Journal of Computer Science 8 (7): 1070-1076, 2012
ISSN 1549-3636

Content Based Medical Image Retrieval with Texture Content Using Gray Level Co-occurrence Matrix and K-Means Clustering Algorithms

[1]Ramamurthy, B. and [2]K.R. Chandran
[1]Department of CS, Sri Ramakrishna Engineering College, Coimbatore, India
[2]Department of CIS, PSG College of Technology, Coimbatore, India

Abstract: Problem statement: Recently, there has been a huge progress in collection of varied image databases in the form of digital. Most of the users found it difficult to search and retrieve required images in large collections. In order to provide an effective and efficient search engine tool, the system has been implemented. In image retrieval system, there is no methodologies have been considered directly to retrieve the images from databases. Instead of that, various visual features that have been considered indirect to retrieve the images from databases. In this system, one of the visual features such as texture that has been considered indirectly into images to extract the feature of the image. That featured images only have been considered for the retrieval process in order to retrieve exact desired images from the databases. **Approach:** The aim of this study is to construct an efficient image retrieval tool namely, “Content Based Medical Image Retrieval with Texture Content using Gray Level Co-occurrence Matrix (GLCM) and k-Means Clustering algorithms”. This image retrieval tool is capable of retrieving images based on the texture feature of the image and it takes into account the Pre-processing, feature extraction, Classification and retrieval steps in order to construct an efficient retrieval tool. The main feature of this tool is used of GLCM of the extracting texture pattern of the image and k-means clustering algorithm for image classification in order to improve retrieval efficiency. The proposed image retrieval system consists of three stages i.e., segmentation, texture feature extraction and clustering process. In the segmentation process, preprocessing step to segment the image into blocks is carried out. A reduction in an image region to be processed is carried out in the texture feature extraction process and finally, the extracted image is clustered using the k-means algorithm. The proposed system is employed for domain specific based search engine for medical Images such as CT-Scan, MRI-Scan and X-Ray. **Results:** For retrieval efficiency calculation, conventional measures namely precision and recall were calculated using 1000 real time medical images (100 in each category) from the MATLAB Workspace database. For selected query images from the MATLAB-Image Processing tool Box-Workspace Database, the proposed tool was tested and the precision and recall results were presented. The result indicates that the tool gives better performance in terms of percentage for all the 1000 real time medical images from which the scalable performance of the system has been proved. **Conclusion:** This study proposed a model for the Content Based Medical Image Retrieval System by using texture feature in calculating the Gray Level Co Occurrence matrix (GLCM) from which various statistical measures were computed in order to increasing similarities between query image and database images for improving the retrieval performance along with the large scalability of the databases.

Key words: Euclidean distance, k-means clustering algorithm, CBIR, GLCM, precision, recall, visual information, texture feature extraction

INTRODUCTION

Content Based Image Retrieval (CBIR) is a method in which various visual contents (called as features) have been considered to search and retrieve images from large scale of image databases based on the user’s requests in the form of a query image (Glatard *et al*., 2004; Remco and Tanse, 2000; Ozden and Polat, 2005; Nandagopalan *et al*., 2008). The following are some of the commercially available image search engines: QBIC, Visual Seek, Virage, Netra, PicSOM, FIRE, AltaVista. But these engines are not in a domain specific one. In this study we have proposed a domain specific based search engine for

Corresponding Author: Ramamurthy, B., Department of CS, Sri Ramakrishna Engineering College, Coimbatore, India

medical Images such as CT-Scan, MRI-Scan and X-Ray. The objective of the study is to permit radiologist to retrieve images of similar features that lead to similar diagnosis purpose from large volumes of databases. In this study, the image feature extraction, classification provides a flexible means of searching and retrieval of an image based on content description of query and database images.

CBIR system is a method for searching and retrieving of images based on their low level features (example texture, color, shape). It is a system which discriminates the dissimilar regions of an image based on their resemblance and decides the resemblance between two images by calculating the distance of these different regions. In CBIR system, any kind of images can be given as input image which depends upon the application requirements.

In medicine, all the data and related health information are stored as visual information in the form of X-rays, ultrasound or other scanned images, for diagnosis and monitoring purposes (John and Graham, 1999).

