

Computational Techniques for the Support of Breast Tumor Diagnosis on Ultrasound Images

Miguel Alemán-Flores¹, Patricia Alemán-Flores², Luis Álvarez-León¹,
José Manuel Santana-Montesdeoca², Rafael Fuentes-Pavón²
and Agustín TrujilloPino¹

¹Departamento de Informática y Sistemas

Universidad de Las Palmas de Gran Canaria, 35017, Las Palmas, Spain

²Sección de Ecografía, Servicio de Radiodiagnóstico

Hospital Universitario Insular de Gran Canaria, 35016, Las Palmas, Spain

Abstract

This work presents a new approach in the application of computer vision techniques to assess the diagnosis of solid breast tumors on ultrasound images. Most works related to medical image analysis for breast cancer detection refer to mammographies. However, the studies carried out by the radiologists prove the significance of some aspects observed on ultrasound images, which are spiculation, calcifications, ellipsoid shape, taller than wide shape, echogenicity, capsule, angular margins, microlobulations, two or three gentle lobulations, acoustic shadow and ramifications. We have developed a common framework for the analysis of such diagnosis criteria, so that they can be measured by the computer and a set of parameters are available for the physicians to decide whether the biopsy is necessary or not. The results produced by the system have been compared to those given by three specialists to adjust the parameters and test their accuracy. The presence or absence of the diagnosis factors coincides for human and computer in most of the nodules. In some cases, the computer indicated the presence of suspicious factors which were not actually worrisome, but the most important aspect is avoiding overlooking the malignant nodules without needing carrying out a biopsy in all of them. We present a set of mathematical methods to extract objective evidence of the presence or absence of the diagnosing criteria. This system is able to extract the relevant features for solid breast nodules with high accuracy and represents a very valuable help in the assessment of radiologists.

1 Introduction

There are certain criteria which are used by the physicians to determine whether a solid breast nodule is benign or malignant. This work uses the computational techniques for the study and classification of shapes and textures to help the physicians in the early diagnosis of breast tumors on ultrasound images. Contrary to other works which deal with mammographies, this one is based on the analysis of ultrasound images, in which a series

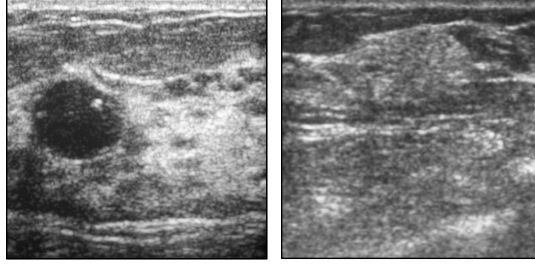


Figure 1: Markedly hypoechoic nodule (left) and hyperechoic nodule (right)

of diagnosis factors can be observed. Some previous works try to adjust the parameters of the ultrasound systems to help in the decision making process [1], segment the tissues [2], or deal with certain particular aspects of the nodules, such as texture [3]. The analysis of ultrasound images that we propose allows performing precise measurements of a series of diagnosing factors in order to characterize each nodule according to the values of the parameters which have been obtained.

Spiculation, taller than wide shape, angular margins, markedly hypoechoic appearance, acoustic shadowing, calcifications, ramifications and microlobulations are considered as malignant findings which lead the physicians to carry out a biopsy. On the other hand, hyperechogenicity, ellipsoid shape, 2 or 3 gentle lobulations as well as a thin and echogenic capsule suggest the benignity of the nodule [4].

It is not always easy for a human observer to measure some of these features and provide objective evidence of the benignity or malignancy of the tumor. However, computer vision techniques may help detecting some of them and supplying numerical measurements of the presence and relevance of these factors, in such a way that the risk of overlooking a malignant nodule can be reduced [5] [6]. The purpose of this work is not the substitution of the human observer or the physician's interpretation, but the addition of a precise, objective and invariable examiner which can draw the specialist's attention to those areas where each diagnostic factor is likely to be present.

