4 \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$= S^{20}_{i,u} + 4S^{21}_{i,u} + 4S^{22}_{i,u} \tag{17}$$

Substituting equations (16) and (17) into equation (13), we have:

$$\begin{aligned} \mathbf{x} &\approx \left(\frac{\sqrt{2}}{4}\mathbf{S}^0_{i,u} + \frac{\sqrt{2}\pi}{64}\mathbf{S}^{10}_{i,u} + \frac{\sqrt{2}\pi}{32}\mathbf{S}^{11}_{i,u} + \frac{\sqrt{2}\pi^2}{2048}\mathbf{S}^{20}_{i,u} + \frac{\sqrt{2}\pi^2}{512}\mathbf{S}^{21}_{i,u} + \frac{\sqrt{2}\pi^2}{512}\mathbf{S}^{22}_{i,u} \right)\mathbf{C} \\ &= \frac{\sqrt{2}}{4}\mathbf{S}^0_{i,u}C + \frac{\sqrt{2}\pi}{64}\mathbf{S}^{10}_{i,u}C + \frac{\sqrt{2}\pi}{32}\mathbf{S}^{11}_{i,u}C + \frac{\sqrt{2}\pi^2}{2048}\mathbf{S}^{20}_{i,u}C + \frac{\sqrt{2}\pi^2}{512}\mathbf{S}^{21}_{i,u}C + \frac{\sqrt{2}\pi^2}{512}\mathbf{S}^{22}_{i,u}C \\ &= J1 + J2 + J3 + J4 + J5 + J6 \end{aligned} \tag{18}$$

Equation (18) presents a re-organised approximation of IDCT, in which all DCT coefficients can be progressively decoded or the signal value x can be progressively reconstructed according to the order of $\mathbf{s}_{i,u}^{0}$, $\mathbf{s}_{i,u}^{10}$, $\mathbf{s}_{i,u}^{11}$, $\mathbf{s}_{i,u}^{20}$, $\mathbf{s}_{i,u}^{21}$ and $\mathbf{s}_{i,u}^{22}$. Consequently, a progressive decoding scheme can be designed in terms of the 6-step decoding corresponding to $\mathbf{s}_{i,u}^{0}$, $\mathbf{s}_{i,u}^{10}$, $\mathbf{s}_{i,u}^{11}$, $\mathbf{s}_{i,u}^{20}$, $\mathbf{s}_{i,u}^{21}$ and $\mathbf{s}_{i,u}^{22}$. For the convenience of description, we denote these steps as $\mathbf{J}1$, $\mathbf{J}2$, ... $\mathbf{J}6$ in equation (18).

To estimate the complexity of this decoding, we list all operations of all order in comparison with IDCT as shown in Table 3.1. Table 3.1 shows that though the total addition is more than that of the conventional IDCT, the multiplication is much less, only half of that of the conventional IDCT. This greatly meets to reduce the requirements of computational complexity.

Table 3.1 The complexity of successive approximation in comparison with IDCT

	$S^{0}_{i,u}$ (J1)	$S^{10}_{i,u}$ (J2)	$S^{11}_{i,u}$ (J3)	$S^{20}_{i,u}$ (J4)	$S^{21}_{i,u}$ (J5)	$S^{22}_{i,u}$ (J6)	IDCT
+	28	8	14	8	14	4	
×	8	4	6	4	6	2	
Accumulation +	28	44	66	82	104	116	56
Accumulation ×	8	12	18	22	28	30	64

For two-dimensional images, the property of separation in DCT/IDCT can be exploited to reach similar conclusions to those for 1D signals. Starting from the definition of 2D IDCT,

$$x(i,j)=\frac{1}{4}\sum_{u=0}^{7}\sum_{v=0}^{7}\alpha(u)\alpha(v)C(u,v)\cos\left(\frac{(2i+1)u\pi}{16}\right)\cos\left(\frac{(2j+1)v\pi}{16}\right)$$
$$=\frac{1}{2}\sum_{u=0}^{7}\alpha(u)\left(\frac{1}{2}\sum_{v=0}^{7}\alpha(v)C(u,v)\cos\frac{(2j+1)v\pi}{16}\right)\cos\frac{(2i+1)u\pi}{16} \qquad (19)$$

equation (18) can be applied along the column and row directions respectively to reach similar simplifications. If we consider equation (19) in matrix form similar to that of (18), we would have the two-dimensional signal $x(i,j)$ being approximately reconstructed as follows:

$$\mathbf{x}=(x(i,j))\approx(\frac{\sqrt{2}}{4}\mathbf{S}^{0}_{j,v}+\frac{\sqrt{2}\pi}{64}\mathbf{S}^{1}_{j,v}+\frac{\sqrt{2}\pi^{2}}{2048}\mathbf{S}^{2}_{j,v})C(u,v)(\frac{\sqrt{2}}{4}\mathbf{S}^{0}_{i,u}+\frac{\sqrt{2}\pi}{64}\mathbf{S}^{1}_{i,u}+\frac{\sqrt{2}\pi^{2}}{2048}\mathbf{S}^{2}_{i,u})^{T}$$
$$=\left(\frac{\sqrt{2}}{4}\mathbf{S}^{0}_{j,v}C(*,v)+\frac{\sqrt{2}\pi}{64}\mathbf{S}^{1}_{j,v}C(*,v)+\frac{\sqrt{2}\pi^{2}}{2048}\mathbf{S}^{2}_{j,v}C(*,v)\right)(\frac{\sqrt{2}}{4}\mathbf{S}^{0}_{i,u}+\frac{\sqrt{2}\pi}{64}\mathbf{S}^{1}_{i,u}+\frac{\sqrt{2}\pi^{2}}{2048}\mathbf{S}^{2}_{i,u})^{T} \qquad (20)$$
$$=C'(u,*)(\frac{\sqrt{2}}{4}\mathbf{S}^{0}_{i,u}+\frac{\sqrt{2}\pi}{64}\mathbf{S}^{1}_{i,u}+\frac{\sqrt{2}\pi^{2}}{2048}\mathbf{S}^{2}_{i,u})^{T}$$
$$=(J_{i,u}1+J_{i,u}2+J_{i,u}3+J_{i,u}4+J_{i,u}5+J_{i,u}6)$$

Equation (20) essentially represents the application of equation (18) twice to equation (19), firstly along the column direction via six steps (J1-J6) to derive $C'(u,*)$ and secondly along the row direction via another six steps to derive $x(i,j)$. To

distinguish the six-step progressive operations on rows from that on columns, $J_{i,u}$ is used to specify the operations relating to i and u in (20).

In summary, the core of the progressive decoding algorithm is designed to have six steps, J1 to J6, applied twice along the row direction and column direction respectively. During the process, the order-0 decoding is implemented in a single step (J1), order-1 decoding in two steps (J2 and J3), and order-2 decoding in three steps (J4, J5 and J6). In a specific decoding design, the progression can be made either in terms of the orders in Taylor series approximation (successive approximation), or in terms of spectral selection relating to the number of DCT coefficients used. Via spectral selection, the progressive decoding does not only reduce the computing cost further, but also reduce the quantity of information being transmitted. When only DC coefficients are received, for example, the initial pixel information can be built up by two single-step decoding in accordance with the matrix $S^0_{j,v}$ and $S^0_{i,u}$. As more AC coefficients $ZZ(k_1)$-$ZZ(k_2)$ $(1 \le k_1 \le k_2 \le 63)$ along the zig-zag route are received during the progression, better quality images can be reconstructed via progressive decoding with higher order approximation matrices. Since successive approximation is considerably more effective than spectral selection at very low bit rates (Pennebaker et al. 1993), however, both successive and spectral selection should be considered in designing the progressive codec algorithm, in order to achieve the best possible performances. To this end, our progressive decoding algorithm is proposed as follows:

*Step 1. Receiving DC coefficient (ZZ(0)) coefficient only, and decode it using **J1** to construct a lowest quality image for further progression;*

*Step 2. Receiving a certain number of AC coefficients $ZZ(k_1)$-$ZZ(k_2)$ $(1 \le k_1 \le k_2 \le 63)$, and decode them using **J1**;*

Step 3. Decoding ZZ(k_1)-ZZ(k_2) ($1 \le k_1 \le k_2 \le 63$) using $\boldsymbol{J2\text{-}J_r}$ *($r \le 6$), depending on the users' requirements.*

Step 4. Repeat Step 2 until all coefficients and all steps, ***J1-J6****, are completed.*

In the above algorithm design, step1 and step 2 covers the progressive design for both compression and decompression. In fact, when only DC coefficients are compressed and transmitted, the decoding end can only have step-1 operation, and thus the progressive decoding has to wait for further data from the encoder. In addition, our extensive experiments show that, when $k_2 \le 5$, the quality of the decoded image by **J1** is already close to that of **J2-J6**. In other words, further progressive decoding of ZZ(1)-ZZ(k_2) by J2-J6 would not make much difference in terms of the image quality, when compared with that of the image reconstructed by J1. Similarly, when $k_2 \le 32$, our progressive decoding by **J3** already reconstructs the image close to that of **J4-J6**. Therefore, the progressive decoding can virtually be terminated at **J3** in this circumstance. Detailed analyses of experimental results are presented in the next section.

3.3.4 Experimental Analysis

In carrying out the experiment the *pepper* image, readily available in the public domain is used. The analysis of the proposed algorithm was carried out in three phases, as described in the following subsections.

3.3.4.1 Phase 1

In the first phase of the experiment, a JPEG compressed image is decoded using both the proposed scheme J1 – J6 and the spectral selection algorithm with 1, 4, 16 and

64 coefficient(s). The experimental results for image sample, *pepper* are displayed in Figures 3.3a. where all reconstructed image samples at each level of decoding are displayed for visual inspection. Images from row 1 to row 5 correspond to the successive approximation J1-J5. The images from column 1 – 4 are decoded with 1, 4, 16 and 64 coefficient(s). From Figure 3.3a, it can be seen from visual inspection, that the reconstructed images do not differ significantly in using the five algorithms (J1 – J5). To illustrate further, the reconstructed images using decoding levels J6 and IDCT are enlarged as in Figure 3.3b. Here, too, by carrying out a visual comparison, it can be seen that the quality of the reconstructed images using decoding level J6 though not as sharp compared with that which have been decoded using conventional IDCT, does not detract in terms of its overall appearance. Hence, it can be said that the same quality of reconstructed images can still be obtained using the decoding levels J1 – J6, but at a savings in terms of the computational costs involved when compared with the existing spectral selection method as shown in Table 3.1.

Figure 3.3a Reconstructed image using decoding levels J1 – J5 to various number of coefficients

Figure 3.3b Reconstructed image using decoding levels J6 and IDCT to various number of coefficients

3.3.4.2 Phase 2

In the second phase of the experiment, the quality of the decoded images was assessed in terms of Peak Signal-to-Noise Ratio (PSNR) values. The PSNR values for J1 – J6 at 64 coefficient level were calculated for the *pepper* image samples. The corresponding PNSR value for the IDCT approach are also calculated as comparison and the results are shown in Table 3.2. The table clearly illustrates how the quality has been improved by each approximation.

Table 3.2 PSNR for pepper image for J1-J6 and IDCT

	J1	J2	J3	J4	J5	J6	IDCT
PSNR	29.78	32.59	40.25	41.27	43.29	53.78	58.92
		2.81	7.67	1.02	2.02	10.49	5.15
		11.60	5.30	40.50	41.40	5.10	11.40

3.3.4.3 Phase 3

In this phase, the optimal number of coefficients was determined by computing the PNSR values using coefficients varying in number from 1 to 64 along the zigzag route. The experimental values are shown in Figure 3.4. The figure shows that J1 and J2 are almost stable after 16 coefficients and that J6 has almost the same performance as IDCT.

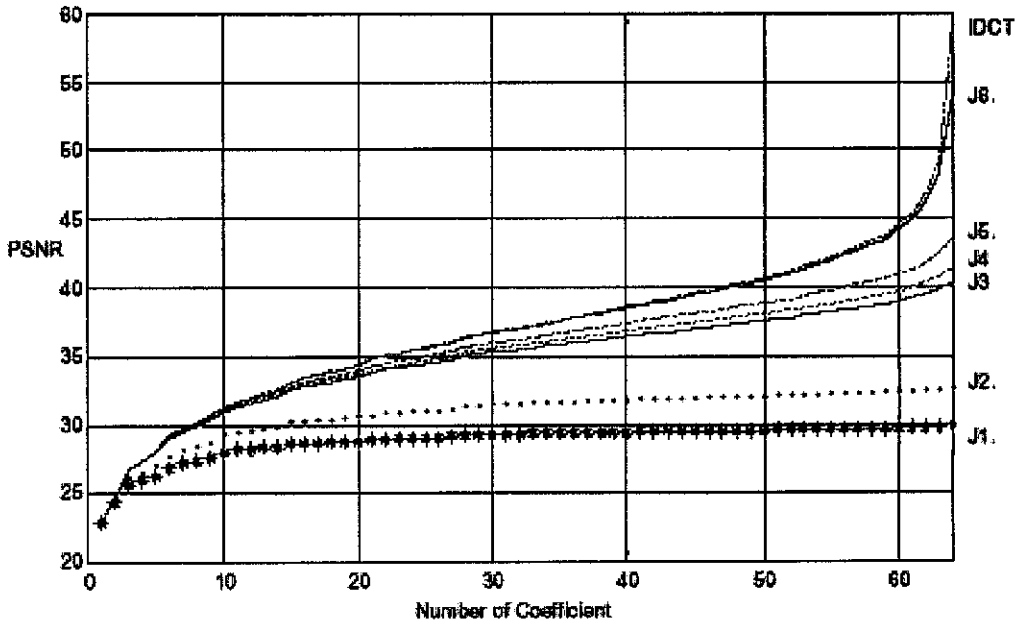


Figure 3.4 Comparison between J1 – J6 and IDCT with respect to the number of coefficients

3.4 Feature extraction using JPEG Coefficient Coding Categories.

3.4.1 Introduction

In this section, a feature extraction scheme using the JPEG coefficient coding categories is proposed. Before going into the details of the feature extraction scheme which will be described in section 3.4.3, a review of related work will be described in section 3.4.2. The experimental design is described in section 3.4.4. Results of the experiments are presented in section 3.4.5.

3.4.2 Related work

As mentioned earlier, in the field of image retrieval, compressed domain indexing techniques can be broadly classified into two categories: transform domain techniques and spatial domain techniques (Mandal et al. 1999b). Although several transform domains are used in the published literature, since the main focus of this

chapter is on the DCT domain, the present survey reviews only work related to that carried out in the DCT domain. Specifically the references made will be those pertaining to the various image indexing and retrieval schemes.

