SEMIAUTOMATED BREAST CANCER CLASSIFICATION FROM ULTRASOUND VIDEO

L. Bocchi, F. Gritti, C. Manfredi

Dept. of Electronics and Telecommunications,
University of Florence, Florence, Italy

E. Giannotti, J. Nori

SOD Diagnostica Senologica
Az. Osp. Univ. Careggi, Florence, Italy

ABSTRACT

The present work presents a new approach, based on the elaboration of a whole video clip of the ecoghraphic acquisition describing the lesion from different viewpoints. The focus of the paper is in comparing performances which can be obtained, using a simple classification algorithm, using the full video clip against the more common single frame approach. The system consists of five modules: preprocessing, semiautomatic segmentation, extraction of morphological features, classification of each frame and integration to obtain the classification of the video clip. Results show that the data integration allows an increase in the correct classification rate from approximately 89%, on a frame-by-frame basis, up to 97% on the whole video clip.

Index Terms— Breast ultrasound, classification, video

1. INTRODUCTION

Breast cancer is the second leading cause of death for women all over the world [1]. An early diagnosis and treatment of this pathology is the key to reduce mortality. UltraSound (US) imaging is commonly used for early detection and diagnosis of breast nodules [2].

The correlation between image features extracted from US images and probability of malignancy has been deeply studied by several researchers. An extensive overview of the techniques, along with possible approaches for automated breast cancer detection from ultrasound images, has been recently published by Cheng et al. [3]. The review reports best results obtained so far by different authors, comparing the relative performance by means of the area under the Receiver Operating Characteristic (ROC) curve (A_z). The comparison indicate that the highest values of A_z are around 0.97. However, the review also highlights the lack of a publicly available benchmark to obtain a fair comparison among the different approaches. Proposed approaches differ also in the classification target: some studies are focused in lesion detection and segmentation [4, 5], while other are focused on the discrimination between malignant and benign lesions [6, 7, 8].

All the approaches that have been proposed so far for designing an automated diagnosis system analyze and classify a single (or a few, for instance [9]) US frame and are therefore restricted to employ only the information which is actually embedded in the single selected frame, relying on the expertise of the radiologist to simplify the problem. Also, research works propose the use of 3D ultrasound, however, at present, there is no clinical evidence of 3D ultrasound in routine practice.

The selection process, performed during the clinical investigation, relies on the experience of the radiologist in identifying the most representative frame(s), which is (are) used to document and support diagnosis. However, manual diagnosis, although summarized in a small number of frames, is performed by the radiologist viewing the whole acquisition, and therefore taking into account all the information perceptually gathered from the whole exam.

Therefore, our research work is focused in gathering and processing all the information contained in a few seconds of an US video clip, composed of hundreds of images of the same lesion, where each frame describes the lesion from a different point of view, in order to support the diagnostic procedure of the radiologist. In the proposed approach, each frame in the clip is processed independently, in order to obtain a tentative classification of the lesion as benign or malignant. Afterward, such classifications obtained from all frames in the clip are integrated to obtain a more reliable classification of the whole clip. The presence of uncertainty in the classification of some frames does reflect the clinical evidence of a lesion presenting different properties when observed from some viewpoints.

2. MATERIALS AND METHODS

The proposed method has been developed using the Matlab 7.0 software. A preprocessing step allows to split the clip into frames, where the appropriate ROI is selected. Each frame is segmented using a semiautomated method based on active contours, and is classified accordingly to its probability of malignancy. In the last step, the individual classifications are combined to give a final assessment of malignancy. The next sections describe each step in detail.

Acquisition and data set

The present study has been developed using a set of US acquisitions collected at the SOD Diagnostica Senologica,

AOUC Careggi in Florence, Italy, during routine diagnosis of breast lesions (masses and nodules). The acquisitions have been performed, during ambulatory visit, often in preparation of biopsies, according to the clinical procedures for diagnosis. Data were collected using a Technos Ultrasound device (Esaote spa) equipped with a 10 MHz linear real-time transducer. Each recording lasts 5s, with spatial and temporal resolution corresponding to the actual settings used by the radiologist to optimize the appearance of the lesion. Settings could vary accordingly to the different lesion depth. The radiologist, after identifying the suspicious lesions, starts the recording, using the dedicated function of the Technos instrument. During the recording, which has a preset duration of 5 seconds, the radiologists explores the lesion from different viewpoints, maintaining the lesion inside the field of view. At present, the data set includes 30 cases, 9 of them having a malignant result, while the remaining ones (21) being benign lesions.