2 Description of the Diagnosis Criteria

In this section, we will describe each one of the criteria that will be considered in the diagnosis of the nodules, both malignant and benign. We will also present some examples of the results provided by the computer-aided system.

First, we will consider the generation of echo by the nodule and the area around it. The hyper or hypoechogenicity of a nodule, i.e. the excessive or defective generation of echoes by that nodule, is measured with respect to the echogenicity of the fat of that particular woman, since it is the relative intensity and not the absolute one that determines whether it is hyper or hypo-echoic. This feature is considered because of the hyperechogenicity of some benign tumors and the marked hypoechogenicity of malignant nodules. In fact, the central part of two-thirds of malignant solid breast nodules is very hypoechoic with respect to fat. In Fig. 1, we can see an example of a markedly hypoechoic nodule (left) and a hyperechoic nodule (right).

Not only the generation of echo within the nodule is relevant, but also how the region

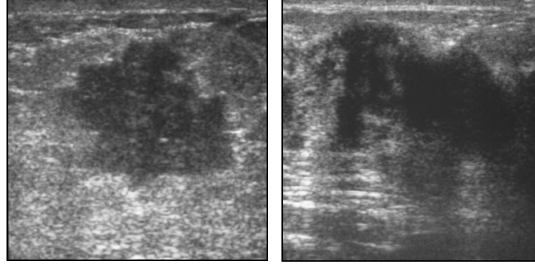


Figure 2: Nodule without acoustic shadowing (left) and with acoustic shadowing (right)

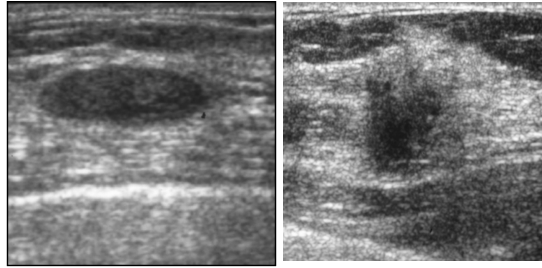


Figure 3: Wider-than-tall nodule (left) and taller-than-wide nodule (right)

under it is affected. A nodule is said to produce acoustic shadow if the ultrasound is attenuated when crossing through it. In this case, the region beyond the nodule, from the point of view of the ultrasound emitting device, is darker than the region at the same level which is not under the nodule, since its emission of echo is not altered by the nodule's being in between. In Fig. 2, we observe how a nodule can leave the generation of echo unaltered (left) or interfere it (right). The generation of shadowing is considered as a malignant factor.

We will now focus on the general shape of the nodule. A nodule is said to be taller than wide when it is greater in the direction normal to the skin than in the parallel ones. When this ratio is quite high, the possibility of being analyzing a malignant nodule increases, since this reflects that the nodule is growing across normal tissue planes. On the other hand, benign nodules usually affect a single tissue and appear perpendicular to the direction of the ultrasound signals. Figure 3 shows a wider-than-tall nodule (left) and a taller-than-wide one (right).

An ellipsoid shape increases the probability of the nodule's being benign, and even if a nodule does not present an ellipsoid-like shape, it is considered as benign if it presents two or three smooth, well circumscribed and gentle lobulations. Figures 4 and 5 present a nodule with elliptical shape and a nodule with two smooth and gentle lobulations, respectively, as well as the ellipses extracted by the computer-aided system to help classifying the nodules.

Contrary to the elliptical shapes described above, malignant breast nodules may present projections from the surface of the nodule which extend radially within a duct towards the nipple (duct extension) and/or within ducts away from the nipple (branch pattern). Figure 6 shows a nodule with ramifications, which have been signaled by the system.

