A Technique for the Automatic Detection and Classification of Mammographic Calcifications using the Watershed Method

ULISSES DE MENDONÇA BRAGA NETO
WINSTON ARAÚJO DE SIQUEIRA NETO

Laboratório de Reconhecimento de Padrões
Departamento de Eletrônica e Sistemas
Universidade Federal de Pernambuco
C.P. 7800 50711-970 Recife, PE
e-mail: `ulisses@npd.ufpe.br`

Abstract. In this work we propose the application of the watershed segmentation method in the automatic detection and classification of mammographic calcifications. Only two simple textbook shape factors, the compactness and the first invariant moment, are used to build a nearest-neighbour classifier. Using mammographic sections of biopsy-proven cases, we demonstrate the effectiveness of the proposed technique.

1 Introduction

Breast cancer is a subject of great concern in medicine, since it is a leading cause of death by cancer of women in middle age and older. Prevention of this disease is unfortunately very difficulty, because the causes of breast cancer are not yet fully understood. Of all the breast screening techniques, mammography has proven to be the only one which allows the diagnosis of breast cancer at its earliest stage, when it is still possible to treat the disease very effectively [2].

One of the major signs for the early diagnosis of breast cancer is the presence of malignant calcifications in mammograms. Automatic methods for the detection and classification of calcifications offer therefore significant help to radiologists in improving the efficiency of screening programs.

The segmentation of the calcifications is a difficult step in automatic methods, because mammographic images usually have poor contrast. Many different approaches to the problem of calcification segmentation can be found in the literature. For instance, Shen *et al.* [11, 12] employed a region growing technique, whereas Chan *et al.* [5] investigated a difference-image technique. Davies and Dance [6] tried a method based on a local-area thresholding process.

In this paper we propose the application of the watershed method in the segmentation of calcifications in mammograms. The watershed method is a powerful Mathematical Morphology tool for segmentation that does not involve any heuristics, unlike classic methods like, for example, region growing.

The criterion for differentiating the benign and malignant calcifications in the classification step is their shape. We make use of only two simple textbook shape factors, the compactness and the first invariant moment, to build a nearest-neighbour classifier. Using mammographic sections of biopsy-proven cases, we demonstrate the effectiveness of the proposed technique.

In the following sections, we first give a detailed explanation of the watershed method, and then describe the shape factors and the nearest-neighbour classifier used. Then we show the results obtained for biopsy-proven calcifications and close up by offering some remarks on the benefits and limitations of the technique proposed, comparing it also with a previous work on the field.

2 The Watershed Method

The watershed method is a powerful Mathematical Morphology tool for segmentation [3, 14]. This tool is useful not only for segmenting gray-level images (as we show in section 2.1 below), but also for binary segmentation, i. e., the problem of separating overlapping particles [4, 10].

2.1 The Watershed Transformation

In order to understand the watershed transformation, it is useful to think of an image as a *relief*, where the bright areas correspond to the elevations of the terrain, and the dark areas to the depressions. If one drop of water is released on any point of the relief, it will slide down until it gets to a regional minimum. The set of all points where the drop of water is bound to end up in a given regional min-

Anais do VIII SIBGRAPI (1995) 273-280

imum M is defined as the catchment basin $W(M)$ associated with the minimum M. The contours of the catchment basins form the watershed line, which is the output of the watershed transformation.

In a completely equivalent manner (and maybe easier to visualize) one may suppose that water is oozing and rising at equal speed from every regional minimum, starting from the lowest minimum and then from each of the others as soon as the waterlevel reaches its altitude. Dams are built in the places where water from different minima would merge. After the relief is completely flooded, the dams rising above the waterlevel constitute the watershed line, which is composed of closed contours that involve each of the regional minima and correspond to the crest lines of the relief. This "flooding interpretation" of the watershed transformation is illustrated in fig. 1.

We remark that the watershed transformation can be defined entirely in terms of Mathematical Morphology operations, however, the most efficient implementations use pixels queues and rely on the flooding interpretation itself [14].

