Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Hybrid Feature Database Creation for Content Based Image Retrieval

Sumathi Ganesan*, T. S. Subashini
Department of Computer Science and Engineering,
Faculty of Engg. and Technology, Annamalai University, Annamalai Nagar, India

Abstract: Content-based image retrieval (CBIR) is useful for manipulating large amount of image databases and archives and for identifying and retrieving similar images. Extraction of invariant features is the basis of CBIR. In this work an effort has been made to create a hybrid feature database of four different classes of X-ray images namely chest, spine, foot and palm. This paper focuses on the problem of texture and shape feature extraction and feature database creation. We investigate texture feature and shape feature for CBIR by combining GLCM and ZM. GLCM is used for texture feature extraction and ZM extracts shape features. The extracted features are classified using SVM into any of the four classes and the X-ray image along with its corresponding features is stored in the database. This database forms the basis for the next step of CBIR namely matching and retrieval.

Keywords: Content based image retrieval, X-ray, Zernike Moments, Gray Level Co-occurrence Matrix and Support Vector Machine.

I. INTRODUCTION

Today medical images are available in digital format and areeffectively used for diagnostics and therapy. Several methodsfrom the computer vision and image processing fields alreadyhave been proposed for the use in medicine. Content basedaccess to medical data has been proposed that would ease themanagement of clinical data. Intensive research is beingcarried out on content-based image retrieval of images forsupporting clinical decision making and its relevance toradiology practice is highly felt nowadays. In addition toenabling similarity-based indexing for identifying andretrieving similar images, CBIR also provides computer-aideddiagnostic support based on image content. Current work inmedical image processing has markedly increased therelevance of CBIR to radiology practice.

In general a CBIR framework consists of two major buildingblocks. The first one represents the visual information present in the image as features/descriptors. The second componentanalyses the similarity between the query image and theimages in the database. This is done by comparing featuresacross different images using mathematical methods. In thiswork an effort has been made to implement the firstcomponent of CBIR by creating an indexed feature database.GLCM is used for texture feature extraction and ZM for shapefeature extraction. SVM classifier is used for imageclassification and the image along with the feature is stored inthe database. This database forms the basis for the secondcomponent of CBIR namely matching and retrieval.

II. PREVIOUS WORK

A main challenge for the development of CBIR systems is theappropriate characterization of images and the creation offeature database. Literature survey shows that different typesof visual features are extracted for CBIR. Color and grey levelhistogram features is used in [1]. The work in [2] uses localand global grey level features. Fourier descriptors are use tocharacterize shapes in [3] and also uses invariant moments andscale space filtering. Features derived from co-occurrencematrices are used in [4]. Markov texture characteristics areused in [5].Segmentation is a preprocessing step in CBIR. Regiongrowing is used as a segmentation method in the workproposed by [6] to detect the region of interest in the X-rayimage. The work in [7] uses Model based segmentation usingK-Means and Support Vector machine to segment masses inmammograms. In contrast, in [8] generalization of K-Meanswhich includes spatial information to refine an initialsegmentation is used. The work in [9] proposes a Markovrandom field (MRF) based technique that is suitable forperforming clustering in an environment which is described bypoor or limited data. In this [10] present a comparativeapproach for classification using neural network classifier andGaussian mixture model. Paper in [11] uses AANN andstatistical features extracted from already marked suspiciousareas by the radiologists to classify the breast mass. Randomforest is an ensemble classifier that consists of many decisiontrees and the work proposed in [12] uses random forestdecision classifier.Classification of breast masses using shape, texture and edgesharpnesswith linear and kernel-based classifiers in [13]. Inthis [14] proposes, Wavelet-based features are extracted fromimages and classified using radial basis function neuralnetwork and support vector machines. Fuzzy modelling methods including the fuzzy K-nearest neighbor algorithm, afuzzy clustering based modeling, and the adaptive networkbased fuzzy inference system are employed in [15]. The workin [16], presents a novel fuzzy classification framework for theautomatic extraction of fuzzy rules from labeled numericaldata. The paper [17] uses and compares Bayesian classificationand K-nearest neighbor classification for cancer detection.

