CHAPTER 2
LITERATURE SURVEY

2.0 INTRODUCTION

The overall objective and scope of the research work proposed in this thesis with a brief introduction has been discussed in the previous chapter. A review of literature is presented in this chapter.

In CBIR, quite often the images that are confronted by one contain too many objects or regions that may not be important. If one can rank different image regions on the basis of the quality of attribute, then the attention can be focused only on those perceptually important regions that have high attribute saliency. Also, this technique can be employed in visual tracking, surveillance, and indexing image databases. Hence the separation or segmentation of the objects or regions based on some homogeneity criteria from the background becomes mandatory for image retrieval based on their contents. The analysis of the image properties enable to develop techniques to classify the images based on its homogeneity criteria. In general the features based on color, shape and texture contributes to the perceptive and analysis of images. The following sections present the survey of the literature made on the extraction of these features, image texture segmentation, CBIR and its application.

2.1 COLOR DESCRIPTOR

The various color descriptors used are color space, representative colors, color histogram, color-structure histogram and color layout. This descriptor defines the color space to be used in a certain application, usually in combination with other descriptors such as dominant color or color histogram. A color space is a method by which one can
specify, create and visualize color. As humans, one may define a color by its attributes of brightness, hue and colorfulness. A computer will describe a color stimulus in terms of the excitations of red, green and blue phosphors on the CRT face plate. A printing press describes a color stimulus in terms of the reflectance and absorbance of cyan, magenta, yellow and black inks on the paper. Using three coordinates, or attributes, which represent its position within a specific color space, usually specifies such a color. These coordinates do not tell us what the color looks like, only where the color is located within a particular color space. Some of the commonly used color spaces are: RGB (Red Green Blue), CMY (Cyan Magenta Yellow), HSL (Hue Saturation and Lightness), YIQ, YUV, YC_{65}, YCC (Luminance - Chrominance), CIELuv and CIELab. The color histogram descriptor is a compound descriptor that expresses the color features by means of a histogram. The color histogram is computed by distributing the color pixel values into color bins. Calculating the distance between the normalized histograms does the matching. The best match is the one with the lowest weighted difference between the corresponding color histogram bins. Representative color is best suitable for representing local (object or image region) features where a small number of colors are enough to characterize the color information in the region of interest. Whole images are also applicable, for example, flag images or color trademark images. Color quantization is used to extract a small number of representative colors in each region or image. The percentage of each quantized color in the region is calculated correspondingly. A confidence measure on the entire descriptor is also defined, and is used in similarity retrieval. B.S. Manjunath et al. (B.S. Manjunath, 2001) made use of the Color Structure Histogram as a color feature descriptor for still image retrieval, i.e. the main functionality
is image-to-image matching. The color-structure histogram descriptor embeds local color structure information into the histogram; that is, the extraction method takes into account the colors in a local neighborhood of pixels, instead of considering each pixel separately. The color-structure histogram descriptor provides additional functionality and improved similarity based retrieval performance compared to the regular color histogram descriptor for natural images.


Also the content of the image can be represented by the characteristic of its dominant color. A compact color descriptor and an efficient indexing method based on this color descriptor are used for the retrieval of images as experimented by Y. Deng et al (Y. Deng, 2001).

2.2 SHAPE DESCRIPTOR

Loncaric et al (Loncaric 1998) illustrated several shape descriptors, which have been widely adopted for CBIR: Fourier descriptors (FD), curvature scale space (CSS) descriptors (CSSD), Zernike moment descriptors (ZMD) and grid descriptors (GD). There are generally two types of shape representations, i.e., contour-based and region based. Contour-based methods need extraction of boundary information, which in some cases may not be available. Region-based methods, however, do not necessarily rely on shape boundary information, but they do not reflect local features of a shape. Therefore, for generic purposes, both types of shape representations are necessary. FD and CSSD are contour-based, while ZMD and GD are region-based.

2.2.1 Fourier descriptors


Shape boundary is a set of coordinates \((x_i, y_i), i = 1, 2, \ldots, L\), which are extracted out in the preprocessing stage by contour tracing technique. The distance between the boundary points and the centroid \((x_c, y_c)\) of the shape was expressed by:

\[
r_i = \left( (x_i - x_c)^2 + (y_i - y_c)^2 \right)^{1/2}, \quad i = 1, 2, \ldots, L
\]

where \(x_c, y_c\) are averages of \(x\) coordinates and \(y\) coordinates respectively. Due to the subtraction of centroid (which represents the position of the shape) from boundary coordinates, the centroid distance representation is invariant to shape translation. Before applying Fourier transform, all the shapes in database are normalized to the same number of boundary points. For the shape signature described above, assuming it is normalized to \(N\) points in the normalization stage, the discrete Fourier transform of \(r_i, i = 0, 1, \ldots, N-1\) is then given by
The coefficients $u_n$, $n = 0, 1, \ldots, N-1$, are usually called Fourier descriptors (FD) of the shape, denoted as $FD_n$, $n = 0, 1, \ldots, N-1$.

