ISSN No. 0976-5697

Volume 3, No. 3, May-June 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Content Base Image Retriveal Using Feature Extraction

Sangram S. Dandge
Department of Information Technology
Prof.Ram Meghe Institute Of Technology & Reasearch,
Badnera. Amravati, India
sangramdandge@gmail.com

Prof. A.P.Bodkhe.
Department of Information Technology
Prof.Ram Meghe Institute of Technology & Reasearch,
Badnera Amravati,India
apbodkhe@rediffmail.com

Abstract: Now a day's people are interested in using digital images. There is a great need for developing an efficient technique for finding the images. A significant and increasingly popular approach that helps in the retrieval of image data from a huge collection is called Content Based Image Retrieval (CBIR). In order to find an image, image has to be represented with certain features. Colour, texture and shape are three important visual features of an image. We will implement an efficient image retrieval technique which uses dynamic dominant colour, texture and shape features of an image. It will be helpful and easy way to retrieve image from huge database. In order to find image from huge database which uses dominant colour it is image is uniformly divide into 8 coarse partition as a first step after above coarse partition, the centroid of each partition is selected dominant colour .Texture of an image of an image is obtain by gray level Co-Occurrence matrix(GLCM) and as per shape .There is no universal definition of what shape is either. Shape is a well-defined concept and there is considerable evidence that natural objects are primarily recognized by their shape. Thus, using matching and comparison algorithms, the colour, texture and shape features of one image are compared and matched to the corresponding features of another image. This comparison is performed using colour, texture and shape distance metrics. In the end, these metrics are performed one after another, so as to retrieve database images that are similar to the query.

Keywords: Digitalize, feature extraction, image feature database, image database, image matching and multidimensional indexing

I. INTRODUCTION

The importance of an effective technique in searching and retrieving images from huge data base, for this proposed Content base image retrieval (CBIR) System [1]. Content Based Image Retrieval is the retrieval of images based on visual features such as color, texture and shape. Reasons for its development are that in many large image databases, traditional methods of image indexing have proven to be insufficient, laborious, and extremely time consuming. These old methods of image indexing, ranging from storing an image in the database and associating it with a keyword or number, to associating it with a categorized description, have become obsolete. This is not in CBIR. In CBIR, each image that is stored in the database has its features extracted and compared to the features of the query image. It involves two steps[2]:

a. Feature Extraction: The first step in the process is extracting image features to a distinguishable extent.

b. Matching: The second step involves matching these features to yield a result that is visually similar.

II. EXISTING SYSTEM

Early techniques were not generally based on visual features but on the textual annotation of images. In other words, images were first annotated with text and then searched using a text-based approach from traditional database management systems. Text-based image retrieval uses traditional database techniques to manage images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard Boolean queries. However, since automatically generating descriptive texts for a wide spectrum of images is not feasible, most text-based image retrieval systems require manual annotation of images. Obviously, annotating images manually is a cumbersome and expensive task for large image databases, and is often subjective, context-sensitive and incomplete. As a result, it is difficult for the traditional text-based methods to support a variety of task-dependent queries [3], [4].

III. PROPOSED SYSTEM

The proposed system is to extract the visual features of a query image and compare them to those of database images. The image features under consideration were color, texture and shape. Thus, using matching and comparison algorithms, the color, texture and shape features of one image are compared and matched to the corresponding features of another image. This comparison is performed using color, texture and shape distance metrics. In the end, these metrics are performed one after another, so as to retrieve database images that are similar to the query [5], [6].

A. Proposed Algorithm in CBIR System:

To describe image from the different aspects in order to obtain better search results and to express more image information, we consider the dominant color, texture and shape features combined. The proposed method is based on dominant color, texture and shape features of image[7]

The retrieval steps can be as follows:

Step 1: Uniformly divide each image in the database and the query image into 8-coarse partitions.

Step 2: For each partition, the centroid of each partition can be selected as its dominant color.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

Step 3: Obtain texture features, we use Gray level co-occurrence matrix (GLCM)
Step 4: Obtain shape features, we use boundary based shape representation
Step 5: Construct a combined feature vector for color, texture and shape.
Step 6: Find the distances between feature vector of query image and the feature vectors of target images.
Step 7: Sort the distances.
Step 8: Retrieve most similar images with minimum distance.

