Active Region Segmentation of Mammographic Masses Based on Texture, Contour and Shape Features

Joan Martí, Jordi Freixenet, Xavier Muñoz, and Arnau Oliver

Institute of Informatics and Applications, University of Girona
Campus de Montilivi s/n 17071 Girona, Spain
{joanm,jordif,xmunoz,aoliver}@eia.udg.es

Abstract. In this paper we propose a supervised method for the segmentation of masses in mammographic images. The algorithm starts with a selected pixel inside the mass, which has been manually selected by an expert radiologist. Based on the active region approach, an energy function is defined which integrates texture, contour and shape information. Then, pixels are aggregated or eliminated to the region by optimizing this function allowing to obtain an accurate segmentation. Moreover, a texture feature selection process, performed before the segmentation, ensures a reliable subset of features. Experimental results prove the validity of the proposed method.

1 Introduction

Breast cancer is considered a major health problem in western countries, and indeed it constitutes the most common cancer among women. Recent studies [1] show that in the European Community, for example, breast cancer represents 19% of cancer deaths and fully 24% of all cancer cases. In absolute terms, this data means that approximately 10% of women will develop breast cancer during the course of their lives. Mammographic screening is the main method to identify early breast cancer, because it allow identification of tumour when it is not yet palpable. However, of all lesions previously diagnosed as suspicion and sent to biopsy, approximately 25% were confirmed malignant lesions, and approximately 75% were diagnosed benign lesions. This high false-positive rate is related with the difficulty in obtaining an accurate diagnosis [2]. In this sense, computerized image analysis are going to play an important role in improving the issued diagnosis. The effort in such computerized schemes have been carried out for detection of two major signs of malignancy, named clustered microcalcifications and masses.

In this sense, in a previous work we proposed the use of selected shape-based features in order to classify clustered microcalcifications between benign and malignant [3]. The computerized analysis of microcalcifications was divided into four steps: 1) digitization of mammograms and enhancement of images; 2) detection and localization of suspicious areas using a region growing segmentation

F.J. Perales et al. (Eds.): IbPRIA 2003, LNCS 2652, pp. 478–485, 2003.

algorithm based on Shen proposal [4]; 3) extraction of shape-based features for every segmented microcalcification; and 4) analysis of the features using Case-Based Reasoning techniques. More recently, we have studied how to characterise clusters of microcalcifications, due to its demonstrated relevance for issuing a diagnosis [5]. It has been observed, in a great number of malignant diagnosed mammograms, that the only indicator used to issue a diagnosis was the number of microcalcifications and their distribution inside every cluster.

On the other hand, masses also contain important signs of breast cancer and are hard to detect as they often occur in dense glandular tissue. In this sense, a number of researchers consider the computer analysis of masses to be more challenging as compared to that microcalcifications because masses are normally indistinguishable from the surrounding tissues [6].

Sahiner et al. [7] analysed and summarized a whole set of mass segmentation methods. Several works are focused on the use of texture and shape features. For instance, Kilday et al. [8] applied seven morphological features, while Petrosian et al. [9] investigated the usefulness of texture features based on spatial gray-level dependence matrices. Rangayyan et al. [10] used an adaptative method of edge profile acutance, and shape measures of compactness, Fourier descriptor, and moment based measure. On the other hand, Wu et al. [11] selected 14 features from a total of 43 for classification of malignant and benign masses and applied artificial neural network.

In this paper we propose a novel method for masses segmentation based on the principle of active region that takes into account texture, contours, and shape features. A set of 80 texture features are extracted and then selected according its homogeneity behaviour in order to choose an appropriated subset for the segmentation. Furthermore, an energy function is then defined which integrates all these sources of information. Then, the active region starts to grow optimizing this function in order to segment the mass region. The remainder of this paper is structured as follows: Section 2 describes the proposed segmentation approach detailing the selection of the texture features as well as the segmentation process based on the active region model. Experimental results proving the validity of our proposal appear in Section 3. Finally, conclusions are given in Section 4.