Given the small number of cases in the dataset, a previously available dataset, composed of static US frames, and used in our previous work [10], has been used for developing and tuning of the single-frame elaboration and classification procedure. All parameters referring to the segmentation procedure have been optimized with respect to the results obtained using only the previous dataset, thus avoiding any bias related to the utilization of the same images during the design and the test phases. The proposed algorithm is reported in the block diagram shown in Figure 1. The main steps of the algorithm are discussed below, while a detailed description of the feature extraction phase and of the neural network, together with an preliminary evaluation of the results obtained on the single frame classification is reported in [10].

Preprocessing

In the preprocessing step each video clip is split in a set of images, which are later cropped according to the selected Region of Interest (ROI), which is defined in this context as the whole field of view of the instrument, with the exclusion of the textual information included in the video record. All video clips have a fixed time length of 5s, which corresponds to 100-150 frames, depending on the instrument setting.

Segmentation

An accurate segmentation of the lesion is a prerequisite for a successful characterization of the malignancy. In this work, we used a segmentation approach based on active contours, or *snakes* [11] using the greedy algorithm [12].

The initialization of the snake has been performed manually on the first frame of the video clip. The developed software allows the radiologist to identify the lesion, at user choice, either by its center and radius, thus creating a circular contour, or by manual selection of a set of points arranged over the lesion contour. In both cases, the algorithm interpolates the provided points to obtain an adequate number (100 points) of control points to drive the snake algorithm. The result of the segmentation is shown to the user who can visually evaluate the obtained contour; if the segmentation is satisfactory, the algorithm goes in to the next frame, otherwise the user is requested to manually correct the segmentation of the lesion. The final positions of the control points of the snake are used to convert the image to binary and to extract the shape of the lesion.

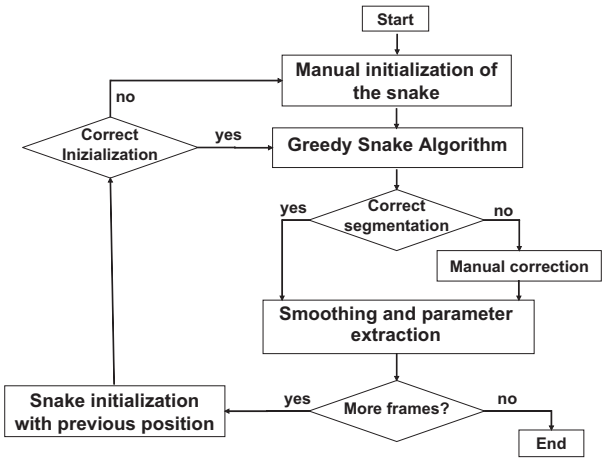


Fig. 1. Block diagram of the segmentation algorithm.

Once the segmentation of the frame is accomplished, the algorithm automatically uses the final position of the snake for the initialization of the contour in the subsequent frame. In few cases the characteristics of the lesion change dramatically from frame to frame. In such cases, the user can perform a manual re-initialization of the snake.

Feature extraction

A basic set of shape features have been extracted from each frame to achieve a frame-by-frame classification. The features have been selected according to the guidelines introduced by the BIRADS protocol, elaborated by the American College of Radiology [13], in order to simplify the interpretation of the outcome by the clinician. In our work , we focus on three parameters (see [10] for details) that describe morphological features (shape and axes ratio), whose regularity is related to benign lesions, and echogenic features (interface halo), which is also a strong predictive sign for benign lesions.

The morphology of the lesion is determined by analyzing the curve representing the distance between the centroid and the lesion margin along the perimeter; a low number of local and absolute minima is related to a regular shape, while a high number of minima denotes morphological irregularity. A b-spline approximation (4th order b-spline, with 10 equally spaced knots) has been applied for smoothing out the noise and extract the overall shape of the curve. The exact number of knots did not prove critical, as the number of minima N_m is practically constant varying the numbers of nodes over a reasonable interval.

The axes ratio $A_r = H/V$ is defined as the ratio between longitudinal axis and latitudinal axis, and is evaluated here

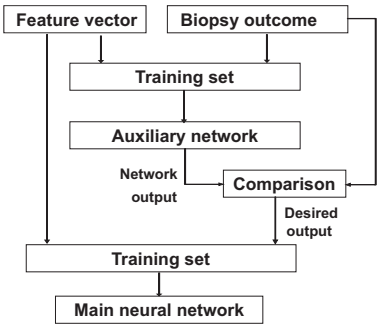


Fig. 2. Block diagram of the training procedure of the classification system.

by measuring the horizontal (H) and vertical (V) axes of the lesion across the centroid. If one of the axes intersects the lesion in more than two points (highly irregular lesions) the considered axis length spans from the first to the last point, thus including any possible gap in between.

