CERTIFICATION OF APPROVAL

CONTENT BASED IMAGE RETRIEVAL (CBIR) SYSTEM

By

MOHAMAD HASBULLAH BIN ABU BAKAR
5035

A Project dissertation submitted to the
Business Information System Programme
Universiti Teknologi PETRONAS
in partial fulfillment of
the requirements for the
Bachelor of Business Information System (Hons)

APRIL 2007

Approved by,

(Ms Vivian Yong Suet Peng)
Supervisor

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein has not been undertaken or done by unspecified sources or persons.

(Mohamad Hasbullah bin Abu Bakar)

ABSTRACT

Advancement in hardware and telecommunication technology has boosted up creation and distribution of digital visual content. However this rapid growth of visual content creations has not been matched by the simultaneous emergence of technologies to support efficient image analysis and retrieval. Although there are attempt to solve this problem by using meta-data text annotation but this approach are not practical when it come to the large number of data collection.

This system used 7 different feature vectors that are focusing on 3 main low level feature groups (color, shape and texture). This system will use the image that the user feed and search the similar images in the database that had similar feature by considering the threshold value. One of the most important aspects in CBIR is to determine the correct threshold value. Setting the correct threshold value is important in CBIR because setting it too low will result in less image being retrieve that might exclude relevant data. Setting to high threshold value might result in irrelevant data to be retrieved and increase the search time for image retrieval.

Result show that this project able to increase the image accuracy to average 70% by combining 7 different feature vector at correct threshold value.

ACKNOWLEDGEMENT

First and foremost I would like to take this opportunity to thanks to God because He has give me the opportunity to finish this project. I would like to express my greatest gratitude to Ms. Vivian Yong Suet Peng, who had give me supportive and constructive advises.

Next, I would like to acknowledge University Teknologi PETRONAS (UTP) specially the final year research project committee for proper organization and management of this course.

Last but not least, I would like to acknowledge my parent for their endless support during this project. They have given me strength, motivation and determination in all my undertaking.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Background of Study

Recent advancement in hardware and telecommunication technology has boosted up creation and distribution of digital visual content. However this rapid growth of visual content creations has not been matched by the simultaneous emergence of technologies to support efficient image analysis and retrieval. To cater these issues, Metadata-based image retrieval had been used. Metadata-based image retrieval relies on the human inputted metadata such as keywords and caption. Images were first annotated with the text and then searched using text/metadata based approach from traditional database management system. Metadata-based system requires human to describe every image in the database. This is not practical for a very large image database or for image that are generated automatically like satellite image or surveillance cameras. Image classification has ability to find the relationship between objects in images. To solve this problem Content-based Image Retrieval had come to the scene. Content-based Image Retrieval techniques used the content of the digital image itself rather the metadata of the digital image.

Various methods had been studied like SVM [1], decision tree and association rule-based approach [2] for image classification by utilizing the image features like Color, Shape and Texture [3]. Semantic of whole image can not represented only by its low-level features. To reach near human perception there are many approaches like Multilevel Association Rules, Region Silent and Data Mining Approach have been developed to map low-level feature to high-level semantic. In general, image classification involves with following mains issues [4]:

i. Extraction of image features
ii. Organization and representation of image features

iii. Building of effective classifier

iv. Semantic structure modeling

According to underlying semantic, images can be classified in two ways [4]: i) Classify by some main object, ii) Classify by multiple objects with their relations.

In order to access a desired image from a repository, it involves a search for images depicting specific types of object or scene, evoking a particular mood, or simply containing a specific texture or pattern. Potentially, images have many types of attribute which could be used for retrieval, including:

- the presence of a particular combination of color, texture or shape features (e.g. blue car);
- the presence or arrangement of specific types of object (e.g. a cat in a basket);
- the depiction of a particular type of event (e.g. a rugby match);
- the presence of named individuals, locations, or events (e.g. A people blowing a candle);
- subjective emotions one might associate with the image (e.g. happiness);
- metadata such as who created the image, where and when.

Each listed query type (with the exception of the last) represents a higher level of abstraction than its predecessor, and each is more difficult to answer without reference to some body of external knowledge. This leads naturally on to a classification of query types into three levels of increasing complexity [5]:

Level 1

Comprises retrieval by *primitive* features such as color, texture, shape or the spatial location of image elements. Examples of such queries might include "find images containing yellow stars arranged in a ring" – or most commonly "find me more pictures that look like this". This level of retrieval uses features (such as a given shade of yellow) which are both objective, and directly derivable from the images themselves, without the need to refer to any external knowledge base. Its use is largely limited to specialist applications such as trademark registration, identification of drawings in a design archive, or color matching of fashion accessories.

Level 2

Comprises retrieval by *derived* (sometimes known as *logical*) features, involving some degree of logical inference about the identity of the objects depicted in the image. It can usefully be divided further into:

a) retrieval of objects of a given type (e.g. "find pictures of a horse");

b) retrieval of individual objects or persons ("find a picture of the KLCC Tower").

To answer queries at this level, reference to some outside store of knowledge is normally required – particularly for the more specific queries at level 2(b). In the first example above, some prior understanding is necessary to identify an object as a bus rather than a lorry; in the second example, one needs the knowledge that a given individual structure has been given the name "the KLCC Tower". Search criteria at this level, particularly at level 2(b), are usually still reasonably objective. This level of query is more generally encountered than level 1.

Level 3

Comprises retrieval by *abstract* attributes, involving a significant amount of high-level reasoning about the meaning and purpose of the objects or scenes depicted. Again, this level of retrieval can usefully be subdivided into:

a) retrieval of named events or types of activity (e.g. "find pictures of Chinese wedding");

b) retrieval of pictures with emotional or religious significance ("find a picture depicting suffering").

1.2 Problem Statement

Images now play a crucial role in fields as diverse as medicine, military, engineering, education and entertainment Image can convey information better than text. Organization with large image database have problem in searching specific image. In a normal image searching system, system will look for image metadata (description of image). This metadata requires human to manually describe every image in the database. This approach is impractical for a large image databases, or for image that are automatically like surveillance cameras. Manual image description consume time and cost and not effective. Metadata searching usually failed to meet image descriptions that match human perception [6]. So there are the need for effective image classification and retrieval system that automatic extract information from the images.

1.3 Objective and Scope of Study

The objectives of this project are to:

- Research on the methodologies that has been applied on developing image retrieval system
- Propose an algorithm for searching and retrieving images.

The system will assist users to find/query specific image from large collection of image that meet user description. This search can be based on the specific object that the user wishes to find in the image until query type level 1.

Chapter 2

Literature Review

2.1 Practical Application of Image Classification and Retrieval.

A wide range of possible applications for CBIR technology has been identified [7]. Potentially fruitful areas include:

- Crime prevention
- The military
- Intellectual property
- Architectural and engineering design
- Medical diagnosis
- Web searching.

Below are the details descriptions of how Image Classification and Retrieval being applied:

Crime prevention

Law enforcement agencies typically maintain large archives of visual evidence, including past suspects' facial photographs (generally known as mugshots), fingerprints, tyre treads and shoeprints. Whenever a serious crime is committed, they can compare evidence from the scene of the crime for its similarity to records in their archives. Example, the Finger Search Engine from East Shore Technologies, Inc. of New York (http://www.east-shore.com/)

The military

Military applications of imaging technology are probably the best-developed, though least publicized. Recognition of enemy aircraft from radar screens, identification of targets from satellite photographs, and provision of guidance systems for cruise missiles are known examples – though these almost certainly represent only the tip of the iceberg.

Many of the surveillance techniques used in crime prevention could also be relevant to the military field.

Intellectual property

CBIR had been used in detecting Intellectual property. Copyright owners able to seek out and identify unauthorized copies of images, particularly if they have been altered in some way.

Architectural and engineering design

Architectural and engineering design share a number of common features – the use of stylized 2- and 3-D models to represent design objects, the need to visualize designs for the benefit of non-technical clients, and the need to work within externally-imposed constraints, often financial. Such constraints mean that the designer needs to be aware of previous designs, particularly if these can be adapted to the problem at hand. Hence the ability to search design archives for previous examples which are in some way similar, or meet specified suitability criteria, can be valuable.

Medical diagnosis

The increasing reliance of modern medicine on diagnostic techniques such as radiology, histopathology, and computerized tomography has resulted in an explosion in the number and importance of medical images now stored by most hospitals. While the prime requirement for medical imaging systems is to be able to display images relating to a named patient, there is increasing interest in the use of CBIR techniques to aid diagnosis by identifying similar past cases.

