Analysis of Sonogram Images of Thyroid Gland Based on Wavelet Transform

M. Bastanfard, B. Jalaeian, and S. Jafari

Abstract—Sonogram images of normal and lymphocyte thyroid tissues have considerable overlap which makes it difficult to interpret and distinguish. Classification from sonogram images of thyroid gland is tackled in semiautomatic way. While making manual diagnosis from images, some relevant information need not to be recognized by human visual system. Quantitative image analysis could be helpful to manual diagnostic process so far done by physician. Two classes are considered: normal tissue and chronic lymphocyte thyroid (Hashimoto's Thyroid). Data structure is analyzed using K-nearest-neighbors classification.

This paper is mentioned that unlike the wavelet sub bands' energy, histograms and Haralick features are not appropriate to distinguish between normal tissue and Hashimoto's thyroid.

Keywords—Sonogram, thyroid, Haralick feature, wavelet.

I. INTRODUCTION

PATHOLOGY of the thyroid gland has followed generations of people since the earliest period of human existence. It is already mentioned by ancient physicians from China, India, and Egypt several thousand years before Christ. Recently, people are stricken with thyroid gland diseases more frequently than with diabetes (about 900 million people suffer only from diseases caused by lack of iodine). The first important step on the way towards health is a successful diagnosis.

A large number of disease processes influence human body tissue in such a manner as to produce abnormalities. These abnormalities are detectable in images produced by sonography or another imaging technique. Sonography is simple non-invasive diagnostic method and it is one of the most applied imaging techniques. Physician can observe the state of human organs at any time it is necessary. Diagnosis is based on physician's knowledge and experience. The character of sonographic images is textural and one can qualitatively characterize texture as having such properties as fineness, coarseness, smoothness, granulation, randomness, lineation or as being mottled or irregular. In case of thyroid gland diseases this qualitative approach is often combined with invasive needle biopsy. It often causes heavy burden and stress for patients. This motivates us to look for a method that would extract quantitative description in an automatic way.

M. Bastanfard, B. Jalaeian, and S. Jafari are with Computer Science and Engineering department of Shiraz University, Shiraz, Iran (fax: +98-711-6271747; e-mail: bastanfard@cse.shirazu.ac.ir, jafaris@shirazu.ac.ir, jalaeian@cse.shirazu.ac.ir).

Such method would be able to replace the invasive procedures annoying for patients. Moreover, description obtained by computer need not be observed by human vision and hence can give more information than just visual inspection.

Hashimoto's disease is the state when function of the thyroid tissue is initially unchanged, but it can not make enough thyroid hormone after certain time. This can result in hypothyroidism. It is named after the Japanese doctor who first described it and it is also called Hashimoto's thyroid, chronic autoimmune thyroid, or lymphocyte thyroid (LT). The disease is five times more common in women than in men. Hypothyroidism caused by Hashimoto's disease results in an overall slowing down of body's functions. A heart may beat more slowly, body temperature may decrease, muscles may weaken, cholesterol level may rise, and one may have difficulty thinking and remembering. In time, this overall slowing down affects most of body's functions, and can seriously affect health. Therefore, it is very important to identify hypothyroidism as early as possible and treat it properly. Human immune system mistakenly identifies stricken thyroid gland as a group of foreign cells and produces antibodies against the thyroid cells. The presence of thyroid antibodies in the blood is in most cases indication of Hashimoto's disease. It can be also recognized by physician from sonogram images. Hypothyroidism is treated with thyroid hormone replacement therapy. The thyroid gland is small and can be seen in its entirety by some transducers, but it is still evaluated by viewing the lobes individually.

In previous woks Haralick features entropy, correlation and homogenous were chosen in [3],[6] as the best subset of 9 Haralick features. This feature set can separate thyroid tissues' sonogram by rating about 72% to 86%. Besides, systems which are based on Histogram [2], where the highest decides of histogram is selected as feature set, denotes recognition rate about 84% to 92%.