2.2 Segmentation of Gray-level Images by the Watershed Method

The transition from one region to another in an image is marked by a sharp change in the grey-level. Now, the gradient operator has the property of detecting such changes: Transition regions become bright in the gradient, whereas relatively uniform regions become dark, so that the crest lines of the gradient correspond approximately to the contours of the regions in the original image. Hence, a procedure for segmenting the regions would be simply to apply the watershed transformation on the gradient of the image.

Unfortunately, in real-world images there is always a considerable amount of noise, which introduces a great number of spurious catchment basins in the determination of the watershed line, yielding an over-segmented image. An initial low-pass filtering of the image improves the result, but is not sufficient to solve the problem.

The most commom procedure to overcome the over-segmentation problem is to modify the gradient *homotopy*, in such a way that its crest lines correspond to the contours of only the *desired* regions, which are specified by providing *markers* for each one of them. A background marker is also required. The modification of the gradient homotopy is then accomplished by imposing the markers as the regional minima of the gradient and then suppressing all the other minima by way of a morphological reconstruction operation. The composition of modification of gradient homotopy and plain watershed transformation is known as the geodesic watershed transformation. We also refer to it simply as the watershed segmentation method.

Note that the the markers may be obtained automatically or specified manually. In the latter case, the method combines computer processing with human expertise. In the case of the mammographic calcifications, for instance, an experienced radiologist can provide the markers very easily using a mouse, discarding artifacts which otherwise could be mistaken by automatic methods as true calcifications.

In fig. 2, we illustrate the concepts discussed above for an image of a single mammographic calcification. Note that the image is blurred (due to magnification by bilinear interpolation) and has low contrast (fig. 2-a). Note also how in the over-segmented result (the output of the plain watershed transformation) the desired calcification contour is bogged down by irrelevant crest lines of the gradient (fig. 2-b). The markers for this example are very simple, a circle (the internal marker) and a circle's border (the external marker), which were placed manually on the image (fig. 2-c). After the homotopy change of the gradient by those simple markers, the watershed line gives the wanted calcification contour (fig. 2-d).

3 Classification of the Calcifications

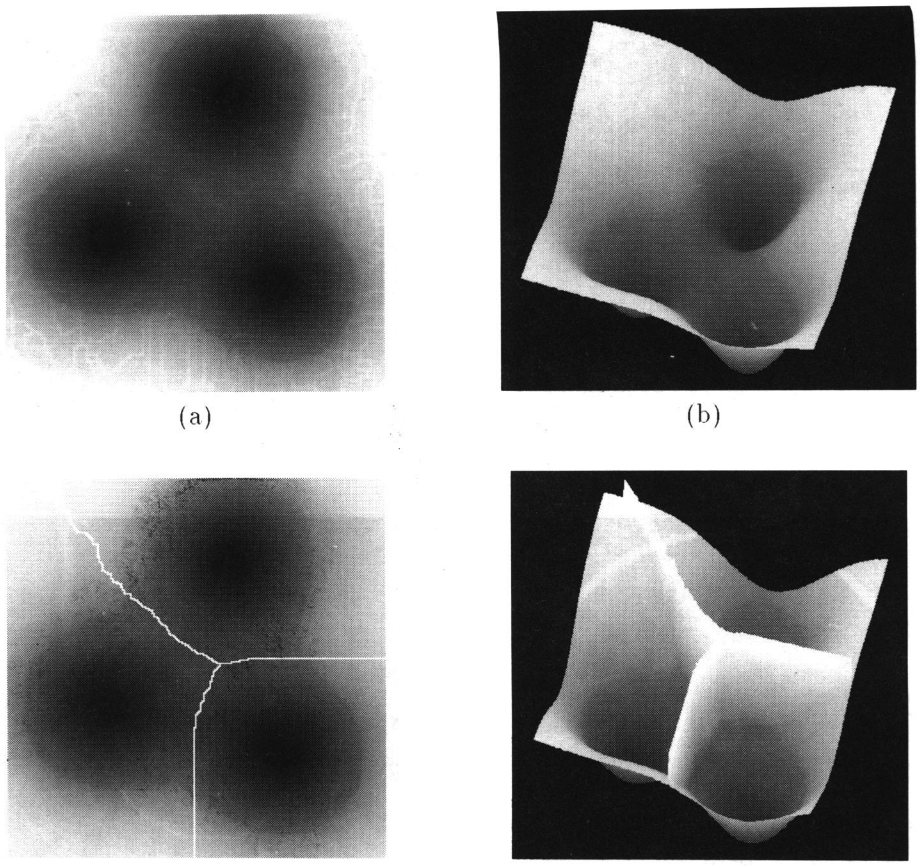
The classification step is based on the shape of the calcifications, which is known to be one of the most significant criteria for differentiating between benign and malignant calcifications. While benign calcifications tend to be roundish and in a small number, malignant calcifications are more irregularly shaped and have angular contours, appearing in a larger number and often in clusters [13].

