A CONTENT-BASED IMAGE RETRIEVAL SYSTEM FOR HAT DATABASE

By
Danhua Li
Bachelor of Science

A project report
presented to Ryerson University
in partial fulfillment of the
requirements for the degree of
Master of Engineering
in the program of
Electrical and Computer Engineering

Toronto, Ontario, Canada

September, 2006

Author's Declaration

I hereby declare that I am the sole author of this thesis. I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

(Danhua Li)

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

(Danhua Li)

Borrower's Page

Ryerson University requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Name	Signature	Address	Date

ABSTRACT

A CONTENT-BASED IMAGE RETRIEVAL SYSTEM FOR HAT DATABASE

Master of Engineering
Department of Electrical and Computer Engineering
Ryerson University

Content-based image retrieval (CBIR) is a technique for indexing and retrieving images based on the low-level features, middle-level features, and high-level features. Low-level feature is extracted from contents of the images such as color, texture and shape; middle-level feature is a region obtained as a result of image segmentation; high-level feature is sematic information about the meaning of image, its objects and their roles, and categories to which the image belongs. In this project, three low-level features texture-based retrieval, color-based retrieval and shape-based retrieval are implemented and compared on hat database. Texture features are obtained from parameters of a two-component Gaussian mixture model (GMM) in the wavelet domain. Color features are extracted from a two-component GMM on HLS color space. Shape features are extracted from the contour by using centroid-contour distance Fourier descriptor. A comprehensive experimental evaluation of the retrieval performance of different feature sets is performed. The experimental results indicate that the shape features based on the centroid-contour distance Fourier descriptor perform much better than the color and texture features for the hat database used in this project.

Acknowledgment

I would like to thank many people who contributed in one way or another to this thesis. I thank my supervisor, Dr. Xiaoping Zhang, who inspired me to enter the exciting area of image processing, and for his guidance and support.

I am very grateful to the students in my lab for their helpful discussion and encouragement.

I would also like to thank the members of my thesis committee, Professor Alagan S. Anpalagan and Professor Javad Alirezaie for their time and interest.

Contents

List of Figures

List of Tables

List of Acronyms

CBIR	Content Based Image Retrieval
CCD	Centroid Contour Distance
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
EM	Expectation Maximization
FCH	Fuzzy Color Histogram
FD	Fourier Descriptor
GFD	Generic Fourier Descriptor
GMM	Gaussian Mixture Model
HSL	Hue Saturation Luminance
HSV	Hue Saturation Value
IBM	International Business Machines
MARS	Multimedia Analysis and Retrieval System
PCA	Principal Component Analysis
PDF	Probability Density Function
QBIC	Query By Image Content
RF	Relevance Feedback
RGB	Red Green Blue
VQ	Vector Quantization
WWW	World Wide Web

Chapter 1

Introduction

Image retrieval systems have been explored with the rapid growth of imaging on the internet in the past decade. The amount of information in the form of images is increasing rapidly with the advent of powerful and inexpensive computers, storage devices, availability of the world wide web (WWW), digital cameras and mobile telephones equipped with imaging technology.

Image retrieval is experiencing a transformation from text-based indexing to content-based indexing. An image storage and retrieval system typically supports Text-based annotations and indexing. These indexes for large image collections are time consuming to create and maintain manually. Also, text indexing for images only provides hit-or-miss type searching. If the user does not specify the right keywords, the desired images may be forever unreachable.

1.0.1 Text-based Image Retrieval System

Traditional image retrieval systems are Text-based. In Text-based systems, keywords are created for each and every image in the database. The images are then searched using these keyword indexes similar to the retrieval of text documents. Although this technology has matured over the years and produces reasonable results in some cases yet there are some drawbacks of Text-based image retrieval systems. Some of the drawbacks of Text-based systems are enumerated below.

1) Accurate text description of an image requires well-trained personnel.

2) Users are restricted to use proper keywords to search the images. This is very crucial because different persons have different understanding of the same image. The choice of keywords to describe the same image might be different for different users and may vary from formal words to slangs. This can also vary when users come from different areas and backgrounds.

3) A lot of tedious work is required to setup and maintain retrieval system because of manual creation of keyword indexes.

4) The performance of Text-based retrieval system is limited by the accuracy of text, understanding of the image, efficiency of the system setup, and maintenance of the database index. As the WWW is expanding at a tremendous speed, it is hard to imagine dealing with the huge collection of images only in a manual way of creating and storing keyword indexes. Content-based indexing and retrieval offers an efficient and more effective alternative to the traditional Text-based image retrieval systems.

1.0.2 Content-based Image Retrieval System

Content-based image retrieval systems use automatic feature extraction from physical characteristics of the images such as texture, color, and shape information. With the powerful personal computers, it is easy to acquire the image features in a reasonable time, and maintain the system at an acceptable cost. Content-based image retrieval system avoids the different understandings of images and dramatically improves accuracy and efficiency. There are two major steps in content-based image retrieval systems:

1) Feature extraction

2) Similarity measurement and search

Feature extraction is to extract features from every image based on low-level characteristics, such as texture, color, and shape, to get perceptually similar query. Similarity measurement and search is to compare the features of the images in the database with that of the query image to find the similar images. In this project color, texture and shape features are extracted from the images. These features are used to index and retrieve hat database. The accuracy of the system is determined in the form of precision-recall plots for the different combination of features.

For color feature extraction, images are first transformed from RGB color space to HSL color space and from spatial domain to wavelet domain. To obtain texture features, a two-component Gaussian mixture model (GMM) is employed and the parameters of the model are used as features. Shape features are extracted by using centroid-contour distance Fourier descriptor.

Similarity measurement in CBIR systems is essentially based on distance measurement between the feature vectors that have been previously extracted from images and that of the query image.

1.1 Related Work

1.1.1 Texture Based Image Retrieval

In one of the earliest work on texture image retrieval, Haralick et al. introduced the idea of the cooccurrence matrix for the representation of texture [1]. Their technique is based on the gray level spatial dependence of texture. The co-occurrence matrix is first constructed based on the orientation and distances between the images and then texture features are obtained from the statistics of this matrix. The work of Tamura et al. was motivated by the psychological studies in human visual perception of texture [2]. They proposed computational approximations to the six visual texture properties namely coarseness, contrast, directionality, linelikeness, regularity, and roughness. Their features for the texture representation are more meaningful because of their connection to the human visual perception.

In early 90's, many researchers used wavelet transform in texture based image retrieval. In [3], Smith et al. used mean and variance of wavelet coefficients as texture features. Flickner et. al. from IBM (International Business Machines) Research developed the QBIC (query by image content) system to explore content-based retrieval methods [4]. QBIC supports different types of queries such as by example images, user-constructed sketches and drawings, selected color and texture patterns, camera and object motion, and other graphical information. The system uses color, texture, shape and motion features. Principal component analysis (PCA) was used to reduce the dimensionality of the high dimensional features such as 20-dimensional moment-based shape feature. Y. Rui, T. S. Huang, S. Mehrotra proposed two sets of features for texture representation in MARS (multimedia analysis and retrieval system)[5]. For the first set of features, they applied 3-level wavelet decomposition to get 10 sub-bands. The standard deviation of the wavelet coefficients in each subband is then used as texture feature. For the second set of texture features, they computed 4 co-occurrence matrices corresponding to 0, 45, 90 and 135 degree angles. They then calculated two statistical features from each of the matrix and called them contrast and inverse difference moment.

Bin Zhang et. al. developed an adaptive system for texture retrieval which automatically selects the best features for a particular query [6]. They used 4 sets of transform based features for their system. The features are calculated by transforming the images using DCT, wavelet, Gabor transforms and computing gray level co-ocurrence matrices. Their results showed that adaptive scheme performs better compared to the systems that use only one fixed representation of texture. Nevel [7] performed texture synthesis via matching the first and second order of a wavelet frame decomposition. Yuan [8][9] et al. used two-component GMM for the distribution of wavelet coefficients in high frequency subbands and estimated the model by applying EM algorithm. They used Kullback divergence to measure the similarity between the images. Liu and Wada proposed a texture characterization method which is robust to geometric distortions such as translation, rotation and scale [10]. They introduced log-polar transform of autocorrelation image to eliminate the effect of geometric distortions. Texture features are then computed from the statistics of wavelet packet decomposed images. Kokare et. al. designed a new set of two-dimensional dual-tree rotated complex wavelet filter and used it in combination with dual-tree-complex wavelet transform. They calculated texture features in 12 different directions [11].

1.1.2 Color Based Image Retrieval

There are many existing techniques for the color based image retrieval. Brunelli [12] used histograms of low level image features, such as color and luminance and defined a histogram capacity curve to deal with the density distribution of histograms in the corresponding spaces. He used this histogram capacity curve to quantify the effectiveness of image descriptors and histogram dissimilarities in Image Retrieval applications. Color histogram is the most commonly used color feature representation [13]. An alternative approach is color moments, which overcomes the quantization effects of color histogram [14]. Smith et al. proposed color sets transform RGB color space to HSV color space[15]. Androutsos[16] et.al proposed vector angular-based distance measurement for color Image Retrieval.

Yang et. al. first transformed the images to LHS color coordinate system from RGB color space [17]. They used luminance component for textural analysis while hue and saturation components were used for chromatic analysis. For the color representation, a histogram of 10 x 10 chromatic bins was constructed where bins of the histogram correspond to different colors in the LHS space. Ju Han and Kai-Kuang Ma proposed the idea of a new color histogram representation, called fuzzy color histogram (FCH) [18]. This approach takes into consideration the color similarity of each pixel's color associated to all the histogram bins through fuzzy-set membership function. This technique overcomes the drawback of conventional color histogram representation because conventional histogram does not consider the color similarity across different bins and the color dissimilarity in the same bin. Lu and Burkhardt proposed a new feature for color image retrieval which is based on DCT-domain vector quantisation (VQ) index histograms (DCTVQIH) [19]. The images are first transformed to YCbCr Color Space from RGB Color Space. After randomly selecting a certain number of images from the database to be the training images, they generated 12 codebooks. Each image in the database is then encoded using these codebooks. They constructed 12 histograms (4 for each color) from the 12 DCT-VQ index sequences.