II. FEATURE SELECTION

Image of size $n \times n$ consists of n2 pixels and can be represented as a point in the n2-dimensional space. Our aim is to reduce the dimension of this space to the least possible one and make diagnosis in this feature space. Such dimension should fulfill the following requirements: (i) information lost during reduction should be as small as possible, (ii) resulting space should provide sufficient information for certain purpose.

A. Linear Discrimination Analysis

Texture is a complicated entity to measure. The reason is that any parameters (features) are likely to be required to characterize it. In addition, when so many features are involved, it is difficult to decide the ones that are most relevant for recognition. Fisher linear discrimination is a method that provides a measure of information about classes represented by features. The principle of the method can be shown in a feature space. There are clusters of points belonging to different classes. Fisher linear discrimination [8] assumes that each cluster can be represented by its mean value and variance (covariance matrix for more than 1-dimensional feature space). The smaller variance inside clusters and higher distances between them, the more appropriate the features are. Therefore, good features are those for which

$$\frac{\text{inter - classes variance}}{\text{intra - classes variance}}$$

is higher than for others. Since the variance inside the individual classes can be different we can use following:

$$\frac{\text{variance between classes}}{\text{higher intra - class variance}}$$

Variance inside classes was not computed as covariance of mean values [8], but as covariance of all feature points, which is more precise in our case. Generally, suppose we have data in n classes. Each class is represented by matrix Xi where columns are feature vectors. Suppose the following:

$$A = \text{cov}([X_1 X_2 X_3 ... X_n]) \qquad (1)$$

$$B_i = \text{cov}(X_i) \qquad (2)$$

$$\lambda_i = \max(eig(B_i^{-1} A)) \qquad (3)$$

For deeper insight into values that can be assumed by Fisher linear discrimination, we will do the following. Suppose

$$p = K_i / \sum_{j=1}^{n} K_j$$

where K_i is a number of vectors in matrix X_i, $\sum_{j=1}^{n} K_j$ is a number of vectors in $[X_1 X_2 X_3 ...X_n]$.
Values of parameter p can be from interval < 0, 1 >.
It was found experimentally that:
1. In case of perfect overlap of classes (they have equal mean values and variance of one class is zero), then F= p.
2. For identical classes (equal mean values and variances (covariance matrices)), then F = 1.
3. If classes are perfectly separable, then F = 1.
4. In case of only partially overlap, then p< F <1.
5. If classes are merely separable, then -1< F <1.The higher F, the better separability.

B. Correlation Coefficient

Another way to compare features is multi-correlation coefficient R(i).

$$R(i) = \text{cov}(x_i, y) / \sqrt{\text{var}(x_i) \text{var}(y)} \qquad (4)$$

Where cov(x,y) is the covariance matrix of two dimensions x , y and var shows the variance.

The coefficient R(i) is also the cosine between vectors xi and y, after they have been centered (their mean subtracted). Although the R(i) is derived from R(i) it may be used without assuming that the input values are realizations of a random variable.

In linear regression, the coefficient of determination, which is the square of R(i), represents the fraction of the total variance around the mean value that is explained by the linear relation between xi and y. Therefore, using R(i)^2 as a variable ranking criterion enforces a ranking according to goodness of linear fit of individual variables. The use of R(i)^2 can be extended to the case of two-class classification, for which each class label is mapped to a given value of y.

Correlation criteria such as R(i) can only detect linear dependencies between variable and target. A simple way of lifting this restriction is to make a non-linear fit of the target with single variables and rank according to the goodness of fit. Because of the risk of over fitting, one can alternatively consider using non-linear preprocessing (e.g., squaring, taking the square root, the log, the inverse, etc.) and then using a simple correlation coefficient.

III. CLASSIFICATION

K-nearest-neighbors classifier was implemented as classifier that makes no assumption about data distribution. Subject is classified to be of class normal (N) if majority of its images is of class normal, and is classified to be of lymphocyte (LT) class when majority of its samples is classified as lymphocyte.