In the work done by (Chang et al. 2004) two features are extracted, a DC feature and an AC feature. The DC feature is calculated from the difference between the quantized DC coefficients of the current block and the previous block. Thus for a given JPEG image of size n x m, where both n and m are multiples of 8, the total number of difference values is (n x m)/(8 x 8). These difference values are then stored in a one-dimensional array in the form of 1's and 0's whereby a 1 is assigned if the difference value is greater than or equal to zero and zero, otherwise. A similar approach is applied for extracting the AC features whereby the first 9 quantized AC coefficients in zigzag order are used. However a limitation of this algorithm is that it only works on sets of images which are all of the same size.

In their work, Climer and Bhatia proposed a method whereby the DCT coefficients of an image are organised into a quad tree structure (Climer et al. 2002). The quad tree is a tree structure in which each node has has either four children or no children at all. If a node has no children, it is called a leaf node; otherwise it is called an internal node. An image feature is then represented by the coefficients on the nodes of the quadtree. During a query process, the query image is processed to produce a quad tree and the roots of the trees in the index are compared with the query root using a distance formula. For distances within a given tolerance, the second level of the corresponding trees are compared. This process repeats down to the leaf nodes and selected images are ordered and returned to the user. Although the system can

effectively extract features from DCT coefficients, the main drawback of this method is that the computation of the distance between images will grow undesirably fast when the number of irrelevant images is big or the threshold value is large.

Feng and Jiang proposed a statistical-parameter-based method (Feng et al. 2003) for retrieving JPEG compressed images. This method is based on statistical features which are computed directly from DCT coefficients without involving a full decoding of JPEG decompression, or inverse DCT. In this case, the mean and variance of the original pixel values in an 8×8 block are derived directly from the DC and AC coefficients. The system then constructs a two-dimensional $m1-\sigma$- space, which satisfies the requirements $0 \le m1 \le 255$ and $0 \le \sigma \le 128$. The mean and variance values are quantized into four non-overlapping sections, labelled 0,1,2,3 respectively, and σ is divided unequally into seven non-overlapping sections labelled 0,1,2,...6. thus forming a vector of 28 elements. As each block is processed, the mean and variance values obtained are classified into one of the 28 subspaces. The final result is a histogram of 28 elements that can be used to characterize an image content.

A DCT-based texture discrimination technique has been proposed by (Reeves et al. 1997) whereby an image is represented by a feature vector formed from the variance of the first 8 AC coefficients. This technique assumes that the eight AC coefficients are the most discriminating coefficients. Hence the run-time complexity of this technique is small, since the length of the feature vector is small.

A method called multi-resolution reordered DCT (MRDCT) was developed by (Huang et al. 1999) in which DCT coefficients are reordered to produce image sub-bands in a multi-resolution decomposition-like form. The absolute mean value and the

standard deviation of each sub-band are then used to construct the feature vector. One limitation of MRDCT is that all the DCT sub-bands were equally treated without discrimination, which might cause inaccuracies while processing

In the work by (Ng et al. 1992) the authors proposed a segmentation technique using the local variance of the DCT coefficients. In this technique, a 3 x 3 DCT is computed at each pixel location using the eight surrounding pixels. The local variance of each DCT coefficient is then computed using a 15 x 15 sliding window. Changes in the local variance are used to segment the image.

Shneier et al.(Shneier et al. 1996) proposed a technique for image retrieval of JPEG compressed images. This technique is based on the mutual relationship between the DCT coefficients of unconnected regions in both the query and target image. Here, a set of 2K windows is selected, and is randomly paired, producing K pairs of windows. For each window, the average of each DCT coefficient is computed resulting in a 64-dimensional feature vector (f). The feature vectors corresponding to a pair of windows are compared and each pair of components is assigned a bit (0 or 1) depending on their similarity. Thus each pair of windows will be assigned 64 bits. The similarity of the query and target image is determined by the overall similarity of the bits in all window pairs.

Smith and Chang(Smith et al. 1994) proposed a method based on the 16 DCT coefficients of a 4 x 4 block in the image. The variance and the mean absolute values of each of these coefficients are then computed over the whole image. The feature for the entire image is then represented by this 32-component vector.

Shen et al. (Shen et al. 1996) proposed a technique to detect regions of interest and edges in JPEG compressed images from the high frequency DCT coefficients. The technique estimates edge orientation, edge offset from centre, and edge strength from DCT coefficients of an 8 x 8 block. The orientation includes horizontal, vertical, diagonal, vertically dominant and horizontally dominant. Their experimental result showed that the DCT based edge detection provides a performance comparable to that of the Sobel edge detection operator applied in the pixel domain.

3.4.3 Feature extraction

To extract indexing keys from the JPEG bit streams, a thorough examination of how a JPEG bit stream is produced is necessary. After the block-based DCT transform coefficients are quantized and arranged in zig-zag scanning order, the JPEG compression scheme uses the run-length coding to specify the number of zero AC coefficients preceding each non-zero valued AC coefficient. In the case of the DC coefficient, only the difference between the DC coefficients of the current block and the previous block is encoded. Both AC and DC coefficients are grouped into a number of categories according their magnitude values as illustrated in Table 3.3. A fixed Huffman coding table is then used to specify the category and the length of its codeword. The base code table used for the DC difference case is illustrated in Table 3.4, and part of the base code table used for the AC coefficient is illustrated in Table 3.5. The latter is incomplete but shown to illustrate how the combinations of run-lengths and AC coefficient values are entropy encoded in JPEG.

Table 3.3 JPEG Coefficient Coding Categories

Range	DC Difference Category	AC Category
0	0	N/A
-1,1	1	1
-3,-2,2,3	2	2
-7,....-4,....7	3	3
-15,...-8,8,...15	4	4
-31,..-16,16,...31	5	5
-63,...-32,32,...63	6	6
-127,...-64,64,...127	7	7
-255,...-128,128,...255	8	8
-511,...-256,256,...511	9	9
-1023,...-512,512,...1023	A	A
-2047,...-1024,1024,...2047	B	B
-4095,...-2048,2048,...4095	C	C
-8191,...-4096,4096,...8191	D	D
-16383,...-8192,8192,...16383	E	E
-32767,...-16384,16384,...32767	F	N/A

Table 3.4 JPEG default DC code (luminance)

C	BC	L	C	BC	L
0	010	3	6	1110	10
1	011	4	7	11110	12
2	100	5	8	111110	14
3	00	5	9	1111110	16
4	101	7	A	11111110	18
5	110	8	B	111111110	20

Where C = Category; BC = base code; L = length

Table 3.5 JPEG default AC code (luminance)

R/C	BC	L	R/C	BC	L
0/0	1010	4	2/1	11011	6
0/1	00	3	2/2	11111000	10
0/2	100	4	2/3	1111110111	13
...			...		
1/1	1100	5	3/1	111010	7
1/2	111001	8	3/2	111110111	11
...			...		

Where R/C=run/category; BC=base code; L=length.

For example, if the difference value of a DC coefficient is within the category C = 2, its magnitude value will be one of C^2 = four possible values -3, -2, 2, or 3 as specified in Table 3.3. Therefore, it will be encoded by a base code of 100 (Table 3.4) followed by C = 2 bits to specify its exact value within that category. This C = 2 bit value will be the C = 2 least significant bits of the value in the case of positive values or the C = 2 least significant bits of the negative difference minus 1, in the case of negative difference values. For an AC coefficient, assuming its magnitude value falls into the category 3, and two preceding zeros are counted by run-length coding. This would give us R/C = 2/3. From Table 3.5, this AC coefficient will be encoded by a base code of 1111110111 followed by 3 bits to specify its exact value out of the possible 8 values given in Table 3.3.

From the above description, it can be seen that JPEG entropy coding scheme is essentially built upon two major factors. One is the number of zeros recorded by run-length coding along the zig-zag scanning order, and the other is the category value reflecting the approximate magnitude of each DCT coefficient. In principle, the magnitude values of the DCT coefficients represent the signal energy within that particular block, which also reflect the texture feature of those pixel values inside that block. As a matter of fact, the MPEG-4 standard refers to the DCT-based entropy coding as texture coding to differentiate it from its shape coding. To this end, we could construct an indexing key by considering category numbers to characterize the texture of each individual pixel block. Hence, the total number of 64 coefficients inside each block enables us to produce 64 texture elements, with values which correspond to the category numbers specified in Tables 3.3. to 3.5. In the case of zero valued AC coefficients, their texture elements are simply taken as zeros. As a result, a vector of 64

elements is constructed to represent the texture feature of the pixel block. To characterize the texture feature of an entire image, such a vector can be used as a building block towards formulation of the indexing key. Therefore, given N blocks of DCT coefficients inside an image, the indexing key can be constructed as follows:

$$\text{Indexing key} = \{c1, c2, ..., c64\} \quad (1)$$

Where $c_i = \frac{\sum_{k=1}^{N} DCTcoefficients_k}{N}$ stands for the i^{th} category number among the 64 DCT coefficients.

The advantage of this design lies in the fact that the indexing key can be directly built up from the entropy codes, and thus no decoding is needed. Specifically, to build the texture key for each image file, all we need to do is to read those bits of base codes from the compressed file, which are used to encode the category number of the DCT coefficients for each individual block, and all other bits are ignored. When reading a JPEG compressed image bit stream, the presence of the EOB (end-of-block) codes will enable us to count the total number of blocks inside each image.

During the retrieval process, the JPEG query image in question will be subjected to the same procedure to read those base code bits and category numbers which are used to calculate its indexing key as given in (1). The keys computed for the query image will then be compared with the keys stored for images inside the database via the Mean Square Error (MSE) method.

There are two main advantages in using the proposed CBIR. As the proposed system does not need full decompression or IDCT, significant improvements on both

computing cost and processing speed can be expected in comparison with conventional techniques developed in pixel domain. However, the limitation of the proposed system is that no experiments were conducted for determining the system's invariance to translation, rotation and scaling.

3.4.4 Experimental Design

To test the proposed algorithm, a test database of around 3000 JPEG images was constructed and classified into a number of themes including cats, animals, flowers, scenes, paintings etc. Out of all these themes, 10 different query images were selected to carry out image retrievals, in which a total of 9 target images were retrieved for every query image. The retrieved images were arranged in an ascending order according to the values of their matched distances, where the first image corresponds to the minimum distance obtained between a retrieved image and the query image. Hence, an ascending order of 9 ranks is formulated.

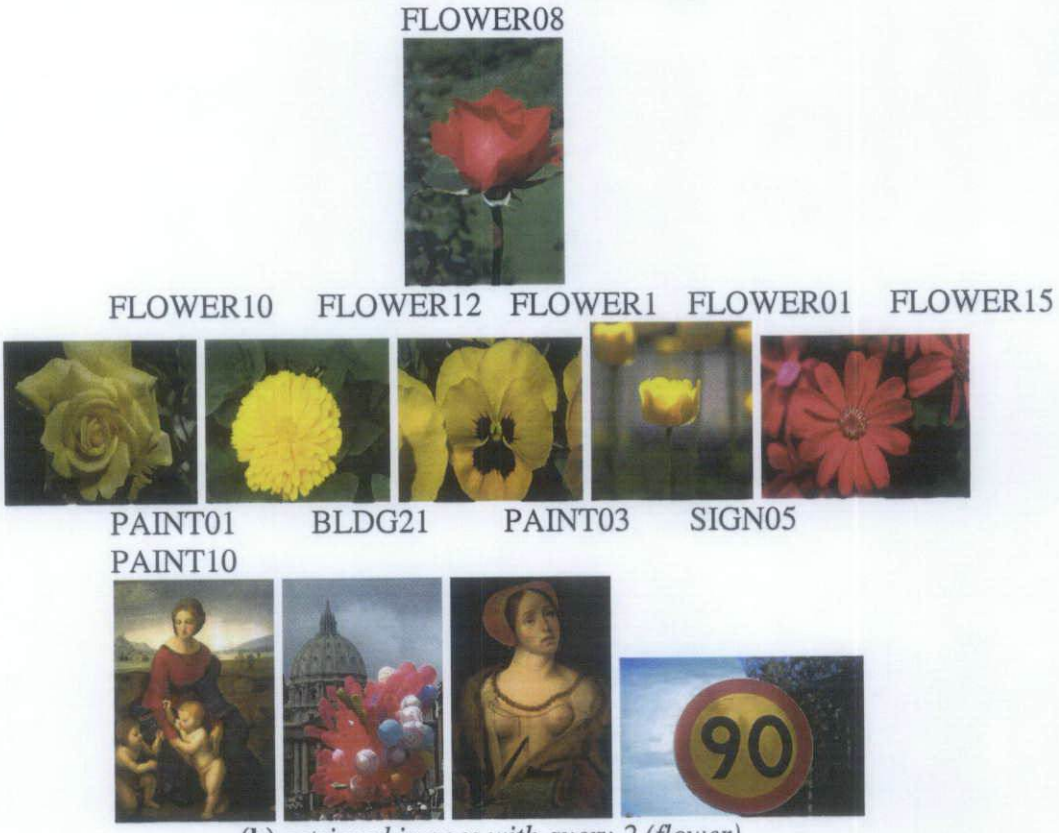
3.4.5 Results

The total experimental results are summarized in Table 3.6. The results show that for rank 1 images, the algorithm is able to correct images 100% of the time, for rank 2 images 90% of the time and up to rank 6 images at least 70% of the time. Two query examples are presented in Figure 3.5 for visual inspection. It should be noted however that the system has yet to be fully tested against a much larger image database. Nonetheless, the initial experimental results do present encouraging signs for our proposed algorithm.

Table 3.6 Relevancy of algorithm over 10 runs. 1 denotes that the retrieved image is relevant, 0 denotes otherwise. The percentage is obtained by dividing *Total relevant images retrieved* by *total run(10)*

Run	1	2	3	4	5	6	7	8	9	10	Total relevant images retrieved	%
Rank 1	1	1	1	1	1	1	1	1	1	1	10	100
Rank 2	1	1	1	1	0	1	1	1	1	1	9	90
Rank 3	1	1	1	0	0	1	1	1	0	1	7	70
Rank 4	1	1	1	1	1	1	1	1	1	1	10	100
Rank 5	0	1	1	1	1	0	1	0	1	1	7	70
Rank 6	0	1	1	1	1	0	1	1	1	0	7	70
Rank 7	0	1	1	1	1	1	0	0	0	0	5	50
Rank 8	0	1	1	1	0	1	0	1	1	1	7	70
Rank 9	1	0	1	0	1	1	0	0	1	0	5	50

(a) *Retrieved images for query-1 (cat)*



(b) *retrieved images with query-2 (flower)*

Figure 3.5 Two example queries, with query image at top and ten retrieved images in descending ranking from left to right starting from the top left.

3.5 Summary

In this chapter two pieces of work related to the JPEG compressed domain are presented. The first is the development of a progressive decoding scheme. From the experimental results it has been found that the scheme saves computational cost without significant degradation of the reconstructed image. The second piece of work involves a feature extraction scheme developed in the JPEG compressed domain. The scheme also saves computational costs since images do not have to be decompressed before comparisons. However it should be noted that the database used for this experiment is considered small (3000 images) in the context of this work. In the following chapter a further extension to this work is carried out using a larger database.