### 2.2.2 CSS descriptors

F. Mokhtarian et al (F. Mokhtarian, 1986) constructed CSS descriptors, which are essentially the descriptors of key local shape features. By dealing shape in scale space, not only the locations of, but also the degree of convexities (or concavities) of shape boundaries are detected. Since curvature is a very important local measure of how fast a planar contour is turning, therefore, curvature scale space is exploited. The CSS descriptors are obtained by first calculating the CSS contour map, the map is a multi-scale organization of the inflection points (or curvature zero-crossing points).

In calculating CSS contour map, curvature is derived from shape boundary points $(x_i, y_i)$, $i = 1, 2, \ldots, L$:

$$k_i = \frac{(\ddot{x}_i \dot{y}_i - \dot{x}_i \ddot{y}_i)}{(\dot{x}_i^2 + \dot{y}_i^2)^{3/2}}$$

where $\dot{x}_i, \ddot{x}_i$ and $\dot{y}_i, \ddot{y}_i$, are the first and second order derivatives at location $i$ respectively. Curvature zero-cross points are then located in the shape boundary. The shape is then evolved into next scale by applying Gaussian smooth:

$$x'_i = x_i \otimes g(i, \sigma), \quad y'_i = y_i \otimes g(i, \sigma)$$

where $\otimes$ means convolution, and $g(i, \sigma)$ is Gaussian function. As $\sigma$ increases, the evolving shape becomes smoother and smoother. New curvature zero-crossing points are located at the new scale. This process continues until no curvature zero-crossing points are found. The final CSS contour map is composed of all zero crossing points $z_c(i, \sigma)$,
where \( i \) is the location and \( \sigma \) is the scale at which the \( z_c \) is obtained. The peaks, or the maxima of the CSS contour map (only those peaks higher than the threshold are considered) are then extracted out and sorted in descending order as CSS descriptors to index the shape.

### 2.2.3 Zernike moments

*M. R. Teague* (M. R. Teague, 1980) has proposed the use of orthogonal moments to recover the image from moments based on the theory of orthogonal polynomials, and has introduced Zernike moments, which allow independent moment invariants to be constructed to an arbitrarily high order. The complex Zernike moments are derived from Zernike polynomials:

\[
V_{nm}(x, y) = V_{nm}(\rho \cos \theta, \rho \sin \theta) = R_{nm}(\rho) \exp(jm\theta)
\]

and

\[
R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s! \left( n + \frac{|m|}{2} - s \right) ! \left( n - \frac{|m|}{2} - s \right) !} \rho^{n-2s}
\]

where \( n \) and \( m \) are subject to \( n-|m| = \text{even}, |m| \triangleq n \). Zernike polynomials are a complete set of complex-valued function orthogonal over the unit disk, i.e., \( x^2 + y^2 = 1 \). Then the complex Zernike moments of order \( n \) with repetition \( m \) are defined as:

\[
A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V'_{nm}(x, y), \quad x^2 + y^2 \leq 1
\]

### 2.2.4 Grid descriptors

In grid shape representation as proposed by *G. J. Lu et al* (G. J. Lu, 1999) a shape is projected onto a grid of fixed size, for example 16×16 grid cells. If the shape covers the grid cells (or covered beyond a threshold) then they are assigned with the value of 1 else
A shape number consisting of a binary sequence is created by scanning the grid in left-right and top-bottom order, and this binary sequence is used as shape descriptors to index the shape. For two shapes to be comparable using grid descriptors, several normalization processes have to be done to achieve scale, rotation and translation invariance. It begins with finding out the major axis, i.e., the line joining the two farthest points on the boundary. Rotation normalization is achieved by turning the shape so that the major axis is parallel with $x$-axis. To avoid multi normalization results for mirrored shape and flipped shape, the centroid of the rotated shape may be restricted to the lower-left part, or a mirror and a flip operation on the shape number are applied in the matching stage. Scale normalization can be done by resizing the shape so that the length of the major axis is equal to the grid width, and by shifting the shape to the upper-left of the grid, the representation is translation invariant. The distance between two sets of grid descriptors is simply the number of elements having different values. For example, the grid descriptors for the two shapes in Fig 2.1.a and Fig 2.1.b, are $001111000\ 01111111\ 11111111\ 11110011\ 00110011$ and $00110000\ 01110000\ 11110000\ 11110000\ 01111100\ 00011100$ respectively, and the distance between the two shapes will be 27.

Fig 2.1 Grid representation of shape
2.2.5 Hough Transform

*D. H. Ballard* (D. H. Ballard, 1981) proposed the Hough transform technique, to isolate features of a particular shape within an image. Because it requires that the desired features which are specified in some parametric form, the classical Hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, etc. The main advantage of the Hough transform technique is that it is tolerant of gaps in feature boundary descriptions and is relatively unaffected by image noise.