1.1.3 Shape Based Image Retrieval

Shape is also an important feature for perceptual object recognition and classification in images. Shape description techniques can be divided into two categories; region-based and contour-based techniques. Region-based shape descriptors represent all the pixel information across the region. Contour-based shape descriptors represent region boundaries. In this project, we concentrate on contour-based descriptors.

Several contour-based techniques including chain codes, curvature scale space descriptor, and Fourier descriptors have been proposed and used in various applications such as object recognition, shape coding and handwriting recognition. Mokhtarian [20] et al. applied the curvature scale space descriptor for retrieval of shapes. Zhang[21] et al. showed that using centroid distance outperforms other shape signatures in shape based retrieval.

P. Parent [23] used curvature as descriptor which is estimated at each point by taking the angular difference between slopes of two line segments to the data points before and after each point. Tangent information is sufficient for the recovery of the trace of a curve in an image. The curvature of the contour defines a shape completely. Curvature measures orientational change at each point along a contour. Choo et. al. applied chain code as Shape descriptor [24]. Dengsheng Zhang and Guojun Lu proposed a generic Fourier descriptor (GFD) to overcome the drawbacks of existing shape representation techniques [25]. GFD is derived by applying a 2D Fourier transform on a polar raster sampled shape image. They claimed that this new shape descriptor is application independent and robust. Baofeng Guo and Jianmin Jiang introduced a new shape descriptor in the wavelet compressed domain. They used morphological operators to refine the significant map of the image and remove the isolated points [26].

Torres et. al. introduced a robust approach to compute contour saliences that is defined as the influence areas of its higher curvature points [27]. They exploited the relation between a contour and its skeleton and modified the original definition to include the location and the value of saliences along the contour. They also proposed a new similarity measure to compare contour saliences. In [28], Badawy and Kamel introduced a shape-based Image

Retrieval technique based on concavity trees. They propose new algorithms for the contour-based concavity tree extraction and concavity-tree matching. Their results indicated that concavity trees boost the retrieval performance of two feature sets by at least 15 percent when tested on a database of 625 silhouette images.

1.2 Motivation

Content based image retrieval systems are natural extension of the keyword search engines. These search engines are available for looking into the huge text files available on the WWW. Similar strategies are required for searching through the large amount of information available in the form of digital images. Automatic indexing and retrieval of these images based on their content is essential for efficient usage of image databases. This is the main motivation of this thesis.

The hat database used in the experimental analysis of the purposed features contains a variety of hats. These hats are pre-classified according to their usage and function. Therefore hats of different colors and textures might fall into the same category because of their common function. Hence color and texture are not the appropriate features for the indexing of the hat database. This is the motivation for selecting shape based features for indexing and retrieval of the hat database.

1.3 Project Outline

The remaining part of this thesis is organized as follows:

Chapter 2 starts with a detailed description of the wavelet transform, then characterize the texture feature from a statistical perspective. Mathematical framework of GMM and expectation-maximization (EM) algorithm are also introduced in this chapter. Chapter 3 presents color based image retrieval. Color spaces are introduced and the advantages and disadvantages of using RGB and HSL color space are also described.

The Chapter 4 concerns the shape feature extraction. Polar coordinates and Fourier descriptor (FD) are introduced in this chapter. A modified Fourier transform based on polar coordinates is proposed. In Chapter 5, how the original data was acquired is illustrated. This chapter also describes the database construction from the original data by edge detection. The efficiency of the features and their different combinations by simulations is investigated. The superiority of the proposed shape matching and retrieval over other perceptual attributes/features such as texture, color and their combinations for this database is demonstrated. The effectiveness color, texture, and shape features for the identification and classification of the hat database is compared. In conclusion, shape features based on centroid-contour distance Fourier descriptor gives the best results for hat database compared with other features.

Chapter 2

Texture Based Image Retrieval

2.1 Introduction

Texture attribute of an image represents the variations in the level of brightness or special frequencies of image pixels. It reflects the structural arrangement of surfaces and their relationship to the surrounding environment. descriptors for the texture can be classified into two categories; statistical based and transform based. The first approach explores the gray level spatial dependence of texture and then extracts meaningful statistics as texture representation. The transform-based approach such as Fourier descriptors (FD) first transforms the image into new domain and then finds the shape features based on the characteristics of the new domain.

Discrete wavelet transform (DWT) decomposes an image into different frequency subbands. DWT can be used in the transform analysis of images to extract features from the image contents (such as shape, sketch and texture). Mean and variance of the coefficients in each subband were used as texture features in the earlier research works. The following sections introduce DWT, GMM and EM algorithm to extract the texture features from the images.

2.2 Wavelet Transform

Wavelet transform replaces the sinusoidal waves in Fourier transform by a family generated by translations and dilations of a window called a wavelet. It takes two arguments; time and scale. The wavelet transform thus has a time frequency resolution which depends on the scale. One dimensional continuous wavelet transform is defined as:

$$\gamma(s,\tau) = \int f(t)\varphi^{*}_{s,\tau}(t)dt \tag{2.1}$$

Where " $*$ " denotes complex conjugation. This equation shows how a function $f(t)$ is decomposed into a set of wavelet coefficients (γ). The variables s and τ called scale and translation respectively are the new dimensions after the wavelet transformation. The inverse wavelet transform is defined as:

$$f(t) = \int\int \gamma(s,\tau)\varphi_{s,\tau}(t)dsd\tau \tag{2.2}$$

The wavelets are generated from a single basic wavelet $\varphi_{s,\tau}(t)$, the mother wavelet, by scaling and translation:

$$\varphi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\varphi(\frac{t-\tau}{s}) \tag{2.3}$$

Where the factor $\frac{1}{\sqrt{s}}$ is used for energy normalization across different scales.

Image can be decomposed by using 2-D wavelet transform. 2-D wavelet transform decomposes the image into 4 subbands at each wavelet scale or level. The wavelet coefficients carry intensity variation information of image along horizontal, vertical and diagonal directions in high frequency subbands, and a scaled down low resolution approximation of the original image in the low frequency subband. 2D-DWT is implemented by using digital filters and downsamplers.

Figure 2.1 shows a graphical view of 3-level wavelet decomposition of an image. The LL is the low resolution approximate subband, while the HL, LH and HH represent the horizontal, vertical and diagonal subbands.

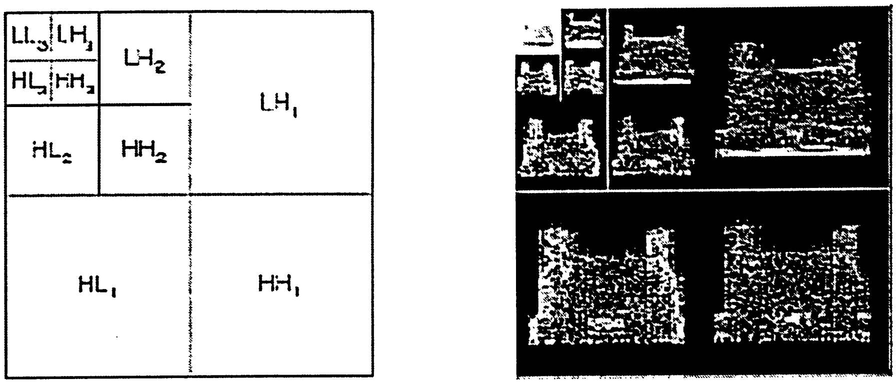


Figure 2.1: Wavelet decomposition

2.3 Gaussian Mixture Model (GMM)

2-D DWT represents an image with both spatial and frequency characteristics. In practice, the probability density function (PDF) of the wavelet coefficients can be described by a peak (centered at 0) and heavy-tailed non-Gaussian density. The density estimation of wavelet coefficients such as mean, and variance, are used as texture features.

There are two classes of density estimation; non-parametric and parametric. Non-parametric methods do not assume a probability density function of the data where the samples are taken from. They need more data to represent the probability density. Histogram, kernel density estimator, nearest neighbor method and adaptive kernel estimators are non-parametric approaches. Parametric methods assume that the data has been generated from certain probability distributions and then estimate the moments of those probability distributions.

GMM is a popular method in the parametric estimation class. Assuming that the data is from a mixture of Gaussian distributions; the means and covariances of the Gaussian Components are to be estimated. Modeling a texture class with a GMM rather than a single Gaussian gives a great deal of added flexibility to the model. It is possible to approximate any arbitrary shape of distribution by a mixture of Gaussians if we have infinite number of components in the mixture. However, the computation of parameters for a large number of

Gaussians in the mixture model makes the model practically infeasible.

Features are extracted from the wavelet coefficients by modeling their distributions as a mixture of Gaussian distributions. Mixture of Gaussians is commonly used model for peaky distributions because they are very hard to approximate by a single distribution. Wavelet transform of the most signals are sparse, resulting a large number of small coefficients and a small number of large coefficients. A typical wavelet coefficient density or histogram is much more peaky at zero and heavy-tailed than a Gaussian. Wavelet coefficients are modeled as a mixture density with hidden state variable.

For a two component Gaussian mixture model for the distribution of wavelet coefficients, the mixing proportion of each distribution π_1, π_2 and variance of each distribution are going to be estimated.

$$P = \pi_1 p_1 + \pi_2 p_2 \tag{2.4}$$

$$\pi_1 + \pi_2 = 1 \tag{2.5}$$

where p_1 and p_2 are both Gaussian components, π_1 and π_2 are the mixing proportions and P is the total probability.

2.4 EM Algorithm

In order to apply GMM classification procedure, the parameters of the model by using a maximum likelihood estimator need to be estimated. The EM algorithm provides a general approach to the problem of maximum likelihood parameter estimation in statistical models with hidden variables. The EM algorithm for the estimation of mean values and covariance matrices of GMM is based on iterated linear regression analysis. EM algorithm is an algorithm for finding maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved latent variables.

