

Image Clustering and Retrieval using Image Mining Techniques

A.Kannan¹ Dr.V.Mohan² Dr.N.Anbazhagan³

¹Associate Professor, Department of MCA, K.L.N. College of Engineering,

Sivagangai District, Tamilnadu, India – 630611

kannamca@yahoo.com

²Professor & Head, Department of Mathematics, Thiagarajar College of Engineering,

Madurai, Tamilnadu, India - 625 015

vmohan@tce.edu

³Reader, Department of Maths, Alagappa University, Karaikudi-04, Tamilnadu, India,

anbazhagan_n@yahoo.co.in

Abstract: Image retrieval is the basic requirement task in the present scenario. Content Based Image Retrieval is the popular image retrieval system by which the target image to be retrieved based on the useful features of the given image. In other end, image mining is the arising concept which can be used to extract potential information from the general collection of images. Target or close Images can be retrieved in a little fast if it is clustered in a right manner. In this paper, the

concepts of CBIR and Image mining have been combined and a new clustering technique has been introduced in order to increase the speed of the image retrieval system.

KEY WORDS: Content Based Image Retrieval, RGB Components, Texture, Entropy.

1.0 INTRODUCTION

In this present scenario, image plays vital role in every aspect of business such as business images, satellite images, medical images and so on. If we analysis these data, which can reveal useful information to the human users. But, unfortunately there are certain difficulties to gather those data in a right way [1]. Due to incomplete data, the information gathered is not processed further for any conclusion.

In another end, Image retrieval is the fast growing and challenging research area with regard to both still and moving images. Many Content Based Image Retrieval (CBIR) system prototypes have been proposed and few are used as commercial systems. CBIR aims at searching image databases for specific images that are similar to a given query image. It also focuses at developing new techniques that support effective searching and browsing of large digital image libraries based on automatically derived imagery features. It is a rapidly expanding research area situated at the intersection of databases, information retrieval, and computer vision. Although CBIR is still immature, there has been abundance of prior work.

The CBIR focuses on Image ‘features’ to enable the query and have been the recent focus of studies of image databases. The features further can be classified as low-level and high-level features. Users can query example images based on these features such as texture, colour, shape, region and others. By similarity comparison the target image from the image repository is retrieved. Meanwhile, the next important phase today is focused on clustering techniques. Clustering algorithms can offer superior organization of multidimensional data for effective retrieval. Clustering algorithms allow a nearest-neighbour search to be efficiently performed.

Hence, the image mining is rapidly gaining more attention among the researchers in the field of data mining, information retrieval and multimedia databases. Spatial Databases is the one of the concepts which plays a major role in Multimedia System. Researches can extract semantically meaningful information from image data are increasingly in demand.

1.1 Comparison of Image Mining with other Techniques

Image mining normally deals with the extraction of implicit knowledge, image data relationship, or other

patterns not explicitly stored from the low-level computer vision and image processing techniques. i.e.) the focus of image mining is in the extraction of patterns from a large collection of images, the focus of computer vision and image processing techniques is in understanding or extracting specific features from a single image.

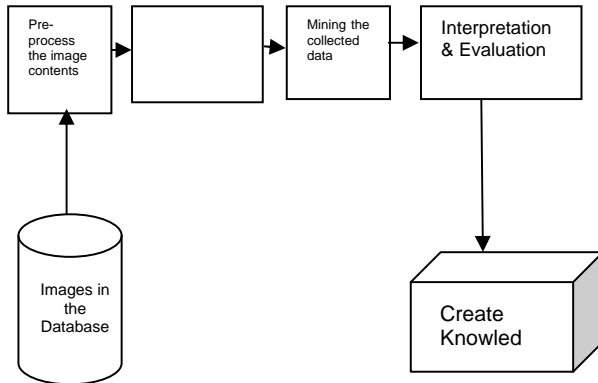


Figure 1 Image Mining Processes [1]

Figure 1.1 shows the image mining process. The images from an image database are first preprocessed to improve their quality. These images then undergo various transformations and feature extraction to generate the important features from the images. With the generated features, mining can be carried out using data mining techniques to discover significant patterns. The resulting patterns are evaluated and interpreted to obtain the final knowledge, which can be applied to applications. [1]

2.0 PROBLEM DEFINITION

In the colour based image retrieval the RGB Colour model is used. Colour images normally are in three dimensional. RGB colour components are taken from each and every image. Then, the mean values of Red, Green, and Blue components of target images are calculated and stored in the database. Based on the RGB component mean values, the images are clustered as Red, Green and Blue major component categories. These three mean values for each image are stored and considered as features.