Hough technique is particularly useful for computing a global description of a feature where the number of solution classes need not be known a priori, given (possibly noisy) local measurements. The motivating idea behind the Hough technique for line detection is that each input measurement (*e.g.* coordinate point) indicates its contribution to a globally consistent solution (*e.g.* the physical line which gave rise to that image point).

A line segment can be described analytically in a number of forms. However, a convenient equation for describing a set of lines uses parametric or normal notion:

\[ x \cos \theta + y \sin \theta = r \]

where \( r \) is the length of a normal from the origin to this line and \( \theta \) is the orientation of \( r \) with respect to the X-axis. For any point \((x,y)\) on this line, \( r \) and \( \theta \) are constant.

In an image analysis context, the coordinates of the point(s) of edge segments in the image are known and therefore serve as constants in the parametric line equation,
while $r$ and $\theta$ are the unknown variables sought. The plot of the possible $(r, \theta)$ values defined by each $(x_i, y_i)$, points in Cartesian image space maps to curves (i.e. sinusoids) in the polar Hough parameter space. This point-to-curve transformation is the Hough transformation for straight lines. When viewed in Hough parameter space, points which are collinear in the Cartesian image space become readily apparent as they yield curves which intersect at a common $(r, \theta)$ point.

The transform is implemented by quantizing the Hough parameter space into finite intervals or accumulator cells. As the algorithm runs, each $(x_i, y_i)$ is transformed into a discretized $(r, \theta)$ curve and the accumulator cells which lie along this curve are incremented. Resulting peaks in the accumulator array represent strong evidence that a corresponding straight line exists in the image. We can use this same procedure to detect other features with analytical descriptions. For instance, in the case of circles, the parametric equation is

$$(x - a)^2 + (y - b)^2 = r^2$$

where $a$ and $b$ are the coordinates of the center of the circle and $r$ is the radius. In this case, the computational complexity of the algorithm begins to increase for three coordinates in the parameter space and a 3-D accumulator. In general, the computation and the size of the accumulator array increase in polynomial manner with the number of parameters. Thus, the modeling of the parameter space with the basic Hough technique described here is only practical for simple curves. In this thesis, the Generalized Hough Transform is used to describe the shape features with reference to template images with different representative shapes. Next a brief review related to various approaches for characterizing textures is presented below.
2.3 TEXTURE ANALYSIS

Ranges of texture analysis techniques that have been proposed by various researchers are available in the literatures; few of them are described in brief in this section. Number of researchers has developed their own techniques for texture analysis. They are categorized here as geometric and topological approaches, second or higher order statistics based approaches, texture with masks and logical operators, texture with stochastic models or random walk, texture based on gradient information, texture based on spectral filters, and other methods.

2.3.1 Geometric and Topological Approaches

P. Kundu et al (P. Kundu, 1993) proposed the use of fuzzy geometric features for texture classification. In their approach, first a set of 2D local membership value extrema is detected for the image. These are used as seed regions and are grown till they do not touch other seed regions. The resulting regions are called regions of influence. A number of features are then calculated from these regions including fuzzy area, perimeter, compactness, height, and width that form the basis of texture classification. On a total of 16 images taken from Brodatz album of size 128x128 pixels, a recognition rate of 90% correct classification was obtained.

2.3.2 Second or Higher Order Statistics

R.M Haralick (R. M. Haralick, 1973) proposed a novel technique for texture image classification. This study is concerned with the task of developing a set of features for classifying or categorizing pictorial data. Texture is chosen as the most suitable feature to represent images. Texture contains important information about the structural arrangement of units and their relationship to the surrounding environment. This
discriminatory information can be used to classify images. The authors present a general procedure for extracting textural properties of blocks of image data. These features are calculated in the spatial domain and it is assumed that the texture information in an image 'I' is contained in the overall or average spatial relationship that the gray-tones in the image have with one another. A set of gray-tone spatial-dependence probability distribution matrices is computed and a set of 14 textural features, which can be extracted from each of these matrices, is suggested. The matrices are constructed by assuming that every pixel, except the peripheral ones, has eight nearest neighbors (horizontally, vertically and diagonally at 45 degrees). It is also assumed that the matrix of relative frequencies adequately specifies the texture context information.

Some of these textural feature measures, obtained from the matrices relate to specific textural characteristics of the image such as homogeneity, contrast and the presence of organized structure within the image. Other measures characterize the complexity of the gray-tone transitions, which occur in the image.

