CHAPTER 1

INTRODUCTION

1.1 GENERAL

Digital images are now the basis of visual information in medical applications. The advent of radiology which employs imaging for diagnosis generates a large number of images. Explosion of World Wide Web (WWW) in the last decade has seen an enormous increase in the usage of digital images and the ease of access to remotely stored databases. This exponential growth in the image stored in databases requires an efficient image indexing and retrieval system. The need to locate the desired image in a large and varied collection has led to design of numerous image retrieval systems. Traditional methods of image indexing are not adequate as the amount of images to be indexed is huge, which makes it impractical and error prone (Enser 1995). Automatic retrieval of images based on features like color, shape and texture is termed as Content Based Image Retrieval (CBIR). CBIR methodologies are similar to the methods used in image processing and computer vision. Image processing encompasses image enhancement, compression and interpretation whereas CBIR emphasizes on retrieval of image from a database in response to queries. CBIR automates indexing and analysis of images. Research on CBIR has gained momentum, as application prospective of CBIR for fast and efficient image retrieval is vast.
1.2 BACKGROUND

Manual keyword annotation and methods of text retrieval was the initial model used for image retrieval. This was feasible when the collection of images were small. However, this approach was not practical when the collections of images were large and a huge amount of manual effort was required, which was also time-consuming and inconsistent in naming keywords among different users. To overcome this problem a new approach called Content Based Image Retrieval (CBIR) was proposed (Zhu et al 2000). Kato (1992) was the first to use the term Content Based Image Retrieval in his experiments, to automatically retrieve images from database using color and shape feature. Swain and Ballard (1991) used simple low level features for retrieval instead of color histograms. Many CBIR systems were developed; IBM Query By Image Content (QBIC) retrieved images based on the color percentage and color layout present in the image. Early researches were focused on retrieval by measuring similarity between the query image and image features in the database (Flickner et al 1995). Later systems improved query formulation by incorporating relevance feedback. Rubner et al (2001) processed images in the database to extract low-level features and created m-dimensional feature vector, where the features in the query image were compared using similarity measures. Shorter the distance between two feature vectors, more the similarities between the two images.

1.3 APPLICATIONS

CBIR finds application in various domains (Gudivada et al 1995). Some of them are discussed in this section:

- **Crime prevention:** Law enforcement agencies maintain documentation of visual evidence, which includes photographs and fingerprints. The fundamental techniques for automatic
fingerprint matching were in use from 1980s, and systems based on this technology were used by police forces around the world. CBIR helps in searching an entire database to find the closest matching records. Face recognition is another widely sought application.

- **Military**: CBIR is widely used in the defense sector for the identification of enemy aircraft from radar screens, study of satellite photographs and the provision of guidance systems for cruise missiles. Many of the techniques used in crime prevention are also relevant to the military field.

- **Intellectual property**: Another prime application area of CBIR is the Trademark image registration and copyright protection, where a new mark/logo is compared with existing marks to make certain that there is no risk of confusion.

- **Journalism and advertising**: Newspapers maintain archives of still photographs to illustrate articles or advertising copy. These archives can often be extremely large. Similarly, broadcasting corporations deal with millions of hours of archive video footage, which makes retrieval impossible by annotating. Automatic measures are required for handling such a large collection of images.

- **Medical diagnosis**: The increasing dependence on diagnostic techniques such as radiology, computerized tomography in the medical field has resulted in a sudden increase in the number of images produced and significance of medical images. The CBIR techniques are used to aid diagnosis by comparing with similar past cases.
- **Cultural heritage:** Museums/art galleries deal essentially in visual objects and the capability to identify objects with aspects of visual similarity is used by researchers attempting to trace historical influences and by art lovers searching for paintings/sculptures suiting their taste.

- **Web searching:** CBIR can effectively locate text and images on the Web, and is now a source for information and entertainment. Expansion of the worldwide web (WWW) has led to rapid growth of text-based search engines. Difficulties in locating images on the Web (Jain 1995) reveal the need for image search tools of power. In contrast, it should be ensured that easy access is not possible with regard to pornographic images. Many systems for content-based image searching on the Web were demonstrated in the last few years.

### 1.4 IMAGE FEATURES

CBIR is based on the extraction of features from the image; various methods are developed to extract the features. The general features which are extracted from the image are as follows:

#### 1.4.1 Pixel Value

The simplest form of an image feature is pixel value. The most basic form of image retrieval approach is to compare query image pixel with database image pixel. Retrieval based only on the pixel value does not give good results because it is tough to identify the pixels which are to be used for comparing the two images.
1.4.2 Local Features

Local features refer to the small pixel blocks obtained by segmenting the image. Finer details are described in local features when compared to global features; thus in various domains local features give good classification results (Shyu et al 1998).