3.1 Shape Factors

In this work we have utilized two simple, well-known shape factors. The *compactness* is a dimensionless descriptor defined as the ratio between the squared perimeter p and the area S of an object A:

$$C = \frac{p^2}{S} \tag{1}$$

The compactness is invariant with respect to scale, translation and, except for digital round-off errors, also to rotation [8]. This descriptor gives a measure of the roughness of an object shape, so it is very appropriate for characterizing calcifications. It is easy to verify that the compactness is minimal for a circle, and that the more complex is the contour of an object, the larger is its compactness.



(a) (b) (c) (d)

Figure 1: Flooding interpretation of the watershed method: (a) original image (b) relief associated with it (c) watershed line superimposed to original image (d) the watershed line as dams built on the relief

The second shape factor we have used is the *first invariant moment*:

$$M = \frac{m_{20} + m_{02}}{m_{00}^2} - \frac{m_{10}^2 + m_{01}^2}{m_{00}^3} \tag{2}$$

where the *moments* of order pq are defined for a digital object A as:

$$m_{pq} = \sum_x \sum_y x^p y^q \,, \quad \forall\,(x,y) \in A \tag{3}$$

Notice that m_{00} is simply the area S of A, whereas m_{10} and m_{01} are equal respectively to $S.\overline{x}$ and $S.\overline{y}$, where $\overline{x}$ and $\overline{y}$ denote the centroid coordinates of A.

It can be shown that the first invariant moment is insensitive to scale, translation and rotation changes [9]. As we show in section 4, the first invariant moment proved to be very effective in discriminating the shapes of benign and malignant calcifications.

3.2 The Nearest-Neighbour Classifier

The computation of the shape factors yields a *feature vector* $\mathbf{x}_i$ for each of the N calcification samples in the *training set*. Now, in order to get better classification results, it is convenient to apply a normalization to the feature vectors, so that each shape factor lies in the interval $[0, 1]$. The resulting normalized feature vectors $\mathbf{x}_i^*$ are given by:

$$\mathbf{x}_i^* = \begin{bmatrix} x_{i1}^* \\ x_{i2}^* \end{bmatrix}, \quad \text{for } i = 1, \ldots, N \tag{4}$$

with

$$x_{i1}^* = \frac{C_i}{\sum_{j=1}^N C_j} \tag{5}$$

$$x_{i2}^* = \frac{M_i}{\sum_{j=1}^N M_j} \tag{6}$$

The nearest-neighbour classifier is a simple nonparametric technique which assigns to a unknown