III. DATA SOURCE

The X-ray images used for this work are from the IRMAImageCLEFmed 2008 database. This collection compilesanonymous radiographs, which have been arbitrarily selectedfrom routine at the Department of Diagnostic Radiology,Aachen University of Technology (RWTH), Aachen,Germany. The imagery represents different ages, genders,view positions and pathologies. Therefore, image qualityvaries significantly. All images were downscaled to fit into a512 x 512 bounding box maintaining the original aspect ratio.

IV. PROPOSED WORK

Database creation is carried out in this work whichforms the initial step in CBIR systems. Four classes of X-rayimages namely chest, foot, spine and palm images are taken upfor the study. The block diagram for the proposed featuredatabase creation system is shown in Fig. 1

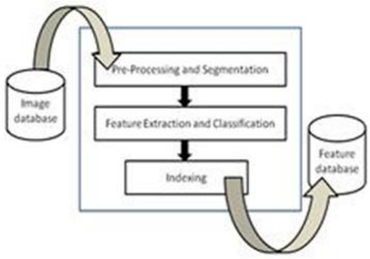

Fig. 1 Block diagram of the proposed work

4.1 Pre-Processing and Segmentation

In this step unwanted noise is removed from the X-ray images. Range filtering is applied to the X-ray images to reduce the distortions and to enhance the image for further processing. The range filter filters the image by replacing every pixel by the difference of maximum and minimum of its neighborhood. Then the filtered images are segmented to avoid the unnecessary computations and misclassifications. The preprocessed images are segmented using adaptive thresholding and connected component labeling [5]. By CCL, various connected regions on the binary image are obtained and only the biggest connected region which is the body part examined is retained and other regions are eliminated[23] and [24]. After segmentation the images are ready for feature extraction.

4.2 Feature Extraction

In this work shape and texture features are extracted by combining Zernike moments and Grey Level Co-occurrence Matrix. GLCM is used for texture feature extraction and ZM extracts shape features.

4.2.1 Shape Feature Extraction

Shape is an important property in the medical image processing domain. X-ray images with similar shapes may belong to the same category. Zernike moments have been reported to be a good descriptor for shape description [9]. Therefore, this study uses the Zernike moments for describing the medical X-ray images because of the methods' properties: (1) the Zernike basis function satisfies the orthogonal property, implying that the contribution of each moment coefficient to the underlying image is unique and independent, i.e. no redundant information overlap between the moments; (2) calculation of the Zernike moments do not require knowledge of the precise boundary of an object. This makes Zernike moments suitable for representing X-ray images with obscure boundaries.

4.2.2 Texture Feature Extraction

Co-occurrence matrix [6] is one the most traditional techniques for encoding texture information. Texture is one of the most important defining characteristics of an image. It describes spatial relationships among grey-levels in an image. A cell defined by the position (i, j) in this matrix registers the probability at which two pixels of gray levels i and j occur in two relative positions. A set of co-occurrence probabilities (such as, energy, entropy, contrast) has been proposed to characterize textured regions [9].

In our work we are using the Gray Level Co-occurrence Matrix (GLCM) for extracting five important texture features reported in the literature viz.,

$$Contrast=\sum_{n=0}^{Ng-1} n^2 \left\{ \sum_{i=1,j=1}^{Ng\,Ng} \sum p(i,j) \right\}, |i,j|$$

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

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

$$Correlation = \sum_{i,j} P(i-\mu i)(j-\mu j)P(i,j)/\sigma_i\sigma_j$$

$$Homogenity = \sum_{i,j} \frac{P(i,j)}{1+|i-j|}$$

where, p_d is the probability matrix obtained through GLCM; $\mu_x$and μ_y are the means and σ_x and σ_y are standard deviations of $\rho_{d(x)}$ and $\rho_{d(y)}$ respectively.

4.3 Classification

In this step totally ten shape and texture features that were extracted is classified using SVM. SVMs are powerful machine learning techniques for classification and regression. They were developed by Vapnik [29], and are based on statistical learning theory. This gave rise to a new class of theoretically elegant learning machines that use a central concept of support vectors and kernels for a number of learning tasks. Kernel machines provide a modular framework that can be adapted to different tasks and domains by the choice of the kernel function and the base algorithm [17]. The support vector machine (SVM) method is considered a good classifier since it is able to predict correctly the class of the new data from the same domain where the learning occurred [18].

