International Journal of Engineering Research and General Science Volume 3, Issue 4, Part-2, July-August, 2015
ISSN 2091-2730

An Efficient Content-Based Image Retrieval System Based On Dominant Color Using a Clustered Database

Joby Elsa Abraham, Kavitha V K

P.G Scholar, Dept of CSE, BMCE, Kollam, Kerala, India.

Email : Jobyeabraham@gmail.com

Abstract— Last years the importance and abundance of image retrieval systems had a big growth. During the searching process the color feature of the images is one of the most significant characteristics. Since the color-based comparison and retrieval has a variety of widely used techniques. In this paper we introduce a Dominant Color extraction scheme through which we extract color feature of the image. After extracting the dominant colors we use quantization to represent the extracted dominant colors within a limit. Based on quantized value we cluster the database and do the indexing to store images. Whenever a query image given to the system it will not search the whole database, just identify the cluster and search the image. Similarity measure used in our paper is a modified Euclidean distance measure. With the proposed feature experiments were carried out in image databases, and that was found the precision of retrieval has significantly improved and the time complexity of retrieval of images is reduced.

Keywords— CBIR, Dominant Colors, Dominant Color Descriptor, Similarity Measure, Quantization, Clustering, Indexing, query image, thresholding, RGB color space.

INTRODUCTION

Over the past few years Content-Based Image Retrieval has become an exciting and in-depth area of research. The relevance of visual information retrieval in many areas such as fashion and design, crime prevention, medicine, law, and science makes this research field one of the important and fastest growing in information technology. Image retrieval has come a long way where it started off with text-based retrieval. However, there are many problems associated with retrieving images based on text such as manual annotation of keywords, differences in perceptions and interpretations, and a few others. Due to this, researchers came up with CBIR where images are retrieved based on automatically derived low-level features (human vision related), middle-level features (objects related), or high-level features (semantic related). Among these features, the low-level features are the most popular due to its simplicity. One of the important low-level features is color as it plays an important role in CBIR due to its robustness to complex background and independent of image size and orientation.

In the current version of the MPEG-7 Final Committee Draft, several color descriptors have been approved including number of histogram descriptors and a dominant color descriptor (DCD). MPEG-7 specifies seven color descriptors [3,4]. It includes dominant colors, scalable color histogram, color structure, color layout, and GoF/GoP color. In [15], the authors have shown that the early perception in human visual system performs dominant color identification, and eliminates the fine details and colors in small areas. Therefore, for macroscopic level, human perceive images as a combination of dominant colors no matter how the exact color distribution. In MPEG-7, DCD provides an effective, compact, and intuitive salient color representation, and describe the color distribution in an image or a region of interesting but it lacks certain semantic information. This feature descriptor contains two main components: (1) representative colors and (2) the percentage of each color. These prominent colors and their percentages may only lead to retrieve many dissimilar images that share the same biggest DC. Usually, the dissimilarity occurs when the background color of an image has the largest percentage. Due to this so many enhancements are done to the MPEG-7 DCD. In [8], a semantic feature is added to the DCD to improve its accuracy in an object-based image retrieval application and it is considered as feature level-based solution to the background dominance problem. In this paper, we will develop an effective color extraction scheme for the image retrieval from the large database.

Besides, MP7DCD's quadratic similarity measure (QSM) that is used by Deng et al.[14]and Yamada et al.[13] has some drawbacks. In [10,12], one can see the first simple changes that were made to improve QSM. Authors in [11] also propose a palette histogram similarity measure to solve QSM problem. Moreover in [6], a new similarity measure was proposed to achieve a good performance compared with QSM and all the aforementioned modifications [10–12]. In [8] they used mutual color ratio (MCR), which alleviates their dependencies on the biggest DC. In this present work, a modification will be applied to all the above dissimilarity measures to improve their performance. In order to increase the retrieval we propose a new clustered and indexed database in this work. In all the former works the database may be clustered or indexed but not clustered and indexed which does not assure good performance. In the present work the new database will be developed which assures a good time complexity

The paper is organized in the following way. Section 2 is concerned with explicating dominant color extraction scheme and the similarity measures. Section 3 is mainly concerned with the proposed color extraction scheme, the newly proposed modification that helps improves the similarity measure and introduces new clustered and indexed database. Section 4 illustrates the extensive experiment that contains visual results. Finally comes the conclusion in Section5.

RELATED WORKS

Last decade has witnessed great interest in research on Content-based Image Retrieval. Many techniques have been done with respect to content-based image retrieval (CBIR). Most proposed CBIR [2, 5] techniques automatically extract low-level features (e.g. color, texture, shapes and layout of objects) to measure the similarities among images by comparing the feature differences. Color, texture and shape features have been used for describing image content. Color is one of the most widely used low-level visual features and is invariant to image size and orientation [1]. As conventional color features used in CBIR, there are color histogram, color correlogram, and dominant color descriptor (DCD).

