Applied Mathematical Sciences, Vol. 5, 2011, no. 8, 367 - 379

Masses Detection Using SVM Classifier Based on Textures Analysis

Fatima Eddaoudi

LIMIARF, Faculty of sciences, Rabat, Mohammed the fifth University, Morocco
eddaoudi_fatima@yahoo.fr

Fakhita Regragui

LIMIARF, Faculty of sciences, Rabat, Mohammed the fifth University, Morocco
regragui@fsr.ac.ma

Abdelhak Mahmoudi

LIMIARF, Faculty of sciences, Rabat, Mohammed the fifth University, Morocco
abdel_mahm@yahoo.fr

Najib Lamouri

LIMIARF, Faculty of sciences, Rabat, Mohammed the fifth University, Morocco
nlamouri2002@yahoo.fr

Abstract

The segmentation of mammograms plays a major role in isolating areas which can be subject to tumors. The identification of these zones is generally done in three stages: pectoral muscle segmentation, hard density zone detection and texture analysis of regions of interest. In this work, we focus on masses detection using SVM classification and textures analysis. As for the first stage, pectoral muscle we use an approach based on contour detection using snakes with an automatic initialization. For the second stage, we use an approach based on maxima thresholding. The region of interesting segmented are classified to normal and abnormal tissue using Haralick features calculated from the cooccurrence matrix. The test of these methods on mammograms of MIAS databases showed better performance in detecting masses compared to the methods proposed in the literature.

Keywords: Cooccurrence matrix, Mammography, Maxima thresholding, Pectoral muscle , Texture analysis

1. Introduction

Breast cancer is the most common form of cancer in the female population, especially in developed and under developed countries. The World Health Organization's International Agency for Research on Cancer in Lyon, France estimates that more than 150,000 women worldwide die of breast cancer each year [22,11]. According to the official statistics of the National Institute of the Oncology Sidi Mohammed Ben Abdellah in Rabat, 16% of women in Morocco develop breast cancer.

As with any form of cancer, early detection and diagnosis of breast cancer is one of the most important factors in recovery from the disease [4]. Currently mammography is the dominant method for detecting breast cancer [7,10], in particular mammography screening programs assisted by computers. But it is still far from being perfect. The large variability in the appearance of breast masses, added to the significant overlap in the appearance of malignant and benign masses, and the fact that abnormalities are often occluded or hidden in dense breast tissue make detection and diagnosis very difficult [25] even by radiologists [13].

Over the past decade and a half, several researchers have studied and proposed methods for computer-aided diagnosis (CAD) of abnormalities related to breast cancer in mammograms. The objective of CAD systems is to help radiologists make a recommendation for patient management. In this article, we propose a new method for mammograms segmentation and classification. The algorithms used are based on already developed methods but we introduce new elements. We segment the mammogram in three areas: the pectoral muscle, fibroglandular tissue and adipose tissue, then we analyze the dense tissues and we classify them in normal tissue and pathological ones.

1.1. Muscle pectoral segmentation

The pectoral muscle represents a predominant density region in the most medio-lateral oblique (MLO) views of mammogram; x-rays cross it less easily [1] and it appears more brightly in the mammogram. The border between the fibroglandular region and the pectoral muscle is fuzzy and irregular because of noise and surrounding breast tissue.

Previous work on automatic detection of abnormalities in breast tissue from mammograms shows that the feature extraction process may be affected by the presence of the pectoral muscle. Gupta and Undrill [19] indicate that mammographic parenchyma and the pectoral muscle may have similar textural characteristics, causing a high number of false positives when detecting suspicious masses. Also, the area overlying the pectoral muscle is a common area for cancers to develop and is carefully checked by radiologists to reduce false negatives. For this and other reasons, it is important to identify and segment out the pectoral muscle.

Several works were devoted to the segmentation of the pectoral muscle. Ferrari and Rangayyan directly tested and segmented pectoral muscle by using Gabor filters [20]. In [2], Oliver Malagelada used a region growing algorithm to locate the muscle and extract it from the breast. Suckling et al. [9] used multiple linked self organizing neural networks to segment the breast into four components: background, pectoral muscle, fibroglandular tissue and adipose tissue. This method had the advantage of simultaneously identifying the background and pectoral muscle. F. Ma, M.Bajger [8] used two image segmentation methods based on graph theory in conjunction with active contours to segment the pectoral muscle in screening mammograms. One method is based on adaptive pyramids (AP) and the other is based on minimum spanning trees MST. Weidong Xu, Shunren Xia [24] used a new algorithm using optimal thresholding and the Hough transform to suppress the pectoral muscle.

