Applied Medical Informatics
Original Research Vol. 33, No. 3 /2013, pp: 33-44

Content Based Medical Image Retrieval for Histopathological, CT and MRI Images

Swarnambiga AYYACHAMY [1,*], Mannivannan VASUKI [2]

[1]SACS MAVMM Engineering college, Madurai, Tamilnadu, India.
[2]Velammal College of Engineering &Technology, Viraganoor, Madurai-09, Tamilnadu, India.
E-mails: aswarnambiga@gmail.com; hello_vasuki@yahoo.co.in

* Author to whom correspondence should be addressed; Tel.: 0452-4522522098

Received: 28 June 2013 /Accepted: 9 September 2013/ Published online: 23 September 2013

Abstract
A content based approach is followed for medical images. The purpose of this study is to access the stability of these methods for medical image retrieval. The methods used in color based retrieval for histopathological images are color co-occurrence matrix (CCM) and histogram with meta features. For texture based retrieval GLCM (gray level co-occurrence matrix) and local binary pattern (LBP) were used. For shape based retrieval canny edge detection and otsu's method with multivariable threshold were used. Texture and shape based retrieval were implemented using MRI (magnetic resonance images). The most remarkable characteristics of the article are its content based approach for each medical imaging modality. Our efforts were focused on the initial visual search. From our experiment, histogram with meta features in color based retrieval for histopathological images shows a precision of 60 % and recall of 30 %. Whereas GLCM in texture based retrieval for MRI images shows a precision of 70 % and recall of 20 %. Shape based retrieval for MRI images shows a precision of 50% and recall of 25 %. The retrieval results shows that this simple approach is successful.

Keywords: Content based image retrieval; medical imaging; medical imaging modalities; color; texture; shape.

Introduction

Medical images support clinical decisions in health care centers. A content based approach is usually followed for medical images. The contents of an image have to be carefully extracted, classified with efficient techniques for easy retrieval. Contents in an image can be of various forms like, texture, color, and shape etc of which shape is regarded as the most efficient metric. The color, texture and shape based features extraction has already proven previously by researches. The reason for carrying image retrieval in medical images is because, the images in the database has some specific characteristic. Medical images more over have a formulaic composition for each modality and anatomic region. Hence the features were selected on the basis of their performance and this is the significance for carrying this work. For effective content based medical image retrieval, the search should be confined. Because medical images differ from one other, and also approaches should be selected according to the imaging modality and it can adopt that modality only. This is the major drawback in CBMIR. But the results found to be very accurate on those methods. On the other hand, it is well known that the diagnosis process in medicine in these days is mainly based on knowledge of diagnosing. There also some difficult issues to be deal with when, image knowledge contents have to be organized for retrieval tasks. Medical imaging in retrieval plays an important role in diagnostic purpose in medicine. Imaging modality is an important visual characteristic that can be used to improve the retrieval performance. Medical images are usually fused, subject to high inconsistency and composed of different minor structures. So there is a necessity for feature extraction and classification of images for easy retrieval. Medical images have a unique characteristic in that their contents always reflect pathological attributes or symptoms for specific diseases. Image classification is often used to group the similar features based on their contents [9] [13]. The main objective of the work is to retrieve the images from huge volume of medical databases with high accuracy by performing feature extraction, classification process. These retrieved medical images play a pivotal role in surgical planning, medical training, and patient diagnoses.

The aim of this experimentation is to include this domain specific knowledge to improve the performance of medical CBMIR system. In this article, color based feature extraction made possible by implementing CCM and histogram techniques and texture based image retrieval made possible by GLCM and LBP. Shape based retrieval by using canny edge detection algorithm with otsu's method. The remainder of this paper is organized as follows. Section 2, describes the methodology for our approach using CBMIR. In section 3 the framework for feature extraction in content based medical image retrieval were presented. A detailed experimental comparative analysis is presented in section 4 followed by discussion and is proceeded by a conclusion in section 5.

Methodology

Main intention of CBIR is efficient retrieval of images from a huge image database based on some automatically extracted features. Visual similarity is a subjective combination of several observable features such as color, texture and shape. These features are extracted from properties such as color, texture and shape of query image and also from the various images in the database. This section describes the methodology used for retrieval. Based on some common parameters evaluated from the feature, the relevancy between the query image and the database image are arranged accordingly. A typical CBIR system for retrieving images from database based on their similarity to the input images consists of four main steps. First step, is giving the query image selected from the database created. Our database is created using histopathological and MR images. Second step is feature extraction using the prescribed method. Third step is calculation of similarity metric for effective retrieval. Finally, the images related to the query image found in the database were retrieved effectively. Before performing retrieval, one has to choose these categories. The database created should also follow these criteria. For different categorization, categories oriented database can be created. Here the database is created by considering the above categories, anatomical regions with 2-D images were considered with monomodal magnetic resonance images .The head, thorax , abdomen, pelvis, perineum , limbs and spine and vertebrae falls under the object category.