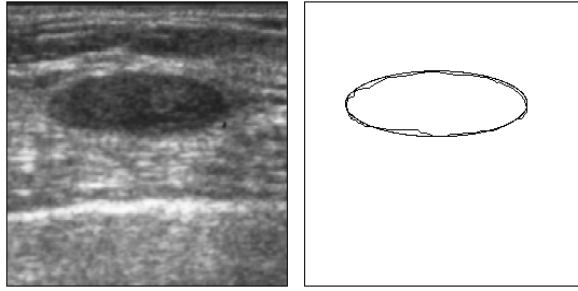


Figure 4: Ultrasound image of a solid breast nodule (left), contour and ellipse extracted from that contour (right)

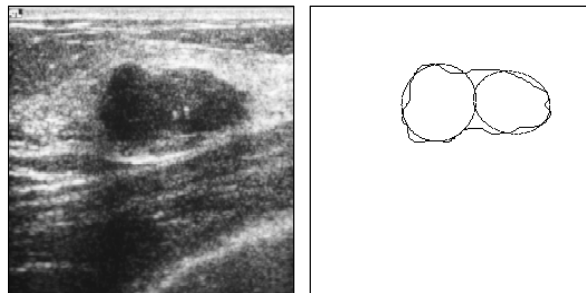


Figure 5: Ultrasound image of a solid breast nodule (left), contour and combination of two ellipses extracted from that contour (right)

Some other factors do not refer to the general shape, but to local variations of the contour. For example, the finding of angular margins makes the nodule suspicious of being malignant. Angular margins refer to the contour of the junction between the hypoechoic or isoechoic solid nodule and the surrounding tissues. The entire surface of the lesion must be surveyed, since they may involve only a part of it. Similarly, microlobulations are observed as small ellipsoid arcs in certain regions of the contour and are frequently associated with angular margins. Figures 7 and 8 show the angular margins and microlobulations extracted for a nodule, respectively.

Spiculation consists of contrasting lines radiating out perpendicularly from the surface

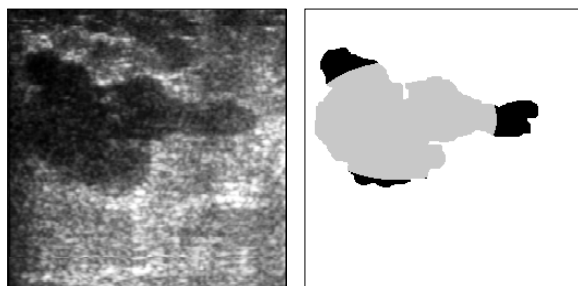


Figure 6: Ultrasound image of a solid breast nodule (left), contour and ramifications extracted by the system (right)

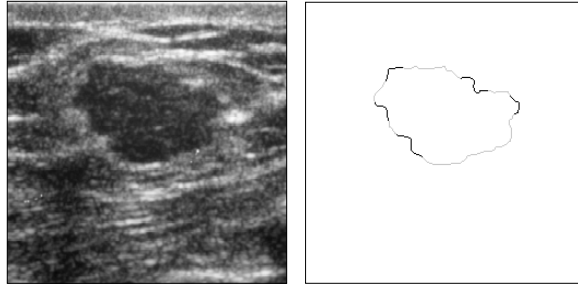


Figure 7: Ultrasound image of a solid breast nodule (left), contour and angular margins extracted from that contour (right)

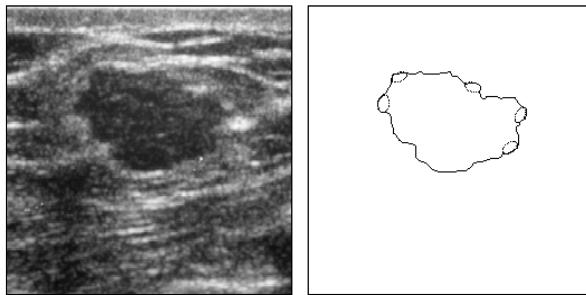


Figure 8: Ultrasound image of a solid breast nodule (left), contour and microlobulations extracted from that contour (right)

of the nodule. Figure 9 shows the spiculated areas extracted for a nodule.