The usefulness of textural features for categorizing images has been tested on three sets of images. In the first instance, photomicrographs of sandstones that are important in the petroleum industry were categorized. The data set, which consisted of 243 images, was divided into five classes. The set was divided into training and test data. A set of 8 variables comprising the mean and variance of the textural features obtained from the matrices was used for classification. The classification result yielded 89% accuracy. In the second instance, a set of aerial photographs was classified using the min-max decision rule. The data set consisted of 170 images divided into 8 categories. Four gray tone spatial dependencies and 11 textural features were used in the classification that
resulted in 82.3% correct classification. In the third instance, a set of satellite images was classified. The 624 samples in the data set were divided into training and test sets of equal sizes. The mean and variance of the four textural features and eight spectral features were used for classification. The result was 83.5% correct classification as compared to the 74 to 77% correct classification obtained by using spectral features. Since this seminal study, the features suggested by Haralick et al have been used in most of texture studies.

L.S Davis et al, (L.S.Davis, 1979 and 1981) describe the generalized co-occurrence matrices for texture discrimination. These do not describe texture directly but rather describe the spatial arrangement of local image features such as edges and lines. The description of Generalized Co-occurrence Matrices (GCM) is based on three attributes: image feature prototype, spatial predicate, and prototype attribute. The prototype edge-pixel can be defined as an ordered triple with three attributes (location, orientation and contrast). Spatial predicate is a mapping from image feature pairs into {true, false} category. The authors compare three prototypes, namely pixel intensity, edge pixel, and extended edge. For each of the three categories, their spatial predicates are defined. The following features are extracted from GCM: contrast, uniformity, entropy, and correlation. In their first study, classification experiments are performed on 30 texture samples. They find that compared to the co-occurrence matrix approach, their method performs much better. They used nearest neighbor classifier, Linear Discriminant Analysis, and the computation of inter-class distances using Bhattacharya method. 128 images of natural texture containing 64x64 pixels each are used for experiments. For the nearest neighbors’ analysis, feature pairings of size 2 are used and for the linear classifier
all features are used. The best nearest neighbor result of 61% correct classification is obtained using the contrast and entropy pair of features on edge-pixel prototype.

*R.F.Walker et al* (R.F. Walker, 1995) proposed an interesting method of improving the quality of co-occurrence matrix features. They classify the features proposed by Haralick and colleagues as being weighted on either the matrix element's value or its spatial location. For example, energy and entropy measures are weighted on the basis of value, and inverse difference moment, shade, inertia, correlation and variance are weighted on the basis of spatial location. The authors propose that it is best to suppress those elements of the matrix that yield little to the discrimination ability. Hence, on the basis of Bhattacharya distance calculation, one can find which elements are the most discriminatory. A discrimination matrix containing these weights can be multiplied with the original matrix to yield a better representation of values that are discriminatory. From these values one can either compute traditional measures or features as weighted sum of elements. Based on new calculations, the authors find that on six out of eight measures, much better performances are observed on a cell nuclei classification task with some improvements as much as 70% greater accuracy.

*A.Al-Janobi* (A.Al-Janobi, 2001) presented a new texture analysis method called Cross Diagonal Texture Matrix (CDTM). This method incorporates the properties of gray level co-occurrence matrices and texture spectrum methods that are both explained. The proposed method is based on characterizing the texture information of an image by separating the eight neighboring pixels around a central pixel in a neighborhood of 3x3 pixels. The eight elements in the texture unit are divided into two groups: the Diagonal Texture Unit (DTU) and Cross Texture Unit (CTU). The members of each unit have a
value of 0, 1 or 2 depending on whether the pixel in that position is less, equal or greater than the central pixel. These two units are combined into a CDTM by taking CTU as x-axis and DTU as y-axis of the matrix. From these units the Haralick's features are extracted for texture information. This method has the advantage that the gray level of the image has no effect on the size of the matrix. In addition, the computational complexity is reduced because of the reduced size of the matrix. Nine images from the Brodatz album have been used for the evaluation of this method. Statistics of order greater than two have also been applied for texture analysis.

2.3.3 Texture with Masks and Logical Operators

M. Unser et al (M.Unser, 1989) developed local linear transforms for texture measurement. They proposed simple and small convolution masks in combination with the computation of the local variance using a moving window in the image. Instead of using different filters, they proposed using four 2x2 Hadamard masks. The first of these masks has equal elements and measures the magnitude. The other three masks have two elements equal to 1 and other two elements equal to -1, which are used to approximate the derivatives in horizontal, vertical and diagonal directions.

V. Manian et al (V. Manian, 2000) presented a new algorithm for texture classification based on logical operators. Operators built from logical building blocks are convolved with texture images. The logical operators are based on order-2 elementary matrices whose building blocks are symbols 0, 1, -1, and matrices of order 1x1. These low order matrices can be operated on using the following operators: row-wise join, and column-wise join. The study selects a total of six operators on the basis of their best
performance. The six operator masks are first convolved with the texture regions and the response is used to compute a standard deviation matrix using a sliding window.