A. K Nearest Neighbors

K-nearest-neighbors is a non-parametric technique for density estimation and classification. This rule classifies new feature vector X by assigning it the label most frequently represented among the K nearest samples. To explain the principle of classification based on this method sufficiently we use Bayes theorem [1]. It requires computing posterior probabilities from class-conditional densities and a priori probabilities for each class. Suppose our data set contains N_k points in class Ck and N points in total, so that $P_k N_k = N$. We then draw a hyper sphere (the cell mentioned above) around the point X which encompasses K points irrespective of their class label.

Suppose this sphere, of volume V , contains K_k points from class C_k. Then approximations for the class-conditional densities can be given in the form

$$P(X | C_k) = K_k / N_k V \qquad (5)$$

The unconditional density can be similarly estimated from

$$P(X) = K / NV \qquad (6)$$

While the priori probabilities can be estimated using (7).

$$P(C_k) = N_k / N \qquad (7)$$

We now use Bayes theorem to give

$$P(C_m \mid X) = P(X \mid C_m)P(C_m)/P(X) = K_k/K \qquad (8)$$

To minimize the probability of misclassifying a new vector X, it should be assigned to the class Ck for which the ratio K_k/K is largest. It means finding a hyper sphere around the point X, which contains K points (independent of their class), and then assigning X to the class having the largest number of representatives inside the hyper sphere. For the special case of K = 1 we simple assign a point X to the same class of which the nearest point from the training set is.

A disadvantage of K-nearest-neighbors method is that the complete set of training samples must be stored and must be searched each time a new feature vector is to be classified. This might result in computational problems while making classification.

B. Leave One Out

Leave-one-out is a method for estimation of an upper bound on the Bayes error. It consists in dividing feature space into design and test set. The test set consists of features from one subject. The design set is created by features from remaining N-1 subjects. Features from the test set are classified by K-nearest neighbors classifier designed on the design set. This is repeated N times to test features from all N subjects. Subject is then classified by majority vote.

After this procedure, leave-one-out error (9) on subjects (LOO subjects) can be computed as number of misclassified subjects divided by the number of all subjects. False negative error for subjects (FN subjects) is number of LT subjects classified as normal number of all LT subjects.

$$\frac{\text{No. LT subjects classified as Normal}}{\text{No. LT subjects}}$$

False positive error for subjects (FP subjects) is number of normal subjects classified as LT number of all normal subjects.

$$\frac{\text{No. Normal subjects classified as LT}}{\text{No. Normal subjects}}$$

Suppose number of normal subjects be mN, number of LT subjects mLT , and number of all subjects M = mN + mLT. Relationship between FN, FP, and LOO error is the following:

$$LOO = \begin{cases} \dfrac{FN+FP}{2} & \text{if mN=mLT=M/2} \\ \dfrac{FN.m_{LT}+FP.m_N}{M} & \text{otherwise} \end{cases} \qquad (9)$$

IV. EXPERIMENTAL RESULTS

A. Materials

Sonogram images are digitized data from ultrasound imaging system Toshiba ECCOCEE (console model SSA-340A, transducer model PLF-805ST at frequency 8MHz). Data are captured in longitudinal cross-sections of both lobes with magnification of 4cm and stored with amplitude resolution of 8 bits (256 grey levels). There are two kinds of images: image with normal tissue and image with lymphocyte thyroid. Real time imaging of the patient's thyroid gland is observed by physician on the monitor of the imaging system while making diagnosis. Since the number of frames observed may play a role in diagnosis, our data consists of approximately 20 images per subject. There are two kinds of images: image with normal tissue and image with lymphocyte thyroid. A set of ultrasonic images which is contained 1405 sonograms and pathologically proven benign lymphocyte tissue from 754 patients and normal tissue from 651 candidates is considered. The patients' ages ranged from 28 to 64 years. Images contain tissue of thyroid gland surrounded by other neighboring tissues. For our purpose we need to have regions with the gland segmented out of the image. Since automatic segmentation is not the aim of this project, the boundary of the thyroid gland is roughly delineated by a physician. For this purpose, simple-to-use interactive tool was implemented. Examples of images with segmented regions of interest can be seen in Fig. 1.
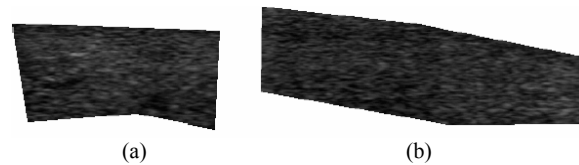