Web searching

Cutting across many of the above application areas is the need for effective location of both text and images on the Web, which has developed over the last 10 years into an indispensable source of both information and entertainment. Text-based search engines have grown rapidly in usage as the Web has expanded; the well-publicized difficulty of locating images on the Web indicates that there is a clear need for image search tools of similar power. Paradoxically, there is also a need for software to *prevent* access to images which are deemed pornographic.

2.2 Image Retrieval Approach

There are many approaches that researcher had developed to come out with an accurate classification and retrieval result. There are several approaches for image classification and image retrieval:

2.2.1 Bayesian Belief Network

Bayesian belief network is graphical representation of a set of random variables and their dependencies. Bayesian network provide an efficient way to represent knowledge [8]. Parent nodes have direct influence on its child nodes. Interactions between nodes are represented by conditional probabilities. Bayesian network also contains

a) prior beliefs: the initials beliefs of nodes in the network
b) Evidences: Observation that are input of the network
c) Posterior probabilities: the final computed belief after the evidences have been propagated through the network.

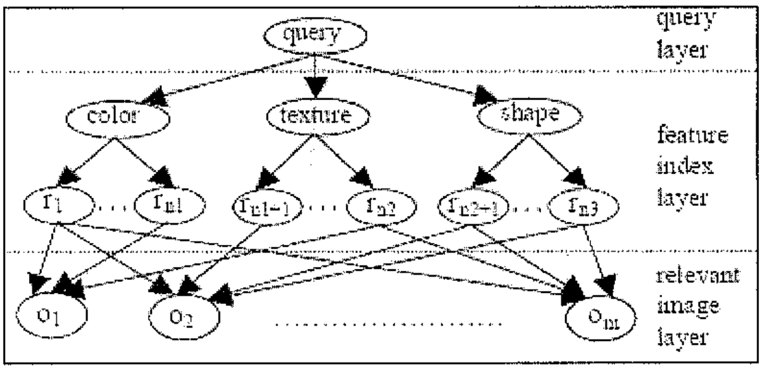

Figure 2.1: Sample of Bayesian Belief Network [8]

Jesse S.Jin and Jing Xin [8] used Bayesian network to optimize user feedback information utilization. Relevant feedback is used to bridge the gap between high level concept and low level features. Bayesian network have the capabilities learning user feedback and used the knowledge to improve retrieval performance and accuracy.

Chen and Edward [9] combine the use of Bayesian network with Inverse Tree Structure for joint classification and variable selection. This approach being used in image analysis that involves millions of variables and few image samples. This approach has advantages on handling high dimensional data and generating stable classifier even with small number of samples. Fang, Geman, Boujemaa [10] used Bayesian network in facial recognition system. The system use query/ answer process to gain knowledge about target (the relevant image). The system updates probability distribution on the target is updated after each respose according to bayes rule. Sueng and Sung [19] had proposed the use of Bayesian network with Scale-Invariant Feature Transform in scene understanding. According to the author, Bayesian network is robust to manage uncertainties and it is a powerful tool to model high-level context like relationship between places and objects.

2.2.2 Artificial Neural Network

An Artificial Neural Network consists of a number of very simple and highly interconnected processors, called neurons, almost similar to the biological neuron in brain. The neurons are connected by weighted links passing signals from one neuron to another. Each neuron received a number of signals through its connection and produce single output. Outgoing connection splits into number of branches that transmit the same signal to the next layer neural network. In supervised learning, where the output had been provided, the network trained by adjusting network weight in attempt to achieve correct result (output). In unsupervised learning, only input variable is introduced to the network. The network is self-organizing – the network organizes internally so that each hidden processing responds strategically to different set of input variable.

Solving a classification problem using neural network involve several steps:

1) Determine the number of input and output node, the number of hidden layer and number of node per hidden layer.
2) Determine the weight and the activation function
3) For each tuple in training set, propagate it through the network and evaluate the output prediction to the actual result. Adjust the weight.
4) For each tuple in real data set, propagate through the network and make appropriate classification.

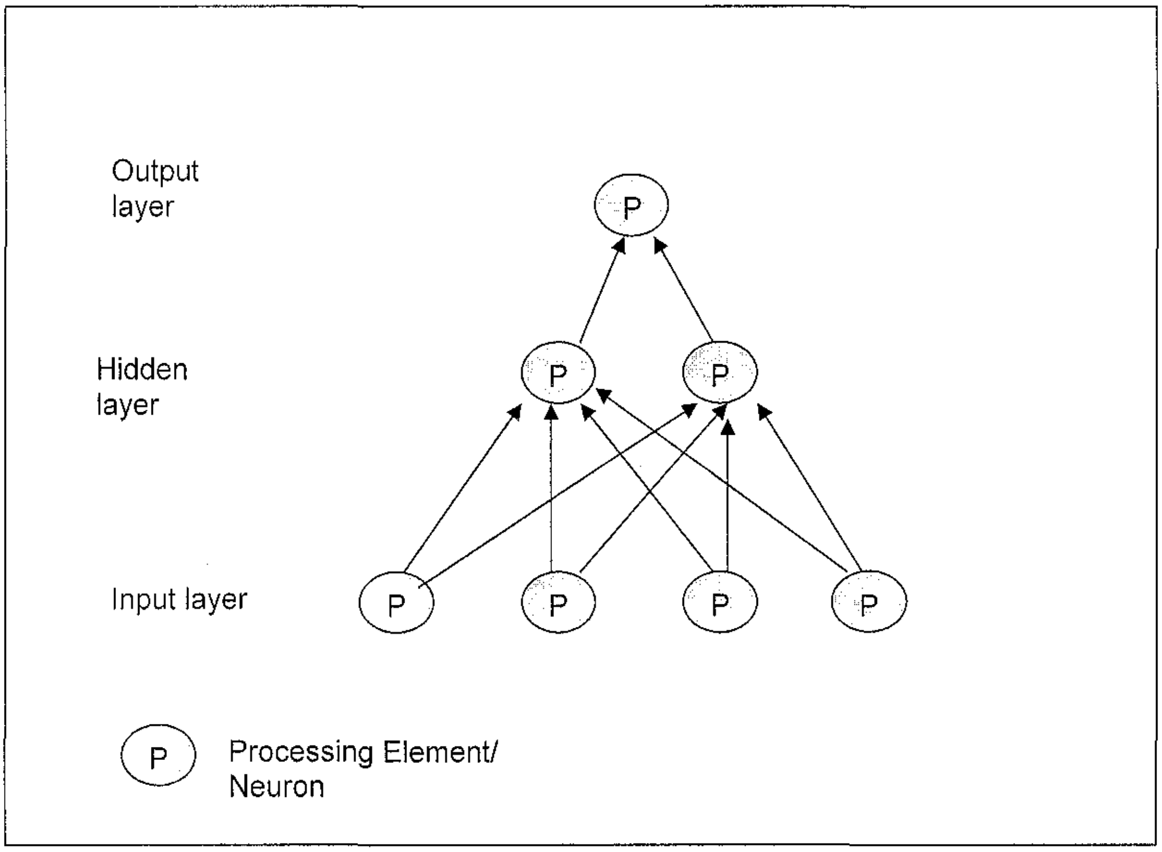

Figure 2.2: Neural Network with One Hidden Layer [12]

Advantages to use neural network for image classification

- Neural Network are more robust because of it capabilities to adjust weight.
- Neural Network has learning capabilities
- There is low error rate and high degree of accuracy

Disadvantages of neural network image classification

- Neural Network is hard to understand
- Require large sample of training data

Fong Tsai, McGarry and Tait [11] used hybrid neural network to bridge image semantic gap. They used hybrid neural network to automate keyword assignment for image classification. The approach composes of 3 components [11]: feature extractor, a content-based classifier and a semantic-based classifier. Two of the components used neural network. Content based classifier use Self-Organizing Maps (SOMs) to extract information form unlabelled samples and semantic-based classifier used Support Vector Machines (SVMs) to classify image into semantic categories.

In [16], CLAIRE had been used to assign keyword to image automatically. CLAIRE is composed of one image processing and three modules of support vector machine (SVM). This system uses two-stage mapping model (TSMM) [16]. By using this classification scheme direct mapping limitations can be solve and this scheme had advantage to compensate and correct error made by color and texture classifier for classifying high level concept in more accurate manner.

