INTENSITY-INVARIANT TEXTURE FEATURES FOR BREAST ULTRASOUND CLASSIFICATION

W. Gómez*, W. C. A. Pereira** and A. F. C. Infantosi**

* Information Technology Laboratory/CINVESTAV-IPN, Ciudad Victoria, Mexico
** Biomedical Engineering Program/COPPE-UFRJ, Rio de Janeiro, Brazil
e-mail: wgomez@tamps.cinvestav.mx

Abstract: This paper presents a methodology for improving breast ultrasound (BUS) classification by using intensity-invariant texture features based on log-Gabor filtering and ranklet transform. The Fisher's linear discriminant analysis (FLDA) classifies 641 BUS images in benign and malignant lesions. The results point out that the FLDA performance improves significantly, in terms of area under ROC curve index, from 0.89 to 0.93 with the proposed scheme.
Keywords: Breast ultrasound, texture features, classification, ranklets, log-Gabor filters

Introduction

Sonographic texture analysis (STA) is helpful in distinguishing benign from malignant breast lesions on ultrasound (BUS). In this sense, radiologists perform STA by depicting qualitatively the speckle patterns variations among breast tissues; hence, the diagnosis depends on their expertise and training.

To improve radiologists' decision, computer-aided diagnosis (CAD) systems have emerged as a "second reader" for analyzing BUS images, whereby the lesions are classified by using quantitative morphological and texture features. The former attempts to quantify shape and contour attributes, whereas the latter describes tissue echo patterns by analyzing locally gray-level variations [1]. Several texture attributes have been employed in BUS classification, which involve gray-level co-occurrence matrix (GLCM) descriptors, autocovariance (or autocorrelation) coefficients, complexity curve descriptors, fractal features, gray-level run-length matrix, block-based coefficients, or posterior acoustic behavior [1].

Currently, the majority of the proposed CAD systems assume that BUS images are identical to those used in training samples with respect to gray-scale range. However, in practical applications, it is very unlikely that images are acquired under the same conditions, since BUS is highly dependent on the equipment and the operator skills. As consequence, the texture features depend on the gray-scale distributions of the BUS dataset used for CAD system development.

Recently, great attention has been devoted to intensity-invariant texture classification, where the input BUS image is transformed to a gray-scale invariant representation for making texture descriptors robust to linear/non-linear monotonic gray-scale transformations, such as brightness variations, contrast enhancement, gamma correction, histogram equalization, etc. [2].

In this paper, we propose a BUS classification approach, where both the ranklet transform and the log-Gabor filters are employed for transforming the BUS data to an intensity-invariant domain. Thereafter, texture features are computed from the transformed data, including GLCM descriptors, autocovariance coefficients, and autocorrelation. Moreover, a feature selection procedure is performed to determine the subset of texture features that improves the classification performance in terms of the area under ROC curve (AUC) index.

Materials and Methods

Image dataset – The 641 BUS images in dataset were acquired during routine breast diagnostic procedures at the National Cancer Institute (INCa, Rio de Janeiro, Brazil) with an ultrasonic scanner Sonoline Sienna equipment (Siemens, Germany), using a 7.5-MHz linear-array probe, and captured directly from the 8-bit video signals. All 228 carcinomas and 413 benign lesions were histopathologically proven by biopsy.

It is worth mentioning that texture analysis is performed on a region of interest (ROI), which contains the gray-level values within a bounding box that encloses the lesion.

Intensity-invariant transformations – Every ROI image from BUS dataset is transformed to an intensity-invariant representation by using two techniques: log-Gabor filtering and ranklet transform. In both cases, a multi-channel decomposition is performed on the input image, in which the outcomes represent purely texture data depicted at different orientations-scales/resolutions.

Log-Gabor filters can be constructed with arbitrary bandwidth, minimal aliasing, and practically without DC component, which contributes to improve the contrast between different textures. The frequency response of the oriented log-Gabor function is expressed as [3]

$$G_{\omega,\phi} = \exp\left[-\frac{\log(\omega/\omega_0)}{2\sigma_\omega^2}\right]\exp\left[-\frac{\log(\phi-\phi_0)}{2\sigma_\phi^2}\right], \quad (1)$$

where (ω,ϕ) denotes the polar coordinates; ω_0 is the filter center frequency; ϕ_0 is the filter orientation angle; and σ_ω and σ_ϕ define the frequency and angular bandwidths, respectively. Then, the filtered image by a log-Gabor function in the frequency domain is computed as

$$F'_{u,v} = F_{u,v} \cdot G_{u,v}, \quad (2)$$

where $G_{u,v}$ is the expression of log-Gabor filter in Cartesian coordinates system, and $F_{u,v}$ and $F'_{u,v}$ are the original and filtered spectra, respectively.