Features are next computed by zonal-filtering using zonal-masks, and are normalized and a feature selection scheme based on distances between feature means and measure of standard deviation is used to find the best features for classification. A total of 39 textures from the Brodatz album (P. Brodatz, 1996) are used in the classification study. On majority of the images, the Euclidean and nearest neighbor classifiers perform the best showing between 90 and 100% accuracy. On mosaic of six Brodatz textures, the average results on comparisons of this technique are as follows: Logical operators (93%), co-occurrence matrix method (70%), Fourier power spectrum method (59%), tree-structured wavelet transform method (61%), Law’s texture features (61%) and Gabor features (63%).

In the paper by P. Kruizinga et al (P. Kruizinga, 1998, 1999) the comparison of the well-known texture operators, the co-occurrence matrix operator and the Gabor-energy operator, with a new biologically motivated nonlinear texture operator, the grating cell operator, is made to evaluate the ability of the operators to detect texture and to separate texture and form information.

### 2.3.4 Texture Similarity and Salience

Some studies are primarily interested in studying texture for determining texture features that is not ultimately needed for classification. For example, the interest is in knowing whether an image has significant amount of texture, and if so, its orientation, location, etc. This can be used for focusing attention at specific parts of the image. Texture based CBIR is also mostly based on matching object textures where nearest
neighbor techniques can be used as proposed by A. Farago et al (A. Farago, 1993). Some of the studies in this area are reviewed in this section.

*F. A. Sadjadi* (F. A. Sadjadi, 1980) presented a comparison of four separability measures for registration of two-dimensional digital images. Image registration is the problem of matching an image denoted as a reference with a usually much broader and differently obtained picture. The main objective of the experiment was to decide which of these image views would perform better when used as a reference in a scene-matching problem. The four separability measures used were Bayes probability of error, Chernoff bound, Bhattacharyya bound and Fisher's criteria. Four sets of image data of 128x128 pixel size and 64 gray levels were used in their experiment. The reference image in each set was selected such that the target area was totally included in it. A matching area of 9 pixels was chosen for the estimation of the correct match statistics. Correlation functions for area, edge, and variations of mean were computed and then separability measures were derived and a comparison of the resulting error probabilities was made. The experimental results show that for the real images, the target looking view performs better, for the synthetic images the down-looking view performs better. The comparisons of the various probabilities of error show that Fisher's criterion produces the highest probability of error.

*M. N. Do et al*, (M. N. Do, 2002) proposed a statistical model for images, the vector Wavelet Domain Hidden Markov Model (WD-HMM), is formed to capture both the subband marginal distributions and the dependencies of wavelet coefficients across scales and orientations. By applying this vector to the steerable pyramid, a steerable model that can be diagonalized to become rotation invariant is obtained. To facilitate the
use of WD-HMMs in the image retrieval application, a fast algorithm to approximate the Kullback–Leibler distance between two WD-HMMs is designed showing better performance compared to the independent sub band model.

*B. B. Chaudhuri et al* (B. B. Chaudhuri, 1993) defined a technique based on Hough transform for calculating texture orientation. In this simple method, an edge image is formed first using Laplacian of Gaussian approach. An orientation histogram of dominant local orientation is constructed from the edge image. The algorithm next detects the peaks and valleys of the histogram. The measure of texture orientation corresponds to the height and width of peaks. The authors show results on 16 texture images from Brodatz album of size 128x128 pixels.

Some of the other most popular texture feature extraction methods are based on the gray level co-occurrence statistics (R.M.Haralick, 1979, C.C.Gotlieb, 1990), texton gradients, edge gradients (D. Marr, 1980), filtering methods like morphological filters, Fourier filters, Random Field Models (R. Chellappa, 1985), Gabor filters (I. Fogel, 1989), (S. E. Grigorescu, 2002) Wavelet Packet approaches (T. Chang, 1993; A. Laine, 1993), Wavelet Frames (M.Unser, 1995), Wavelets like Gaussian (M.Cheriet, 1998; D. Charalampidis, 2002), fractal dimension (L M. Kapalan, 1999), and Local binary patterns (T. Ojala, 2002). Each method is superior in discriminating the texture of its characteristics and there is no unique method available for detecting all textures. In the study made by *T. Randen, et al* (T.Randen, 1999) they observed that the particular texture classification technique is restricted to a limited real texture. Every texture has a band of frequency components in it where the selection of filter banks for extracting the features becomes a non-trivial task.
S. Kulkarni et al (S. Kulkarni, 2002) proposed and investigated an auto-associator texture feature extractor and two hybrid intelligent techniques such as an auto-associator-Multi Layer Perceptron (MLP), and statistical-MLP for texture feature extraction and classification. They showed that the auto-associator is capable of separating texture classes very well and without any feedback from the user. The feature extraction and classification techniques were tested on a large database of texture patterns namely the Brodatz texture database. The results obtained were analyzed and compared with other intelligent and conventional techniques.