2.2.3 Visual Attention

Visual Attention is the ability to rapidly detect the interesting parts of given scene. This technology being expire by how human look at something. This model relies on the low level characteristics of image (color, intensity and orientations) to determine saliency and where attention should be directed, and in what order the attention should be moved around the image [18] .This approach allows rapid selection of subset of available sensory information before further process [13]. The advantage of this approach is it reduces computation cost of high level tasks like segmentation and object recognition which are complex.

One of the famous attention model architecture is Itti Model. In Itti model [21], extract color, intensity and orientation feature from image. These features are assembled onto multiscale representation using Gaussian and Laplacian pyramids. In each feature center-surround operators are applied to generate multiscale feature maps. An approximation to lateral inhabition is then employed to transform multiscale feature maps to conspicuity/noteworthy maps. The conspicuity maps are linearly combined to determine the saliency of the scene. Figure 2.3 will summarize these operations.

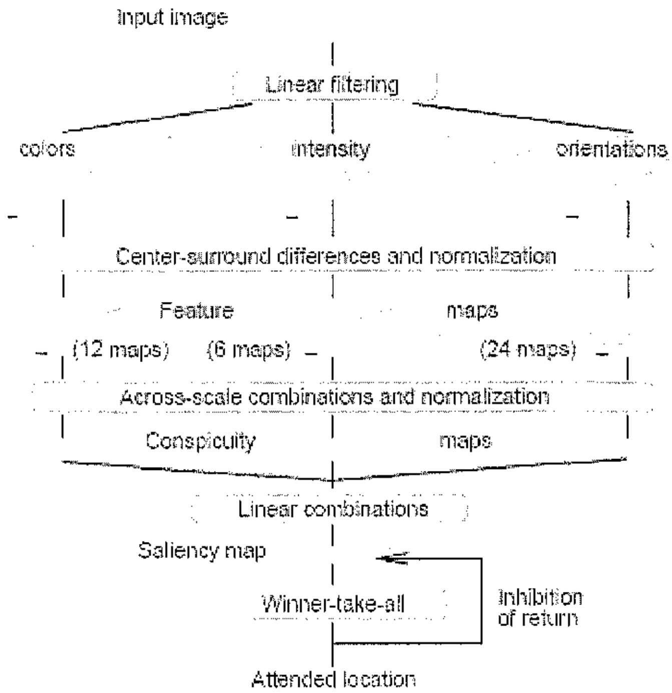

Figure 2.3: Itti Visual Attention Model[21]

In [14], visual attention being used for image adapting on small display. This research paper use visual attention mode to determine what object is important in images that need to be display is small display such as handphone, smartphone and pocket PC. This research paper had outline a very important issue that need to be consider when using visual attention model, that is image from different application do not catch attention in the same way. For example, human figure in sport picture attract most attention compare to human figure in scenery picture.

Nabil, Neculai, Heinz and Pierre [19] used combination visual attention model and seeded region growing (SRG) algorithm in seed selection process for performing color image segmentation. SRG is a segmentation technique which performs a segmentation of an image by referring to a set of point, called seeds. By using the seed SRG find homogeneous region. In the model propose visual attention being used to provide salient location. This point will be used as the starting point for SRG algorithm. Visual attention model and SRG algorithm speed up segmentation task.

Xiaodi Hou and Liqing Zhang [20], propose a model that describes how to quantify information capacity of attention. This paper try to answer the question of how many information can absorb given duration of time. This paper had show clear dependency of response time and spatial resolution attention. The more time given, one can obtain higher resolution (higher spatial frequency information, of presented visual stimuli). This model can be used to deduct/ predict the frequency of incoming information given observation time. This model can also be used to predict the shortest viewing time when given the resolution of stimulus.

Hua, Lu, and Zhang [22], used visual attention model in Photo2Video system. This system automatically converts photographic series into video. This system used contrast based visual attention model to detect focus of attention in image other than human faces. Robert and Fred [23], propose an approach that model eye's focusing mechanism. This approach used the saliency feature of an image to determine where to focus in an image. The advantage of this approach, the user does not need to assume specific feature to be presented in an image.

Chapter 3
Methodology

3.1 Introduction

This project was aspired by Distributed Content-based Visual Information Retrieval System on Peer-to-Peer (P2P) Network (DISCOVIR) project [32]. DISCOVIR is a software which enables user to search for and share multimedia files based on their content with anyone on the internet. DISCOVIR is built on top of Limewire [33] open source code. Most of the feature extraction libraries used in this CBIR system was taken from the DISCOVIR library. This project uses the libraries and applies it as a standalone application.

3.2 System and Software Design

During the design stage, the system architecture will need to be established to identify and describe the fundamental software abstractions and their relationships. The system architecture is shown in figure 3.1.

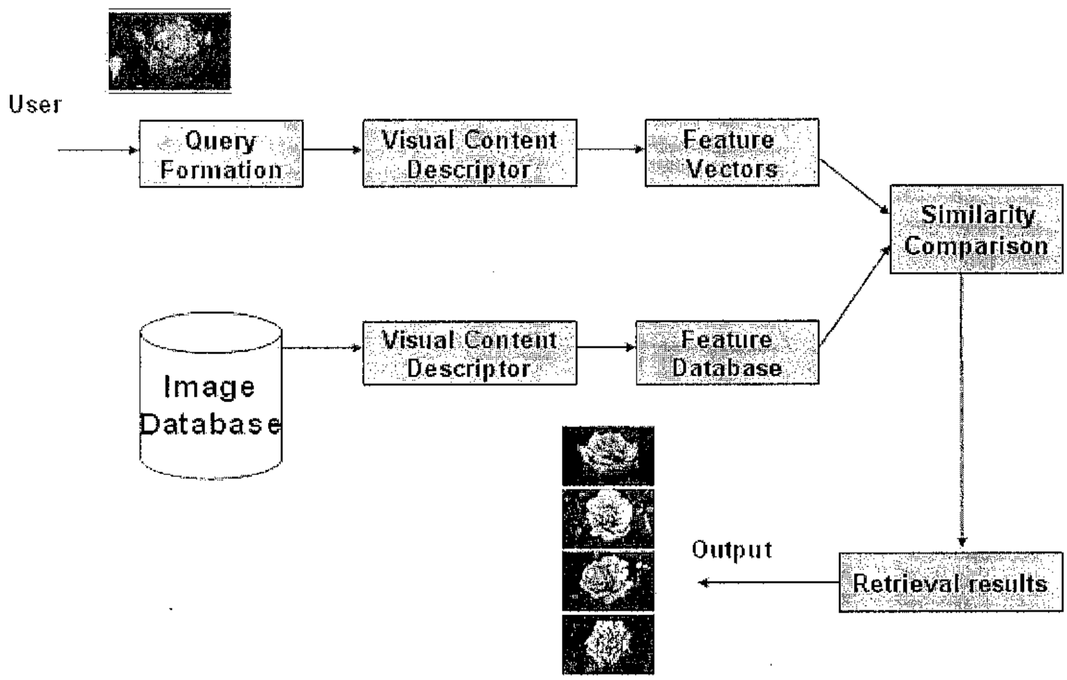

Figure 3.1: System architecture

3.2.1 Query Formulation

At Query Formulation stage, an interface was used to feed in the image that needs to be query in order to find similar images. User will select the image file and feed the system. The system will display the image that the user selects on the screen.

3.2.2 Visual Content Descriptor

Visual Content Descriptor will extract 7 different image feature vectors from an image. There are 3 main groups of feature vector that will be extracted from an image (color, shape and texture). Below are the main group and the type of feature vectors:

- Color
    - Color Coherent Vector
    - Global Color Histogram
    - Color Moment
- Texture
    - Cooccurence
    - Autocorelation
    - Edge Frequency
- Shape
    - Eccencity

3.2.3 Feature Vector

Feature Vector is an n-dimension vector of numerical feature that represent image. Feature vector represent the characteristic of images based on the type of feature extractor method.

3.2.4 Similarity Comparison

At Similarity Comparison stage, feature vector of the queried image will be compared with the feature vector of the images in the database. Query image will be compare one by one with images form the database.

3.2.5 Feature Database

Feature database describes the feature vector of images in the database.

3.2.6 Retrieve Result

In this stage, an interface will display the image that similar to the image queried on the screen.