Finally, the presence of punctate echogenic calcifications within a solid breast nodule is worrisome for malignancy. Calcifications are identified as highly contrasting white points inside the nodule. Figure 10 shows an example of a nodule with calcifications and where they are located by the computer.

In the following section, we show how the ultrasound images have been processed to analyze these factors on them.

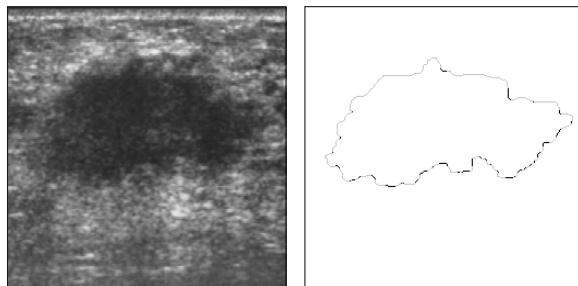


Figure 9: Ultrasound image of a solid breast nodule (left), contour and spiculations extracted from that contour (right)

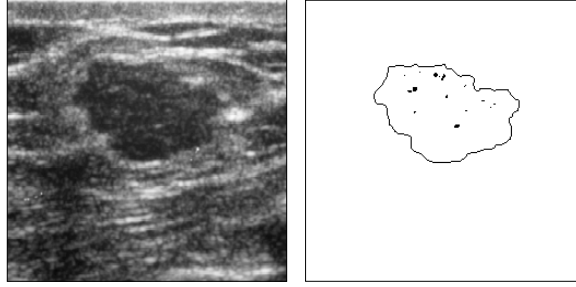


Figure 10: Ultrasound image of a solid breast nodule (left), contour and calcifications extracted from that contour (right)

3 Computer Processing of the Diagnosis Criteria

In this section, we will show how the criteria described above have been processed by the computer-aided system. First, we will explain how the nodule has been segmented, then we will present the processing of the intensity-related criteria, and finally we will explain how the shape of the nodule has been analyzed.

3.1 Image Segmentation

Before measuring the diagnosis criteria on the nodule, it must be located and delimited. Thus, we must segment the interesting area and the background in order to identify the regions which will be processed in the further studies. The segmentation of the images to separate the nodule from the surrounding tissues is performed by means of a truncated median filter, a balloon algorithm and a dilation process. Due to the presence of Rayleigh distributed noise, which characterizes ultrasound images, the mode filter could be appropriate to denoise the images, since it eliminates the spurious negligible points. However, its performance depends strongly on the size of the window in which the filter is applied and, if it is not large enough, it may result in modeless regions or quite arbitrary values. That is why we have used the truncated median filter, which produces quite good results in a few iterations. In this filter, the median is calculated in each iteration and the further values are eliminated, so that median and mode approach. For a given pixel (x, y) , the process is as follows: let f_n be the histogram of the grayvalues in the $w \times h$ window from $x - (w - 1)/2$ to $y + (h - 1)/2$ (n ranges from 0 to 255), and let m be the median of such histogram, i.e. $\sum_{k=0}^{k=m-1} f_k < (w * h)/2$ and $\sum_{k=0}^{k=m} f_k \geq (w * h)/2$, those values which differ from m more than the minimum of m and $255 - m$ are eliminated and the median is recalculated [6]. After a few iterations, the median can be considered as a good approximation of the mode. Figure 11 shows some iterations of the application of this filter on an ultrasound image.