B. Extraction of dominant color of an image:

Dominant color descriptor (DCD) is one of the color descriptors proposed by MPEG-7 in [7] that has been extensively used for image retrieval. Among the color descriptors, the salient color distributions in an image or a region of interest are described by DCD. DCD provides an intuitive, effective and compact representation of colors presented in an image. The efficiency of computation for dominant color extraction is significantly improved by this approach[1][8]

Yang et al. presented a color quantization method for dominant color extraction, called the linear block algorithm (LBA), and it has been shown that LBA is efficient in color quantization and computation. For the purpose of effectively retrieving more similar images from the digital image databases (DBs). According to numerous experiments, the selection of color space is not a critical issue for DCD extraction. Therefore, for simplicity and without loss of generality, the RGB color space is used. Firstly the image is uniformly divided into 8 coarse partitions, as shown in Fig.1 If there are several colors located on the same partitioned block, they are assumed to be similar. After the above coarse partition, the centroid of each partition is selected as its quantized color. Let X=(XR, XG,XB) represent color components of a pixel with color components Red, Green, and Blue, and Ci be the quantized color for partition i.

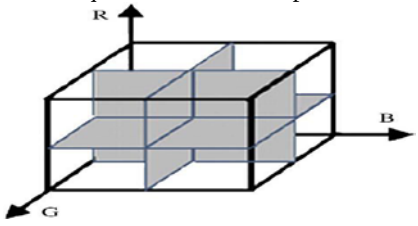

Figure 1. The coarse division of RGB color space.

C. Extraction of texture of an image:

A variety of techniques can be used for measuring texture similarity. The best established can be rely on comparing values of what are known as Second-order statistics calculated from query and stored images. Essentially, these can calculate the relative brightness of selected pairs of pixels from each image. From these, we can calculate measures of image texture. Alternative methods of texture analysis for retrieval can include the use of Gabor filters. Texture queries/specifying can be formulated in a similar manner to color queries, by selecting examples of desired textures from a palette, or by supplying an example query image. The system then retrieves images with texture measures similar in value. as a extraction of texture of an image we use Grey-Level Co-occurrence Matrix (GLCM)[9][10].

The GLCM is a tabulation of often different combination of pixel brightness values (gray levels) occurs in an image. it is very successful method . GLCM uses grey-level co-occurrence matrix to sample statistically the way certain grey level occur in relation to other grey levels. Grey level matrix is a matrix whose elements measures the relative frequencies of occurrence of grey level combinations among pairs of pixels with specified spatial relationship .several texture features such as entropy, energy, contrast and homogeneity can be extracted from the co-occurrence matrix of grey levels of an image. all these feature supported by MATLAB.

GLCM is composed of the probability value, it is defined by $P(i,j|d,\theta)$ which expresses the probability of the couple pixel at θ direction and d interval. when θ and d is determined, $P(i,j|d,\theta)$ is showed by $P_{i,j}$. Distinctly GLCM is a symmetry matrix and its level is determined by the image gray level. Elements in the matrix are computed by the equation shown below.

$$P(i,j|d,\theta) = \frac{P(i,j|d,\theta)}{\sum\sum P(i,j|d,\theta)}$$

The texture feature are computed for an image when d=1 and $\theta = 0^0, 45^0, 90^0, 135^0$. In each direction four texture feature calculated. They are as texture feature descriptor. they include

Energy $E = \sum_i \sum_j P(i,j)$
Entropy $S = \sum_i \sum_j P(i,j) \log P(i,j)$
Contrast $I = \sum_i \sum_j (i-j)^2 P(i,j)$
Homogeneity $H = \sum_i \sum_j \frac{P(i,J)}{1+|i-j|}$

D. Extraction of shape of an image:

We will consider shape as something geometrical. Impressions of shape can be conveyed by color or intensity patterns, or texture, from which a geometrical representation can be derived. Therefore, edge orientation at only region contours as shape information. Shape is the most obvious requirement at the primitive level. Unlike texture, shape is a fairly well-defined concept and there is considerable evidence that natural objects are primarily recognized by their shape.

First we identified the object of an images, then extracting the features of an object , the same set of a feature extracting from query image. and retrieving those stored image whose feature are mostly closed/ match with query image[11],[12]

E. Measurement of a Shape:

a. First we take a color image is convert into gray image and gray image is convert to binary image.
b. Maximum area containing object is detected.
c. First detect the edge and then extract the edge of an image.
d. To reduce the number of points we use down sampling.
e. Same procedure for query image
f. lastly comparing and matching the values most closed image retrieved.

