Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

A New Methodology Based on q-Entropy for Breast Lesion Classification in 3-D Ultrasound Images

Paulo S. Rodrigues (PhD), Gilson A. Giraldi (PhD), Marcia Provenzano (MD), Marcelo D. Faria (PhD), Ruey-Feng Chang (PhD) and Jasjit S. Suri (PhD)

Abstract— **Classification of breast lesions is clinically most relevant for breast radiologists and pathologists for early breast cancer detection. This task is not easy due to poor ultrasound resolution and large amount of patient data size. This paper proposes a five step novel and automatic methodology for breast lesion classification in 3-D ultrasound images. The first three steps yield an accurate segmentation of the breast lesions based on the combination of (a) novel non-extensive entropy, (b) morphologic cleaning and (c) accurate region and boundary extraction in level set framework. Segmented lesions then undergo five feature extractions consisting of: area, circularity, protuberance, homogeneity, and acoustic shadow. These breast lesion features are then input to a Support Vector Machine (SVM)-based classifier that classifies the breast lesions between malignant and benign types. SVM utilizes B-spline as a kernel in its framework. Using a data base of 250 breast ultrasound images (100 benign and 150 malignant) and utilizing the cross-validation protocol, we demonstrate system's accuracy, sensitivity, specificity, positive predictive value and negative predictive value as: 95%, 97%, 94%, 92% and 98% respectively in terms of ROC curves and A_z areas, better in performance than the current literature offers.**

I. Introduction

As per American Cancer Society [1], breast cancer ranks second in the list of women's cancer. Even though the rate of breast cancer has risen since 1980, the mortality rates have declined by 2.3% since 1990. The reduction in mortality rate is due to early detection and improvement in technology for treatment.

With the advancement of ultrasound acquisition (see [2]) and the digital nature of 3-D volume acquisition, computer-aided diagnosis (CAD) has become even more powerful, reducing the physician's work load (see [3] and references there in). The process of breast lesion classification and efficient CAD-based methods from utlrasound is generally categorized into several techniques such as those in [4], [5], [6] and [7]. On these lines, in [8] was proposed a CAD system using fuzzy inference for breast sonography and adopted six different criteria to classify lesions such as: lesion shape, border, edge shadows, internal echoes, posterior echoes, and halo. However, their system accuracy, sensitivity, and specificity were only 60.3%, 82.1% and 42.9%. On the other hand, texture framework in US image processing goes back to more than two decades [9], [10]. Since the implementations were presented on a smaller set of data, there is a fundamental weakness with texture-based strategies. The settings of ultrasound machine parameters have to be fixed for acquiring ultrasound images. On the contrary, if ultrasound parameter setting changed, the CAD performance was very unstable [6]). Moreover, a CAD system trained by images from one ultrasound machine needs to be trained again for a different ultrasound machine due to different image resolution and image quality. Hence, in [6] is proposed nearly setting-independent features based on shape information. Their system was very robust and powerful because the statistical data using ROC curve were all greater than 0.95. Note that all above strategies where used on 2-D ultrasound images, which has its own limitations such as, it did not allow 3-D spatial processing during the ultrasound data acquisition. As a result, the shape and structure information of the breast lesions could not be reconstructed. This helps in the determining the growth of the cancer and is spatial relationships. Recently, 3-D ultrasound [11] has shown promising signs that overcome the limitations of traditional 2-D ultrasound allowing to view the anatomy in 3-D interactively.

This paper proposes a five step novel and automatic methodology for breast lesion classification in 3-D ultrasound images. The first three steps yield an accurate segmentation of the breast lesions based on the combination of (a) novel non-extensive entropy, (b) morphologic cleaning and (c) accurate region and boundary extraction in level set framework. For the first step, we propose a new algorithm based on a new kind of entropy, called non-extensive entropy, which extends the well known concept of Boltzman-Gibbs entropy from statistical mechanics. Our algorithm, named NESRA (Non-Extensive Segmentation Recursive Algorithm), is a recursive version of the non-recursive procedure proposed by Albuquerque [12]. Segmented lesions then undergo five feature extractions consisting of: area, circularity, protuberance, homogeneity, and acoustic shadow. These features are then input to a Support Vector Machine (SVM) framework that classifies the lesions between malignant and benign types. Using a data base of 250 breast ultrasound images and utilizing the cross-validation protocol, we demonstrate system's accuracy, sensitivity, specificity, positive predictive value and negative predictive value as: 95%, 97%, 94%, 92% and 98% respectively in terms of ROC curves and A_z areas, better in performance than the current literature offers.

