CHAPTER 1

INTRODUCTION

Digitized information has gained tremendous interest for more than two decades because of rapid growth of computer technologies. One of the major problems in information technology is easy accessible through digital media, navigation and retrieving accurate and relevant information from big data, since the steadily increasing amount of information has been made.

The starting of this decade witnessed a digital revolution in imaging, not least for applications targeting the mass market. The digital cameras are used dramatically and therefore the amount of digital images increased. Many people will have private collections with thousands of digital images. Naturally, people share images with each other, and publish them on the Internet for instance. These images are collections of important contribution to the internet. Large collections of medical images managed and stored every day in modern hospitals which is a best example. The digital images are used in the workflow and databases by the newspapers, image providers, and other companies in the graphic design industry. Surveillance cameras can produce tremendous amounts of image material, which is an example in the security industry.

Digital cameras, image scanners and high capacity public networks, cheap storage, the size of digital image collection is increasing rapidly with the development of Internet and the availability of efficient image capturing devices. By the users of various domains including remote sensing, fashion, publishing, crime prevention, medicine, architecture etc.,, there is a great need of image searching, retrieval and browsing tool. Fields like biomedicine, crime prevention, architecture, engineering, military, commerce, education and entertainment, in many application fields image utilization is used. In these areas imaging provides important support.

Many general-purpose image retrieval systems have developed to accomplish image retrieval task. They are split into two frameworks (a) Text based image retrieval (b) Content based image retrieval.

1.1 Text-based Image Retrieval

In early days, as the basis for multimedia or image retrieval, the image retrieval systems, the multimedia or images were manually annotated with text and textual queries were used. In text-based retrieval, by taking advantage of keywords, classification codes or subject heading, images are labeled to search and retrieve images. Usage of these textual labels, retrieval has been quickly applied. In a structured way, these strings are stored and can be searched same as in structured query language (SQL). Though, there were some significant problems, because of high cost of manual text annotation for large collections and the lack of information that can be captured effortlessly in the text. In text-based retrieval there is understandably no general query pattern as different users query through different keywords for the same images. With descriptive metadata it requires sophisticated techniques and tremendous effort to annotate all images, which is another problem in this. As well, this method cannot prevent incorrect or missed results based on mistakenly labeled images.

Lastly, in digital images the spatial information can rarely be accurate and completely described in textual annotations. For many image categories, this might be impossible. The development of various techniques of image retrieval are led by these difficulties of indexing, matching and retrieval multimedia information.

1.2 Content-based Image Retrieval

To the problem of digital image search in large databases, content-based image retrieval (CBIR) is the application of computer vision techniques. Past from several years, to retrieve images efficiently CBIR emerged as powerful tool, optically similar to query image. The basic idea behind this approach is representation of image as feature vector. This approach is very effective to measure the similarities between the images with distance and their corresponding feature vectors according to some metrics. In CBIR technique, to search and retrieve digital images image contents are used. To address the problems associated with text-based image retrieval, these technique are introduced. To index and retrieve images efficiently CBIR is mainly used so that it reduces the need for (automated and manual) annotations in the indexing process. To represent images these will find the correct features, and the similarity metric that groups visually similar image together are said to be a important milestone in construction of any kind of CBIR system.

The images are indexed by their visual contents under Content-based image retrieval, like color, texture, shape, structure, motion and combination of these.

The general concepts are rather similar even though CBIR systems differ in details. The main stages in CBIR is to extract features of every image based on the image content and defining comparison rules for images.

1.3 Challenges in CBIR

Designing a robust classifier of object poses in visible light images under many variations is the most notable tedious job, in which it includes:

• Firstly, within the class almost every class of natural objects has huge variations. For instance, for every human face, face shape, skin color and hairstyle, etc., will changes considerably and there will be differences in beard, mustache and wearing glasses and also has sometimes will result in further changes. But in cars, the model, the color and the size will changes significantly between one to another. A robust classifier will try to reach independence of these variations.

• Secondly, one common issue is the background clutter. From various background settings, the images may have taken such as outdoor scenes in cities and indoor environments. The classifier must be able to differentiate between object class from those complex backgrounds.

• Thirdly, ranging from direct sunlight and shadows during the day to artificial or dim lighting inside buildings or at night the illumination condition varies a lot. They are still extremely ineffective even though some solutions for illumination invariance have adopted, when compared to human visual systems in being adaptive to such changes. Hence, to the changes of illumination and lighting conditions a robust classifier of object poses should also provide extensive invariance.

• Finally, since only part of the object in the image is obtainable for processing, partial occlusions may further degrade the classification performance. Though, in this work, it is assumed that no partial occlusion happens and all the objects in the images are fully displayed.

