RETRIEVAL AND CLASSIFICATION OF ULTRASOUND IMAGES OF OVARIAN CYSTS COMBINING TEXTURE FEATURES AND HISTOGRAM MOMENTS

Abu Sayeed Md. Sohail[1], Md. Mahmudur Rahman[2], Prabir Bhattacharya[3], Srinivasan Krishnamurthy[4], Sudhir P. Mudur[5]

[1,5]Dept. of Computer Science and Software Engineering, Concordia University, Montreal, Canada
[2]U.S. National Library of Medicine, National Institutes of Health, Bethesda, Maryland
[3]Dept. of Computer Science, University of Cincinnati, Cincinnati, Ohio
[4]Dept. of Obstetrics and Gynecology, Royal Victoria Hospital, Montreal, Canada
E-mails: {[1]a_sohai, [5]mudur}@encs.concordia.ca, [3]bhattapr@ucmail.uc.edu, {[2]mahrahm73, [4]krishnaindu}@gmail.com

ABSTRACT

This paper presents an effective solution for content-based retrieval and classification of ultrasound medical images representing three types of ovarian cysts: *Simple Cyst, Endometrioma,* and *Teratoma*. Our proposed solution comprises of the followings: extraction of low level ultrasound image features combining histogram moments with Gray Level Co-Occurrence Matrix (GLCM) based statistical texture descriptors, image retrieval using a similarity model based on Gower's similarity coefficient which measures the relevance between the query image and the target images, and use of multi-class Support Vector Machine (SVM) for classifying the low level ultrasound image features into their corresponding high level categories. Efficiency of the above solution for ultrasound medical image retrieval and classification has been evaluated using an in-progress database, presently consisting of 478 ultrasound ovarian images. Performance-wise, in retrieval of ultrasound images, our proposed solution has demonstrated above 77% and 75% of average precision considering the first 20 and 40 retrieved results respectively, and an average classification accuracy of 86.90%.

Index Terms— Ultrasound Medical Image Retrieval, Classification of Ultrasound Ovarian Images, Statistical Texture Descriptors, Histogram Moments

1. INTRODUCTION

This paper addresses the issue of efficiently representing ultrasound images and their retrieval and classification towards developing a computer-aided diagnosis system for ultrasound ovarian abnormalities. Rapid developments in the field of imaging have made it relatively easy to acquire medical images during the diagnostic process. Since the volume of images is increasing very fast, retrieving and analyzing images with similar cases/situations from a large database is an important problem that is of current interest. Medical images have been frequently used in developing and analyzing the performance of image retrieval and classification systems. This has eventually lead medical domain to be explored and cited as one of the principal application domains of image retrieval and classification in terms of potential for high impact. A large number of approaches have already been proposed for content-based retrieval and classification of medical images including radiology images, X-ray images, CT images of lung, dermatology images, MRI images of heart and brain, ultrasound images of kidney and breast [1, 2]. However, to the best of our knowledge, no published research has yet reported the application of content-based image retrieval and image classification technique over ultrasound images of ovarian abnormalities.

Accurate profiling of ultrasound ovarian images is very important to arrive at a diagnosis, but their inherent heterogeneity makes it rather difficult. Presently, visual ultrasound examination is considered as the most widely accepted and practiced diagnostic modality for non-invasive assessment of ovarian cysts and other types of ovarian abnormalities [3]. A number of ultrasound-based algorithms have been proposed for this purpose, but recognition of inherent patterns through visual ultrasound observation remains the best way for assessing their nature and category. However, this method largely depends on accumulation of practical experience in identifying the morphology and characteristics of various types of ovarian abnormalities present in their corresponding ultrasound images. As a result, inexperienced ultrasound operators always encounter difficulties in differentiating among different types of cysts, which eventually lead to a lower rate of correct diagnosis. Since incorrect diagnosis can either result in unnecessary biopsies/surgery, or worse, missed cases, there is a need for inexperienced operators to be given supporting tools to help increase their diagnostic accuracy. A computer based system for retrieval and classification of ultrasound images could serve the purpose of such a decision support tool in the diagnosis of ovarian abnormalities.

The subsequent discussions of this paper have been organized as follows:

Procedures for extracting and combining histogram moments and GLCM based texture descriptors have been discussed in Section 2. Section 3 gives a brief theoretical formulation of the techniques used for image similarity matching and image classification. Section 4 discusses the experimental results achieved by an implementation of our proposed solution and lastly, Section 5 concludes the paper summarizing our main contributions and planned future work.

