VISUAL FEATURE EXTRACTION FOR CONTENT-BASED IMAGE RETRIEVAL

Asifa Bhatti
Department of Computer System
Mehran University of Engineering and Technology, Jamshoro,
Pakistan
muet.07cs55@yahoo.com
Saad Masood Butt
UniversitiTeknologi PETRONAS, Malaysia
saadmasoodbutt668@yahoo.com
MumtazHussainSoomro
UniversitiTeknologi PETRONAS, Malaysia
Mumtaz.muet@gmail.com
AzuraOnn
Department of Management and Human Resource, UniversitiTenagaNasional, Malaysia
azura@uniten.edu.my
Moaz Masood Butt
Computer and InformationSciences Department
Bahria University Islamabad Pakistan
moazbutt786@hotmail.com

ABSTRACT: *In this age of computers, virtually all spheres of human life including commerce, government, academics, hospitals, crime prevention, surveillance, engineering, architecture, journalism, fashion and graphic design, and historical research use images for efficient services. In the medical profession, X-rays and scanned image database are kept for diagnosis, monitoring, and research purposes. In architectural and engineering design, image database exists for design projects, finished projects, and machine parts. Most leading academic institutions and many researchers of information-technology-giants focuses on having a general purpose and an efficient CBIR system which is used for searching large image collection. Images are more expressive than words, to express the contents; images are used by most of the websites. With the help of simple searching for an image, image can easily be identified in a small collection of images. But this is not the case with large and random image collections. Effective and efficient retrieval techniques of images are needed because of the explosive growth of digital images. There are two basic approaches used for the information retrieval i-e text-based and content-based image retrieval technique. Text based image retrieval (TBIR) means to search the images with the help of textual metadata associated with the image to specify its characteristics. While CBIR search relies totally on the image searching and retrieval by extracting the contents of the image. The main idea behind the CBIR is efficiency enhancement during image indexing and retrieval, reducing the need for human interruption during image indexing and retrieval process.Content- based image retrieval system is most widely used in area of Architectural and engineering design , Crime prevention , Medical diagnosis, Art collections, Retail catalogs, Photograph archives, The military, Entertainment and Geographical information and remote sensing systems. It is also efficient for simple user searching on web. In this paper the proposed system includes the design of an image database and extraction of visual features from the images for the content based image retrieval system. And then the validation and analysis of the system by using color histogram technique and the Euclidian distance measure on the designed database images. The proposed content based image retrieval system consists of design of image database, feature extraction and distance measure techniques for the storage and automatic retrieval of images.*

1. INTRODUCTION

In the past decade, the information has been developed in computer understandable formats. In the meanwhile, much of the information in older books, journals and newspapers has been digitized and made it computer understandable. Music, satellite, images, pictures, books, newspapers, and magazines and big archives of films have been made accessible for computer users. Currently all institutions make great efforts to digitize existing stocks into multimedia database. Internet enables the users to search and retrieve this vast information. The greatest challenge of World Wide Web is to make available more information about a specific topic; most of the difficult task it is to extract accurate and most suitable information. Mostly users know about their required information, but they don't know how and where to find it. Search engines facilitate the users to search for the relevant information.

In textual representation we can define the above image as it is a picture which depicts a Sunset scene, having some other objects also just like lake or river, a camping site, a person standing, Kayak, National park, a picnic. There are mainly three categories of user consideration about the retrieval when using the system.

1. Search by association - There are various systems and techniques, which are used to browse through massive image

collection from sources that are unspecified by the user, are known as the 'search by association'. There is no specific aim of the users other than finding interesting things. These systems are highly interactive.

2. Target search- Another class of users searches for a specific image is known as 'target search'. The user may have a summarized theme of image in mind. Just like when a user searches through an art catalogues.

3. Category search- is known as the third category of applications. It aims at searching an image which belongs to a specific class.

In this age of computers, virtually all spheres of human life including commerce, government, academics, hospitals, crime prevention, surveillance, engineering, architecture, journalism, fashion and graphic design, and historical research use images for efficient services. A large collection of images is referred to as image database. An image database is a system where image data are integrated and stored [1]. Image data includes the raw images and information extracted from images by automated or computer assisted image analysis. In the medical profession, X-rays and scanned image database are kept for diagnosis, monitoring, and research purposes. In architectural and engineering design, image database exists for design projects, finished projects, and machine parts. Images are more expressive than words. Most of the websites use images to express the contents. With the help of simple searching for an image, image can easily be identified in a small collection of images. But this is

not the case with large and random image collections. Effective and efficient retrieval techniques of images are needed because of the explosive growth of digital images. There is a problem when searching for the images that is relevant to a user requirement in a large image collection. To solve this problem, text-based and content-based are the two techniques adopted for search and retrieval in an image database. Text based image retrieval (TBIR) means to search the images with the help of textual metadata associated with the image to specify its characteristics. While CBIR search relies totally on the image searching and retrieval by extracting the contents of the image.