(a)   (b)

Fig. 1 Manually segmented regions of interest.
a) Normal tissue, b) Lymphocyte tissue

B. Feature set

Set of features which is used for classifying the thyroid gland tissues is contained wavelet [4] sub bands' energy, entropy and correlation of Haralick features [5],[7].

Image's histogram is a distribution formed by the simplest features: individual pixels which are used widely as a texture feature in previous works, could be replaced by the energy band of wavelet transform with more performance and separability for classification. The energy of 1st ,2nd,3rd and 4th sub bands in the 3rd level as well as energy of 2nd, 3rd and 4th sub bands in the first and second levels of wavelet transform computed respect to (5) .

$$E = \sum_t W(t)^2 \qquad (10).$$

The wavelet's coefficients have a role as shadowing which is pointed in morphological feature set.

The Haralick features should be calculated according to gray level Co-occurrence matrix [5],[7]. Features which are used here in this investigation are defined as:
▫ Texture Entropy

$$E = -\sum_m C_m \log C_m \qquad (11).$$

▫ Texture correlation

$$Co = \sum_j (i-\mu)(j-\mu)C/\delta^2(i)\delta^2(j) \qquad (12)$$

It is obvious from the fact that it is the principal direction in the sonogram, the direction in which ultrasonic wave propagates through the tissue. Features are derived from co-occurrence matrix for distance vector d=[11,0]. Analogous to (9), FN images and FP images can be computed for proposed feature set. All this process can be repeated over different

number of neighbours (K). FN subjects and FP subjects versus K for proposed feature set is shown in Fig. 2. LOO curve of subjects is given for comparison.

Here we can evaluate separability provided by the previous works' feature set in comparison with our proposed feature set. We computed Fisher linear discrimination for them. This yielded in F_{har}=0.996. For comparison, Fisher linear discrimination was also computed on data represented by histograms and proposed feature set. It resulted in F_{hist}=15.226 and $F_{proposed}$=16.63.
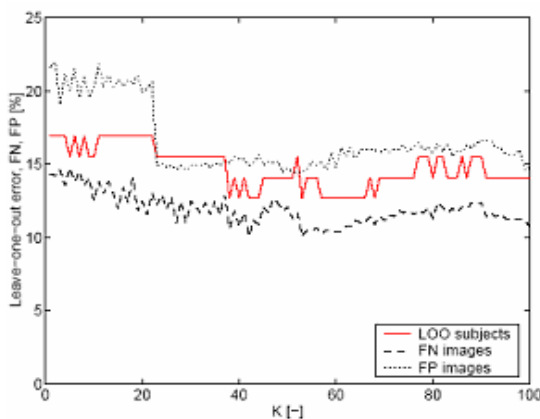

Fig. 2 FN, FP and LOO curve for proposed feature set versus different K

Another way to compare features is multi-correlation coefficient (4). Multi-correlation coefficient can acquire values from interval [0, 1]. The higher R(i), the bigger dependence. For Haralick features, histograms and proposed set R(i)=0.006, R(i)= 0.559 and R(i)= 0.71.respectively.

C. Results

We aimed to classify sonogram images of thyroid gland into two classes: normal tissue and lymphocyte thyroid (LT). For this purpose, images were represented by features that should extract important textural character and information of image.