1.4.3 Global Features

Global features consider the whole image and most systems use global feature like color histogram, which gives percentage of color present in the whole image. The global features like color or shape provide an overall idea and not the details of image. Global features are advantageous as extraction and matching is done with high speed (Glatard et al 2004).

1.4.4 Low Level Features

The visual content of the image like color, texture, shape which can be extracted from the image are termed as low level features. Notable features that can be extracted for CBIR include:

1.4.4.1 Color

Color was the most widely used feature in CBIR (Stricker et al 1995). Instead of color model RGB, CBIR uses HSV (Hue, Saturation and Value) and HLS (Hue, Lightness and Saturation) color for measuring the color similarity between two images (Sural et al 2002). The color in the image is generally represented in color histograms. It is also represented in the form of color coherence vector or color correlogram.

CBIR methods for retrieving images based on color are done by comparing color similarity between the query image and image collection.
Each image in the database or collection is analyzed, and its corresponding color histogram is stored in the database. The color histogram represents the proportion of pixels of every color in the image. Input query can be either by specifying desired color proportion or by a query image for which color histogram is calculated. The system retrieves images whose color histogram matches those of the query most closely. The color retrieval procedure was first proposed by Swain and Ballard (Swain et al 1991), based on this, many variations of the color retrieval process have been developed successfully (Stricker et al 1995).

1.4.4.2 Texture

The texture is a natural property of surface and is characterized by repetition of patterns or patterns over a region in an image. Texture is extracted by comparing the color contrast of each pixel in a group of pixels or region. The texture is extracted either by spectral or statistical methods. The similarity between two textures is done by co-occurrence matrix (Haralick et al 1973).

Texture similarity comparison is useful for distinguishing distinctive physical composition of an image. The texture is compared using values known as second-order statistics calculated from query and stored image. The measure of image texture is given by contrast, coarseness, directionality and regularity (Tamura et al 1978). Input query is similar to the color retrieval procedure with the system retrieving images of similar texture value of the query. A texture thesaurus with automatically derived codewords representing main classes of texture in the database is created. Thus, the system using the thesaurus retrieves the images based on similarity (Wei-Ying Ma et al 1998).
1.4.4.3 Shape

Shape is the geometric information of an object present in the image, where the shape does not change even when location, scale or orientation of the object is changed (Zhang 2001). Segmentation of the image is done to obtain regions or objects. The shapes are either obtained by using boundary information with edge detection or region based using the whole shape region (Freixenet et al 2002).

Shape based retrieval ability is an urgent requirement. While shape, unlike texture, is fairly well-defined with evidence existing that natural objects are recognized by shape (Biederman 1987). Many features of object shape (but independent of size/orientation) are computed for identified objects in stored images and queries answered by computing query image features and retrieving stored images whose features match the query image closely.

1.4.5 High Level Image Features

High level features are based on the semantic understanding of the image and derived attributes are from low level features.

1.4.5.1 Semantic features

Semantic refers to the meaning of the image content. This is a high-level concept when compared to low-level visual features like color, texture and shape (Wang et al 2001). Computers find it difficult to understand and process semantic meaning of images, as visual features need not necessarily match perceptual semantics of images. To improve the accuracy of semantic retrieval, Relevance Feedback (RF) is used where the user evaluates the results (Jing et al 2004). RF is a repetitive process, by which the
performance of CBIR is considerably enhanced by adjusting the query using the user’s feedback on the retrieval results.

1.4.5.2 Low level feature extraction in frequency domain

Frequency domain analysis is commonly used for extracting features from a frequency domain for image recognition. The features are obtained from edge detection in the first step and in the next step DCT or wavelet transforms are used. The wavelet transforms is useful computational tool for image processing. The transforms compress the digital images and are also useful for ‘cleaning’ signals and images by reducing unwanted noise and blurring. This enhances the classification information. Fast Fourier Transforms, DCT, DST and wavelet analysis are discussed in this section.