Chapter 4 Region-based image retrieval

4.1 Introduction

Region-based image retrieval (RBIR) is one of the categories of the existing general purpose Content-Base Image Retrieval (CBIR) system identified by (Wang et al. 2001), the others being histogram based and colour-layout based. The use of RBIR methods are primarily to overcome limitations of global features by representing images at the object-level, which is intended to be close to the perception of the human visual system. The underlying principle behind region-based image retrieval is the partitioning of an image into several regions, which may be based on colour, texture, shape, etc. Image features are extracted from these regions and the set of all features then becomes the image index or image key. During a query process, features extracted from the query image are matched against the corresponding features of other images in the image database. A key issue in RBIR systems is the segmentation technique used to partition images. Indirectly related to the image segmentation issue is also the issue of the number of segments or regions, because as the number of regions increases the computational complexity involved also increases. Other problems related to region-based segmentation are *noise problems, different length problem* and *boundary value problem* (Kim et al. 2003).

This chapter is organised as follows. In Section 4.2 a survey of RBIR systems is presented. Section 4.3 describes the RBIR system that I have developed working in the JPEG compressed domain. Information on the image database set up for the

experimental work is summarized in Section 4.4, followed in Section 4.5 by the experimental design, which highlights the different ways of obtaining the similarity measures. Experimental results are presented in Section 4.6 followed by a summary in Section 4.7.

4.2 Survey of RBIR systems

Interest in region-based approaches came about because of the need to overcome the deficiencies of CBIR systems which use global features in carrying out image retrieval. A consequence of using global features is that no spatial information is preserved. Proponents of region-based approaches suggest that an image can be best represented as a set of regions instead of as just a single region. Ideally when an image is decomposed into regions, each region represents a distinct object. However, image segmentation is not a very easy thing to accomplish. Below are described some of the region-based approaches currently in use.

In the Blobworld system for image retrieval (Carson et al. 1999), images are segmented into regions (blobs) by fitting a mixture of Gaussians to the pixel distribution in a joint colour-texture-position feature space. Each region is made up of a composition of colour and texture. In carrying out a query, the user first selects a category, thus limiting the search space. Given an initial image, the user selects a region (blob), and indicates the importance of the blob, the blob's color, texture, location, and shape. One or more regions can be used for querying. The results obtained show that the system yields good results when querying for distinctive objects. However the segmentation process in this system takes about 5 minutes per image on a 300MH Pentium II PC.

In the NeTra toolbox for navigating large image databases (Ma et al. 1997), a given image is segmented into a number of relatively homogeneous regions which are represented by colour, texture, and shape attributes. The segmentation is done automatically. Each of the regions in the database is indexed separately. To initiate a query, the user is required to select regions and the corresponding features to evaluate similarity. However, this could lead to problems, since the user is placed with the burden of deciding on the parameters to be used for the retrieval process.

In the Simplicity Semantic-sensitive Integrated Matching for Picture Libraries system (Wang et al. 2001) the authors make use of semantics to classify images into the following categories: Textured versus Non-textured and Graph versus Photographs. A textured image is defined as an image of a surface, a pattern of similarly-shaped objects, or an essential element of an object. For the Graph versus Photographs, they defined an image as that of a photograph if it is a continuous-tone image, whereas a graph image is defined as an image containing mainly text, graph, and overlays. Segmentation is carried using the k-means algorithm using, six image features. A threshold value is then used for assigning the images into one of the two classes i.e. either textured or non-textured or whether it is classified as graph or photograph. This approach is robust to inaccurate segmentation as it uses a region-matching scheme that integrates the properties of all regions in the images. However, the computational cost can be expensive as the number of regions increases.

Messer and Kittler use local colour and texture properties in their region-based image retrieval system (Messer et al. 1999). Along with the region size and location, the mean colour and texture properties of each region are computed and stored as the index

to the image. The final image signature is in the form of a 33-dimensional feature vector. However the database used in this experiment has an image collection of 3481 images. On top of that no ground truth for the images are available. As such the retrieval results obtained could possibly be subjective.

VisualSEEk (Smith et al. 1996b) is a fully automated content-based image query system, which integrates feature-based image indexing with spatial query methods. The system is able to retrieve single or multiple regions by their spatial location. A colour-set back-projection technique is used to extract colour regions. The similarity between two images is computed by taking into account colour, location, size and relative positions of regions. During the query process, the user sketches regions, positions them on the grid and assigns them properties of colour, size and absolute location. The user may also assign boundaries for location and size. However the system also suffers the same drawback similar to the Netra system in that the user has to make the selection of parameters for the retrieval process.

Wavelet-Based Index*ing of Image*s U*sing* R*egion* F*ragmentation (Windsurf), is another region-based image retrieval system* (Ardizzoni et al. 1999). In this system the Discrete Wavelet Transform (DWT) is used to extract colour and texture features from an image. The k-means clustering algorithm is then used to partition the images into a set of "homogeneous regions. Similarity between images is assessed by first computing the similarity scores between regions and then combining the results at the image level. The Windsurf approach was then compared with an established work (Stricker et al. 1995) and was found to perform better than its benchmark. However a major limitation

of the this approach is its low speed during the retrieval phase, since the regions from the database are sequentially scanned.

Another approach to region-based image retrieval has been proposed by (Natsev et al. 2004), called WALRUS (WAveLet-based Retrieval of User-specified Scenes), a similarity retrieval algorithm that is robust to scaling and translation of objects within an image. In order to extract regions of an image and their signatures, WALRUS considers multiple sliding windows of varying sizes in the image and computes a wavelet signature for each. It then clusters the windows based on the proximity of their signatures and considers each cluster as a region, with its centroid as the representative signature for the region. Each region is then inserted in a spatial index R*-tree (Beckmann et al. 1990). The similarity measure between a pair of images is then defined to be the fraction of the area of the two images covered by matching regions from the images.

In the FRIP system (Ko et al. 2005), the segmentation process is done at two levels. At the first level, an image is segmented using three types of adaptive circular filters based on the amount of image texture information the image possesses. At the second level, small patches of images are merged into similar adjacent regions by region merging and region labelling. The filtering process is repeated until 30 or less regions have been obtained. From each region, five vectors are extracted, each representing colour, texture, area, shape and location features. The five feature vectors are normalized to [0–1] using the Gaussian normalization method (Rui et al. 1998) as soon as features are extracted from the segmented regions before they are used for distance estimation. During the query process, after regions are segmented, the user selects one

or multiple regions that he/she wants to search. The user then makes a further selection of the user-specified constraints, such as: color-care/do not care scale area-care/do not care, shape-care/do not care, and location care/do not care. The overall matching score is calculated according to the pre-defined similarity measure. However giving the user some degree of freedom in selecting query regions and attributes, does inhibit usability because the segmented regions do not necessarily correspond to semantic objects and, therefore, identifying relevant regions/attributes is difficult.

4.3 RBIR in JPEG compressed domain

The author has developed a region-based image retrieval system in the JPEG compressed domain. For this work, the number of regions is fixed at 16 based on the JPEG Coefficient Coding Categories (Table 3.3). There are two stages in the proposed algorithm design. The first one segments the image and the second one develops an indexing key based on the regions obtained.

4.3.1 Image Segmentation

The segmentation referred here is different to that generally used by the image processing community in that it acts in the compressed domain. The segmentation approach is to divide the image into regions in accordance with the JPEG coding category rather than by analysing image content. From the JPEG DC Difference Categories given in Table 4.1 (copied from Table 3.3 for the convenience of the reader), it can be seen that there are 16 cases.

Table 4.1 JPEG DCT Coefficient Coding Categories.

Range	DC Difference Category	AC Category
0	0	N/A
-1,1	1	1
-3,-2,2,3	2	2
-7,....-4,....7	3	3
-15,...-8,8,...15	4	4
-31,..-16,16,...31	5	5
-63,...-32,32,...63	6	6
-127,...-64,64,...127	7	7
-255,...-128,128,...255	8	8
-511,...-256,256,...511	9	9
-1023,...-512,512,...1023	A	A
-2047,...-1024,1024,...2047	B	B
-4095,...-2048,2048,...4095	C	C
-8191,...-4096,4096,...8191	D	D
-16383,...-8192,8192,...16383	E	E
-32767,...-16384,16384,...32767	F	N/A

Given an image with N blocks of 8×8, the DC of the first block value is taken and is compared with the Range values specified in the above table. Based on the range to which it belongs, the DC value is then replaced with the corresponding value from the DC Difference Category. For the remaining 8x8 blocks, the difference between the DC values of the previous block and the current block is taken. The result is compared with the range values and a new DC Category value is subsequently derived. In the case of the AC coefficients, the value is compared with the range values and it is then replaced with the corresponding AC category value. If we put together all the 8x8 pixel blocks which have the same DC Difference category values, a set of up to 16 regions is defined. So, as a result of this grouping, we are able to divide each JPEG compressed image into 16 regions, some of which may be null. The practical relevance of this segmentation scheme is that the number of regions is fixed, hence the computational complexity issue associated with increasing number of regions mentioned earlier, does not arise. A significant aspect of this process is that this division is carried out

completely in the compressed domain by just looking up the category number of the JPEG bit streams. When an indexing key is constructed for every region, as described next, we will have multiple keys to characterize the texture feature of the compressed image, and thus better performance can be expected. Upon completion of the first stage, the image keys are then constructed.

4.3.2 Image key construction

Assuming that an image I has been segmented from 4.3.1, I can be represented as:

$$I = \{R_1, R_2, ...R_N\}, \quad (1)$$

where R_i represents the i'th region and $N_{(max)} = 16$.

A particular region R_i, is represented as follows:

$$R_i = \{B_1, B_2, ...B_M\} \quad (2)$$

In other words the region R_i is made of M blocks $B_j$of size 8×8.

Therefore the key for region R_i is a vector of 64 elements given as:

$$key_i = \{c_1, c_2, ...c_{64}\} \quad (3)$$

c_i is derived from the following formula

$$c_i = \frac{\sum_{j=1}^{M} DCTCoefficients_j}{M}, \quad (4)$$

that is the average of the coefficients over all the blocks in a region.

Correspondingly, for N regions inside an image, we will have N indexing keys, which can be represented as:

$$Key = \{k_1, k_2, ...k_N\} \quad (5)$$

4.4 Database organization

To test the proposed algorithm, a test database of around 5000 JPEG images of varying sizes was constructed. Though some images are of type greyscale, most images are colour. Out of the 5000 images, 571 images are included for which their ground truth had been established by manual inspection by the author (see Table 4.2). The 571 images are classified into five categories namely scenes, cats, painting, plants and signs.

The scene category essentially represents outdoor scenery containing images of the sky, sea and valleys and mountains. The cat category contains images whose main focus are images of cats. The painting category contains captured images of paintings done by human artists. The plant category mainly contains images of flowers. Finally the signs category contains sign-posted images. An example image from each category is given in Figure 4.1.

Table 4. 2 The categories of image with known content

Category	No. of images	No. of query images
Cats	70	7
Painting	61	6
Plants	120	12
Scenes	300	30
Signs	20	2
Total	571	57

cats	
painting	
plants	
scenes	
signs	

Figure 4.1 Example images from each category of image in the database.

4.5 Experimental Design

In carrying out the image retrieval process, all images are indexed at run-time. Based on the feature extraction scheme as described above, four different approaches were used in measuring the retrieval performance. The Euclidean distance measure was employed in all the experiments and the minimum distance value denotes the degree of similarity. For the purpose of clarification the approaches are designated as metrics 1, 2,3 and 4.

For metric 1, a many-to-many comparison procedure is carried out, where all the images involved have multiple keys. In other words every region of the query image will be compared with all regions that are found in the database images. This results in a maximum of 256 distance measurements (16×16) per comparison. The measurement with the minimum distance value is selected as the final distance measurement.

In metric 2, the "most influential region" comparison procedure is carried out whereby a single key is generated for both the query image and the database images. This key is the one associated with the region which has the largest number of blocks which is regarded as the most influential.

In metric 3, a single key is assigned to the query image and to the database images. However, this key is obtained by combining features from all texture elements together, weighted by the number of blocks inside each region according to the following formula:

$$Key = \{c_1, c_2, \ldots . c_{64}\} \text{ and } \quad c_i = \frac{\sum_{j=1}^{N} w_j c_j}{\sum_{j=1}^{N} w_j} \qquad (4)$$

Finally for metric 4, a one-to-many procedure is adopted, whereby the query key is a single key derived from the most influential region whereas the database images are represented by multiple keys representing the multiple regions. Therefore, when comparing the regions, at most 16 distance measurements will be obtained. Out of the

16 measurements, the one with the minimum value is selected as the final distance measurement.

In order to evaluate the effectiveness of the above four metrics, two other algorithms were included, named P1 and P2. P2 is an image feature extraction based on an earlier work, mentioned in chapter 3. P1 is a feature extraction based on the work done by (Shneier et al. 1996), whose technique is based on the mutual relationship between the DCT coefficients of unconnected regions in both the query image and target image. Based on the user-supplied input K, a set of $2K$ windows is selected and randomly paired producing K pairs of windows. For each window the average of each DCT coefficient is computed resulting in a 64-D feature vector (f). The vectors corresponding to a pair of windows are then compared and each pair of components is assigned a bit (0 or 1) depending on their similarity. Thus each pair of windows will be assigned 64 bits. The similarity of the query and target image is determined by the overall similarity of all the bits in all window pairs.

From the five categories specified in Section 4.4, 57 different query images were selected to carry out image retrieval, in which a total of 10 target images were retrieved for every query image. The retrieved images are arranged in ascending order according to the values of their distances from the query image in the feature space. The ranked images were then checked as relevant or not relevant and the results tabulated.

4.6 Experimental Results

In this section detailed results from the six algorithms are given.

4.6.1 Results for Metric 1

As stated earlier, the metric involves multiple key comparisons i.e. all regions from the query image is compared to all regions of the database images. Using this metric, it can be seen that Rank1 and Rank2 retrievals for all the five image categories are correct (Table 4.3). In other words the top 2 ranked images retrieved in all the queries were relevant. Inspecting the Scene category it can also be seen that this category has been successfully retrieved up to rank 8.