2.4 IMAGE SEGMENTATION TECHNIQUES

In this section the studies that are based on finding object/regions in the images are reviewed along with those studies that deal with color segmentation to highlight how they have been used for CBIR. Image segmentation has been approached from a wide variety of perspectives. In summary the segmentation is based on histogram thresholding, edges, tree/graph, region growing and clustering the pixels based on some homogeneity criteria. One of the segmentation techniques is the clustering of the pixels providing the number of categories available on the images based on the knowledge available on them such as the type of features. An adaptive clustering algorithm for image segmentation is proposed by T. N. Pappas (T.N. Pappas, 1992), as a generalization of the K-means clustering algorithm to include the spatial constraints and to account for the local intensity variations in the image.
2.4.1 Histogram Thresholding

R. Ohlander et al (R. Ohlander, 1978) proposed a thresholding technique that is very useful on segmenting outdoor color images. This is based on constructing color and hue histograms. The picture is thresholded at its most clearly separated peak. The process iterates for each segmented part of the image until no separate peaks are found in any of the histograms. The criterion to separate peaks was based on the ratio of peak maximum to peak minimum to be greater than or equal to two.

In a number of applications, histogram threshold is not possible simply because the histogram may be unimodal. In some cases, the images may be of such quality that any preprocessing may not improve the contrast between objects sufficiently and hence one may not achieve two or more peaks in the histogram for selecting thresholds for segmentation. Unimodal distributions are typically obtained when the image consists of mostly of a large background area with small, but significant regions. This often happens in medical imaging applications. Similarly in aerial scenes with many different objects, the histogram may only have one peak because of the vast range of intensities for each object and an overlap between these.

2.4.2 Tree/Graph Based Approaches

K. Cho et al (K. Cho, 1997) proposed an approach for segmentation, which is derived from the consensus of a set of different segmentation outputs on one input image. Instead of statistics characterizing the spatial structure of the local neighborhood of a pixel, for every pair of adjacent pixels their collected statistics are used for determining local homogeneity. Several initial segmentations are derived from the same input image by changing the probabilistic component of the hierarchical Region Adjacency Graph
(RAG) pyramid based technique. From the ensemble of these initial segmentations, for every adjacent pixel pair a co-occurrence probability is derived, which captures global information (about the image) at the local level (pixel level). The final segmentation of the input image is obtained by processing the co-occurrence probability field with the same RAG pyramid technique. The pixel pairs with high co-occurrence probability are then grouped together based on the consensus about local homogeneity. This technique can also be used to extract the high confidence homogeneous regions from the co-occurrence probability field. Bayesian networks were then used to extract features from the images. The features extracted were variance of the width of the region, ratio of average width to length and the average gray level. Then post-processing of over-segmented images is done based upon a priori information about the sought features. The RAG of the final segmentation provides the spatial relationship between regions and can be used for further interactive analysis of the image. This segmentation method is completely unsupervised. Experiments were performed on an aerial image, and images of a boat, a pentagon, and a house.

2.4.3 Region Growing

A range of image segmentation algorithms has been presented based on region growing. In this section some of the relevant studies that have used region-growing algorithms are reviewed. Region growing algorithms take one or more pixels, called seeds, and grow the regions around them based upon a certain homogeneity criteria. If the adjoining pixels are similar with some criteria to the seed, they are merged with them within a single region. The process continues until all the pixels in the image are assigned to one or more regions.
Y. L. Chang et al (Y.L.Chang, 1994) proposed a region-growing framework for image segmentation. This process is guided by regional feature analysis and no parameter tuning or a priori knowledge about the image is required. The image is first divided into many small primitive regions that are assumed to be homogeneous. These primitive regions are then merged to form larger regions until no more merges are possible. Two regions are merged if they pass the homogeneity test and also if the value of the edge connecting them is weak. The focus of this study is on investigating how different merge criteria affect the quality of segmentation and the processing time. The experiments designed to evaluate the merge criteria are based on four important aspects of segmentation output: region merge ability, boundary accuracy, merge rejections, and number of iterations required. In these experiments, 300 images of size 50x50 pixels were used. Each image had two equal sized regions that share a simple straight 50-unit boundary. The data was randomly generated for the two regions from a pair of normal distributions having different means and variances. The best results using the fast merge method gives the correct classification rate of 86% (with less than 5 regions in the image). The authors concluded that the algorithm automatically computes segmentation thresholds based on local feature analysis. The algorithm is robust and produces high quality segmentation on a wide range of textured and gray scale images. This framework can also be easily adapted to different image applications by substituting the suitable features. The main limitation of this algorithm is however the limited applicability of the adaptive homogeneity tests on very small regions and order dependency of its segmentation results.
Segmentation based on watershed is another technique by which the homogeneous regions are separated out from the rest of the image by computing the gradient and applying the morphological operations. Some of these techniques are presented by P. T. Jackway et al (P. T. Jackway, 1996), and L. Shafarenko et al (L. Shafarenko, 1997).