The main objective of the study is to retrieve the images from the huge volume of medical databases with high accuracy by performing feature extraction, classification process. So that the retrieved images are used for various medical diagnostic purposes (Ramamurthy and Chandran, 2011).

Related work: Medical images play an important role in the medical field (for example in surgical planning, medical training and patient diagnoses). In large hospitals and medical laboratories, there are thousands of images to be managed every year. For images classifying, indexing and retrieval in manual method is very expensive one and time consuming because those medical images vary person to person. For a successful CBIR system in the medical domain, classification and indexing schemes to be very efficient for searching and retrieval of the images from the database. The following systems are some of the existing works related to image classification and retrieval methods in the content-based image retrieval.

The ASSERT system uses a physician-in-the-loop method to retrieve HRCT of lung images. It allows the users to describe the pathology-bearing regions and to determine anatomical marker for each image. This system extracts 255 features of texture, shape, edges and grayscale properties in pathology-bearing regions. A multi-dimensional hash table is constructed to index the HRCT images. It has been implemented exclusively for lung images only not to general (Shyu *et al*., 1999).

The IRMA system is exclusively developed for medical applications for retrieval of medical images. This system has seven successive steps in order to perform the image retrieval process in an effective manner which includes categorization, registration, feature extraction, feature selection, indexing, identification and retrieval (Thies *et al*., 2005).

Content-Based Retrieval and Classification of Ultrasound Medical Images of Ovarian Cysts: In this study author presents a combined method of content-based retrieval and classification of ultrasound medical images. In this study, the authors have used histogram moments and GLCM based texture features in their study for retrieving and classifying ultrasound images. To retrieve images, similarity measurements have been used to determine image similarity between the query image and database images using a similarity model based on Gower’s similarity coefficient. For image classification, Fuzzy k-Nearest Neighbor (k-NN) classification method has been used. However this method gives more accurate and efficient result, but it has its own characteristics for classifying and retrieval of the images (Sohail *et al*., 2010).

The Spine Pathology and Image Retrieval System (SPIRS) (Long *et al*., 2005; Thoma *et al*., 2006), has been developed for pathologically sensitive retrieval of digitized spine x-rays and associated person metadata that come from the second US. National Health and Nutrition Examination Survey using localized vertebral shape-based CBIR methods at the U. S. National Library of Medicine Centre. In the SPIRS system, the images in the collection must be homogeneous.

The Image Map (Petrakis *et al*., 2002) is a system that considers how to handle multiple organs of interest. However, it works based on spatial similarity. Consequently, a problem caused by the user is likely to occur and therefore, the retrieved image will represent an unexpected organ.

Proposed model: The proposed model of this study is shown in Fig. 1 in which the Medical images (such as x-ray, MRI Scan, CT scan) are given as input into the system. Then, given input images are segmented by using the method described in (Ozden and Polat, 2005).

In this proposed study, only the texture regions of the image are considered for feature extraction. For each image in the image database, feature vector value has been developed and which are stored in feature database. When a query image is submitted by the user, the same texture feature extraction and feature vector value construction process has been applied to the query image in order to obtain the feature vector value to the query image. For similarity comparison between the query image and the database image, the Euclidean distance method is used. The closest Euclidean distance values to the database images are ranked and retrieved.

Texture: Texture is a natural property of surfaces and it provides visual patterns of the image. It has repeated pixel of information and it contains vital information regarding the structural arrangement of the surface (example clouds, leaves bricks). It also gives the relationship between the surface and external environment.

In this study, the extraction process of texture feature is performed by computing the Gray Level Co-Occurrence Matrix (GLCM). In this process, the graycomatrix function is used to create a GLCM. The graycomatrix function creates a gray level co matrix by calculating how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. The spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent) (Haralick *et al*., 1973). The outcome of GLCM for each element (I, J) is computed by summing the pixel with the value I occurred in the particular spatial relationship to a pixel with value j in the input image (Partio *et al*., 2002; Park *et al*., 2004). GLCM features are extracted using one distance d = {1} and four directions θ = {0, 90, 180 and 270°}.