1.2. Textures analysis of mammograms

Many techniques are presented in literature detailing mammogram texture analysis, but in the interest of space, we restrict our discussion to the ones that are closely related to our research. The cooccurrence matrix is the most common method to analyze mammographic image textures. Ribeiro and al. use texture [14] features and association rules to classify mammograms. In the same way, Sumeet Dua [6] used a unique weighted association rule based classifier; images are pre-processed to reveal regions of interest. Texture components are extracted from segmented parts of the image and discretized for rule discovery. Association rules are derived between various texture components extracted from segments of images and employed for classification based on their intra- and inter-class dependencies. These rules are then employed for the classification of a commonly used mammography dataset.

In this work, we focus on these aspects using a segmentation of region of interests. As for the first stage, pectoral muscle we use an approach based on contour detection using snakes with an automatic initialization. For the second stage we use an approach based on maxima thresholding to detect dense tissues witch are going to be our regions of interesting. They are analyzed using texture features released from the cooccurrence matrix and then, classified to normal and abnormal cases using SVM. The test of these methods was performed on mammograms of MIAS databases [5]. To test our approach we used 95 mammographic images : calcification (20 cases); circumscribed masses (22 cases); speculated masses (19 cases); ill-defined (15 cases) and 19 normal mammograms. The size of all the images is 1024 pixels x 1024 pixels.

The rest of this article is divided in two sections. The first one describes the proposed method for mammogram segmentation. This section presents first the new method for pectoral muscle segmentation based on the method developed by S.M. Kwok and R. Chandrasekhar, followed by a description of the segmentation of dense tissues regions. Details of the textures analysis of dense tissues already located in order to classify them in normal and pathological cases. Results and conclusion follow in the last section.

2. Proposed method

To detect masses our method consists first in segmenting the pectoral muscle by applying active contours and second localizing dense tissues by using the maxima method. The textures of the located zones are analyzed through the cooccurrence matrix and the haralick features to classify them in normal or abnormal tissues using the SVM. Figure. 1 below describes these steps.

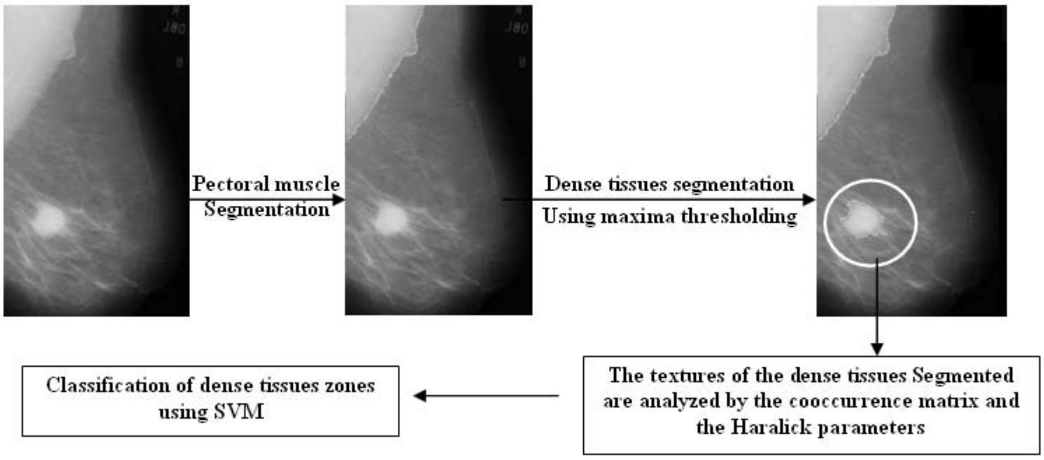


Figure 1 : The proposed method for mammogram masses detection

2.1. Pectoral muscle detection

The active contours method is one of the better performing methods for detecting the pectoral muscle. Its principal disadvantage is the need for an initialization close to the required result. Thus, it will be necessary to work out an automatic initialization in addition to one evolution of the contour algorithm which is also fast and effective.