Table 4.3 Number of images correctly retrieved per category – Metric 1. The percentage is obtained by dividing the total number of relevant images retrieved in each category to the total number of images for all categories (57)

Categories	CATS	PAINT	PLANTS	SCENE	SIGN	Total	%
No.of images in each category	7	6	12	30	2	57	
RANK1	7	6	12	30	2	57	100.00
RANK2	7	6	12	30	2	57	100.00
RANK3	7	3	12	30	1	53	92.98
RANK4	7	3	9	30	1	50	87.72
RANK5	4	1	9	30	0	44	77.19
RANK6	7	3	5	30	1	46	80.70
RANK7	3	3	8	30	2	46	80.70
RANK8	1	1	4	30	0	36	63.16
RANK9	3	2	7	28	0	40	70.18
RANK10	4	2	5	27	0	38	66.67

4.6.2 Results for Metric 2

The main reason for applying this metric is to see if a single "most influential region" can give a good representative feature vector. However the performance of this metric is not very good when compared to Metric 1 as shown by the percentages obtained for each rank (Table 4.4). It performs well for the Cat category, successfully

retrieving all images up to rank 5. However it fares poorly on the Sign category, where only up to rank 2 are images successfully retrieved.

Table 4.4 Number of images correctly retrieved per category – Metric 2. The percentage is obtained by dividing the total number of relevant images retrieved in each category to the total number of images for all categories (57)

Categories	CATS	PAINT	PLANTS	SCENE	SIGN	Total	%
No.of images in each category	7	6	12	30	2	57	
RANK1	7	6	12	30	2	57	100.00
RANK2	7	4	11	30	2	54	94.74
RANK3	7	2	10	30	0	49	85.96
RANK4	7	5	3	29	0	44	77.19
RANK5	7	2	3	30	0	42	73.68
RANK6	4	1	5	23	0	33	57.89
RANK7	2	2	6	25	0	35	61.40
RANK8	3	2	5	24	0	34	59.65
RANK9	2	0	5	23	0	30	52.63
RANK10	1	1	9	20	0	31	54.39

4.6.3 Results for Metric 3

This metric, which uses a system of weights performs better than Metrics 1 and 2 taking into account ranks 1, 2 and 3. (Table 4.5). Unfortunately, rank 1 is the worst of these three ranks with 98.25 success rate. Also, under the Scene category, it has the best retrieval efficiency so far, whereby images up to rank 9 are successfully retrieved.

Table 4.5 Number of images correctly retrieved per category – Metric 3. The percentage is obtained by dividing the total number of relevant images retrieved in each category to the total number of images for all categories (57)

Categories	CATS	PAINT	PLANTS	SCENE	SIGN	Total	%
No.of images in each category	7	6	12	30	2	57	
RANK1	7	6	11	30	2	56	98.25
RANK2	7	6	12	30	2	57	100.00
RANK3	7	6	12	30	2	57	100.00
RANK4	7	4	9	30	2	52	91.23
RANK5	4	4	5	30	2	45	78.95
RANK6	7	2	6	30	2	47	82.46
RANK7	3	0	7	30	1	41	71.93
RANK8	1	0	6	30	0	37	64.91
RANK9	3	1	4	30	1	39	68.42
RANK10	4	1	6	28	2	41	71.93

4.6.4 Results for Metric 4

For this metric, a one-to-many approach is carried out, whereby the query key is a single key represented by the most influential region and the database images are represented by multiple keys. The results obtained for this metric (Table 4.6) is similar to Metric 3.

Table 4.6 Number of images correctly retrieved per category – Metric 4. The percentage is obtained by dividing the total number of relevant images retrieved in each category to the total number of images for all categories (57)

Categories	CATS	PAINT	PLANTS	SCENE	SIGN	Total	%
No.of images in each category	7	6	12	30	2	57	
RANK1	7	5	12	30	2	56	98.25
RANK2	7	6	12	30	2	57	100.00
RANK3	7	5	9	30	1	52	91.23
RANK4	7	4	9	29	1	50	87.72
RANK5	7	3	8	28	1	47	82.46
RANK6	6	4	9	28	1	48	84.21
RANK7	5	5	9	21	1	41	71.93
RANK8	6	2	8	23	0	39	68.42
RANK9	5	1	6	19	1	32	56.14
RANK10	4	3	5	21	0	33	57.89

4.6.5 Results for P1

The results obtained using this algorithm based on the work of (Shneier et al. 1996) are not good compared to the other algorithms (Table 4.7). In fact it has the lowest correct retrieval percentages compared to the rest.

Table 4.7 Number of images correctly retrieved per category – Algorithm P1. The percentage is obtained by dividing the total number of relevant images retrieved in each category to the total number of images for all categories (57)

Categories	CATS	PAINT	PLANTS	SCENE	SIGN	Total	%
No.of images in each category	7	6	12	30	2	57	
RANK1	3	5	9	24	1	42	73.68
RANK2	2	0	9	28	1	40	70.18
RANK3	3	3	6	29	0	41	71.93
RANK4	2	1	5	29	0	37	64.91
RANK5	2	0	4	26	0	32	56.14
RANK6	0	0	0	19	0	19	33.33
RANK7	1	0	1	11	0	13	22.81
RANK8	0	0	1	4	0	5	8.77
RANK9	0	0	0	4	0	4	7.02
RANK10	0	0	0	3	0	3	5.26

4.6.6 Results for P2

For the last algorithm base on an earlier work, the results compares favourably with algorithms 1 and 3 (Table 4.8). This algorithm too gives excellent results for the Scene category whereby up to rank 8 images are successfully retrieved.

Table 4.8 Number of images correctly retrieved per category – Algorithm P2. The percentage is obtained by dividing the total number of relevant images retrieved in each category to the total number of images for all categories (57)

Categories	CATS	PAINT	PLANTS	SCENE	SIGN	Total	%
No.of images in each category	7	6	12	30	2	57	
RANK1	7	6	9	30	2	54	94.74
RANK2	7	6	12	30	2	57	100.00
RANK3	7	6	12	30	2	57	100.00
RANK4	7	5	10	30	2	54	94.74
RANK5	6	4	8	30	2	50	87.72
RANK6	4	2	8	30	2	46	80.70
RANK7	4	0	6	30	0	40	70.18
RANK8	2	2	5	30	0	39	68.42
RANK9	5	1	2	29	1	38	66.67
RANK10	5	0	10	29	2	46	80.70

Tables 4.3 – 4.8 are combined to produce one final table (Table 4.9) which gives the overall results for all the algorithms. Based on the retrieval of the top ranked images, Metric 1 gives the best results by virtue of being able to retrieve the top two ranked images. However, representing the performances of the algorithms by a single measure, the mean retrieval overall the ranks, the P2 algorithm gives the best performance with an average retrieval rate of 84.4%, followed by Metric 3 with an average retrieval rate of 82.8%. The average performance of the P1 Algorithm is only 41.4% compared with the next lowest, Metric 2 which has an average retrieval rate of 71.8%. A sample query is shown in Figure 4 .2.

Table 4.9 Overall retrieval performance of all algorithms.

	Metric 1	Metric 2	Metric 3	Metric 4	Algorithm P2	Algorithm P1
RANK1	100.00	100.00	98.25	98.3	94.74	73.68
RANK2	100.00	94.74	100.00	100.0	100.00	70.18
RANK3	92.98	85.96	100.00	91.2	100.00	71.93
RANK4	87.72	77.19	91.23	87.7	94.74	64.91
RANK5	77.19	73.68	78.95	82.6	87.72	56.14
RANK6	80.70	57.89	82.46	84.2	80.70	33.33
RANK7	80.70	61.40	71.93	71.9	70.18	22.81
RANK8	63.16	59.65	64.91	68.4	68.42	8.77
RANK9	70.18	52.63	68.42	56.1	66.67	7.02
RANK10	66.67	54.39	71.93	57.9	80.70	5.26
Average	81.9	71.8	82.8	79.8	84.4	41.4

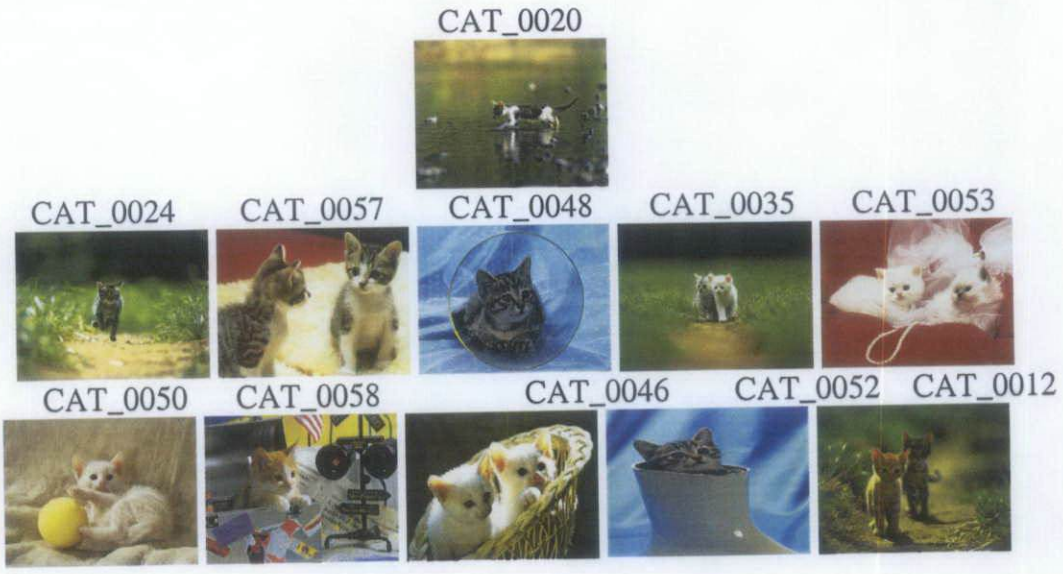


Figure 4. 2 An example query, with query image at top and ten retrieved images in descending ranking from left to right starting from the top left.

4.7 Summary

In this chapter, a Region-Based Image Retrieval System in the compressed domain has been developed. To the best of our knowledge, there is no implementation of a region based image retrieval in the compressed domain. In this implementation there are at most sixteen regions (see Table 4.1), therefore key comparison during the query stage would be fast. Several variations in computing the similarity measures have been carried out to see the best representation within a region-based system.

The first approach, Metric 1, involves the use of multiple keys, representing the regions in an image. The rationale for this approach is the belief that all regions should be considered in determining the similarity measure. As it turns out, this belief has been justified, since the approach gives the best overall results, in terms of top-ranked images retrieved. However, the approach takes the longest amount of time compared to the other 3 metrics, which is understandable, since it involves multiple comparisons.

The rationale for the second approach, Metric 2, was to have an image represented by its most "influential" or dominant region. In this case, "influential" is defined as the region with the most number of blocks. By having a single key to represent an image, this approach is expected to take a much shorter time during the retrieval phase as compared to the first approach, since only a single comparison is needed. However, "influential" regions of two images does not necessarily have to be the same. This is reflected in the overall results obtained, whereby it receives a ranking of 5 out of 6 in terms of overall average retrieval efficiency.

For Metric 3, weights represented by the number of blocks in each region are used so that all regions will contribute towards the generation of the final key. In other words, a bigger-sized region will have a greater influence on the final key being generated as compared to a smaller-sized region. The retrieval performance using this metric is comparable to that of Metric 1, where

For Metric 4, a one-to-many approach was taken, whereby a single key derived from metric 2 was compared with multiple keys for the target images. This metric proved to be better than by metric 2, in terms of average retrieval performance.

However this system does have its limitation in terms of the query process, in that it does not offer a query facility involving part(s) of a region. Instead the whole image is submitted as a query image. This is due to the fact that the regions were derived in accordance with the JPEG coding category rather than by analysing image content. Based on the concepts presented in this chapter and the previous chapter a hybrid feature extraction scheme is developed in the next chapter.

Chapter 5 Image Classification

5.1 Introduction

Classification can be regarded as the grouping or dividing of objects into classes to form an ordered arrangement of items. Each arrangement of items will have a defined range of characteristics, relationships and distinctive differences. Classification systems may be taxonomic, mathematical, observed, or inferred, depending upon the purpose to be served. The purpose of classification is to describe the structure and relationships of objects to each other and to similar objects, and to simplify these relationships in such a way that general statements can be made about classes of objects, thereby achieving economy of memory and ease of manipulation. The construction of a classification procedure from a set of data for which the true classes are known has also been variously termed *pattern recognition*, *discrimination*, or *supervised learning* (in order to distinguish it from *unsupervised* learning or *clustering* in which the classes are inferred from the data).

Image classification, then is the task of classifying images into (semantic) categories based on the available training data. This categorization of images into classes can be helpful both in semantic organizations of digital libraries and in obtaining automatic annotations of images. In general, classification comprises of four steps:

- Pre-processing of images
- Training of image features

- Selection of suitable method for comparing the image patterns with the target patterns.
- Assessing the accuracy of the classification.

There are many examples of image classifiers including the k-nearest neighbour, the decision tree, the Bayesian net, the maximum likelihood analysis, the linear discriminant analysis, the neural network and the SVM, etc. For this work the latter two have been chosen. The neural network was chosen because of its popularity and the SVM was chosen because despite being relatively new it has been known to give good classification results. The study of neural networks, or more specifically Artificial Neural Networks (ANN), is a multidisciplinary field with wide ranging applications to areas such as finance, industry, agriculture and computer science. The SVM is a training algorithm for learning and classification based on the statistical learning theory, originally developed by Vapnik (Vapnik 1995).

Two main objectives are set for this work, namely to find the best image features from five which have been constructed and also to determine a best classifier. In order to achieve these objectives, the image features are classified using the two classifiers and a comparison is then made between the performances of all five features. Figure 5.1 below is a representation of the proposed classification scheme. Two sets of images are used; one containing the training set and the other containing the testing set. Both sets of images will be subjected to the feature extraction module before being passed to the image classifier. The output from the image classifier will be the category that the testing images belongs to.

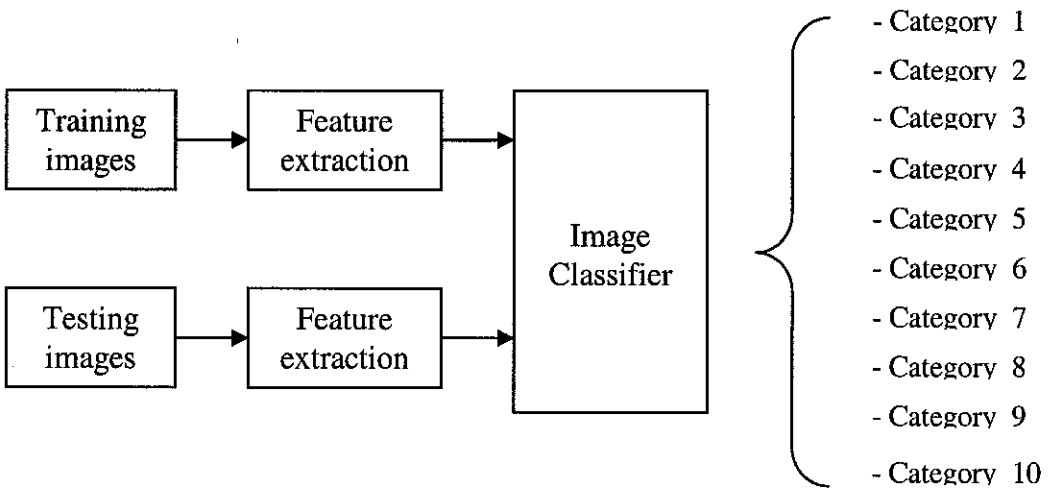


Figure 5.1 Proposed Image classification scheme

The remaining content of this chapter is as follows. Brief descriptions of neural networks and support vector machines are given in sections 5.2 and 5.3 respectively. This is followed by a survey of the use of these two classifiers in image retrieval in Section 5.4 as well as comparisons between the two classifiers. The five individual feature extraction schemes chosen for this work are described in section 5.5. Section 5.6 contains a description of the experimental design, including the database organizational setup as well as the parameters associated with the two classifiers. Results and conclusion follows in Sections 5.7 and Section 5.8 respectively.