After computation of Gray level co occurrence matrix, a number of statistical texture measures based on GLCM are derived which are suggested by Haralick.

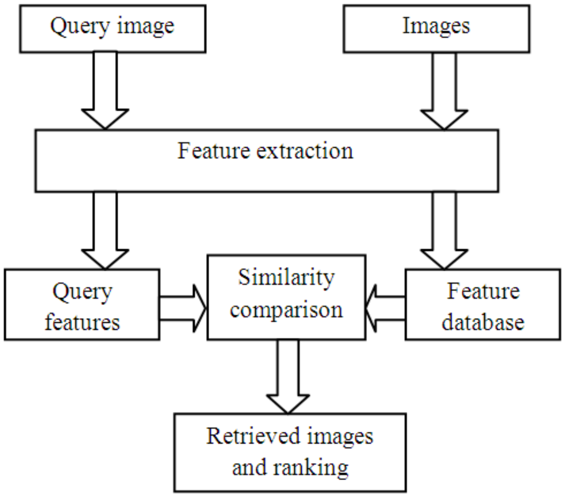


Fig. 1: Proposed model of the work

For generating texture features, second order method has been used that are derived from the co-occurrence probabilities. These probabilities represent the conditional joint probabilities of all pairwise combinations of gray levels in the spatial window of interest given two parameters: inter pixel distance (δ) and orientation (θ). The probability measure can be defined as Eq. 1:

$$\Pr(x) = \{Cij \mid (\delta, \theta)\} \quad (1)$$

where, Cij (the co-occurrence probability between gray levels i and j) is defined as Eq. 2:

$$Cij = \frac{Pij}{\sum_{i,j=1}^{G} Pij} \quad (2)$$

Where:

Pij = Represents the number of occurrences of gray levels

i and j = Within the given image window, given a certain (δ, θ) Pair

G = The quantized number of gray levels

The sum in the denominator thus represents the total number of gray level pairs (I, j) within the window.

Graycomatrix computes the GLCM from a full version of the image. By default, if I a binary image, graycomatrix scales the image to two gray-levels. If I is an intensity image, graycomatrix scales the image to eight gray-levels. (Haralick *et al*., 1973)

A texture is distinguished by a 14 statistical measurement value suggested by Haralick *et al.* (1973). The following formulas are used to calculate the features and which are shown in Eq. 3-6 (Yin *et al*., 2008):

$$\text{Energy} = \sum_{i,j} P(i,j)^2 \quad (3)$$

$$\text{Entropy} = -\sum_{i,j} P(i,j)\log(P(i,j)) \quad (4)$$

$$\text{Correlation} = \sum_{i,j} \frac{(i - \mu i)(j - \mu j)p(i,j)}{\sigma_i \sigma_j} \quad (5)$$

$$\text{Homogeneity} = \sum_{i,j} \frac{P(i,j)}{1 + |i - j|} \quad (6)$$

Algorithm: For calculating GLCM measures for each pixel:

1. Read the input image.

2. Convert the data type to double and Zero pad the image
3. Extract a 3×3 window image from the input image and compute the co-occurrence texture measure
4. Estimate the texture parameters for the obtained texture image
5. Repeat the step3 and step4 by moving the window till the end of the image
6. Display various texture parameters by normalizing them

Classification: Classification is a technique to detect the dissimilar texture regions of the image based on its features. It can be used to cluster the feature sets of the image that characterized as different regions. A frequently used clustering algorithm is the k-means algorithm. In this study, we have used k-means algorithm for classifying the texture regions of the image so that different regions of the texture image have been identified in order to increase the performance of the retrieval by comparing the classified texture image with user's query image (AlaguRaja and SathyaBama, 2010).

K-means clustering: K-means clustering is a simple algorithm to clustering the texture regions of an image. For K clusters $\{C_1, C_2 \ldots C_K\}$ each with n_k patterns aim to find cluster centers m_k to minimize the cost function E_K^2 where Eq. 7 and 8:

$$M_K = \frac{1}{n_k}\sum_{x\hat{I}C_k} X \qquad (7)$$

$$E_k^2 = \sum_{k=1}^{k}\sum_{x\hat{I}C_k} || X - m_K ||^2 \qquad (8)$$

The initial cluster midpoints are selected randomly and the algorithm is applied repeatedly until a fixed state level is arrived.