An important result in segmentation of the pectoral muscle was found by S.M. Kwok and R. Chandrasekhar [21]. The pectoral edge is first estimated by a straight line which is validated for correctness of location and orientation. This estimate is then refined using iterative "cliff detection" to delineate the pectoral margin more accurately. This method performs well, especially in the estimation of the straight line which, in almost all the cases, is very close to the position of the pectoral muscle contour. We saw that it will be important to use a part of this method to find a good initial contour for our active contours step.

2.1.1 S.M. Kwok and R. Chandrasekhar method

The S.M. Kwok and R. Chandrasekhar method is done in two stages:

- Estimate the straight line which represents the pectoral muscle
- Modify that detected line so that it follows the curve of the muscle.

We adopt the following notations:

- R1 : Initial region of interest (ROI) = ¼ of Image

- AB : the approximation of muscle pectoral border

Straight line estimation in ROI

To start, one eliminates all the gray levels less than 15% of the maximum intensity (I_{max}) in the initial region of interest R1, then one determines a threshold **t**, as being the median value of pixels coming from the area of interest.

After calculating the median values of the areas representing the background and the pectoral muscle, respectively denoted by μ_b and μ_o, one recomposes the value of t_{cal}. If the t_{cal} is equal to **t**, the thresholding is finished; if not the ROI is segmented using t_{cal} and one recomposes μ_b and μ_o until **t** and t_{cal} are very close. The diagram in Figure 2 describes the method in more detail.

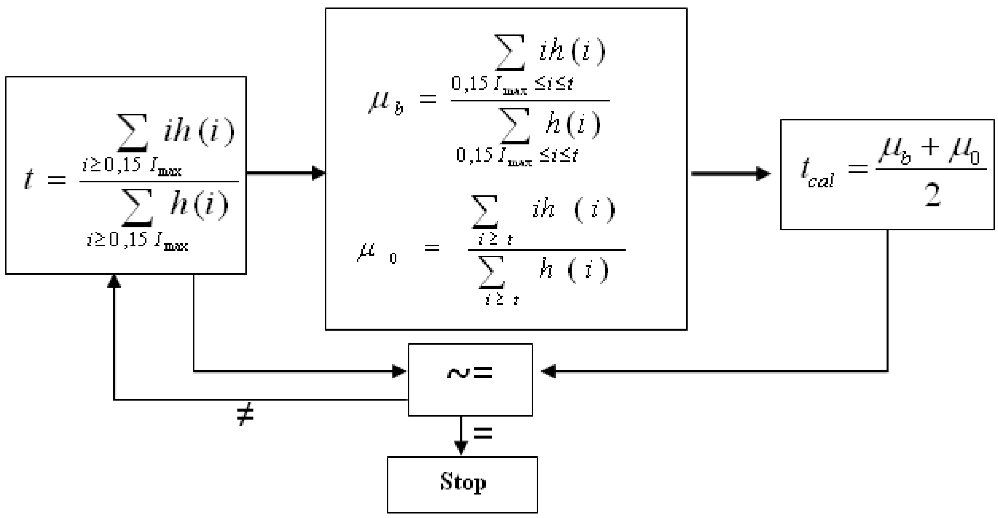


Figure 2 : Threshold Calculation Algorithm

Pixel selection and AB line validation

The calculated contour representing the pectoral muscle is deviated on the right to form a concave line. To avoid this problem, one applies a gradient test to eliminate the concave parts. One uses a sliding window of width 20 mm and length equal to that of the ROI. One progressively slides the window and one calculates the gradient of the segment. If the gradient is positive, one eliminates the part and one forms a new discontinuous contour. One only uses the continuous parts to trace the straight line representing the pectoral muscle.

For this straight line to be valid, it is necessary that the intersection between AB and the edges of image be in the breast area. If AB is not valid, one shrinks the ROI.

The new ROI R2 is defined so that line AB is the diagonal of this zone. In this ROI, we define the line AB in the same way as described above. If it is always valid, one concludes that found line AB is the best approximation of the contour of the pectoral muscle. Figure 3 shows some calculated contours.