EM alternates between performing an expectation (E-step), which estimates the distributions of the hidden variables given the data and the current value of the parameters; and a maximization (M-step), which modifies the parameters in order to maximize the joint distribution of the data and the hidden variables. E-step computes the expectations as given in the following equations [29]:

$$p(x_i|m,\theta) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left(-(x_i - \mu_m)^2/\sigma_m^2\right) \tag{2.6}$$

$$< z_{im} >= \frac{p(x_i|m,\theta)}{\sum_j^M p(x_i|j,\theta)\pi_j} \tag{2.7}$$

Where m indicates the component of the mixture model from which the data point x_i was generated. θ is the parameter vector of the Gaussian distribution with mean (μ) and variance (σ^2) as its components. M is the total number of components in the GMM, j is one of M components. Z_i represents the hidden variables and indicates which component has generated a particular data value. Since the values of Z_i for each x_i are unknown, these variables are called hidden variables. z_{im} is the m-th component of vector Z_i and $< . >$ is the expectation operator.

M-step maximizes the expected complete log-likelihood with respect to the parameters of the model i.e. μ_m, π_m and σ_m; $m = \{1, ..., M\}$ by an iterative procedure and get the parameters. Differentiating the expected log-likelihood function and maximizing with respect to the parameters, the following update equations [29] are obtained:

$$\mu_m = \frac{\sum_i^N < z_{im} > x_i}{\sum_i^N < z_{im} >} \tag{2.8}$$

$$\sigma_m^2 = \frac{\sum_i^N < z_{im} > (x_i - \mu_m)^2}{\sum_i^N < z_{im} >} \tag{2.9}$$

$$\pi_m = \frac{\sum_i^N < z_{im} >}{N} \tag{2.10}$$

Where N is the total number of data points. Maximum likelihood estimation of a finite mixture of Gaussians can be estimated trivially if the mixing distributions are known.

2.5 Texture Image Retrieval Implementation on Hat Database

In order to isolate texture and color information, the image is mapped from the RGB color space to the HSL color space. The luminance channel contains the gray scale component, and consequently the texture information. The H channel contains the color information.

The wavelet feature vector from the L channel using two levels of DB2 wavelet decomposition is extracted. The two levels of horizontal, vertical or diagonal detail coefficients are analyzed. Gaussian mixture model is applied to each of the detailed wavelet subbands to extract mean, variance and mixing probabilities of the two Gaussian components. There are 6 detail subbands for 2 level wavelet decomposition of the images and hence the dimension of the texture feature vector is 36.

A measurement of similarity between two images is obtained by computing the distance between their 36-dimensional feature vectors. Euclidean distance between the feature vectors to determine the texture similarity is used. Smaller the distance greater is the similarity.

Figure 2.2: Sample image retrieval result by applying texture feature

Figure 2.2 shows the sample retrieval result obtained by using the texture feature. From this sample retrieval result, it can be observed that most of the retrieved hats are made of fibre similar to the query image in texture but are different in styles and color. Therefore it is no possible to obtain better retrieval efficiency by using just the texture feature for this hat database. The detailed experimental results can be found in chapter 5.

2.6 Chapter Summary

This chapter provides a brief overview of the wavelet transform and explains the EM algorithm which is used to estimate the parameters of the GMM. In order to extract the texture feature, the images from RGB color space are transformed to HSL color space. The luminance (L) channel is then used for the extraction of texture feature vector. The intensity images formed by the luminance (L) channel are decomposed by applying 2D DWT. DB2 wavelet kernel is used and the decomposition is performed to 2 levels giving 6 detail coefficient subbands and 1 approximate coefficient subband. A two component GMM is applied to each of the detail subband and parameters of GMM are determined using the EM algorithm. The feature vector is composed of mean, variance and mixing probability of each Gaussian component. Although the texture is a very important characteristic of the images, yet it is observed that the retrieval efficiency is very low by using just the texture feature on the hat database. This is due to the fact that hats in the database are very similar in texture and most of them are smooth textured. The hats are more distinctive in color and shape. Hence color and shape features are extracted also which will be explained in the next chapters.

Chapter 3

Color Based Image Retrieval

3.1 Introduction

Color perception is response of the human visual system to different wavelengths of light. GMM in HSL color space is applied to extract the color features. The color information is usually represented by points in 3-dimensional color space. Color space representing colors along human perceptual dimensions is crucial in grouping colors based on color perceptual similarity. There are many color spaces created for different applications such as RGB, HSV, HSL, CMY, CMYK, and indexed color space. The retrieval performance of color features is related to the used color space.

The discussion is limited to RGB and HSL color spaces because RGB and HSL are the most popular ones for image analysis and indexing. The discussion starts with RGB color space as RGB color space is the most commonly used color space in computer graphics, primarily because it is directly supported by the most color displays and scanners. Images acquired from scanners and digital cameras usually use RGB color space. Then HSL color space is introduced and described the advantages of using HSL color space in image retrieval. The transformation of RGB color space to HSL color space will also be discussed.

3.2 RGB Color Space

RGB stands for red, green and blue. RGB color space is a color system based on trichromatic theory. According to the trichromatic theory, three fundamental lights, red, green and blue can produce any visible color. The idea is based on the structure of human eye that contains three color sensitive sensors. RGB is a convenient color model for computer graphics and is the most widely used color system in monitors, televisions, scanners and digital cameras.

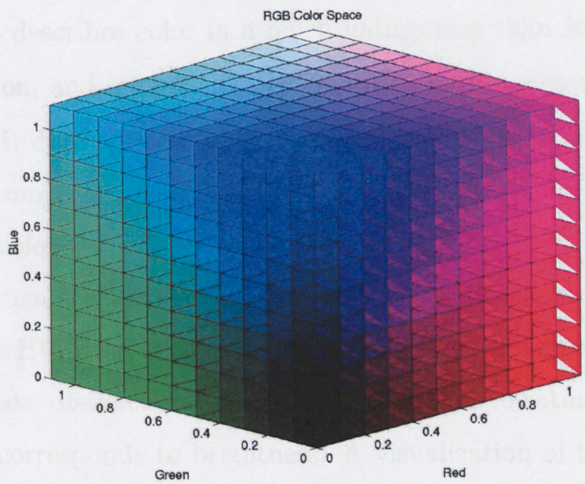


Figure 3.1: RGB color space

RGB color space is often illustrated by a unit cube. Each color is assigned to one of the three orthogonal coordinate axes in 3D space representing red, green and blue. Along each axis of the color cube the colors range from no contribution of that component to a fully saturated color. Any color within the cube is specified by three numbers. The diagonal line of the cube from black (0,0,0) to white (1,1,1) represents all the greys, that is, all three components red, green, and blue are the same.

However, RGB color space is not a good choice for extracting color features from the images. First of all, RGB color spaces are device dependent and additive, which indicates different device gives different color when displaying. Second, RGB cube is smaller and represents fewer colors than those humans can see. Moreover, all the three channels represent color information, changing the color and keeping the brightness constant affects the ratio of the three channels. Similarly when the brightness changes the ratio of the three channels remains the same but the values of three channels are changed. Therefore in RGB color space, color information is dependent on luminance which is not desirable for robust color features.

3.3 HSL Color Space

HSL color space describes color in a much nature way than RGB color space. HSL stands for hue, saturation, and luminance. Instead of using three channels for red, green and blue individually, HSL defines continuing change of colors. Hue measures the central tendency of wavelength; Luminance measures the brightness of the color and saturation measures the intensity of the color.

HSL space is usually depicted as a double cone or double hex cone. In HSL color space the two apexes of the HLS double hexcone correspond to black and white. The angular parameter corresponds to hue, distance from the axis corresponds to saturation, and distance along the luminance axis corresponds to brightness. A visualization of the HSL color space is shown in Figure 3.2.

A hue value of 0 represents red color, green is at a value corresponding to 120, and blue is at a value corresponding to 240. Horizontal planes through the cones as shown in 3.2 are hexagons. The primaries and secondaries (red, yellow, green, cyan, blue, and magenta) occur at the vertices of the hexagons. The saturation component in HSL color space describes color intensity. A saturation value of 0 (in the middle of a hexagon) means that the color is "colorless" (gray) and a saturation value at the maximum (at the outer edge of a hexagon) means that the color is at maximum "colorfulness" for that hue angle and brightness.

The luminance component in HSL color space describes brightness or luminance. A luminance value of 0 represents black. A maximum value for lightness means that the color is white, regardless of the current values of the hue and saturation components. The brightest and the most intense in HSL color space occurs at a lightness value of exactly

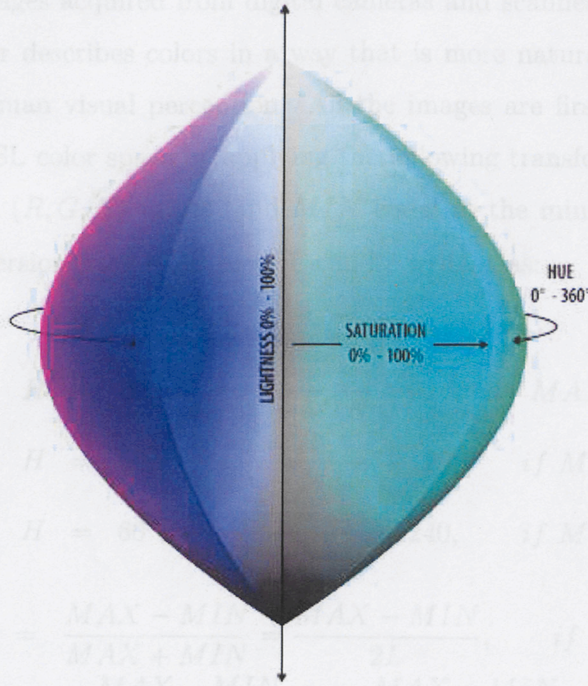


Figure 3.2: HSL color space

half the maximum. HSL color space separates the luminance and the chrominance. In this project, the color feature is extracted from hue channel and the texture feature from luminance channel in the wavelet domain of HSL color space.