Color Based Retrieval for Histopathological Images

Color image is divided into $N \times N$ image sub-block, for anyone image sub-block $T_{(i,j)}$ $(1 \leq i \leq N, 1 \leq j \leq N)$, by main color image extraction algorithm to calculate the main color $C_{(i,j)}$.

The statistic features extracted from CCM are as follows: Through this method, 8 dimensional texture features for component R, G in RGB color space and H in HSV color space is obtained. Each component correspond to two statistic values E and S as Function [RE, RS, GE, GS, HE, HS, VE, VS].

For E = ∑for(i=1, D) ,

$$\text{Energy} = \sum\sum[m(i,j0)]^2 \quad (1)$$

For S = ∑ ∑(i, D),(j, D),

Entropy S = ∑∑m(i,j).log[m(i,j)] (2)

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

Histogram with Meta Features Retrieval for Histopathological Images

Histogram features have been traditionally used in content based image retrieval to calculate similarity measure and to rank the images [2] . The following histogram features were used.
Gray histogram: Luminance intensities in a 256 gray scale.
Color histogram: In the RGB color model with a partition space of 8×8×8.
A set of meta features are calculated from the information of each histogram as follows (k is a index for histogram bins).
Skewness: P the third central moment,

μ^3/σ^3 (3)

Kurtosis: the fourth central moment

μ^4/σ^4 (4)

Mean:

$\sum_k kh(k)$ (5)

All Meta features were calculated on each of the two histogram features, which amounts to a total of 8 meta features per image.

Texture Based Retrieval

Texture in CBIR is classified under two categories. First, an image can be considered to be a mosaic which consists of different texture regions. These regions can be used as examples to search and retrieve similar areas. Second, if medical images are represented in gray level, texture becomes a crucial feature, which provides indications about scenic depth, the spatial distribution of tonal variations, and surface orientation. For texture based medical image retrieval two approaches were followed. The first approach is based on Gray level co occurrence matrix and the second approach is based on local binary pattern.

GLCM based retrieval for MR images

GLCM creates a matrix with the directions and distances between pixels. Its level is determined by image gray level [1]. GLCM is composed of the probability value and is given in below equation (6). GLCM texture feature is determined by the correlation of the couple pixels in gray-level at different positions.

Two features were selected, it includes energy and entropy and is given in equation (1) and (2) and is used in feature extraction using GLCM. The matrix $P(i,j\ d,\theta)$denotes the distance between pixels (x_1,y_1) and (x_2,y_2).Elements in the matrix were computed by following equation.

$$p(i,j.,d,\theta)=\frac{p(i,j,d,\theta)}{\sum i\sum j\, p(i,j,d,\theta)} \quad (6)$$

LBP for MR images

Local Binary Pattern (LBP) is an intensity invariant texture descriptor with low computational complexity. The original LBP operator describes the texture in the image by thresholding the neighborhood with the gray value of its center pixel and then representing the result as a binary code [5]. The LBP operator as two-dimensional surface textures can be described by two complementary measures. They are local spatial patterns and grey scale contrast. The original LBP operator gives labels for the image pixels by thresholding the 3 x 3 neighborhood of each pixel with the center value and summing the threshold values weighted by powers of two which is a binary number [4]. The LBP value for the center pixel (a,b) of the image f (a,b) is calculated using the equation (7) . Where U(x) is the thresholding function and is defined in equation (8).

$$\mathrm{LBP}(a,b)=\sum_{i=0}^{7}U(f(a,b)-f(ai,bi))2^{i} \qquad (7)$$

$$U(x)=f(x)=\begin{cases}1, & x\geq 0\\ 0, & x<0\end{cases} \qquad (8)$$

Shape Based Retrieval

Shape is an important and most powerful feature used for image classification, indexing and retrievals [3]. In this work, the edge information in the image is obtained by using the canny edge detection with otsu's method.