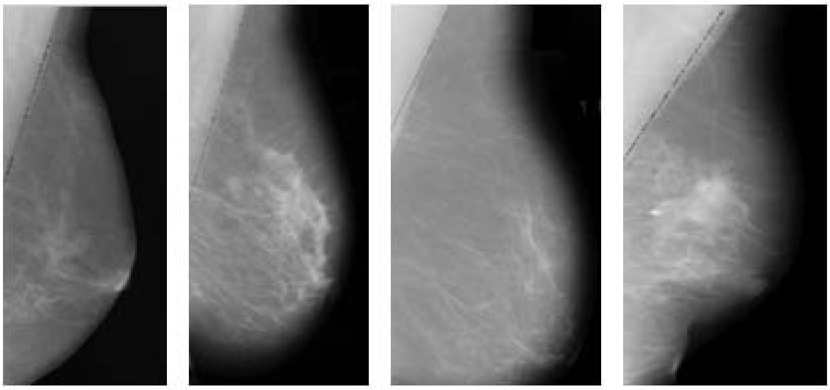


Figure 3: Estimate of the contour of the pectoral muscle by a straight line

Cliff detection

The method suggested by S.M.Kwok, R.Chandrasekhar for the detection of the curve of AB is done in four stages:

- Defining the search path: For each pixel of AB, we define a search path, denoted by S_i of length 2d and perpendicular to AB.
- Extracting intensity profiles: Along the search path, we extract the intensity profile.
- Determining cliff location.
- Smoothing of the determined curve.

2.1.2. Proposed method

Figure 4 presents our approach. As mentioned above, we use the straight line from S.M. Kwok, R. Chandrasekhar method as a good initialization for the active contours. Figure 5 illustrates pectoral muscle segmentation using our approach and cliff's method. To evaluate the performance of our method, on all the 95 mammograms we manually marked the contour of the pectoral muscle. On all these mammograms we applied our method and that of SM Kwok, R. Chandrasekhar. If the segmentation results are much closed to the marked borders we accept it then we count the good segmentation for the two methods. Table 1 shows the results found. Even though the results are comparable, our method is more efficient in terms of computation time saving.