The goal of SVM modeling is to find the optimal hyper plane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. The vectors near the hyper plane are the support vectors. Fig. 2 presents an overview of the SVM process. In this work, the classification is employed for four different classes and each class is trained and tested with the test image.

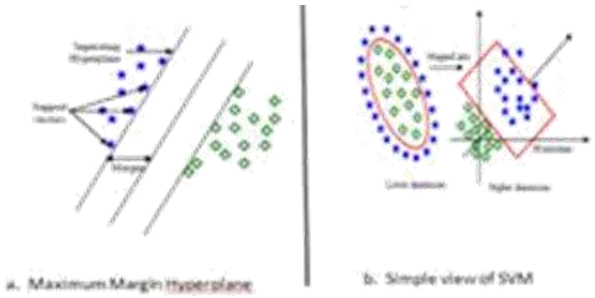

Fig. 2. Overview of SVM

4.4 Hybrid Feature Database Creation

The given image is preprocessed, segmented and classified into any of the four class of X-ray using the ten hybrid shape and texture features extracted. A simple two bit coding structure is used to represent the four classes of X-rays and this code is used to index into the feature database. The feature database is created in access and linked with the corresponding X-ray image using OLE. This forms the basic step in CBIR and this database will be used in the retrieval process based on matching and similarity metrics.

V. EXPERIMENTAL RESULTS

This work is done in MATLAB. First the X-ray images are preprocessed and segmented using range filter and CCL and the results are shown in Fig. 3(a), 3(b) and 3(c).

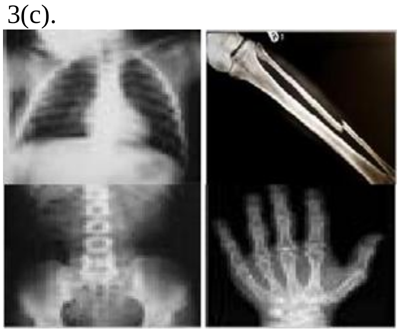

Fig. 3(a). Input X-ray Images

Fig. 3(b). Filtered X-ray Images

Fig. 3(c). Segmented X-ray Images

Then five shape and five texture features are extracted using ZM and GLCM. The results are shown in Figs. 4 and 5 from which it can be understood that the combination of the shape and texture features are more suitable for the classification of X-ray images under test-study. Its accuracy has also been found to be increased as compared to its rivals [31 & 32].

Fig. 4. Zernike Moments

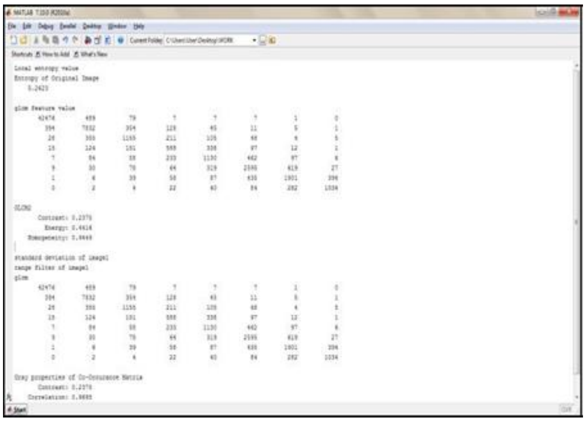

Fig. 5. Texture Features using GLCM

In this step the features that were extracted are fed to the SVM classifier and classified into any of the four classes of X-rays. The accuracy of the classifier obtained is 94.3%. Fig. 6 shows the snapshot of the SVM which correctly classified the hand X-ray images.