Algorithm: For K-means clustering:

1. Initialize cluster centers randomly in texture image
2. For all the pixels in the image do the following
   a) Compute the Euclidean distance of the feature vector from the cluster for every other cluster.
   b) Assign the pixel to that cluster whose center yields the minimum distance from the feature vector
3. Update the cluster centers by computing the mean of the feature vectors of the pixels belonging to that cluster
4. Between two consecutive updates, if the changes in the cluster centers are less than a specified value, then stop
   Else go to step 2

Similarity comparison: For similarity comparison, we have used Euclidean distance, d is using the following Eq. 9:

$$d = \sqrt{\sum_{i=1}^{N}(F_Q[i] - F_{DB}[i])^2} \qquad (9)$$

Where:

$F_Q[i]$ = The ith query image feature
F_{DB} [i] = The corresponding feature in the feature vector database

Here N refers to the number of images in the database (Nandagopalan *et al*., 2008).

MATERIALS AND METHODS

The main phases of CBIR processes are Image Pre-processing, Feature Extraction, Classification and Retrieval.

Preprocessing: In the preprocessing stage, the main goal is to prepare the image for feature extraction by using the traditional segmentation process that has been carried out. After preprocessing step images are used for feature extraction process.

Feature extraction: In this step, the main goal is to extract texture feature by Using Gray Level Co-Occurrence Matrix (GLCM). The main objective of the system is maximum utilization of GLCM for extracting texture feature. In this system, texture features are extracted using GLCM technique and which has been given as input image to the classification phase.

Classification: Classification is a technique to detect the dissimilar texture regions of the image based on its features. It can be used to cluster the feature sets of the image that characterized as different regions. Frequently used clustering algorithm is k-means algorithm which has been used in this study in order to differentiate the different texture regions of the image. Finally, this output image only has been considered for measuring the similarity between the query image and database images.

Table 1: Precision and recall values in %

	Texture	
Query image	Precision	Recall
1	53.0	17.0
2	69.0	26.0
3	40.0	10.0
4	70.0	33.0
5	55.0	23.0

Retrieval: For retrieval, Euclidean distance calculation has been calculated between the query image and database images. This Euclidean distance value only then considered to compare similarities between the query image and database images. Finally, based on closest distance query image to the database images, images were ranked and retrieved in order to get an exact desired image of the user.

By performing above mentioned phases, retrieval efficiency was calculated using traditional precision and recall parameters which are presented in Table 1.

RESULTS

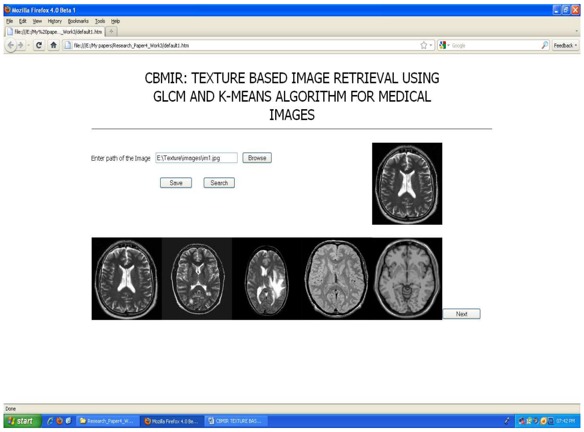
Experimental setup and results: A Intel® Core 2 Duo CPU Workstation with 2GB RAM computer is used for conducting the experiments. The main browser tool Mozilla Firefox 4.0 Beta1 version was used for developing User Interface components as a front end, MATLAB-Image Processing tool Box-Workspace was used as the feature database for storage as back end and for image processing study, other MATLAB utilities were used. For mathematical equations, Math Type tool was also used for writing documents. Initially, the MATLAB workspace database with 1000 real time medical images were used for testing the proposed CBIR system. The sample snapshot of the CBIR Model user interface for medical Images is shown in Fig. 2 (Ramamurthy and Chandran, 2011).

The following Fig. 3 shows the graphical representation of the result values given in Table 1.