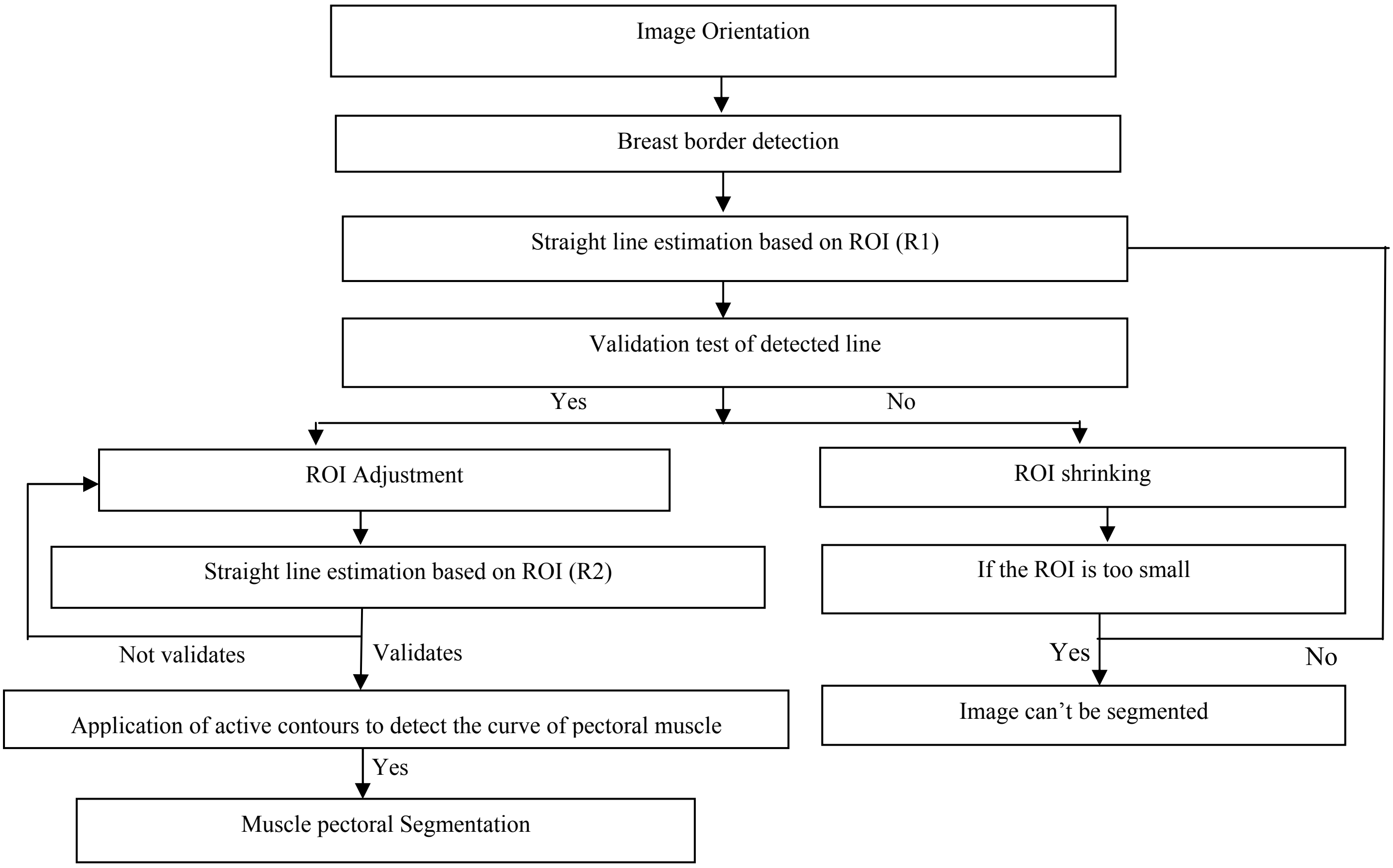


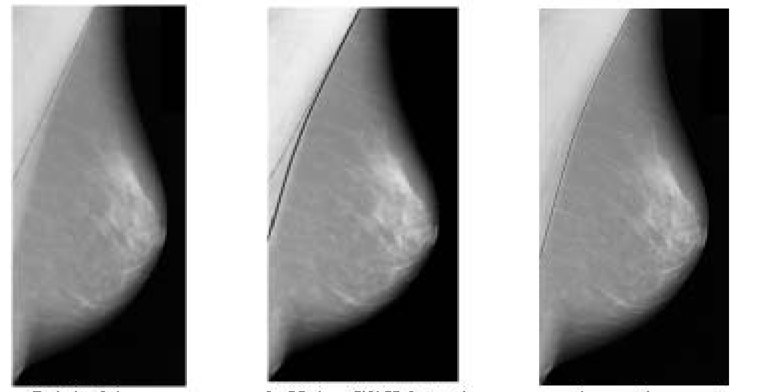
Figure 4: Algorithm of pectoral muscle segmentation

	Number of cases	Number of good segmentation
Muscle pectoral segmentation using S.M. Kwok, R. Chandrasekhar method	95	89
Muscle pectoral segmentation using active contours	95	93

Tableau 1 : Results of muscle pectoral segmentation using active contours method and R. Chandrasekhar method

2.2. Dense tissues zone segmentation

After eliminating the pectoral muscle, we limit our zone of interest to the fibroglandular region. The following stage is the localization of the zones that may contain tumors. Any mammographic image segmentation is based on the principle that the pixels inside pathological tissue have characteristics different from the pixels of normal tissue. These characteristics can include intensity, texture, morphological parameters like form, size, etc.



a- Original image b-Using Cliff detection c- using active contours

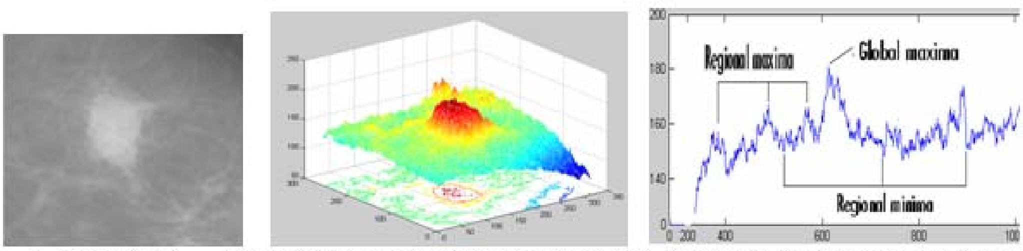
Figure5: Detection of pectoral muscle curve using S.M. Kwok, R. Chandrasekhar method and active contours

A bibliographical study made by A. Oliver [2] presents several works on mammograms segmentation. We note that the majority of the segmentation techniques are based on the intensity and histogram analysis. Displeased, radiologists in their reading of the mammograms use the gray levels as important indicator of pathology [21]. For these reasons, we decided to take the density expressed as a gray level value as the first characteristic for mammographic image segmentation.