3.3 Feature Vector Extraction

This system will extract 14 different feature vectors that will be used to train Bayesian network and these feature vectors will be extracted from the image that being queried by users. These features are divided to 3 main categories Color Extractor, Texture Extractor and Shape Extractor.

3.3.1 Color Extractor

Color Histogram

Color Histograms are frequently used to compare image. A color histogram H is a vector (h_1, $h_2 \ldots h_n$), in which each bucket hi contains the number of pixels of color j in the image. Images in this project are represented in RGB color space. Global/ Local Color Histogram H, is the compact summary of a given image, I. This system query images in database that similar to I, and return the image I', that have the most similar color histogram H_I. This system compares the color histogram by comparing the sum of squared differences (L_2-distance) or the sum of absolute value of differences (L_1-distance). The system determines the most similar image to I, based on the less L_2-distance or L_1-distance of image I'. Below are the mathematical formulas involved in determining the color histogram.

Notation:

- M - number of pixels that an image has, we assumed it for the ease of explanation
- $H(h_1, h_2, ..., h_n)$ - a vector, in which each component h_j is the number of pixels of color j in the image
- n - Number of distinct (discretized) color
- I - an image
- H_I - the color histogram of image I
- A, B, C – represent the three color channels.
- Prob – probability

Color Histogram is defined by,

$$h_{A,B,C}(a,b,c) = M.\text{Prob}(A = a, B=b, C=c)$$

Global Color Histogram is defined by,
$H(h_1, h_{2,...} h_n)$

In this system, we measure the distance between of 2 images for color histogram as follow:

$$H_I - H_{I'} = \sqrt{\frac{\sum_{j=1}^{n}(H_I[j] - H_{I'}[j])^2}{n}} \qquad (1)$$

L_2-distance

$$\|H_I - H_{I'}\| = \sum_{j=1}^{n}(H_I[j] - H_{I'}[j])^2 \qquad (2)$$

L_1-distance

$$|H_I - H_{I'}| = \sum_{j=1}^{n} |H_I[j] - H_{I'}[j]| \quad (3)$$

Color Histogram had advantages in it robustness to rotation and color histogram is insensitive to the changes of camera viewpoint. The main drawback of global color histogram is, it ignore an image texture and color [30].

Average RGB

Average RGB compute the average values if R (red), G (green) and B(blue) channel of each pixel in an image. RGB is also known as "additive primaries" because colors are added together to produce the desired color.

Notation

- I - an image
- w - width of image I
- h - height of image I
- I(x,y) - the pixel of image I at row y, column x
- R(p), G(p), B(p) - the red, green and blue color component of pixel p
- r_a, g_a, b_a - the average red, green and blue component of image I_a
- $d(I_a,I_b)$ - the distance measure between image I_a and I_b

The following are the equations for computing R, G, B component of an Image I:

$$r = \sum_{x=1,y=1}^{x=w,y=h} \frac{R(I(x,y))}{w \times h} \quad (4)$$

$$g = \sum_{x=1,y=1}^{x=w,y=h} \frac{G(I(x,y))}{w \times h} \quad (5)$$

$$b = \sum_{x=1, y=1}^{x=w, y=h} \frac{B(I(x,y))}{w \times h} \tag{6}$$

In this system, we measure the distance between 2 images as follow (by using weighted Euclidean distance):

$$d(I_a, I_b) = \sqrt{\frac{(r_a - r_b)^2 + (g_a - g_b)^2 + (b_a - b_b)^2}{3}} \tag{7}$$

Color Coherent Vector

Color Coherent is the degree to which pixels of that color are members of large similarly-colored regions. This feature measures the classified image as either coherent or incoherent. Coherent pixels refer to a part of some sizable contiguous region, while incoherent pixels are not. This system blurs the image by replacing the pixel values with the average value in a small local neighborhood. With this, we eliminate the small variation between neighboring pixels. Then we choose the proper colorspace to ensure there only n distinct colors in the images. After that the pixel is classified to coherent or incoherent. A pixel is coherent if the size of its connected component exceeds a fixed value T (2% of number of pixels) and the pixels that do not meet the criteria are considered incoherent.

Notation

- C (Connected Component) - a maximal set of pixels such that for any two pixels p, p' belongs to C, there is a path in C between p and p'. (Formally, a path in C is a sequence of pixels $p = p_1, p_2, \ldots pn = p'$ such that each pixel p_i is in C and any two sequential pixels p_i; p_{i+1} are adjacent to each other. Two pixels are to be adjacent if one pixel is among the eight closest neighbors of the other; in other words, diagonal neighbors is included.)
- I - an image
- n - number of distinct(discretized) color

- a_j - (number of coherent pixels of the j's discretized color) / (total number of pixels in the image)
- b_j - (number of incoherent pixels of the j's discretized color) / (total number of pixels in the image)

The equation for distance measure of image I_a and I_b (weighted Euclidean distance):

$$\sqrt{\frac{\sum_{j=1}^{n}((a_j - a'_j)^2 + (b_j - b'_j)^2)}{2n}} \qquad (8)$$

Color Moment

Color moment computes index containing only the dominant features instead of storing the complete color distributions. To compare between images this system used a weighted sum of the absolute differences between corresponding moments (average, variance and skewness).

Notation

- Pij - The value of the i-th color channel at the j-th image pixel
- N - Number of image pixels of an image
- r- Number of color channels
- w - User specified weights

These are the equations involve in this feature extraction:

Average Moment – This equation defines the average moment of a specific image at color channel i.

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} \qquad (9)$$

Variance - This equation defines the variance of a specific image at color channel i.

$$\sigma_i = (\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^2)^{\frac{1}{2}} \qquad (10)$$

Skewness - This equation defines the skewness of a specific image at color channel i.

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N}(p_{ij} - E_i)^3\right)^{\frac{1}{3}} \tag{11}$$

The equation for finding distance between 2 color distribution H and I

$$d_{mom}(H, I) = \sum_{i=1}^{r} w_{i1}|E_i - F_i| + w_{i2}|\sigma_i - \varsigma_i| + w_{i3}|s_i - t_i| \tag{12}$$

3.3.2 Texture Extractor

Autocorrelation

Autocorrelation measure the coarseness (rough or loose in texture) of an image by evaluating the linear relationship between texture primitives. Large primitives give rise to the coarse texture (e.g. rock surface) and small primitives give rise to fine texture (e.g. silk surface). Autocorrelation function will decreases slowly with the increasing distance if the primitives are large. If the texture consists of small primitives, the autocorrelation function will decrease rapidly.

Notation

- f(i,j) - the gray level value of the pixel in row i and column j
- M, N - image dimensions
- p, q- positional difference in i, j direction

This is the Autocorrelation function:

$$C_{ff}(p,q) = \frac{MN}{(M-p)(N-q)} \frac{\sum_{i=1}^{M-p}\sum_{j=1}^{N-q} f(i,j)f(i+p,j+q)}{\sum_{i=1}^{M}\sum_{j=1}^{N} f^2(i,j)} \tag{13}$$

- Usually, (p,q) varied from (0,0) to (8,8) in a step of two, which give a total of 25 features

Co-occurrence Matrices

Co-occurrence matrix is a statistical method using second order statistics to model the relationships between pixels within the region by constructing Spatial Gray Level Dependency (SGLD) matrices. The Gray-level co-occurrence matrix (GLCM) is the two dimensional matrix of joint probabilities Pd,r(i,j) between pairs of pixels, separated by a distance, d, in a given direction, r. There are 3 features used by this system to estimate the similarities between different Co-occurrence Matrices. There are energy, entropy and homogeneity.