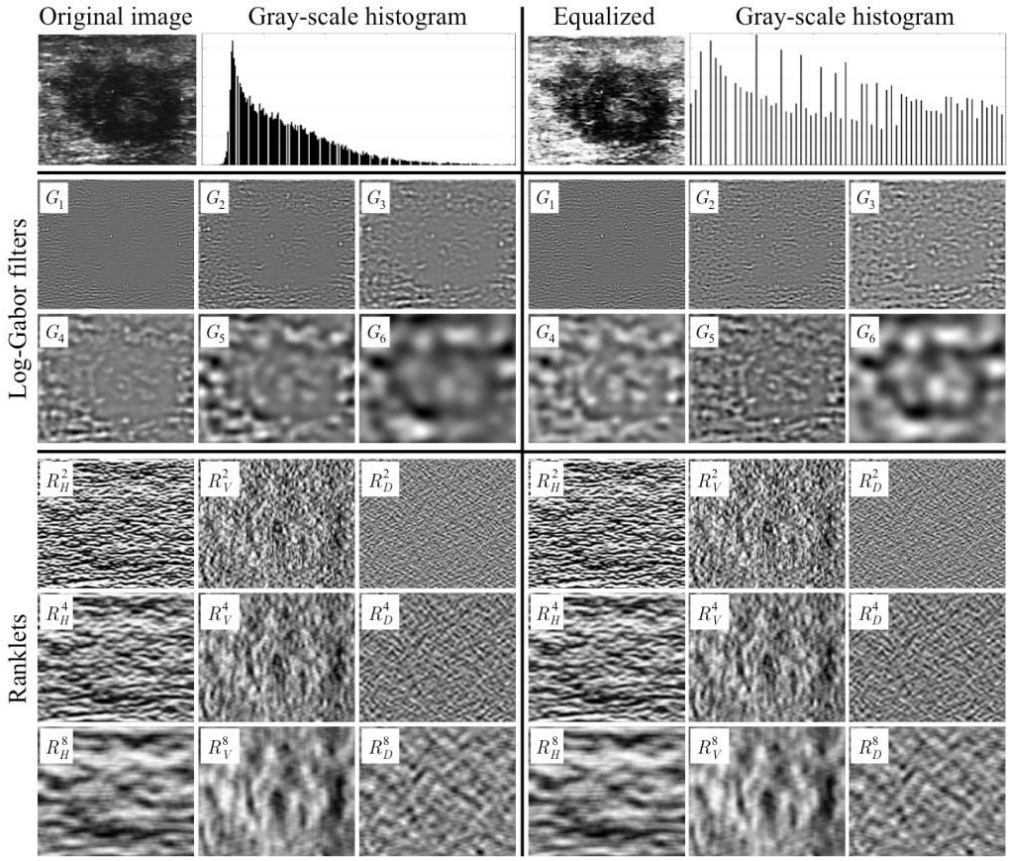

Figure 1: The log-Gabor filters (G_ω, with scales ω=1, 2,…,6) and ranklets (R_t^r, with orientations t={H, V, D} and resolutions r ={2, 4, 8}) are derived from an original BUS image and its enhanced version by histogram equalization.

Thereafter, the inverse Fourier transform of each filtered spectrum is obtained to recover $n_\omega \times n_\phi$ texture images, where n_ω=6 and n_ϕ=24 are the number of scales and orientations, respectively. Both parameters were determined heuristically in a previous study developed by our research group [4].

However, for reducing the feature space dimensionality, the filtered images at the same scale are averaged over all orientations to create six texture images.

The ranklet transform involves three stages: multiresolution, orientation-selective, and nonparametric analysis [2]. It considers the relative rank of the pixels, within a local region, instead of their gray-scale values; hence, it could be defined as an invariant operator to linear/nonlinear monotonic intensity transformations of the original image. The ranklet coefficient is defined as

$$R_t = \frac{\sum_{p\in T_i} \pi(p) - \frac{N}{4}\left(\frac{N}{2}+1\right)}{N^2/8} - 1, \quad t = \{V, H, D\}, \qquad (3)$$

where T_i is the half part of a local region of size N=r×r, with resolutions r ={2, 4, 8} in pixels, for a specific orientation, t, that could be vertical (V), horizontal (H), or diagonal (D); and $\pi(p)$ denotes the pixel ranks in T_i. The ranklet transform yields $n_r \times n_t$ texture images, where n_r=3 and n_t=3 are the number of resolutions and orientations, respectively.