3.4 RGB Color Space to HSL Color Space Conversion

The popular RGB color space is efficient for image display and acquisition. RGB color space was used in the earlier color-based segmentation algorithm implementations. However, RGB color space is not suitable for color feature indexing and discrimination as explained in the above sections. In HSL color space, color image processing is performed independently on color channels. It is easier to compensate for artifacts and color distortions. The capacity of luminance and chromatic components of a color is extremely useful in processing images under different illumination situations such as shading, highlights and strong contrast.

The color images acquired from digital cameras and scanners are usually in RGB color space. HSL color describes colors in a way that is more natural to an artist because it is based on the human visual perception. All the images are first converted from the RGB color space to HSL color space by applying the following transformation. Let MAX be the maximum of the (R, G, B) values, and MIN equal to the minimum of those values. The formula for conversion from RGB to HSL can be written as:

$$H = 60 \times \frac{G-B}{MAX-MIN} + 0, \quad if\ MAX = R \tag{3.1}$$

$$H = 60 \times \frac{B-R}{MAX-MIN} + 120, \quad if\ MAX = G \tag{3.2}$$

$$H = 60 \times \frac{R-G}{MAX-MIN} + 240, \quad if\ MAX = B \tag{3.3}$$

$$S = \frac{MAX-MIN}{MAX+MIN} = \frac{MAX-MIN}{2L}, \quad if\ L \leq \frac{1}{2} \tag{3.4}$$

$$S = \frac{MAX-MIN}{2-\{MAX+MIN\}} = \frac{MAX-MIN}{2-2L}, \quad if\ L \geq \frac{1}{2} \tag{3.5}$$

$$L = \frac{1}{2}(MAX + MIN) \tag{3.6}$$

In this thesis, hue channel is used to extract the color features. The distribution of hue channel values is modeled with a two component GMM. EM algorithm is applied to acquire the parameters of the two component GMM.

3.5 Color Image Retrieval Implementation

All the images are first mapped from the RGB color space to the HSL color space. The color information is contained in hue (H) channel. The luminance (L) channel contains the gray scale component, and consequently the texture information. The mean value of the Hue channel is calculated and used as a feature. The parameters of Gaussian mixture model including mean values of two Gaussian components, variances and mixing probabilities are also extracted to be the components of the feature vector. The total length of the color feature vector is 7.

A measurement of similarity between two images is obtained by computing the distance between their 7-dimensional feature vectors. The distance measure between the two feature vectors is based on Euclidean distance with correction on mean value when the difference is greater than *pi*. The value of *pi* has been taken correct to 2 decimal places as 3.14 for implementation purposes.

Let X and Y be the color feature vectors of two images. When $|X - Y|$ is less than pi, the distance is given by:

$$D(X,Y) = (\sum(|X - Y|)^2)^{1/2} \tag{3.7}$$

When $|X - Y|$ is greater than pi, the distance is given by:

$$D(X,Y) = (\sum(|X - Y| - pi)^2)^{1/2} \tag{3.8}$$

Figure 3.3 shows the sample experimental result by using the 7-dimensional color features vector. The detailed results are presented in chapter 5.

Figure 3.3: Sample image retrieval result by applying color feature

3.6 Chapter Summary

In this chapter, the procedure of color feature extraction was presented. A discussion on the RGB and HSL color spaces was presented and the advantages of using the HSL color space over the RGB color space were enumerated. The color features are obtained from the hue (H) channel after the transformation of the images from the RGB to HSL color space. The transformation procedure from RGB to HSL color spaces was explained and the relevant equations were given. The mean of the H channel forms one component of the color feature vector. The other components of the color feature vector are parameters of the GMM. The values of the H channel are modeled by two component GMM and model parameters (mean, variance and mixing probability of each component) are extracted by EM algorithm. The dimension of the color feature vector is 7. The similarity measure based on the Euclidean distance was defined in the last section. The retrieval result obtained by using only the color feature vector for a sample query was also given in the last section. In the next chapter, the extraction of the shape feature vector will be discussed.

Chapter 4

Shape Based Image Retrieval

Ideally, shape descriptor should be invariant to translation, scale and rotation. It is also required that the shape descriptor should also be robust to noise. For application to the hat database investigated in this project, both rotational and scale invariance is required. This is due to the fact that all the images are centered before hand that makes translation invariance not critical. However, it is noticed that scale is not an important characteristic of hat images because the images are normalized in the preprocessing.

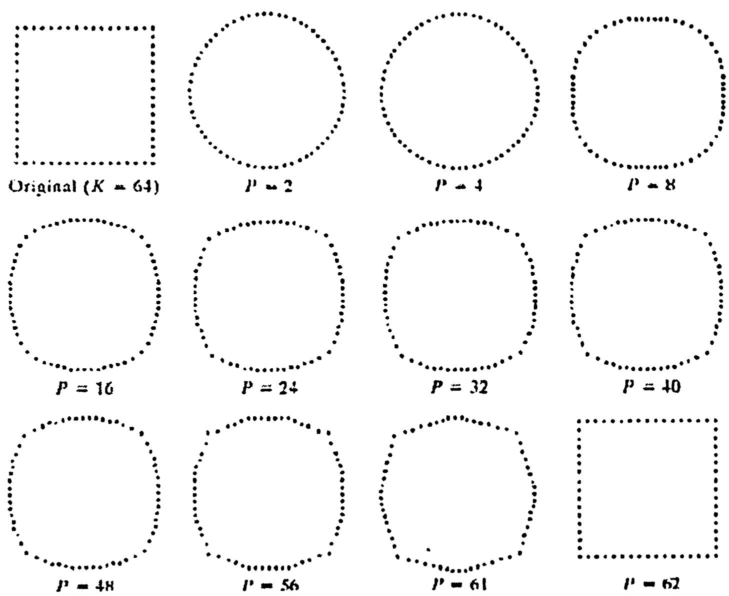
The boundary of an image can be represented digitally by N points, $(x(k), y(k))$, for $k = \{0, 1, 2, \ldots, N-1\}$, where N is the length of the boundary. These boundary points can be represented in terms of complex numbers by considering the image as a complex plane. The X-axis represents the real axis while Y-axis is taken as the imaginary axis. Complex coordinate function $z(k) = x(k) + jy(k)$, for $k = \{0, 1, 2, \ldots, N-1\}$, expresses the boundary points as a one dimensional signal. The discrete Fourier transform (DFT) of $z(k)$ is defined as:

$$a(u) = \frac{1}{K} \sum_{0}^{K-1} z(k) e^{\frac{-j2\pi uk}{K}} \tag{4.1}$$

for $u = 0, 1, 2, , K-1$. The complex coefficients $a(u)$ are called Fourier descriptors (FD) [32].

Global shape is captured by the first few low frequency terms in FD descriptors, while higher frequency FDs capture finer details of the shape. Fourier descriptors overcome the noise sensitivity in the shape signature representations and it is easy to normalize and preserve information.

The following figure shows the reconstruction of an image from different number of Fourier coefficients [33]. It can be observed that by increasing the number of coefficients for reconstruction, the finer details of the shape become visible in the reconstructed images. The global shape of the object can be captured by only a few low frequency Fourier descriptors.



This boundary consist of 64 points, P is the number of descriptors used in the reconstruction of the boundary

Figure 4.1: Reconstruction of image from Fourier coefficients

4.1 Complex Coordinates Fourier Descriptor

A complex coordinates function is the complex number generated from the boundary coordinates:

$$z(k) = [x(k) - x_c] + j[y(k) - y_c] \tag{4.2}$$

where $(x(k), y(k))$, $k = \{0, 1, 2, \ldots, N-1\}$ are the coordinates of the pixel on the boundary and (x_c, y_c) are the coordinates of the center point. Complex Fourier descriptors allow scalability in describing a curve. With a low frequencies subset of descriptors, the reconstructed curve approximates the outline of a shape. By increasing the number of components in the description, high frequencies are also included, and sharp curves or details can be reconstructed.

4.2 Centroid-Contour Distance Fourier Descriptor

In polar coordinates the distance of a point on the boundary to the centroid can represent the shape of the image. In this way, it also changes two-dimensional image into one-dimensional signal. r, θ plot is a boundary descriptor in polar coordinates relative to the centroid. In polar coordinated, each point on the boundary is determined by the distance from the centroid and the angle from a certain reference. Each point $(x(k), y(k))$ is converted to polar coordinates $r\langle\theta$ by using the following equations:

$$\begin{aligned} r(k) &= \sqrt{x(k)^2 + y(k)^2} \\ \theta(k) &= \arctan \frac{y(k)}{x(k)} \end{aligned}$$

Where $\theta = [0, \ldots, 2\pi]$ is the angle of the vector (from centroid to the boundary point) with positive X-axis measured in the counter clockwise direction. r is the distance of the point from the centroid. The problem with this approach is that both r and θ are rotation variant. Another problem with this approach is that r might have multi-values unless the shape is very simple. The usual way to deal with multi-values is to discard all other values except the outermost value.

In this project centroid-contour distance (CCD) and FD are combined to improve the retrieval efficiency. 1-D discrete Fourier transform of CCD, $r(k) = \sqrt{x(k)^2 + y(k)^2}$, is given by:

$$a(u) = \frac{1}{K} \sum_{0}^{K-1} r(k) e^{\frac{-j2\pi uk}{K}} \tag{4.3}$$

where $u = \{0, 1, 2, \ldots, K-1\}$. The benefits of the centroid-contour distance Fourier descriptor are invariance to the start point of the boundary and rotational invariance but it does not have scale invariance. Rotation of the object in cartesian space results in circular shift in polar space; and the circular shift does not change the spectra distribution on polar space. Because shift or translation in spatial domain results in phase change in spectral domain, the spectra becomes shift or translation invariant by ignoring the phase. The scale invariance is achieved by scaling all the images in the database to the smallest size image in the database. The polar Fourier spectra is more concentrated around the origin of the polar space. This is particularly well-suited for shape representation, because for efficient shape representation, the number of spectra features selected to describe the shape should not be large.