2.4.4 Texture based Segmentation

M. Unser (M. Unser, 1995) used the wavelet transforms for characterizing the texture properties at multiple scales. His analysis used an over complete wavelet decomposition yielding a description that is translation invariant. He experimented with 12 Brodatz images and has shown that the discrete wavelet frame approach is superior to standard (critically sampled) wavelet transform feature extraction. Also he proved this method is performing well compared to more traditional single resolution techniques.

In the paper by D. Dunn et.al (D. Dunn, 1995) the techniques for segmenting the textured images by transforming texture differences into detectable filter-output discontinuities at texture boundaries using Gabor filter is given. They modeled the Gabor outputs as Rician random variables (often approximated well as Gaussian rv’s) and developed a decision-theoretic algorithm for selecting optimal filter parameters. To improve segmentations for difficult texture pairs, they proposed a multiple-filter segmentation scheme, motivated by the Rician model. Their experimental results indicated that this method is superior to previous methods in providing useful Gabor filters for a wide range of texture pairs. They considered eight-bit, 512 x 512 images drawn from Brodatz and synthetic textures. In the segmented output images, it exhibits errors, and the errors are typically confined to the vicinity of the texture boundary.
P. L. Palmer et al (P. L. Palmer, 1997) exploited the fact that a textured region has a high density of edge pixels associated with it to present an algorithm that can locate boundaries of the textured regions.

W. Y. Ma et al (W. Y. Ma, 2000) demonstrated a framework for detecting image boundaries based on Edge Flow technique, which utilizes a predictive coding scheme to detect the direction of change in various image attributes and construct an edge flow field. By propagating the edge flow vectors, the boundaries can be detected at image locations, which encounter two opposite directions of flow in the stable state. This approach requires very little parameter tuning. But the use of the texture feature increases the processing time significantly in performing image segmentation.

L. M. Kaplan et al (L. M. Kaplan, 1999) considered the Hurst parameter for two-dimensional (2-D) fractional Brownian motion (fBm), which provides a single number that completely characterizes isotropic textured surfaces, whose roughness is scale invariant. The extended self-similar processes that provide a generalization of fBm is described by a number of multi scale Hurst parameters. In contrast to the single Hurst parameter, the extended parameters are able to characterize a greater variety of natural textures where the roughness of these textures is not necessarily scale invariant. The effectiveness of multi scale Hurst parameters as features for texture classification and segmentation were studied. For texture classification, the performance of the generalized Hurst features is compared to traditional Hurst and Gabor features. Their experiments showed that classification accuracy for the generalized Hurst and Gabor features are comparable even though the generalized Hurst features lower the dimensionality by a factor of five.
Y. Deng et al in their paper (Y. Deng, 2001) have proposed a method for unsupervised segmentation of color-texture regions in images. The method they used consists of two independent steps: color quantization and spatial segmentation. In the first step, colors in the image are quantized to several representative classes that can be used to differentiate regions in the image. Their corresponding color class labels, thus forming a class-map of the image, are used to replace the image pixels. Applying the criterion to local windows in the class-map results in the image, in which high and low values correspond to possible boundaries and interiors of color-texture regions. A region growing method is then used to segment the image based on the multiscale images. An additional region-tracking scheme is embedded into the region growing process to achieve consistent segmentation and tracking results, even for scenes with non-rigid object motion. The limitations in their work are caused by the varying shades due to the illumination.

2.5 CONTENT BASED IMAGE RETRIEVAL (CBIR)

Typically, a CBIR system extracts visual features from a given query image, which are then used for comparison with the features of other images stored in the database. The similarity function is thus based on the abstracted image content rather than the image itself. One should note that given the ever-increasing volume of image data that is available, the approach of relying on human-assisted annotations as a means of image abstraction is not feasible. Many research articles are available in the literature on CBIR based on color, shape and texture and the following sections review a few of them.
2.5.1 CBIR using Binary Signatures

Significant research has been focused on determining efficient methodologies for retrieving images in large image databases as studied by M.A. Nascimento et al (M. A. Nascimento, 2002). This paper addresses the design and implementation of a new image abstraction technique based on compact signature bit-strings and an appropriate similarity metric. Performance evaluation of image retrieval on a large heterogeneous database of up to 50,000 images demonstrated that the proposed technique is able to outperform the use of Global Color Histograms by up to 55%, and that of Color Coherence Vectors by up to 20% in terms of retrieval effectiveness – this relative advantage was also observed when using classical Precision vs. Recall curves. Perhaps more important is the fact that the proposed approach saves 75% (87.5%) of storage space when compared to Global Color Histograms (Color Coherence Vectors).