1.4.5.2.1 Fourier discrete cosine transform

Like Fourier related transforms, Discrete Cosine Transform (DCT) states a function as a sum of sinusoids with different frequency and amplitudes and also operates on a function at a finite number of discrete data point. The Discrete Fourier Transform (DFT) uses both cosine and sine functions, whereas DCT uses only the cosine functions. A DCT expresses a sequence of finite several data points in terms of a sum of cosine functions oscillating at different frequencies. The use of cosine functions is more efficient as fewer functions are needed to approximate a signal. The DCT for N data item is defined in Equation (1.1) as following:

\[
F(u) = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{i=0}^{N-1} \Lambda(i) \cos \left[\frac{\pi u i}{2N} (2i + 1)\right] f(i)
\]

(1.1)

and the corresponding inverse DCT transform is \( F^{-1}(u) \).
The general equation for 2D (N By M image) DCT is defined in Equation (1.2),

\[ F(u, v) = \frac{1}{\sqrt{N}} \left[ \frac{1}{\sqrt{M}} \right] \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} \Lambda(i) \Lambda(j) \cos \left[ \frac{\pi u}{2N} (2i + 1) \right] \cos \left[ \frac{\pi v}{2M} (2j + 1) \right] f(i, j) \]  

(1.2)

and the corresponding inverse 2D DCT transform is \( F^{-1}(u, v) \).

where

\[ \Lambda(i) = \frac{1}{\sqrt{2}} \quad \text{for} \ i = 0 \quad \text{and} \ j = 0 \quad \text{otherwise} \]  

(1.3)

1.4.5.2.2 Fast Fourier Transform

The Discrete Fourier Transform (DFT) transforms the original function into frequency domain representation. The traditional DFT is computational expensive in the image transformation, Fast Fourier Transforms (FFT) improves the speed of the calculation. The signals transformed by FFT are composed of two parts which are magnitude spectrum and phase spectrum. The FFT gives the same solution as DFT but the number of operations required to compute the transforms is reduced.

1.4.5.2.3 Discrete sine transform

Discrete Sine Transform (DST) extracts a feature vector from every image using pixels which are one length away from each other. Discrete
Sine Transform and Discrete Fourier Transform (DFT) are similar the difference being the use of real numbers. The Discrete Sine Transform is represented in Equation (1.4),

\[
X_k = \sum_{n=0}^{N-1} x_n \sin \left[ \frac{\pi}{N} \left( n + \frac{1}{2} \right) (k + 1) \right]; \quad 0 \leq k \leq N-1
\]  (1.4)

where \( x_n \) is the original vector on \( N \) real numbers. DST operates on real data with odd symmetry and thus output data is shifted by half a sample but Discrete Sine Transform is preferred choice over Fast Fourier Transform because of simplicity and reduced computation time for medical image coefficients. The inverse of Discrete Sine Transform is given by the Equation (1.5).

\[
X_k = \frac{2}{N+1} \sum_{n=0}^{N-1} x_n \sin \left[ \frac{\pi}{N} \left( n + \frac{1}{2} \right) (k + 1) \right]; \quad 0 \leq k \leq N-1
\]  (1.5)

1.4.5.2.4 Discrete wavelet transform

The Discrete Wavelet Transform (DWT) was elaborated by Engelhart (Englehart 1998). These techniques have shown better performance than the others in this area because of its multilevel decomposition with variable trade-off in time and frequency resolution. The DWT is a transformation of the original temporal signal into a wavelet basis space. Performance of time-frequency wavelet representation is through repeated filtering of signal with a pair of filters which cut a frequency domain in two. Signals are decomposed into approximation and detail signals by DWT. The approximation signal is subsequently divided into new approximation and detail signals. This process is carried out iteratively producing a set of approximation signals at different detail levels (scales) and a final gross approximation of the signal.
The detail $D_j$ and the approximation $A_j$ at the level $j$ can be acquired by filtering the signal. An $L$-sample high-pass filter $g$, and an $L$-sample low-pass filter $h$ is used to filter the signal. Both approximation and detail signals are down sampled by a factor of two.

This can be expressed in Equations (1.6) and (1.7) as follows:

$$A_j[n] = H \left\langle A_{j-1}[n] \right\rangle \sum_{k=0}^{L} h[k] A_{j-1}[2n-k]$$

(1.6)

$$D_j[n] = G \left\langle D_{j-1}[n] \right\rangle \sum_{k=0}^{L} g[k] A_{j-1}[2n-k]$$

(1.7)

where $A_n[n], n = 0, 1, 2, ..., N - 1$ is the original temporal sequence, while $H$ and $G$ represent the convolution/down sampling operators. Sequences $g[n]$ and $h[n]$ are associated with wavelet function $\psi(t)$ and the scaling function $\varphi(t)$ through inner product is shown in Equations (1.8) and (1.9) as

$$g[n] = \langle \psi(t), \sqrt{2} \psi(2t-n) \rangle$$

(1.8)

$$h[n] = \langle \varphi(t), \sqrt{2} \varphi(2t-n) \rangle$$

(1.9)

where

$$\psi(t) = \frac{\sin(2\pi t) - \sin(\pi t)}{\pi t}$$

(1.10)

A feature extraction approach based on DWT applied by Englehart (1998) consists of four different phases:
1. Perform full DWT decomposition of the EMG signals, until scale \( j = \log_2 (N) \), with the Coiflet wavelet of order 4 (C4).