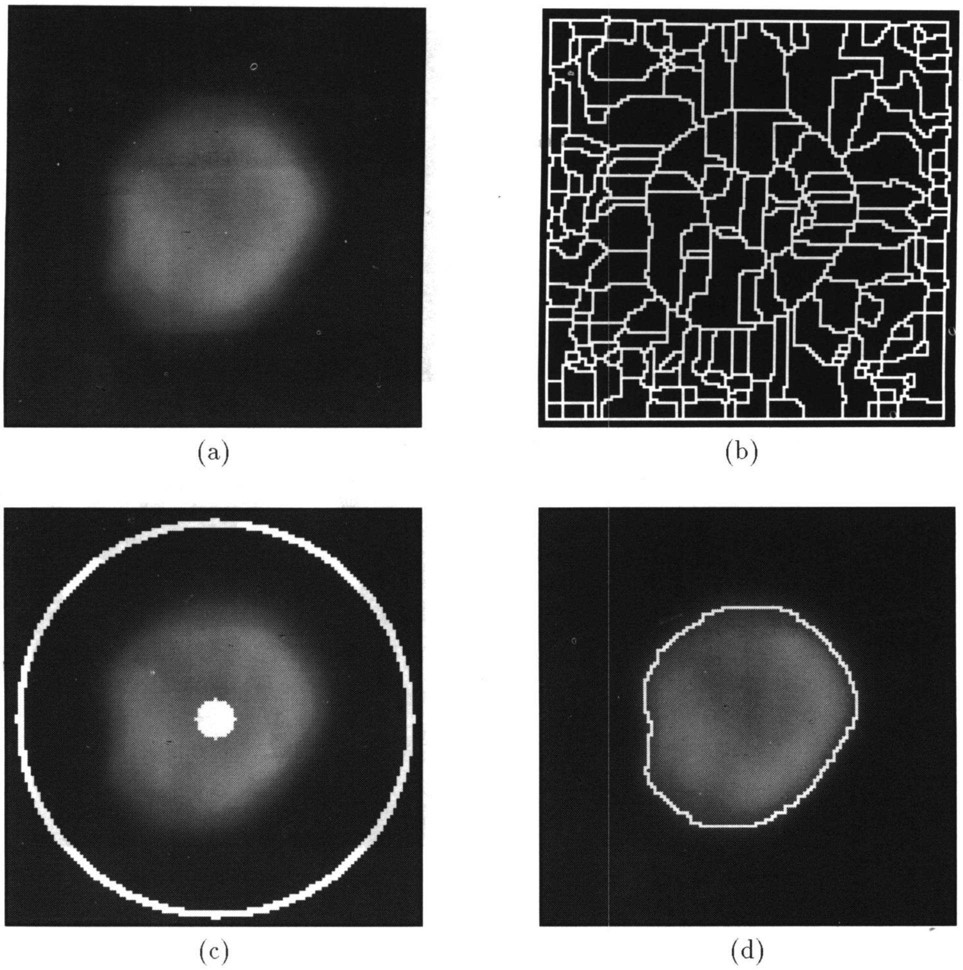


Figure 2: Example of segmentation by the watershed method: (a) Original image (b) Over-segmentation produced by the plain watershed transformation (c) Internal and external markers (d) Desired segmentation produced by the geodesic watershed transformation superimposed to the original image

calcification sample the class (in our case, benign or malignant) of the nearest sample in the training set, according to the euclidean metric. Despite the simplicity of the nearest-neighbour classifier, it can be shown that its error rate is nevertheless never greater than twice the minimal Bayes' rate [7].

Hence, the classification procedure is simply to compute the squared (to avoid the square root numerical complexity) euclidean distance

$$d(\mathbf{x}_i^*, \mathbf{y}^*) = \sum_{j=1}^{2} (x_{ij}^* - y_j^*)^2 \qquad (7)$$

between each normalized feature vector in the training set and the unknown normalized feature vector $\mathbf{y}^*$ and assign to the latter the class c of the training sample $\mathbf{x}_k^*$ for which $d(\mathbf{x}_k^*, \mathbf{y}^*)$ is smallest.

4 Experimental Results

In this section we present the results obtained with the application of the techniques described in the previous sections for the automatic detection and classification of mammographic calcifications.

We have based our implementation on the Khoros image processing system, running in a Sun SPARC10 workstation under Unix and X11R5. Khoros is a very popular open platform developed at New Mexico University and freely available through anonymous ftp. Khoros provides a large number of basic operators for image processing and analysis, and also allows C programmers to contribute their own operators, offering automatic code generators that simplify the task of integration to the system. We have utilized a set of such contributed operators for Mathematical Morphology image processing, the MMach

Khoros toolbox [1]. MMach provides all the basic morphological operators, as well as directly implemented operators for the plain watershed transformation and the modification of homotopy. MMach is freely available through anonymous ftp, at São Paulo University. For the classification step, we used the **vshape** Khoros program for the computation of the shape factors, and implemented a new Khoros operator for the nearest neighbour classifier.