1.4 Need of CBIR

From a large-scale image database, with the explosive growth in the internet and the image records, retrieving images becomes one of the most active research fields. A lot of research work on image retrieval have been carried out during last decade [1]. To give all images text annotations manually become very tedious and impractical because of the large number of image records and it generates the need of an efficient image retrieval system [2]. The visual similarity between a query image and database images has been measured by content-based image retrieval. According to the requests of the user, visual contents are normally known as features which are used by CBIR to search images from large-scale image databases, which is given in the form of a query image [3]. One of the main advantage of CBIR is that it does not suffer from the subjectiveness of textual description. In internet, multimedia, medical image archives, crime prevention, entertainment, and digital libraries, CBIR has diverse applications and it is very important field in image processing [4]. For image retrieval, a number of previous works have done addressing different techniques [5].

1.5 Major CBIR Techniques

Many CBIR systems are developed, but the main problem is to retrieve images based on their pixel content will exist largely unsolved.

1.5.1 Query Techniques

Different types of user queries are used by different implementations of CBIR. Query by example is a query technique with an example image which is involved in providing the CBIR system. Depending on the application, the underlying search algorithms may vary, but with the provided example result images should share common elements. Options for providing example images to the system include:

- By the user, a preexisting image may be supplied or from a random set it is chosen.
- A rough approximation of the image they are looking for is drawn by the user, for instance with blobs of color or general shapes. The difficulties that can arise when trying to describe images with words have been removed by query technique.

- **Semantic Retrieval**

With a user making a request like "find pictures of Abraham Lincoln" the semantic retrieval has been started. For computers this type of open-ended task is very difficult to perform Lincoln may not always be facing the camera or in the same pose. Therefore, Lower-level features like texture, color, and shape are used generally by many CBIR systems. Either in combination with interfaces that allow easier input of the criteria or with databases to match features that have already trained (such as faces, fingerprints, or shape matching) used by these features. In general, in order to identify higher-level concepts image retrieval requires human feedback.

- **Relevance Feedback (Human Interaction)**

Relevance feedback is an interactive process in content-based image retrieval, in which it builds a bridge to connect users with a search engine. According to a user's preference, by updating a query and similarity measures it leads to much improved retrieval performance. The main difficult task is to combine CBIR search techniques available with the wide range of potential users and their interest. To understand the user intent, an aspect of making CBIR successful relies entirely on the ability of the algorithm. CBIR systems will make use of relevance feedback, by marking images in the results as "relevant", "not relevant", or "neutral" to the search query, where the user progressively refines the search results, then repeating the search with the new information.

- **Machine Learning**

Machine learning is a kind of artificial intelligence (AI) technique in this without explicitly programmed, it provides computers the ability to learn. On the development of computer programs, the machine learning focuses completely in which it can change when exposed to new data. The more common in CBIR are Machine learning and application of iterative techniques.

- **Other query methods**

Other query methods contain browsing for instance images, querying by direct specification of image features, querying by image region (rather than the entire image), querying by multiple example images, querying by visual sketch, navigating customized or hierarchical categories, and multimodal queries (e.g., combining touch, voice, etc.,)

1.5.2 Content Comparison using Image Distance Measures

In content-based image retrieval the most common method for comparing two images (typically an example image and an image from the database) is using an image distance measure. In various dimensions like color, texture, shape, and others, an image distance measure will compares the similarity of two images. For instance, with respect to the dimensions that considered a distance of zero signifies an exact match with the query. A value greater than zero might be gathered intuitively which indicates various degrees of similarities between the images. Based on their distance to the queried image search results are sorted.

1.5.2.1 Color

Based on color similarity a computing distance measures is been achieved by determining a color histogram for each image, within an image holding specific values the proportion of pixels is been identified. Based on the colors they contain, the image is examined without regard to image size or orientation, it is known as most widely used techniques. Though, to segment color proportion by region and by spatial relationship among several color regions, several research have attempted to make this task successful.

Color is the most important feature for image representation, and it is also used widely in image retrieval. The color feature extracted from image or regions is done only after the color space is specified. The most often used color extraction techniques are:

- Color Histogram
- Color Moments
- Color Coherence Vector(CCV)

1.5.2.2 Texture

Textures are said to be the complex visual patterns which is made up of entities, or sub patterns that includes characteristics like brightness, color, slope, size etc.,. The texture in an image can be regarded as a similarity grouping. A similarity group can be regarded as texture in an image.