2 Proposed Segmentation Method

The method is grounded on the observations: 1) that masses in mammographic images have approximately uniform textures across their interiors, 2) the mass edges coincide with maxima in the magnitude of the gray level gradient, and 3) that changes in masse profile shape are small.

Taking these observations into account our proposal is based on the definition of an energy function, which integrates all these sources of information. Roughly speaking, the method starts from a connected set of pixels known to occupy the mass interior, which have been provided by the user (expert radiologist). Then, the algorithm is composed by two basic stages. Firstly, best texture features to segment the mass are selected; and secondly, the region grows by optimizing the energy function, which ensures the homogeneity inside the region, the presence of edges at its boundary, and the similarity of the region shape with those of previously-determined regions.

Furthermore, our proposal is based on the active region model, which has been recently introduced as a way to combine region and boundary information. This model is a considerable extension on the active contour model since it incorporates region-based information with the aim of finding a partition where the interior and the exterior of the region preserve the desired image properties. The underlying idea is that the region moves through the image (shrinking or expanding) in order to contain a single, whole region. The works of Chakraborty et al. [12], Hibbard [13] and Sahiner et al. [7] are good examples of active regions applied to the segmentation in medical images.

2.1 Texture Features Selection

One of major problems of texture segmentation is the selection of adequate texture features to model the homogeneity of regions, which are able to provide us the information required to perform the segmentation. In order to solve this difficulty, we use the knowledge provided by the user selecting an area of the image which is known of belonging to the region, and features which are homogeneous in this neighbourhood are selected.

Co-occurrence matrices proposed by Haralick et al. [14] are used in this work. Some of the most typical features, contrast, energy, entropy and homogeneity, are computed for distances from one to five and for $0°$,$45°$,$90°$ and $135°$ orientations, providing a set of 80 features. Then, the homogeneity of each feature inside the initial region is tested by measuring its match with a normal distribution using a skewness and kurtosis test, which many authors recommend by its simplicity [15]. Hence, a small subset of k texture features which present an homogeneous behaviour inside the region are selected for the next step of the segmentation process.

2.2 Active Region Segmentation

The combination of region, edge and shape information represents more accurately boundaries in medical images than either region growing or edge detection alone was compellingly argued by Chakraborty et al. [12]. With the aim of integrating all these kinds of information in an optimal segmentation, the global energy is defined with three basic terms. Region terms measures the homogeneity in the interior of the region by the probability that these pixels belong to the region. Boundary term measures the probability that boundary pixels are really edge pixels. And finally, shape term measures the similarity of the contour shape with those of previously determined cases.

Some complementary definitions are required: let $\rho(R)$ be a partition of the image into two non-overlapping regions, where R_0 is the region corresponding to the background region and R_1 corresponds to the mass region. Let ∂R_1 be the current region boundaries of the growing region R_1. The energy function is then defined as

$$E(\rho(R)) = \alpha \sum_{i=0}^{1} - \log P_{Region}(j : j\epsilon R_i|R_i) + \beta(- \log P_{Boundary}(j : j\epsilon\partial R_1)) \\ +(1 - \alpha - \beta)(- \log P_{Shape}(R_i)) \tag{1}$$

where α is a model parameter weighting the region term and β weights the boundary term. The influence of these parameters on the segmentation results will be analised in Section 3.

Region Information

The region term measures the homogeneity of the pixels into a texture region which is modelled by a multivariate Gaussian distribution. Hence, the probability of a pixel j characterized by the selected texture features $\vec{x_j}$ of belonging to the region R_1 is

$$P_{Region}(\vec{x_j}|R_1) = \frac{1}{\sqrt{(2\pi)^k|\Sigma_1|}} \exp\{-\frac{1}{2}(\vec{x_j} - \vec{\mu_1})^T \Sigma_1^{-1}(\vec{x_j} - \vec{\mu_1})\} \tag{2}$$

where $\vec{\mu_1}$ is the mean vector of the region and Σ_1 its covariance matrix. The background is treated as a single region having uniform probability distribution P_0.