Then the top ranked images are re-grouped according to their texture features. In the texture-based approach the parameters gathered are on the basis of statistical approach. Statistical features of grey levels were one of the efficient methods to classify texture. The Grey Level Co-occurrence Matrix (GLCM) is used to extract second order statistics from an image. GLCMs have been used very successfully for texture calculations [9]. The different texture parameters like entropy, contrast, dissimilarity, homogeneity,

standard deviation, mean, and variance of both query image and target images are calculated. From the calculated values the required image from the repository is extracted.

Then, the pre-processed images in the database are classified as low-texture, average-texture and high-texture detailed images respectively based on some factor like MLE (Maximum Likelihood Estimation) estimation. The classified images are then subject to colour feature extraction. The retrieved result is pre-clustered by Fuzzy-C means technique. This is followed by GLCM texture parameter extraction where the texture factors like contrast, correlation, mean, variance and standard variance are mined. The resulted values of both the query image and target images are compared by Euclidean distance method.

2.1 Proposed Solution

In this, a new method for image classification is formulated in order to reduce the searching time of images from the image database. The coarse content of image is grouped under three categories as:

- (i) High-texture detailed Image
- (ii) Average-texture detailed Image
- (iii) Low-texture detailed Image

Thereby, we can reduce the search space by one third of what was earlier. If we go more number of groups or less number of groups, they may reveal unnecessary overlapping overhead problems or may produce approximate results.

So, the main focus on this classification is by making use of “textures” present in an image. This is because this texture-based classification is simple, easy and efficient for real time applications as compared to classifications based on Entropy method as well as segmentation based techniques.

2.2 Image Retrieval

Image Retrieval from the image collections involved with the following steps

- Pre-processing
- Image Classification based on some true factor
- RGB Components processing
- Preclustering
- Texture feature extraction

- Similarity comparison
- Target image selection

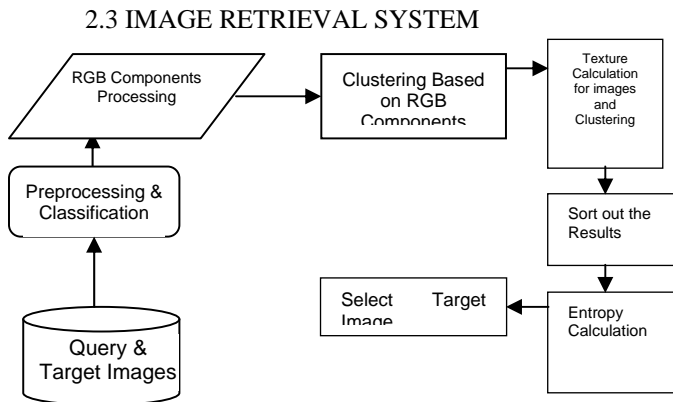


Figure 2.0 Block Diagram of Image Retrieval System

2.4 Pre-processing & Noise Reduction Filtering

Pre-processing is the name used for operations on images at the lowest level of abstraction. The aim of the pre-processing is an improvement of the image that suppresses unwilling distortions or enhances some image features, which is important for future processing of the images. This step focuses on image feature processing. Filtering is a technique for modifying or enhancing an image. The image is filtered to emphasize certain features or remove other features. The noise in the images is filtered using linear and non-linear filtering techniques. Median filtering is used here to reduce the noise.[12]

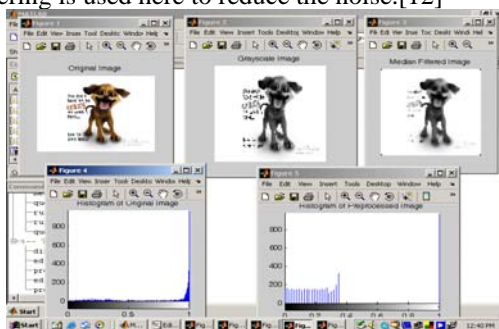


Figure 2.1 Results for Pre-processing Image

2.6 RGB Components Processing

An RGB colour images is an M*N*3 array of colour pixels, where each colour pixel is a triplet corresponding to the red, green, and blue components of an image at a spatial location. An RGB image can be viewed as the stack of three gray scale images that, when fed into the red, green, blue inputs of a colour monitor, produce the colour image on the

screen. By convention the three images form an RGB images are called as red, green and blue components.