2.5.2 CBIR using Local Color Histogram

Global color histograms are well known as a simple and often easy way to perform color-based image retrieval. However, it lacks spatial information about the image colors. The use of a grid of cells superimposed on the images and the use of local color histograms for each such cell improves retrieval in the sense that some notion of color location is taken into account. In such an approach however, retrieval becomes sensitive to image rotation and translation. J.Han et al (J. Han, 2002) has brought out an improved color histogram using fuzzy techniques for improving the retrieval efficiency. A. Mojsilovie et al (A. Mojsilovie, 2000) proposed a color histogram with dominant color features used along with a set of rules for the matching and retrieval of images.
2.5.3 CBIR using Shape Descriptors

*S. Brandt et al* (S. Brandt, 2000) made a study on the shape-describing features for general CBIR. They formed various types of statistical feature vectors from the edges in non-segmented images. The best results were obtained with decimated magnitude spectrum of the edge image indicating that both local and global information are important clues of the image shape. *S. Belongie et al* (S. Belongie, 2002) presented a scheme for matching the images by measuring the similarity between objects under shape context where the distribution of shape information is captured at reference points.

*J. Hsieh et al* (J. Hsieh, 2003) have presented a template-based method for retrieving images from picture libraries with more semantic understanding introducing two important features to represent the content of an image; that is, a set of templates and the spatial relations between them. Spatial relation information associated with the corresponding templates can help the retrieval system capture more high-level concept in retrieving images. But identification of the templates from the images makes the issue.

2.5.4 CBIR using Texture

*B. S. Manjunath et al* (B. S. Manjunath, 1996) proposed the retrieval of texture images based on Gabor filter and extended to the retrieval of satellite images and air photo. They also compared the performance of Gabor features with the conventional Pyramid Structured wavelet features, Tree structured wavelet transform features and the Multi resolution simultaneous autoregressive model features showing the robustness of the Gabor features.
2.5.5 CBIR for Facial Recognition

Among the face recognition techniques based on the constituent such as nose, mouth; the template matching techniques and Principal Components Analysis based methods are popular. In the template matching techniques, during the training phase a template or model for an object in the database is created. In the testing phase, the testing images are classified using that model created in the training phase.

Other techniques includes the usage of deformable template methods (A.L.Yuille, 1992) (K.M.Lam, 1996), Gabor wavelet based methods (H.Okada, 1998), Elastic graph matching techniques (L.Wiskott, 1997), fisher face technique based on Linear Discriminant Analysis (M.Kirby, 1990), Principal Component Analysis (PCA) based methods, proposed by M. Turk et al (M.Turk, 1991). A discriminant analysis method is proposed by T.Kim et al (T. Kim, 2005) to classify a nonlinear structure for face recognition by finding a set of local linear transformations so that the locally linearly transformed classes maximize the between-class covariance and minimize the within-class covariance in a single global space. X. He et al (X. He, 2005) have proposed the Laplacianfaces for face recognition and the performance is compared with PCA and Linear Discriminant Analysis (LDA) based methods.

Also S. Lawrence et al (S. Lawrence, 1996) proposed a convolutional neural-network approach for face recognition where multi layer neural architecture with supervised training is presented and its performance is compared with Back Propagation Network (BPN). Though the retrieval rate is very good and is fast compared with BPN, the training of the images is still slow.
A frontal face authentication algorithm based on morphological dynamic link architecture on facial grid is proposed by C. L. Kotropoulos et al (C. L. Kotropoulos, 2000) by determining the number of test client and impostor claims to give statistically significant results.

2.5.6 CBIR for Medical Applications

Recently, with the emergence of Picture Archiving and Communication Systems (PACS) (A. Marsh, 1997) (E. L Siegel, 1999), there has been an added interest to integrate all the information related to patients (texts, images, charts, temporal data, etc.) in unified systems. The huge amount of digital images generated in hospitals and health care centers leads to the need of automatic storage and retrieval of them. Therefore, a PACS should incorporate properties allowing retrieving these images in a timely manner. Moreover, in order to effectively aid the physicians in their analysis and diagnosis, the retrieval should bring images that match the criteria given by the specialists.

2.6 CONCLUSION

This chapter has dealt with the survey of the various techniques for extraction of the primitive features color, shape and texture. Though the color is the basic feature of any image, the shape and texture features mostly influence the perception of the image. Hence the shape and texture features extraction techniques are emphasized in this thesis. The existing techniques of CBIR are mainly based on the features collected on the whole image or the part of the image to obtain the signature for representing them using any one techniques in their own aspect. As such there is no work has been reported which combines the effect of the features using different methods. In this thesis an effort has been taken to exploit the effect of the texture features collected from combined moment
and Gabor wavelet based features, combined crude and orthogonal wavelet based features, spatial and spectral features. Also the texture features are collected from the regions bounded by the edge pixels obtained thorough optimal threshold rather than collecting features from the whole image or a block of pixels. Also the shape descriptors based on FD, CSS, CSSD, ZMD, GD and Hough Transform are discussed. Each shape representation is limited by its use. Hence the usage of shape features based on GHT with relevant template shapes with modified procedure is proposed in the thesis. Then the study on the retrieval of images using its content characterized by the primitive features such as color, shape and texture is made with a relevant application in medical field.

The principles and study of feature extraction methods for this thesis work is dealt in the next chapter.