2. LOW LEVEL ULTRASOUND IMAGE FEATURES

Our proposed solution for ultrasound image retrieval and classification consists of three sequential processing stages: (i) feature extraction and fusion, (ii) image retrieval by similarity matching, and (iii) image classification using SVM. As automated segmentation of ultrasound ovarian images is still considered as an open problem, our method requires human interaction in selecting the Region of Interest (ROI) for feature extraction from ultrasound images. Once the region has been manually specified through a number of ROI boundary points on the image surface, the

remaining processes of feature extraction, feature fusion, image retrieval and classification is performed in an automated manner without requiring any further input from the user.

2.1. Histogram Moments for Ultrasound Images

One of the most widely used visual features in image retrieval and classification is the histogram based feature. The histogram of an image f_n is an N-dimensional vector $H(f_n, i)$; $i = 0,1,2,\cdots, N-1$, where N is the number of gray levels and $H(f_n, i)$ is the number of pixels with gray level value i. Histogram feature is relatively robust to background complication. Besides, it is also insensitive to changes such as image size, rotation and slight transition, each of which has little or no impact on the distribution of the gray levels in an image. However, the disadvantage of using histogram feature is that different images can have similar overall histogram in a large image database. Consequently, this can lead to poor performance in terms of retrieval and classification accuracy. Another drawback of using histogram features for ultrasound image classification is that an ultrasound image can have histogram with many empty bins due to the quantization process involved in the imaging system. As a result, slight changes in illumination may cause a shift in the histogram probability density function (pdf), which can introduce huge changes between the set of features obtained from two similar images. The moment-based approach can minimize these problems since it smoothes the histogram pdf. The histogram distribution can be interpreted as a probability distribution and can be characterized by its moments. Stricker and Orengo [4] introduced a method using the central moment approach. Mandal et al. [5] used orthogonal Legendre moments for histogram indexing, which resulted in a better performance than regular or central moments. Legendre moments are based on orthogonal Legendre polynomials. A Legendre polynomial is defined as follows:

$$P_k(x) = \frac{1}{k!\,2^k}\frac{d^k}{dx^k}(x^2-1)^k = \sum_{j=0}^{k} a_{kj}x^j \qquad (1)$$

The value of a_{kj} can be expressed as:

$$a_{kj} = \sum_{\substack{j=0 \\ j,k\ odd \\ j,k\ even}}^{k} (-1)^{(k-j)/2}\frac{(k+j)!}{2^k\left(\frac{k-j}{2}\right)!\left(\frac{k+j}{2}\right)!\,j!}$$

The k-th Legendre moment of a function $f(x)$ is defined by

$$\lambda_k = \frac{2k+1}{2}\int_{-\infty}^{\infty} P_k(x)f(x)dx \qquad k \geq 0, k \in \mathbb{Z}$$

Replacing the value of $P_k(x)$ from Eq. (1), and applying the definition of centrals moments, λ_k can be expressed as:

$$\lambda_k = \frac{2k+1}{2}\sum_{j=0}^{k} a_{kj}M_j$$

It can be observed that the Legendre moment of any order depends only on the regular moments of the same order and lower and can be calculated easily using the above equation.

Generally, the image grey levels range from 0 to 255. Since, the Legendre polynomials are orthogonal only in the interval $[-1, 1]$, the dynamic range of the pdf has to be mapped onto this interval. The histogram pdf function $f(x)$ can then be written as an infinite series expansion in terms of the Legendre polynomials as:

$$f(x) = \sum_{k=0}^{\infty} \lambda_k P_k(x)$$

From this equation, the image *pdf* can be reconstructed using first $(N+1)$ moments as follows:

$$f'(x) = \sum_{k=0}^{N} \lambda_k P_k(x)$$

The reconstructed pdf function $f'(x)$ is then free from any quantization effect and has no empty bins. The optimum number of moments required for reconstructing the image pdf accurately is an important concern. Usually, 10-16 moments can give a good representation of an image pdf when it does not contain any sharp peak [5]. Ultrasound ovarian images usually contain sharp peaks with empty bins and therefore cannot be represented with few moments. To quantify the reconstruction efficiency as well as the no. of moments required to better represent ultrasound images, we have calculated the Signal to Error Ratio (SER) defined by Mandal et al. [5] as follows:

$$SER = \frac{\int f^2(x)dx}{\int [f(x)-f'(x)]^2 dx}$$

Where, $f(x)$ and $f'(x)$ are the original histogram and the moment reconstructed histogram respectively. Table 1 shows the percentage of images having a SER greater than a certain threshold value using 8, 16, 32, 64, 128 moments.

Table 1: SER of the re-constructed image histogram pdf with finite number of moments. Using 128 moments, 98% of the reconstructed histograms have $SER > 8\ dB$.