The layout of the paper is as follows: Section II presents a theoretical background of this work; Section III present our proposed protocol for breast US classification and Section IV presents practical results; and Section V highlights some conclusions.

1-4244-0033-3/06/$20.00 ©2006 IEEE.

II. Theoretical Background

A. Tsallis Entropy

The entropy is an idea born under the thermodynamics concept, but Shannon redefined the concept of entropy of Boltzmann/Gibbs (now called BGS entropy) for the context of system information. Generically speaking, systems that have statistics of the type BGS are called extensive systems. Such systems have an additive property, defined as follows. Let P and Q be two random variables, with probability densities functions $P = (p_1, \ldots, p_n)$ and $Q = (q_1, \ldots, q_m)$. If S is the BGS entropy associated with P or Q (and if P and Q are independent), the entropy of the composed distribution verify the so called additivity rule: $S(P*Q) = S(P)+S(Q)$, where the composed distribution (also called direct product) between P and Q is defined as $P * Q = \{p_i q_j\}_{i,j}$, with $1 \leq i \leq n$ and $1 \leq j \leq m$.

This traditional form of entropy is well known and for years has achieved relative success to explain several phenomenon *if* the effective microscopic interactions are *short*-ranged (i.e., close spatial connections) *and* the effective spatial microscopic memory is *short*-ranged (i.e., *close* time connections) *and* the boundary conditions are *non(multi)fractal*. Roughly speaking, the standard formalism are applicable whenever (and probably only whenever) the relevant space-time is non(multi)fractal. If this is not the case, some kind of extension appears to became necessary. Then, recent developments, based on the concept of non-extensive entropy, also called Tsallis entropy, have generated a new interest in the study of Shannon entropy for Information Theory [12], [13]. Tsallis entropy (or q-entropy) is a new proposal for the generalization of Boltzmann/Gibbs traditional entropy applied to nonextensive physical systems.

This non-extensive characteristic has been applied through the inclusion of a parameter q, which generates several mathematical properties and has the following general equation:

$$S_q(p_1, \ldots p_k) = \frac{1 - \sum_{i=1}^{k} (p_i)^q}{q - 1} \tag{1}$$

where k is the total number of possibilities ($0 \leq p_i \leq 1$) of the system and the real number q is the entropic index that characterizes the degree of non-extensiveness. These characteristics give to q-entropy flexibility in explanation of several physical systems that can not be explained by the traditional BGS entropy.

The main characteristic of Tsallis entropy, which is useful in our work, is the so called q-additive property, given by:

$$S_q(P + Q) = S_q(P) + S_q(Q) + (1 - q) S_q(P) S_q(Q) \tag{2}$$

In this equation, the term $(1 - q)$ stands for the degree of non-extensiveness.

In Albuquerque *et al.* [12] is proposed an algorithm using the concept of q-entropy to segment US images. Since this concept may be naturally applied over any statistical distribution, in this paper we propose a natural extension of the algorithm proposed in [12] which yields to a recursive procedure when, for each distribution P and Q we applied again the concept of q-entropy. We named our extended algorithm as NESRA (Non-Extensive Segmentation Recursive Algorithm) and also propose to apply it for an initial segmentation of our five step protocol.

B. The Non-Extensive Segmentation Recursive Algorithm (NESRA)

We can apply the concept of q-entropy separately over the two distributions P and Q. Them we normalize them as $P : \frac{p_1}{p_A}, \frac{p_2}{p_A}, \ldots, \frac{p_t}{p_A}$ and $Q : \frac{q_{t+1}}{q_B}, \frac{q_{t+2}}{q_B}, \ldots, \frac{q_k}{q_B}$ where $p_A = \sum_{i=1}^{t} p_i$ and $q_B = \sum_{i=t+1}^{k} q_i$.

