ISSN: 2320-1363

An Efficient Image Retrieval Based on Color, Texture (GLCM & CCM) features, and Genetic-Algorithm

Sapthagiri.k*1, Manickam.L*2

M.Tech (CS) Student, Department of ECE, NIE, Macherla, Guntur (Dist), AP, India

Associate Professor, Department of ECE, NIE, Macherla, Guntur (Dist), AP, India

Abstract

With the rapid development of multimedia and network technology, people can access a large number of multimedia information. For people who want to make full use of multimedia information resources, the primary question is how to query the multimedia information of interest. Text query can be applied to multimedia information retrieval, but it has inherent deficiencies. One hand, text annotation of multimedia information will spend a lot of manpower and resources and it is inefficient. On the other hand, annotated text is usually a person's perception of multimedia information. It is subject to impact of individual difference and state of human and environment, and the described results may be more one-sided. In addition, it is clearly incomplete to describe content-rich multimedia information with a small amount of text. Content Based Image Retrieval (CBIR) techniques appeared in 1990s. It solves the above problems well. It uses low-level features like color, texture and shape to describe image content, and breaks through the limitation of traditional text query technique. This paper proposes an image retrieval method based on multi-feature similarity score fusion using genetic algorithm. 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. For comparison, other three methods are implemented. They are image retrieval based on color feature, texture feature and fusion of color-texture feature similarity score with equal weights.

Keywords: CBIR, HSV color space, Texture features (GLCM & CCM), Multi-features.

ISSN: 2320-1363

I. Introduction

As processors become increasingly powerful and memories become increasingly cheaper the deployments of large image database for a variety of applications have now become realizable. Databases of art works satellite and medical imaginary have been attracting more and more users in various professional fields for example, geography, medicine, architecture, adverting, design, fashion, and publishing. Effectively and efficiently accessing desired images from large ad varied images databases is now a necessity.

Currently rapid and effective searching for desired images from large-scale image database becomes an important and challenging research topic. Image retrieval technology overcomes the defects of traditional text based image retrieval technology such as heavy workload and strong subjectivity. With a rapid development of multimedia and network technology people can access a large number of multimedia information. For people who want to make full use of multimedia information resources, the primary question is how to query the multimedia information of interest. The query can be applied to multimedia information retrieval but it has inherent deficiencies. One hand text annotation of multimedia information will spend a lot of manpower and resources and it is inefficient. The other hand annotated text is usually a person's perception multimedia information. CBIR system can be implemented based on single feature. Single feature describes the content of an image from a specific angle. It may be suitable for some images, but is also may be difficult to describe other images. Moreover describing an image with single feature is also incomplete. Representing an with multi feature from multi-angle is expected to achieve better results. The problem how to organize multi-sources information in a suitable way to achieve the intended results attracts extensive attention from the researchers in the field. The rest of this paper is organized as follows section describes the extraction method of color feature and texture feature. Section introduced the strategy to fuse multi-feature similarity score. The method assigning weights of multi feature similarity score by genetic algorithm.

2. Related Work

2.1 Image feature extraction

The image content is mainly embodied in color, texture and shape etc. the color feature and shape feature describe the image content from different angles. More features will provide more information on the image content. This paper focuses on

ISSN: 2320-1363

fusion method of multi feature similarity score. This paper discusses the fusion method of two features similarity score. Without lose of generality the used features are color feature and texture feature. The following part describes the used extraction method of color feature and texture feature.

2.2 Color feature extraction

Several methods for retrieving images on the basics of color similarity have been describe in the literature but most are variations on the same basic idea. Each image is added to the collection is analyzed to compute a color histogram which shows the proportion of pixels of each color within the image. The color histogram for each image is then stored in database. At search time the user can either specify the desired proportion of each color or submit an example image from which a color histogram is calculated. The way matching process then retrieves those images whose color histogram match those of the query most closely. The matching technique most commonly used, histogram intersection and HSV. The HSV color model forms a uniform color space, which uses a linear gauge. The perspective 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. In this paper the color histogram in HVS color space is taken as the color feature describing image content. For calculating color histogram in HSV color space, HSV color space must be quantified. According to human cognitive about color three components of HSV space are quantified in non uniform manner.

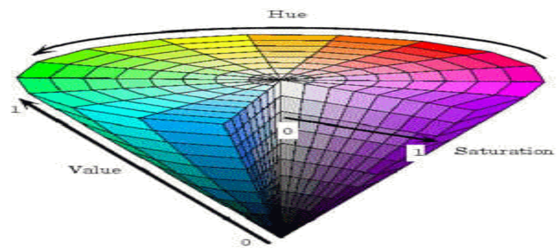


Fig1. Quantized HSV

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

$$C = 16H + 4S+V \qquad (1).$$

Where C is an integer between 0 and 255. Thus the color feature can be obtained by calculating histogram of an image in HSV space.