Text based image retrieval is not accurate because contents of the images and human perception makes it difficult to the user to specify the digital images searched byusing keywords that should match the images vast range the users may need.Need for an efficient database to handle huge amount of image collection so that the user can automatically search for the images. CBIR methods need to be validated using specialized image corpora (DB). The problem of recognizing an object of unknown position and rotation within a scene can be solved by using color histogram.

2. Proposed system

The creation of Content-Based Image Retrieval systems involves research on image processing and database design, handling problems that vary from storage issues to friendly user interfaces. Particularly it is very difficult task to handle the image collection. In this chapter an evaluation of the CBIR system is given by calculating the color histogram of images and then finding the most similar images from the collected image dataset. The main goal is the design of image database following the basic idea behind the CBIR systems, after the database design extraction of color information of all the images by calculating the color histogram. And finally use of Euclidian Distance Measure for the similarity calculation. We have used MATLAB 2008a, and image features are stored in the .mat file. When the feature vectors of the database images and query image are compared, the image having smaller distance is considered as the matching image and returned to the user. In this case search does not depend on the perfect match, but the images having least distance value will be displayed according to the values. By going through preprocessing techniques, database of images is designed.

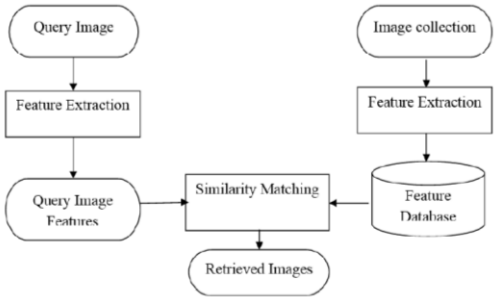

Figure 1: Proposed Content-Based Image Retrieval system

Fig1 shows the proposed CBIR system's block diagram in which when user searches an image, the features of that image is automatically extracted and stored in the feature vector, on the other side the features of already collected images are derived and saved into the array consisting of pixel values at every point of the image. When query image is provided to the system, its features are calculated and compared with the database image features by using the Euclidian Distance Measure for the similarity calculation. The images having smaller distance will be displayed to the user.

There are three steps taken into account for the system evaluation.

1. Database Design
2. Implementation of color histogram for feature extraction
3. Use of Euclidian distance measure for classification

3. Feature Extraction

As we have already discussed the basic theme of Content-based image retrieval system, it purely depends upon the image features. Image contents just like color, shape and textures. By using feature extraction techniques, the information is extracted from the images. After feature extraction, distance measure techniques are applied in order to search for most similar images. Visual features are classified into three types, primitive, logical and abstract features. Primitive features are color, shape and texture. Logical features are just like recognition of the objects shown in the scenes and abstract features such as importance of the scene rendered. Primitive features are most commonly used features in currently available systems. The systems using local features and segments still don't use abstract features.

Color

Color is a most effective feature that is used in CBIR systems and it is used by most of the available systems. Journalists, artists, advertisers use large, varied databases in stock photography. Generally the images use RGB color space.

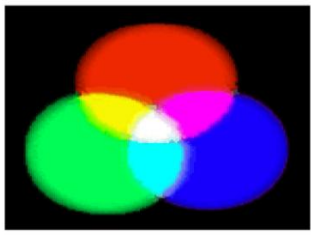

Figure 2: RGB color formulation

Red, green and blue are three basic colors that are used by RGB color spaces. By adding these three colors make it able to reproduce all other colors. Frequently used spaces just like CIE lab, Luv and HSV are adjusted according to the human conceptualization. In medical field, grey level features that are also known as absolute color possess narrow expressive power.

4. Algorithm of Proposed Method

The following proposed algorithm is used to test the system for color histogram technique for content based image retrieval system.