Them, we can apply the q-entropy to each new normalized distribution and compute its q-additivity as follows:

$$\begin{aligned} S_{P+Q}(t) = \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_A})^q}{q - 1} + \frac{1 - \sum_{i=t+1}^{k} (\frac{q_i}{q_B})^q}{q - 1} + \\ (1 - q) . \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_A})^q}{q - 1} . \frac{1 - \sum_{i=t+1}^{k} (\frac{q_i}{q_B})^q}{q - 1} \end{aligned} \tag{3}$$

In order to find the optimal partition ($t = topt$) which maximizes the S_{P+Q} additive sum, we can achieve with cheap computational effort the t which maximizes

$$t_{opt} = argmax[S_P(t) + S_Q(t) + (1 - q) . S_P(t) . S_Q(t)] \tag{4}$$

The NESRA algorithm then computes recursively the q-additive rule separately in each partition P and Q. Now, after the appropriate normalization we have P split into two new distributions P_1 and P_2. Analogously, Q is split into Q_1 and Q_2. If P_{A1} and P_{A2} is the normalized sum of the P_1 and P_2 distribution respectively, and Q_{B1} and Q_{B2} are the same for Q_1 and Q_2, the new additive rule can be computed as:

$$\begin{aligned} S_q(P + Q) = (\frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_{A1}})^q}{q - 1} + \frac{1 - \sum_{i=t+1}^{\varrho} (\frac{p_i}{p_{A2}})^q}{q - 1} + \\ (1 - q) \cdot \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_{A1}})^q}{q - 1} \cdot \frac{1 - \sum_{i=t+1}^{\varrho} (\frac{p_i}{p_{A2}})^q}{q - 1}) + \\ (\frac{1 - \sum_{i=\varrho+1}^{\upsilon} (\frac{q_i}{q_{B1}})^q}{q - 1} + \frac{1 - \sum_{i=\upsilon+1}^{k} (\frac{q_i}{q_{B2}})^q}{q - 1} + \\ (1 - q) \cdot \frac{1 - \sum_{i=\varrho+1}^{\upsilon} (\frac{p_i}{q_{B1}})^q}{q - 1} \cdot \frac{1 - \sum_{i=\upsilon+1}^{k} (\frac{q_i}{q_{B2}})^q}{q - 1}) + \\ (1 - q) \cdot (\frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_{A1}})^q}{q - 1} + \frac{1 - \sum_{i=t+1}^{\varrho} (\frac{p_i}{p_{A2}})^q}{q - 1} + \\ (1 - q) \cdot \frac{1 - \sum_{i=1}^{t} (\frac{p_i}{p_{A1}})^q}{q - 1} \cdot \frac{1 - \sum_{i=t+1}^{\varrho} (\frac{p_i}{p_{A2}})^q}{q - 1}) \cdot \\ (\frac{1 - \sum_{i=\varrho+1}^{\upsilon} (\frac{q_i}{q_{B1}})^q}{q - 1} + \frac{1 - \sum_{i=\upsilon+1}^{k} (\frac{q_i}{q_{B2}})^q}{q - 1} + \\ (1 - q) \cdot \frac{1 - \sum_{i=\varrho+1}^{\upsilon} (\frac{q_i}{q_{B1}})^q}{q - 1} \cdot \frac{1 - \sum_{i=\upsilon+1}^{k} (\frac{q_i}{q_{B2}})^q}{q - 1}) \end{aligned} \tag{5}$$

where t, ϱ and υ are the new optimal partition for the first partition and second and third recursive partitions. This is

accomplished computing $t = t_{opt}$ which maximizes

$$\begin{aligned} t_{opt} = argmax[&(S_{P_1} + S_{P_2} + (1-q)\cdot S_{P_1}\cdot S_{P_2}) \\ &\cdot(S_{Q_1} + S_{Q_2} + (1-q)\cdot S_{Q_1}\cdot S_{Q_2})\cdot(1-q) \\ &\cdot(S_{P_1} + S_{P_2} + (1-q)\cdot S_{P_1}\cdot S_{P_2}) \\ &\cdot(S_{Q_1} + S_{Q_2} + (1-q)\cdot S_{Q_1}\cdot S_{Q_2})] \end{aligned} \quad (6)$$

The Equation (5) is simple, although with several terms. By developing a third recursion would yielding the number of terms up to sixteen, which is difficult to analyze. However, the growing of recursion number does not enlarge the algorithm complexity or computation, since there is a dropping of the states to be computed at each recursion.

