Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 5, May 2014, pg.523 – 529

RESEARCH ARTICLE

Implementation of Content Base Image Retrieval Using Clustering Technique

[1]Junaid Khan

jkhan34@gmail.com

[2]Prof. S.S. Kulkarni

kulkarniss@gmail.com

ABSTRACT

The purpose of this report is to describe our research and solution to the problem of designing a Content Based Image Retrieval, CBIR system. It outlines the problem, the proposed solution, the final solution and the accomplishments achieved. Due to the enormous increase in image database sizes, as well as its vast deployment in various applications, the need for CBIR development arose. Firstly, this report outlines a description of the primitive features of an image; texture, colour, and shape. These features are extracted and used as the basis for a similarity check between images. The algorithms used to calculate the similarity between extracted features, are then explained. Our final result was a built software application, with an image database, that utilized texture and colour features of the images in the database as the basis of comparison and retrieval. The structure of the final software application is illustrated. Furthermore, the results of its performance are illustrated by a detailed example.

Index Terms: Algorithms, Design, Experimentation, Human Factors, Model, Method

1. INTRODUCTION

The aim of this project is to review the current state of the art in content-based image retrieval (CBIR), a technique for retrieving images on the basis of automatically-derived features such as color, texture and shape. Our findings are based both on a review of the relevant literature and on discussions with researchers in the field.

The need to find image from a collection a desired is shared by many professional groups, including journalists, design engineers and art historians. While the requirements of image users can vary considerably, it can be useful to characterize image queries into three levels of abstraction: primitive features such as color or shape, logical features such as the identity of objects shown and abstract attributes such as the significance of the scenes depicted. While CBIR systems currently operate effectively only at the lowest of these levels, most users demand higher levels of retrieval.

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases. "Content-based" means that the search will analyze the actual contents of the image. The term 'content' in this context might refer colors, shapes, textures, or any other information that can be derived form the image itself. Without the ability to examine image content, searches must rely on metadata such as captions or keywords. Such metadata must be generated by a human and stored alongside each image in the database Problems with traditional methods of image indexing [6] have led to the rise of interest in techniques for retrieving images on the basis of automatically-derived features such as color, texture and shape – a technology now generally referred to as *Content-Based Image Retrieval* (CBIR). However, the technology still lacks maturity, and is not yet being used on a significant scale. In the concepts which are presently used for CBIR system are all under research [10].

Researcher at content base image retrieval has gain tremendous momentum over last decades. A lot of worked carried out on image retrieval by researcher expending on depth or breadth [1]-[5]. The term Content Based Image Retrieval (CBIR) seems to have originated with the work of Kato [6] for the automatic retrieval of the images from a database, based on the color and shape present. Since then, the term has widely been used to describe the process of retrieving desired images from a large collection of database, on the basis of syntactical image features (color, texture and shape). The techniques, tools and algorithms that are used, originate from the fields, such as statistics, pattern recognition, signal processing, data mining and computer vision. CBIR is the most important and effective image retrieval method and widely studied in both academia and industry arena. In this we propose an image retrieval system, called Wavelet-Based Color Histogram Image Retrieval (WBCHIR), based on the combination of color and texture features. The color histogram for color feature and wavelet representation for texture and location information of an image. This reduces the processing time for retrieval of an image with more promising representatives. The extraction of color features from digital images depends on an understanding of the theory of color and the representation of color in digital images. Color spaces are an important component for relating color to its representation in digital form. Absence of hard evidence on the effectiveness of CBIR techniques in practice, opinion is still sharply divided about their usefulness in handling real-life queries in large and diverse image collections.

2. 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 [8].

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 [11.]

3. 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.