-Calculate color histogram for TR1 and TR2 images.
-Feature vectors are generated for TR1 and TR2 and stored.
-Give Query image as an input.
-Calculate Color histogram
-Calculate features
-Generation of feature vector for the query image.
-Similarity measurement between query image and the images in the databaseis calculated by using Euclidian distance measure.
-From the image collection, the top closest images, to query image areretrieved.

5. Experimental Arrangement

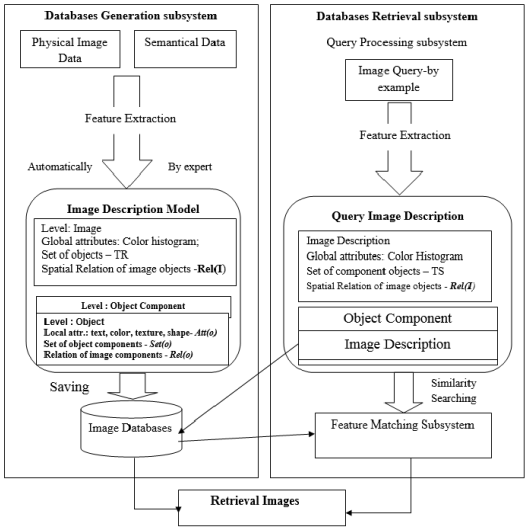

Figure 3: Experimental Arrangement

All visual information system researchers focus on the Content Based Image Retrieval systems. New visual query specification methods, new indexes for data setting, and new ways for similarity determination between the saved in the Image Databases (IDB) image description data and describing queries are used for image search and retrieval on the base of their contents. The CBIR system is explained in detail in proposed experimental arrangement, in which we have two sections. One is Database Generation subsystem and another is Database Retrieval subsystem.Within the above proposed model an image is described in terms of simple and of complex objects. Each object is described by Attributes (semantic and syntacticalsuch as color, texture, shape. Image is denoted by I. The images inserted in IDBtogether with their feature description data form a property presentation that is saved in a data structure of IDB. The features in the form of coded vectors can be used as indexes for direct organization of the access to IDB or for data clustering. The measure evaluates the similarity degree between two images and is determining in the process indexing and similarity integration of two-dimensional index vectors of the image and the query. The attributes of both query and database image is compared.

6. Experimental Results

Color Histogram

Color Histogram is widely suggested technique in the literature for efficient feature extraction, but the validation of the technique is yet not achieved for the variety of databases. Red plane shows the pixel distribution of red color in the images, green plane shows the distribution of green color in the house image and in same manner the blue plane chows the distribution of blue in the image.

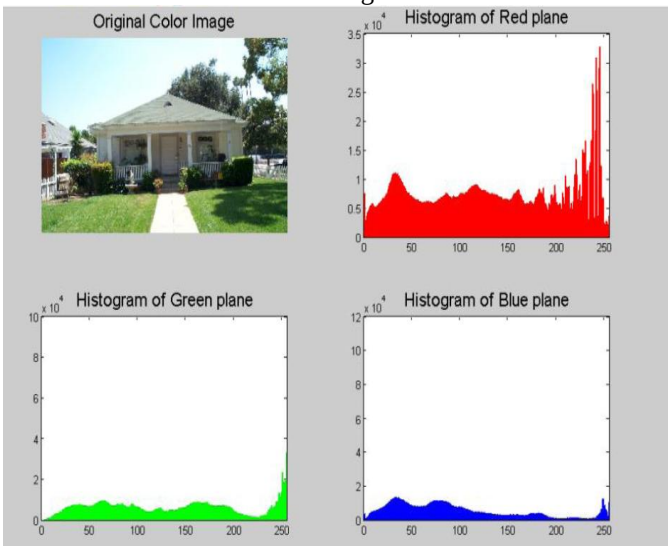

Figure 4: Color Histogram

The features extracted from the above histogram of the image, are stored into the .mat file of matlab, shown in the following figure.