Color histogram is the most commonly used color representation, but it does not include any spatial information. Color correlogram describes the probability of finding color pairs at a fixed pixel distance and provides spatial information. Therefore color correlogram yields better retrieval accuracy in comparison to color histogram. Color autocorrelogram is a subset of color correlogram, which captures the spatial correlation between identical colors only. Since it provides significant computational benefits over color correlogram, it is more suitable for image retrieval. DCD is MPEG-7 color descriptors [3,4]. DCD describes the salient color distributions in an image or a region of interest, and provides an effective, compact, and intuitive representation of colors presented in an image. However, DCD similarity matching does not fit human perception very well, and it will cause incorrect ranks for images with similar color distribution [7, 16]. In [6], 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), Lu et al. [9] uses the color distributions, the mean value and the standard deviation, to represent the global characteristics of the image, and the image bitmap is used to represent the local characteristics of the image for increasing the accuracy of the retrieval system. In [8] they used a weighted DCD for content based image retrieval for increasing the performance of MPEG-7 DCD. In our present work we will replace the DCD by a Dominant Color extraction scheme.

In other hand, similar to the dynamic quantization-based histogram, MP7DCD [13] also uses QSM with some modification to measure the dissimilarity between the query image and database images. However, QSM is not void of serious drawbacks. For instance, it does not match human color perception [6,11]. Therefore, some extensions to QSM have been proposed in Ma et al. [10], Mojsilovic et al. [12]. Po and Wong [11] propose a merging palette histogram for similarity measure (MPHSM). Yang et al. [6] propose a similarity measure that simulates human color perception. In [8] a modified DC-based similarity measure is proposed. In our work new similarity measure is proposed by the separation of RGB values.

Addition to this, in our present work we will introduce a new clustered and indexed database to increase the performance level and this will help to reduce the image retrieval time complexities. In the former works they used only clustering or indexing approach and not both approach together. As we use both approaches it increases the retrieval performance. Other great advantage we considered in our work is reduced time complexity for the retrieval.

PROPOSED SYSTEM

When we are considering Content-based Image Retrieval the foremost operation is feature extraction. The proposed method is based on Dominant Color feature of an image, so the first step is to identify the Dominant Color. Using the quantization technique we limits the dominant color from the image into a certain limit. The images are grouped into different clusters using the quantized values and indexed. Color indexing is used to index the images in the clusters and stores the image into the database. Whenever the query image arrives as the first step we will do the feature extraction and quantization. Most advantage in our work is reduced time complexity in retrieving images from the database because the clusters are identified for query image. So instead of searching the whole database, only need to search the corresponding cluster. A block diagram for the proposed system is shown in the Fig 1.