Depending on how many textures detected in the image the textures are represented by Texel's, in which they are placed into number of sets. These sets define the texture and also tells where in the image the texture is located. A technique to compute a characteristic of digital image that is able to express numerically its texture properties is known as Extraction of feature in image texture analysis.

To represent texture is a difficult concept. By modeling texture as two-dimensional gray level variation the identification of specific textures in an image achieved primarily. To estimate the degree of contrast, regularity, coarseness and directionality, the relative brightness of pairs of pixels are computed. The problem is in detecting patterns of co-pixel variation and connected them with particular classes of textures like silky, or rough.

The remaining methods for classifying textures include:

- Co-occurrence matrices
- Laws texture energy
- Wavelet Transform
- Orthogonal Transforms
- Autocorrelation based texture features

The major drawback of the existing CBIR technique due to the semantic gap between the low-level visual features used to represent images and the high-level semantic meaning behind images, the performance of CBIR algorithms are very limited. In this work, to reduce the semantic gap between high-level features and low–level features, a content-based image retrieval system is proposed based on the combination of shape descriptor and HOG features.

1.5.3 Shape Descriptor (High-Level Feature)

From visual cues contained in an image, the image features are derived. As an alphanumeric data in a different formats has been represented such as graphs or the vectors, for the visual context it stands as compact surrogates. Two types of visual features can be distinguished:

i) Color and texture cues are exploited by Photometric features and from raw pixel intensities they may derive directly.

ii) Geometric features make use of shape-based cues.

1.5.3.1 Shape

To describe a given shape feature, some set of numbers that are produced known as a shape descriptor. To quantify shape, a descriptor attempts in such a way that agree with human intuition and these descriptors are in the form of a vector. Some of the shape based techniques which are used commonly are:

- *Fourier Descriptor:* Fourier Descriptor (FD) is considered as a valid description tool even though it is an old technique. To compute, robust to noise and compact, in either contour or regions the shape description and classification using FD are simple. By applying fourier transform on a shape signature the FD is obtained and said to be a one-dimensional function derived from shape boundary coordinates. The Fourier descriptor of the shape is also known as the normalized fourier transformed coefficients.

- *Wavelet Transform:* Wavelet transform is a hierarchical planar curve descriptor, it decomposes a curve in which the coarsest scale components carry the global approximation information and by the same time the finer scale components contain the local detailed information. The desirable properties like multi-resolution representation, invariance, uniqueness, stability, and spatial localization exist in the wavelet descriptor.
- *Region-based Fourier Descriptor:* The region-based FD is also known as Generic FD (GFD), which is used for general applications. On shape images, by applying a Modified Polar Fourier Transform (MPFT) the GFD is derived. The polar shape image treated as a normal rectangular image to apply MPFT. In first step, the approximated normalized image turned counter clockwise by an angular step sufficiently tiny. In second step, into new matrix as row matrix, the positive x direction and the pixel values starting from the image Centre are copied and pasted. In the third step, reputation of step first and second is done until the image rotated by 360 degree. The outcome of these steps is that an image in polar space plots turns to Cartesian space.

1.5.4 Histogram of Oriented Gradients Features

One of the most widely used features in object detection and classification is edge of the image. Edges give dominant gradient magnitudes over the entire image, so this is the main idea behind which the HOG approaches using image gradients and the edge features represented in various ways. The pedestrian detection system [6,7] using histogram of oriented gradients (HOG) is a most popular approach, within dense and overlapping image regions, to compute feature descriptors based on histogram of dominant orientations the gradients are used.

Histogram of oriented gradients are used for the purpose of object detection as feature descriptors in localized parts of an image. The occurrences of gradient orientation play a very important roles. This technique is same as scale-invariant feature transform (SIFT) but there are only few differences like, in operating on a dense grid of uniformly spaced cells and for improved accuracy it uses local contrast normalization on overlapping blocks. HOG [6] descriptors work well on pedestrian detection in both images and videos and on a variety of common animals and vehicles in static pictures, a great success has been achieved. The appearance and shape of local objects is the basic idea behind HOG, which can be described by the distribution of intensity gradients within an image as the votes for dominant edge directions. These features are obtained initially by dividing the image into small contiguous regions of same size, known as cells, then for the pixels within each cell, histogram of gradient directions are collected and finally all these histograms are combined. To increase the detection accuracy against varied illumination and shadowing, by computing a measure of the intensity around a greater region of the image, local contrast normalization are applied known as block, to normalize all cells within the block resultant are used.

Therefore, the HOG feature descriptor contains some critical advantages on other techniques. Initially it operates on localized cells so that those changes would only appear in larger spatial regions, so to geometric and photometric transformations this method results in significant invariance. Next the individual body movement of pedestrians HOG features are more tolerant, because as long as they keep a roughly upright position which in particular is suitable for human detection.