4.3 Shape Image Retrieval Implementation

Shape feature is extracted from the edge image of the hats. The edge maps of the image are obtained by applying the Canny edge detector which will be explained in chapter 5 section 5.4. The images are centered and then transferred to the polar coordinates. Fourier descriptors of 5, 10, 30, 45, 90, 180, 360 points are applied respectively. The lengths of corresponding feature vectors are 5, 10, 30, 45, 90, 180, 360 respectively. Similarity between two images is measured by computing the Euclidean distance between their respective multi-dimensional feature vectors. Figure 4.2 shows the sample retrieval result by using only the shape feature vector. More detailed results can be found in the next chapter.

4.4 Chapter Summary

In this chapter, the procedure to extract the shape feature vector for the indexing and retrieval of the hat image database was discussed. The images are processed using the Canny edge detector to obtain the edge maps. The coordinates of the boundary point are represented in polar coordinates to obtain a one dimensional centroid-contour distance function from the two dimensional edge map. A one dimensional Fourier transform is applied to the centroid-contour distance. The FD of CCD is rotational invariant and also invariant to the start point. However, it is not scale invariant. Scale invariance is achieved by preprocessing

Figure 4.2: Sample image retrieval result by applying shape feature

all the images and scale them to a standard size. Low frequency Fourier descriptors are used as features for the shape. The retrieval result obtained by using only the shape feature vector for a sample query was also presented in the last section. In the next chapter, a detailed experimental analysis of the retrieval results using texture, color and shape features and their various combinations will be performed.

Chapter 5

Experimental Results

5.1 System Overview

The image retrieval system consists of five major components: user interface, image database, feature database, feature extraction and search engine. These components are explained below:

User interface: The user interface is the graphical interface for users to input their selections, and for the computer to display the results. Users can browse the images by file name from the database and can present a query image to the system for retrieval. The query image is displayed on the screen. The users can press the Retrieval button to search the images similar to the query image. The 20 most similar images are displayed to the user.

Image database: The image database is composed of 86 color hat images with a variety of color, texture, and shape, which are classified into 10 categories according to their functions. The images are stored in RGB color format.

Feature database: The color, texture and shape features for each of the image in the database are extracted off line and saved in the feature database. When user presses the Retrieval button, the feature distance of the query image and all the other images in feature database are calculated and ranked, and the similar images are identified.

Feature extraction: Feature extraction is the process of extracting feature vectors from the image to represent image. There are three kinds of feature used in this project: color, texture, and shape. The feature extraction for the images in the database is extracted off line.

Search engine: Search engine is a program where the retrieval is preformed. It takes the feature of the query image, calculates the distance between the feature vectors in the database and the feature vector of the query image. It then ranks the images according to the distances between their corresponding feature vectors. The smallest the distance, the largest is the similarity value.

The following block diagram shows the system architecture and the associations between different system components.

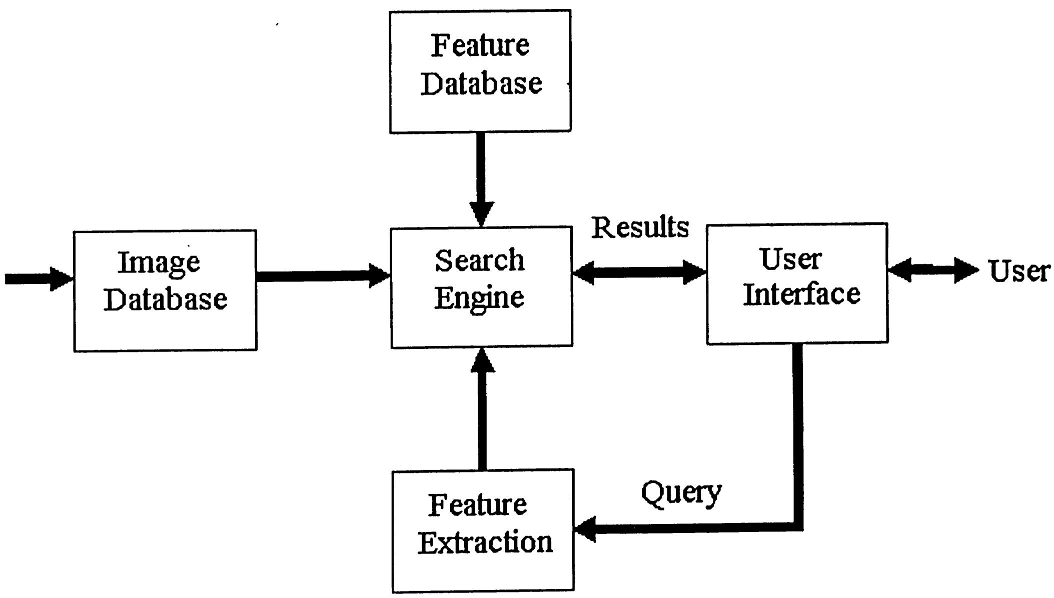


Figure 5.1: Block diagram of image retrieval system

Among the five system components described above, the feature extraction plays the most important role in image retrieval system. The following section describes our feature extraction procedure in an algorithmic format.

5.2 Feature Extraction

In this project, color, texture and shape features are used for the indexing and retrieval of hat image database. The extractions of all three feature sets have been explained in the preceding chapters. In this section, the feature extraction process is summarized and the algorithm is explained with the help of a block diagram. The following is the block diagram of the feature extraction process.

The algorithm consists of the following steps:

- The input RGB image is converted to HSL color space for feature extraction. The conversion formulae from RGB to HSL color space are given in Equations 3.1-3.6 in chapter 3, section 3.4.

- The hue channel is then used for the color feature extraction. The hue channel is fed to the GMM block where the parameters of a two component GMM model are extracted which form the color feature vector. The parameters of the GMM are calculated in an iterative manner using the update equations 2.8-2.10 given in chapter 2, section 2.4. The model parameters are initialized before starting the iterative EM algorithm. The mixing probabilities are intialized to be equal to 0.5 for the each of the component. The EM algorithm is run a few times with different initial values of variances and means of the two components to avoid local minima.

- The luminance channel is fed to a DWT block where a 2-dimensional 2 level db2 discrete wavelet decomposition is performed. The wavelet coefficients are then supplied to the GMM block to extract the parameters of a two component GMM model to represent the texture of the images.

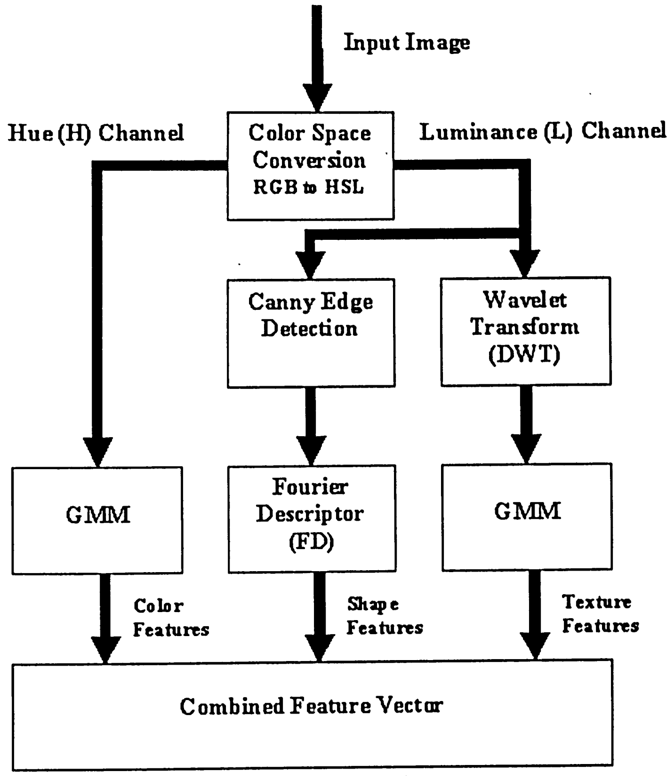


Figure 5.2: Block diagram of feature extraction

- For the extraction of the shape feature vector, the gray scale image is obtained from the luminance (L) channel. The grey scale image is then supplied to the Canny edge detection block to find the image edge map. The edge map is then represented as a complex 1-D signal and 1-D Fourier transform is taken from the centroid-contour distance of the edge image.

- The bottom block combines the color feature, texture feature, and the shape feature sets into one feature vector.

Index	Name of category	number of images
1	Cold colored baseball hat	9
2	Hard hat	9
3	Hard hat with cord	16
4	Kids' hat	5
5	Kids' hat with ear protection	7
6	Knitted hat	13
7	Skates ice hat with line patterns	9
8	Skates hat	5
9	Warm colored baseball hat	7
10	Western cowboy hat	6

Table 5.1: Classification of the hat database

5.3 Database Composition

The database consists of 86 hat images. These images are classified into 10 categories namely cold colored baseball hat, hard hat, hard hat with cord, kids' hat, kids' hat with ear protection, knitted hat, skates ice hat with line patterns, skates hat, warm colored baseball hat and western cowboy hat. Table 5.1 shows all the 10 categories of hat with the corresponding number of image in each category.

5.4 Edge Detection

The original images are of different size and ratio of dimensions. In order to obtain the comparable shape features of hat, the images are centralized and their sizes are normalized. In order to calculate contour-based shape descriptor the edge of the region needs to be located. Edges characterize boundaries in images and are areas with strong intensity contrasts or jump in intensity from one pixel to the next. By using image edges for shape description, the amount of data for processing reduces significantly. Moreover, edge detection also filters out useless information while preserving the important structural properties in an image. Edge detection algorithms fall into two major categories; gradient and Laplacian. The gra-

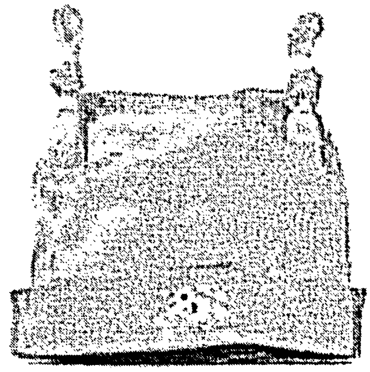


Figure 5.3: Sample image of a hat

dient method detects the edges by using the maximum and minimum values of the first derivative of pixel values in the image. The Laplacian method uses zero crossings in the second derivative of the pixel values in the image. An edge has the one-dimensional shape of a ramp and calculating the derivative of the image can highlight its location[30].