Similarity Retrieval

Medical CBIR calculates the visual similarities between a query image and images in a database for exact matching for diagnosis. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on the distribution of features. Different similarity distance measures will affect the retrieval performances of an image retrieval system significantly. The Euclidean distance measure is best for CBMIR based matching. To calculate the Euclidean distance between the query image and the images of the database the formula given in equation (11) is used. This technique can be implemented using MATLAB software. Similarity measure is to retrieve images. Here Euclidean similarity measure is implemented. CBIR system ranks similarity in descending order and then returns relevant images that are most similar to the query images. The direct Euclidean distance between an image P and query image Q can be given as follows,

Euclidean Distance = $\sqrt{\sum(\nu_{pi}-\nu_{qi})^2}$ where i= 1 to n (9)

Experimental Results and Discussion

A computer supported by Intel core processor has been used for conducting the experiments. The developed user interface components with database of medical images serving as front end. MATLAB 7.10.0- image processing toolbox- workspace was used as feature database for storage as backend and for image processing work other MATLAB 7.10.0 utilities were used for mathematical equations. Math type tool was also used for writing document. Initially, MATLAB 7.10.0 workspace database with 1500 medical images were used for testing the proposed CBMIR system. These images were acquired without any imaging artifacts. This might be seen lesser, But the view carried by this is, no toolkit has been used, no browsers were needed, an easy interpretation and representation is possible only with limited number of images. But this is not the restriction in this work. The scenario for a physician diagnosing a patient and observing a new result of an imaging exam wishes to fetch from the database visually similar images that have been already pre-diagnosed by him/her or others. The important section in this work is this experimental setup. The results obtained were given below. The results show the effectiveness of our experimental setup.

Discussion

The color based medical image retrieval is made possible by feature extraction in step (1) and (2). Table (1) shows the retrieval performance of retrieved images and table (2) shows the evaluation made form table (1). It is evident that histogram with meta features outperforms CCM with precision of 60 % and recall of 30% with higher efficiency and lesser error rate. Fig (1) shows the retrieved images using CCM and is evident that its performance is lesser found from the visual observance and for evaluation. The reason for CCM result found to be worse than histogram with meta features is, CCM uses only a feature set of 8 features and is exactly in RGB and in HSV space. Moreover this also concentrates on color and texture region. This concentration of texture is similar to GLCM. Whereas histogram with meta features also uses a feature set of 8 features but the features were directly extracted from the images not considering the space. This meta features calculated from the images concentrate on whole image. The histopathological images differ from other medical images by its distribution of concentration of texture, color and region of interest. This is reason for the higher efficiency in the retrieval of histogram with meta features for histopathological images. In texture based image retrieval GLCM outperforms LBP with a precision of 70% and recall of 20% moreover with higher efficiency and lesser error rate. LBP depends on intensity related texture descriptor whereas GLCM depends on texture orientation of pixels. This characteristic of texture orientation makes GLCM to perform better in retrieval of MR images where concentration of texture found to be greater. In shape based medical image retrieval shape alone cannot be a feature considered for retrieval. Usually for medical images shape based retrieval is not widely used as a major criteria. So for this curious problem faced by the researches [3], in this work have combined canny edge which is suitable for shape based medical image retrieval and otsu's method which is significantly used for multivariable thresholding is used. This has proven the effectiveness with a precision of 50% and 25% of recall. Fig 1, 2 and table 1, 2 shows the details of retrieved images using color based retrieval for histopathological images. Fig 3, 4 and table 3, 4 shows the details of retrieved images using texture based retrieval for MRI images. Fig 5, 6 and table 5, 6 shows the details of retrieved images using shape based image retrieval for MRI images.