The echogenic feature describes the possible presence of a hyperechogenic halo surrounding the segmented lesion, which is an indication of benignity. The halo has been evaluated by defining an annular region of fixed width outside the perimeter of the lesion. Assuming the number of pixel in the ring is N_r and the average gray level of the lesion is g_l, we evaluated the number N_b of pixels P_i belonging to the ring which satisfy the relation $g(P_i) > g_l$, where $g(P)$ is the gray level of pixel P. The echogenic feature is then defined as $e = N_b/N_r$. This definition reduces the dependency of the parameter from the instrument settings, by avoiding the introduction of a threshold or other parameters.

Classification

The proposed approach assumes that each video clip describes the lesion from different angles, some of which may appear benign, other malignant. This observation points out the possible presence of benign-looking frames in a clip related to a malignant lesion and, on the other side, the presence of malignant-looking frames in a clip of a benign lesion, defined here as "misleading frames". This problem has been handled using a classifier based on a two-step training process, aimed to identify and adequately manage the presence of misleading frames.

The classifier has been designed using a neural network, composed of three inputs (corresponding to the feature vector $f = \{N_m, A_r, e\}$, normalized in the [0,1] range), a hidden layer of ten units, and one output unit. All units have the hyperbolic tangent as activation function. The number of units in the hidden layer has been selected experimentally on a preliminary test [10]. The training has been performed using the standard back-propagation rule in 300 iterations, using an adaptive learning rate, with momentum. In the training phase, the classifier is coupled to an auxiliary network, with the same structure, devoted to detect misleading frames (Figure 2).

The training procedure consists of two steps. During the first step, the auxiliary network is trained to classify each frame of the training set, using the outcome of the biopsy as teaching input (where a malignant lesion is labeled with +1 and a benign one with -1). After this training step, the auxiliary network output is tested and compared with the desired input. The outcome of the test is then used to build a teaching input for the primary network, using a four levels logic. The four levels describe frames as "reliable" or "misleading" depending whether the auxiliary network correctly classifies them. Reliable frames are labeled with ± 1, respectively for malignant and benign lesions. Frames which were incorrectly classified by the auxiliary network are assumed to be misleading ones, and have been assigned a lower degree of truth, with a level equal to ± 0.5, for frames incorrectly classified as benign or malignant, respectively. This approach gives a "fuzzy" classification of each frame which has been used to train the final classifier. Once the training has completed, the auxiliary network is discarded, and the test samples are directly fed to the main network.

Data integration

The final step requires the integration of the results obtained for each frame into a single classification of the whole video clip. The classification outcome of each frame should contribute to the final classification, but different frames may give different contributes to the final result. We assumed that the information provided by each frame is roughly inversely proportional to the "uniqueness" of the image contained in the frame: if frames are similar each other, the information provided by each of them is small. On the other side, if frames considerably differ from each other, they carry a larger amount of information.

An estimate of the "uniqueness" of the information associated to each frame has been performed by thresholding the correlation matrix C_{ij} between each pair of frames using a threshold value $\theta = 0.85$, which corresponds to a strong similarity between the two frames. Then, for each frame j we evaluated the number n_j of consecutive entries of the correlation matrix which are above the selected threshold, i.e. if $C_{ij} > \theta$ for each i between i_1 and i_2, then $n_j = i_2 - i_1 + 1$. We defined the weights μ_j as:

$$\mu_j = 1/n_j \tag{1}$$

According to this, we obtained an estimate of the degree of malignancy of the clip (out_c) using a weighted sum of the classification out_j of each frame j belonging to the clip:

$$out_c = \sum_{j=1}^{N} \mu_j \cdot out_j \Big/ \sum \mu_j \tag{2}$$

3. RESULTS

Given the limited number of cases included in our dataset, the validation of the system has been performed using the leave-

Table 1. Manual segmentations required for each video clip.

Manual segmentations	0	1-4	5-8	>9
% of video clips	40%	30%	20%	10%

Table 2. Average and standard deviation of the classification, over 10 runs of the leave-one-out method

	Accuracy	**Sensitivity**	**Specificity**	A_z
frame	$89 \pm 7\%$	$89 \pm 4\%$	$90 \pm 2\%$	$95 \pm 1\%$
clip	$97 \pm 1\%$	$94 \pm 4\%$	$99 \pm 1\%$	$98 \pm 1\%$

one-out method. One of the video clips, in turn, has been removed from the data set for creating the training set. After the training is complete, the resulting network has been tested on the removed clip.