	No. of Moments				
SER	8	16	32	64	128
> 8 dB	52%	79%	91%	97%	98%
> 9 dB	41%	72%	86%	97%	97%
> 10 dB	35%	67%	80%	93%	95%
> 11 dB	31%	63%	77%	91%	94%
> 12 dB	25%	57%	72%	88%	92%

As can be observed from Table 1, use of 64 and 128 moments demonstrates the best re-construction capability. However, calculation of 128 moments is computationally at least twice as expensive as calculating 64 moments. In addition, use of 64 moments instead of 128 has a significant impact on reducing the dimension of the extracted feature vector. Therefore, we opted for using the first 64 moments as image histogram based features.

2.2. GLCM Based Texture Feature from Ultrasound Images

The co-occurrence probabilities of GLCM provide a second-order method for generating texture features [6]. These probabilities represent the conditional joint probabilities of all pair wise combinations of grey levels in the spatial window of interest with respect to two parameters: inter-pixel distance (d) and orientation (θ). The probability measure can be defined as: $P(x) = \{C_{ij}|(d,\theta)\}$, where C_{ij}, the co-occurrence probability between grey level i and j, is defined as: $C_{ij} = \frac{P_{i,j}}{\sum_{i,j=1}^{G} P_{i,j}}$. Here $P_{i,j}$ represents the number of occurrences of grey level i and j within the given window, given a certain pair of (d,θ) pair; and G is the quantized number of grey levels. The sum in the denominator thus represents the total number of grey level pairs (i,j) within the window. For extracting GLCM based texture features from ultrasound ovarian images, we obtained four co-occurrence matrices from each image using $\theta = \{0, 45, 90, 135\}$ degree and $d = 1$ pixel. Then, 14-statistical texture descriptors have been calculated from each of these co-occurrence matrices as

proposed by Haralick et al. [6]. These descriptors are angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, two information measures of correlation, and maximal correlation coefficient. Therefore, a total of 56 texture features were extracted from each image using the four co-occurrence matrices.

2.3. Feature Fusion and Normalization

After extracting the histogram moments and GLCM based texture feature from an image, they are organized into a single feature vector. Each feature vector x_k, consisting of 120 features $(64 + 56)$, is then normalized as: $\hat{x}_k = \frac{x_k - \mu_k}{\sigma_k}$; where, μ_k and σ_k are the mean and standard deviation of feature vector x_k.

3. RETRIEVAL AND CLASSIFICATION FORMULATION

3.1. Similarity Model for Image Retrieval

For retrieving ultrasound images, Gower's similarity coefficient [7] based similarity model has been used. In this method, a combination of features to constitute a global similarity is done as an average of each of the individual similarities on each feature. The model is defined as follows [8]:

$$GS_{ij} = \frac{\sum_{k=1}^{n} S_{ij}^{(k)}}{\sum_{k=1}^{n} \delta_{ij}^{(k)}} \tag{2}$$

Here, $S_{ij}^{(k)}$ is the result of comparing image i and j on their feature k, and $\delta_{ij}^{(k)}$ represents the possibility of comparing image i and j on their feature k. In Eq. (2), $\delta_{ij}^{(k)} = 1$ if image i and j can be compared on feature k, otherwise, $\delta_{ij}^{(k)} = 0$. If the image i and image j can be compared across all the considered features, $\sum_{k=1}^{n} \delta_{ij}^{(k)} = N$, which is the dimension of the feature vector. So, global similarity GS_{ij} between images i and j is defined as an average of the similarities on each feature between image and i and j. The quantity $S_{ij}^{(k)}$ can be defined as follows:

$$S_{ij}^{(k)} = 1 - \frac{|x_{ik} - x_{jk}|}{R_k} \tag{3}$$

Where R_k represents a normalization factor and is calculated as: $R_k = Max(x_{ik}) - Min(x_{ik})$ where $x_{ik};\ i = 1,2,\cdots,n$ is the set of values taken by each of the image i of the sample considered for the feature k. $S_{ij}^{(k)} = 1$ if image i and j are identical and $S_{ij}^{(k)} = 0$ if they are completely different. $S_{ij}^{(k)}$ can take a positive value between 0 and 1 if the two images have a certain degree of similarity according to feature k. Using Eq. (3) and considering that all features can be compared, global similarity GS_{ij} between two images i and j, as defined in Eq. (2), can be re-written as:

$$GS_{ij} = \frac{1}{N} \sum_{k=1}^{n} \left(1 - \frac{|x_{ik} - x_{jk}|}{R_k}\right) \tag{4}$$