The HOG descriptors are divided into two categories they are: static and motion HOGs. Static HOGs computes from the appearance channel and over individual images. Motion HOGs computes from the motion channel and over a set of consecutive images of a video sequence, which can be used to detect the videos. Static and motion HOGs are completely based on oriented histograms.

1.6 Objectives of the present study

The major challenge in CBIR is to support the high level query by utilizing the low level features. To address this challenge the present study of my research is concentrated on the following objectives:

1. To study the topological structure of an image and its various properties.
2. To design a new shape descriptor (a fast operator) based on shape and color to extract the shape features.
3. To eliminate the semantic gap between high level and low-level feature the shape descriptor(hierarchical part template) and HOG (tree modeling) approach are combined.
4. To perform qualitative analysis of the algorithms, two classification methods Support Vector Machine and K-Nearest Neighbors algorithm are used.

1.7 Principle of CBIR

The contents of image are used by content-based image retrieval to represent and access the images. The off-line feature extraction and online image retrieval, a typical content-based image retrieval system has been divided. A conceptual framework for content-based image retrieval is given in the figure 1.1.

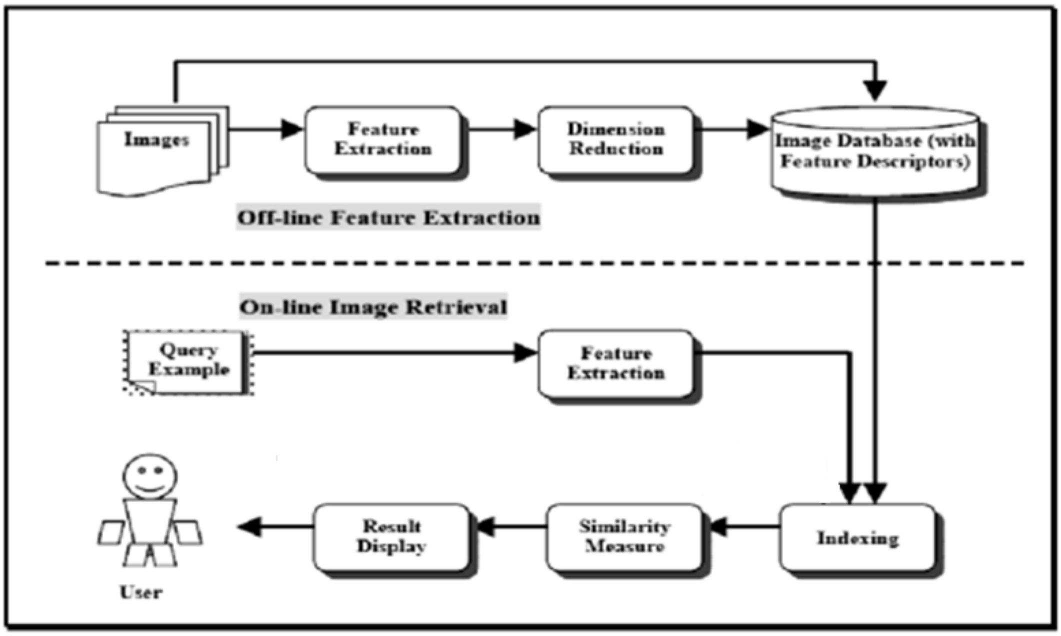

Figure 1.1 A Conceptual Framework for Content-Based Image Retrieval

In off-line stage, based on its pixel values the system automatically extracts visual attributes like color, shape, texture, and spatial information of each image in the database and stores them in a different database within the system known as feature database. Compared to the image data the feature data (also called as image signature) for each of the visual attributes of each image is very tiny in size, therefore in the image database, the feature database contains an abstraction (compact form) of the images. The significant compression of image representation is one of the main advantage of a signature over the original pixel values. However, to gain an improved correlation between image representation and visual semantics, is one of the main important reason for using the signature [63].

The user submits a query image example to the retrieval system to search a desired images in an on-line image retrieval. With a feature vector the system represents this example. Between the feature vectors of the query example the distances (i.e., similarities) and those of the media in the feature database are computed and ranked. To provide an efficient way of searching the image database by applying an indexing scheme retrieval which has been conducted. Finally, the system ranks the search results and later returns the results which is almost similar to the query examples.

1.8 Motivation

It is now possible to create large and extensive databases of digital imagery because of recent advancements in digital storage technology. However, the search in very large databases containing several thousand images is said to be an intended usage. For the research purpose 1000 images of Wang dataset and 5000 images from the Corel-5K bench mark dataset is used. In image classification and image retrieval systems including high-level semantic information is the main motivation of this research.