We have utilized mammographic sections of biopsy-proven cases, chosen from the Radiology Teaching Library of the Foothills Hospital, Calgary, Canada. A total of seventy-four calcifications were selected for this study, divided in twenty-one benign calcifications from two sections and fifty-three malignant calcifications from one section. In figs. 3-a and 4-a we show respectively one of the benign sections and the malignant section used in this study.

We have applied the watershed method described in section 2 for the segmentation of the selected calcifications. The markers required by the geodesic watershed transformation were obtained in a semi-automatic fashion. The internal markers were single pixels placed manually inside each calcification with the mouse. From the internal markers we got the external marker, as the result of the geodesic watershed transformation of the *inverse* of the input mammographic section using the internal markers as the marker set. Considering the relief model of images (see section 2), the external marker correspond to the *valleys* in the background which separate the internal markers. The external marker thus obtained isolates each of the internal markers in disjoint regions, as can be seen in figs. 3-b and 4-b. Note that to each internal marker exactly one region is assigned.

The input mammographic sections were lowpass filtered and then their morphological gradient was computed (the low-pass pre-filtering has the property of enhancing the gradient). The geodesic watershed transformation, using the markers described above, was applied to the gradient, yielding the desired calcification contours, which can be seen superimposed to the original images in figs. 3-c and 4-c. For the computation of the shape factors defined in section 3.2, it is necessary to fill up the contours in order to obtain the binary regions correspondent to the calcifications. This last step in the segmentation procedure in accomplished by using the *close-holes* morphological operator [1]. The final result of the calcification segmentation process is seen in figs. 3-d and 4-d (the labels identify some of the calcifications for a posterior discussion of the method).

The shape factors defined in section 3.1, namely the compactness and the first invariant moment, were then computed for each of the segmented regions representing the calcifications. For estimating the probability of error of the nearest-neighbour classifier (section 3.2) we used the "leave-one-out" procedure. It consists of taking as test sample each one of the available samples in turn, the training samples being the remaining ones in the original set. In our case, this means that a total of seventy-four classification experiments were performed, in which each one of the calcifications were classified against the other seventy-three. This allowed us to compute a correct-classification rate for each of the classes (benign and malignant) and the global classification rate, which turned out to be the following:

$$\begin{aligned} \text{Benign classification rate} &= 90.48\% \\ \text{Malignant classification rate} &= 98.11\% \\ \text{Global classification rate} &= 95.95\% \end{aligned}$$

Only one of the malignant and two of the benign calcifications were misclassified, that is, three failures in a total of seventy-four experiments, a near-optimal result. If the malignant calcification labelled 3 in fig. 4-d is removed from the original set of seventy-four calcifications, the classification rates obtained become 100%.

The above classification result is an implicit function of the shape factors used and specially of the accuracy of the segmentation method employed. By comparison, a previous work on mammographic calcification detection and classification (which uses a similar nearest-neighbour classifier) [12], describes that using as segmentation method the classic region growing algorithm, and as shape factors the compactness, fourier descriptors, and a combination of boundary moments specially designed for this application, it was possible to get a 100% classification rate on a set of 143 calcifications. In our approach, by using as segmentation tool the watershed method, we were able to replace their two latter involved shape factors by the simple textbook first invariant moment and still get a very close result. We remark also that, unlike the previous work mentioned, our technique does not rely on any heuristically selected factors.

By having access to more mammographic sections, which would enable us to have an approximately equal number of benign and malignant calcifications, and by counting on the assistance of a experienced radiologist to select the internal markers in the segmentation step, we believe that our results can be even better.

A sensitive point in the application of the watershed method to the segmentation of mammographic calcifications is the very low constrast of the images.