3.2. SVM for Image Classification

Support Vector Machine (SVM) belongs to the class of maximum margin classifiers. They perform pattern recognition between two classes by finding a decision surface that has the maximum distance to the closest points in the training set known as support vectors [9]. Unlike other classifiers, SVM controls its generalization ability by minimizing the error rate on the training set and their capacity. Let a training set of points be $x_i \in \mathbb{R}^n, i = 1,2,\cdots,N$; where each point of x_i belongs to one of the two classes identified by the label $y_i \in \{-1, 1\}$. Assuming linearly separable data, the goal of maximum margin classification is to separate the two classes by a hyperplane such that the distance to the support vectors is maximized. This hyperplane is known as the *Optimal Separating Hyperplane* and is expressed as:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i x_i \cdot x + b \tag{5}$$

The coefficients α_i and the b in Eq. (5) are the solutions of a quadratic programming problem. Classification of a new data point x is performed by computing the sign of the right side of Eq. (5). Each data point x is separated from the hyperplane as:

$$d(x) = \frac{\sum_{i=1}^{l} \alpha_i y_i x_i \cdot x + b}{\left\| \sum_{i=1}^{l} \alpha_i y_i x_i \cdot x \right\|} \tag{6}$$

The sign of d is the classification result for x, and $|d|$ is the distance from x to the hyperplane. Intuitively, the farther away a point is from the decision surface, *i.e.* the larger $|d|$, the more reliable the classification result. The entire construction can be extended to the case of nonlinear separating surfaces. Each point x in the input space is mapped to a point $z = \Phi(x)$ of a higher dimensional space, called the *feature space*, where the data are separated by a hyperplane. The key property in this construction is that the mapping $\Phi(\cdot)$ is subject to the condition that the dot product of two points in the feature space $\Phi(x) \cdot \Phi(y)$ can be rewritten as a kernel function $K(x, y)$. Using the kernel function, the decision surface is defined by:

$$f(x) = \sum_{i=1}^{l} y_i \alpha_i K(x, x_i) + b \tag{7}$$

where the coefficients α_i and b are the solutions of a quadratic programming problem and $f(x)$ is independent of the dimensionality of the feature space. Finally, multi-class classification was performed by arranging $\frac{q(q-1)}{2}$ 2-class SVMs in "pair-wise" top down tree structured approach proposed in [10]. Here, q represents the number of classes in the dataset, which is 3 for this particular problem.

4. PERFORMANCE ANALYSIS

The process of creating an ultrasound image database of various ovarian abnormalities is in progress. At present, the database includes 478 ultrasound ovarian cyst images collected during regular clinical practice at Royal Victoria Hospital, Montreal. We have used this database to measure the performance of our proposed solution for ultrasound ovarian image retrieval and classification. The images in the database were classified into three types of ovarian cysts: *Simple Cyst* (187-images), *Endometrioma* (154-images) and *Teratoma* (137-images). This categorization was performed by one or more domain experts and the categorization decisions were further verified by consulting the associated proven pathological diagnosis.

4.1. Retrieval Performance

To evaluate the retrieval performance, we have randomly selected 50 images of each category as the query images. We adopted the "Query by Example" method for submitting the query to the retrieval system where the query was specified by providing an

example image to the system. A retrieved image was considered a match if it belongs to the same category as that of the query image. For quantitative evaluation, retrieval performances of each category (simple cyst, endometrioma, and teratoma) were compared by calculating "Precision" values for $N = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140\}$ retrieved results as:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Figure 1 demonstrates the precision curves drawn by calculating the average precision values from the retrieved images of each category. As can be observed from this graph, the best overall retrieval performance has been achieved in retrieving the ultrasound images of simple cyst. The average precision value lies above 77% for the first 20 retrieved images and above 75% for the first 40 retrieved images, which indicates very satisfactory and consistent retrieval performance.

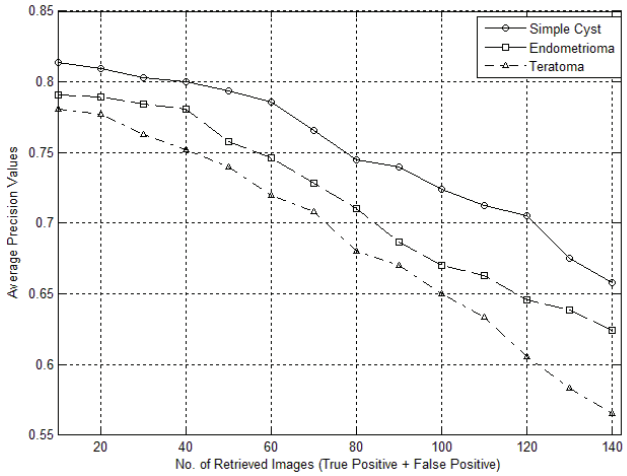

Figure 1: Performance of the proposed method in retrieving ultrasound images of ovarian cysts.

