World Academy of Science, Engineering and Technology
International Journal of Biomedical and Biological Engineering
Vol:4, No:4, 2010

Effectiveness of Dominant Color Descriptor Technique in Medical Image Retrieval Application

Mohd Kamir Yusof

***Abstract*—**This paper presents a dominant color descriptor technique for medical image retrieval. The medical image system will collect and store into medical database. The purpose of dominant color descriptor (DCD) technique is to retrieve medical image and to display similar image using queried image. First, this technique will search and retrieve medical image based on keyword entered by user. After image is found, the system will assign this image as a queried image. DCD technique will calculate the image value of dominant color. Then, system will search and retrieve again medical image based on value of dominant color query image. Finally, the system will display similar images with the queried image to user. Simple application has been developed and tested using dominant color descriptor. Result based on experiment indicates this technique is effective and can be used for medical image retrieval.

***Keywords*—**Medical Image Retrieval, Dominant Color Descriptor.

I. Introduction

NOWADAYS, hospitals and medical research centrals produce lot of digital images of diverse modalities everyday [1][2]. Besides that, growth of the World Wide Web, medical images are now available in large numbers in online repositories and atlases [1][3]. Suitable technique for retrieving process is needed to handle lot of images in databases effectively and efficiently. Currently, the utilization of medical images is limited. This limitation due to the lack of effective search method i.e. text based searches.

The traditional method for image search is text-based retrieval. This technique executed based on one or more keywords specified by users. The problem with this technique is query request cannot be easily described by keywords. One of the solution have been tested is CBIR (Content-Based Image Retrieval) [4]. In CBIR system, the images are indexed by their visual contents as the features. The characteristics such as color, texture, shape, and color layout (both color features and spatial relations) are also considered [5]. In CBIR system also, the image features also is stored in database. In CBIR, when a query image is given, systems automatically extracted the features of query image and match the features in the feature database by a pre-established algorithm. After that, system will display a group of similar images with query image. Generally, the processes in CBIR system as follow:

1. User is required enter keyword for query image
2. System search and retrieve the query image
3. Search image from database based on image feature (such as color, texture, shape, and color layout)
4. Display a group similar image based on query image

However, this CBIR system is not enough in term of accuracy. That way, dominant color descriptor is proposed in medical image environment. The dominant color descriptor is one of techniques in image retrieval. This technique is smart and powerful in term of accuracy. The literature about dominant color description will describe in section 2.

II. Previous Work

Dominant Color Descriptor (DCD) is technique for image retrieval. This technique is providing an effective, compact, and powerful color representation for image retrieval [5]. A DCD specifies a small number of dominant color values and the statistical properties i.e. distribution and variance [7]. The structure of a DCD, F, is defined as:

$$F = \{p_i, c_i, \mathbf{v}_i, s\}, i=1,2,\ldots,N_{DCD} \quad (1)$$

where NDCD is the number of dominant colors, s is the spatial coherency value that represents the overall spatial homogeneity of the dominant colors, p*i* is the percentage of pixels in the image corresponding to the *i*th dominant color, $\mathbf{v}_i$ is a vector representing the *i*th dominant color, and the c_i is optional which is the variation of the dominant color values pixels around $\mathbf{v}_i$. In the most of application, s is set to zero.

This technique will calculate the distance to measure the similarity or dissimilarity between query image and any image in database.

A. Similarity Matching

Two additional types of similarity measurements are optimal distance and Euclidean distance. Suppose there two DCDs

$$F_1 = \{p_{1i}, c_{1i}, v_{1i}, s_1\}, i = 1, 2, \ldots, N_1 \quad (2)$$
$$F_2 = \{p_{2j}, c_{2j}, v_{2j}, s_2\}, i = 1, 2, \ldots, N_2 \quad (3)$$

where i is the F_1's subscript, and j is the F_2's subscript.

Comparison distance between query image and any image in database is a very time-consuming task. However, dynamic programming [8, 9] makes such a computation in a reasonable time. The optimal distance can be describe the similarity of two DCDs very well, however, when the measurement is used in an image retrieval system, there are takes much longer time than the computation of the Euclidean distance; the another disadvantages is that it is sensitive to noise. The Euclidean distance of two DCDs is

$$D2 = \arg \min_{ij} \sum_{ij} [(p1i - p2j)^2 + (c1i - c2j)^2 + \|v1i - v2j\|^2] \quad (4)$$

where N = max (N_1, N_2) if N_1 is not equal to N_2. Without losing generality, let $N_1 > N_2$, then the elements in F_2 between N_2 and N are set to 0's.

This technique has been tested using Corel Photo Gallery [38]. There various images in different folders in the Corel image database, and each folder has 100 similar images. The

subset is made up of 5000 images in 50 folders. The features are extracted from each image and they are stored in an image feature database. Based on experimental has been done by previous researcher [5], the result indicates dominant color descriptor technique is effective for Corel Photo Gallery.

Based on previous work, this technique has a potential to implement in another domain. In this paper, this technique will used for medical image domain. The implementation for medical image retrieval using dominant color description will describe in section 3.