This may lead in some cases to bad segmentation results, like the incomplete benign calcification labelled 2 in fig. 3-d, which "looks like" malignant, and probably also to the troublesome malignant calcification labelled 3 in fig. 4-d, mentioned above, which "looks like" benign.

The calcification labelled 1 in fig. 3-d poses a different kind of problem (see also the original calcification in fig. 3-a). It is probably a combination of *overlapping* calcifications. This is not surprising, because mammography is a projective screening technique, losing information on the axis paralell to view. Whether one considers these kind of objects as a single or multiple calcifications, any segmentation technique, including the watershed method, will produce undesirable contours which may lead to misclassification as malignant due to the angularities present.

5 Conclusion

We have demonstrated in this work that the watershed method, a powerful Mathematical Morphology tool for segmentation, is very effective for the automatic detection and classification of mammographic calcifications. Using a nearest-neighbour classifier and two simple textbook shape factors, the compactness and the first invariant moment, we were able to get classification rates, as given by the "leave-one-out" procedure, of nearly 100%, despite the poor contrast of mammographic images and the presence of overlapping calcifications. By having access to more mammographic sections and by counting on the assistance of a experienced radiologist we believe that it is possible to get even better results.

Acknowledgements

The first author of this work has been supported by a grant of FACEPE (Pernambuco State Science and Technology Foundation). The authors would like to thank Professor Rangaraj Rangayyam from the University of Calgary, Canada, for providing us with the mammographic images used in this work and also for calling our attention to his previous work on the subject.

References

[1] J. Barrera, G.J.F. Banon, and R.A. Lotufo. A Mathematical Morphology toolbox for the KHOROS system. Technical Report RT-MAC-9403, Instituto de Matemática e Estatística - Universidade Estadual de São Paulo, São Paulo, SP, Brasil, Jan 1994.

[2] L.W. Basset and R.H. Gold (eds.). *Breast cancer detection: mammography and other methods in breast imaging.* Grune & Stratton, Orlando (Second edition), 1987.

[3] S. Beucher. *Segmentation d'images et morphologie mathématique.* Doctorate Thesis - École Nationale Supérieure des Mines de Paris, France, 1990.

[4] U.M. Braga Neto. *Reconstrução Volumétrica e Análise de Imagens Tridimensionais por Morfologia Matemática.* MSc thesis, State University of Campinas, Brazil, 1994.

[5] H.P. Chan, K. Doi, S. Galhotra, C.J. Viborny, H. Macmahon, and P.M. Jokich. Image feature analysis and computer aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography. *Medical Physics*, 14(4):538–548, 1987.

[6] D.H. Davies and D.R. Dance. Automatic computer detection of clustered calcifications in digital mammograms. *Phys. Med. Biol.*, 35(8):1111–1118, 1990.

[7] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis.* John Wiley & Sons, NY, 1973.

[8] R.C. Gonzalez and R.E. Woods. *Digital Image Processing.* Addison Wesley Publishing Co., 2nd edition, 1992.

[9] M.K. Hu. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, pages 179–187, Feb 1962.