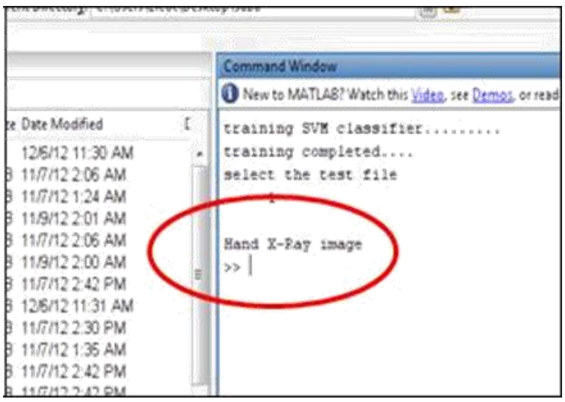

Fig .6. SVM Classifies the given X-ray Image as Hand

Based on the classification results the images along with its hybrid features are stored in the Access database. The user interface is done in Visual Basic which is shown in Fig. 7 and Fig. 8.

Fig. 7. Loading Image

Fig. 8. User Interface showing the X- ray image and its corresponding features stored in the feature database created

VI. CONCLUSION

In this work feature database of four different classes of X-ray images namely chest, spine, foot and palm was created by extracting both shape and texture features. The texture feature and shape feature for CBIR was investigated by combining GLCM and ZM. GLCM is used for texture feature extraction and ZM extracts shape features. SVM classifier was used to categorize X-rays into any of the four class.

After classification the X-ray image along with its hybrid features are stored in Access database for further retrieval The work is done in MATLAB and Visual basic is used to create the user interface. This database forms the basis for the next step of CBIR namely matching and retrieval.

REFERENCES

[1] S. Antani, L. R. Long, G. R. Thoma, A biomedical information system for combined content-based retrieval of spine X-ray images and associated text information, in: Proceedings of the 3rd Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2002), Ahmadabad, India, 2002.

[2] Baek Hwan Cho, Hwanjo Yu, Jongshill Lee, Young JoonChee, In Young Kim, and Kim S.I., “Nonlinear support vector machine visualization for risk factor analysis using mammograms and localized radial basis function kernels,” IEEE Transactions on Information technology in Biomedicine, vol. 12, no. 2, pp. 247–256, March 2008.

[3] S.Belciug, “Bayesian classification vs. k-nearest neighbor classification for the non- invasive heptatic cancer detection,” in Proceedings of the 8th International conference on artificial Intelligence and Digital Communications, Research Notes in Articial Intelligence and Digital Communications, 2008, pp. 31–35.

[4] C. Brodley, A. Kak, C. Shyu, J. Dy, L. Broderick, A. M. Aisen, Content{based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision, in: Proceedings of the 10th National Conference on Artificial Intelligence,Orlando, FL, USA, 1999, pp. 760-767.

[5] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classification, John Wiley & Sons, Singapore, 2003.

[6] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Pearson Education, Singapore, 2002.

[7] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernatics, 3(6):610– 621, 1973.

[8] IoannisGadaras and LudmilMikhailov, “An interpretable fuzzy rule-based classification methodology for medical diagnosis,” Artificial Intelligence in Medicine, vol. 47, no. 1, pp. 25–41, September 2009.

[9] Lehmann, T “IRMA X- Ray Libraryurl: http://irma-project.org.

[10] Leonardo de Oliveira Martins, Geraldo Braz Junior, Aristofanes Correa Silva, Anselmo Cardoso de Paiva, and Marcelo Gattass, “Detection of masses in digital mammograms using k-means and support vector machine,” Electronic Letters on Computer Vision and Image Analysis, vol. 8, no. 2, pp. 39–50, 2009.

[11] H. D. Li, R. A. Clark, and J .A. Thomas, “Computer-aided diagnosis of masses with full-field digital mammography,” Academic Radiology, vol. 9, no. 1, pp. 4– 12, 2002.

[12] M. E. Mattie, L. Staib, E. Stratmann, H. D.Tagare, J. Duncan, P. L. Miller, PathMaster: Content-based cell image retrieval using automated feature extraction, Journal of the American Medical Informatics Association 7 (2000) 404-415.

[13] H. Muller, A. Rosset, J.P. Vallee, A. Geissbuhler, Integrating content-based visual access methods into a medical case database, in: Proceedings of the Medical Informatics Europe Conference (MIE 2003), St. Malo, France, 2003.