3. Methodology

3.1 Texture feature extraction based on GLCM

ISSN: 2320-1363

The statistical properties of image co-occurrence matrix are taken as texture features of an image. Firstly color image is converted to grayscale images 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, correlation and local stationary. All these statistical properties are calculated in four directions, so we can get 20 texture features. At last, we calculate the means and variance of these five kind of statistical properties, and took the results as the ultimate texture features denoted as.

$$T=(\mu_1, \mu_2, \mu_3, \mu_4 : \sigma_1, \sigma_2, \sigma_3, \sigma_4) \qquad (2)$$

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 $p(i,j|d,\theta)$ which expresses the probability of the couple pixels 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, it's level is determined by the image gray level. Elements in the matrix are computed by the equation showed as follows:

$$p(i,j|d,\theta) = p(i,j|d,\theta)/\sum_i . \sum_j p(i,j|d,\theta) \qquad (3)$$

GLCM expresses the texture feature according the correlation of the couple pixels gray level at different positions. It quantificational describes the texture feature. In this paper four features is selected including energy, contrast, entropy, inverse difference.

Energy $$E=\sum_{x} . \sum_{y} p(x,y)2 \qquad (4)$$

It is a gray scale image texture measure of homogeneity changing, reflecting the distribution of images gray-scale uniformity of weight and texture.

Contrast $$I = \sum . \sum (x-y)2 \; p(x,y) \qquad (5)$$

Contrast is the main diagonal near the moment of inertia. Which measure the value of matrix is distributed and images of local changes in number, reflecting the image clarity and texture of shadow depth. Contrast is large means texture is deeper.

Entropy $$S= - \sum_{x} . \sum_{y} p(x,y) log p(x,y) \qquad (6)$$

Entropy measures image texture randomness, when the space co-occurrence matrix for all values are equals it achieved the minimum value on the other hand if the value of co-occurrence matrix is very uneven its value is greater. Therefore the maximum entropy implied by the image gray distribution as random.

3.2 Feature extraction based on CCM

ISSN: 2320-1363

Assuming color images is divided into N×N image sub-block, for anyone image sub-block $T_{(i,j)}(1\leq i\leq N, 1\leq j\leq N)$, using the main color image extraction algorithm to calculate the main color $C_{(i,j)}$. for any two 4-connected image sub-block $T_{(i,j)}$ and $T_{(i,j)}(|i-k|)=1$ and j=l; or |j-l|=1 and i=k), if its corresponds to the main color and in the HSV space to meet the following condition.

The set S corresponds to the color connected region. For each color connected region{R_i} (1 ≤I ≤M), the color component R, G in RGB color space and H in HVS color space are respectively extracted the CCM at the direction δ=1;θ=0°,45°,90°,135°. The statistics features extracted from CCM are as follows:

Energy $E=\sum_{i=1}^{D}.\sum_{j=1}^{D}[m(i,j)]^2$ (7)

Contrast $I=\sum_{i}^{D}.\sum_{j}^{D}(i-j)^2\times m(i,j)$ (8)

Entropy $S=-\sum_{i}^{D}.\sum_{j}^{D}m(i,j)\times\log[m(i,j)]$ (9)

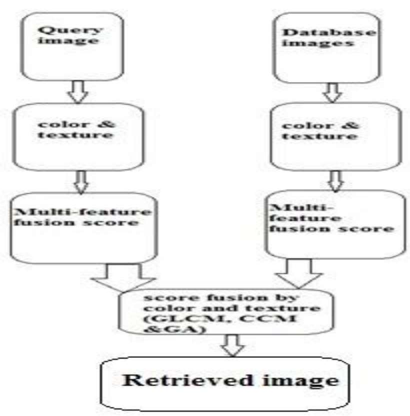
Where if m(i,j)=0, log[m(i,j)]=0

Through this method we can get a 16 dimensional texture feature for component R, G, H and I, each component correspond to four statistics values E,I,S and H.

IV. Experiment and analysis

In this paper experimental data set contains 1000 images from corel database of images, divided into 10 categories, each category has 100 images. Experimental images covers a wealthy of content, including landscapes, animals, plants, monuments, transport (cars, planes) and so on. Selection of each type in the 80 images as training samples, 20 samples for testing.

In section 3 we study two kinds of feature extraction techniques: feature extraction techniques based on the HSV color space and texture features extraction technology. At texture features extraction techniques, we introduce two different extraction methods the gray co-occurrence matrix (GLCM) and CCM. Color and texture are just in part describing the characteristics of images. Image database varies some images dramatic ups and downs in gray-level showing a very strong texture characteristics, and some images from a number of smooth but the colors are different regional composition.