Figure 1 illustrates the intensity-invariant texture images derived from log-Gabor filters and ranklet transform applied on both the original BUS image and its enhanced version by histogram equalization, that is, with a monotonic gray-scale transformation. The transformed texture images are quantized to 64 gray levels [5].

Texture features – They are computed from the transformed ROI image, which are based on GLCM, autocovariance coefficients, and autocorrelation.

The GLCM represents the joint frequencies of all pairwise combinations of gray levels i and j separated by distance d and along direction θ, and it is defined as

$$C(i,j) = \left\| \{(x_1,y_1),(x_2,y_2)\} \middle| x_2 - x_1 = d\cos\theta, \; y_2 - y_1 = d\sin\theta, I(x_1,y_1) = i, I(x_2,y_2) = j \right\|, \qquad (4)$$

where (x_1,y_1) and (x_2,y_2) are pixel locations; $I(\cdot)$ is the gray-level of the pixel; and $\|\cdot\|$ is the number of pixel pairs that satisfy the condition. The GLCM parameters are defined as d ={1, 2, 4, 8} in pixels and θ={0º, 45º, 90º, 135º}; thus, for each analyzed image, 16 GLCMs are computed. Besides, in order to reduce the dimensionality of the feature space, the GLCMs of a same distance are averaged over all orientations to build four matrices. Finally, six common texture features are extracted from every GLCM: contrast, correlation, entropy, sum average, sum entropy, and homogeneity [5].

The normalized autocovariance coefficients depict the inner-intensity variance within the ROI, which is computed as [6]

$$\tilde{S}_{\Delta m,\Delta n} = S_{\Delta m,\Delta n} / S_{0,0}, \qquad (5)$$

where the autocovariance matrix is

$$S_{\Delta m,\Delta n} = \frac{\sum_{x,y}\left[I(x,y)-\bar{I}\right]\left[I(x+\Delta m, y+\Delta n)-\bar{I}\right]}{(M-\Delta m)(N-\Delta n)}, \qquad (6)$$

where $\bar{I}$ is the mean value of $I(x,y)$ and M and N are the image width and height, respectively. In this work, $\Delta m=\Delta n=5$; hence, for each image a 5×5 autocovariance matrix is produced. However, the first coefficient (0,0) is removed from the feature space because its value is always the unity.

The autocorrelation of image $I(x,y)$ is defined as [7]

$$\tilde{R} = \sum_{n=0}^{N-1} \overline{R}(n) / \overline{R}(0), \quad (7)$$

where the autocorrelation in depth and its sum in the lateral direction are

$$R(m,n) = \sum_{k=0}^{N-1-n} I(m,n+k)^2 I(m,k)^2, \quad (8)$$

$$R(n) = \sum_{m=0}^{M-1} R(m,n). \quad (9)$$

Note that the feature space dimensionality, Λ, is defined by the number of texture images, obtained after image transformation, times the number of texture features. With respect to log-Gabor filters Λ=294, whereas for ranklet transform Λ=441.

Feature space ranking – We are interested in selecting a subset of texture features by removing irrelevant and redundant ones while maintaining acceptable classification accuracy. This goal could be achieved by ranking the Λ-dimensional feature set according to the minimal-redundancy-maximal-relevance (mrMR) criterion followed by the selection of the first λ features (such that $\lambda<\Lambda$) with the best classification performance.

The mrMR criterion measures the mutual information (MI) among variables, where the minimal redundancy condition selects the features such that they are mutually exclusive, whereas the maximal relevance condition quantifies the level of dependency between an individual feature and the target class (i.e., benign or malignant) [5].

Classifying performance – Once the whole Λ-dimensional feature set is ranked according to mrMR criterion, distinct λ-dimensional feature subsets are created by adding iteratively the top λ features to the classification procedure until all of them are considered, that is, λ=1, 2,…, Λ.

The Fisher's linear discriminant analysis (FLDA) is frequently used for classifying BUS images because it is parameter-free and easy to learn [1]. FLDA creates a classification rule, by using training data, which yields the largest mean differences between the two desired classes: benign and malignant.