In this project, Canny edge detection is used. The Canny edge detector first smoothes the image to maximize the signal to noise ratio by Gaussian convolution. It achieves good localization to accurately mark edges with 2-D first derivative operator. Edges give rise to ridges in the gradient magnitude image. The algorithm then minimize the number of responses to a single edge by tracking along these regions and suppressing any pixel that is non-maximum. The gradient array is reduced by hysteresis. Hysteresis is used to track along the remaining pixels that have not been suppressed. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is made a non-edge by setting it to zero. If the magnitude is above the higher threshold, it is made an edge. And if the magnitude is in between the 2 thresholds, then it is set to zero unless there is a path from this pixel to a pixel with a gradient above higher threshold. The image is converted to a binary image. Figure 5.4 shows the edge map of hat image in figure 5.3.

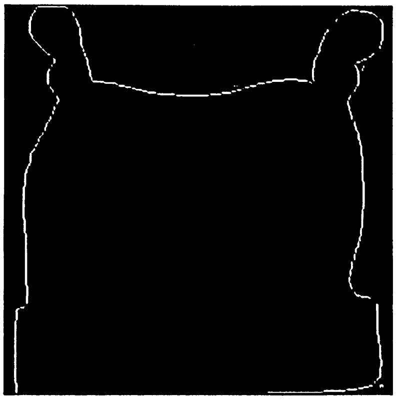


Figure 5.4: Edge map of the hat shown in figure 5.3

5.5 Feature Similarity Measurement

A metric distance between the feature vectors of the query and the test image is commonly used for similarity measurement. The Minkowski-form distance is defined as:

$$D(X,Y) = (\sum |X - Y|^p)^{1/p} \tag{5.1}$$

where X and Y are vectors of dimension N. The above equation is the general form of the distance metric. If $p = 1$, then the distance is known as the City-block or Manhattan distance:

$$D(X,Y) = (\sum |X - Y|) \tag{5.2}$$

When $p = 2$, the distance metric is called Euclidean and is defined as:

$$D(X,Y) = (\sum |X - Y|^2)^{1/2} \tag{5.3}$$

The distance between the query feature and every other feature vector in the feature database is calculated. Images are then ranked according to their distances from the query feature vector. The images corresponding to the lowest 20 distances are displayed for the users of the search engine. From these top 20 matches, the number of image belonging to the same class as that of the query image are found. This gives us the retrieval efficiency of the system.

5.6 Precision-Recall Plot

The performance measures to evaluate the effectiveness of the image retrieval system are precision and recall. The precision of the retrieval system is the ratio of relevant documents that belong to the correct species to the total retrieved documents (k). The recall is the ratio of all relevant documents in the k retrieved images to the total number of relevant documents in the image database. There is always a precision-recall tradeoff. If the information retrieval system returns the entire set of relevant documents, then recall will be 100%, but precision will be very low. If the system returns just one document from the correct class, then precision will be 100%, but recall will be very low. Precision-recall Plot is the the standard measure of performance for information retrieval systems. In this project, the most relevant documents are the ones from the same species as that of the query. The tradeoff can be visualized by plotting precision versus recall as K is increased from one to some maximum value.

5.7 Retrieval Results by using Texture and Color Features and their Combination

Seven color features and 36 texture feature are used for the comparative study. Different combinations of the features are used for evaluation of their relative effectiveness. It is noticed that the results are the lowest for color features among the three. Figure 5.5 shows the retrieval results for the color features. The combination of texture and color, and texture only are similar, both are higher than color feature result. Figures 5.6 and 5.7 show the results for texture and combination of texture and color features respectively. It can be seen from these plots that for the same precision level, the recall rate is much higher for texture features compared to the color features. For example, for the precision level of 0.18, the recall for color features is 0.37 while for the texture features is 0.44.

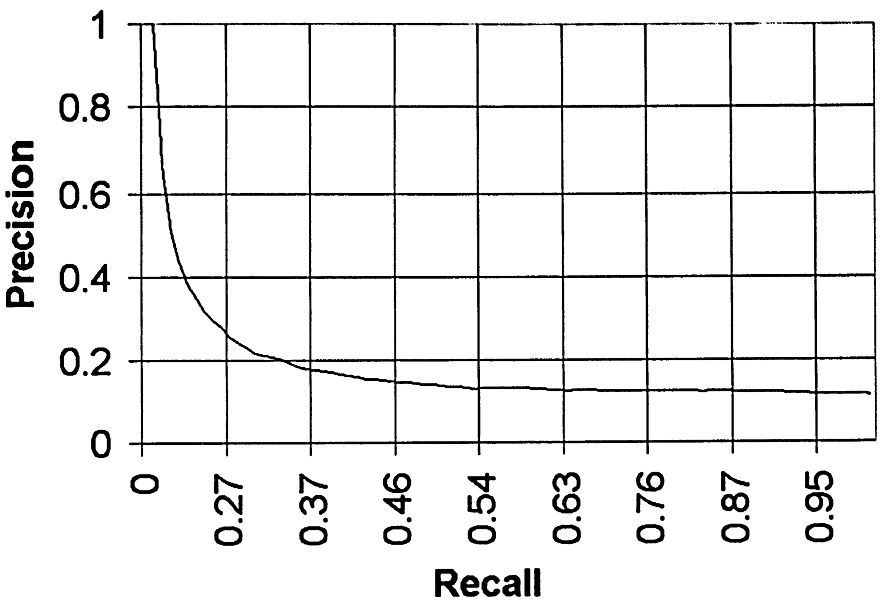


Figure 5.5: Precision-recall plot for color features

5.8 Shape-based Feature Retrieval Results

Shape feature is calculated by using centroid-contour descriptor with N point Fourier transform. Starting with even 5 point FD, it is observed that the retrieval results are much better than those of color, texture and their combinations. In order to optimize the retrieval results, the results obtained from 5, 10, 15, 30, 45, 90, 180 and 360 point FD are examined. The results are shown in the following precision-recall plots. It can be seen from the following figures that the precision-recall curve goes up as by increasing the number of point from 5 to 10. However by increasing the number of points beyond 10, the precision-recall curve decreases gradually. It is observed from the precision-recall plots shown in figures 5.8, 5.9, 5.10, 5.11, 5.12, 5.13 and 5.14 that the retrieval efficiency is the highest when 10 point FD

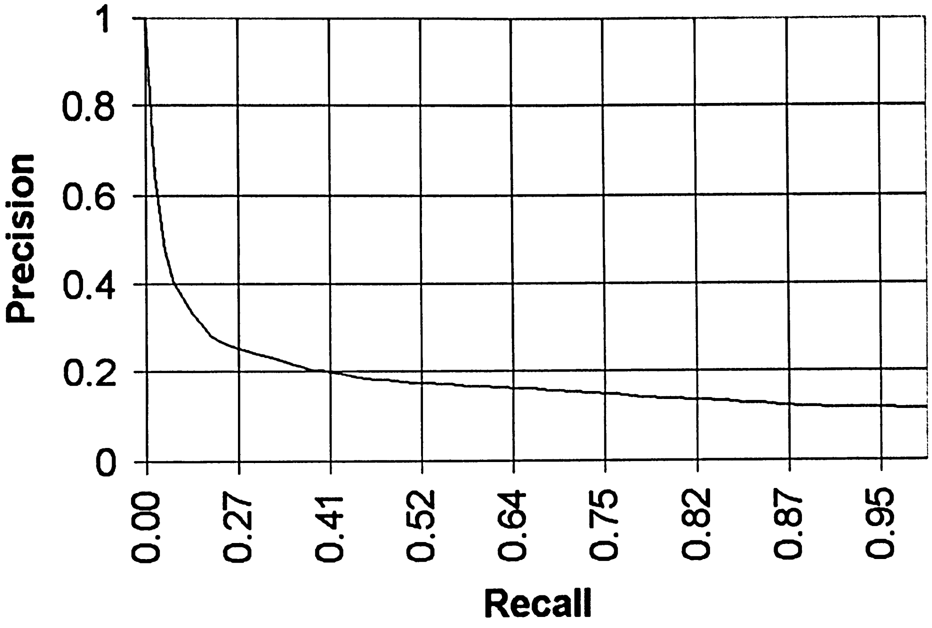


Figure 5.6: Precision-recall plot for texture features

is used.

5.9 Discussion on Results

It is observed from the precision-recall plots that the shape features in the form of FD perform much better than the color or texture features. The retrieval efficiency of the color features is the lowest. The retrieval efficiency of the texture features is slightly better than the color features but much lower than the shape based features. The combination of texture and color features perform slightly better than the texture only features. This indicates that the color features have negligible contribution towards the performance of the retrieval system. However, these results are not general in nature. The selection of features for the indexing and retrieval of images largely depends on the nature of the database and the purpose of the search engine.