ISSN: 2320-1363

Fig 3.2.1. Algorithm scheme

Mainly image library is used to evaluate the proposed algorithm. It contains dozens categories, each of which has about 60 images. The total number of images is about 1000. They are flower, waterfowl, automobile, beach, mountain, loin, bridge, butterfly, plane, building and tiger and so forth. The color feature and texture feature of every image are extracted to build feature database. Precision rate and recall rate are employed to evaluate the performance of the proposed method. The numbers of returned images are 5, 10, 15, 20, 25, 30, 35, 40, 45 and 50.

P=r/n= No. of relevant images retrieval / total no. of images retrieved (10)

R=r/m= No. of relevant images retrieval / total no. of images in DB (11)

For comparison the image retrieval methods based on color feature, texture feature and two feature similarity score fusion with equal weights are implemented. The precision rate and recall rate relationship of these methods.

Here I combined color features and two different kinds of texture features separately. The following experiments are part of search part of search results and test statistics. Figure 3, 4 for the same retrieval images were compares figure 5 for the classic picture plain conducted a search Table 1 shows the image retrieval result. In this table, recall and precision are compared. Obviously in three feature extraction mode, integrated feature has effective performance, especially, color combined with CCM is better than others.

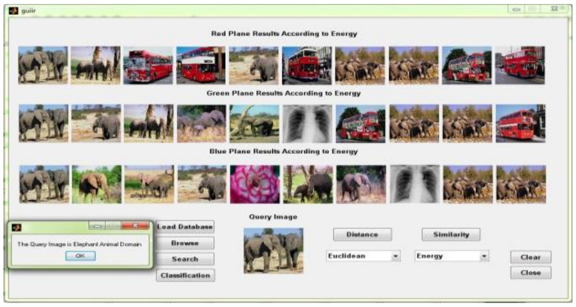


Fig3.2.2 Retrieval result based on distance and similarity of text query image.

Fig3.2.3. Retrieval result based on HSV color space.

ISSN: 2320-1363

Fig3.2.4 Retrieval result based on color + GLCM.

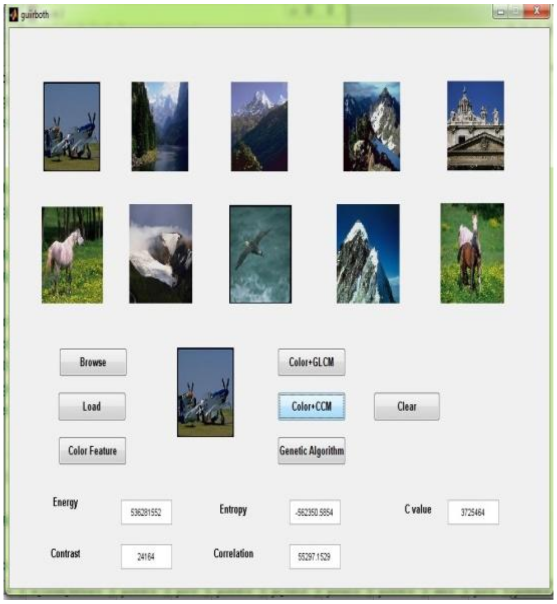


Fig3.2.5. Retrieval result based on color + CCM.

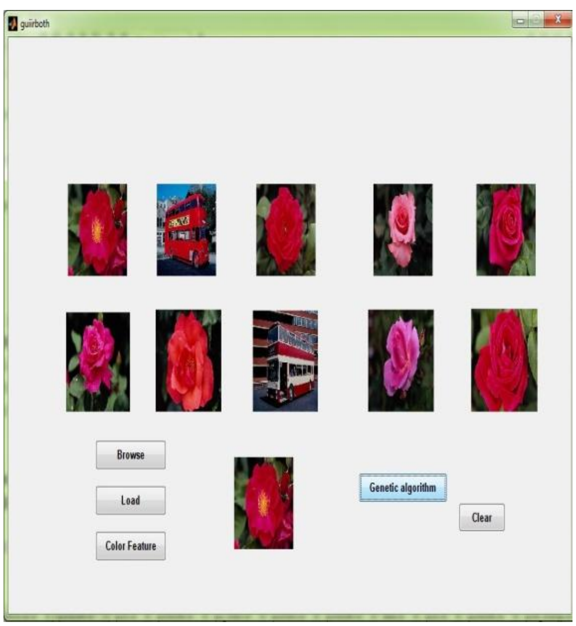
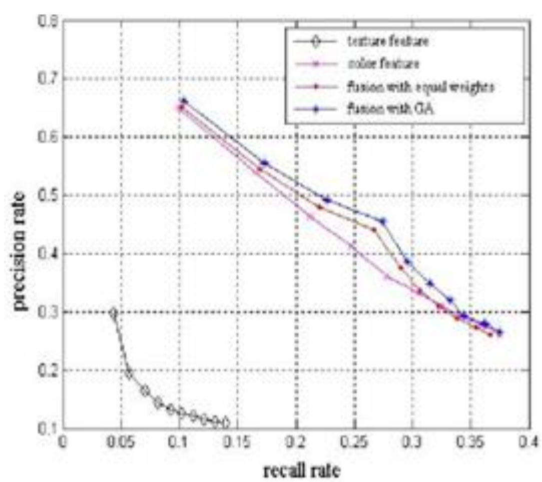