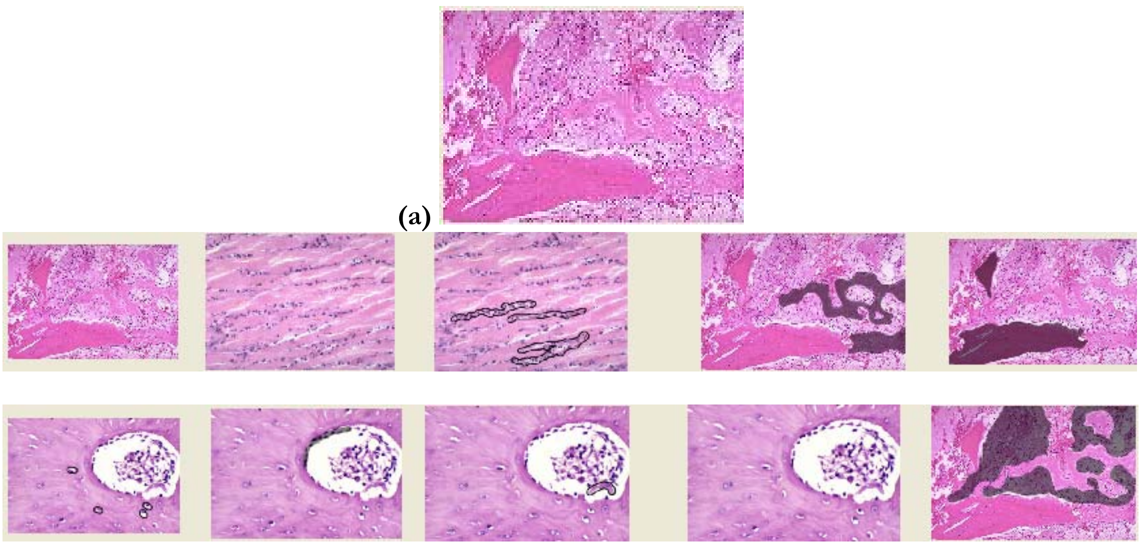
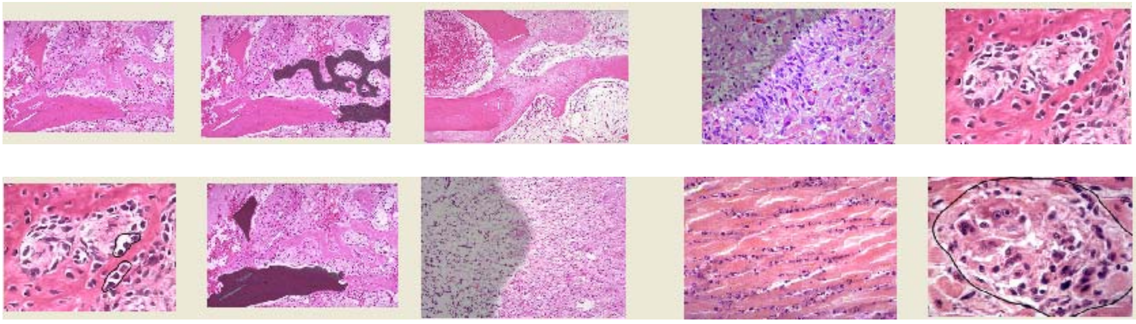
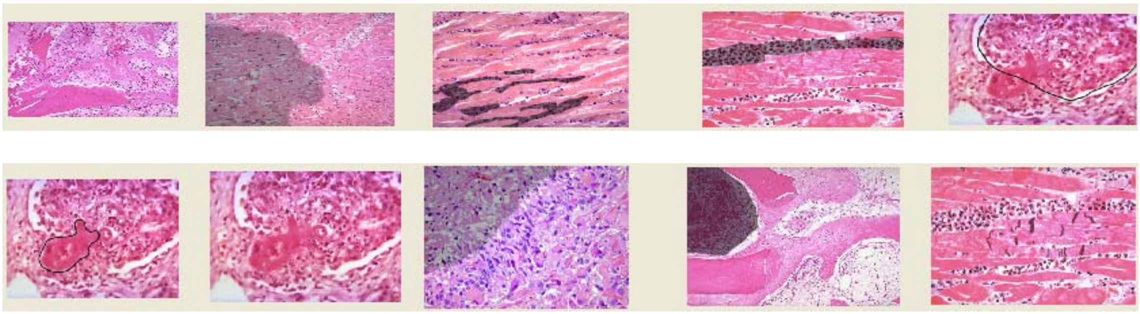


Figure 1. The retrieved images using CCM (a) Shows the extraction of color features.

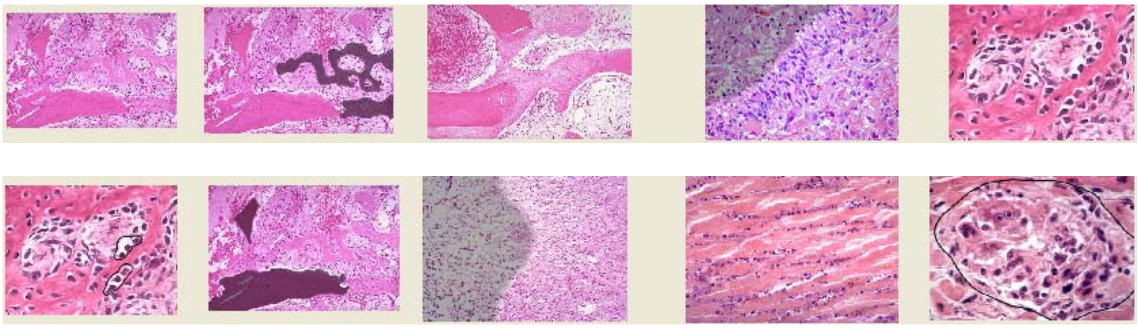
(b)