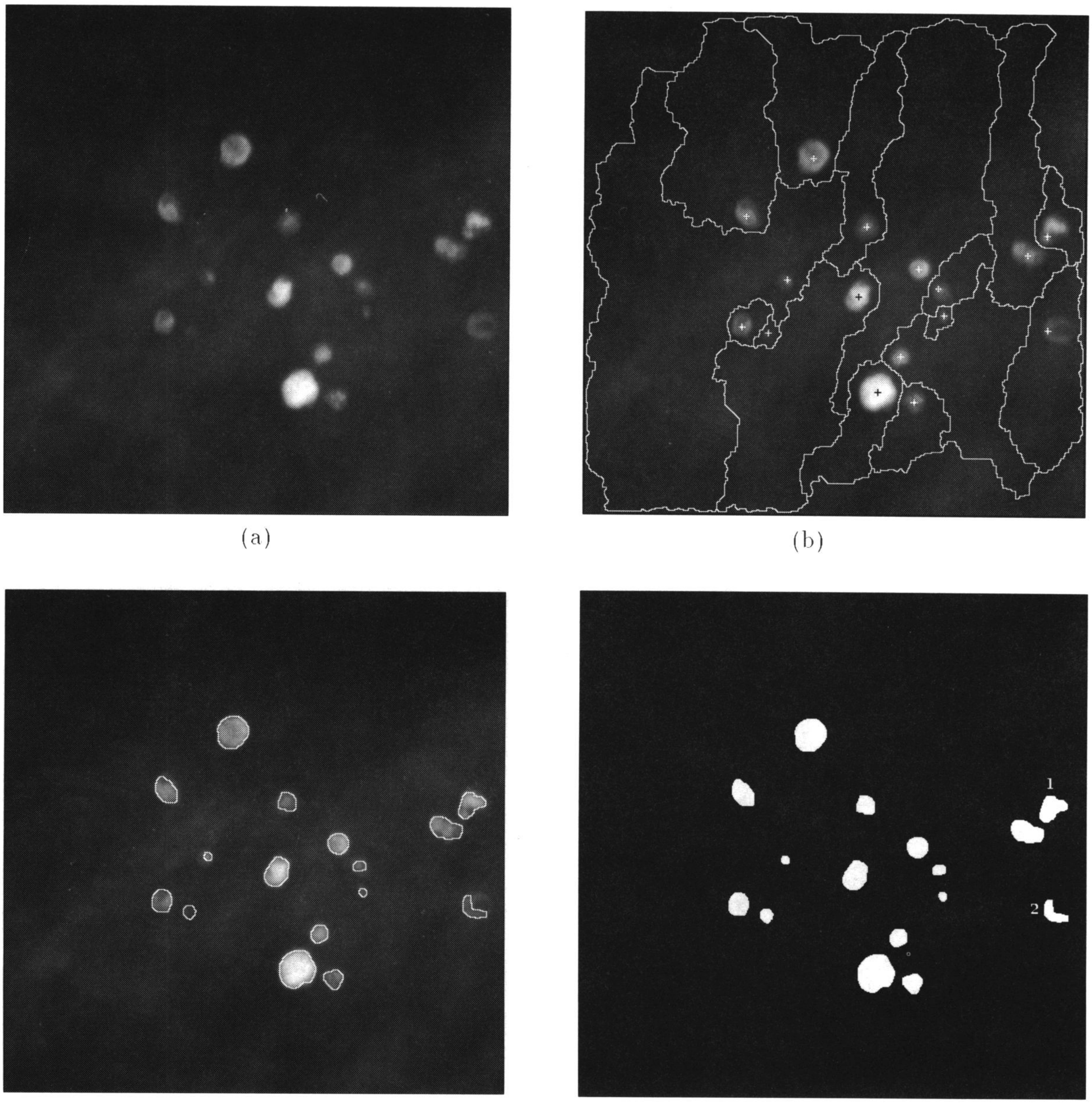
[10] J. Serra. *Image analysis and mathematical morphology.* Academic Press, Inc., 1982.

[11] L. Shen, R.M. Rangayyam, and J.E.L. Desautels. Detection and classification of mamographic calcifications. *Intenational Journal of Pattern Recognition and Artificial Inteligence*, 7(6):1403–1416, 1993.

[12] L. Shen, R.M. Rangayyam, and J.E.L. Desautels. Application of shape analysis to mamographic calcifications. *IEEE Transactions on Medical Imaging*, 13(2):263–274, June 1994.

[13] E.A. Sickles. Breast calcifications: Mammographic evaluation. *Radiology*, 160:289–293, 1986.

[14] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pat. Anal. Mach. Intel.*, 13(6):583–598, Jun 1991.