5.2 Neural Networks (NN)

Neural networks are a powerful tool for solving complex, non-linear classification problems, due to their ability to be trained in order to handle unknown information hidden in the data (Haykin 1999). (Egmont-Petersen et al. 2002) have described 6 tasks that the NN is used for in the image processing chain namely; preprocessing, data reduction and feature extraction, image segmentation, object recognition, image understanding and optimisation.

The name, neural network is a biological term which refers to the collection of neurons or tiny brain cells found in the human brain. The use of Artificial Neural Networks (ANNs) is really an attempt to build simplified but useful models of these biological structures both in architecture and operation. The basic biological neuron consists of synapses, the soma, the axon and dendrites. Synapses are connections between neurons that allow electric signals to jump across from neuron to neuron. These electrical signals are then passed across to the soma which performs some operation and sends out its own electrical signal to the axon. The axon then distributes this signal to dendrites. Dendrites carry the signals out to neighbouring synapses, and the cycle repeats. Each artificial neuron has a certain number of inputs, each of which have a *weight* assigned to them. The weights simply are an indication of how 'important' the incoming signal for that input is. The *net value* of the neuron is then calculated as the sum of all the inputs multiplied by their specific weights. Each neuron has its own unique threshold value, and if the *net value* is greater than the threshold, the neuron fires (or outputs a 1), otherwise it stays inert (outputs a 0). The output is then fed to all the neurons it is connected to. The weights and threshold values are set by a training algorithm using a set of training data.

An ANN is characterized by its topology, activation function, and learning rules. The topology is the architecture specifiying the number of neurons and how they are connected, the activation function is a characteristic of each neuron, and the learning rule is the strategy for learning. The standard network architecture consists of several layers of neurons, an input layer, one or more hidden layers, and an output layer. Input layers take the input and distribute it to the hidden layers. The function of the hidden layers is to carry out computational work and output the results to the output layer. The NN learning strategy can be divided into feed forward models and the feedback models.

In the feed forward model, data from neurons closer to the input layer are propagated forward to neurons closer to the output layer through the network connections, whereas in the feedback model, data from neurons closer to the output layer are brought back to neurons closer to the input layer. Figure 5.2 shows the structure of a feed forward model ANN.

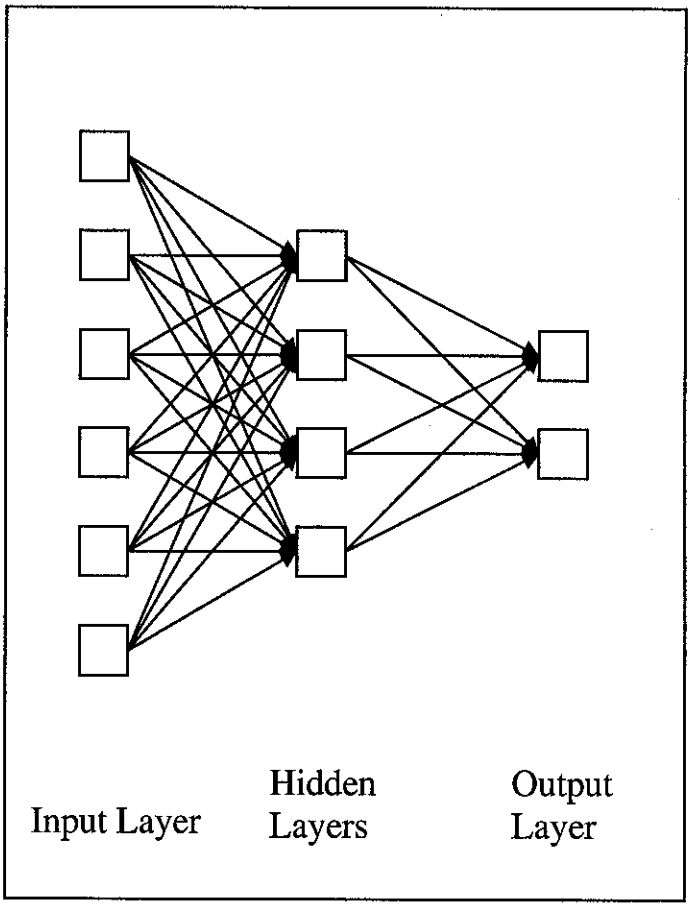


Figure 5.2 Structure of a feed forward Neural Network model ANN.

The training of a NN can be done in a supervised or unsupervised manner. In supervised training the actual output of a neural network is compared to the desired output. Weights, which are usually set to random values to begin with, are then adjusted by the network training algorithm so that the next iteration, or cycle, will produce a closer match between the desired and the actual output. The learning method tries to

minimize the current errors of all processing elements. This global error reduction is created over time by continuously modifying the input weights until acceptable network accuracy is reached. Furthermore, with supervised learning, the artificial neural network must be trained before it becomes useful. Training consists of presenting input and output data to the network. This data is often referred to as the training set. That is, for each input set provided to the system, the corresponding desired output set is provided as well. In most applications, actual data must be used. This training phase can consume a lot of time (Ma et al. 2000). In prototype systems, with inadequate processing power, learning can take weeks. This training is considered complete when the neural network reaches a user defined performance level. This level signifies that the network has achieved the desired statistical accuracy as it produces the required outputs for a given sequence of inputs. Training sets need to be fairly large to contain all the information needed if the network is to learn the features and relationships that are important.

The most widely used supervised training procedure is the back-propagation (BP) algorithm which has been successfully used in many fields, especially for pattern recognition. A back-propagation neural network is a multilayer feed forward neural network known for its learning capability (Widrow et al. 1994). It consists of a layer of input neurons, a layer of output neurons and one or more hidden layers. Each layer consists of neurons that are fully connected to the neurons of adjacent layers using weights. The term back-propagation refers to the manner in which the computed error at the output layer is back-propagated from the output layer to the hidden layer and finally to the input layer in the training phase which involves a number of iterations defined by the user. Every iteration constitutes two sweeping actions: forward activation to produce a solution and a backward propagation of the computed error to modify the weights.

Thus, based on a set of inputs and the corresponding output targets, the weights are successively adjusted during the training phase until a stopping criterion is met. Since the nature of the error space cannot be known a priori, neural network analysis often requires a large number of individual runs to determine the best solution. Once a neural network is 'trained' to a satisfactory level it may be used as an analytical tool on other data. To do this, the user no longer specifies any training runs and instead allows the network to work in forward propagation mode only. New inputs are presented to the input pattern where they propagate into and are processed by the middle layers as though training were taking place, however, at this point the output is retained and no back-propagation occurs. The output of a forward propagation run is the predicted model for the data which can then be used for further analysis and interpretation. As mentioned earlier the training phase is an integral characteristic of the ANN methodology as it serves as a precursor to an optimised neural network structure. An ANN is optimised when the best hidden layer nodes and the optimum learning time are reached (Ma et al. 2000). Deciding on the number of hidden layers is also a critical issue in neural network design. Too few of them could result in poor quality prediction. On the other hand too many hidden layers could result in data overfitting problems whereby it works well on the trained data but not on other data. A network should be large enough to learn and small enough to generalise (Horn 1997; Tzafestas et al. 2000). Informative literature on BP can be found in (Rumelhart et al. 1986)

5.3 Support Vector Machines (SVM)

SVMs are a relatively new learning algorithm that is based on statistical learning theory. It was developed by Vapnik (Vapnik 1995) and uses a structural risk minimisation principle. Given an input vector, the SVM classifies it into one of two

classes. This is done by seeking the optimal separating hyper-plane that not only separates the data without errors but also maximises the margin, i.e. it maximises the distance between the closest vector in both classes to the hyperplane. This is of course with the assumption that the data are linearly separable. For the case when the data is not linearly separable, SVM will map the data into a higher dimensional feature space where the data is linearly separable. This is done by the use of kernel functions. For this study the following kernel functions were employed, namely the polynomial, radial basis and linear functions. Informative literature on SVMs can be found in the books by (Cristianini et al. 2000; Vapnik 1995).

5.4 Literature survey

There have been many uses of machine learning in image retrieval. (Sheikholeslami et al. 2002) applied a back-propagation neural network to feature vectors generated using colour and texture. For each generated feature vector a ranked list of retrieved images are obtained using two different similarity measures. The two lists are then combined to obtain a final list of ranked images. The work done by (Laaksonen et al. 2000) involves the use of self-organising map (SOM) for classification. Relevance feedback is also incorporated into the system. Five feature vectors are generated for each image based on two colour features, two shape features and one simple texture feature. For each feature vector, a corresponding Tree-structured Self Organising Map (TS-SOM) was created during the query process. During an image-based query, the five feature vectors computed from the images are passed to the respective TS-SOMs producing the best-matching units for each map. The results from these maps are then combined and presented to the user. Based on the user responses, the whole process is then repeated again. Other work involving the use of SOM was

carried out by (Chan et al. 2004) where multiple feature vectors were generated using colour, edge, region and texture. The final feature vector is a combination of the four feature extraction methods which are then fed to the neural network for classification.

Work done by (Guo et al. 2002b), involves an image retrieval scheme using relevance feedback. A boundary is created separating the image into those that are inside a boundary and those that are outside a boundary. Two machine learning techniques, the Support Vector Machine (SVM) and Adaboost (Freund et al. 1997) were compared in order to find out which technique has the better learning capability in terms of boundary learning. The positive and negative examples used to learn the boundary were provided by the user through relevance feedback. In this work it was reported that SVM provides the better generalisation capability.

(Qu et al. 2003) have used Multi-Layer Perceptron (MLP), Radial Basis Function (RBF) type ANNs and SVM as the classifiers for their automatic solar detection. From their experimental results it was found that SVM gave the best classification rate for solar flares. Work done by (Acir et al. 2004; Distante et al. 2003; Huang et al. 2004; Pal et al. 2004), though not related to content-based image retrieval, have shown that SVM outperformed neural networks in the classification task.

5.5 Feature Extraction design

Five different feature extraction techniques were carried out in this experiment. These involve feature extraction from the pixel domain as well as the compressed domain and the corresponding algorithms were implemented using the C programming language. The five features, Feature 1 to Feature 5, are described briefly in the

following sub-sections 5.5.1 to 5.5.5. Features 2, 3 and 4 are used for comparison with the two new features, Feature 1 and Feature 5.

5.5.1 Feature 1

Feature 1 is based on the compressed domain and was developed earlier by the author in (Baharudin et al. 2003). Details of the scheme can be found in 3.4.3.

5.5.2 Feature 2

Feature 2 is based on the well-known colour histogram scheme. Originally proposed by Swain and Ballard (Swain et al. 1991), it has been developed further by Sawhney et al. (1994), (Hafner et al. 1995; Han et al. 2002; Sawhney et al. 1994; Smith et al. 1996a). Using this approach, the representation of the image is obtained using three primaries of the colour space, of which the RGB colour space is the most common. The three-color channels are then discretised into a specified number of intervals from which the total number of bins is obtained. For example, if each channel is discretised into 4 intervals, the total number of bins obtained will be 64 ($4 \times 4 \times 4$). A colour histogram is then constructed based on the occurrences of each pixel in the respective bins. For the example used above, a final vector image of size 64 will be obtained, which contains the respective frequencies within each bin. These feature vectors (indexes) will then be stored as the indexes (keys) of the images. For this work our feature vector also contains 64 elements.

5.5.3 Feature 3

Feature 3 is based on another well-known feature, binary texture. The local binary pattern (LBP) operator is a simple texture descriptor proposed by (O'Docherty et al. 1991; Ojala et al. 1996; Ojala et al. 2002). It works on the basis that certain local binary patterns are fundamental properties of the texture of local image regions and that

the histogram obtained from the binary patterns can be regarded as a representative texture feature for a given image. Given an image, a binary code is obtained for each pixel by thresholding its nearest neighbours using the pixel's value. A histogram is then constructed based on the occurrences of the different binary patterns produced. It should be noted that the total number of bins used as well as the size of the bins can vary. In the author's implementation, the RGB components are extracted from the image file and stored into three two-dimensional arrays. In other words each array will store the pixel values representing the three components, R, G and B. For each individual array, the surrounding pixel values are determined, producing an 8-bit value. From this 8-bit value, the decimal equivalent is then obtained (see Fig. 5.3). This process is then repeated for the other two arrays. The Red, Green and Blue values are then quantised using 64 bins.

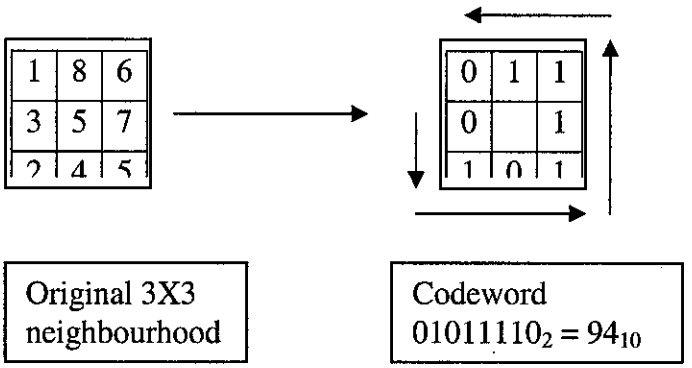


Figure 5.3 Computation of LBP feature

5.5.4 Feature 4

Feature 4 is a combination of the colour histogram, Feature 2, and the binary texture histogram, Feature 3. Since both features are vectors with 64 elements, this feature is a vector of 128 elements in which the first 64 elements are elements from Feature 2 and the next 64 are elements from Feature 3.

5.5.5 Feature 5

Finally feature 5 is a hybrid feature extraction scheme that is developed by the author based on the combination of features 1, 2 and 3. For this hybrid feature extraction, feature I combined Feature 1 with Feature 2 and Feature 3. The new image signature is made up of 192 vector elements (64 from Feature 1, 64 from Feature 2 and the remaining 64 elements from Feature 3).