Boundary Information

The second term in equation 1 depends on the coincidence of the region boundary with the image edges appearing as coherent features in the scalar gradient of the gray levels. Hence, we can consider $P_{Boundary}(j)$ as directly proportional to the value of the magnitude gradient of the pixel j.

Shape Information

Region shape is specified from Fourier descriptors, which use the Fourier transform over the points that define the contour of the region, where each point (x, y) is defined as a complex number $(x + jy)$. Thus, we have a sequence of complex numbers that represent the region contour. Nevertheless, the Fourier transform algorithm requires an input array whose length is an integral power of 2. So, if the number of points of the contour does not satisfy this condition, we need to follow the contour until the condition is true. Due to the periodic rate of the Fourier transform, this will have no effect on the result. There are other changes that must be applied to the Fourier descriptors to eliminate their dependence on position, size, orientation and starting point of the contour. A change in the position of the contour alters only the first descriptor, so we initialize this descriptor to null. A change of size only requires multiplying by a constant.

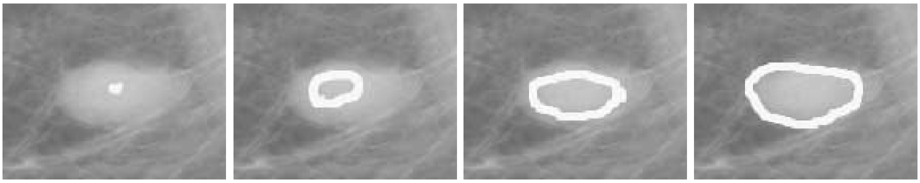


Fig. 1. Sequence of the region growth. The region starts to grow from the interior area selected by the user, competing for the image pixels in order the segment the whole mass

A rotation on the region requires multiplying each coordinate by $\exp(j\phi)$ where ϕ is the rotation angle.

$$NFD(k) = \begin{cases} 0 & k = 0 \\ \frac{A(k)}{A(1)} & k = 1, 2, ..., N/2 \\ \frac{A(k+N)}{A(1)} & k = -1, -2, ..., -N/2 + 1 \end{cases} \tag{3}$$

With these Fourier descriptors we achieve a set of points that characterizes the region. However, a solely value would be advisable. In this sense, we use the measure defined by Shen [4] that gives a single descriptor for each region.

$$FF = \frac{\sum_{k=\frac{-N}{2}+1}^{N/2} \frac{\| NFD(k) \|}{|k|}}{\sum_{k=\frac{-N}{2}+1}^{N/2} \| NFD(k) \|} \tag{4}$$

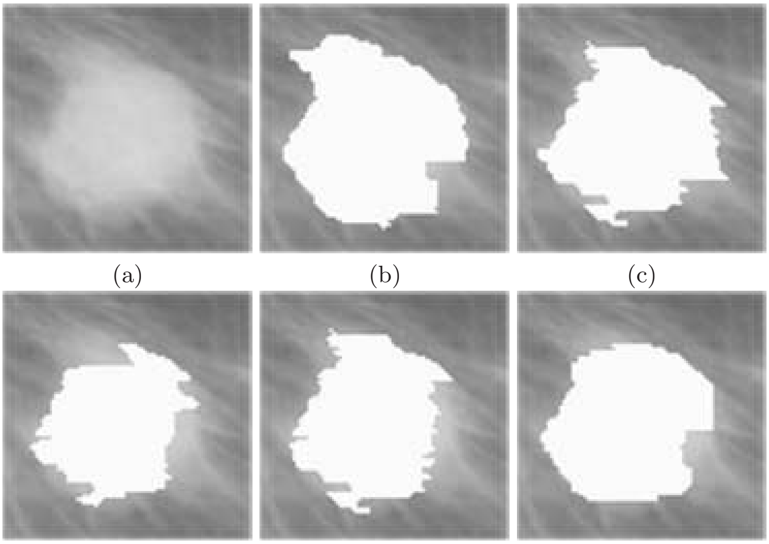
Hence, the match of the region shape and the model shape, which is related to previously determined cases, is given by the difference between their corresponding FF descriptors.