(a) (b)

(c) (d)

Figure 3: Segmentation of calcifications in one of the benign mammographic sections: (a) Original image (b) Internal markers (represented as little crosses) and external markers superimposed to the original image (c) Contours produced by the geodesic watershed transformation superimposed to the original image (d) Filled up regions correspondent to the selected calcifications

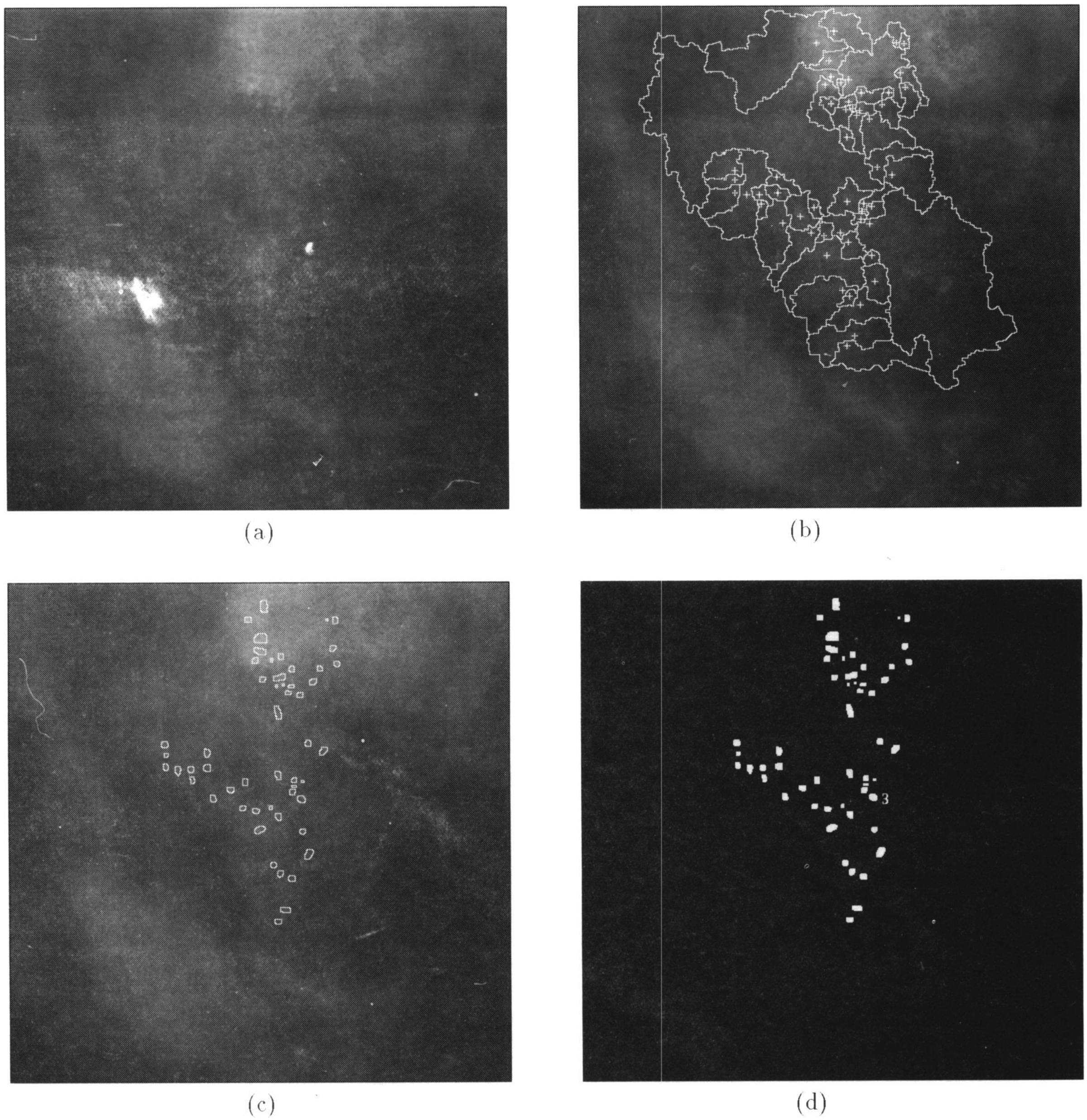


Figure 4: Segmentation of calcifications in the malignant mammographic section: (a) Original image (b) Internal markers (represented as little crosses) and external markers superimposed to the original image (c) Contours produced by the geodesic watershed transformation superimposed to the original image (d) Filled up regions correspondent to the selected calcifications