From a seed into the nodule, which must be entered to the system by the user, a balloon algorithm allows filling the interesting area as connected intensity-similar pixels. The pixels which are connected to the already selected ones, whose intensity is similar to theirs and inside a certain interval are included into the selected region. If given two thresholds for intensity and difference, no more points can be incorporated, the thresholds are increased until no point is added in a given iteration or the added region is larger to

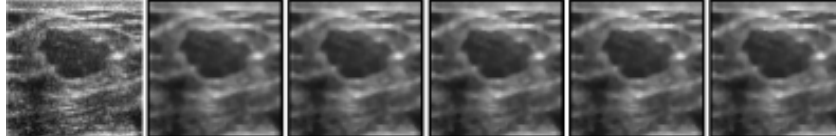


Figure 11: Iterative filtering of an ultrasound image using the truncated median filter

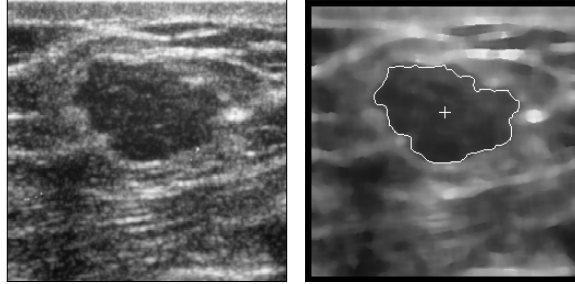


Figure 12: Ultrasound image of a nodule (left), contour and center on the filtered image (right)

the previous one, since the latter might mean that the background has been connected to the nodule.

Due to the difficulty that many of these images present, the user is allowed to correct the result of the automatic segmentation in case it does not fit the area covered by the nodule. Thus, adding and subtracting regions in the segmentation will produce a more reliable delimitation of the nodule to proceed to the further analysis. Fig. 12 shows an example of the ultrasound image of a nodule (left), the filtered image, the contour and the center of the selected region (right).

3.2 Intensity Related Features

In this section, we consider those factors which refer to the graylevels of the image, which represent the echo produced at each point when the ultrasound reaches it. These are the echogenicity of the nodule, the presence of acoustic shadow under the nodule and the observation of a thin echogenic capsule around the nodule.

Using the binary segmented image as a mask, we calculate the mean intensity of the region covered by the nodule. This is its mean echogenicity and, similarly, a fat region is selected by the user to extract its echogenicity and compare them. This comparison will help determining whether the nodule is hypo or hyperechogenic.

To determine whether the nodule is producing acoustic shadowing, we compare the region under the nodule and the region at the same level which is not covered by the nodule. As done before, we calculate the mean intensity (echogenicity) of both regions. The comparison of both mean intensities allows determining whether the nodule is absorbing the ultrasound or not.

By using the structure tensor method, we have estimated the magnitude of the gradient in every point inside the nodule. Those points where the intensity and the magnitude of

the gradient is quite high have been selected as calcifications (see Fig. 10). In order to estimate the magnitude of the gradient, we first calculate an initial estimation for every point using the following mask for the horizontal component u_x :

$$\frac{1}{4h} \begin{pmatrix} -(2 - \sqrt{2}) & 0 & (2 - \sqrt{2}) \\ -2(\sqrt{2} - 1) & 0 & 2(\sqrt{2} - 1) \\ -(2 - \sqrt{2}) & 0 & (2 - \sqrt{2}) \end{pmatrix} \quad (1)$$

and the following one for the vertical component u_y :

$$\frac{1}{4h} \begin{pmatrix} -(2 - \sqrt{2}) & -2(\sqrt{2} - 1) & -(2 - \sqrt{2}) \\ 0 & 0 & 0 \\ (2 - \sqrt{2}) & 2(\sqrt{2} - 1) & (2 - \sqrt{2}) \end{pmatrix} \quad (2)$$

If we have an estimation of the gradient (x_n, y_n) for a certain point, the vector (a, b) would be parallel to that gradient if the following product is 0:

$$ay_n - bx_n = (a, b)(x_n, y_n)^\perp = \|(a, b)\| \|(x_n, y_n)\| \cos(\alpha) \quad (3)$$

where α is the angle formed by the directions of $(x_n, y_n)^\perp$ and (a, b) . Thus, if we want to find the direction which best represents the estimations for a set of points, we must calculate the vector (a, b) which minimizes the following sum:

$$\sum_{i=0}^N (ay_i - bx_i)^2 = (a, b) \begin{pmatrix} \sum_{i=0}^N y_i^2 & -\sum_{i=0}^N x_i y_i \\ -\sum_{i=0}^N x_i y_i & \sum_{i=0}^N x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \quad (4)$$