2. Square the DWT coefficients.

3. Apply PCA for dimensionality reduction techniques.

4. Determine the optimal number of features per channel based on the target classifier.

1.5 CBIR PROCESS

In the generic CBIR process, the process involves three stages. The first stage involves extraction of features from the images in the database. The extracted features are further indexed and compiled into the database. In the second stage, the query image in input is extracted for features. The final stage involves the comparison of the extracted feature from query with the feature database, and the image is retrieved. The block diagram of the CBIR process is shown in Figure 1.1.

![Figure 1.1 Block diagram of CBIR process](image-url)
**Image Database:** The images are uploaded into database from which the relevant images are to be retrieved.

**Feature Extraction:** All the images in the database are processed to extract features. Generally low level features like color, shape and texture are used as features.

**Feature Indexing:** The features extracted are further indexed for easy comparison.

**Feature Database:** The indexed features are stored in a feature database. Any new image included in the database is processed and its feature is indexed in the feature database.

**Query image:** When images are to be retrieved from the image database based on the content of image, a query image for which similar images are required is given as input.

**Comparison:** The features extracted from the query image are compared with the features indexed in the feature database. The similarity is compared using distance metrics, decision tree and Neural Networks.

**Image Retrieval:** The images whose features are most similar to the query image features are retrieved from the image database.

1.6 **GENERAL ISSUES IN CBIR**

Incorporating adaptable techniques to process images of varied characteristics and classes is the biggest issue for CBIR systems. Factors like image resolution, illumination variation, and occluded objects affect the performance of the systems. The low-level features do not coincide with the high level concepts like events, emotions conveyed by an image. This is
termed as semantic gap (Smeulders et al 2000), which gives the difference between the low level features and high level concepts. Bridging the semantic gap between the low-level features and semantics of the given image is a challenge in CBIR systems. Memory and disk space requirements needed for storing the images and processing is an issue faced by CBIR systems. High dimensionality of feature vectors leads to high computational cost which affects the usability and efficiency of the systems.

1.7 UNSUPERVISED LEARNING

In unsupervised learning, clustering of feature vectors in feature space is done on the basis of similarity. Thus, in a set of clusters, each cluster belongs to a particular class. The system extracts information from the input image and matches it to get the cluster it belongs to. Unsupervised classifier has several issues like the number of classes not known beforehand, defining the similarity between two feature vectors and its measure. Another issue is choosing the right algorithm that will cluster the vectors on the basis of similarity measure, as different algorithms lead to different clusters (Cordon et al 2006).

1.8 SUPERVISED LEARNING

In supervised training, a set of training data and class labels are provided to the classifier. The classifier exploits this prior knowledge (Haykin 1999). Unlike the unsupervised learning, the number of classes, their location and extension in the feature space are known in supervised learning. There are many techniques to design a classifier using supervised learning. Some of the most commonly used techniques are discussed.
1.8.1 Bayesian Classifier

Naïve Bayes is a simple density estimation method for construction of a classification method. Based on prior knowledge, a Bayesian classifier classifies patterns to class C, to which it in all probability belongs. (Duda et al 1973). Knowledge of K, prior probability of each class and class conditional probability density functions, where \( x \) is a feature vector should be available initially to compute the likelihood. Then, Bayes theorem calculates the probability that \( x \) belongs to class given by the Equation (1.11).

\[
P(C_i|x) = \frac{P(x|C_i)P(C_i)}{\sum_{i=1}^{N} P(x|C_i)P(C_i)}
\]

(1.11)

The Bayes probabilistic classifier is also known as Bayesian classification approach (Cox et al 1996), as Naïve Bayes classifier is based on probability models incorporating strong independent assumptions. Though these might not work in a real situation, many complex applications have still been implemented successfully.

1.8.2 k-Nearest Neighbor

k-Nearest neighbor classification (Cover et al 1967) is very simple, conceptually and computationally, with good classification accuracy (Hall 2008). The classification is based on the majority vote of k-nearest classes. The feature vectors are represented as points in feature space and then the distance is determined between the point and the points in training data set. Then the classifier takes only k-nearest neighbor classes so that majority vote is taken to predict the best-fit class for a point (Alexander Thomasian et al 2008).
1.8.3 Decision Tree

Partitioning input space is how a decision tree works. Generally, a tree structure, whose internal nodes and leaf nodes represent tests and classes (Quinlan 1996). The root node is the starting point for classification of a new test point. The root’s different links present possible outcomes, the next step being to make decisions on subsequent nodes. This is continued till leaf nodes are reached. Common tree building algorithms are ID3 and Classification and Regression Tree (CART) (Breiman et al 1984).