The motivations to use the q-entropy are: 1) managing only a simple parameter q to search for good segmentations; 2) as suggested in [12], the mammographic images may have a non-extensive behavior; 3) it is simple and makes the implementation easy having a low computational overload.

III. 3-D BREAST IMAGE SEGMENTATION AND CLASSIFICATION

As we highlight above, in this paper we proposed a five step protocol in order to classify 3D breast images. After application of NESRA algorithm, we must to extract the lesion area from the background. The Figure 1 shows an image example of an original benign lesion (a) and the NESRA result (b). This result was obtained only with one NESRA recursion, where we used $q = 0.0001$, which suggest a non-extensive behavior of the gray scale distribution.

In the second step we use a morphological chain approach in order to extract the ROI from the background. This is accomplished through the following rule. Considering the binary image generated by NESRA (e.g Figure 1-b), let α and β be the total ROI's area and the total image area, respectively. If $\alpha \geq \xi\beta$ an erosion is carried out, and if $\alpha \leq \delta\beta$ a dilation is carried out. After, assuming that the ROI has a geometric point near to the image center, we apply a region growing algorithm which defines the final ROI's boundary. In our experiments, we fixed $\xi = 0.75$ and $\delta = 0.25$ to correctly extracting most of all ROIs. The result of this morphological rule applied in the image of Figure 1-b is shown in Figure 1-c.

The region generated by the morphological chain rule is a coarse representation of the lesion region. Then, we apply a level set framework using as initialization this region's boundary [14]. The result can be seen in Figure 1-d, which was accomplished with only 10 iterations.

The next step is the feature extraction of the ROI. In our work, two radiologists stated five features that have high probability to work well as a discriminator of malignant and benign lesions. Then, we have used these features and tested them in order to achieve the best combination in terms of performance.

The first feature is the normalized lesion area. Since malignant lesions generally have large areas in relation to benign ones, this characteristic is a power discriminant. The second characteristic is the region circularity. Since benign

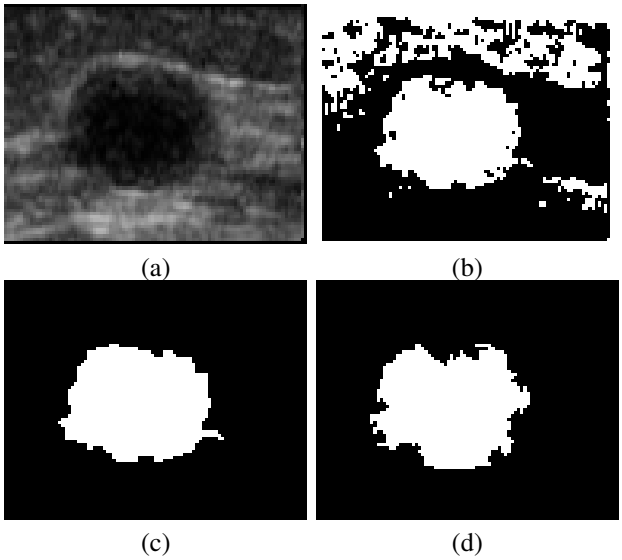

Fig. 1. (a) original ultrasound benign image; (b) NESRA segmentation with two recursions and $q = 0.0001$; (c) ROI after morphological chain application; (d) ROI after level set application.