Notation

- Pd,r(i,j) - joint probabilities between pairs of pixels in a given direction
- d - distance between pairs of pixels in a given direction
- r - a given direction

a. Energy

Energy is a measure of textural uniformity of an image [24]. Energy has its highest value when grey level distribution has either a constant or a periodic form. This feature can be defined as:

$$\sum_i \sum_j P_{d,r}^2(i,j) \qquad (14)$$

b. Entropy

Entropy measure the disorder of an image [24]. Entropy reaches its highest value when all elements in P matrix (the image matrix), are equal. Entropy is very large when the image is not textually uniform and GLCM elements have smaller value.

$$\sum_i \sum_j P_{d,r}(i,j) \log P_{d,r}(i,j) \qquad (15)$$

c. Contrast: (typically k = 2)

$$\sum_i \sum_j |i-j|^k P_{d,r}^{\lambda}(i,j) \qquad (16)$$

d. Homogeneity

Inverse different moment is used to measure the image homogeneity [24]. This parameter achieves it highest value when most of occurrence in GLCM are concentrated near the main diagonal.

$$\sum_i \sum_j \frac{P_{d,r}(i,j)}{|i-j|} \qquad (17)$$

Edge Frequency

Edge Frequency refer to the number of edges exist in an image. Below are the equation to find the Edge Frequency:

Notation

- B(a,r) - the number of primitives
- r - the number of primitives of all directions having length
- a - gray level
- M,N - image dimensions
- L - Numbers of gray levels
- Nr - the maximum primitive length in the images
- K - the total number of runs

Short primitive emphasis:

$$\frac{1}{K} \sum_{a=1}^{L} \sum_{r=1}^{N_r} \frac{B(a,r)}{r^2} \qquad (18)$$

Long primitive emphasis:

$$\frac{1}{K} \sum_{a=1}^{L} \sum_{r=1}^{N_r} B(a,r) r^2 \qquad (19)$$

Gray level uniformity:

$$\frac{1}{K}\sum_{a=1}^{L}\left[\sum_{r=1}^{N_r} B(a,r)r^2\right]^2 \qquad (20)$$

Primitive length uniformity:

$$\frac{1}{K}\sum_{a=1}^{L}\left[\sum_{r=1}^{N_r} B(a,r)\right]^2 \qquad (21)$$

Primitive percentage:

$$\frac{K}{\sum_{a=1}^{L}\sum_{r=1}^{N_r} rB(a,r)}=\frac{K}{MN} \qquad (22)$$

Primitive Length

A primitive is a continuous set of maximum number of pixels in the same direction that have the same gray level. Each primitive is defined by its gray level, length and direction. Primitive length use length of texture primitives in different directions as texture description.

Notation

- B(a,r) - number of primitives of all directions having length r and gray level a
- M, N - image dimensions
- L - number of image gray levels
- N_r - max length of primitive
- K - total number of primitives

$$K = \sum_{a=1}^{L} \sum_{r=1}^{N_r} B(a, r) \tag{23}$$

Short primitive emphasis - define the amounts of short texture primitives

$$\frac{1}{K} \sum_{a=1}^{L} \sum_{r=1}^{N_r} \frac{B(a, r)}{r^2} \tag{24}$$

Long primitive emphasis - define the amounts of long texture primitives

$$\frac{1}{K} \sum_{a=1}^{L} \sum_{r=1}^{N_r} B(a, r) r^2 \tag{25}$$

Gray-level uniformity - define the distribution of gray-level of textures

$$\frac{1}{K} \sum_{a=1}^{L} (\sum_{r=1}^{N_r} B(a, r))^2 \tag{26}$$

Primitive length uniformity - define the distribution of primitive length

$$\frac{1}{K} \sum_{r=1}^{N_r} (\sum_{a=1}^{L} B(a, r))^2 \tag{27}$$

Primitive percentage - define the concentration of texture primitives

$$\frac{K}{\sum_{a=1}^{L} \sum_{r=1}^{N_r} r \; B(a, r)} = \frac{K}{MN} \tag{28}$$

3.3.3 Shape Extractor

Zernike Moment

Zernike moments are used as shape descriptor for complex shapes such as trademarks that are difficult to be defined with single contour [27]. Zernike moment is invariant to rotation.

Notation

- n – the order of Zernike moment
- m – constrained by n
- *(k1, k2)* – Cartesian coordinate position
- (ρ, θ) - polar coordinate

Basic Zernike function is defined by:

$$V_{nm}(k_1,k_2) = V_{nm}(\rho,\theta) = R_{nm}(\rho)\exp(jm\theta),$$
$$n \geq 0, n-|m|: even, \ and \ |m| \leq n, \tag{29}$$

R_{nm} (p) is defined by:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!(\frac{n+|m|}{2}-s)!(\frac{n-|m|}{2}+s)!} \rho^{n-2s} \tag{30}$$

Euclidean Distance is used to distance measure of two shape descriptor.

Invariant Moment

This shape descriptor based on the theory of moments. The moment invariants are moment-based descriptors of planar shapes, which are invariant under general translational, rotational, scaling, and reflection transformation [28].

Notation:

p and q – order of the moment

Equation for Invariant Moment given f (i, j)

$$m_{pq} = \sum_i \sum_j i^p j^q \tag{31}$$

Pseudo Zernike Moment

Pseudo Zernike has an important attribute when it comes to the shape description: their magnitudes are invariant under rotation [29]. This means, it robust to rotation. This feature is very important when we try to retrieve image.

Notation:

- n – nonnegative integer
- m – an integer that takes positive, negative or zero value such that *n-m* ≥ *0.*

Basic Pseudo Zernike Moment Equation (1)

$$V_{nm}(x, y) = R_{nm}(\rho)e^{jm\theta},$$

Where $\rho = \sqrt{x^2 + y^2}$, $\theta = \tan^{-1}(y/x)$ (32)

The radial Pseudo Zernike polynomial R_{nm}(p) can be define as:

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|} \frac{(-1)^s(2n+1-s)!\rho^{n-s}}{s!(n+|m|+1-s)!(n-|m|-s)!}. \tag{33}$$

Pseudo Zernike Moment Equation for Digital Image

$$\hat{A}_{nm} = \frac{n+1}{\pi} \sum_{x_i} \sum_{y_j} h_{A_{nm}}(x_i, y_j) f(x_i, y_j),$$

where $x_i^2 + y_j^2 \leq 1$, and

$$h_{A_{nm}}(x_i, y_j) = \int_{x_i - \frac{\Delta x}{2}}^{x_i + \frac{\Delta x}{2}} \int_{y_j - \frac{\Delta y}{2}}^{y_j + \frac{\Delta y}{2}} V_{nm}^*(x, y)\,dx\,dy, \tag{34}$$

Eccentricity

Eccentricity refers to ratio of the short axis' length to the long axis' length of the best fitting ellipse of the shape [25].

Notation:

-

$$Eccentricity = \frac{I_{min}}{I_{max}} = \frac{u_{20} + u_{02} - \sqrt{(u_{20} - u_{02})^2 + 4u_{11}^2}}{u_{20} + u_{02} + \sqrt{(u_{20} - u_{02})^2 + 4u_{11}^2}}$$

where $u_{p,q} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y)$, (35)

is the (p, q)order central moments of the shape.

Edge Direction Histogram

Edge Direction Histogram is used to compare the edge information between the distorted image block and the reference one [26].

Notation:

- i – no of bin to represent edge direction
- n_e – total number of edge point

Edge Direction Histogram equation defines as following:

$$H(i) = \frac{H(i)}{n_e} \; ; \tag{36}$$

3.4 Distance Measure

Generally, Distance Measure is used to calculate the distance between each pixel in images. This method is very important to compare similarity between images. In this project, the system used L2 distance (Euclidean distance) to determine distance measure. This system will consider two images is similar if the distance measure between 2 images is equal or less than the threshold value that had been be pre-define by the system.

3.5 Development Tools

The system will be developed using JAVA programming language. Database used Apache Derby Database.

Chapter 4

Result and Discussion

4.1 Dataset Description

Dataset to test this system, Corel Image Dataset had been used [31]. 250 images from the dataset had been selected randomly. There are 4 main image groups (dinosaur, scene, elephant, horse and flower).

4.2 Experiment set up to decide threshold value

To find the optimum threshold value that fits all group of images, 7 image features will be used. These feature vectors are divided into 3 main groups. There are:

- **Color Extractor** – Color Coherent Vector, Global Color Histogram and Color Moment
- **Texture Extractor** – Cooccurence, Autocorrelation, Edge Frequency
- **Shape Extractor** – Eccencity

Below are the steps taken to determine the optimum threshold value for each feature vector.

1. The system will store 250 different random images from different groups will be store in database.
2. Set the threshold value
3. Feed the system will random images. Choose random images in each group.
4. Observe and note the number of relevant image being display and the number of image being display by system.
5. Repeat the steps 1-4 with different threshold value (Please refers to table 4.1 for the threshold values).
6. After finished testing with different threshold value, repeat steps 1-5 to test different feature vectors. (Please refer to table 4.1).

