CHAPTER -1

An Introduction to Content Based Image Retrieval

1.1 Introduction

With the advancement in internet and multimedia technologies, a huge amount of multimedia data in the form of audio, video and images has been used in many fields like medical treatment, satellite data, video and still images repositories, digital forensics and surveillance system. This has created an ongoing demand of systems that can store and retrieve multimedia data in an effective way. Many multimedia information storage and retrieval systems have been developed till now for catering these demands. The most common retrieval systems are Text Based Image Retrieval (TBIR) systems, where the search is based on automatic or manual annotation of images. A conventional TBIR searches the database for the similar text surrounding the image as given in the query string. The commonly used TBIR system is Google Images. The text based systems are fast as the string matching is computationally less time consuming process. However, it is sometimes difficult to express the whole visual content of images in words and TBIR may end up in producing irrelevant results. In addition annotation of images is not always correct and consumes a lot of time. For finding the alternative way of searching and overcoming the limitations imposed by TBIR systems more intuitive and user friendly content based image retrieval systems (CBIR) were developed. A CBIR system uses visual contents of the images described in the form of low level features like color, texture, shape and spatial locations to represent the images in the databases. The system retrieves similar images when an example image or sketch is presented as input to the system. Querying in this way eliminates the need of
describing the visual content of images in words and is close to human perception of visual data. Some of the representative CBIR systems are Query by Image Content (QBIC) [Flickner et al. (1995)], Simplicity [Wang et al. (2001)], and Blob world [Carson et al. (2002)].

![Figure 1.1: Architecture of a typical CBIR system](image)

In a typical CBIR system (Figure 1.1), image low level features like color, texture, shape and spatial locations are represented in the form of a multidimensional feature vector. The feature vectors of images in the database form a feature database. The retrieval process is initiated when a user query the system using an example image or sketch of the object. The query image is converted into the internal representation of feature vector using the same feature extraction routine that was used for building the feature database. The similarity measure is employed to calculate the distance between the feature vectors of query image and those of the target images in the feature database. Finally, the retrieval is performed using an indexing scheme which facilitates the efficient searching of the image database. Recently, user’s relevance feedback is also incorporated to further improve the retrieval process in order to produce perceptually and
semantically more meaningful retrieval results. In this chapter we discuss these fundamental techniques for content-based image retrieval.

1.2 Visual Content Descriptor

Natural images depicting a complex scene may contain a variety of visual artifacts. CBIR systems represent the visual contents of images in the form of a feature descriptor. A good descriptor should not only be invariant to rotation, scaling and illumination variations but also has high discriminating capability. However, there is a tradeoff between invariance and discriminating power of visual features. Employing features having wide variety of invariance may result in losing the capability to discriminate between most essential properties. Study of invariance is largely investigated in the field of computer vision but is relatively new in image retrieval.

A feature descriptor may be local or global. Local descriptors are extracted using a part or region of an image while a global feature uses the visual content of the whole image. A CBIR system which uses region features to represent images is known as Region Based Image Retrieval systems (RBIR). On the other hand CBIR systems utilizing global features for describing images are classified as Global CBIR systems. Local and Global features of an image largely represent color, texture, shape and spatial relationships of different objects in the image. Some widely used color, texture, shape and spatial relationship features are discussed in the following subsections.

1.2.1 Color Features

Color is the most commonly used feature of an image. The perceived color at any pixel of an image is obtained by mixing three preliminary colors in appropriate proportion. The three
dimensional color provides more discriminating information than the single dimensional gray level values. Before extracting color descriptor a proper color space must be determined first. Commonly used color spaces for image retrieval application are RGB, CIE L\*a\*b\*, CIE L\*u\*v\*, HSV and opponent color space. There is no agreement over which color space is best but one of the desirable characteristic of color space for image retrieval task is its uniformity. Uniformity means that the physical distance between any two color pair in the color space must be equal to the perceived distance between them. Some commonly used color descriptors are color moments, color histogram, color coherence vector and color correlogram.

1.2.2 Texture Features

There is no specific definition of texture however one can define texture as the visual pattern that has properties of homogeneity not resulting from the presence of only a single color or intensity. Various techniques for texture analysis have been investigated in the field of computer vision and pattern recognition. The texture extraction techniques can be classified into two categories: statistical and structural. Statistical approaches use intensity distribution of image to extract statistical parameters representing texture of image. Commonly used statistical methods include Fourier power spectra, Co-occurrence matrices, Shift-invariant principal component analysis (SPCA), Tamura feature, Wold decomposition, Markov random field, Fractal model, and Multi-resolution filtering techniques such as Gabor and wavelet transform. Structural methods, including morphological operator and adjacency graph, describe texture by identifying structural primitives and their placement rules. They tend to be most effective when applied to textures that are very regular.
1.2.3 Shape Features