F. Distance Calculation:

The distances between feature vector of query image and the feature vectors of target images can be computed using Euclidean distance.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

Calculate Euclidean Distance then get Euclidean (D)

$D(I,J) = (\sqrt{\sum(fi\ (I) - fj\ (J))2}\ )$

We denote D(I,J) as the distance measure between query image I and image J in the database f_i (I) as the number of pixel in bin i of I.where f_i and f_j represent the average valves of feature vector respectively

G. Distance calculation:

The distances between feature vector of query image and the feature vectors of target images can be computed using Euclidean distance.

H. Image retrieval:

At last, most similar images with minimum distance will be retrieved. That is after the matching and comparing the combined feature vector and query image the most matching image retrieved.

IV. GENERAL SCHEMA DIAGRAM OF CONTENT BASED IMAGE RETRIEVAL

The block diagram consists of following main blocks - digitizer, feature extraction, image database, feature database, and matching and multidimensional indexing. Function of each block is as follows.

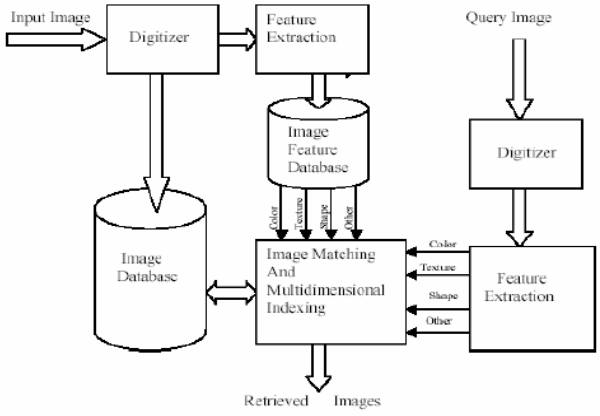

Figure 2. General schema of Content Based image Retrieval.

a. ***Digitizer:*** To add new images in image database or query images which are acquired from Cameras, X-ray imaging system, microdensitometer's, image dissectors, vision cameras etc. are needed to be digitized, so that computer can process those images.

b. ***Image Database:*** The Comparison between Query image and images from image database can be done directly pixel by pixel which will give precise match but on the other hand, recognizing objects entirely at query time will limit the retrieval speed of the system, due to the high expense of such computing. Generally this crude method of comparison is not used, but image database, which contains raw images, is required for visual display purpose[15].

c. ***Feature Extraction:*** To avoid above problem of pixel-by-pixel comparison next abstraction level for representing images is the feature level. Every image is characterized by a set of features such as Texture, Color, Shape and others. Extract these features at the time of injecting new image in image database. Then summarize these features in a reduced set of k indexes and store it in Image feature database. The query image is processed in the same way as images in the database. Matching is carried out on the feature database.

d. ***Image matching and Multidimensional Indexing:*** Extracted features of query image are compared with features, which are stored in image feature database. To achieve fast retrieval speed and make the retrieval system truly scalable to large size image collections and effective multidimensional indexing is indispensable part of the whole system. The system selects the N images having the greatest overall similarities [7],[8].

V. RESULT

As per the result on Content base image retrieval (CBIR) System is as shown below. The below Screeshotshots shows the directory/file contain lots of images. as per retrieving a particular image we passing query as a image it can comparing and matching the whole images on the base on color, texture and shape ,The most matching image retrieved from the database.

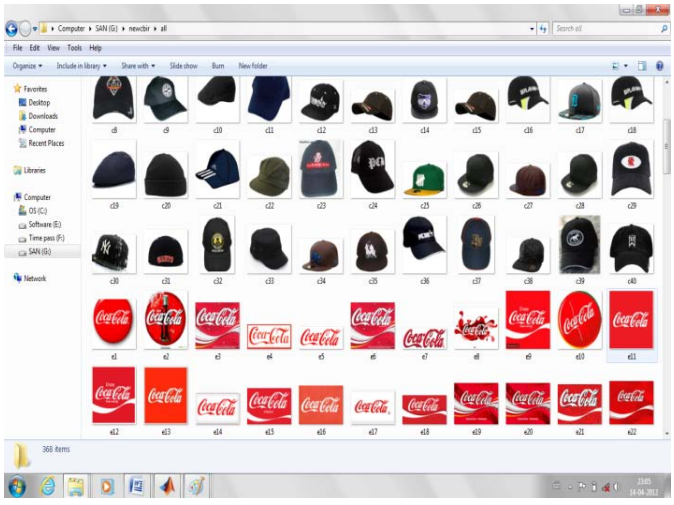

Figure 3. Screenshot no 1: Directory contains lots of images.

