International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)
Volume 3, Issue 2, February 2014

Multifeature similarity score image retrieval

S.Muralikaruppaiyan M.E., Communication systems
Sudharsan Engineering College.
R.Devi M.TECH., Software Engineering
Kings College of Engineering.

***Abstract*—This project deals with efficient retrieval of images from large databases based on the multi-feature similarity score fusion using genetic algorithm. With the increasing popularity of the use of large volume image data base in various applications, it becomes imperative to build an efficient retrieval to browse through the entire data base. Unlike previous approaches which concentrate on extracting a single concise feature, our technique combines features that represent both the color and texture in images. Single feature describes image content only from one point of view, which has a certain one-sided. Fusing multi-feature similarity score is expected to improve the system's retrieval performance. In this paper, the retrieval results from color feature and texture feature are analyzed, and the method of fusing multi-feature similarity score is described. For the purpose of assigning the fusion weights of multi-feature similarity scores reasonably, the genetic algorithm is applied.**

***Keywords: Genetic algorithm, Color feature, Text feature*.**

I. Introduction

Application of World Wide Web (www) and the internet is increasing exponentially, and with it the amount of digital image data accessible to the users. A huge amount of Image databases are added every minute and so is the need for effective and efficient image retrieval systems. There is a rapid increase in the size of digital image collections together with the fast growth of the Internet in the recent years. Digital images have found their way into many application areas, including Geographical Information System, Office Automation and Medical Imaging. The large numbers of image collections have posed increasing technical challenges to computer systems to store/transmit and manage image data effectively. The storage and transmission challenge is tackled by image compression. The challenge to retrieve stored images is studied in the context of image database and has been attempted by researchers worldwide from a wide range of disciplines from computer vision image processing to traditional database areas for over a decade.

.

II. FEATURE EXTRACTION OF HSV COLOR

HSV color space is widely used in computer graphics, visualization in scientific computing and other fields. In this space, hue is used to distinguish colors, saturation is the percentage of white light added to a pure color and value refers to the perceived light intensity. The advantage of HSV color space is that it is closer to human conceptual understanding of colors and has the ability to separate chromatic and achromatic components.

A. Color feature extraction

HSV color model forms a uniform color space, which uses a linear gauge. The perceived distance between colors is in proportion to Euclidean distance between corresponding pixels in HSV color model, and conforms to eye's feeling about color. So it is very suitable for color based image similarity comparison. It is essential to quantify HSV space component to reduce computation and improve efficiency. At the same time, because the human eye to distinguish colors is limited, do not need to calculate all segments. Unequal interval quantization according the human color perception has been applied on H, S, and V components

According to human cognitive about color, three components of HSV space are quantified in non-uniform manner. Hue is quantized into 16 bins and is among [0, 15]. Saturation is quantized into 4 bins and is among [0, 3]. Value is quantized into 4 bins and is among [0, 3]. Among those three components, human cognitive about color is mainly based on hue, and then saturation, finally value. So, quantized results are coded as

$$C=16H+4S+V$$

III. Texture feature extraction

The statistical properties of image co-occurrence matrix are taken as texture features of an image. Firstly, color image is converted to gray scale image, and the image co-occurrence matrix is gained. Then, the following five statistical properties are calculated to describing image content. They are contrast, energy, entropy and inverse difference.

B. Gray level co-occurrence matrix

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a special spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. The texture filter functions, described in using Texture filter functions, cannot provide information about shape i.e., the spatial relationship of pixels in an image.

By calculating how often with the intensity value I occur in a specific spatial relationship to a pixel with the value j, gray-level co-occurrence matrix can be creating. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right, but you can specify other spatial relationship between to pixels. Each element (I, j) in the resultant GCLM is simply the sum of the number of times that

ISSN: 2278 – 909X

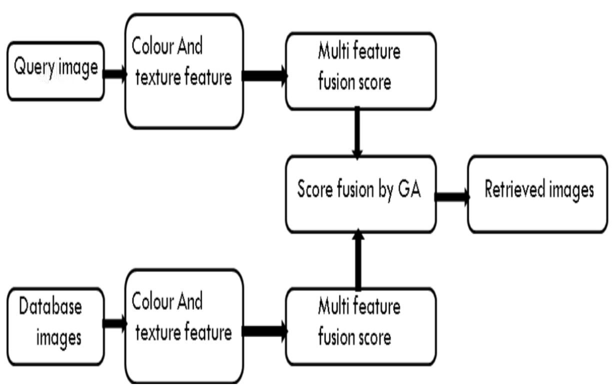
the pixel with value I occurred in the specified spatial relationship to a pixel with value j in the input image. The number of gray levels in the image determines the size of the GLCM. The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture images. For example, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse with respect to the specified offset.