(c)



(d)



(e)

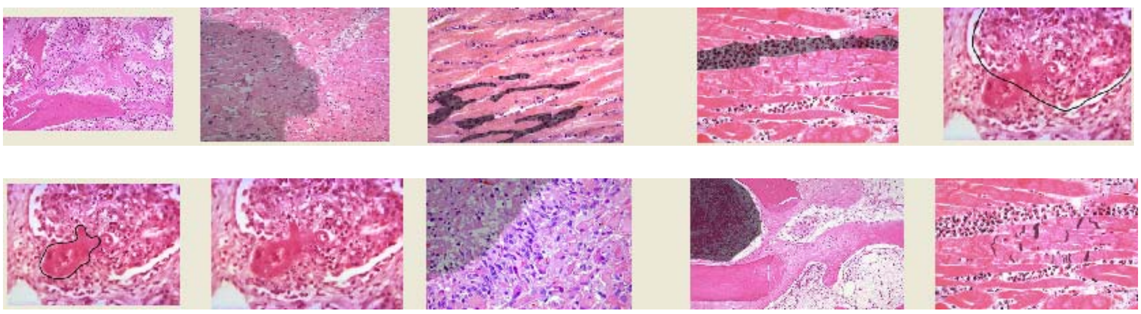


Figure 1. The retrieved images using CCM (b) shows GLCM based feature extraction (c) shows CCM based feature extraction (d) shows COLOR features + GLCM based feature extraction (e) shows COLOR features+ CCM based feature extraction. The top most image is the query image.

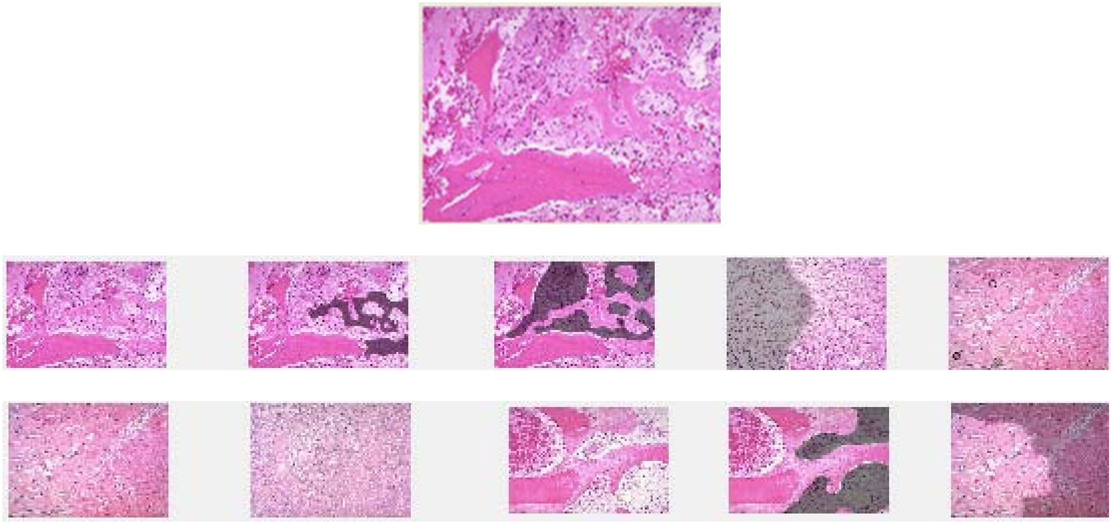


Figure 2. Retrieved images using histogram with multi feature extraction. The top most image is the query image.

Table 1. Retrieval performance of color based retrieved images

Color based feature extraction	Relevant image	Irrelevant image	Actual image	Time taken for retrieval in seconds
CCM	4	5	1	3
Histogram with multi feature extraction	6	3	1	2

Table 2. Evaluation of the retrieval performance of color based retrieved images

Color based feature extraction	Precision	Recall	Efficiency	Error rate
CCM	40	20	40	90
Histogram with multi feature extraction	60	30	60	30