Optimization

The energy function is then optimized by a region competition algorithm [16] which takes the neighbouring pixels to the current region boundary ∂R_1 into account to determine the next movement. Specifically, the region aggregates a neighbouring pixel when this new classification decreases the energy of the segmentation. Intuitively, the region begin to move and grow, competing for the pixels of the image until an energy minimum is reached. A sequence of the region growth is shown in Figure 1.

3 Experimental Results

Twenty mammographic images including sixteen circumscribed-benign and four circumscribed-malignant were selected from the Mammographic Image Analysis Society (MIAS, UK) database. The spatial resolution of the image is 50μm x



(a) (b) (c) (d) (e) (f)

Fig. 2. Segmentations results using different weights (α) and (β) on the terms of the energy function. (a) Original mass, and segmentations obtained considering (b) texture, (c) contour, (d) texture and shape, (e) contour and shape, and (f) all three terms

50μm. The optical density is linear in the range $0 - 3.2$ and quantized to 8 bits. The boundary of each mass was traced by an expert radiologist specialized on mammography and were used as the basis for visual evaluation of the segmented results obtained by the proposed mass detection algorithm.

The influence of the three terms which are considered in the energy function (region, boundary and shape) was analised. Figure 2 shows segmentation results obtained with different weights of these terms. As is stated, all sources provide us useful information to perform the segmentation. However, best results have been achieved considering all three terms together.

An expert radiologist was the responsible to provide an initial placement inside the mass to segment. Some segmentation results obtained then by our proposal are shown in Figure 3. As is stated, the technique allows to correctly segment the masses and results have been considered very positive for radiologists.

4 Conclusions and Further Work

This paper has presented a segmentation method for the identification of masses in mammographic images. The technique is based on the integration of texture, contour and shape information in the segmentation process. Hence, an energy function which considers all these sources together has been defined. Then, the

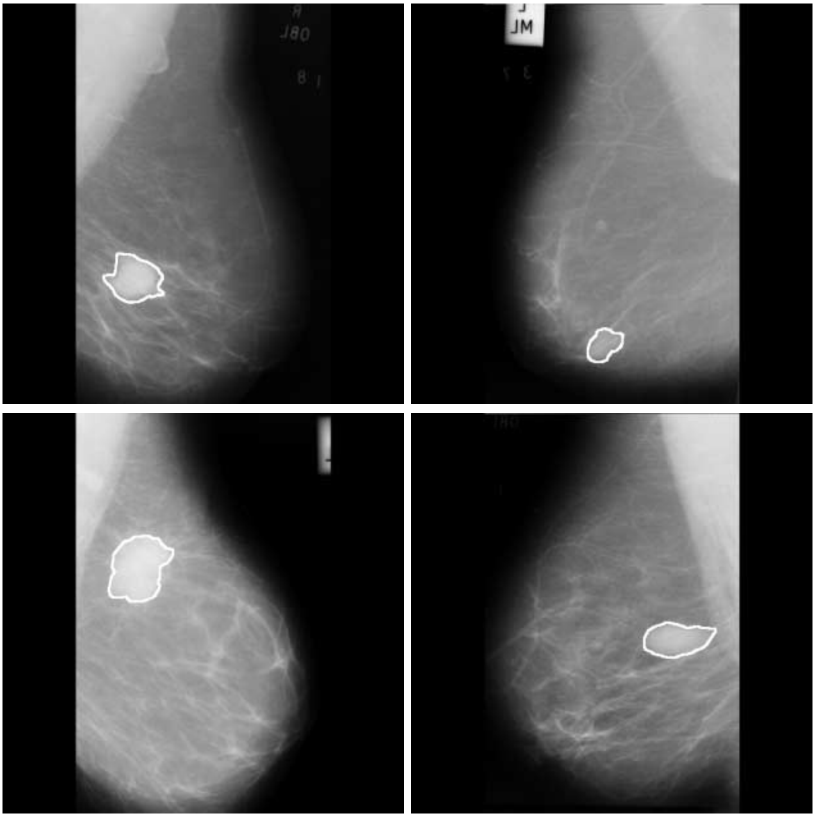


Fig. 3. Mass segmentation results

growing of the region by optimizing this function allows to obtain an accurate segmentation. Furthermore, the a-priori selection of the most adequate texture features to segment the mass has been described.