C. Text feature extraction based on GLCM

GLCM creates a matrix with the directions and distances between pixels, and then extracts meaningful statistics from the matrix as texture features. GLCM texture features commonly used are shown in the following

GLCM is composed of the probability value, it is defined by P (i, j|d, ø) is showed by P i, j. distinctly GLCM is a symmetry matrix; its level is determined by the image gray-level. Elements in the matrix are computed by the equation showed as follow:

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



Block diagram of the process

IV. Multi-feature similarity score fusion

The physical meaning of different, features are different, and value ranges are totally different, so similarity scores of different features cannot be compared. So, before multi-feature similarity score are fused, they should be normalized. Similarity scores can be normalized through the following ways. Let Q be the query image. By calculating distances

Can be gotten, where i=0, 1…., N, N is the number of images in database. Thus, similarity score normalization can be implemented as

$$S_{Ni}=\frac{S_i-\min\{S_i\}}{\max\{S_i\}-\min\{S_i\}}$$

Thus results of multi-feature similarity score is

$$S_{Fi}=\frac{S_{NCi}\cdot W_C+S_{NTi}\cdot W_T}{W_C+W_T}$$

Where the SFI is the fused similarity score, S_{NCi} is the normalized color feature similarity score, SN_{Ti} is the normalized texture feature similarity score, W_c is the weight of color feature similarity score, and WT is the weight of texture feature similarity score. By assigning appropriate values to WC and WT, a fine similarity score fusion can be gained.

D. Similarity score fusion using genetic algorithm

Assigning the weighs of the similarity scores affects directly the retrieval performance of the system. It can be considered as an optimization problem to assign reasonably the weights of color feature similarity score and texture feature similarity score. That is to find the optimum in weight value space. So, this problem can be resolved by genetic algorithm.

E. Determination of fitness function

The fitness of individuals can be evaluated as follows. According to the weights WC and WT of N individuals, we can get N groups of images retrieval results. For every group, the top M images are considered. Total number of images is MN. By calculating occurrence frequency of images of every group in all images, the fitness of every individual is evaluated. Specific operations are as follows. Let N_{ikj} denote if k_{th} image A_{ik} of i_{th} group G_i is in j_{th} group G_j or not. That can be formulated as

$$N_{ikj}=\begin{cases}1, & A_{ik}\in G_j\\ 0, & A_{ik}\notin G_j\end{cases}$$

Then the occurrence frequency of k_{th} image A_{ik} of i_{th} group G_i in all MN images is

$$N_{ik}=\sum_{j=0}^{N}N_{ikj}$$

The occurrence frequency of all images of i_{th} group G_i in all MN images is

$$N_i=\sum_{k=1}^{M}N_{ik}$$

The normalized version of it is

$$P_i=\frac{N_i}{\sum_{l=0}^{N}N_l}$$

ISSN: 2278 – 909X

The bigger Pi indicates that the images in i_{th} group G_i possess a high proportion in all MN images, and the solution is considered a good one.

V. EXPERIMENT AND ANALYSIS

In order to measure retrieval effectiveness for an image retrieval system, precision and recall values are used. For comparison, the image retrieval methods based on color feature , texture feature and two-feature similarity score fusions with equal weights are implemented. Precision and recall for image retrieval can be defined as follows

$$\text{Precision} = \frac{\text{Number of relevant images}}{\text{Total number of images retrieved}}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in database retrieved}}$$