2.2.1. Fibroglandular region Segmentation using the maxima method

- <u>Method principle</u>

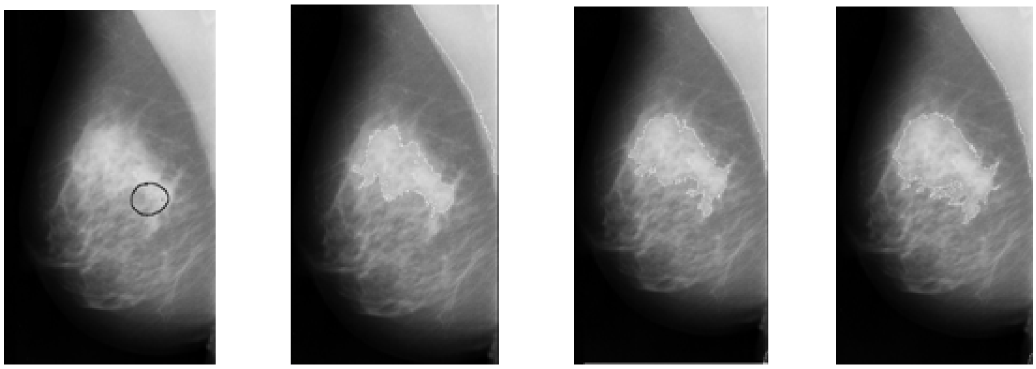
A gray level image can be presented in three dimensions, in Figure 6, the x and y axes represent the spatial position of the pixel while the z axis represents its intensity. In this representation the gray level intensity is similar to the level in topographic charts [18]. An image can have several regional maxima but only one global maximum. The global maxima can be used as a marker for the detection of objects. The results of the segmentation can be improved by thresholding this maximum. To test the importance of segmentation by the maxima method, we tested it on the whole databases. Theses results showed that the choice of the threshold depends on the density of breast tissue. Simulation results showed that for a threshold below 50, nothing is detected while a threshold greater than 60 may lead to satisfactory results for dense breast.



a - Original image b- Representation of an image in three dimensions c- Global and regional maxima and minima

Figure 6 : Maxima method

Best results are obtained with a threshold equal to 60; Figure 7 shows some results. Consequently, the segmentation performed on the entire set of mammograms permits us to limit the zone of interest to 30% of the total analyzed area. Subsequently, we used a threshold equal to 60 to segment the mammograms. Only regions detected by this threshold will be analyzed and classified into normal or pathological cases. The performance of this segmentation will indirectly be evaluated by estimating its effect on the detection of lesions.



a- Original image* b- Threshold = 50 c- Threshold = 60 d – Threshold = 70

*The manual contour in the original image marks the pathology

Figure 7 Segmentation by maxima with various thresholds: 3 values of the threshold are used. The results showed the presence of the pathology within the zone of interest

2.3. Texture analysis

2.3.1. Cooccurrence matrix

Texture information constitutes one of the important parameters on which a radiologist bases his or her analysis of a mammogram. This characteristic makes it possible to differentiate normal tissue from pathological tissue [17]. The tissues seen in a mammogram have completely random, non-homogeneous structures; this is why statistical analysis is preferred [12]. Among the statistical methods most used for the characterization of textures, the cooccurrence matrix makes it possible to determine the frequency of appearance of two pixels separated by a distance d at angle θ from the horizontal.

This matrix contains a very large amount of information that is not easily to handle. Therefore, it is not used directly but through measurements known as indices of texture [23]. Haralick proposed 14 indices]. In this article we use: Average, variance, energy, contrast, correlation, normalized correlation, entropy, homogeneity, diagonal moment, shade cluster, and prominence cluster [16].

2.3.1.1. Choice of Window size

To calculate the cooccurrence matrix for a given pixel, we use a block of pixels surrounding the pixel. The size of this block or window of analysis must satisfy two criteria: it must be as small as possible to reduce the risk of mixing different textures and at same time as large as possible to extract robust and significant statistics [15,2].

To study the effect of the size of the block of analysis on the quality of the segmentation, we considered several sizes. We noticed that even if the size of the block is increased, the number of the gray level contained in the block does not change much, whereas the computing time increases enormously. For this reason, we will limit ourselves to the following block sizes: 5x5, 7x7, 9x9, 11x11, 13x13.

2.3.1.2 SVM classification based on Haralick vector

Textural information is extracted from the Haralick vector calculated from the cooccurrence matrix. Haralick vector calculation is based on the choice of the analysis window size, the displacement size (d) and the direction according to which the two selected pixels are compared.