For every component of the sum in the matrix, there are two eigenvalues and eigenvectors. One of the eigenvalues is 0, and its corresponding eigenvector is the estimation of the orientation of the gradient (x_i, y_i) :

$$\begin{pmatrix} y_i^2 & -x_i y_i \\ -x_i y_i & x_i^2 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (5)$$

Since any two-dimensional vector can be expressed as a linear combination of the two linearly independent eigenvectors, the maximum variation will be achieved by the maximum eigenvector. We can see that this vector is $(-y_i, x_i)$, and its corresponding eigenvalue is $(x_i^2 + y_i^2)$, i.e. the square of the norm of the gradient estimation:

$$\begin{pmatrix} y_i^2 & -x_i y_i \\ -x_i y_i & x_i^2 \end{pmatrix} \begin{pmatrix} -y_i \\ x_i \end{pmatrix} = (x_i^2 + y_i^2) \begin{pmatrix} -y_i \\ x_i \end{pmatrix} \quad (6)$$

Consequently, $(x_i^2 + y_i^2)$ is the highest eigenvalue, and its square root is the norm of the gradient.

The capsule of the nodule is extracted by eroding the selected region and analyzing the eroded points. The echogenicity of the capsule is measured, as the previous ones, from the mean intensity of the region.

3.3 Shape Related Features

This section deals with those factors which refer to the shape of the nodule. The proportion of the width and the height of the nodule, the presence of angular margins, the ramifications which may appear, the number and size of lobulations and the ellipsoid shape are considered as benign or malignant factors.

On the binary segmented image, we calculate the mean width and the mean height of the nodule and the quotient of both measures indicates its disposition. When studying whether the nodule has ellipsoid shape or not, we extract the ellipsoid for which the distance from the points on the contour is minimum. This yields the one which best represents the set of contour points (see Fig. 4). When the mean distance is low enough, the nodule has a clearly ellipsoid-like shape. On the other hand, the higher the distance, the more dissimilar the nodule and the ellipsoid are.

The usual techniques to estimate the ellipse location in an image are based on the Hough transform or the principal axis estimation [7][8]. We propose an algorithm to estimate, in a very accurate way, the ellipse location in an image by minimizing the mean distance from a set of points to the ellipse. This minimization criterion has a more physical meaning and provides a very accurate estimation.

An ellipse depends on 5 parameters $\Theta = (x_0, y_0, a, b, \theta)$, where (x_0, y_0) is the center, (a, b) are the sizes of the principal axes and θ is the orientation. We parametrize an ellipse in the following way:

$$e_{\Theta}(t) = \begin{pmatrix} x_0 + \cos(\theta)a \cos(t) + \sin(\theta)b \sin(t) \\ y_0 - \sin(\theta)a \cos(t) + \cos(\theta)b \sin(t) \end{pmatrix} \quad t \in [0, 2\pi] \quad (7)$$

In the following lemma, we compute the distance from a point to an ellipse:

Lemma 1 *Let $(x, y) \in R^2$, and let $e_{\Theta}(t)$ be an ellipse. The distance from the point (x, y) to the ellipse is denoted by $d((x, y), e_{\Theta}(\cdot))$. Such distance is attained in a point $t_0 \in [0, 2\pi]$. If $z = \cos^2(t_0)$ then z is a root of the polynomial*

$$l^2 z^4 + 2mlz^3 + (k + 2ln + m^2)z^2 + (2mn - k)z + n^2 = 0 \quad . \quad (8)$$

where

$$\begin{aligned} \tilde{x} &= \cos(\theta)(x - x_0) - \sin(\theta)(y - y_0) \\ \tilde{y} &= \sin(\theta)(x - x_0) + \cos(\theta)(y - y_0) \\ l &= -(b^2 - a^2)^2, \quad m = (b^2 - a^2)^2 + a^2\tilde{x}^2 - b^2\tilde{y}^2, \quad k = 4\tilde{x}^2\tilde{y}^2a^2b^2, \quad n = -\tilde{x}^2a^2 \end{aligned}$$

Proof: A straightforward calculation.