The average values for the RGB components are calculated for all images

$$\text{Red average} = \frac{\text{sum of all the Red Pixels in the image } R(P)}{\text{No. Of pixels in the image } P}$$

$$\text{Green average} = \frac{\text{sum of all the Green Pixels in the image } G(P)}{\text{No. Of pixels in the image } P}$$

$$\text{B average} = \frac{\text{sum of all the Blue Pixels in the image } B(P)}{\text{No. Of pixels in the image } P}$$

Where R (P) = RED component pixels,
 G (P) = GREEN component pixels,
 B (P) = BLUE component pixels,
 P =No. of pixels in the image.

After calculating the mean values of Red, Blue and Green components, the values are to be compared with each other in order to find the maximum value of the components. For eg., if the value of Red component is High than the rest of the two, then we can conclude that the respective image is Red Intensity oriented image and which can be clustered into Red Group of Images.

Whenever the query image is given, calculate the RGB components average values. Then compare this with the stored values.



Figure 2.2 Result of RGB Components Clustering Images

2.7 Entropy Classification

The texture represents the energy content of the image. If an image contains more and high textures, then the energy will be high as compared to that of average and low texture images. There are several texture parameters to be considered [12]. However, here, the texture parameter Entropy is highly focused

and which is to be calculated for the query and target images. Entropy is a statistical measure of randomness that can be used to characterize the texture of the input image. Entropy is defined as $-\sum(hc . \log_2(hc))$

where hc is the histogram counts obtained from the histogram calculation.

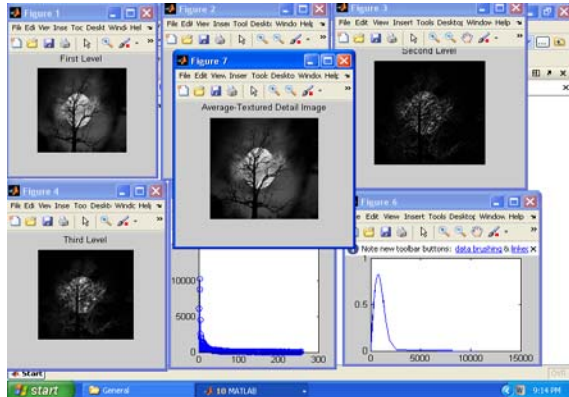


Figure 2.3 Processes for High and Average Texture Analysis

2.8 Image Clustering

Clustering will be more advantage for reducing the searching time of images in the database. Fuzzy C-means (FCM) is one of the clustering methods which allow one piece of data to belong to two or more clusters. In this clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely too just one cluster. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the centre of cluster. FCM groups data in specific number of clusters.

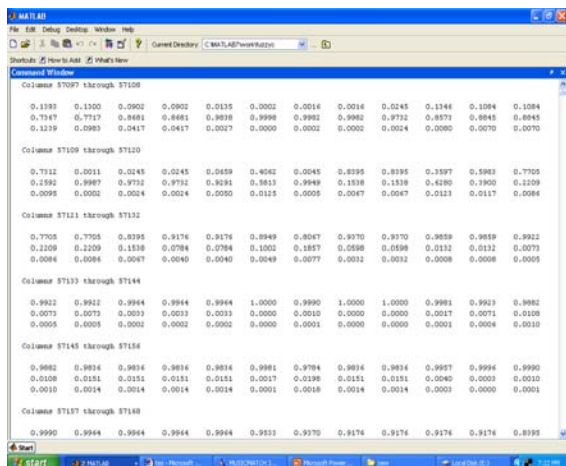


Figure 2.4 Results of Clustering Processes

2.9 Similarity Comparison and Image Retrieval

The given query image is pre-processed and the features of the given query image to be calculated in the usual way. Then, the entropy value of the given query image is calculated based on the calculation given in Sec.2.7. A threshold constant value is to be added with the entropy value of the query image. Hence, the result will be compared with the concerned cluster and the target images are to be retrieved based on the constraints. The results of this process are shown in the figure 2.5.