Let μ_1 and μ_2 be the mean vectors of Class 1 (benign) and Class 2 (malignant), respectively, and **S** denotes the pooled covariance matrix; then, the FLDA output for some data **x** in test set is defined as [8]:

$$D(\mathbf{x}) = [\mathbf{x} - \tfrac{1}{2}(\mu_1 + \mu_2)]^T \mathbf{S}^{-1}(\mu_1 - \mu_2). \quad (10)$$

Discrepancy between the FLDA predicted value and the actual class could be used to assess the discrimination power of texture features [8].

Then, for every λ-dimensional feature subset, 50 bootstrap replications are performed to build the training sets, whereas the test sets considered patterns non-included into training data.

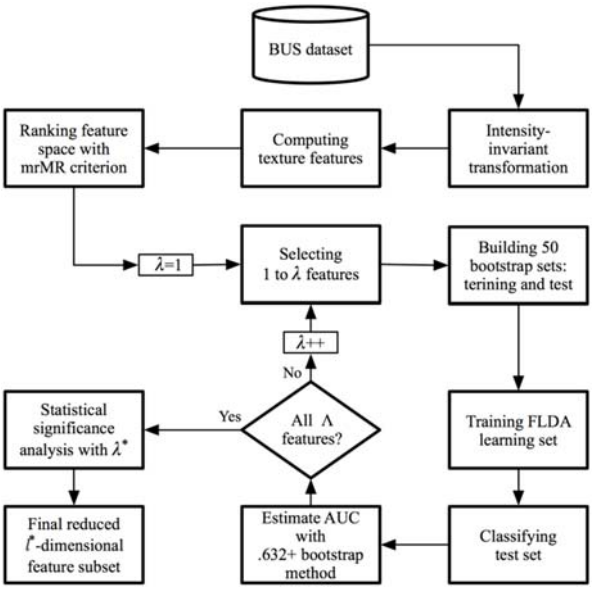

Figure 2: Flow chart of classification performance procedure for distinct texture feature subsets.

The .632+ bootstrap estimator uses the classified test sets by FLDA to determine the discrimination power of each λ-dimensional subset in terms of the AUC value, which is a quantitative index of the overall performance of a classification system. It is usually provided within the range [0, 1], where unity stands for perfect discrimination.

Finally, the λ-dimensional feature subset with the highest $AUC_{.632+}$ value (i.e., the best performance), denoted as λ^*, is selected as "reference subset". Next, a statistical analysis attempts to find the smallest l^*-dimensional feature subset, such that $l^*<\lambda^*$, that performs statistically similar to the reference subset given a 95% confidence interval. To perform such statistical analysis, the one-way analysis of variance (ANOVA1, α=0.05) test is employed and the correction for multiple testing on the basis of the same data is made by the Tukey-Kramer method [8].

Figure 2 illustrates the schematic block diagram of classification performance procedure considering distinct λ-dimensional texture feature subsets.

Results

Four distinct texture feature sets were evaluated: without intensity-invariant transformation (i.e., original dataset) (FS_1), log-Gabor filtering (FS_2), ranklets (FS_3), and combining log-Gabor and ranklets features (FS_4).

Figure 3 summarizes the classification performance, in terms of $AUC_{.632+}$ value, attained by each texture feature space and distinct dimensionalities. One can note that the best performance was reached by λ^*-FS_4 (pointed with symbol '▼'). Besides, the ANOVA1 test, with Tukey-Kramer correction, was used to perform pairwise comparisons between the best feature space and the remaining groups.

It is noticeable that the reduced feature subsets, with intensity-invariant transformation, performed statistically similar than λ^*-FS_4, that is: l^*-FS_2, λ^*-FS_2, l^*-FS_3, λ^*-FS_3, and l^*-FS_4 (marked with symbol '♦').

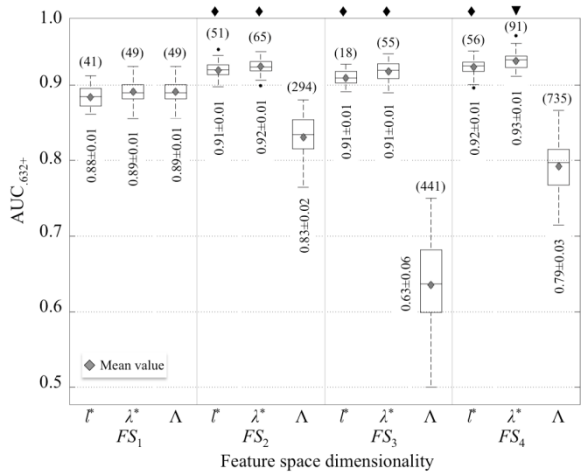

Figure 3: Classification performance results of sets FS_1, FS_2, FS_3, and FS_4 for the dimensionalities l^*, λ^*, and Λ classified by FLDA. The symbol '♦' indicates not statistically difference between groups related to the best one, marked with symbol '▼'. It is included the performance mean±standard deviation values and, in parenthesis, the dimensionality of each feature set.