Given a set of points $X = \{(x_i, y_i)\}_{i=1, \dots, N}$, we define the mean distance from an ellipse to the set of points X as

$$d(X, e_{\Theta}(\cdot)) = \frac{1}{N} \sum_{i=1, \dots, N} d((x_i, y_i), e_{\Theta}(\cdot)) \quad (9)$$

Next, we minimize the distance from X to the ellipse with respect to the parameters of the ellipse Θ . We propose a gradient descent method to perform the minimization procedure. We start with an initial approximation for the ellipse parameters which could

be provided by the Hough transform or another simpler method. Then, we iterate the scheme:

$$\Theta^{k+1} = \Theta^k - \lambda \nabla_{\Theta} d(X, e_{\Theta^k}(\cdot)) \quad (10)$$

The step λ is adapted in each iteration using the following criterion: We initialize $\lambda = \lambda_0$, we compute Θ^{k+1} from Θ^k using (10). If $d(X, e_{\Theta^{k+1}}(\cdot)) < d(X, e_{\Theta^k}(\cdot))$, we update $\lambda = 10\lambda$. If $d(X, e_{\Theta^{k+1}}(\cdot)) > d(X, e_{\Theta^k}(\cdot))$, we update $\lambda = 0.1\lambda$ and we recompute Θ^{k+1} . We stop the iterative scheme when $\Theta^{k+1} - \Theta^k$ is small enough, or when we cannot reduce $d(X, e_{\Theta^k}(\cdot))$ after a number of attempts. If the nodule has not ellipsoid shape, we must determine if it has two or three gentle lobulations. We proceed as in the previous case, but instead of searching for an ellipsoid which fits the whole contour, we search for large lobulations which fit a large region of the contour each, which together cover the whole contour, whose axes do not differ strongly in their lengths, and whose centers are clearly separated (see Fig.5).

To analyze whether the nodule presents any of both kinds of ramifications, we locate those points whose distance to the ellipsoid which best fits the nodule is large. Since the ellipsoid covers the most representative part of the nodule, when it has ramifications, they appear as branches projecting out from this ellipsoid. The number and dimensions of these ramifications is strongly considered to characterize the nodule.

In order to prevent the ramifications from influencing the main ellipse's position and dimensions, an iterative scheme has been used in which the points which are very far from the ellipse are eliminated and the ellipse is recalculated. After a few iterations, the ellipse which is extracted is much more representative of the central region and the ramifications are more clearly identified (see Fig. 6).

We can locate angular margins by searching for pseudo-straight lines on the contour of the nodule. In order to adjust a set of N points to a straight line, we proceed as follows:

Let (u, v) be a point, and let $r = x + ay + c$ be a line, the distance from (u, v) to $x + ay + c$ is:

$$d(m, r) = \frac{|u + av + b|}{\sqrt{1 + a^2}} \quad (11)$$

If we minimize with respect to a and b , we extract the values which best extract the supporting line for a set of planar points.

When a point joins two straight lines with different orientations, an angle is signaled. In order to select those angles which are significant for the study, the lines must fit the contour, i.e. the mean distance from the points on the contour to the line must be low, they must be long enough in order to neglect small variations in the contour's orientation, and the difference in the orientation must be less than $\pi/2$, i.e. they form an acute angle (see Fig. 7). For every candidate to be a pseudo-corner, they line which adjust the points before it in the contour and the line which adjust the points after it, considering the contour as a sequence, must be extracted and compared.