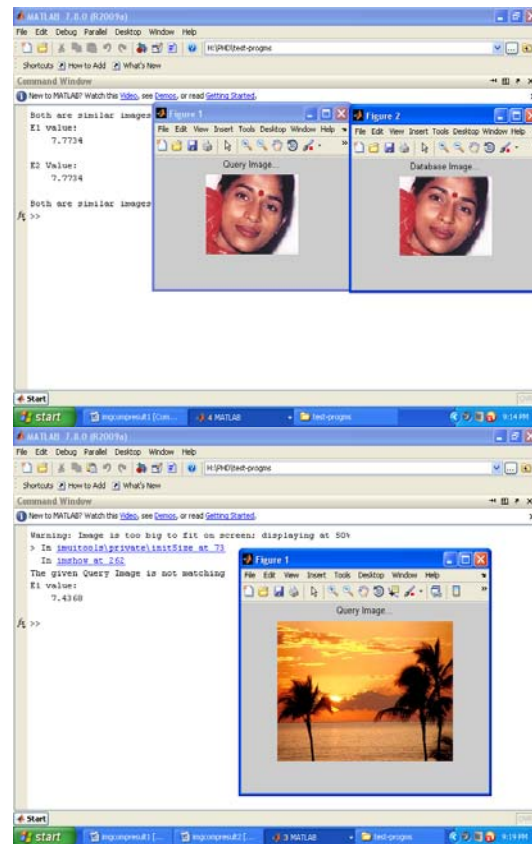


Figure 2.5 Results of Image Retrieval for the given Query Image

3.0 Performance Evaluation of Proposed CBIR System

Evaluation of retrieval performance is a crucial problem in Content-Based Image Retrieval (CBIR). Many different methods for measuring the performance of a system have been created and used by researchers. We have used the most common evaluation methods namely, Precision and Recall

usually presented as a Precision vs Recall graph. Precision and recall alone contain insufficient information. We can always make recall value 1 just by retrieving all images. In a similar way precision value can be kept in a higher value by retrieving only few images or precision and recall should either be used together or the number of images retrieved should be specified.

With this, the following formulae are used for finding Precision and Recall values.

$$\text{Precision} = \frac{\text{No. of Relevant Images Retrieved}}{\text{Total number of Images Retrieved}}$$

$$\text{Recall} = \frac{\text{No. of relevant images retrieved}}{\text{Total no. of relevant images in the database}}$$

4.0 CONCLUSION

The main objective of the image mining is to remove the data loss and extracting the meaningful information to the human expected needs. The images are preprocessed with various techniques and the texture calculation is highly focused. Here, images are clustered based on RGB Components, Texture values and Fuzzy C mean algorithm. Entropy is used to compare the images with some threshold constraints. This application can be used in future to classify the medical images in order to diagnose the right disease verified earlier.

REFERENCES

1. Image Mining: Trends and Developments, Ji Zhang Wynne Hsu Mong Li Lee
2. U. M. Fayyad, S. G. Djorgovski, and N. Weir: Automating the Analysis and ataloging of Sky Surveys. *Advances in Knowledge Discovery and Data Mining*, 471-493, 1996.
3. W. Hsu, M. L. Lee and K. G. Goh. Image Mining in IRIS: Integrated Retinal Information System (Demo), in *Proceedings of ACM SIGMOD International Conference on theManagement of Data*, Dallas, Texas, May 2000.
4. A. Kitamoto. Data Mining for Typhoon Image Collection. In *Proceedings of the SecondInternational Workshop on Multimedia Data Mining (MDM/KDD'2001)*, San Francisco, CA,USA, August, 2001.
5. C. Ordonez and E. Omiecinski. Discovering Association Rules Based on Image Content. *Proceedings of the IEEE Advances in Digital Libraries Conference (ADL'99)*, 1999.
6. O. R. Zaiane, J. W. Han et al. Mining MultiMedia Data. *CASCON'98: Meeting of*

Minds, pp 83-96, Toronto, Canada, November 1998.

7. M. C. Burl et al. Mining for Image Content. In *Systemics, Cybernetics, and Informatics / Information Systems: Analysis and Synthesis*, (Orlando, FL), July 1999.
8. M. Datcu and K. Seidel. Image Information Mining: Exploration of Image Content in LargeArchives. *IEEE Conference on Aerospace*, Vol.3, 2000.
9. Yixin Chen, James Z.Wang, Robert Krovetz "Cluster Based Retrieval Of Images by Unsupervised Learning", *IEEE Transaction on Image Processing*, Vol 14, pp.1187-1199, No.8, August 2005.
10. D.S. Zhang and G.Lu, "Content Based Image Retrieval Using Texture Features", In *Proc. of First IEEE Pacific-Rim Conference on Multimedia (PCM'00)* pp.392-395, Sydney, Australia, December2000.
11. Hewayda M.Lofty, Adel S.Elmaghraby "CoIRS: Cluster Oriented Image Retrieval System" *IEEE conference on tools with Artificial Intelligence*, 2004.
12. "Comparison of Feature Selection Techniques for Detection of Malignant Tumor in Brain Images", M. Sasikalal and N. Kumaravel2, *IEEE Indicon 2005 Conference*, Chennai, India, 11 -13 Dc 2005