3.1 Image Retrieval

Image Retrieval from the image collections involved with the following steps

- Feature extraction
- Image Classification based on some true factor
- RGB Components processing
- Pre-clustering

• Texture feature extraction

• Similarity comparison

• Target image selection

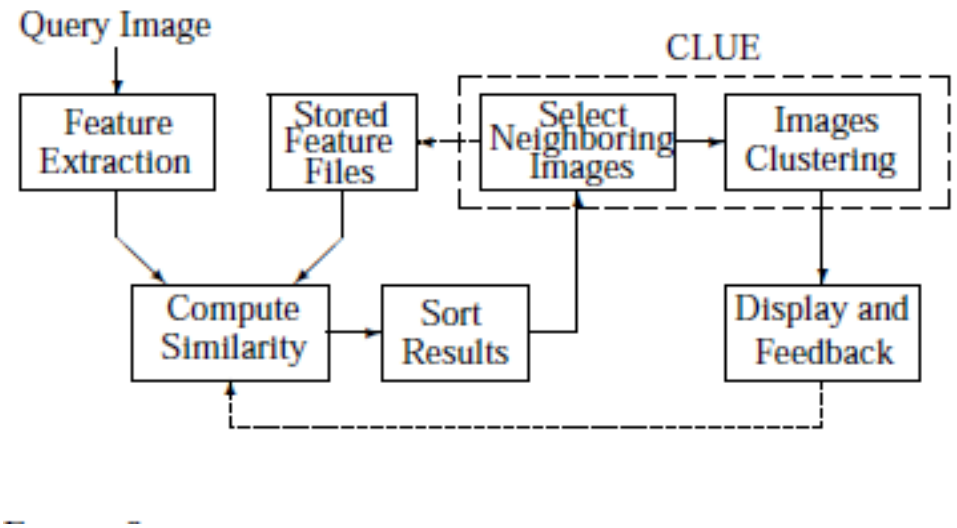

Figure 1.Block Diagram of Image Retrieval System

Feature extraction is the name used for operations on images at the lowest level of abstraction. The aim of the Feature extraction 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. [12]

3.2 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.

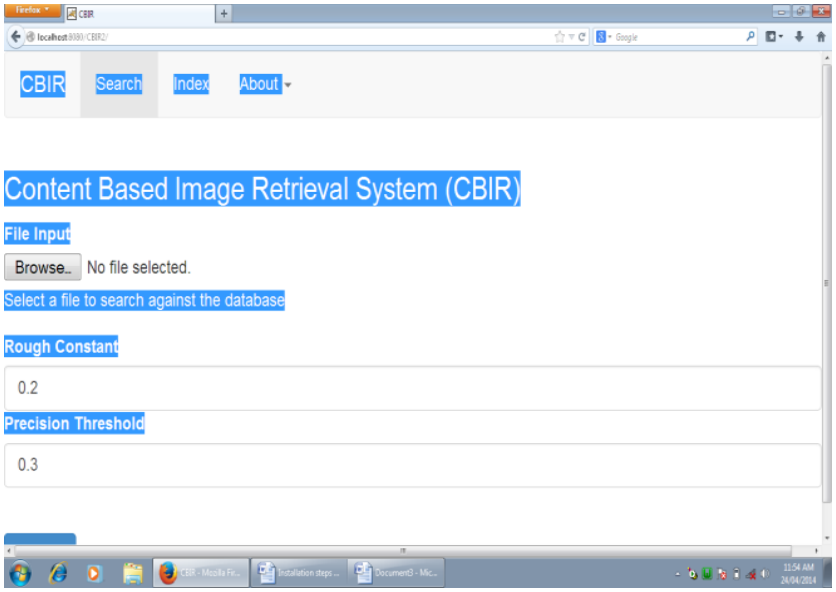

Figure 2. Result of RGB Components Clustering Images

3.4 Image Clustering

Clustering will be more advantage for reducing the searching time of images in the database. K-means clustering algorithm 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. K-M-C groups data in specific number of clusters.

3.3 THE K-MEANS ALGORITHM

Algorithm: k-means. The k-means algorithm for partitioning based on the mean value of the objects in the cluster.
Input: The number of clusters k and a database containing n objects.
Output: A set of k clusters that minimizes the squared-error criterion.
Method:
(1) Arbitrarily choose k objects as the initial cluster centers:
(2) Repeat
(3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
(4) Update the cluster means, i.e., calculate the mean value of the objects for each cluster;
(5) Until no change.