The following Fig. 4 shows Sample Image Results used in the Model for Texture feature extraction and it gives about 90% of the accuracy of energy, entropy, correlation and homogeneity measurements.

The following table shows precision and recall values for the query images presented in Fig. 4.

Retrieval efficiency: For retrieval efficiency calculation, precision and recall values were calculated for randomly selected query images from 1000 medical images from MATLAB Workspace database. Benchmark formulas have been used to calculate the precision and recall values:


Mozilla Firefox 4.0 Beta 1
File Edit View History Bookmarks Tools Help
file:///E:/My%20pape..._Work3/default1.htm
file:///E:/My papers/Research_Paper4_Work3/default1.htm
Google
Feedback
Browse
Save
Search
Next
Done
start
Research_Paper4_W...
Mozilla Firefox 4.0 Be...
CBMIR TEXTURE BAS...
07:42 PM

CBMIR: TEXTURE BASED IMAGE RETRIEVAL USING GLCM AND K-MEANS ALGORITHM FOR MEDICAL IMAGES

Enter path of the Image: E:\Texture\images\im1.jpg

Fig. 2: Sample snapshot of CBIR Model for medical images

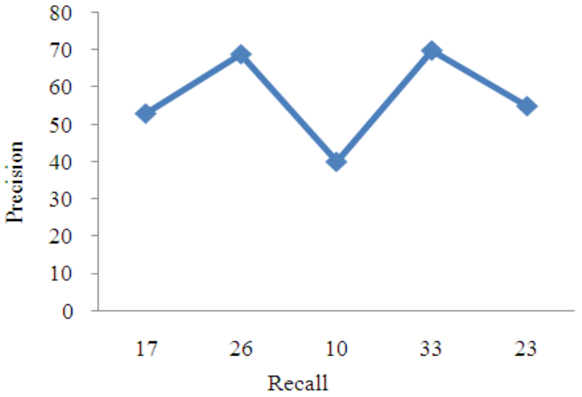


Fig. 3: Graphical representation for the precision and recall values given in Table 1

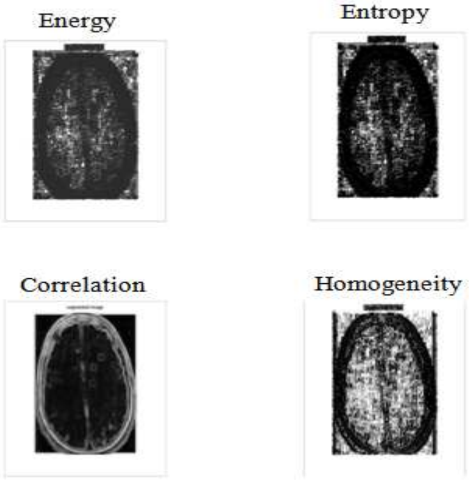


Fig. 4: Sample image results used in the model for texture feature extraction

$$\text{precision} = \frac{\text{No.of relevant images retrieved}}{\text{Total No.of images retrieved}}$$

$$\text{recall} = \frac{\text{No.of relevant images retrieved}}{\text{Total No.of relevant images in the Database}}$$

DISCUSSION

Initially, the MATLAB workspace database with 1000 real time medical images were used for testing the proposed CBIR system. These images were taken from the CasImage database in different categories for testing purpose. In feature extraction phase, the results were quite good and it gives about 90% of the performance for texture feature extraction. In classification phase, to characterizing and recognizing the different regions of the image, most of the images perform well and it gives about 60-70% of the performance of this phase. In retrieval phase, almost all the images perform well and traditional Euclidean distance method was used for retrieval and it gives better performance result. For measuring retrieval performance, traditional parameters such as precision and recall measurements were used and their results are presented in Table 1 and corresponding graphical representation also presented in Fig. 3. For precision, the system gives minimum 40% to a maximum 70% of the result and to recall it gives minimum 10% to a maximum 33% of the result. It shows that precision gives better performance in relevant image retrieval out of the retrieved images and recall gives its own performance in relevant image retrieval out of total images available in the database. Also the system provides a very Good User Interface (GUI) for retrieving the images. Hence the system has been developed successfully in an effective manner by achieving the targeted output. The system is designed with a flexible and consistent flow for easy understanding.