Shape feature provides the most important semantic information about an image. Shape features are usually described using part or region of an image. The accuracy of shape features largely depends upon the segmentation scheme used to divide an image into meaningful objects. However, fast and robust segmentation is difficult to achieve. This limits the shape features only to those retrieval applications where objects or region of images are readily available. The shape descriptors are categorized into two classes: boundary based descriptor and region based descriptor. Some boundary based representative shape description techniques are chain codes, polygonal approximations, Fourier descriptor and finite element model. On the other hand state of the art region based descriptors are statistical moment and area. A good shape feature should be invariant to translation, rotation and scaling. A detail review of shape matching techniques used in image retrieval application can be found in[Mingqiang et al.(2008)].

1.2.4 Spatial Information

The performance of a image retrieval system can be improved by considering spatial locations of different objects in the image. The spatial location of objects and their relationship can provide useful discriminating information in image retrieval applications. For instance, parts of blue sky and ocean may have similar color histograms, but their spatial locations in images are different. The spatial location matching can be implemented by matching the images based on fixed location similarity. In this approach a similar object lying in different regions of an image cannot be detected. For instance; image having tiger in the left part may not get similarity with images having tiger in the right part of images. To overcome this problem systems compare all region of image with the query object or region. This may result in the increase of response time
of the system. The most commonly used techniques for finding spatial location similarity includes 2D strings, spatial quad-tree and symbolic images.

1.3 Similarity Measure

The degree of similarity between query and target images is calculated based on the value of similarity measure. The images are ranked according to their similarity value and presented as output of CBIR system. Often, the choice of similarity measure affects the performance of retrieval system. Many similarity measures have been developed over the years based on the quantitative estimates of the distribution of features in the image. Some of the most commonly used similarity measures employed in CBIR are Euclidean distance, Minkowski- form distance, Histogram intersection distance, Quadratic- form distance, Mahalanobis distance and Kullback Leibler (KL) divergence distance.

1.4 Performance Evaluation

The performance of a retrieval system is evaluated based on several criteria. Some of the commonly used performance measures are average precision, average recall, average retrieval rate and Average Normalized modified Retrieval Rate (ANMRR). All these parameter are computed using precision and recall values computed for each query image. The precision of the retrieval is defined as the fraction of the retrieved images that are indeed relevant for the query:

\[
\text{Precision} = \frac{\text{No. of relevant images retrieved}}{\text{Total no. of images retrieved from the database}}
\] (1.1)

The recall is the fraction of relevant images that is returned by the query:
Recall = \frac{\text{No. of relevant images retrieved}}{\text{Total no. of relevant images in the database}} \quad (1.2)

A good retrieval system should have high values for precision and recall.

1.5 Motivation of the Work

In the last two decades, CBIR systems have been improved a lot. However, there still remain some problems which have not been answered satisfactorily. First and foremost problem is of semantic gap, which exist between low level feature representation of images and the actual visual perception of the image. Researchers all over the globe are working in the direction of narrowing down this semantic gap. Semantic gap is a big problem which can be seen as a collection of many small problems. In this work, we have identified such problems and tried to provide an effective solution to these problems.

1.6 Objective of the Thesis

This thesis is devoted to improving existing techniques involved in feature extraction, similarity matching and reducing the overall computation time of image retrieval system while increasing the accuracy. The main contributions of the thesis are listed below:

1. Design and development of a multistage model for image retrieval to improve the retrieval accuracy by filtering down irrelevant images at each stage.

2. The accuracy of region based image retrieval system is improved by introducing novel region codes based matching scheme.

3. The state of the art LBP based texture descriptors are improved to develop more robust texture features based on LBP framework.
4. The structural and statistical approaches of texture description are utilized to develop a single feature which can thoroughly describe the correlation between color and texture properties of image.

5. A novel region based image retrieval system is developed using LBP based representation of color, texture and shape features. The notion of relative spatial location of different regions in the images is also exploited using improved region codes based matching scheme.

1.7 Organization of the Thesis

This thesis is organized as follows:

Chapter 2: This chapter provides the detailed review of the state of the art related CBIR techniques

Chapter 3: Describes the proposed multistage framework for image retrieval.

Chapter 4: This chapter presents the region based image retrieval using the proposed region codes based matching scheme.

Chapter 5: Deals with developing the local binary pattern based noise invariant features for describing texture.

Chapter 6: Presents the detailed description of proposed short run length feature for image retrieval which can describe the detailed correlation of color and texture in the image.

Chapter 7: In this chapter the proposed LBP based descriptor and region codes based matching scheme is integrated to develop an effective region based image retrieval system.

Chapter 8: Concludes thesis and presents future work.