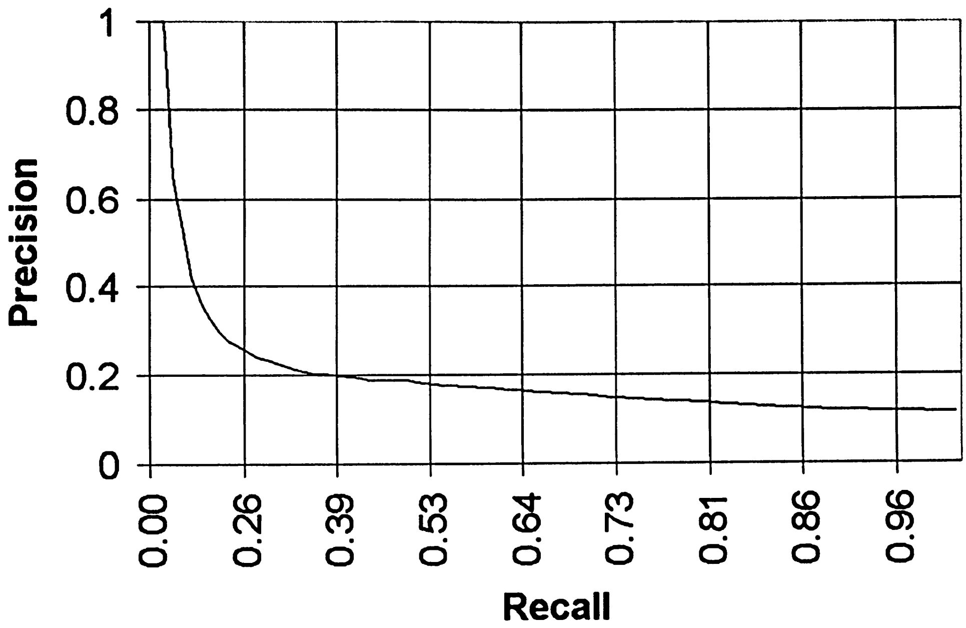


Figure 5.7: Precision-recall plot for the combination of texture and color features

Generally hat classification is performed by the function of the hats. The hats of different colors might fall in the same category. For example in the hard hat category there are 16 hats all with cords and having similar shape but they are of two different colors blue and pink. Therefore, color cannot be used to distinguish this category from other categories. Therefore shape feature is the most suitable feature in this case.

5.10 Chapter Summary

In this chapter, detailed experimental results were presented with the help of precision-recall plots. Texture, color and shape features are their various combinations were tried to get better retrieval efficiency. A brief overview of the Canny edge detection algorithm was also given in this chapter. It is observed that the shape features in the form of FD are the most suitable set of features for the hat image database among the features used in this project.

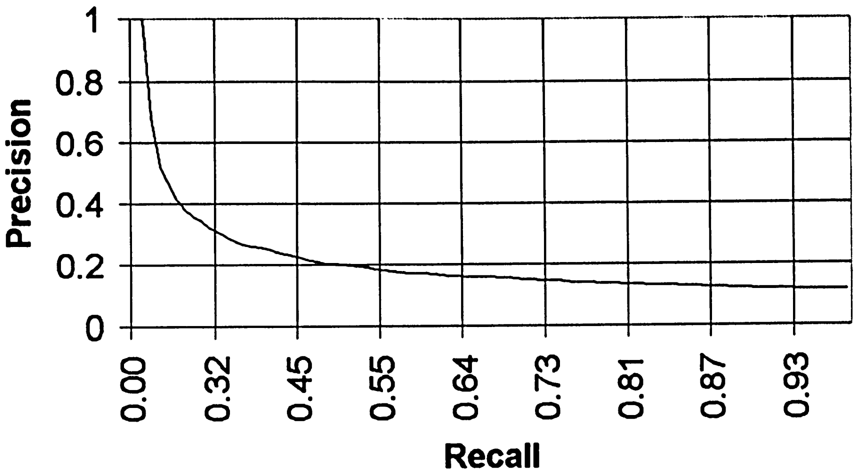


Figure 5.8: Precision-recall plot for 5 FD features

A discussion on the experimental results was presented in the last section of the chapter. In the next chapter, some conclusions based on the various techniques and experimental results performed in this project will be drawn.