5.6 Experimental Design

For the experiment three sets of images were used to form a small database, a medium-sized database and a large database. Firstly I selected from our image collection of about 10,000 images a total of 5000 images for which I had established the ground truth. Most of these images were downloaded from http://wang.ist.psu.edu/docs/related (Wang et al. 2001). The 5000 images were then classified into ten categories as listed in Table 5.1. For each set of images I assigned 90% of the images for training and the remaining 10% for testing in a manner similar to the Jackknife technique applied by (Romdhani 1996). Set 1 is made up of 1000 images of which 900 are used for training and 100 for testing. Set 2 is made up of 3000 images of which 2700 are used for training and 300 are used for testing. Finally Set 3 is made up of 5000 images of which 4500 are used for training and 500 are used for testing. The images in the training and testing set were then subjected to the five feature extraction methods and the resulting output feature vectors used as input to the BP network and SVM.

5.6.1 Parameters used for neural network and SVM.

For the back-propagation neural network, the configuration used was a two-layer network with 16 hidden layers and 10 output neurons representing the ten categories. The number of hidden layers was determined empirically. The transfer function in the input layer is tan-sigmoid, and the transfer function in the output layer is linear. The training function used is the scaled conjugate grading algorithm. The learning rate and error rate were set at 0.01 and 0.005 respectively.

Table 5.1 Numbers of images used in each category and database for training and testing

	Set 1		Set 2		Set 3	
	Training	Testing	Training	Testing	Training	Testing
Bikes	90	10	270	30	450	50
Buildings	90	10	270	30	450	50
Cars	90	10	270	30	450	50
Cats	90	10	270	30	450	50
Flowers	90	10	270	30	450	50
Girl	90	10	270	30	450	50
Mountain	90	10	270	30	450	50
Sky	90	10	270	30	450	50
Sunset	90	10	270	30	450	50
Texture	90	10	270	30	450	50
Total	900	100	2700	300	4500	500

5.7 Results

As mentioned earlier in this chapter there are two important objectives for the work described here. One is to compare the performance of a hybrid feature extraction

technique with earlier work as well as other known feature extraction techniques. The other is to compare the suitability and performance of SVM as compared to a back-propagation trained ANN for carrying out classification work.

5.7.1 Comparison of image features using a BP ANN

The classification rate performance of the machine learning algorithms to the small, medium and large database sets is shown in Table 5.2 In terms of individual image features, when applied to BP and using Set 1, it can be seen that Feature 5 (hybrid feature extraction scheme) yields the best result with a classification rate of 58%. This is followed by Feature 3, Feature 2 and Feature 1, with classification rates of 38%, 35% and 31% respectively. The same pattern of performance is obtained for Set 2, with Feature 5 obtaining the highest percentage in classification rate (70.3%). In Set 3, again the top two positions belong to Feature 5 and Feature 4 respectively. However, this time Feature 1 (50.4%) produces a better performance than Feature 2 (49.4%) and Feature 3 (35.2%).

5.7.2 Comparison of image features using SVM

Referring to Table 5.3, for Set 1, the best results is given by Feature 5 with a classification rate of 73%. Second is Feature 2 with a rate of 58%, followed by Feature 4, Feature 3 and Feature 2 respectively. In the results for Set 2, again Feature 5 is first with a rate of 76%, followed by Feature 4 (61%). The next three positions are occupied by Feature 1 (57%), Feature 2 (51.3%) and Feature 3 (42%). The results in Set 3 again show Feature 1 in the first position with a yield of 73.6%. As a matter of fact the standing of the other features follows that of Set 2.

Table 5.2 Percentage classification rates by individual features for a BP ANN classifier.

BP ANN	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
Set 1	31.0	35.0	38.0	54.0	58.0
Set 2	33.7	35.3	37.7	61.0	70.3
Set 3	50.4	49.4	35.2	59.0	65.0

Feature 1: entropy-coding
Feature 2: colour histogram
Feature 3: local binary pattern
Feature 4: Feature 2 + Feature 3
Feature 5: Feature 1 + Feature 2 + Feature 3

Table 5.3 Percentage classification rates by individual features for a SVM classifier.

SVM	Feature 1	Feature 2	Feature 3	Feature 4	Feature 5
Set 1	58.0	39.0	40.0	55.0	73.0
Set 2	57.0	51.3	42.0	61.0	76.0
Set 3	53.6	52.4	42.2	60.0	73.6

Feature 1: entropy-coding
Feature 2: colour histogram
Feature 3: local binary pattern
Feature 4: Feature 2 + Feature 3
Feature 5: Feature 1 + Feature 2 + Feature 3

5.7.3 Comparison between SVM and BP

Tables 5.2 and 5.3 are combined to produce Table 5.4 to provide a better view of the performance between the two machine learning algorithms. From the results obtained in Table 5.4, it is clearly evident that SVM outperforms BP regardless of the image features used as well as the sizes of the image database used. This is clearly consistent with the findings mentioned in Section 5.4 of this thesis.

Table 5.4 Percentage classification rate performance of BP ANN and SVM machine learning algorithms

		BP	SVM
Feature 1	Set 1	31.0	58.0
	Set 2	33.7	57.0
	Set 3	50.4	53.6
Feature 2	Set 1	35.0	39.0
	Set 2	35.3	51.3
	Set 3	49.4	52.4
Feature 3	Set 1	38.0	40.0
	Set 2	37.7	42.0
	Set 3	35.2	42.2
Feature 4	Set 1	54.0	55.0
	Set 2	61.0	61.0
	Set 3	59.0	60.0
Feature 5	Set 1	58.0	73.0
	Set 2	70.3	76.0
	Set 3	65.0	73.6

Feature 1: entropy-coding
Feature 2: colour histogram
Feature 3: local binary pattern
Feature 4: Feature 2 + Feature 3
Feature 5: Feature 1 + Feature 2 + Feature 3

5.8 Summary

In this chapter five feature extraction shemes are introduced. Feature 1, is a feature extraction scheme based on an earlier work, described in 3.4.3. Features 2 and 3 are based on two well-known feature extraction scheme, namely the colour histogram scheme and the binary texture scheme. Feature 4 is a combination of the colour histogram scheme and the binary texture scheme. Finally, feature 5 is the proposed hybrid feature extraction scheme which is a combination of features 1, 2 and 3. Putting it in another way, features 1, 2 and 3 are derived from single image features, whereas features 4 and 5 are made up of a combination of image features. Based on the results obtained it can be said that using multiple image features gives a better performance than using just a single feature. This can be seen from Tables 5.2 and 5.3 whereby features 4 and 5 clearly outperforms features 1, 2 and 3. This could be due to the fact that the images used in the experiments are better represented by using multiple features

rather than a single feature i.e. by using colour or texture alone. By comparing features 4 and 5, it appears that using three image features gives a better result than by using two image features. The obvious question here would probably be in determining the optimal number of image features to be used.

There are two main achievements of this work. Firstly the proposed hybrid feature extraction scheme (Feature 5) has proven to outperform the other image features used in this experiment. The classification rate obtained was the highest in all the three sets it was used, irrespective of the machine learning algorithm used. The second achievement is in confirming the SVM as a better image classifier compared to a back-propagation ANN

Chapter 6 Conclusions and Further Work

6.1 Conclusion

CBIR is still an active research area despite its first appearance more than twenty years ago. This can be seen in the significant number of papers still being published. Research into CBIR is carried in the pixel and compressed domain and within each domain, there are various methods of implementing a CBIR system.

CBIR systems rely on the use of image features, and many have been developed ranging, from, single properties like colour, texture or shape to combination of such properties. Currently there are also various methods of measuring the "similarities" between images (both general and specific). Based on commonly used measurements like the standard Euclidean distance, Hausdorff distance, Earth Movers Distance (EMD) and Mahalanobis distance, other variations have also been developed. Within CBIR systems, different query methods have been developed namely target search, category search and search by association. The suitability of each type of query is application dependent. The multitude of similarity measures and different image databases being used make it difficult to compare the performance of different published systems.

The major work done in this thesis is presented in chapters 3, 4 and 5. In chapter 3, a novel feature extraction scheme operating in the JPEG compressed domain is presented and tested on a database of 3000 images. It is based on the JPEG DCT coefficient coding categories. The proposed system was able to extract rank 1 images

100% of the time, rank 2 images 90% of the time and up to rank 6 images at least 70% of the time.

In chapter 4, the implementation of region-based image retrieval in the compressed domain is presented. The implementation is an extension of the work mentioned in the previous chapter. Four algorithms are developed and compared with two others selected for benchmark purposes. Based on the results obtained, the best performance is obtained by using multiple-key representations of images.

In chapter 5, an experiment is carried out with two main aims. The first is to compare the performance of a hybrid feature extraction scheme with four other feature extraction schemes. From Table 2.1, which gives a summary of existing CBIR systems, it can be seen that most of the systems are developed using multiple features rather than just a single feature, which has been the premise of our feature extraction algorithm mentioned in this chapter. The second is to compare the performance of two machine learning algorithms for carrying out an image classification task. The two proposed machine learning algorithms are the back-propagation Artificial Neural Network and the Support Vector Machine. The experimental results obtained have shown that the hybrid feature extraction scheme outperforms the other feature extraction schemes and that the Support Vector Machine is the better classifier.

6.2 Summary of contributions

The summary of the contributions of this thesis are as follows:

- A feature extraction scheme based on the JPEG coefficient coding categories presented in chapter 3 of the thesis

- ❖ A region-based image retrieval in the compressed domain whereby four algorithms were developed and tested presented in chapter 4
- ❖ A hybrid image feature extraction scheme combining features developed in chapter 3 with image features adopted from colour histogram and the local binary pattern presented in chapter 5
- ❖ Verification of the suitability Support Vector Machines as a classifier for CBIR presented in chapter 5

6.3 Future work.

This thesis in itself does not attempt to solve all the problems faced by the CBIR community. There are still other areas that can be explored as part of future work.

One area that can be explored further is the adoption of an ontology-based approach towards image retrieval. Though it is primarily used in text retrieval, it can also be applied to CBIR as mentioned in the following references. To be explored is the incorporation of Information Retrieval (IR) technology into the CBIR system. Since manual annotation is a non-trivial task, research in this area can be carried out to explore the possibilities of automatically annotating the images. Some of the work that has been done in this area can be found in (Breen et al. 2002; Fu et al. 2004; Maillot et al. 2004; Yang et al. 2004)

Another area to be explored further is the incorporation of relevance feedback (RF) which was originally developed for information retrieval. It is a supervised learning technique used to improve the effectiveness of information retrieval systems. The main idea of relevance feedback is basically to use the responses (both positive and negative) provided by the user to improve the system's performance. For a given query,

the system first retrieves a list of ranked images according to predefined similarity metrics, which are often defined as the distances between feature vectors of images. From the list of retrieved images, the user then selects a set of positive and/or negative examples from it, and the system subsequently refines the query and retrieves a new list of images. The main issue however, is how to incorporate the positive and negative examples to refine the query and how to adjust the similarity measure according to the feedback. RF has been applied successively, as seen by the following references (Ciocca et al. 1999; Duan et al. 2005; Giacinto et al. 2004; Kherfi et al. 2003; MacArthur et al. 2002; Peng 2003; Rui et al. 1997; Stejic et al. 2003; Won Kwak et al. 2003; Zhao et al. 2003; Zhou et al. 2002).

Looking at Table 2.1, it can be seen that all the CBIR systems mentioned have the common feature that they all use more than one image feature for image representation. This has also been confirmed from the results obtained in the previous chapter of this thesis. Thus an obvious direction is do an extensive study of different feature representations in order to find a set of well-balanced features which, on the average, perform as well as possible.

The use of wavelets in CBIR is also becoming popular (Bashar et al. 2003; Chang et al. 1993; Guo et al. 2002a; Huang et al. 2003; Kokare et al. 2004; Mandal et al. 1996; Mandal et al. 1999a). Wavelets have been widely used in many image processing applications including compression, enhancement, reconstruction and image analysis. Basically, a wavelet transformation provides a multi-scale decomposition of the image data. In the field of image retrieval, it is widely used for describing image

texture. Important issues related to wavelet based storage include the choice of decomposition (i.e. choice of filters) appropriate for the different image databases

Another minor area that can be explored further is the addition of an interface system to the current work. However at this juncture and given the time frame, it is not considered to be a high priority as it is regarded as only giving cosmetic value to the system.

In conclusion, the challenge faced by the CBIR community is an inter-related one. In other words, once a feature extraction scheme has been developed there are other aspects that need to be looked into like the similarity measures and the performance measures. It would extremely useful if a suitable benchmarking environment could be developed and taken up by the CBIR research community. This would involve the setting up of a common database and a common set of performance measures.

References

Acir, N., and Guzelis, C. "Automatic recognition of sleep spindles in EEG by using artificial neural networks," *Expert Systems with Applications* (27:3), 2004/10 2004, pp 451-458.

Ardizzoni, S., Bartolini, I., and Patella, M. "Windsurf: A Region-base image retrieval using wavelets," *Proc 4th Int'l Workshop on Database and Expert Systems Application*) 1999.

Armstrong, A., and Jiang, J. "An Efficient Image Indexing Algorithm in JPEG Compressed Domain," Int'l. Conf. of IEEE Consumer Electronics(ICCE'2001), 2001.

Bach, J.R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., and Shu, C.F. "The Virage image search engine: An open framework for image management," Proc. SPIE Storage and Retrieval for Image and Video Databases., 1996, pp. 76-87.

Baharudin, B., and Jiang, J. "A fast imaging indexing technique in JPEG compressed domain," Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on, 2003, pp. 882-885.

Bashar, M.K., Matsumoto, T., and Ohnishi, N. "Wavelet transform-based locally orderless images for texture segmentation," *Pattern Recognition Letters* (24:15), 2003/11 2003, pp 2633-2650.

Baudrier, E., Millon, G., Nicolier, F., and Ruan, S. "A new similarity measure using Hausdorff distance map," Image Processing, 2004. ICIP '04. 2004 International Conference on, 2004, pp. 669-672 Vol. 661.

Beckmann, N., Kriegel, H.-P., Schneider, R., and Seeger, B. "The R| -Tree: An Efficient and Robust Access Method for Points and Rectangles," Proc. ACM SIGMOD, 1990, pp. 322-331.

Benazza-Benyahia, A., and Souidene, W. "A fast retrieval of texture images coded with lifting schemes," Systems, Man and Cybernetics, 2002 IEEE International Conference on, 2002, p. 4 pp. vol.4.

Blaser, A. "Database Techniques for Pictorial Applications," in: *Lecture Notes in Computer Science*, Springer Verlag GmbH, 1979, Vol 81.

Breen, C., Khan, L., and Ponnusamy, A. "Image classification using neural networks and ontologies," Database and Expert Systems Applications, 2002. Proceedings. 13th International Workshop on, 2002, pp. 98-102.