1.8.4 Neural network

Artificial networks consist of a large set of interconnected neurons, working parallely to perform learning tasks and its success is due to its learning ability. Multi Layer Perceptron (MLP) a favored supervised learning network model consists of an input layer, one or more hidden layers and an output layer with connections between layers formed by connecting nodes from a specific layer to neurons in the next.

Each connection has a scalar weight which is adjusted during the training phase. The outputs are got from the output nodes. Each connection has a training phase adjustable scalar weight with outputs from output nodes. A feature vector x is input at input layer with the output representing a discriminator between its class and other classes. During training, training examples are fed, predicted outputs computed with the latter being compared to the target output. Errors if any are measured and propagated back through the network, and weights adjusted. Soo Beom Park et al 2004 used color features based neural classification.
1.8.5 Support Vector Machines

Support Vector Machines (SVM) are machine learning algorithms, which show better performance in many domains. SVM Image classification finds optimal separating training sets based hyperplane between classes (Vapnik 1998) with the former being designed for accurate data generalization. The optimal hyperplane leaves the maximum margin from both classes and so the main idea is to locate the hyperplane with maximum margin towards a sample object. In other words, the greater the margin, lesser the possibility about misclassification of any feature vector. Input vectors in image classification are mapped into high dimensional feature space through non linear transformation. Kernel function is used to construct an optimal hyperplane using kernel function. A one-against-all or one-against-one approach is used for multi-class classification in SVM. In the former SVMs are constructed between each class versus other classes and in the latter SVMs are constructed between pairs of classes.

The SVM algorithm is as follows:

For linear SVMs, K is linear; the output of SVM can be expressed in Equation (1.12) as

\[ o = w^* x - b \]  

(1.12)

and

\[ w = \sum_i \alpha_i y_i x_i \]

where

\( o \) is the SVM output

\( w, x, x_i \) are vectors and

\( b \) is the threshold.
Training of SVMs is done by finding $\alpha_i$, expressed as minimizing a dual quadratic form in Equation (1.13).

$$\min_{\nu_i, \nu_j} = \min_{\nu_i, \nu_j} \frac{1}{2} \sum_i \sum_j y_i y_j K(x_i, x_j) \nu_i \nu_j - \sum \nu_i$$

Subject to box constraint

$$0 \leq \nu_i \leq C$$

and linear equality constraint is expressed in Equation (1.14) as

$$\sum_i y_i \nu_i = 0.$$  \hspace{1cm} (1.14)

The $\nu_i$ are the Lagrange multipliers.

1.9 OBJECTIVE OF THE THESIS

This research investigates the efficacy of the various classification algorithms used in content based image retrieval and proposes a Neural Network classifier to improve the classification accuracy. The proposed classification model is tested using medical images. The work is classified into the following areas:

- To investigate CBIR using Discrete Cosine Transform (DCT) for feature extraction and existing classification methods. To investigate feature reduction techniques to improve the classification accuracy.
- To propose an improved kernel for Support Vector Machine (SVM) and investigate the classification accuracy.
- To propose an improved Neural Network classifier for content based image retrieval using Fuzzy to improve the classification accuracy.
To propose an optimization technique for the proposed Neural Network classifier using Genetic algorithm to optimize the momentum and learning rate.

1.10 THESIS ORGANIZATION

This thesis is organized into six chapters. The First chapter is an introduction to the subject of CBIR, the problems associated with it, background of the problem and contributions of this thesis. Second chapter deals with Literature survey on retrieval by semantic image feature, feature extraction techniques used for image retrieval, image Segmentation techniques, relevance feedback, review of image retrieval methods and CBIR using neural network, decision tree, SVM.

In the third chapter, the generic approach to content based image retrieval using DCT and classification techniques is studied. In chapter four, an improved image retrieval using adaptive RBF kernel for SVM classification technique with optimization is investigated. Chapter Five discusses an improved Neural Network classifier for content based image retrieval using Fuzzy for medical image classification. The proposed model is benchmarked with other models including Naïve Bayes, CART and Multilayer Perceptron Neural Network. To further improve on the classification accuracy, the proposed neural network is optimized using genetic algorithm and chapter six concludes this thesis with directions for future scope of work.