lesions generally have more circular areas compared with the malignant ones, also this can be a good discriminant. Then, the standard deviation of the normalized distances of each boundary point to the ROI's geometric center is taken as a circularity measure. Since Malignant lesions have high average lobe areas, the third feature is the size distribution of the lobes in a lesion. A boundary's lobe is a protuberant region on the boundary. We compute the convex hull of the ROI and the lobe as a protuberance between two valleys. The next feature is related to the homogeneity of the lesion. Malignant lesions tend to be less homogeneous than benign ones. Then, we take the BGS entropy over the gray scale histogram relative to the maximum entropy as the fourth discriminant feature. As higher the relative entropy more homogeneous is the lesion region. The last feature is related with an interesting characteristic of the lesions: the acoustic shadow. When benign lesions have many water particles, the formation of an acoustic reinforcement below it is more probable. On the other hand, when the lesion is more solid (a malignant characteristic), there is a tendency in forming an acoustic shadow. Then, by comparing the region inside and below the lesion may give an idea of the region malignance. Then, we compute the gray scale histograms of both regions (inside and below the lesion) and compare them. As more darkness is the region below the lesion more is the acoustic reinforcement and, consequently, higher is the probability to have a benign lesion. Obviously, this is not a general rule and then we must to combine several features in order to have more precise inferences.

Then, as final step of our proposed protocol, we use a Nonlinear Support Vector Machine (SVM) framework in order to classify the lesion area by combining these five US features. SVM is a very good discriminant tool with successful application in several areas. In our work, these features compose a five-dimensional space used as SVM's

input. Also, since this data do not have a linear separation, we have used a B-Spline curve as the SVM's kernel.

IV. Performance Evaluation

In order to test our proposed method we used a 50 pathology-proven cases data base – 20 benign and 30 malignant –. Each case is a sequence of 5 images of the same lesion. Then, we tested 100 images of benign lesion and 150 of malignant ones. Since the detection of a malignant lesion between five images of the same case indicates a malignant case, it is reasonable to consider 250 different cases.

Since our data base is small we have improved the results through a cross-validation method. Then, these ultrasonic images are randomly divided into five groups. We first set the first group as a testing group and use the remaining four groups to train the SVM. After training, the SVM is then tested on the first group. Then, we set the second group as a testing group and the remaining four groups are trained and then the SVM is tested on the second. This process is repeated until all five groups have been set in turn as testing group.

To estimate the performance of the experimental result, five objective indices are used. These indices are accuracy, sensitivity, specificity, positive predictive value and negative predictive value. In our experiment, accuracy of SVM with B-Spline kernel for classification malignancies is 95.2% (238/250), the sensitivity is 97% (97/100), the specificity is 94% (141/150), the positive predictive value is 91.51% (97/106) and the negative predictive value is 97.92% (141/144)[1]. We also show the ROC analysis. In Figure 2 we

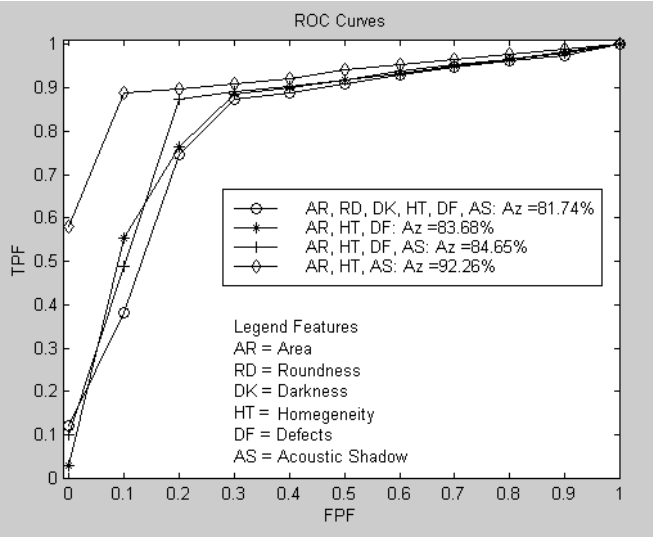

Fig. 2. ROC curves for the proposed CAD System based on NESRA algorithm for several features combinations.

plot several ROC curves for different combinations of the lesion features. The curves are normalized.