This paper focuses to use wavelet sub band energy and entropy and correlation of Haralick features on dataset consisting of 1405 images and to use classification that suppose no assumption on data distribution: K-nearest-neighbors rule. Classification was done using leave-one-out methods, which give estimates of upper and lower bounds on the Bayes error. Dependence of these errors over K was analyzed and data structure discussed. Achieved results could be seen in Table I. In addition, the results which are based on proposed feature set, comparison with Haralick features and Histogram are shown in Fig. 3.

V. CONCLUSION

Results for Haralick features were not encouraging: leave-one-out error was up to 45%. Classification of histograms achieved the following results: leave-one out error approached 12%. Whereas the proposed feature set has better performance and separability rate. Experiment shows that the leave one out

error drops up to 7.1% for proposed set. The results were confirmed by parametric methods, Fisher linear discrimination and multi correlation coefficient.

We conclude that Haralick features (texture entropy, texture correlation and texture probability of run length of 2) are not efficient enough for our purpose. They provide feature space in which two clusters of normal and LT tissue are no separable.

TABLE I
COMPARISON OF REC. RATE VIA VARIOUS FEATURE SETS

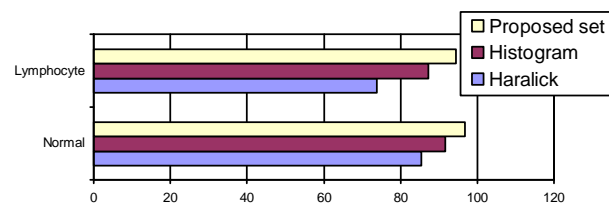| Tissue	HARALICK	Histogram	Proposed Set
Lymphocyte	73.9%	87.3%	94.4%
Normal	85.4%	91.6%	96.8%

Fig. 3 Recognition rate of proposed set, in comparison with previous works

Our results suggest that information needed to distinguish normal from LT tissue can be obtained using wavelet transform. Automatic classification based on wavelet sub bands' energy can improve the diagnosis reliability in distinguishing between normal tissue and Hashimoto's thyroid. Better results could be obtained by using entropy and correlation of Haralick features in addition.

REFERENCES

[1] C. M. Bishop. Neural Networks for Pattern Recognition. Clarendon Press, 1995.
[2] E. A. Toufik. Automatic classification of the thyroid gland diseases by a histogram. Master's thesis, Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic, Jan 2001.
[3] M. Švec and R. Šara. Analýza textury sonografických obrazů difuzních process u parenchymu štítné žlazy. Research Report CTU–CMP–1999–12, Center for Machine Perception, FEE CTU in Prague, Dec 1999.
[4] Mallat, S. G. A Theory for multi resolution signal decomposition: The Wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11, No. 7 1989, pp. 674-693.
[5] R. Šara, D. Smutek, P. Sucharda, and Š. Svačina. Systematic construction of texture features for Hashimoto's lymphocytic thyroiditis recognition from sonographic images. In S. Quaglini, P. Barahona, and S. Andreassen, editors, Artificial Intelligence in Medicine, LNCS, Berlin-Heidelberg, Germany, 2001.
[6] R. Šara, M. Švec, D. Smutek, P. Sucharda, and Š. Svačina. Diffusion process classification in thyroid gland parenchyma based on texture analysis of sonographic images: Preliminary results. In Svoboda T., editor, Proceedings of the Czech Pattern Recognition Workshop 2000. Czech Pattern Recognition Society Praha, Feb 2000, pp 45–47.
[7] R. Šara. Sonograph images: Texture analysis [online]. C 1998, last revision 9th of November 2000. http://cmp.felk.cvut.cz/~sara/Sono/sono.html.
[8] Z. Kotek, I. Brˇuha, V. Chalupa, and J. Jelínek. Adaptivní a učící sesystémy. SNTL, 1980.