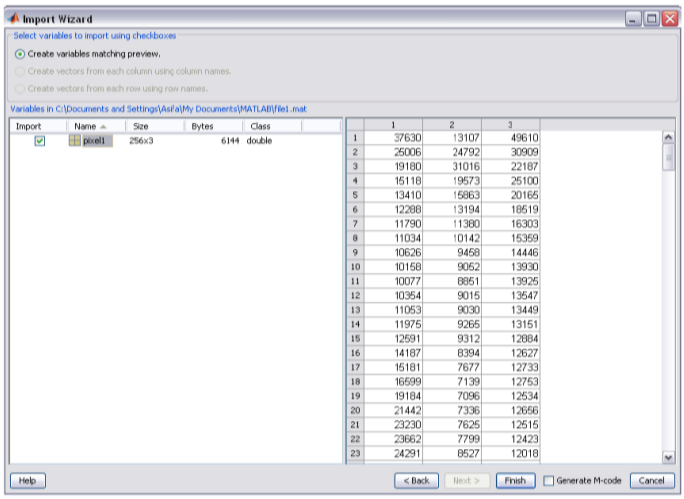

Figure 5: .mat file

The feature vector is then brought into the main memory of matlab for further calculations. The size of every image is 256*256, that is why the size is 256*3 as shown in above figure. Red color pixels are shown in the first, the second column indicates the pixel distribution of green color in each bin and in same fashion the last column shows the blue color pixel distribution in every bin. The features are derived directly from the image, as it is the basic idea behind Content- based image retrieval systems.

7. RESULTS

These are experimental results after applying the CBIR techniques on our datasets. When one of the image from TS1 is given as query image, tested on Training set1, the result is shown in following figure.

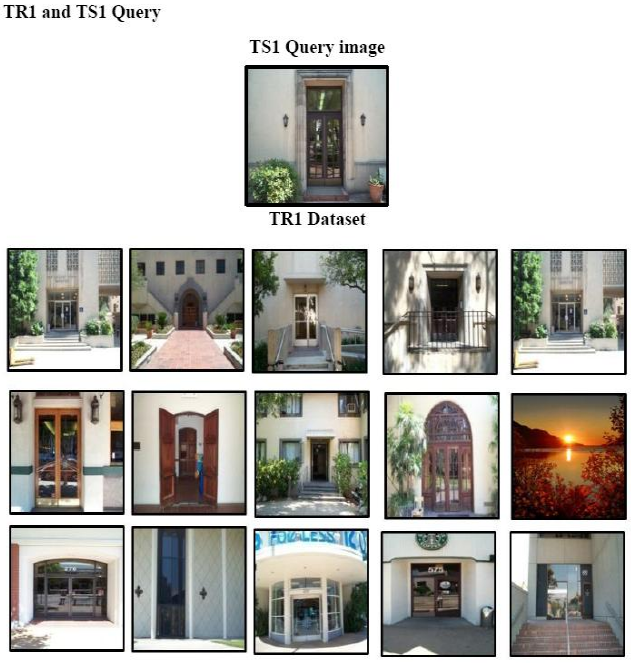

Figure 6: TS1 and TR1 Query Result

When query image is entered to training set1, and by using Euclidian distance measure we measured the distance between query image and dataset images. Top 15 images having smaller distance value are retrieved, since the TR1 contains door and some random images, the query returned us 14 door images and one sunset image. It gives us 93.33%. In the following figure, the previous table data is represented graphically. Horizontal
axis denotes the Testing Sets i-e TS1, TS2, TS3, TS4. The vertical axis denotes the
Training Sets i-e TR1 and TR2. Graph shows the % of each query when we tested
every testing set
Data with Training sets.
TS1 and TR1 resulted 93.33%
TS1 and TR2 resulted 80%
TS2 and TR1 resulted 86.66%
TS2 and TR2 resulted 93.33%
TS3 and TR1 resulted 73.33%
TS3 and TR2 resulted 86.66%
TS4 and TR1 resulted 93.33%
TS4 and TR2 resulted 73.33%

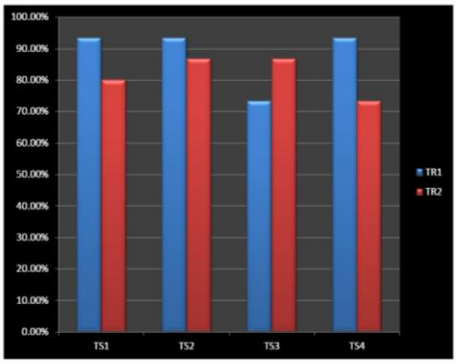

Figure 7: Graphical Representation

8. CONCLUSION AND FUTURE WORK

The basic aim of this thesis was the design of an image database and extraction of visual features from the images for the content based image retrieval system. And then the validation and analysis of the system by using color histogram technique and the Euclidian distance measure on the designed database images. The proposed content based image retrieval system consists of design of image database, feature extraction and distance measure techniques for the storage and retrieval of images. Color histogram represents the distribution of pixels of each color within the image. Ease of computation and simplicity are the attractive characteristics of conventional color histogram and is faster to compute compared to other invariants. It is a widely
suggested technique in the literature for efficient feature extraction, but the validation of the technique is yet not achieved for the variety of databases. Color histogram is used for the images in which the position of objects is not known and having unknown rotation within the scene. The features of training set images are calculated first and stored in .mat file. Training set consists of 500 images. Then the features of testing set images are derived. Testing set consist of 120 images. By applying the Euclidian distance measure technique, distance between query Image and database

image is calculated. The images having least distance value are displayed. For the 20 iterations, the proposed system provides average of 85%.