1.9 Advantages

The Shape and HOG feature descriptor holds some major advantages over other techniques:

- Visual features, like color, and shape information, of images automatically extracted.
- Similarities of images based on the distances between features.
- For the individual body movement of pedestrians HOG features are more tolerant, as long as they keep a roughly upright position so this descriptor is mainly suitable for human detection.
- The compactness of the representation is the huge advantage of key point based method for shape descriptor, compared to the total number of image pixels, the

dimension of feature descriptors based on key points are normally very small, therefore accelerating the classification process is done.

- Instead of the traditional keyword-based approach, the possibility of an automatic retrieval process is one of the major advantages of the CBIR approach.
- Faster calculation and applicability are provided by shape descriptors.

1.10 Limitations

- There is no issue on small set of images but sometimes collision may occur for large set of images.
- To describe image properly, HOG features alone cannot be more effective hence another features like shape feature and color set representations must be incorporated.
- To noise and variations, contour based shape descriptors are sensitive because they use small parts of shapes.

1.11 Applications

There are many applications of shape and HOG descriptor, which are as given below:

- Finger print identification
- Digital libraries
- Crime prevention
- Medical applications
- Bio diversity information systems
- Historical research
- Military (radar, aerial)

1.12 Commercial Implementations

For image retrieval, a selection of some large and interesting commercial services are:

- The best-known Internet search engine is Google (www.google.com). Google's Image search is huge, but the retrieval is mainly depend on text, keywords, etc., however, they are certainly interested in searching image based on its content as seen by Jing and Baluja [8].

- Picsearch another big player in image retrieval, with an image search service holding more than three billion pictures. According to recent research Picsearch propose higher retrieval accuracy (evaluated on hand-labeled ground truth) compared to many of their competitors, for instance Google.
- Cydral (www.cydral.com) and Exalead (www.exalead.com/image) are the instance of other search engines for providing similar image search services, for color or gray scale images and images containing faces, both with the possibility to search. Common challenge for search engines very huge databases which contains uncountable numbers of image domains.
- TrackMyPicture (www.trackmypicture.com) and Photopatrol (www.photopatrol.eu) are two other search engines with similar tracking functions. However, they target rather narrow consumer groups with a Swedish and a German webpage.

The number of commercial implementations are steadily increasing rapidly, but the list will be even longer within few months. In plethora of applications like publishing and advertising, historical research, fashion and graphic design, architectural and engineering design, crime prevention, medical diagnosis, geographical information and remote sensing systems the image retrieval [61] is based on its content. Before trademark finally approved for use in the commerce department there is a need to find out if such or similar ones ever existed or not. In hospitals, to search and review similar X-rays or scanned images of a patient before proffering a solution some ailments requires the medical practitioner.

Web is the most important application, it is devoted to images as big fraction, and searching for a specific image is indeed a daunting task. Numerous commercial and experimental CBIR systems are now obtainable, and with CBIR facilities, many web search engines are now equipped, as for example Alta Vista, Yahoo and Google [62].

1.13 Thesis Outline

In ensuring optimal knowledge transfer and effective research presentation, a well defined research work plays a major role. The overall thesis manuscript has been divided into six individual chapters by considering the research objectives.

A brief outline of the presented thesis work is as given below:

Chapter-2 Literature Survey

For future optimization or enhancements of any kind of research, literature survey or review can be considered as the backbone. This chapter narrates all relevant literature studies for the presented research work which includes the discussion for all shape descriptor, HOG features and other available techniques are noted for Content-Based Image Retrieval system.

Chapter-3 A study on the topological structure of an image and its various properties

The first phase of the research is to learn the different properties of the image. Thereafter apply all the edge detecting operators to find which operator gives the best edge shape for using the edge detector as image signature. Three distance metrics such as Euclidean, Canberra and Manhattan distance are used to evaluate image retrieval based on the edge of the shape.

Chapter-4 A fast operator for detection and precise location of distinct points

The second phase of the research is shape identification and edge mapping have been discussed. Mainly the intended shape descriptor technique for CBIR is based on key points using templates.

Chapter-5 HOG approach to provide low-level features and image structure

The third phase of the research is incorporating the shape descriptor model into another generalized algorithm called HOG descriptor which reduce the semantic gap between high-level and low-level features.

Chapter-6 Conclusion and future work

The significances and conclusions of the proposed research have been mentioned in this chapter. This chapter discusses the respective conclusions and overall system models and its implemented outcomes. A summary of the methodologies along with a direction to carry out the future work is stated.