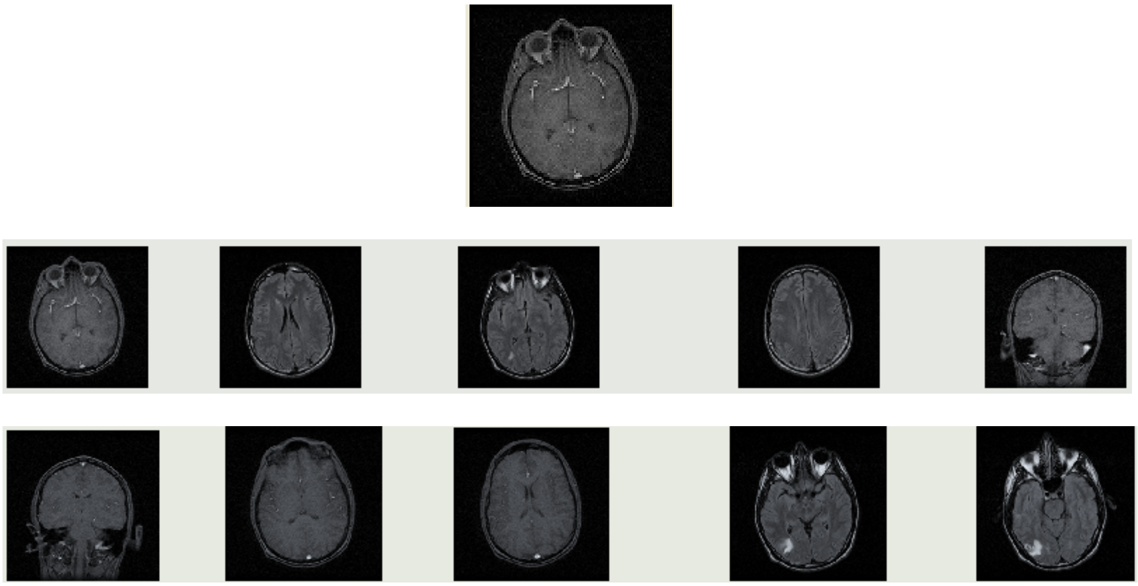
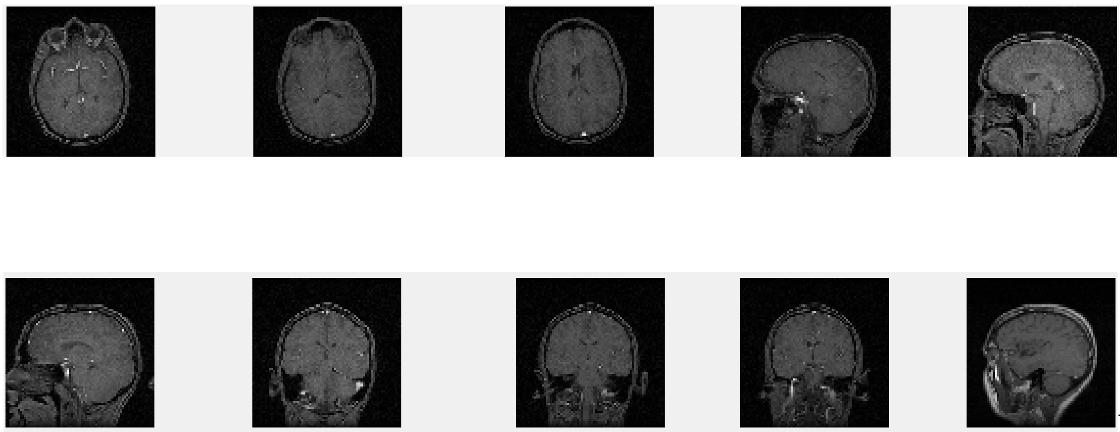


Figure 3. GLCM based feature extraction for retrieval of medical images



F

Figure 4. LBP (local binary pattern) based medical image retrieval and the topmost image is the query image common for both GLCM and LBP.

Table 3. Retrieval performance of texture based retrieved images

Texture based feature extraction	Relevant image	Irrelevant image	Actual image	Time taken for retrieval in seconds
GLCM	7	2	1	2
LBP	3	6	1	2

Table 4. Evaluation of the retrieval performance of texture based retrieved images

Texture based feature extraction	Precision	Recall	Efficiency	Error rate
GLCM	70	20	70	20
LBP	30	15	30	60