The proposed system is only based on the retrieval of images using color feature, its performance can be enhanced by using the combination of shape, texture and color features. Another future work can lead to Fuzzy Color Histogram (FCH) and Invariant Color Histogram (ICH). By considering the color similarity of each pixel's color associated to all the histogram bins through fuzzy-set membership function. Fuzzy color histogram can be used to solve the problem of spatial relationship between bins which are closer to each other. When comparing with the CCH, in which each pixel is assigned into only one 67 of the bins, fuzzy color histogram sets each pixel's total membership value to all the histogram bins.

REFRENCES

[1] NeelimaSadineni, DavarapalliAnandam, Venkata Ramesh Pokala, Pasupuletu Rajesh &Kolakalurisirinivasa Rao, "Efficient Image Retrieval Based On Texture Features Using Concept of Histogram", Global Journal Of Computer Science And Technology, Volume 12 Issue 5 Version 1.0 March, 2012.

[2] Igor Francisco AreiasAmaral, "Content-Based Image Retrieval for Medical Applications", Faculty of Science, University of Porto, October, 2010.

[3] S.Selvarajah and S.R. Kodituwakku, "Analysis and Comparison of Texture Features for Content Based Image Retrieval", International Journal of Latest Trends in Computing (E-ISSN: 2045-5364) Volume 2, Issue 1, March 2011.

[4] S.Nagendram, Gh.Ridhika Rani, G.S Sharma, "A Study on Content Based Image Retrieval and Storage Methods for Medical Images", CLEAR International Journal of Research in Science and Technology, Vol-01, December 2011.

[5] P.S.Suhasini , Dr. K.Sri Rama Krishna, Dr. I. V. Murali Krishna, " CBIR Using Color Histogram Processing", Journal of Theoretical and Applied Information Technology, India, 2005 - 2009 JATIT.

[6] S.R.Surya, " Survey on Cotent Based Image Retrieval", PSGR Krishnammal College for women Coimbatore-641004, Vol. 2, No. 5 Oct-Nov 2011.

[7] John Eakins Margaret Graham, " Content-based Image Retrieval", University of Northumbria at Newcastle, October, 1999.

[8] Swati V. Sakhare&Vrushali G. Nasre, "Design of Feature Extraction in Content Based Image Retrieval (CBIR) using Color and Texture", BapuraoDeshmukh College of Engg, Sevagram, Wardha, International Journal of Computer Science & Informatics, Volume-I, Issue-II, 2011.

[9] D. Zhang, "Improving Image Retrieval Performance by Using Both Color and Texture Features," In Proc. of IEEE 3rd International Conference on Image and Graphics (ICIG04), Hong Kong, China,Dec.18-20, 2004, pp.172-175.

[10] Hiremath P. Sand JagadeeshPujari, "Content Based Image Retrieval using Color Boosted Salient Points and Shape features of an image", International Journal of Image Processing, Volume (2): Issue (1), 2008.

[11] OnurKucuktunc, DaniyaZamalieva, "Fuzzy Color Histogram-based CBIR System", Bilkent University, Department of Computer Engineering, Ankara Turkey, March, 2009.

[12] HanifeKebapci, BerrinYanikoglu, " Plant Image Retrieval Using Color, Shape and Texture Features", The Computer Journal Advance Access published April 9, 2010.

[13] John Eakins Margaret Graham, "Content-based Image Retrieval", University of Northumbria at Newcastle, October, 1999.

[14] Arjunan R.V., "Image Classification in CBIR systems with Color Histogram Features", International Conference on Advances in recent Technologies in Communication and Computing 2009.

[15] Ricardo da Silva Torres, Alexandre Xavier Falcão, "Content-Based Image Retrieval: Theory and Applications", Institute of Computing, State University of Campinas, Campinas, SP, Brazil, 2006.