Feature Vector Group	Feature Vector	Threshold Values		
		Tested 1	**Tested 2**	**Tested 3**
Color	Color Coherent Vector	0.02	0.04	0.06
	Global Color Histogram	0.025	0.05	0.075
	Color Moment	0.5	1.0	1.5
Texture	Cooccurence	2.5	5	7.5
	Autocorelation	0.02	0.04	0.06
	Edge Frequency	10	20	30
Shape	Eccentricity	0.02	0.04	0.06

Table 4.1: Threshold value

The accuracy of each feature in retrieving image is based on the following formula:

Accuracy of feature (%) =

(Total number of relevant image/ Total number of display by the system) x 100

4.3 Result and Discussion

This section show the accuracy of image retrieve based on the threshold value

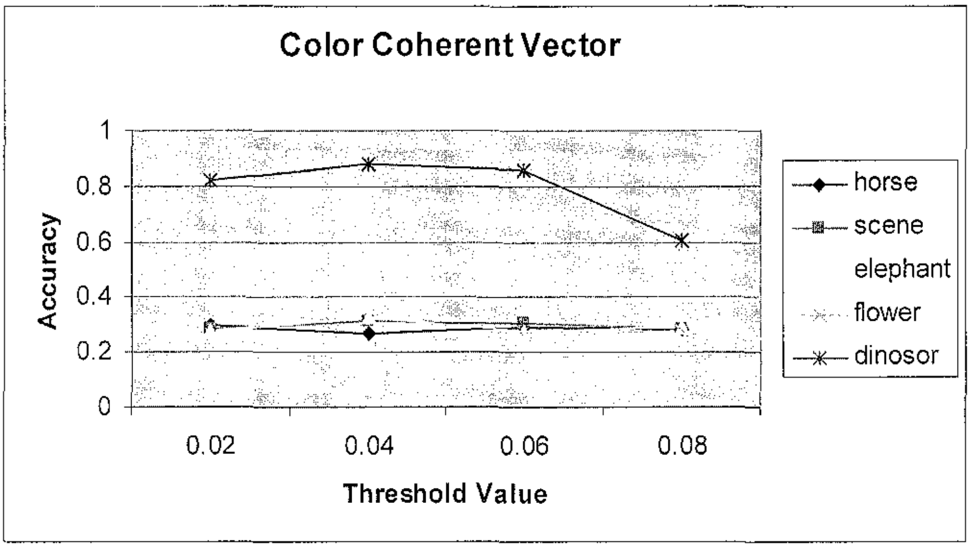

Figure: 4.1: Color Coherent Vector Accuracy Graph

Color Coherent Vector					
	horse	scene	elephant	flower	dinosaur
0.02	0.301703	0.282197	0.276243	0.232227	0.82417582
0.04	0.269565	0.311715	0.326797	0.228175	0.88235294
0.06	0.292105	0.307216	0.276243	0.260309	0.85714286

Table 4.2: Color Coherent Vector Threshold

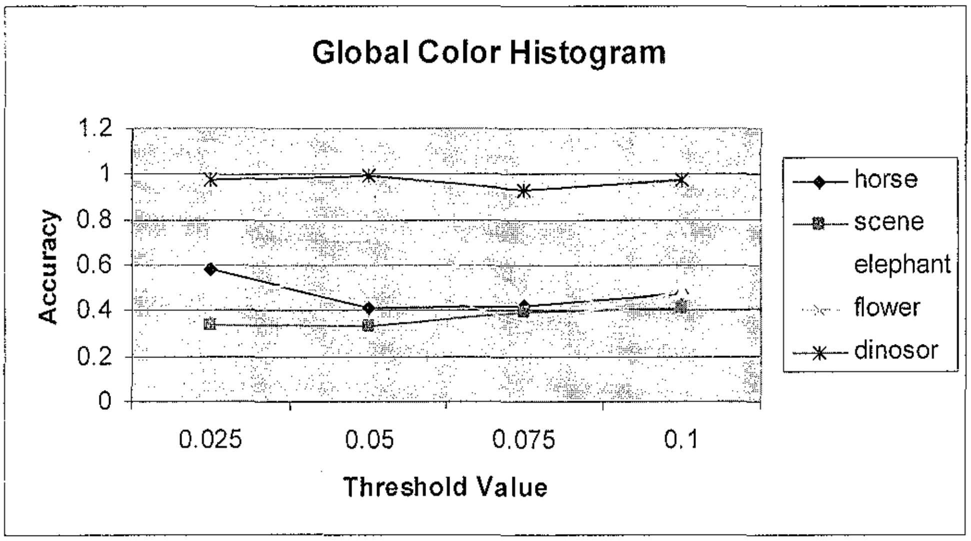

Figure: 4.2: Global Color Histogram Accuracy Graph

Global Color Histogram					
	horse	scene	elephant	flower	dinosaur
0.025	0.580882	0.341216	0.455	0.722222	0.98039216
0.05	0.415929	0.334448	0.571429	0.700599	0.99337748
0.075	0.419355	0.392727	0.338095	0.515419	0.93037975
0.1	0.476744	0.410714	0.486275	0.338542	0.98013245

Table 4.3: Global Color Histogram Threshold

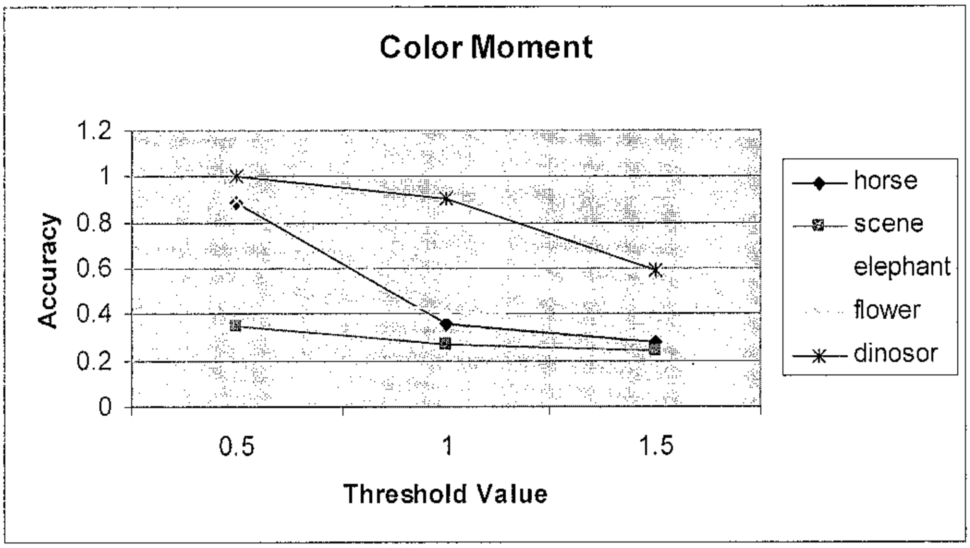

Figure: 4.3: Color Moment Accuracy Graph

Color Moment					
	horse	scene	elephant	flower	dinosaur
0.5	0.885714	0.347826	0.54	0.896552	1
1	0.355049	0.271084	0.410569	0.433198	0.90361446
1.5	0.276471	0.240631	0.338942	0.326039	0.59504132

Table 4.4: Color Moment Threshold

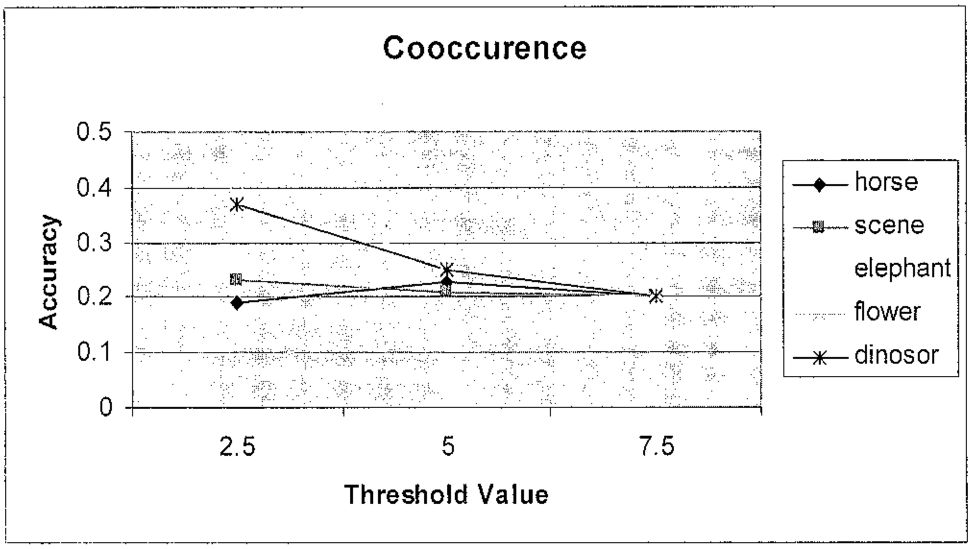

Figure: 4.4: Cooccurence Accuracy Graph

Coocurence					
	horse	scene	elephant	flower	dinosaur
2.5	0.189003	0.232323	0.069444	0.389262	0.36979167
5	0.227642	0.208738	0.185759	0.260054	0.24875622
7.5	0.200803	0.201613	0.206612	0.209205	0.2033195