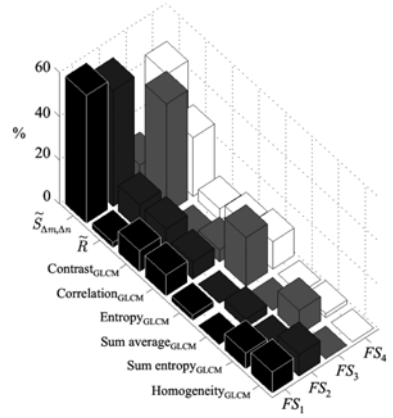

Figure 4: Percentage of texture characteristics regarding l^*-dimensional subsets for each evaluated feature space.

Moreover, these feature subsets outperformed all instances in FS_1. Also, the feature selection procedure improves significantly the classification performance of FS_2, FS_3, and FS_4, since irrelevant and redundant features were removed.

Finally, regarding the l^*-dimensional feature subsets, in Figure 4, it is noticeable that autocovariance coefficients and autocorrelation are the features with the largest discrimination power when applying intensity-invariant transformations.

Discussion and conclusions

Commonly, texture features are computed directly from original gray-scale BUS images for classification purposes. However, it has been demonstrated that texture attributes are sensitive to gray-level distributions, for instance, given two images with similar textures but different brightnesses the texture quantification may be different. Therefore, recently, Min-Chun *et al.* [2] employed gray-scale invariant texture features based on ranklet transform and 144 GLCM characteristics for classifying 470 BUS images (315 benign tumors and 155 carcinomas). Despite the authors reported improvement in classification rates (from 0.83 to 0.91 AUC values), they did not perform a feature selection procedure for determining a texture subset with the highest classification performance.

On the other hand, we explored ranklets as well as log-Gabor filters for depicting intensity-invariant texture features. Also, common texture features widely used in BUS classification (GLCM features, autocovariance and autocorrelation coefficients) were implemented. Besides, a feature selection procedure, based on mrMR criterion, was applied to remove irrelevant and redundant features. We found that combining ranklets and log-Gabor features, the classification performance is about $AUC_{.632+}$=0.93, with 91 texture characteristics. Additionally, the most discriminant features are based on autocovariance coefficients and autocorrelation.

Besides, these results point out that the method presented by Min-Chun *et al*. [2] was outperformed, since the authors obtained lower performance (i.e., $AUC_{.632+}$=0.91) with more characteristics (i.e., 144 texture features) derived from a single texture technique (i.e., GLCM with ranklets). Consequently, our proposed approach should require less computational effort for classifying breast lesions accurately.

Therefore, it is convenient using intensity-invariant texture features in combination with a feature selection procedure and different features computed from distinct texture descriptors for increasing the performance of CAD systems for BUS images.

References

[1] H. D. Cheng, *et al*., "Automated breast cancer detection and classification using ultrasound images: A survey," Pattern Recogn, vol. 43, pp. 299-317, 2010.

[2] Y. Min-Chun, *et al*., "Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis," IEEE T Med Imaging, vol. 32, pp. 2262-2273, 2013.

[3] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J Opt Soc Am A, vol. 4, pp. 2379-2394, 1987.

[4] W. Gómez, *et al*., "Breast ultrasound despeckling using anisotropic diffusion guided by texture descriptors," Ultrasound Med Biol, *in press*, 2014.

[5] W. Gómez, *et al*., "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," IEEE T Med Imaging, vol. 31, pp. 1889-1899, 2012.

[6] R. F. Chang, *et al*., "Improvement in breast tumor discrimination by SVM and speckle-emphasis texture analysis," Ultrasound Med Biol, vol. 29, pp. 679-686, 2003.

[7] K. Horsch, *et al*., "Computerized diagnosis of breast lesions on ultrasound," Med Phys, vol. 29, pp. 157-164, 2002.

[8] W. Gómez, *et al*., "Improving classification performance of breast lesions on ultrasonography," Pattern Recogn, *in press,* DOI: 10.1016/j.patcog.2014.06.006, 2014.