Table 2: Comparison of classification accuracy (%) achieved using different classifiers in classifying ultrasound ovarian images.

	Simple Cyst	Endometrioma	Teratoma	Average
SVM-RBF	89.84	87.66	83.21	86.90
SVM-Polynomial	86.10	82.47	80.29	82.95
SVM-Sigmoid	87.70	85.71	81.02	84.81
Neural Network	82.89	79.87	75.91	79.56
k-NN	86.63	81.17	79.56	82.45

4.2. Classification Performance

Due to its capability of providing high classification accuracy over small training sets as well as generalization performance on data that is highly variable and difficult to separate, SVM has been chosen in this work for classifying ultrasound images into 3-categories: simple cyst, endometrioma, and teratoma. Features extracted from 200 images of the databases have been used to train the classifier applying "K-Fold Cross Validation" technique with $K = 5$. The choice of kernel function is among the most important customizations that can be made while adjusting an SVM classifier to a particular application domain. By performing experiments with SVM using a range of Polynomial, Gaussian Radial Basis Function (RBF) and Sigmoid Kernels, we have found that RBF kernel significantly outperforms the others, boosting the overall recognition accuracy. We have also compared the performance of the proposed ultrasound image classification method with two other popular classification techniques namely, k-Nearest Neighbor (k-NN) and Neural Network (NN). Results of these comparisons have been summarized in Table 2.

5. CONCLUSION

In this paper, we presented a solution for retrieving and classifying ultrasound images of three types of ovarian cysts. In addition, performance of combining two different types of image feature has been evaluated in content-based retrieval and classification of ultrasound images. As the experimental results show, the proposed solution demonstrates significant prospect in support of using histogram moments and GLCM based texture feature together for retrieving and classifying ultrasound images. Our future plan is to investigate the classification and retrieval performance of this system over ultrasound images of other types of ovarian cysts. The proposed methods can be adopted for developing a Computer-Aided Diagnosis (CAD) system, which can serve as a tool to provide decision support in the diagnosis of ovarian abnormalities. By querying the CAD system with a new image and consulting the retrieved results along with their proven pathological diagnosis, the physician would gain more confidence in his/her decision or even sometimes see the scope of considering other possibilities towards improving the overall diagnosis accuracy.

6. REFERENCES

[1] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A Review of Content-Based Image Retrieval Systems in Medical Applications-Clinical Benefits and Future Directions," *Intl. Journal of Medical Informatics,* vol. 73, no. 1, pp. 1-23, Feb. 2004.

[2] T.M. Lehman, M.O. Güld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, B.B. Wein, "Content-Based Image Retrieval in Medical Applications," *Methods of Information in Medicine*, vol. 43, no. 4, pp. 354-61, 2004.

[3] J.R. Van Nagell, P.D. Depriest, E.S. Donaldson, H.H. Gallion, E.J. Pavlik, and R.J. Kryscio, "Ovarian Cancer Screening in Asymptomatic Postmenopausal Women by Transvaginal Sonography," *Cancer,* vol. 68, no. 3, pp. 458-462, 2006.

[4] M. Stricker, and M. Orengo, "Similarity of Color Images," in proc. *SPIE: Storage and Retrieval for Image and Video Databases III*, vol. 2420, pp. 381-392, San Jose, Feb. 1995.

[5] M. K. Mandal, T. Aboulnsar, and S. Panchanathan, "Image Indexing Using Moments and Wavelets," *IEEE Trans. Consumer Electronics*, vol. 41, pp. 557-565, 1996.

[6] R.M. Haralick, K. Shanmugan, and I.H. Dinstein, "Textural Features for Image Classification," *IEEE Trans. Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, May 1973.

[7] J.C. Gower, "A General Coefficient of Similarity and Some of Its Properties," *Biometrics*, vol. 27, no. 4, pp. 857-871, 1971.

[8] N. Abbadeni, "Content Representation and Similarity Matching for Texture-Based Image Retrieval," in proc. *5th ACM SIGMM international Workshop on Multimedia information Retrieval (MIR'03),* California, Nov. 2007, 2003, pp. 63-70.

[9] V.N. Vapnik, The Nature of Statistical Learning Theory, 2nd edn, Springer-Verlag, New York, 2000.

[10] J. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin dags for multiclass classification," *Advances in Neural Information Processing Systems,* vol. 12, pp. 547–553, 2000.