Table 4.5: Coocurence Threshold

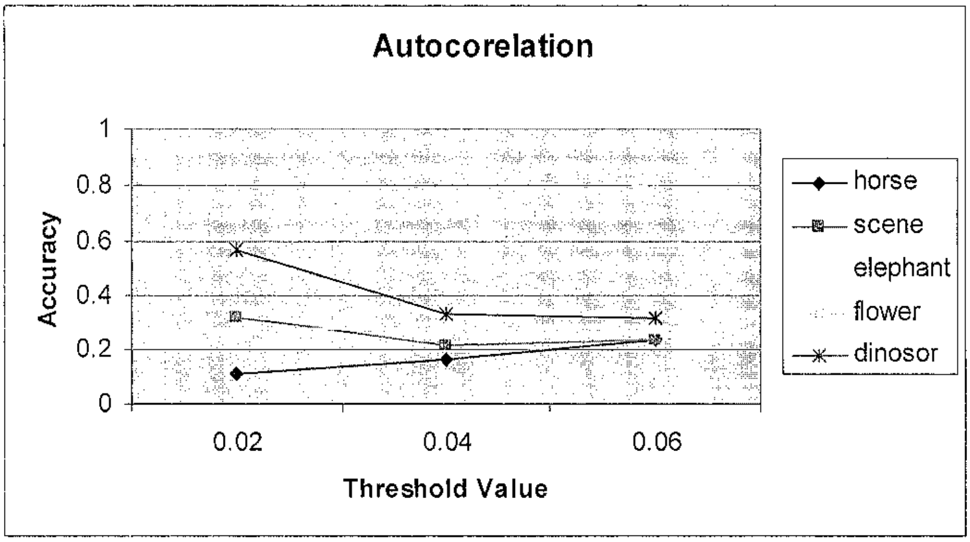

Figure: 4.5: Autocorrelation Accuracy Graph

Autocorrelation					
	horse	scene	elephant	flower	dinosaur
0.02	0.111111	0.323077	0.273684	0.875	0.56382979
0.04	0.162281	0.215789	0.278351	0.882353	0.33497537
0.06	0.238636	0.239286	0.280277	0.793103	0.32258065

Table 4.6: Autocorrelation Threshold

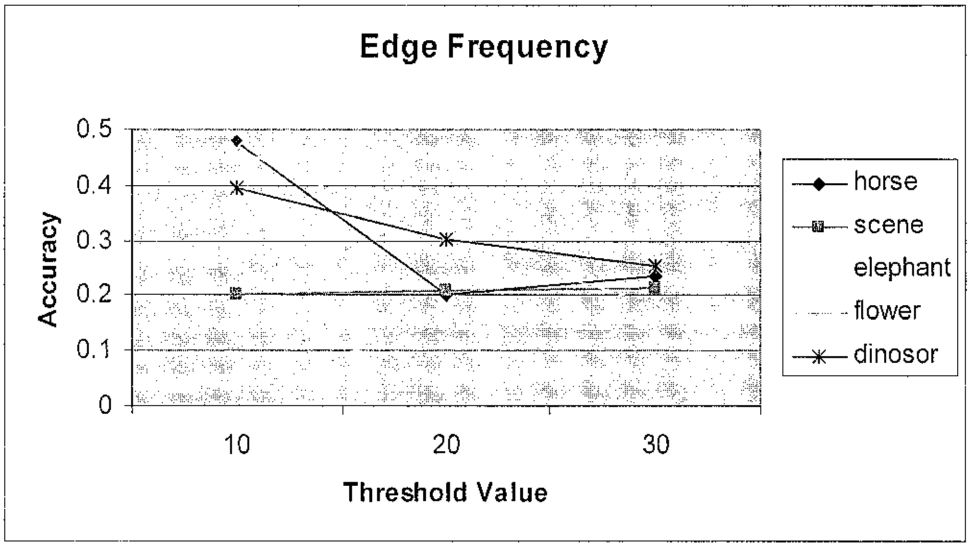

Figure: 4.6: Edge Frequency Accuracy Graph

Edge frequency					
	horse	scene	elephant	flower	dinosaur
10	0.47619	0.2	0.030303	0.468085	0.3943662
20	0.200873	0.209302	0.154229	0.297753	0.30215827
30	0.233516	0.213483	0.136111	0.182353	0.25297619

Table 4.7: Edge Frequency Threshold

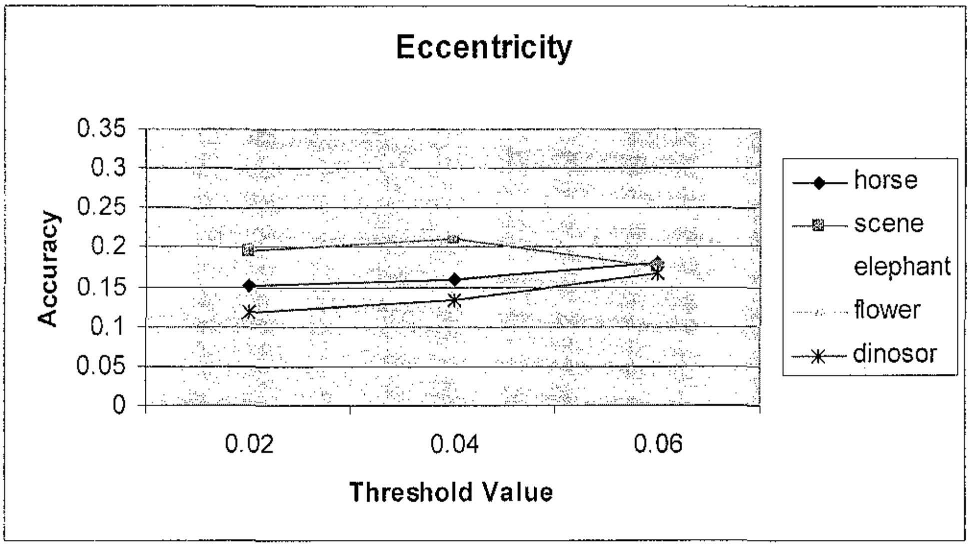

Figure: 4.7: Eccentricity Accuracy Graph

Eccentricity					
	horse	scene	elephant	flower	dinosaur
0.02	0.151724	0.194805	0.232877	0.288288	0.11827957
0.04	0.15894	0.211111	0.224242	0.305019	0.13385827
0.06	0.179739	0.174603	0.220238	0.266667	0.16793893

Table 4.8: Eccentricity Threshold

Based on the experiment above, the optimum values for each feature vector are shown in table 4.9.

Feature Vector	**Threshold Value**
Color Coherent Vector	0.05
Global Color Histogram	0.05
Color Moment	1
Cooccurence	5
Autocorelation	0.04
Edge Frequency	20
Eccentricity	0.04

Table 4.9: Feature Vector and Optimize Threshold Value

Based on the optimize threshold value, below are the accuracy of the retrieved result.

Picture Group	Average Accuracy (%)
Horse	27.6596
Elephant	59.375
Flower	73.0769
Dinosaur	95.4545
Scene	32.3077

Table 4.10: Accuracy Table based on Image Groups

Based on the result, the dinosaur image has the highest accuracy and horse image the lowest accuracy. This is because dinosaur images white back ground compare to other images. This distinctive attribute increase the accuracy when the system try to retrieve accuracy. Horse images have the lowest accuracy because of the characteristic of the images itself. Majority of horse image contain green background (grass and tree) which have the similar attribute with most of flower and scene attribute.