III. IMPLEMENTATION DCD FOR MEDICAL IMAGE

Dominant color descriptor will be used for retrieve medical image. Architecture for medical image retrieval should be designed before implementation and testing phase.

A. Architecture for Medical Image Retrieval

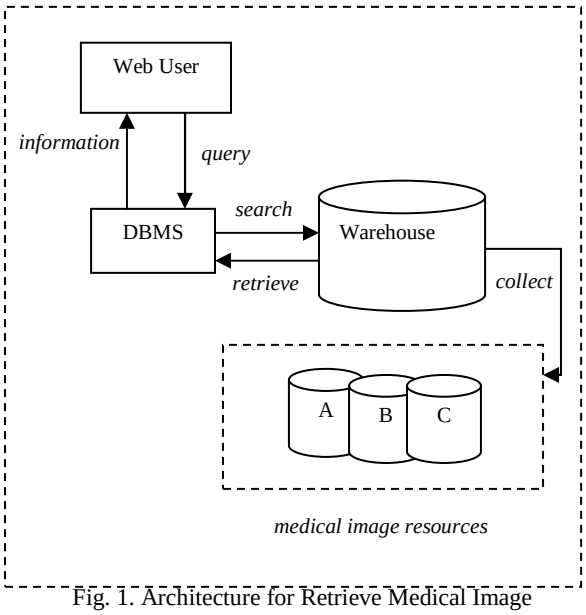

Fig. 1. Architecture for Retrieve Medical Image

Fig. 1 represented architecture for medical image retrieval. Web user will send a query to Database Management System (DBMS). The DBMS will search and retrieve data from warehouse. A warehouse is place for data storing. The data in warehouse is collection from medical image databases. System automatically updates and retrieves data from medical image databases and store in warehouse. After DBMS finish for searching and retrieving, system will display result to web user.

B. Architecture for Similarity Medical Image

Architecture for similarity medical image divided into two parts which are search query image and search similar image.

i. Search Query Image

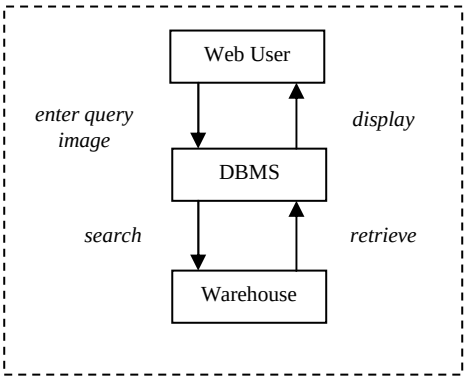

Fig. 2. Search Query Image

Fig. 2 shows first stage for search a query image. First, web user is needed to enter keyword i.e. "*M12DS*" into text field. System will search and retrieve image from warehouse. Then, the image is defined as a "*query image*". Section B (ii) will describe how system search similar image in database with queried image

ii. Search Similar Image

By dominant color descriptor technique, dominant color for queried image will be calculated using (eq. 2). Then, system will retrieve any image in warehouse. System will calculate value of dominant color for each image in database. Suppose in this module, the similarity values have to pass a threshold, which is set as A = 0.50. System will list all image if the value dominant color image equal or less than A. Simple formula for this process can be representing as below:

$$X = \{x_1, x_2, x_3, \ldots, x_4, x_n\} \quad [5]$$

$$Y = \{x_1, x_4, x_6, x_7\} \quad [6]$$

Equation 1 represents images in database, and then equation 2 represents images have a value dominant color equal or less than A = 0.50.

IV. EXPERIMENTAL RESULT

Simple application has been developed and tested using dominant color descriptor technique to retrieve medical image. In this application, PHP is used as a programming language and MySQL is used to store and manage database.

Fig. 3. Interface for Biomedical Application

Fig. 3 shows the main interface for biomedical application image. This technique can retrieve similar images with query image from medical database.

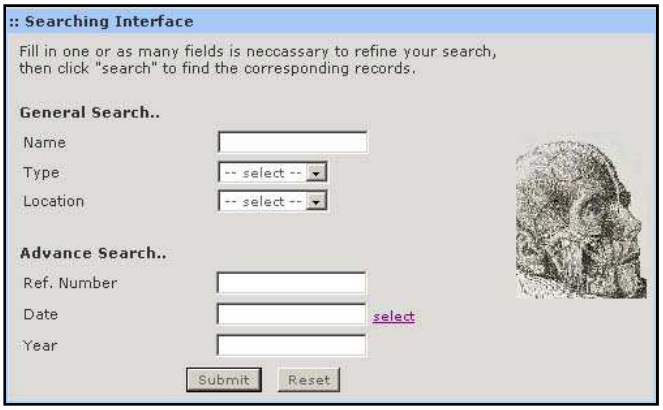

Fig. 4. Searching Interface

Fig. 4 shows interface for searching. User is needed to enter information either general information or advance information. Suppose the user key in reference number equal to M3786. After that, the user is needed to click button submit. Other method, the user also can press button enter on keyboard. Then, system will automatically search image from database based on reference number has been key in. After query image is found, system will assigned the image as a queried image.