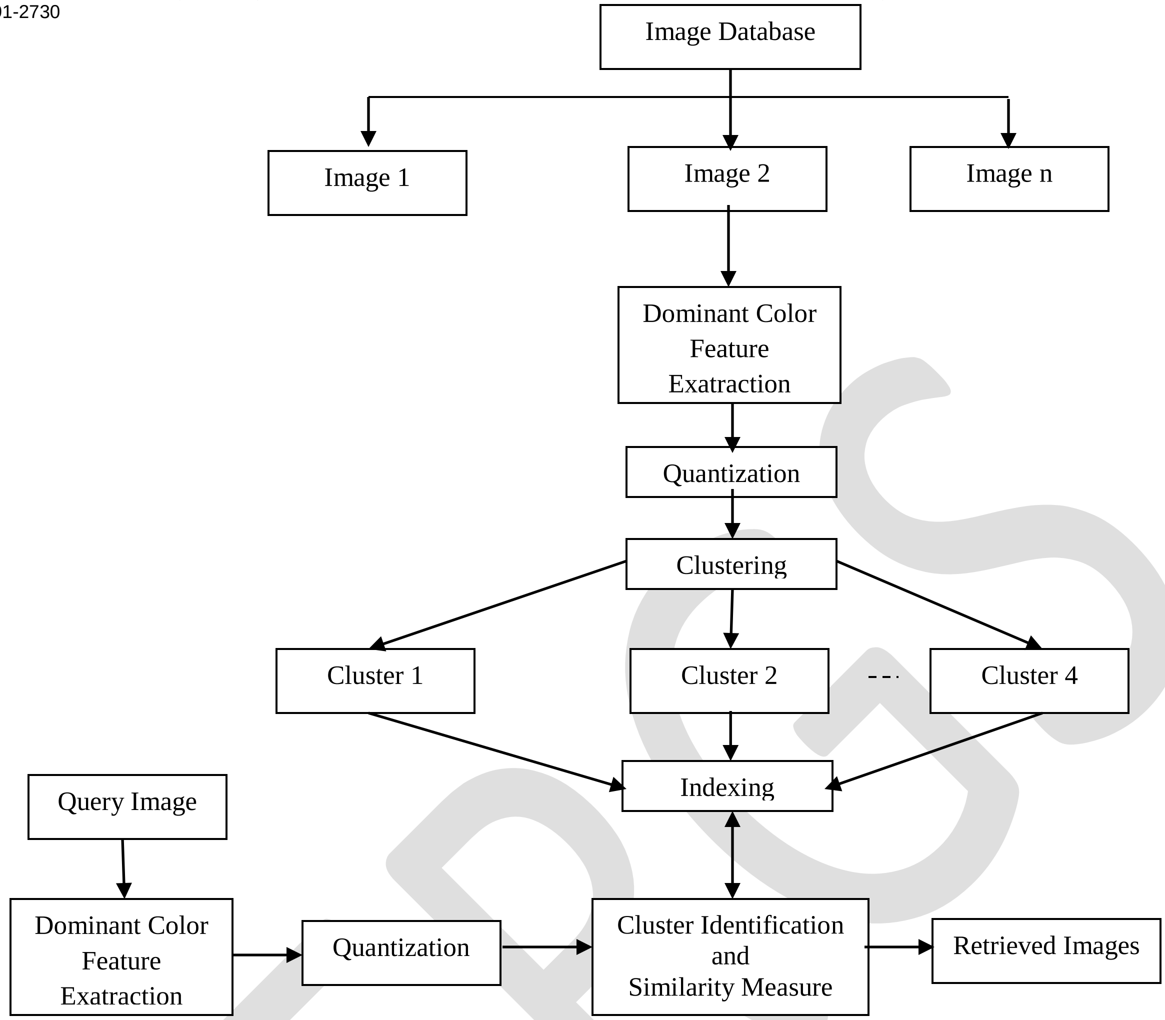

Fig 1 A Block Diagram for Proposed System

Dominant Color Extraction

When we are extracting the color feature of an image, first need is to fix a color space. We are using a RGB color space for the color feature extraction. The separation of RGB is done by using the thresholding technique and stores into R, G and B. Mean value of red, blue and green color are calculated an stored into $m_i(l)$. Among these values the dominant colors are identified by calculating the percentage of each red, green and blue component separately. The color has the highest percentage considered as the dominant color. Dominant Colors are extracted using the formulas given below:

$$DC_r(l) = \frac{m_r(l)}{m_r(l)+m_g(l)+m_b(l)} \times 100$$

$$DC_g(l) = \frac{m_g(l)}{m_r(l)+m_g(l)+m_b(l)} \times 100$$

$$DC_b(l) = \frac{m_b(l)}{m_r(l)+m_g(l)+m_b(l)} \times 100$$

In the above I is the image, $DC_r(l)$, $DC_g(l)$ and $DC_b(l)$ are dominant colors in R, G and B components and $m_r(l)$, $m_g(l)$ and $m_b(l)$ mean values of red, blue green colors respectively. We choose the color with the highest percent as the dominant color and quantized it into eight color values of a fixed limit.

Clustering and Indexing

Clustering done as just the grouping of the images based on the quantized color values. We are giving a different threshold value and cluster the images into different small cluster. Indexing is done to the each cluster as a labeling part. We use color indexing with the help of a color histogram. The images are stored in the different cluster and labeled through indexing. Whenever a query image arrives, extract the dominant color from the image and do the quantization. The cluster in which the image belongs to is identified by using the index values.

Similarity Measure

Searching large databases of images is a challenging task especially for retrieval by content. Most search engines calculate the similarity between the query image and all the images in the database and rank the images by sorting their similarity. Similarity measure is used to subtract the two images. We use the modified Euclidean distance here. Let I_1 and I_2 are two images, then the similarity measured as

$$SIM(I_1, I_2) = \sqrt{\left(\mathrm{DC_r(I_2)} - \mathrm{DC_r(I_1)}\right)^2 + \left(\mathrm{DC_g(I_2)} - \mathrm{DC_g(I_1)}\right)^2 + \left(\mathrm{DC_b(I_1)} - \mathrm{DC_b(I_2)}\right)^2}$$

In the above measure $\mathrm{DC_r}$, $\mathrm{DC_g}$ and $\mathrm{DC_b}$ are dominant color values in red, green and blue component.