[1]TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative; Acuracy = (TP+TN)/(TP+TN+FP+FN); Sensitivity = TP/(TP+FN); Specificity = TN/(TN+FP); Positive Predictive Value = TP/(TP+FP); and Negative Predictive Value = TN/(TN+FN).

V. Discussion and Conclusions

Geometrical and Textured information from lesion area in ultrasound images provide important discriminant for computer-aided diagnosis systems. Since ultrasound images generally have complex characteristics between pixels it is interesting to study them from the point of view of non-extensive entropy. Then, our work proposes a SVM diagnostic system which uses as input a five-dimensional feature space characteristics. These characteristics are based on geometrical and textured information and should be combined in order to tuning the system. In our experiments, the best combination was achieved when only area, homogeneity and shadow were used.

According to Fig. 2, the better feature combination occurs when we have only area (AR), homogeneity (HT) and acoustic shadow (AS), which generates $A_z = 95\%$.

Acknowledgment

The authors are grateful to CAPES, CNPq and FAPERJ (Brazilian agencies for Scientific Financing) for the support of this work.

References

[1] American Cancer Society, Cancer Facts & Figures, 2004.
[2] J. S. Suri and R. M. Rangayyan, *Emerging Technologies in Breast Imaging and Mammography*: ASP Press, 2006.
[3] J. S. Suri and R. M. Rangayyan, *Recent Advances in Breast Imaging, Mammography and Computer Aided Diagnosis of Breast Cancer*: SPIE Press, April, 2006.
[4] D.R. Chen and W. J. Kuo and R.F. Chang and W. K. Moon and C. C. Lee, "Use of the bootstrap technique with small training sets for computer-aided diagnosis in breast ultrasound," *Ultrasound Med. Biol,* vol 28, no. 7, pp. 897-902, July, 2002.
[5] R. F. Chang and W. J. Wu and W. K. Moon and D. R. Chen, "Improvement in breast tumor discrimination by support vector machines and speckle-emphasis texture analysis", *Ultrasound Med. Biol,* vol. 29, no. 5, pp. 679-686, 2003.
[6] C. M. Chen and Y. H. Chou and K. C. Han and G. S. Hung and C. M. Tiu and H. J. Chiou and S. Y. Chiou, "Breast Lesion on Sonograms: Computer-Aided Diagnosis with Neural Setting-Independent Features and Artificial Neural networks", *Journal of Radiology,* vol. 226, pp. 504-514, 2003.
[7] W. J. Kuo and R. F. Chang and W. K. Moon and C. C. Lee and D. R. Chen, "Computer-aided diagnosis of breast tumors with different US systems", *Acad. Radiol.*, vol. 9, no. 7, pp. 793-799, July, 2002.
[8] A. Sawaki and K. Shikamoto and H. Satake and T. Ishigaki and S. Koyama and Y. Obata and M. Ikeda, "Breast Ultrasonography: diagnostic efficacy of computer-aided diagnostic system using fuzzy inference", *Radiat. Med.*, vol. 17, no. 1, pp. 41-45, January, 1999.
[9] B. S. Garra and B. H. Krasner and S. C. Horii and *et al.*, "Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis",*Ultrasound Imaging*, vol. 15, no. 4, pp. 267-285, 1993.
[10] D. R. Chen and R. F. Chang and Y. L Huang, "Computer-Aided Diagnosis Applied to US of Solid Breast Nodules by Using Neural Networks", *Journal of Radiology.*, vol. 213, pp. 407-412, 1999.
[11] R. F. Chang *et al.*, "Computer-aided Diagnosis for 2-D/3-D Breast Ultrasound, in Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer", SPIE, Bellighan, WA, 2006.
[12] M. P. Albuquerque and M. P. Albuquerque and I. A. Esquef and A. R. G. Mello, "Image thresholding using Tsallis entropy", *Pattern Recognition Letters*, vol. 25, pp. 1059-1065, 2004.
[13] C. Tsallis, "Nonextensive Statistical Mechanics and its Applications", *Series Lecture Notes in Physics*, Springer, Berlin, 2001.
[14] J.S. Suri *et al.* "Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review," IEEE Trans. Inf. Technol. Biomed., vol. 6, no. 1, pp. 8-28, Mar. 2002.