The second process, the system will search and retrieve a similar image by using DCD technique. This technique will calculate value of dominant color for query image. Suppose the value of dominant color for query image is 0.50.

System will retrieve all images from database, and calculate dominant color for each image in database. After that, system will display image that similar with query image based on dominant color value. Figure 5 indicates three images which are value of dominant color less or equal to 0.50.

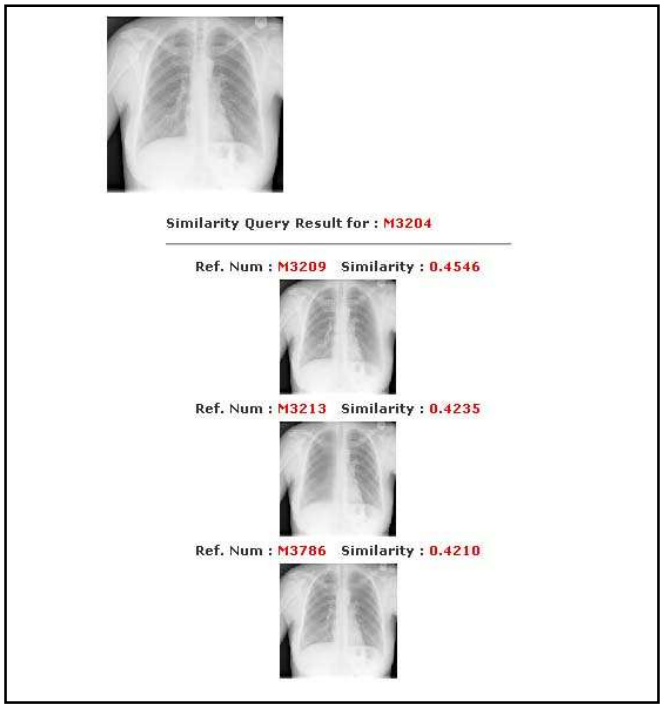

Fig. 5. Display a Result after Searching and Retrieving Process

Based on experiment has been done, the result indicate Dominant Color Descriptor is effective to retrieve medical

V. Conclusions and Future Work

This paper suggests Dominant Color Descriptor (DCD) technique to retrieve medical image. This technique is effective based on experiment has been done. This technique has a potential to be improved with added new function or algorithm. Future work will approach DCD technique for another specific domain. Beside that, enhancement is needed to improve capability of DCD in term of time for retrieving process when involve with large medical image database.

Acknowledgment

Special thanks are owned to my friend Che Mat Ismail for support and advice.

References

[1] Miller H, Michoux N, Bandon D, Geissbulher A. A Review of Content-Based Image Retrieval System in Medical Application Clinical Benefits and Future Directions. Int J Med Inform 2004, 73:1-23.

[2] Tagare LHY, Hanka R, Ip HSS. Medical Image Databases: A Content-Based Retrieval Approach. J. Am Med Informat Assoc 1997, 4: 184-98.

[3] Florea F, Muller H, Rogozan A, Geissbuhler A, Darmoni S. Medical Image Categorization with MedIC and MedGIFT. Proc Med Inform Europe (MIE 2006). P. 3-11.

[4] Y. Rui, T.S. Huang, S.F. Chang, Image Retrieval: Current Technique, Promising Directions and Open Issues, J. Visual Commun. Image Representation 10 (1999) 39-62.

[5] Rui Min, H.D. Cheng. Effective Image Retrieval Using Dominant Color Descriptor and Fuzzy Support Vector Machine.Pattern Recognition 42 (2009) 147-157.

[6] Y. Rui, T.S. Huang, M. Ortega, S. Mehrotra. Relevance feedback: a power tool for interactive cantent-based image retrieval, IEEE Trans. Circuit Syst. Video Technol. 8 (1998). 644-655.

[7] B.S. Manjunath, P. Salembier, T. Sikora,. Introduction to MPEG-7: Multimedia Content Description Interface, Wiley, Chichester, 2002.

[8] A. Lew, H. Much, Dynamic Programming a Computational Tool, Studies in Computational Intelligence, vol 38, Springer, Berlin, New York, 2007, p. xviii, 379 p.

[9] Rand Cooperation, R.E. Bellman, Dynamic Programming, Princeton University Press, Princeton, NJ, USA, 1957.

[10] J.Z. Wang, Li. G. Wiederhold, Simplicity: Semantics-Sensitive Integrated Matching for Picture Libraries, IEEE Trans. Image Process. 10 (2001) 140-147.

Mohd Kamir Yusof obtained his Master of Computer Science from Faculty of Computer Science and Information System, Universiti Teknologi Malaysia in 2008. Currently, he is a Lecturer at Department of Computer Science, Faculty of Infomatics, Universiti Darul Iman Malaysia (UDM), Terengganu, Malaysia.