Fig3.2.6. Retrieval result based on multi features (Genetic algorithm).

Table 3.2.1Retrieval result contrast

Retrieval mode	Recall (%)	Precision (%)
color	32.1	44.6
Color + GLCM	35.6	47.5
Color + CCM	38.8	50.2
Genetic algorithm	44.1	55.2

Figure.8 showed that image retrieval method based on multi feature similarity score fusion using genetic algorithm ranked the first. The method with equal weight followed. The method based on color feature is better than the method based on texture feature. The relationship showed on figure 8 reflects mainly the impact of multi feature similarity score fusion based image retrieval method performance of these two image retrieval methods based on multi feature similarity score fusion doesn't increase much.



ISSN: 2320-1363

Fig3.2.7. Relationship diagram of precision rate and recall rate.

4. Conclusion

This paper presents an approach based on HSV color space and texture characteristics of the image retrieval through the quantification of HSV color space, I combine color feature and gray level co-occurrence matrix as well as CCM separately, using normalized Euclidean distance classifier. This paper proposed an image retrieval method based multi-feature similarity score fusion. Then using genetic algorithm multi-feature similarity score are fused and better image retrieval results are gained. However the location of an image in retrieval result reflects directly the similarity of it and query image. So this factor should be taken into account when evaluating the fitness of individual, which is my future work.

5. References

[1]Rui,Y.,Huang,T.S.,Mehrotra,S. [Sharad],"Retrieval with relevance feedback in MARS", In Proc of the IEEE Int'1 Conf. on Image Processing, New York, pp. 815-818, 2012.

[2] H. T. Shen, B. C. Ooi, K. L. Tan, "Giving meanings to www images" Proceedings of ACM Multimedia, pp. 39-48, 2011.

[3] B S Manjunath, W Y Ma, "Texture feature for browsing and retrieval of image data", IEEE Transaction On PAMI, Vol 18, No. 8, pp.837-842,2006.

[4] Y. A. Aslandogan and C. T. Yu, "Techniques and Systems for Image and Video Retrieval,"IEEE Transactions on Knowledge and Data Engineering, Vol. 11, Issue 1, pp. 56-63, Jan/Feb 2009.

[5] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, Issue 12, pp. 1349-1380,

[6] M. J. Swain and D. H. Ballard, „„Color indexing, International Journal of Computer Vision, vol. 7, Issue 1, pp. 11-32,2001.

[7] Irena Valova, Boris Rachev and Michael Vassilakopoulos, "Optimization of the Algorithmfor Image Retrieval by Color Features", International Conference on Computer Systems and Technologies-CompSysTech, pp 1-4, 2006.

[8] Sarfraz and M. Ridha "Content-Based Image Retrieval Using Multiple Shape Descriptors", IEEE/ACS International Conference On Computer Systems and Applications, pp.730-737, 2007.

[9] I. Markov, N. Vassilieva, "Image Retrieval: Color and Texture Combining Based on Query-Image," ICISP 2008, LNCS 5099, Springer-Verlag Berlin Heidelberg, 2008, pp. 430–438.

[10] M. Jovic, Y. Hatakeyama, F. Dong, K. Hirota, "Image Retrieval Based on Similarity Score Fusion from Feature Similarity Ranking Lists," 3rd International Conference on Fuzzy Systems and

ISSN: 2320-1363

Knowledge Discovery (FSKD 2006), LNAI 4223, Springer-Verlag Berlin Heidelberg, 2006, pp. 461–470.

He was working as a Associate professor, Head Of Department (HOD) of Electronics and Communication Engineering in Newton's institute of engineering macherla, in Andhra Pradesh, India. His special field of interest in image processing.

Sapthagiri.K was born in Durgi in the Andhra Pradesh, India on March 23, 1990. He graduated from EVM College of engineering & technology under JNTUK University and pursuing M.Tech from Newton's institute of engineering under JNTUK University. He was mainly interested MATLAB in image processing and signal processing