Brezmes, J., Fructuoso, M.L.L., Llobet, E., Vilanova, X., Recasens, I., Orts, J., Saiz, G., and Correig, X. "Evaluation of an electronic nose to assess fruit ripeness," *Sensors Journal, IEEE* (5:1) 2005, pp 97-108.

Carson, C., Belongie, S., Greenspan, H., and Malik, J. "Blobworld: image segmentation using expectation-maximization and its application to image querying," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (24:8) 2002, pp 1026-1038.

Carson, C., Thomas, M., Belongjie, S., Hellerstein, J.M., and Malik, J. "Blobworld: A System for Region-Based Image Indexing and Retrieval," *Proc Visual Information Systems*:June) 1999, pp 509-516.

Castro, R., Mandal, M.K., Ajemba, P., and Istihad, M.A. "An electronic nose for multimedia applications," *Consumer Electronics, IEEE Transactions on* (49:4) 2003, pp 1431-1437.

Chan, S.W.K., and Chong, M.W.C. "Unsupervised clustering for nontextual web document classification," *Decision Support Systems* (37:3), 2004/6 2004, pp 377-396.

Chang, C.-C., Chuang, J.-C., and Hu, Y.-S. "Retrieving digital images from a JPEG compressed image database," *Image and Vision Computing* (22:6), 2004/6/1 2004, pp 471-484.

Chang, S.K., and Hsu, A. "Image information systems: where do we go from here?," *IEEE Transactions on Knowledge and Data Engineering* (5:5) 1992, pp 431-442.

Chang, T., and Kuo, C.C.J. "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans on Image Processing* (2:4), Oct 1993, pp 429-441.

Chen, J.L., and Kundu, A. "Rotation and gray scale invariant texture identification using wavelet decomposition and hidden Markov model," *IEEE Trans on Pattern Analysis and Machine Intelligence* (16:2), February 1994, pp 208-214.

Chiang, R.H.L., Chua, C.E.H., and Storey, V.C. "A smart web query method for semantic retrieval of web data," *Data & Knowledge Engineering* (38:1), 2001/7 2001, pp 63-84.

Ciocca, G., and Schettini, R. "A relevance feedback mechanism for content-based image retrieval," *Information Processing & Management* (35:5), 1999/9 1999, pp 605-632.

Climer, S., and Bhatia, S.K. "Image database indexing using JPEG coefficients," *Pattern Recognition* (35:11), 2002/11 2002, pp 2479-2488.

Cole, M., Sehra, G., Gardner, J.W., and Varadan, V.K. "Development of smart tongue devices for measurement of liquid properties," *Sensors Journal, IEEE* (4:5) 2004, pp 543-550.

Cristianini, N., and Shawe-Taylor, J. *An Introduction to Support Vector Machines* Cambridge University Press, Cambridge, 2000.

Daqi, G., Qin, M., and Guiping, N. "Simultaneous estimation of odor classes and concentrations using an electronic nose," Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, 2004, pp. 1353-1358 vol.1352.

Distante, C., Ancona, N., and Siciliano, P. "Support vector machines for olfactory signals recognition," *Sensors and Actuators B: Chemical* (88:1), 2003/1/1 2003, pp 30-39.

Duan, L., Gao, W., Zeng, W., and Zhao, D. "Adaptive relevance feedback based on Bayesian inference for image retrieval," *Signal Processing* (85:2), 2005/2 2005, pp 395-399.

Egmont-Petersen, M., de Ridder, D., and Handels, H. "Image processing with neural networks--a review," *Pattern Recognition* (35:10), 2002/10 2002, pp 2279-2301.

Enser, P.G.B. "Query Analysis in a Visual Information Retrieval Context.," *Journal of Document and Text Management,* (1:1) 1993, pp 25-39.

Feng, G., and Jiang, J. "JPEG compressed image retrieval via statistical features," *Pattern Recognition* (36:4), 2003/4 2003, pp 977-985.

Forsyth, D.A. "Benchmarks for storage and retrieval in multimedia databases.," *SPIE Storage and Retrieval for Media Databases III* (4676) 2002.

Freund, Y., and Schapire, R.E. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,," *Journal of Computer and System Sciences* (55:1), 1997/8 1997, pp 119-139.

Fu, H., Chi, Z., Feng, D., and Song, J. "Machine learning techniques for ontology-based leaf classification," Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th, 2004, pp. 681-686 Vol. 681.

Fudos, I., and Palios, L. "An efficient shape-based approach to image retrieval," *Pattern Recognition Letters* (23:6), 2002/4 2002, pp 731-741.

Furht, B., Smoliar, S.W., and Zhang, H. *Video and Image Processing in Multimedia Systems* Kluwer Academic Publishers, 1995.

Gagaudakis, G., and Rosin, P.L. "Incorporating shape into histograms for CBIR," *Pattern Recognition* (35:1), 2002/1 2002, pp 81-91.

Gardener, J. "Electronic tongues," MEMS Sensor Technologies, 2005. The IEE Seminar and Exhibition on, 2005, p. 26 pp.

Giacinto, G., and Roli, F. "Bayesian relevance feedback for content-based image retrieval," *Pattern Recognition* (37:7), 2004/7 2004, pp 1499-1508.

Gong, Y., Zhang, H., Chuant, H., and Sakauuchi, M. "An image database system with content capturing and fast image indexing abilities," Proceedings of the International Conference on Multimedia Computing and Systems, 1994, p. 121–130.

Grauman, K., and Darrell, T. "Fast contour matching using approximate earth mover's distance," Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2004, pp. I-220-I-227 Vol.221.

Greenspan, H., Dvir, G., and Rubner, Y. "Context-dependent segmentation and matching in image databases," *Computer Vision and Image Understanding* (93:1), 2004/1 2004, pp 86-109.

Gunther, N.J., and Beratta, G.B. "Benchmark for image retrieval using distributed systems over the internet: BIRDSI," *Internet Imaging III, SPIE*) 2001, pp 252-267.

Guo, B., and Jiang, J. "A modified shape descriptor in wavelets compressed domain," Image Processing. 2002. Proceedings. 2002 International Conference on, 2002, pp. I-936-I-939 vol.931.

Guo, G.-D., Jain, A.K., Ma, W.-Y., and Zhang, H.-J. "Learning Similarity Measure for Natural Image Retrieval With Relevance Feedback," *IEEE Transactions on Neural Networks* (13:4) 2002b, pp 811-820.

Guocan., F., B., B., and J., J. "A Comprehensive Progressive Decoding for JPEG Compressed Image," Electronic Imaging, Santa Clara, California, USA, 2002.

Hafner, J., Sawhney, H.S., Equitz, W., Flickner, M., and Niblack, W. "Efficient Color Histogram Indexing for Quadratic Form Distance Function," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (17:7) 1995, pp 729-736.

Han, J., and Ma, K.-K. "Fuzzy Color Histogram and Its Use in Color Image Retrieval," *Image Processing, IEEE Transactions on* (11:8) 2002, pp 944-952.

Han, J.W., and Guo, L. "A shape-based image retrieval method using salient edges," *Signal Processing: Image Communication* (18:2), 2003/2 2003, pp 141-156.

Hauptmann, P., Borngraeber, R., Schroeder, J., and Auge, J. "Artificial electronic tongue in comparison to the electronic nose. State of the art and trends," Frequency Control Symposium and Exhibition, 2000. Proceedings of the 2000 IEEE/EIA International, 2000, pp. 22-29.

Haykin, S. *Neural Networks: A Comprehensive Foundation* Prentice-Hall International, Englewood Cliff, NJ, 1999.

Helsingius, M., Kuosmanen, P., and Astola, J. "Image compression using multiple transforms," *Signal Processing: Image Communication, Vol.15* (15) 2000, pp 513-529.

Hidderly, R., and Rafferty, P. "Democratic Indexing: An Approach to the Retrieval of Film.," Proceedings of Library and Information Studies: Research and Professional Practice., Taylor Graham., Queen Margaret College, Edinburgh, 1997.

Hitchcock, F.L. "The distribution of a product from several sources to numerous localities.," *Journal of Mathematical Physics* (20) 1941, pp 224-230.

Hollink, L., Schreiber, A.T., Wielinga, B.J., and Worring, M. "Classification of user image descriptions," *International Journal of Human-Computer Studies* (61:5), 2004/11 2004, pp 601-626.

Horn, D. "Neural Computation Methods and Applications: Summary Talk of the AI," *Journal of Nuclear Instruments and Methods in Physics Research* (389) 1997, pp 381-387.

Huang, J., Kumar, S.R., Mitra, M., Zhu, W.-J., and Zabih, R. "Image Indexing Using Color Correlograms," *Internat. Conf. Computer Vision Pattern Recognition*) 1997, pp 762-768.

Huang, X.-Y., Zhong, Y.-J., and Hu, D. "Image retrieval based on weighted texture features using DCT coefficients of JPEG images," Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on, 2003, pp. 1571-1575 vol.1573.

Huang, Y.L., and Chang, R.F. "Texture features for DCT-coded image retrieval and classification," *Proc IEEE Acoustics, Speech and Signal Processing* (6) 1999, pp 15-19.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., and Wu, S. "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision Support Systems* (37:4), 2004/9 2004, pp 543-558.

Jain, R. "Visual Information Management Systems," US NSF Workshop, 1992.

Jeong, S., Won, C.S., and Gray, R.M. "Image retrieval using color histograms generated by Gauss mixture vector quantization," *Computer Vision and Image Understanding* (94:1-3), 2004/0 2004, pp 44-66.

Jermyn, I.H., Shaffrey, C.W., and Kingsbury, N.G. "Evaluation methodologies for image retrieval systems," Proceedings of ACIVS 2002 (Advanced Concepts for Intelligent Vision Systems), Ghent, Belgium, 2002.

Jorgensen, C. "Towards an Image Testbed for Benchmarking Image Indexing and Retrieval Systems," Multimedia Content-Based Indexing and Retrieval Workshop 2001 (MMCBIR 2001), Rocquencourt, France, 2001, pp. 101-106.

Kageyama, R., Kagami, S., Inaba, M., and Inoue, H. "Development of soft and distributed tactile sensors and the application to a humanoid robot," Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on, 1999, (2) pp. 981-986.

Keister, L.H. "User Types and Queries: Impact on Image Access Systems, Challenges in Indexing Electronic Text and Images, R. Fidel, Editor. American Society for Information Science. 1994, p. 7-19.,") 1994.

Kermani, B.G., Schi!man, S.S., and Nagle, H.T. "Using neural networks and genetic algorithms to enhance performance in an electronic nose.," *IEEE Transactions on Biomedical Engineering* (46:4) 1999, pp 429-439.

Kherfi, M.L., Ziou, D., and Bernardi, A. "Combining positive and negative examples in relevance feedback for content-based image retrieval," *Journal of Visual Communication and Image Representation* (14:4), 2003/12 2003, pp 428-457.

Kim, C.-R., and Chung, C.-W. "A multi-step approach for partial similarity search in large image data using histogram intersection," *Information and Software Technology* (45:4), 2003/3/15 2003, pp 203-215.

Kim, W.-Y., and Kim, Y.-S. "A region-based shape descriptor using Zernike moments," *Signal Processing: Image Communication* (16:1-2), 2000/9 2000, pp 95-102.

Ko, B., and Byun, H. "Integrated region-based image retrieval using region's spatial relationships," Pattern Recognition, 2002. Proceedings. 16th International Conference on, 2002, pp. 196-199 vol.191.

Ko, B., and Byun, H. "FRIP: A Region-Based Image Retrieval Tool Using Automatic Image Segmentation and Stepwise Boolean AND Matching," *IEEE Transactions on Multimedia* (7(1):February) 2005, pp 105-113.

Kokare, M., Chatterji, B.N., and Biswas, P.K. "Comparison of similarity metrics for texture image retrieval," TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region, 2003, pp. 571-575 Vol.572.

Kokare, M., Chatterji, B.N., and Biswas, P.K. "Cosine-modulated wavelet based texture features for content-based image retrieval," *Pattern Recognition Letters* (25:4), 2004/3 2004, pp 391-398.

Krishna, G.M., and Rajanna, K. "Tactile sensor based on piezoelectric resonance," *Sensors Journal, IEEE* (4:5) 2004, pp 691-697.

Kumar, C.S., and Wei, F.S. "A bilingual speech recognition system for English and Tamil," Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on, 2003, pp. 1641-1644 vol.1643.

Laaksonen, J., Koskela, M., Laakso, S., and Oja, E. "PicSOM - content-based image retrieval with self-organizing maps," *Pattern Recognition Letters* (21:13-14), 2000/12 2000, pp 1199-1207.

Laaksonen, J., Koskela, M., and Oja, E. "PicSOM - Self-Organizing Image Retrieval with MPEG-7 Content Descriptors," *Neural Networks, IEEE Transactions on* (13:4), July 2002 2002, pp 841-853.

Lee, K.-L., and Chen, L.-H. "An efficient computation method for the texture browsing descriptor of MPEG-7," *Image and Vision Computing* (23:5), 2005/5/1 2005, pp 479-489.

Leung, C.H.C., and Ip, H.H.S. "Benchmarking for content-based visual information search," Proc 4th Int. Conf. Visual Inform. Syst., Lyon, France, 2000, pp. 442-456.

Lindquist, M., and Wide, P. "Virtual water quality tests with an electronic tongue," Instrumentation and Measurement Technology Conference, 2001. IMTC 2001. Proceedings of the 18th IEEE, 2001, pp. 1320-1324 vol.1322.

Lvova, L., De Angelis, G., Montieri, C., Primadei, T., Martinelli, E., Mazzone, E., Pede, A., Paolesse, R., Di Natale, C., and D'Amico, A. "An 'electronic tongue' system based on an array of metallic potentiometric sensors," Sensors, 2004. Proceedings of IEEE, 2004, pp. 233-235 vol.231.

Ma, Q., Yan, A., Hu, Z., Li, Z., and Fan, B. "Principal Component Analysis and Artificail Neural Networks Applied to the Classification of Chinese Pottery Neolithic Age," *Analytica Chimica Acta* (406) 2000, pp 247-256.

Ma, W.Y., and Manjunath, B.S. "NeTra: A Toolbox for Navigating Large Image Databases," *Proc. IEEE Int'l Conf. Image Processing*), October 1997 1997, pp 568-571.

MacArthur, S.D., Brodley, C.E., Kak, A.C., and Broderick, L.S. "Interactive Content-Based Image Retrieval Using Relevance Feedback," *Computer Vision and Image Understanding* (88:2), 2002/11 2002, pp 55-75.

Maillot, N., Thonnat, M., and Hudelot, C. "Ontology based object learning and recognition: application to image retrieval," Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on, 2004, pp. 620-625.