It is well known that classification based on SVM strongly depend on the training phase and the values of displacement orientation and the window size used to calculate the training vectors [3,25]. Several studies were undertaken to determine a distance or an optimal orientation [2]. In these considerations, we study the effect of these parameters on classification performance using a large number of combinations of parameters applied to our mammograms. Table 2 presents all the combinations of parameters used for classification.

Size of analysis block	Displacement (d)	Orientations
5x5	1,2	0,45°,90°,135
7x7	1, 2, 3	0,45°,90°,135
9x9	1, 2, 3, 4	0,45°,90°,135
11x11	1, 2, 3, 4, 5	0,45°,90°,135
13x13	1, 2, 3, 4, 5, 6	0,45°,90°,135

Table 2: Combination of all parameters used for training and classification by SVM

It is worth mentioning that the use of gray levels information often used by radiologists is not considered in the classification based on texture analysis. In this work, we combined gray level information and Haralick parameters which improve significantly the mammogram classification.

3. Results and discussion

We tested the proposed method on the original non segmented images of the entire set of mammograms of the database. Considering the different parameters combinations, we drew the following conclusions:

- with all the combinations of parameters using a size block of 11x11 or 13x13 during training, no detection of the masses is possible;
- The use of diagonal displacement (following 45° orientation) gives bad classification results.
- The block sizes of 11x11 and 13x13 don’t give good classification results.

Block size	7x7	9x9	9x9	7x7	7x7	7x7
Displacement	3	4	4	3	3	3
Orientation	90°	0° ;90°	90°	0°,90°	0°,90°	0°,90°

Table 3: Combinations of parameters which give acceptable classification results

- Successful combinations are presented in Table 3.
- Best results were obtained by the following combinations: 9x9 block with orientation 90° and displacement equal to 4.

Using the last combination of parameters, we tested our classification method using 95 mammograms. The results, obtained with original mammograms, showed that 65 malignant out of 76 were classified true positive while 13 mammograms out of 19 were classified true negatives which correspond to a classification rate of 77% in average. These rates were significantly improved, achieving 95% in average, when we applied the classification method on pre-segmented mammograms by the maxima thresholding (see Table 4).

			True positives	True negatives	False positives	False negatives
Classification performance using original mammograms	Number of cases	95	65/76	13/19	11/76	6/19
	Pourcentage %	100	85.52	68.42	14.47	31.57
Classification performance using pre-segmented mammograms	Number of cases	95	73/76	18/19	3/76	1/19
	pourcentage	100	96.05	94.73	3.94	5.26

Table 4 : Classification results using SVM

4. Conclusions

In this article, we used a series of methods to isolate the zones susceptible to tumors. The algorithm developed by S.M. Kwok and R. Chandrasekhar to estimate the contour of the pectoral muscle by a straight line constituted a good starting point for our method based on active contours. The mixing of the two approaches gave very satisfactory results.

We chose the cooccurrence matrix to analyze the breast area. Not wishing to neglect gray level information, we pre-segmented the mammogram by the maxima method. The two methods are complementary and gave good performances.

The last part of this article was devoted to mammograms classification by SVM based on Haralick parameters. Like all the supervised classification methods, the results strongly depend on the training phase and the classification parameters. In our case, the window size, the displacement and the orientation used for the calculation of cooccurrence matrix significantly affect the results. We tested all possible combinations of parameters and concluded that a window size of 9x9 according to a 90° orientation with a step equal to 4 gave the best results. Using this combination of parameters, we tested our classification method using 95 mammograms; we find a classification rate equal to 77% in average. These rates were significantly improved, achieving 95% in average, when we applied the classification method on pre-segmented mammograms by the maxima thresholding.