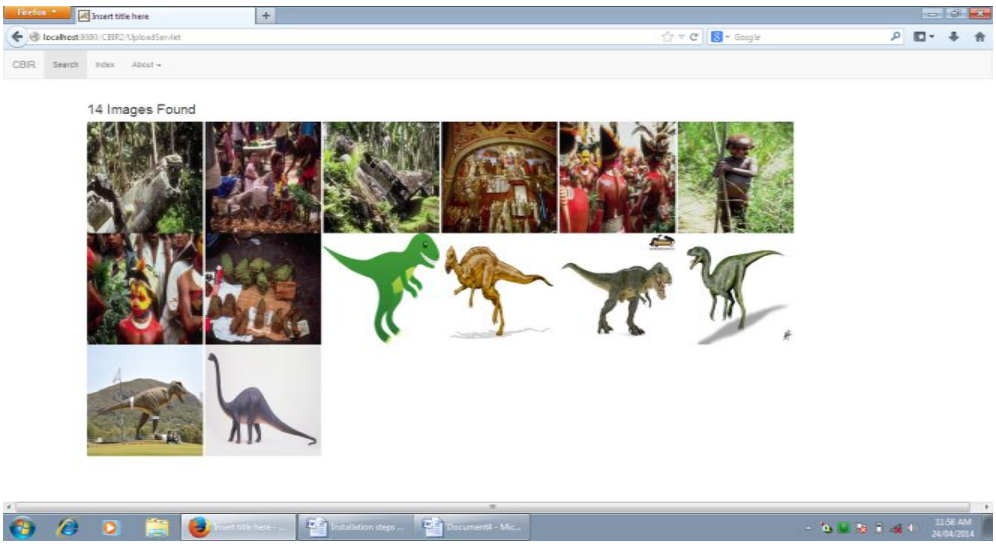

Figure 3.0 Results of Image Retrieval for the given Query Image

4. CONCLUSION

The main objective of the image clustering is to remove the data loss and extracting the meaningful information to the human expected needs. The images are pre-processed with various techniques. Here, images are clustered based on RGB Components, Texture values and K- mean algorithm. This application can be used in future to classify the medical images in order to diagnose the right disease verified earlier.

REFERENCES

[1] R. Datta, D. Joshi, J. Li and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age", ACM computing Survey, vol.40, no.2, pp.1-60, 2008.

[2] J. Eakins and M. Graham, "Content-Based Image Retrieval", Technical report, JISC Technology Applications Programme, 1999.

[3] Y. Rui, T. S. Huang and S.F. Chang, "Image Retrieval: Current Techniques, Promising Directions

[4] A. M. Smeulders, M. Worring and 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, 22(12): pp. 1349-1380, 2000.

[5] Y. Liu, D. Zang, G. Lu and W. Y. Ma, "A survey of content-based image retrieval with high-level semantics", Pattern Recognition, Vol-40, pp-262-282, 2007.

[6] A. Natsev, R. Rastogi and K. Shim, "WALRUS: A Similarity Retrieval Algorithm for Image Databases", In Proceeding. ACM SIGMOD Int. Conf. Management of Data, pp-395–406, 1999.

[7] S. Ardizzoni, I. Bartolini, and M. Patella, "Windsurf: Region based Image Retrieval using Wavelets", In IWOSS'99, pp. 167–173, 1999.

[8] G. V. D. Wouwer, P. Scheunders and D. V. Dyck, "Statistical texture characterization from discrete wavelet representation", IEEE Transactions on Image Processing, Vol.8, pp-592–598, 1999.

[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] R. C. Gonzalez and E.W. Richard, Digital Image Processing, Prentice Hall. 2001.

[11] Yang Changchun, Yi Li, "A Data Mining Model and Methods Based on Multimedia Database," Internet Technology and Applications, 2010 IEEE International Conference on, vol., no., pp.1-4, 20-22 Aug. 2010.

[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.