EXPERIMENTAL RESULTS

To show the practical relevance of the proposed CBIR system, retrieval experiments are conducted. The technique is evaluated based on recall and precision. A recall rate can be defined as the number of relevant documents retrieved by a search divided by the total number of existing relevant documents (which should have been retrieved). A precision rate on the other hand is the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search.

For the experiments we selected database corel images of ground truth. It consists of 10 categories with each category comprising of 100 images. The categories include mountain, elephant, dinosaurs, African people, buses, horses, flowers, buildings, food, and beaches. The size of the images is 256×384 and other corel database with Image size 128×85 which contain butterflies, cars, etc

Fig 2 Images in the Database

The total number of relevant correctly retrieved images and the total number of retrieved images obtained from each of the query performed is recorded. For each query, the recall and precision values will be interpolated. The average precision at each recall level for each image category is then calculated. The time complexity also calculated and time taken for the retrieval is less than other systems.

Fig 3 Retrieval Result

CONCLUSION

In this paper we introduced a Content-based Image Retrieval System based on dominant color using a clustered database. The system uses quantization technique to limits the dominant color from the image into a certain level. In our work the images were grouped into different clusters using the quantized values and indexed. Whenever the query image given to the system, it identifies the cluster to which the image belongs. So instead of searching the whole database it searches the corresponding cluster and retrieves the images. Experiments shows that the retrieval time complexity is reduced compare to other systems and the precision rate is also good.

REFERENCES:

[1] Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang, "Image retrieval: ideas, influences, and trends of the new age", ACM Computing Surveys 40 (2) (2008) 1–60.

[2] W. Niblack et al., "The QBIC Project: Querying Images by Content Using Color, Texture, and Shape," in Proc. SPIE, vol. 1908, San Jose, CA, pp. 173–187, Feb. 1993.

[3] B. S. Manjunath, Jens-Rainer Ohm, Vinod V. Vasudevan,, and Akio Yamada, "Color and Texture Descriptors", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 11, NO. 6, JUNE 2001.

[4] Thomas Sikora "The MPEG-7 Visual Standard for Content Description—An Overview" IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 11, NO. 6, JUNE 2001.

[5] M. Sticker, and M. Orengo, "Similarity of Color Images," in Proc. SPIE Storage and Retrieval for Image and Video Databases, pp.381-392, Feb. 1995.

[6] Nai-Chung Yang, Wei-Han Chang, Chung-Ming Kuo, Tsia-Hsing Li, "A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval", Journal of Visual Communication and Image Representation 19 (2) (2008) 92–105.

[7] Chia-Hung Wei, Yue Li, Wing-Yin Chau, Chang-Tsun Li, "Trademark image retrieval using synthetic features for describing global shape and interior structure", Pattern Recognition 42 (3) (2009) 386–394.

[8] Ahmed Talib, Massudi Mahmuddin, Husniza Husni, Loay E. George, "A weighted dominant color descriptor for content-based image retrieval", J. Vis. Commun. Image R. 24 (2013) 345–360.

[9] P. Howarth and S. Ruger, "Robust texture features for still-image retrieval", IEE. Proceedings of Visual Image Signal Processing, Vol.152, No. 6, December 2005.

[10] W.Y. Ma, Y. Deng, B.S. Manjunath, "Tools for texture/color based search of images", Proc. SPIE Hum. Vis. Electron. Imaging II 3106 (1997) (1997) 496–507.

[11] L.M. Po, K.M. Wong, "A new palette histogram similarity measure for MPEG-7 dominant color descriptor", IEEE Int. Conf. Image Process. (ICIP'04) 3 (2004) 1533–1536.

[12] A. Mojsilovic, J. Kovacevic, J. Hu, R.J. Safranek, S.K. Ganapathy, "Matching and retrieval based on the vocabulary and grammar of color patterns", IEEE Trans.Image Process. 9 (1) (2000) 38–54.

[13] A. Yamada, M. Pickering, S. Jeannin, L.C. Jens, "MPEG-7 Visual Part of Experimentation Model Version 9.0-Part 3 Dominant Color". ISO/IEC JTC1/SC29/WG11/N3914, Pisa, 2001.

[14] Y. Deng, B.S. Manjunath, C. Kenney, M.S. Moore, H. Shin, "An efficient color representation for image retrieval", IEEE Trans. Image Process 10 (1) (2001)140–147.

[15] K.-M. Wong, L.-M. Po, K.-W. Cheung, "Dominant color structure descriptor for image retrieval", IEEE Int. Conf. Image Process., 2007. ICIP 2007 6 (2007) 365–368.

[16] ISO/IEC 15938-3/FDIS Information Technology, "Multimedia Content Description Interface", Part 3 Visual Jul. 2001, ISO/IEC/JTC1/SC29/WG11 Doc. N4358