Figure 4. Screenshot no 2: The retrieving most matching image.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html

The above figure 3 and 4, shows that we are passing query as image after it can match the whole images on base on color, texture and shape the most matching image retrieved from the database.

VI. CONCLUSION

The image retrieval system started with retrieving images using textual annotations but later introduced image retrieval based on content this came to be known as Content Based Image Retrieval System. Using this proposed design images based on visual features such as color, texture, shape, as opposed to depending on image descriptions or textual indexing can be retrieved. In this seminar is proposed an image is stored in the database has its features extracted and compared to the features of the query image by using Extraction and matching.

VII. REFERENCES

[1]. S. Brandt, J. Laaksonen, and E. Oja, "Statistical shape features in content-based image retrieval", In Proceedings of 15th International Conference on Pattern Recognition, volume 2, pages 1066–1069, Barcelona,Spain, September 2000.

[2]. N.-S. Chang and K.-S. Fu, "Query by pictorial example", IEEE Transactions on Software Engineering,6(6):519–524, November 1980.

[3]. Srinivasa Kumar Devireddy, "Content Based Image Retrieval" In Georgian Electronic Scientific Journal: Computer Science and Telecommunications 2009|No.5(22)

[4]. P. Koikkalainen and E. Oja, "Self-organizing hierarchical feature maps", In Proc. IJCNN-90, International Joint Conference on Neural Networks, Washington,DC, volume II, pages 279–285, IEEE Service Center,Piscataway, NJ, 1990.

[5]. J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "Self organizing maps as a relevance feedback technique in content-based image retrieval", Pattern Analysis & Applications, June 2001, In press.

[6]. J. T. Laaksonen, J. M. Koskela, S. P. Laakso, and E. Oja, "PicSOM - Content- based image retrieval with self-organizing maps",Pattern Recognition Letters,21(13-14):1199–1207, December 2000.

[7]. Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design", IEEE Transactions on Communications, COM-28(1):84–95, Jan. 1980.

[8]. R.W. Picard, T. P. Minka, and M. Szummer, "Modeling user subjectivity in image libraries", Technical Report #382, M.I.T Media Laboratory, 1996.

[9]. G. Salton and M. J. McGill, "Introduction to Modern Information Retrieval", Computer Science Series, McGraw-Hill, 1983.

[10]. M. Stricker and M. Orengo, "Similarity of color images" In Storage and Retrieval for Image and Video Databases III (SPIE), volume 2420 of SPIE Proceedings Series, pages 381–392, San Jose, CA, USA,February 1995

[11]. ZhouBing, Yang Xin-xin, "A Content-based Parallel Image Retrieval System" In 2010 International Conference On Computer Design And Appliations (ICCDA 2010)

[12]. Wang Xiaoling,"A Novel Circular Ring Histogram for Content-based Image Retrieval" in 2009 First International Workshop on Education Technology and Computer Science

[13]. Markus Koskela, Jorma Laaksonen, and Erkki Oja ,"COMPARISON OF TECHNIQUES FOR CONTENT-BASED IMAGE RETRIEVAL" In 2001 Norsk forening for bildebehandling og mønstergjenkjenning (NOBIM). Reprinted with permission from Proceedings of the 12th Scandinavian Conference on Image Analysis (SCIA 2001). Bergen, Norway, pages 579-586.

[14]. Rouhollah Rahmani, Sally A. Goldman, Hui Zhang, Sharath R. Cholleti, and Jason E. Fritts, "Localized Content-Based Image Retrieval" In IEEE Trancations on Pattern analysis and machine Intelligence, vol.30,no.11.november2008

[15]. Ya-Li Qi, "Classification for Trademark Image Based on Normal Cloud Model" In 2009 International Conference on Information Management, Innovation Management and Industrial Engineering.

CONFERENCE PAPER
National Conference on Information and Communication Technology for Development
Organized by PRMITR, Amravati (MS) India
http://mitra.ac.in/forthcoming.html