References

[1] A.Boucher, P.E.Jouve, F.Cloppet, N.Vincent, Segmentation du muscle pectoral sur une mammographie , Congrès des jeunes chercheurs en vision par ordinateur, ORASIS'09, Trégastel, France, (2009)
[2] A.Oliver Malagelada, Automatic Mass Segmentation in Mammographic Images, PHD Thesis, Universiy of Girona, Girona, April, (2007)
[3] A.Rojas Domínguez, N.AsokeK, Toward breast cancer diagnosis based on automated segmentation of masses in mammograms", Pattern Recognition, V42, (2009), 1138-1148
[4] A.Strauss, A.Sebbar, S.Désarnaud, P.Mouillard, M.Le Gal, Cancer du sein : des progrès en détection automatique des microcalcifications, Nouvelles Méthodes de Traitement de l'Information en Médecine, Paris, Springer-Verlag, France, (1992), Volume 5
[5] Available at : http://peipa.essex.ac.uk/info/mias.html
[6] D.Sumeet, S.Harpreet, H.W.Thompson, Associative classification of mammograms using weighted rules, Expert Systems with Applications, ELSEVIER, (2009), 36, 9250–9259
[7] E.E.Júlia de Oliveira, T M.Deserno, A.Araújo, Breast Lesions Classification applied to a reference database, 2nd International Conference: E-Medical Systems, October 29-31, (2008), Tunisia
[8] F. Ma, et al, Two graph theory based methods for identifying the pectoral muscle in mammograms, Pattern Recognition, (2007), doi: 10.1016/j.patcog.2006.12.011
[9] J.Suckling, D.Dance, E.Moskovic, D.Lewis, S.Blacker, Segmentation of mammograms using multiple linked self-organizing neural networks. Med Phys, (1995);22:145–52
[10] K.Bentley, Mammography image quality: Analysis of evaluation criteria using pectoral muscle presentation, Radiography (2007), doi:10.1016/j.radi.2007.02.002
[11] K.Thangavel, M. Karnan, R.Siva Kumar, A.Kaja Mohideen, Automatic Detection of Microcalcificatio in Mammograms–A Review. International Journal on Graphics Vision and Image Processing. 5, (2005)
[12] L.Sheng, The analysis of digital mammograms: speculated tumor detection and normal mammogram characterisation, PhD thesis, Purdue University, (1999)
[13] M.G.Linguraru, J.M.Brady, M.Yam, Filtering hint Images for the Detection of Microcalcifications, Proceedings of the 4th International Conference on Medical Imagine Comuting and Computer-Assisted Intervention, Lecture Notes in Computer Science 2208, Springer-Verlag Berlin Heidelberg, (2001), 629-636
[14] M.X.Ribeiro, A.J.M.Traina, A.G.R Balan, C.Jr.Traina: A framework to support mammogram diagnosis. IEEE CBMS, Slovenia: Maribor, (2007), 47–52

[15] N.Idrissi, La navigation dans les bases d'images : prise en compte des attributs de textures, PHD thesis, October, Faculty of Siences, FSR, Rabat Maroc,(2008)
[16] O.D'hondt, Analyse spatiale de texture non stationnaire dans les images SAR, PHD Thesis, University de Renne I, Paris, (2006)
[17] P.KESTENER; Analyse multifractale 2D et 3D à l'aide de la transformation en ondelettes: Application en mammographie et en turbulence développée, PhD thesis; Université BORDEAUX I ; Ecole Doctorale des sciences physiques et de l'ingénieur, (2003)
[18] P.Soille, Morphological Image Analysis: Principles and Applications, Springer-Verlag, (1999), 170-171
[19] R.Gupta, P.E.Undril, The use of texture analysis to delineate suspicious masses in mammography, Phys.Med.biol.40,(1995), pp.835-855
[20] R.J.Ferrari, R.M.Rangayyan, J.E.L.Desautels, R.A.Borges, A.F.Frère, Automatic indentification on the pectoral muscle in mammogramms. IEEE Transaction on Medical Imaging, (2004), 23(2):232-245
[21] S.M. Kwok, R. Chandrasekhar, Y. Attikiouzel, M.T. Rickard. Automatic pectoral muscle segmentation on mediolateral oblique view mammograms. IEEE transaction on Meical Imaging, (2004), 23(9) :1129-1140
[22] S.Suri.Jasjit, R.M.M.Rangayyan, Recent Advances In Breast Imaging, Mammography and Computer-Aided Diagnosis of Breast Cancer, Spie Press, Belligham, Washington USA, (2006) ISBN: 9780819460813
[23] T.Iftenet, A;Safia, Comparison Between Cooccurrence Matrix And Wavelet Transform For The Textural Classification Of The Spot's HRV (XS) Images, Télédétection,vol.4, n°1, Contemporary Publishing International, 2004, 39-49
[24] Xu.Weidong, Xia.Shunren, D.Huilong, A Novel Computer-Aided Diagnosis System of the Mammograms, Intelligent Computing in Signal Processing and Pattern Recognition, (2006), 345, 639-644
[25] W.Liyang, Y.Yongyig, M.N.Robert, Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis", Pattern Recognition, V42, (2009), 1126-1132

Received: August, 2010