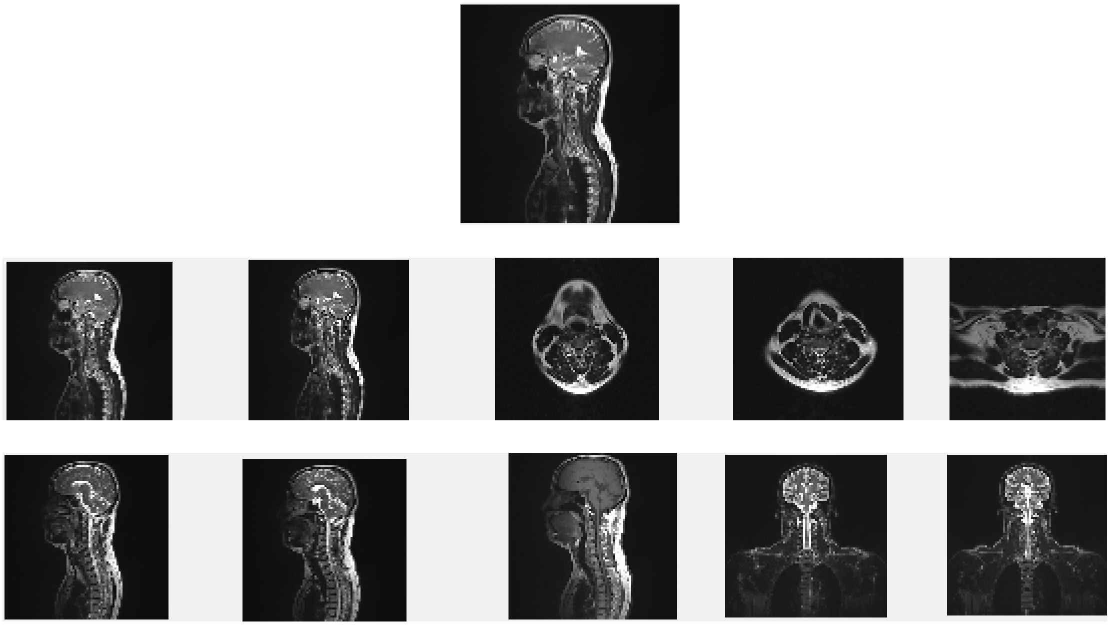


Figure 5. Retrieved images of CANNY EDGE with OTSU's method and topmost image is the query image.

Table 5. Retrieval performance of shape based retrieved images

Texture based feature extraction	Relevant image	Irrelevant image	Actual image	Time taken for retrieval in seconds
CANNY EDGE with OTSU's method	5	3	2	5

Table 6. Evaluation of the retrieval performance of shape based retrieved images

Texture based feature extraction	Precision	Recall	Efficiency	Error rate
CANNY EDGE with OTSU's method	50	25	50	30

Conclusion

The goal of this experimentation is to provide an effective means for organizing large collections of medical images. This requires an intelligent system, medical image retrieval using CBMIR which have the ability to recognize, capture, and understand the complex content of medical images. Despite of our accurate experimental results medical content-based image retrieval still has a long way to go and more efforts are expected to be devoted to this area. Ultimately, a well-organized image database is created, accompanied by an intelligent retrieval mechanism. This will support clinical treatment, and provide a basis for better medical research and education.

List of abbreviations

Computed Tomography (CT)
Magnetic Resonance Imaging (MRI)
Gray Level Co-occurrence Matrix(GLCM)

Color Co-occurrence Matrix (CCM)
Local Binary Pattern (LBP)

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors like to thank the editors and reviewers for their useful comments in improving the quality of the paper.

References

1. Kong FH. Image Retrieval Using Both Color and Texture Features, Proc of the Eighth Inte'l Conf. on Machine Learning and Cybernetics, Baoding, 2009.
2. Caicedo J, Gonzalez F, Romero E. A Semantic Content-Based Retrieval Method for Histopathology Images, [Internet] 2008 [cited 2013] Available from: http://www.bioingenium.unal.edu.co.
3. Ramamurthy B, Chandran KR. Content based Image Retrieval for Medical Images using Canny Edge Detection Algorithm. International Journal of Computer Applications 2011;6(17):23-37.
4. Gao X, Qian Y, Hui R, Loomes M, Comley R, Barn B, Chapman A, Rix J. Texture-based 3D image retrieval for medical applications. IADIS International Conference e-Health 2010:101-108.
5. Callins C, Rajamani V. Comparison of Local Binary Pattern Variants for Ultrasound Kidney Image Retrieval. International Journal of Advanced Research in Computer Science and Software Engineering 2012;10(2):224-228.
6. Manjunath BS, Ma WY. Texture feature for browsing and retrieval of image data. IEEE Transaction on PAMI 1996; 8(18):837-42.
7. Rui Y, Alfred C, Huang TS. Modified descriptor for shape representation, a practical approach, In: Proc of First Int's workshop on I mage Database and Multimedia Search 1996.
8. Wei CH, Li CT, Wilson R. A content-based approach to medical image database retrieval. Idea Group Inc., 2006.
9. Tsumura M, Miyake Y. Why multi-spectral imaging in medicine? Journal of Imaging Science and Technology 2004;48(2):125-29.
10. Qian G, Sural S, Gu Y, Pramanik S. Similarity between Euclidean and cosine angle distance for nearest neighbor queries. In [Proceedings] 2004, ACM Symposium on Applied Computing 1232-7.
11. Muller H, Michous N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications - Clinical benefits and future directions. International Journal of Medical Informatics 2004;73(1):1-23.
12. Muller H, Rosset A, Vallee JP, Geissbuhler A. Comparing features sets for content-based image retrieval in a medical-case database. [Proceedings] 2004, IS&T/SPIE Medical Imaging 2004: PACS and Imaging Informatics, 99-109.
13. Nishibori M. Problems and solutions in medical color imaging. Second International Symposium on Multi-Spectral Imaging and High Accurate Color Reproduction 2000:9-17.