Besides angular margins, we may find microlobulations. For this feature, a similar technique as the one used when extracting the general ellipsoid is applied. However, in this case, we search for segments of the contour which can be approximated by a part of an ellipsoid and whose size and axes are small. In order to avoid the duplication of the lobulations by extracting overlapping ellipsoids, only those whose centers are clearly different are selected (see Fig. 8).

To locate the areas where spiculation appears, we extract the orientation of a window around every contour point. As done with the calcifications, the structure tensor method has been used, though, in this case, not only the magnitude of the gradient, but also its orientation is significant. When the orientation in that region is quite dissimilar to the one of the contour in that point, it is marked as a spiculation zone (see Fig. 9). Non-spiculated nodules present a clearly defined contour along which the orientation varies softly. Similarly as done when we calculated the magnitude of the gradient for the localization of the calcifications, we now extract the magnitude and the orientation. We have already described how we can extract the magnitude. The orientation can be estimated with the eigenvector associated to the lowest eigenvalue of the central matrix. For that purpose, we apply Jaccobi's method to extract the eigenvalues and eigenvectors of the following matrix:

$$\begin{pmatrix} \sum_{i=0}^N y_i^2 & -\sum_{i=0}^N x_i y_i \\ -\sum_{i=0}^N x_i y_i & \sum_{i=0}^N x_i^2 \end{pmatrix} \quad (12)$$

4 Conclusion

In this work, we have presented a new approach in the field of breast cancer early detection. We have dealt with a very difficult type of images, those generated by ultrasound devices, which present a dense and intense kind of noise and whose interpretation is quite difficult, even for a human observer.

The definition of a series of objective and reliable criteria for the discrimination between benign and malignant solid breast nodules and their relation to visual characteristics of a region have made it possible to apply those techniques used in shape and texture analysis to the description of those nodules.

A wide range of techniques, such as truncated median, seed segmentation, shape adjustment, ellipse location or structure tensor, have allowed reaching very satisfactory results in terms of quantitative and qualitative description of the criteria that the physicians take into account when diagnosing a nodule.

The possibility of providing objective numerical measurements of the presence of the diagnosis criteria represents a remarkable progress in the classification of the nodules.

These results emphasize the importance of the introduction of computer vision techniques in the processing of medical images, not to substitute the role of the physicians, whose skills can only be imitated in a hard and limited way, but to aid the decision making process.

References

- [1] Kuo, W.J., Chang, R.F., Moon, W.K., Lee, C.C., Chen, D.R.: Computer-aided Diagnosis of Breast Tumors with Different US Systems. *Acad Radiol* 9 (2002) 793-799
- [2] Kaufhold, J., Chan, R., Karl, W.C., Castanon, D.A.: Ultrasound Tissue Analysis and Characterization. *Battlefield Biomedical Technologies*, H.H. Pien editor, Proc. SPIE V 3712 (1999)

- [3] Chen D.R., Chang, R.F., Juang, Y.L.: Computer-aided Diagnosis Applied to US of Solid Breast Nodules by Using Neural Networks. *Radiology* 213 (1999) 407-412
- [4] Stavros, A.T., Thickman, D., Rapp, C.L., Dennis, M.A., Parker, S.H., Sisney, G.A.: Solid breast nodules: use of sonography to distinguish between benign and malignant nodules. *Radiology* **196** (1995) 123-134
- [5] Sonka, M., Hlavac, V., Boyle, R.: *Image processing, analysis, and machine vision*. PWS-ITP (1999)
- [6] Nixon, M.S., Aguado, A.S.: *Feature extraction and image processing*. Oxford Newnes (2002)
- [7] Ahn, S.J., Rauh, W., Warnecke, H.J. Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola. *Pattern Recognition*, Elsevier **34** (2001) 2283-2303
- [8] Halíř, R., Flusser, J. Numerically stable direct least squares fitting of ellipses. In *Proc. of WSCG 98* (1998).