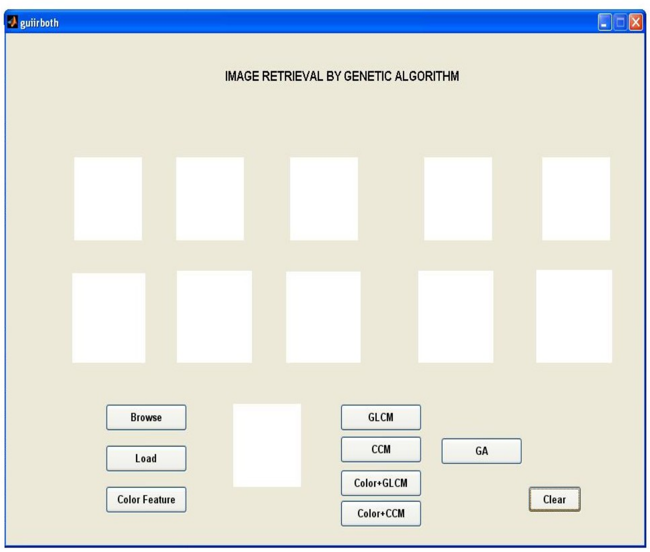


Fig: 1GUI with basic push buttons

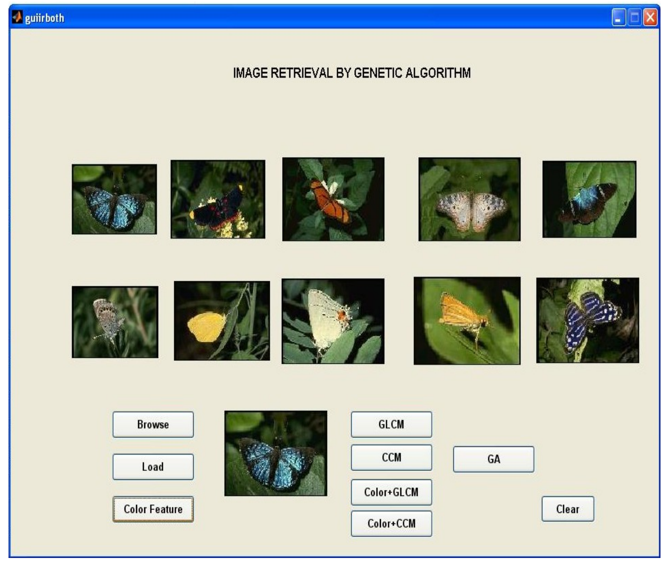


Fig: 2 GUI for image retrieved by color feature

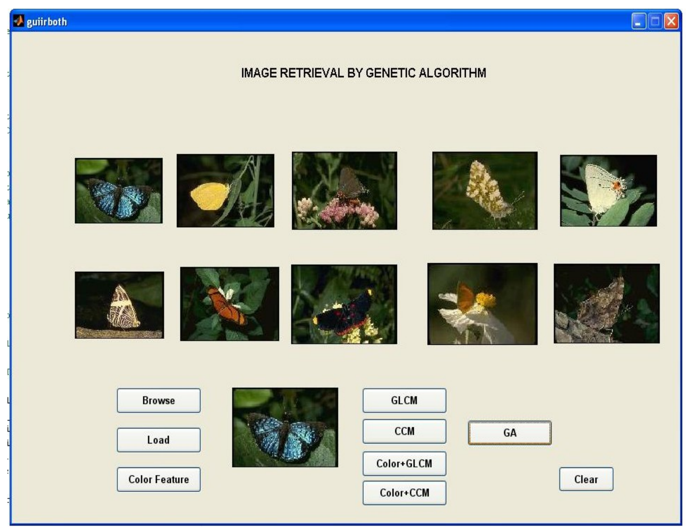


Fig: 3 GUI for image retrieved by GA

VI. CONCLUSIONS

More precise and faster retrieval techniques are needed to access the large image archives being generated, for finding similar images. In order to achieve efficient image retrieval system, a genetic algorithm, multi-feature similarity score fusion of color and texture features are proposed. Precision and recall parameters are used to analyze the image retrieval performance. High resolution images are taken long time to extract the color and text features. In future the above problem will be avoided by using appropriate algorithms for fusing different features.

References

[1] H. Yu, M. Li, H.-J. Zhang, J. Feng, “Color texture moments for

ISSN: 2278 – 909X

content-based image retrieval," In International Conference on Image Processing, pp. 24-28, 2002

[2] Irena Valova, Boris Rachev "Retrieval by color features in image databases", International Conference on Computer Systems and Technologies - CompSysTech' 2006

[3] T.M.Lehmann, M.O. Guld, C.Thies, B.Fischer, K.Spitzer, D.Keysers," Content based image retriveal in medical applications" , computerized medical imaging and graphics, vol.29, no.2,pp.143-155,2005

[4] MladenJovi´c,Yutaka Hatakeyama,Fangyan Dong, and Kaoru Hirota,"Image Retrieval Based on Similarity ScoreFusion from Feature Similarity Ranking Lists", Fuzzy systems and Knowledge Discovery, pp. 461-470, 2006

[5] SelimAksoy and Robert M.Haralick,"Textural Features for image database Retrieval", IEEE Transactions on pattern Analysis and Machine Intelligence,18(80:837-842,August 1988

[6] M.V.Sudhamani, C.R.Venugopal,"Grouping and Indexing Color Features for Efficient Image Retrieval", International Journal of Applied Mathematics and Computer Sciences Volume 4 Number 3 March 25, 2007

[7] Szabolcs Sergyán,"Color Content-based Image Classification" ,5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics january 25-26, 2007

AUTHORS

First Author Mr.S.Muralikaruppaiyan born in Pudukkottai district, Tamilnadu, India. He received his B.E degree from Shri Andal Alagar College of Engineering and he has completed his masters degree in Sudharsan Engineering College. He has presented several papers in various conferences.

Second Author Ms.R.Devi born in Thanjavur district, Tamilnadu, India. She received her B.Tech degree from SASTRA University andshe has completed her masters degree in Periyar Maniammai University. She has presented several papers in various conferences.

ISSN: 2278 – 909X