14. Lehmann TM, Guld MO, Thies C, Plodowski B, Keysers D, Ott B, Schubert H. IRMA - Content-based image retrieval in medical applications. 14th World Congress on Medical Informatics, 2004:842-8.
15. Feng D, Siu WC, Zhang HJ. (Eds.). Multimedia information retrieval and management: Technological fundamentals and applications. Berlin, Springer; 2003.
16. Antani S, Lee DJ, Long LR, Thoma GR. Evaluation of shape similarity measurement methods for spine X-ray images. Journal of Visual Communication and Image Representation 2004;15(3):285-302.
17. Beckmann N, Kriegel HP, Schneider R, Seeger B. The R*- tree: An efficient and robust access method for points and rectangles. ACM SIGMOD International Conference on Management of Data 1990:322-31.
18. Brodley CE, Kak A, Dy J, Shyu CR, Aisen A, Broderick L. Content-based retrieval from medical image database: A synergy of human interaction, machine learning, and computer vision. Sixteenth National Conference on Artificial Intelligence, 1999:760-7.
19. Rahman M, Wang T, Desai B, Medical image retrieval and registration: towards computer assisted diagnostic approach. IDEAS Workshop on Medical Information Systems: The Digital Hospital, 2004.
20. Deselaers T, Keysers D, Ney H. Features for image retrieval: an experimental comparison [Internet] [Cited 2013]. Available from: http://www.lis.ei.tum.de/research/bv/topics/mmdb/empeg7html
21. Singha M, Hemachandran K. Content based image retrieval using color and texture. Signal & Image Processing : An International Journal (Sipij) 2012;1(3):780-784.
22. Yagediez A. Revisiting intensity based image registration applied to Mammography. IEEE Transaction on Information Technology in Biomedicine 2011;5(15): 716-25.
23. Braveen M, Dhavachelvan P. Evaluation of content based image retrieval systems based on color feature. International Journal of Recent Trends in Engineering 2009;2(1):57-62.
24. Muller H. User interaction and performance evaluation in content based visual information retrieval. [PhD thesis]. Computer Vision and Multimedia Laboratory, University of Geneva, Geneva, Switzerland 2002.
25. Smith JR. Integrated spacial and feature image systems: Retrieval, compression and analysis. [PhD thesis]. Graduate School of Arts and Sciences, Columbia University, 2960 Broadway, New York, NY, USA 1997.
26. Yang A, Caixinglin C. A improved medical image registration framework based on mutual information. IEEE computer society. Global Congress on Intelligent System 2009:588-92.
27. Meyer CR, Boes JL, Kim B, Bland PH, Zasadny KR, Kison PV, Koral K, Frey A, Wahl RL. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. Medical Image Anal 1997:195-206.
28. Studholme DL, Hill G, Hawkes DJ. Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. Medical Phys 1997:25-35.
29. Hafner J, Sawhney H, Equitz W, Flickner M, Niblack W. Efficient color histogram indexing for quadratic form distance functions. IEEE Transactions on Pattern Analysis and Machine Intelligence 1995;7(17):729-36.
30. Jain A, Vailaya A. Image retrieval using color and shape. Pattern Recognition 1996;29(8):1233-44.
31. Niblack W, Barber R, Equitz W, Flickner M, Glasman E, Pektovic D, Yanker P, Faloutsos C, Taubin G. The QBIC Project: Querying images by content using color texture and shape. Proc. SPIE Int. Soc. Opt. Eng., in Storage and Retrieval for Image and Video Databases 1993;1908:173-87.

32. Mojsilovis A, Gomes J. Semantic based imagecategorization, browsing and retrieval in medical imagedatabases, [Conference] 2000 IEEE International Conference on ImageProcessing (ICIP' 2000), Rochester, NY, USA.
33. Liu Y, Dellaert F. Classification-driven medical image retrieval. 1997 ARPA Image Understanding Workshop.