Experimental results over 20 images from the Mammographic Image Analysis Society database demonstrate the effectiveness of the proposed algorithm in estimating mass regions and their boundaries with high accuracy.

Further work is focused on two different directions. First, the inclusion of an automatic module of mass detection and seed placement in order to perform an unsupervised segmentation. And second, the extension of our proposal to deal with spicular masses.

References

[1] J. Esteve, A. Kricker, J. Ferlay, and D. Parkin. Facts and figures of cancer in the european community. Technical report, International Agency for Research on Cancer, Lyon, France, 1993. 478

[2] L. Basset and R. Gold. *Breast Cancer Detection: Mammograms and Other Methods in Breast Imaging*. Grune & Stratton, New York, 1987. 478

[3] J. Martí, J. Español, E. Golabardes, J. Freixenet, R. García, and M. Salamó. Classification of microcalcifications in digital mammograms using case-based reasoning. In *International Workshop on Digital Mammography*, pages 285–294, Toronto, Canada, June 2000. 478

[4] L. Shen, R. M. Rangayyan, and J. E. L. Desautels. Application of shape analysis to mammographic calcifications. *IEEE Transactions on Medical Imaging*, 13:263–274, June 1994. 479, 482

[5] J. Freixenet, P. Planiol, J. Martí, R. García, J. Batlle, and R. Bassaganyas. Cluster-shape characterization of microcalcifications. In *IEE Proceedings of Medical Applications of Signal Processing*, pages 5/1–5/5, London, UK, October 2002. 479

[6] H. Li, M. Kallergi, L. Clarke, V. Jain, and R. Clark. Markov random field for tumor detection in digital mammography. *IEEE Transactions on Medical Imaging*, 14(3):565–576, September 1995. 479

[7] B. Sahiner, N. Petrick, H. Chan, L. M. Hadjiiski, C. Paramagul, M. A. Helvie, and M. N. Gurcan. Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization. *IEEE Transactions on Medical Imaging*, 20(12):1275–1284, December 2001. 479, 480

[8] J. Kilday, F. Palmieri, and M. D. Fox. Classifying mammographic lesions using computer-aided image analysis. *IEEE Transactions on Medical Imaging*, 12:664–669, December 1993. 479

[9] A. Petrosian, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsit. Computer-aided diagnosis in mammography: classification of mass and normal tissue by texture analysis. *Physics in Medicine and Biology*, 39:2273–2288, 1994. 479

[10] R. M. Rangayyan, N. M. El-Faramawy, J. E. L. Desautels, and O. A. Alim. Discrimination between benign and malignant breast tumors using a region-based measure of edge profile acutance. In *International Workshop on Digital Mammography*, pages 213–218, 1996. 479

[11] Y. Wu, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, and C. E. Metz. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology*, 187(1):81–87, 1993. 479

[12] A. Chakraborty, L. H. Staib, and J. S. Duncan. Deformable boundary finding in medical images by integrating gradient and region information. *IEEE Transactions on Medical Imaging*, 15:859–870, 1996. 480

[13] L. S. Hibbard. Maximum a posteriori segmentation for medical visualization. In *IEEE Workshop on Biomedical Image Analysis*, pages 93–102, Santa Barbara, California, June 1998. 480

[14] R. M. Haralick, K. Shanmugan, and I. Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, November 1973. 480

[15] L. Wilkinson. Statistical methods in psychology journals. *American Psychologist*, 54(8):594–604, 1999. 480

[16] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, September 1996. 482