Mandal, M.K., Aboulnasr, T., and Panchanathan, S. "Image indexing using moments and wavelets," *IEEE Trans on Consumer Electronics* (42:3), August 1996, pp 557-565.

Mandal, M.K., Aboulnasr, T., and Panchanathan, S. "Fast Wavelet Histogram Techniques for Image Indexing," *Computer Vision and Image Understanding* (75:1-2), 1999/7 1999a, pp 99-110.

Mandal, M.K., Idris, F., and Panchanathan, S. "A critical evaluation of image and video indexing techniques in the compressed domain," *Image and Vision Computing Journal* (17) 1999b, pp 513-529.

Manohar, M., and Tilton, J.C. "Model-based vector quantization with application to remotely sensed image data," *IEEE Transactions on Image Processing* (8:11) 1999, pp 1630-1638.

Messer, K., and Kittler, J. "A region-based image database system using colour and texture," *Pattern Recognition Letters* (20:11-13), 1999/11 1999, pp 1323-1330.

Mostafa, J., and Dillon, A. "Design and Evaluation of a User Interface Supporting Multiple Image Query Models.," Proceedings of the 59th Annual Conference of the American Society for Information Science: Global Complexity, Information,

Chaos and Control., American Society for Information Science., Baltimore, Maryland, USA, 1996.

Mukai, T. "Soft areal tactile sensors with embedded semiconductor pressure sensors in a structured elastic body," Sensors, 2004. Proceedings of IEEE, 2004, pp. 1518-1521 vol.1513.

Mukhopadhyay, R., Ma, A., and Sethi, I.K. "Pathfinder networks for content based image retrieval based on automated shape feature discovery," Multimedia Software Engineering, 2004. Proceedings. IEEE Sixth International Symposium on, 2004, pp. 522-528.

Muller, H., Muller, W., Squire, D.M., Marchand-Maillet, S., and Pun, T. "Performance evaluation in content-based image retrieval: overview and proposals," *Pattern Recognition Letters* (22:5), 2001/4 2001, pp 593-601.

Nagle, H.T., Schiffman, S.S., and Gutierrez-Osuna, R. "The how and why of electronic noses.," *IEEE Spectrum* (35:9) 1998, pp 22-34.

Natsev, A., Rastogi, R., and Shim, K. "WALRUS: A Similarity Retrieval Algorithm for Image Databases," *IEEE Transactions on Knowledge and Data Engineering* (16:3), March 2004, pp 301-316.

Nezamabadi-pour, H., and Kabir, E. "Image retrieval using histograms of uni-color and bi-color blocks and directional changes in intensity gradient," *Pattern Recognition Letters* (25:14), 2004/10/15 2004, pp 1547-1557.

Ng, I., Tan, T., and Kittler, J. "On local linear transform and Gabor filter representation of texture," *Proc. Intl Conf. Pattern Recognition*) 1992, pp 627-631.

Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D., Yanker, P., and Faloutsos, C. "The QBIC project: querying images by content using color, texture and shape.," *SPIE Storage and Retrieval for Image and Video Databases I* (1908) 1993, pp 173-187.

O'Docherty, M., and Daskalakis, C. "Multimedia information systems: the management and semantic retrieval of all electronic datatypes," *The Computer Journal* (34:3) 1991, pp 225-238.

Ojala, T., Pietikainen, M., and Harwood, D. "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition (29),* 1996, pp 51-59.

Ojala, T., Pietikainen, M., and Maenpaa, T. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell* (24:7) 2002, pp 971-987.

Pal, M., and Mather, P.M. "Assessment of the effectiveness of support vector machines for hyperspectral data," *Future Generation Computer Systems* (20:7), 2004/10/1 2004, pp 1215-1225.

Pass, G., and Zabith, R. "Histogram refinement for content-based image retrieval," *IEEE Workshop on Applications of Computer Vision*) 1996, pp 96-102.

Pastra, K., Saggion, H., and Wilks, Y. "Extracting relational facts for indexing and retrieval of crime-scene photographs," *Knowledge-Based Systems* (16:5-6), 2003/7 2003, pp 313-320.

Peng, J. "Multi-class relevance feedback content-based image retrieval," *Computer Vision and Image Understanding* (90:1), 2003/4 2003, pp 42-67.

Pennebaker, W.B., and Mitchell, J.L. *JPEG still image data compression standard* Van Nostrand Reinhold, New York, 1993.

Pentland, A., Picard, R.W., and Sclaroff, S. "Photobook: Tools for Content-Based Manipulation of Image Databases," *Proc of SPIE: Storage and Retrieval fro Image and Video Databases II* (2185:Feb) 1994, pp 34-47.

Podder, S.K., Shaban, K., Sun, J., Karray, F., Basir, O., and Kamel, M. "Performance improvement of automatic speech recognition systems via multiple language models produced by sentence-based clustering," Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003 International Conference on, 2003, pp. 362-367.

Puzicha, J., Buhmann, J.M., Rubner, Y., and Tomasi, C. "Empirical evaluation of dissimilarity measures for color and texture," Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, 1999, pp. 1165-1172 vol.1162.

Qu, M., SHih, F.Y., Jing, J., and Wang, H. "Automatic Solar Flare Detection Using MLP, RBF and SVM," *Solar Physics* (217:1), October 2003, pp 152-172.

Reeves, R., Kubik, K., and Osberger, W. "Texture characterization of compressed aerial images using DCT coefficients," *Proc SPIE, Storage and Retrieval for Image and Video Databases V* (3022:February) 1997, pp 398-407.

Rigoll, G. "Maximum mutual information neural networks for hybrid connectionist-HMM speech recognition systems," *Speech and Audio Processing, IEEE Transactions on* (2:1) 1994, pp 175-184.

Romdhani, S. "Face Recognition Using Principal Component Analysis," *MSc. Thesis, University of Glasgow*) 1996.

Rubner, Y., Guibas, L.J., and Tomasi, C. "The Earth Mover's Distance, Multidimensional Scaling and Colour Image Retrieval," *Proc DARPA Image Understanding Workshop*, 1997, pp 661-668.

Rui, Y., Huang, T.S., and Chang, S.-F. "Image Retrieval: Current Techniques, Promising Directions, and Open Issues," *Journal of Visual Communication and Image Representation* (10:1), 1999/3/1 1999, pp 39-62.

Rui, Y., Huang, T.S., and Mehrotra, S. "Content based image retrieval with relevance feedback in MARS," *IEEE Int'l Conference on Image Processing, Santa Barbara*:October) 1997, pp 815-818.

Rui, Y., Thomas, S., Ortega, M., and Mehrotra, S. " Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans.Circuits Syst. Video Technol* (8:5), September 1998, pp 644-655.

Rumelhart, D.E., and McClelland, J.L. "Parallel Distributed Processing: Explorations in the Microstructure of Cognite," The MIT Press, 1986, pp. 318-362.

Saber, E., and Tekalp, A.M. "Region-based Shape Matching for Automatic Image Annotation and Query-by-Example.," *Journal of Visual Communication and Image Representation* (8:1), March 1997, pp 3-20.

Sawhney, H.S., and Hafner, J.L. "Efficient color histogram indexing," *Proceedings of ICIP '94* (2) 1994, pp 66-70.

Scarlogg, S., Taycher, L., and Cascia., M.L. "Image Rover: A Content-based Image Browser for the WorldWideWeb," In Proc. of the IEEE Workshop on Content-based Access of Image and Video Libraries, Puerto Rico, 1997, pp. 10-18.

Schmid, C., and Mohr, R. "Local grayvalue invariants for image retrieval," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (19:5) 1997, pp 530-535.

Sebe, N., Tian, Q., Loupias, E., Lew, M.S., and Huang, T.S. "Evaluation of salient point techniques," *Image and Vision Computing* (21:13-14), 2003/12/1 2003, pp 1087-1095.

Sheikholeslami, G., Chang, W., and Zhang, A. "SemQuery: Semantic Clustering and Querying on Heterogeneous Features for Visual Data," *Knowledge and Data Engineering, IEEE Transactions on* (14:5) 2002, pp 988-1002.

Shen, B., and Sethi, I.K. "Direct feature extraction from compressed images," *Proc SPIE Storage & Retrieval for Image and Video Databases IV.* (2670) 1996, pp 404-414.

Shneier, M., and Mottaleb, M.A. "Exploiting the JPEG compression scheme for image retrieval," *IEEE Trans. On Pattern Analysis and Machine Intelligence* (18:August) 1996, pp 849-853.

Smeulders, A.W.M., Kersten, M.L., and Gevers, T. "Crossing the Divide between Computer Vision and Databases in Search of Image Databases," Proc Fourth Working Conf Visual Databse Systems, 1998, pp. 223-239.

Smeulders, A.W.M., Santini, S., Gupta, A., and Jain, R. "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. On Pattern Analysis and Machine Intelligence* (22:12), December 2000 2000, pp 1349-1380.

Smith, J.R. "Integrated Spatial and Feature Image Systems: Retrieval, Compression and Analysis.," in: *Graduate School of Arts and Sciences*, Columbia University, 1997.

Smith, J.R., and Chang, S.F. "Transform features for texture classification and discrimination in large databases," *Proc IEEE Intl. Conf. On Image Processing* (3) 1994, pp 407-411.

Smith, J.R., and Chang, S.-F. "Tools and techniques for color image retrieval," *Proceedings, IS&T/, SPIE Symposium on Electronic Imaging: Science and Technology (EI '96)—Storage and Retrieval for Image and Video Databases IV* (2670) 1996a, pp 426-437.

Smith, J.R., and Chang, S.-F. "VisualSEEK: a fully automated content-based image query system," *ACM Multimedia*:Nov. 1996) 1996b, pp 87-98.

Stejic, Z., Takama, Y., and Hirota, K. "Genetic algorithm-based relevance feedback for image retrieval using local similarity patterns," *Information Processing & Management* (39:1), 2003/1 2003, pp 1-23.

Stricker, M., and Orengo, M. "Similarity of Color Images," Storage and Retrieval for Image and Video Databases III, I&ST/SPIE, San Jose, CA, USA, 1995, pp. 381-392.

Surong, W., Liang-Tien, C., and Rajan, D. "Efficient image retrieval using MPEG-7 descriptors," Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, 2003, pp. III-509-512 vol.502.

Swain, M.J. "Interactive indexing into image databases," Proc. SPIE: Storage Retrieval Image Video Databases, 1993, pp. 95-103.

Swain, M.J., and Ballard, D.H. "Color Indexing," *International Journal of Computer Vision* (7:1) 1991, pp 11-32.

Tamura, H., Mori, S., and Yamawaki, T. "Image database systems: A survey," *Pattern Recognition* (17:1) 1984, pp 29-43.

Tilton, J.C., Manohar, M., and Newcomer, J.A. "Earth science data compression issues and activities," *Remote Sens. Rev.* (9) 1994, pp 271-298.

Tong, F.H.F., and Zhang, D. "A new progressive color image transmission scheme for the world wide web," *Computer networks and ISDN systems* (30) 1998, pp 2059-2064.

Town, C., and Sinclair, D. "Language-based querying of image collections on the basis of an extensible ontology," *Image and Vision Computing* (22:3), 2004/3/1 2004, pp 251-267.

Tzafestas, E., Nikolaidou, A., and Tzafestas, S. "Performance Evaluation and Dynamic Node Generation Criteria for 'Principal Component Analysis' Neural Network," *Mathematics and Computer Simulation* (51) 2000, pp 145-156.

Vapnik, V. *The nature of statistical learning theory* New York: Springer-Verlag., 1995.

Voyles, R.M., Jr., Fedder, G., and Khosla, P.K. "Design of a modular tactile sensor and actuator based on an electrorheological gel," Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on, 1996, pp. 13-17 vol.11.

Wang, J.Z., Li, J., and Wiederhold, G. "SIMPLIcity: Semantics-Sensitive Integrated matching for Picture Libraries," *IEEE Trans. On Pattern Analysis and Machine Intelligence* (23:9), September 2001, pp 947-963.

Widrow, B., Rumelhart, D.E., and Lehr, M.A. "Neural networks: applications in industry, business, and science," *Communications of the ACM*) 1994, pp 93-105.

Won Kwak, J., and Cho, N.I. "Relevance feedback in content-based image retrieval system by selective region growing in the feature space," *Signal Processing: Image Communication* (18:9), 2003/10 2003, pp 787-799.

Yang, C., Dong, M., and Fotouhi, F. "Learning the Semantics in Image Retrieval - A Natural Language Processing Approach," Computer Vision and Pattern Recognition Workshop, 2004 Conference on, 2004, p. 137.

Yang, Z., and Kuo, C.C.J. "Survey on Image Content Analysi, Indexing, and Retrieval Techniques and Status Report of MPEG-7," *Tamkang Journal of Science and Engineering* (2:3) 1999, pp 101-118.

Yuan, T., Yu, N., and Li, X. "Image retrieval with EMD for new perceptual color feature," Neural Networks and Signal Processing, 2003. Proceedings of the 2003 International Conference on, 2003, pp. 965-968 Vol.962.

Zhao, T., Tang, L.H., Ip, H.H.S., and Qi, F. "On relevance feedback and similarity measure for image retrieval with synergetic neural nets," *Neurocomputing* (51), 2003/4 2003, pp 105-124.

Zhou, X.S., and Huang, T.S. "Relevance feedback in content-based image retrieval: some recent advances," *Information Sciences* (148:1-4), 2002/12 2002, pp 129-137.

Zhuge, H. "Retrieve images by understanding semantic links and clustering image fragments," *Journal of Systems and Software* (73:3), 2004/0 2004, pp 455-466.

Author's contribution

1. Guocan F., Baharudin B., "A Comprehensive Progressive Decoding for JPEG Compressed Image", IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology, Santa Clara, California, USA

2. Baharudin B., Jiang J., "Compressed Image Retrieval via Entropy Codes", Postgraduate Research Conference in Electronics, Photonics, Communications & Networks and Computer Science, Exeter, United Kingdom, 2003.

3. Baharudin B., Jiang J., "A Fast Imaging Indexing Technique in JPEG Compressed Domain, ICECS 2003, Sharjah, Vol 2, pp 882-885

4. Baharudin B., Ipson S., Jiang J., "A Fast Region-Based Indexing Technique in JPEG Compressed Domain", CACSUK 2004, Liverpool, United Kingdom, 2004.

5. Baharudin, B., Ipson S., Jiang J., "Region_based Indexing Technique in JPEG Compressed Domain", IEE Int'l Conf , VIE 2005, pp 287-291, Glasgow, United Kingdom, 2005.

6. Qahwaji R., Baharudin B., Jiang J., "A Feasibility Study Towards the use of Support Vector Machines for Content-Based Image Retrieval", submitted to ICECS 2005, Tunisia.

7. Qahwaji R., Baharudin B., Jiang J., "Using a Hybrid Feature Extraction Scheme for Image Classification", submitted to Pattern Recognition.