CONCLUSION

This study proposed a model for the Content Based Medical Image Retrieval System by using texture feature in calculating the Gray Level Co Occurrence matrix (GLCM) from which various statistical measures were computed in order to increasing similarities between query image and database images for improving the retrieval performance. The system has been developed successfully in an effective manner by achieving the targeted output. The system is designed with a flexible and consistent flow for easy understanding.

REFERENCES

AlaguRaja, R.A. and B.S. Bama, 2010. Two day workshop on recent trends in satellite/image processing. Conducted by Thiagarajar College of Engineering, Madurai, India.

Glatard, T., J. Montagnat and I.E. Magnin, 2004. Texture based medical image indexing and retrieval: Application to cardiac imaging. Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, Oct. 10-16, ACM, New York, USA., pp: 135-142. DOI: 10.1145/1026711.1026734

Haralick, R.M., K. Shanmagan and I. Dinstein, 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernetics, 3: 610-621. DOI: 10.1109/TSMC.1973.4309314

John, E. and M. Graham, 1999. Content-based image retrieval. University of Northumbria in Newcastle.

Long, L.R., S.K. Antani and G.R. Thoma, 2005. Image informatics at a national research center. Comput. Med. Image. Graphics, 29: 171-93. DOI: 10.1016/j.compmedimag.2004.09.015

Nandagopalan, S., B.S. Adiga and N. Deepak, 2008. A universal model for content-based image retrieval. World Acad. Sci. Eng. Technol., 46: 644-647.

Ozden, M. and E. Polat, 2005. Image segmentation using color and texture features. Kırıkkale University, Turkey.

Park, M., J.S. Jin and L.S. Wilson, 2004. Detection of abnormal texture in chest x-rays with reduction of ribs. Proceedings of the Pan-Sydney Area Workshop on Visual Information Processing, (VIP' 05), ACM, Australian Computer Society, Inc. Darlinghurst, Australia, pp: 71-74.

Partio, M., B. Cramariuc, M. Gabbouj and A. Visa, 2002. Rock texture retrieval using gray level co-occurrence matrix. Tampere University of Technology.

Petrakis, E.G.M., C. Faloutsos and K.I. Lin, 2002. Image Map: An image indexing method based on spatial similarity. IEEE Trans. Knowl. Data Eng., 14: 979-987. DOI: 10.1109/TKDE.2002.1033768

Ramamurthy, B. and K.R. Chandaran, 2011. CBMIR: Shape-Based Image Retrieval using canny edge detection and k-means clustering algorithms for medical images. Int. J. Eng. Sci. Technol., 3: 1870-1877.

Remco, C. and V.M. Tanse, 2000. Content based image retrieval systems. A Survey. Technical Report UU-CS-2000-34, pp: 1-62.

Sohail, A.S.M., P. Bhattacharya, S.P. Mudur, S. Krishnamurthy and L. Gilbert, 2010. Content-based retrieval and classification of ultrasound medical images of ovarian cysts. Artif. Neur. Netw. Patt. Recog., 5998: 173-184. DOI: 10.1007/978-3-642-12159-3_16

Shyu, C.R., C.E. Brodley, A.C. Kak, A. Kosaka and A.M. Aisen *et al*., 1999. ASSERT: A physician-in-the-loop content-based retrieval system for HRCT image databases. Comput. Vis. Image Understand., 75: 111-132. DOI: 10.1006/cviu.1999.0768

Thies, C., M.O. Guld, B. Fischer and T.M. Lehmann, 2005. Content-based queries on the CasImage database within the IRMA framework. Lecture Notes Comput. Sci., 3491: 781-792.

Thoma, G.R., L.R. Long and S. Antani, 2006. Biomedical Imaging research and development: Knowledge from images in the medical enterprise. U.S. National Library of Medicine.

Yin, D., J. Pan, P. Chen and R. Zhang, 2008. Medical image categorization based on Gaussian mixture model. Proceedings of the IEEE International Conference on Biomedical Engineering and Informatics, May 27-30, IEEE Xplore Press, Sanya, pp: 128-131. DOI: 10.1109/BMEI.2008.210