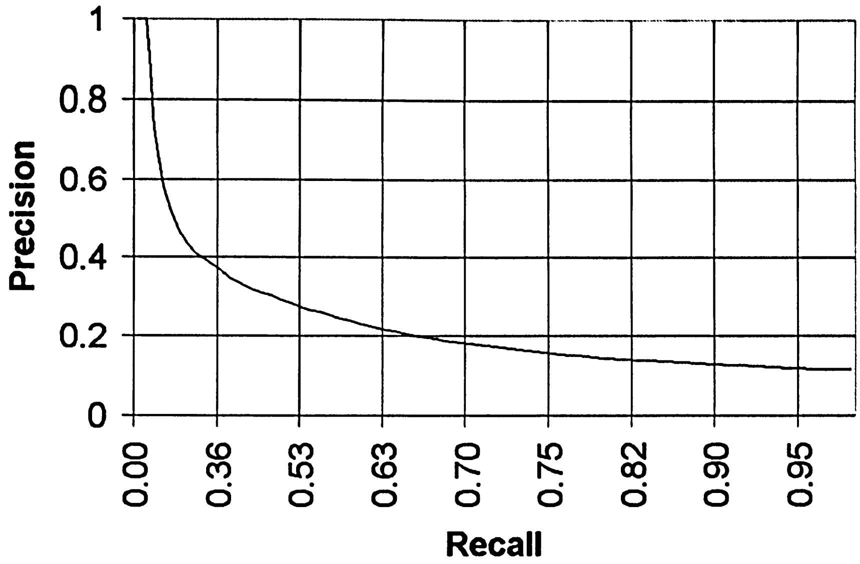


Figure 5.9: Precision-recall plot for 10 FD features

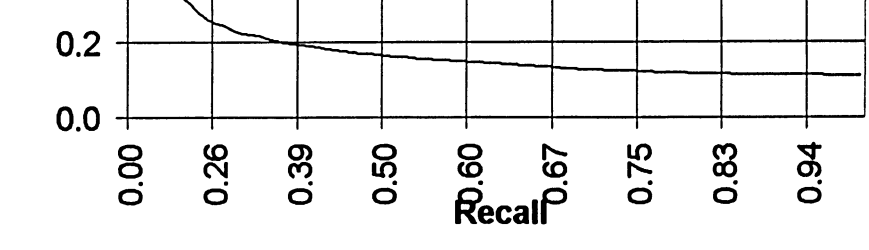


Figure 5.10: Precision-recall plot for 30 FD features

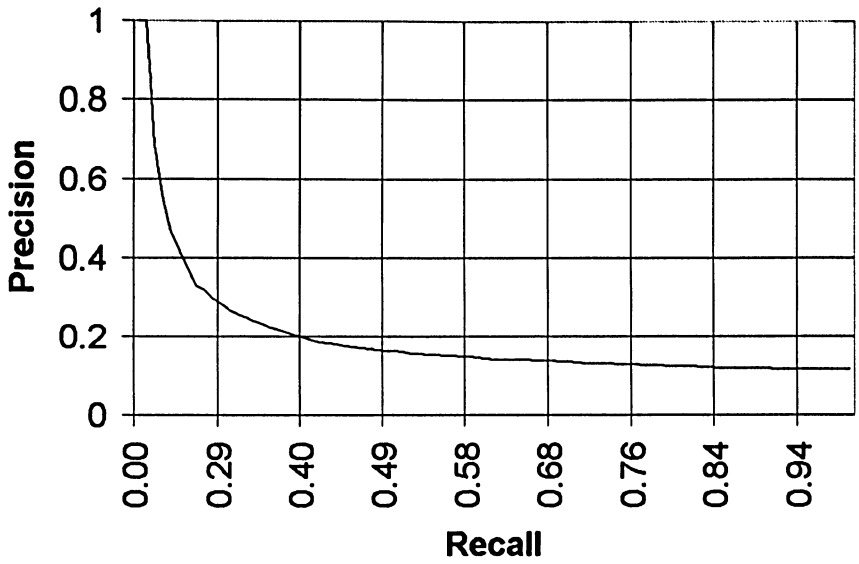


Figure 5.11: Precision-recall plot for 45 FD features

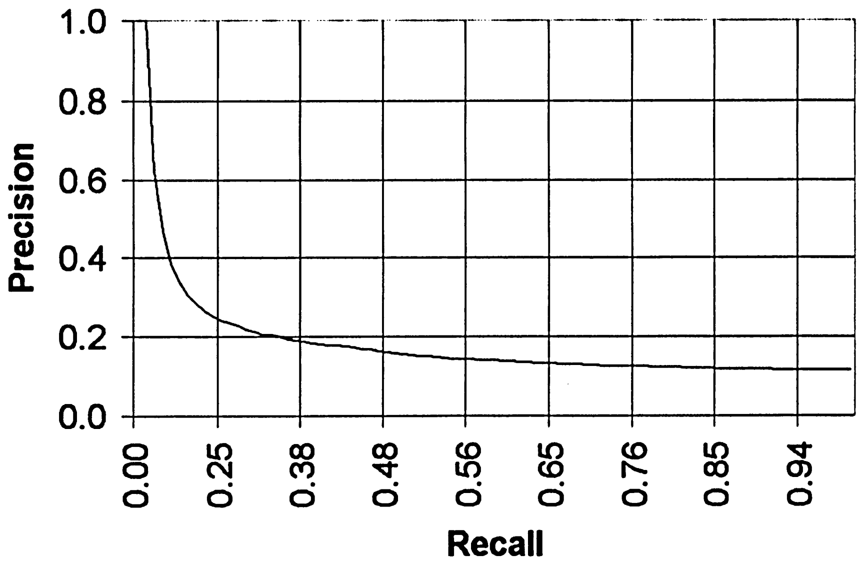


Figure 5.12: Precision-recall plot for 90 FD features

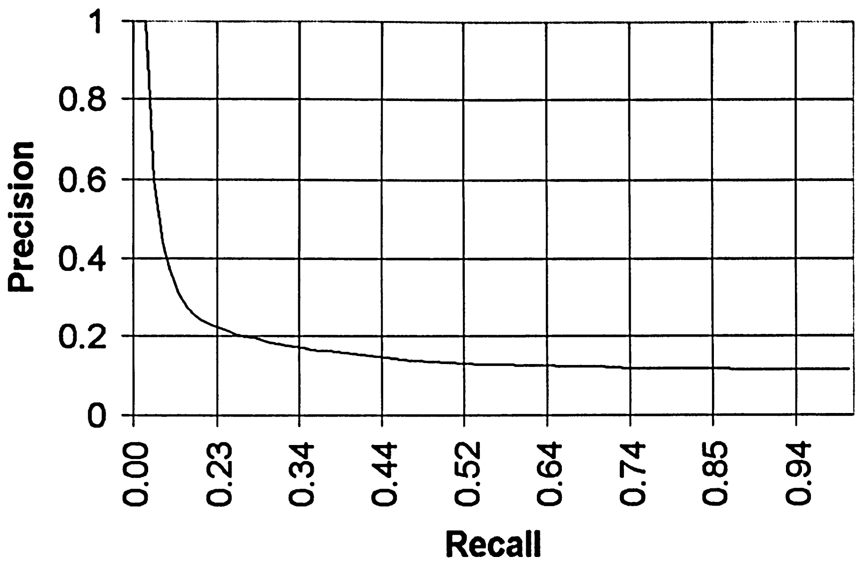


Figure 5.13: Precision-recall plot for 180 FD features

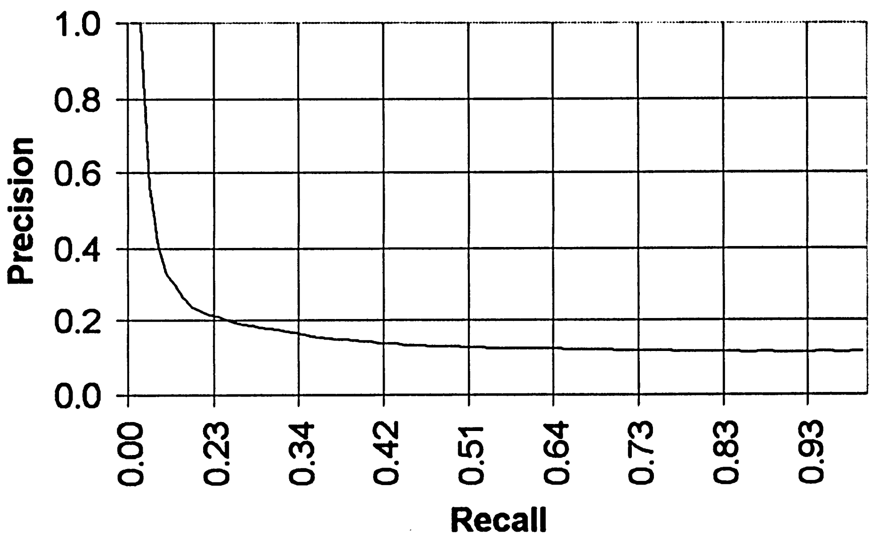


Figure 5.14: Precision-recall plot for 360 FD features

Chapter 6

Conclusions and Future Work

6.1 Conclusions

There are many different image retrieval approaches based on texture, color and shape. In general, there is no single approach which can be described as superior to other for all databases. For this particular hat database used in this project, it is observed that the shape is the most distinct perceptual attribute for all the hats. This can been seen from the results presented in the preceding chapters, centroid-contour distance Fourier descriptor performs far more better than texture based and color based techniques on this specific database.

It is also noted that although shape features used in the project perform reasonably well for the hat image retrieval yet it is also dependent on the pre-classification of the hat database. The hat database is classified into 10 categories according to their usage. This type of classification divides the database into different classes that contain a variety of hats. The within class variation in color and texture of hats makes it harder to design an image retrieval system. Moreover, it is pertinent to mention here that the concept of similarity between the images is highly subjective and user dependent. Different users might be looking and searching the database with a different purpose and perspective. To satisfy the needs of different users, it is required to incorporate their input into the system in some way.

6.2 Future Work

It is very hard to design a system that can be used by a variety of users for different purposes. In order to capture the users' ideas of image similarity, Relevance Feedback (RF) can be implemented in the retrieval system. Several researchers have implemented the idea of RF to improve the retrieval results. The basic idea is to get the feedback from the user by displaying images in the first round. The users are then provided with a user interface to mark the images as relevant or irrelevant. This input can then be used in a variety of ways to improve the results in the next round of searches. The process can be repeated until there is no more improvement in results or the user is satisfied with the results. Some of the ways in which the user feedback can be used are:

- Selection of features based on user input
- Changing the weights corresponding to the features and putting higher weights to the more relevant features.
- Computing and using new set of features
- Modifying the query

In this project, RF has not been used. This work may be extended to include the idea of RF to improve the efficiency of the system. Another possible extension of the work is to experiment with different types of similarity measures such as Kullback Divergence.

Bibliography

[1] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," IEEE Trans. on Sys. Man. and Cyb. SMC-3(6), 1973.

[2] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," IEEE Trans. on Sys., Man. and Cyb. SMC-8(6), 1978.

[3] J. R. Smith and Shih-Fu Chang, "Frequency and spatially adaptive wavelet packets," Proc. International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 2233-2236, 1995.

[4] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Qian Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele and P. Yanker, "Query by image and video content: the QBIC system," Computer, vol. 28, no.9, pp. 23-32, Sep 1995.

[5] Y. Rui, T. S. Huang and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," Proc. International Conference on Image Processing, vol.2, pp. 815-818, Oct 1997.

[6] B. Zhang, C. I. Tomai and A. Zhang, "An adaptive texture image retrieval system using wavelets," Proc. 7th International Conference on Control, Automation, Robotics and Vision, vol.3, pp. 1210-1215, Dec. 2002.

[7] A. V. Nevel, "Texture synthesis via matching first and second order statistics of a wavelet frame decomposition," Proc. IEEE International Conference on Image Processing, 1998.

[8] H. Yuan, X. P. Zhang and Ling Guan, "A statistical approach for image feature extraction in the wavelet domain," Proc. IEEE CCECE vol. 2, pp. 1159-1162, May 2003.

[9] H. Yuan and X. P. Zhang, "Texture image retrieval based on a Gaussian mixture model and similarity measure using a Kullback divergence," Proc. IEEE International Conference on Multimedia and Expo, vol. 3, pp. 1867-1870, June 2004.

[10] Z. Liu and S. Wada, "Translation, rotation and scale invariant texture characterization method for retrieval," Proc. 4th IEEE International Symposium on Signal Processing and Information Technology, Dec. 2004.

[11] M. Kokare, P. K. Biswas and B. N. Chatterji, "Texture image retrieval using new rotated complex wavelet filters," IEEE Transactions on Systems, Man and Cybernetics, Part B, vol. 35, no. 6, pp. 1168-1178, Dec. 2005.

[12] R. Brunelli and O. Mich, "Histograms analysis for image retrieval," Pattern Recognition vol. 34, Issue 8, pp. 1625-1637, August 2001.

[13] A. Rao, R. K. Srihari and Z. Zhang, "Spatial color histograms for content-based image retrieval," Proc. 11th IEEE International Conference on Tools with Artificial Intelligence, pp. 183-186, Nov. 1999.

[14] T. Mostafa, H. M. Abbas and A. A. Wahdan, "On the use of hierarchical color moments for image indexing and retrieval," Proc. IEEE International Conference on Systems, Man and Cybernetics, vol. 7, Oct. 2002.

[15] J. R. and Shih-Fu Chang, "Single color extraction and image query," Proc. International Conference on Image Processing, vol. 3, pp. 528-531, Oct. 1995.

[16] D. Androutsos, A. N. Venetsanopoulos and K. N. Plataniotis, "Efficient color image indexing and retrieval using a vector-based scheme," Proc. IEEE Second Workshop on Multimedia Signal Processing, pp. 15-20, Dec. 1998.

[17] C. C. Yang and M. C. Chan, "Color image retrieval based on textural and chromatic features," Proc. IEEE International Conference on Systems, Man and Cybernetics, vol.4, pp.922-927, 1999.

[18] J. Han and Kai-Kuang Ma, "Fuzzy color histogram and its use in color image retrieval," IEEE Transactions on Image Processing, vol. 11, no. 8, pp. 944-952, Aug 2002.

[19] Z. M. Lu and H. Burkhardt, "Colour image retrieval based on DCT-domain vector quantisation index histograms," Electronics Letters, vol. 41, no. 17, pp. 29-30, August 2005.

[20] F. Mokhtarian, "Silhouette-based isolated object recognition through curvature scale space," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, Issue 5, pp. 539-544, May 1995.

[21] D. Zhang and G. Lu, "Content-based shape retrieval using different shape descriptors: a comparative study," Proc. IEEE International Conference Multimedia and Expo, 2001.

[22] C. Zahn and R. Roskies, "Fourier descriptors for plane closed curves," IEEE Transations on Computer, vol. C-21, no. 3, pp. 269-281, March 1972.

[23] P. Parent and S.W. Zucker, "Trace inference, curvature consistency and curve detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 823-839, Aug. 1989.

[24] C. Y. Choo and H. Freeman, "An efficient technique for compressing chain-coded line drawing images," Proc. Twenty-Sixth Asilomar Conference on Signals, Systems and Computers, vol. 2, pp. 717-720, Oct. 1992.

[25] D. Zhang and G. Lu, "Generic Fourier descriptor for shape-based image retrieval," Proc. IEEE International Conference on Multimedia and Expo, vol. 1, pp. 425-428, 2002.

[26] B. Guo, J. Jiang, "A modified shape descriptor in wavelets compressed domain," Proc. International Conference on Image Processing, vol.1, pp. I-936-I-939, 2002.

[27] R. S. Torres, E. M. Picado, A. X. Falcao and L. F. Costa, "Effective image retrieval by shape saliences," Proc. XVI Brazilian Symposium on Computer Graphics and Image Processing, pp. 167-174, Oct. 2003.

[28] O. El-Badawy and M. Kamel, "Shape retrieval using concavity trees," Proc. 17th International Conference on Pattern Recognition, vol. 3, pp. 111-114, Aug. 2004.

[29] A. D'Souza, "Using EM to estimate a probablity density with a mixture of Gaussians," *http://www-clmc.usc.edu/ adsouza/notes/mix_gauss.pdf*

[30] J. Canny, "Canny; A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, Nov. 1986.

[31] G. Gagaudakis and P. L. Rosin, "Shape measures for image retrieval," Proc. International Conference on Image Processing, vol. 2, pp. 757-760, Oct. 2001.

[32] R. L. Cosgriff, "Identification of Shapes," Rep. 820-11, ASTIA AD 254 792, Ohio State University Research Foundation, Columbus, Ohio, Dec. 1960.

[33] R. C. Gonzalez and R. C. Woods, "Digztal Image Processing," Reading, MA: Addison-Wesley, 1992.