In a first step, a preliminary evaluation has been carried out to assess the performances of the segmentation algorithm, focusing on the evaluation of the number of manual interventions required to achieve a successful segmentation of the clip. Results are shown in Table 1. On the overall, only 3.3% of the frames (102 out of 3132) required manual segmentation.

The performances of the classifier has been tested as concerns both the outcome of the frame-by-frame classification and the overall classification of the clip (Tab. 2). In both cases, to increase the significance of the results by reducing the impact of the random initialization of neural networks, ten iterations of the leave-one-out method has been performed. Reported results include the true positive (TP) and true negative (TN) rates, the accuracy (number of correct classifications CC), and the area under the ROC curve (A_z).

Results obtained for the frame-by-frame classification indicate an accuracy of the system equal to 0.89 ± 0.07 over the ten iterations of the training procedure. The integration step significantly improves the obtained results, with an accuracy equal to 97%. A detailed analysis of each run indicates in 9 out of 10 runs there is only one misclassified clip, while in the other one all the clips have been correctly classified, corresponding to 100% accuracy.

4. CONCLUSIONS

The main focus of the paper is in comparing the single frame classification scheme, commonly used in CAD systems, to a clip classification. Each frame, representing a single slice of the 3D lesion, does not necessarily present the typical properties of a malignant lesion. The proposed approach, therefore, aims to integrate the information coming from different sections without the additional technology involved in a volumetric acquisition of the breast using 3D ultrasound imaging.

Results show that this approach can provide a significant improvement over the single-frame classification commonly used in computer aided diagnosis systems. Although the results obtained on a single frame classification show that the proposed approach gives results not better than results obtained with previously published approaches, the integration of the data over a whole video clip substantially improves the results, giving a correct classification rate close to 100%, reached in one-out-of-ten runs of the leave-one-out algorithm.

5. REFERENCES

[1] American Cancer Society, "Cancer facts & figures 2010," Atlanta, 2010.

[2] A. Stavros *et al.*, "Solid breast nodules: use of sonography to distinguish between benign and malignant lesions," *Radiology*, vol. 196, no. 1, pp. 123–134, 1995.

[3] H. Cheng *et al.*, "Automated breast cancer detection and classification using ultrasound images: A survey," *Pattern Recognition*, vol. 43, no. 1, pp. 299 – 317, 2010.

[4] K. V. Mogatadakala *et al.*, "Detection of breast lesion regions in ultrasound images using wavelets and order statistics," *Med Phys*, vol. 33, no. 4, pp. 840–849, 2006.

[5] M. A. Kutay, A. P. Petropulu, and C. W. Piccoli, "Breast tissue characterization based on modeling of ultrasonic echoes using the power-law shot noise model," *Pattern Recogn. Lett.*, vol. 24, pp. 741–756, February 2003.

[6] P. M. Shankar *et al.*, "Application of the compound probability density function for characterization of breast masses in ultrasound b scans," *Physics in Medicine and Biology*, vol. 50, no. 10, p. 2241, 2005.

[7] J. H. Song *et al.*, "Artificial neural network to aid differentiation of malignant and benign breast masses by ultrasound imaging," in *Med. Im. 2005*, vol. 5750, no. 1. SPIE, 2005, pp. 148–152.

[8] B. Sahiner *et al.*, "Malignant and Benign Breast Masses on 3D US Volumetric Images: Effect of Computer-aided Diagnosis on Radiologist Accuracy1," *Radiology*, vol. 242, no. 3, pp. 716–724, March 2007.

[9] N. P. Gruszauskas *et al.*, "Performance of breast ultrasound computer-aided diagnosis dependence on image selection," *Academic radiology*, vol. 15, no. 10, pp. 1234–45, 2008.

[10] L. Favilli and L. Bocchi, "Automatic system for the analysis and the discrimination of breast nodules in ultrasound imaging," in *World Congr on Med Phys & Biom Eng, Munich, Germany*, vol. 25/4, 2009.

[11] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[12] D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation," *CVGIP: Image Underst.*, vol. 55, pp. 14–26, January 1992.

[13] American College of Radiology, "Bi-rads: ultrasound, 1st ed." in *Breast imaging reporting and data system: BI-RADS atlas*, Reston, VA, 2003.