Chapter 5

Conclusion and Recommendation

5.1 Conclusion

Content Based Image Retrieval has become one of the most important technologies in image processing. Textual or meta-data image retrieval have it limitation because it is labor intensive and time consuming. A similar image can be assigned different textual annotation. This can lead to inconsistency in standard of naming images.

Content Based Image Retrieval has it advantages when retrieval an image because it does not depend on the textual but it depends on the content of the image.
This project utilizes the low level feature of 7 different feature vectors that can be used to improve the accuracy of image search.

5.2 Recommendation

Since this project serves to demonstrate the basic function of image retrieval based on the feature of the image, there are more functionality that can be added into this system.

Firstly, there can be improvement in the speed of querying images. Current system consume a lot time when the system try to retrieve large number of images. This system takes average 20 second to display result. This is because the system extract feature vector and load it into memory. After the system finish, the system will load the next feature vector into the memory and process it. This process consume large amount of computer memory and CPU process. Therefore, there should be new design of the application to speed up the application. The design should directly display the result when the image query is similar to the image database. Current system displays the result in a group. This mean the system will display image result when the system have search all images in the database.

Secondly, the proposed algorithm is still in development stage. At this stage, the system used whole image element when comparing with other image. There are some areas that can be improved to increase the accuracy the image being retrieved. For example by using visual attention approach where the system will focus on the image that is capture the attention of a people. This approach is more relevant compare to the approach that read whole image.

REFERENCE

1 Oliver Chapelle, Patrik Haffner, and Vladimir N. Vapnik, "Support Vector Machines for Histogram-Based Image Classification," IEEE Transactions on Neural Networks, Vol. 10, Issue 5, September 1999.

2 Vincent S. Tseng, and M. S. Wang, "A New Method for Image Classification by Using Multilevel Association Rules", IEEE International Workshop on Managing Data for Emerging Multimedia Applications, 2005.

3 Alexander Kaplan, and Ehud Rivlin , "Robust Feature matching across widely separate Color Images, "Proceedings of the 7th International Conference on Pattern Recognition, Vol. 2, pp. 136-139, August 2004.

4 Vincent. S. Tseng, Chon-Jei Lee, and Ja-Hwung Su, "Classify By Representative or Associations (CBROA): A Hybrid Approach for Image Classification", MDM/KDD 2005 Chicago, pp. 61-69 Chicago, Illois, USA, August 21 2005.

5 Eakins, J P (1996) "Automatic image content retrieval – are we getting anywhere?" *Proceedings of Third International Conference on Electronic Library and Visual Information Research (ELVIRA3)*, De Montfort University, Milton Keynes, pp 123-135

6 Dr. Fuhui Long, Dr Hongjiang Zhang and Prof. David Dagan Feng, *Fundamentals of Content-based Image Retrieval,* <http://research.microsoft.com/asia/dload_files/group/mcomputing/2003P/ch01_Long_v40-proff.pdf- >.

7 Gudivada V N and Raghavan V V (1995a) "Content-based image retrieval systems" *IEEE Computer* 28(9), 18-22

8 Jing Xin and Jesse S. Jin, "Relevance Feedback for Content-Based Image Retrieval Using Bayesian Network", Proceedings of the Pan-Sydney area workshop on Visual information processing, pp. 91 - 94, 2004.

9 Rong Chen and Edward H. Herskovits,, "A Bayesian Network Classifier with Inverse Tree Structure for Voxelwise Magnetic Resonance Image Analysis", Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining KDD '05, August 2005.

10 Yuchun Fang, Donald Geman and Nozha Boujemaa, "An interactive system for mental face retrieval", Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, pp. 193 - 200, 2005.

11 Chih-Fong Tsai, Ken McGarry and John Tait, "Image classification using hybrid neural networks", Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 431 - 432, 2003.

12 Efraim Turban, Jay E. Aronson, Ting-Peng Liang. *Decision Support Systems and Intelligent System* (7th ed.). Prentice Hall.

13 Ouerhani, N and Hugli, H, "Computing visual attention from scene depth", Pattern Recognition, 2000. Proceedings. 15th International Conference on Pattern Recognition - Volume 1, pp. 375 - 378, 2000.

14 Li-Qun Chen, Xing Xie, Xin Fan,Wei-Ying Ma, Hong-Jiang Zhang and He-Qin Zhou, "A visual attention model for adapting image on small display", http://research.microsoft.com/~xingx/mmsj1.pdf

15 Simone Frintrop and Erich Rome, "Simulating Visual Attention for Object Recognition", www.nada.kth.se/~frintrop/paper/frintrop_rome_ecovis04.pdf

16 Chin Fong Sai, Ken Mcgarry and John Tait, "CLAIRE: A modular Support Vector Image Indexing and Classification System", ACM Transactions on Information Systems, Vol. 24, No. 3, Pages 353–379, July 2006.

17 Seung-Bin Im and Sung-Bae Cho ,"Context-Based Scene Recognition Using Bayesian Networks with Scale-Invariant Feature Transform", J. Blanc-Talon et al. (Eds.): ACIVS 2006, LNCS 4179, pp. 1080 – 1087, 2006.

18 Geoffrey Underwood, Tom Foulsham,Editha van Loon, and Jean Underwood, "Visual Attention, Visual Saliency, and Eye Movements During the Inspection of Natural Scene", J. Mira and J.R. 'Alvarez (Eds.): IWINAC 2005, LNCS 3562, pp. 459–468, 2005.

19 Nabil Ouehani, Neculai Archip, Heinz Hugli, and Pierre-Jean Erard "Visual Attention Guided Seed Selection for Color Image Segmentation", W. Skarbek (Ed.): CAIP 2001, LNCS 2124, pp. 630{637, 2001.

20 Xiaodi Hou and Liqing Zhang , "A Time-Dependent Model of Information Capacity of Visual Attention", I. King et al. (Eds.): ICONIP 2006, Part I, LNCS 4232, pp. 127–136, 2006.

21 Laurent Itti, Christof Koch and Ernst Nieur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis", (find the resource).

22 Xian-Sheng Hua, Lie Lu, and Hong Jiang Zhang, "Automatically Converting Photographic Series into Video", Volume 16, Issue 7, July 2006 Page(s):803 - 819 Digital Object Identifier.

23 R. Shilston and F. W. M. Stentiford, "An attention-based focus control system," Int. Conf. on Image Processing, Oct. 8-11, Atlanta, 2006

24 M.Partio, B.Cramariuc, M.Gabbouj and AriVisa, "Rock Texture Using Gray Level Co-Occurrence Matrix", www.muvis.cs.tut.fi/Documents/Norsig2002final.pdf

25 Liu Wenyin, Tao Wang and HongJiang Zhang, "A Hierarchical Characterization Scheme for Image Retrieval" , research.microsoft.com/china/papers/Hierarchical_Characterization_Scheme_Image_Retrieval.pdf

26 Guan-Hao Chen, Chun-Ling Yang, Lai-Man Po and Sheng-Li Xie, "Edge-Based Structural Similarity for Image Quality Assesment", www.ee.cityu.edu.hk/~lmpo/publications/2006_ICASSP_EBSSIM.pdf

27 Hae-Kwang Kim, Jong-Deuk Kim, Dong-Gyu Sim, Dae-Il Oh, "A modified Zernike moment shape descriptor invariant to translation, rotation and scale for similarity-based image retrieval", Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on Volume 1, 30 July-2 Aug. 2000 Page(s):307 - 310 vol.1.

28 Xiaolong Dai and Siamak Khorram, "A Feature-Based Image Registration Algorithm Using Improved Chain-Code Representation Combined with Invariant Moments", IEEE Transactions on Geoscience and Remote Sensing, Vol. 37 No. 5, September 1999.

29 Yongqing Xin, Liao, S., Pawlak, M., "Geometrically robust image watermarking via pseudo-Zernike moments", Electrical and Computer Engineering, 2004. Canadian Conference on Volume 2, 2-5 May 2004 Page(s):939 - 942 Vol.2.

30 G. Pass, R. Zabih, "Histogram Refinement for Content-Based Image Retrieval", Applications of Computer Vision, 1996. WACV '96., Proceedings 3rd IEEE Workshop on 2-4 Dec. 1996 Page(s):96 – 102.

31 The Corel Corporation World Wide Web home page. http://www.corel.com

32 http://www.cse.cuhk.edu.hk/~miplab/discovir/

33 http://www.limewire.com/