[14] H. Osta, R. Qahwaji, and S. Ipson, “Wavelet-based feature extraction and classification for mammogram images using RBF and SVM,” in Proceeding Visualization, Imaging, and Image Processing, 2008, p. 630.

[15] N. Petrick, H. P. Chan, B. Sahiner, and M. A. Helvie, "Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms," MedicalPhysics, vol. 26, no. 8, pp. 1642–1654, 1999.

[16] T. Pun, G. Gerig, O. Ratib, Image analysis and computer vision in medicine, Computerized Medical Imaging and Graphics 18 (2) (1994) 85-96.

[17] M.Rahman, B. C. Desai, and P. Bhattacharya. Supervised machine learning based medical image annotation and retrieval. In Image CLEFmed, pages 692–701, 2005.

[18] U. RajendraAcharya, E. Y. K. Ng, Y. H. Chang, J. Yang, and G. J. L. Kaw, "Computer-based identification of breast cancer using digitized mammograms,"Journal of Medical Systems, vol. 32, no. 6, pp. 499–507, December 2008.

[19] B. Ramamurthy, K. R. Chandran, S. Aishwarya, and P.Janaranjani, "CBMIR: Content Based Image Retrieval using Invariant Moments, GLCM and Grayscale Resolution for MedicalImages", European Journal of Scientific Research, ISSN 1450-216X ol.59 No.4 (2011),pp. 460-471.

[20] RinkuPancha and BrijeshVerma, "Digital mammograms using auto-associator neural network," Bioinformatics and Biomedical Applications, vol. 4234, pp. 127–136, 2006.

[21] Sean N. Ghazavia and Thunshun W. Liao,"Medical data mining by fuzzy modeling with selected features," Artificial Intelligence in Medicine, vol. 43, no. 3, pp. 195–206, July 2008.

[22] Shyu C, Brodley C, Kak A, Kosaka A, Aisen A, Broderick L. Local versus global features for content-based image retrieval; Proc IEEE Workshop on Content-Based Access of Image and Video Libraries; 1998. pp. 30–34.

[23] T.S.Subashini, V.Ramalingam, S.Palanivel, "Automated assessment of breast tissue density in digital Mammograms", in International Journal of Computer Vision and Image Understanding (Elsevier), no.114, pp. 33-43, 2010.

[24] T.S.Subashini, V.Ramalingam, S.Palanivel, "Breast mass classification based on cytological patterns using RBFNN and SVM", in Internationa Journal of Expert Systems with Applications (Elsevier), no. 36, pp. 5284-5290, 2009.

[25] M. Suliga, "Markov random field-based clustering applied to the segmentation of masses in digital mammograms," Computerized Medical Imaging and Graphics, vol. 32, no. 6, pp. 502–512, September 2008.

[26] L. H. Tang, R. Hanka, H. H. S. Ip, R. Lam, Extraction of semantic features of histological images for content{based retrieval of images, in: Proceedings of the IEEE Symposium on Computer{Based Medical Systems (CBMS 2000), Houston, TX, USA, 2000.

[27] Tingting Mu, Asoke K. Nandi, and Rangaraj M. Rangayyan, "Classification of breast masses using selected shape, edge-sharpness, and texture features with linear and kernel-based classifiers," Journal of Digital Imaging, vol. 21, no. 2, pp. 153–169, June 2008.

[28] K.Vaidehi T. S.Subashini V. Ramalingam S. Palanivel M. Kalaimani , "Transform based approaches for Palmprint Identification", International Journal of Computer Applications (0975 – 8887) Volume 41– No.1, March 2012.

[29] V.Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[30] L. Vibha, G. M. Harshavardhan, K. Pranaw, P. DeepaShenoy, K. R. Venugopal, and L. M. Patnaik, "Classification of mammograms using decision trees," in Proceedings of the 10th International Database Engineering and Applications Symposium, 2006, p.263 266.

[31] Sumathi Ganesan and T.S.Subashini, "An approach toward the Efficient Indexing and Retrieval on Medical X-ray Images", International Journal of Computer Applications, 2013.

[32] Sumathi Ganesan and T.S.Subashini, "A Content Based Approach to Medical X-ray Image Retrieval using Texture Features", International Journal of Computers & Technology, 2014.

