

A fast algorithm for intra-breast segmentation of digital mammograms for CAD systems

Renato Campanini¹, Enrico Angelini¹, Emiro Iampieri¹, NicoLanconelli¹, Matteo Masotti¹,
Matteo Roffilli^{1,2}, Omar Schiaratura², Massimiliano Zanoni¹

¹ Department of Physics, University of Bologna and INFN – Bologna, Italy

² Computer Science, University of Bologna, Italy

Abstract. Current CAD systems always demand better performance, both in terms of the best sensitivity-specificity tradeoff and of the processing time. In order to decrease the false positive rate and to increase the time responsiveness of our CAD system, we present a powerful algorithm that performs an intra-breast segmentation. Starting from a digital mammogram and by using a combination of different techniques, the proposed algorithm produces a binary map showing the locations of the breast with higher probability of containing a suspect mass. Experimental trials performed on 317 mammograms show that our algorithm is able to reduce the intra-breast searching area down to 6% of the entire breast area, while maintaining almost the same negative rate. Typical dimension of images is about 2800×2000 pixels at 13 bits of gray level resolution. The set of images was gathered from FFDM units coming from two different hospitals.

1. Introduction

An important step in mass detection is the segmentation of suspect regions within the breast. The aim of such technique is basically to exclude the background area from the further processing and to find out within the breast region the suspect zones. Based on such segmentation, a significant reduction of both the number of false-positives and the computation time can be achieved. Segmentation methods are usually based on gray-level

thresholding, or on more sophisticated techniques such as region growing (Bischof and Adams 1994) or active contours (Sahiner et al. 2004).

The algorithm presented in this paper has been used as a first step in a CAD system for the detection of masses developed by our group (Campanini et al. 2004). The CAD is based on a Support Vector Machine classifier, which allows a featureless detection of the lesions. That system does not rely on any feature extraction task; on the contrary, the algorithm automatically learns to detect the lesions by the examples presented to it during the training phase. By using the presented segmentation technique as a pre-selector for the regions to be analyzed by the CAD, it is possible to greatly decrease the area to be scanned. That allows the reduction of the computation time and of the false-positive rate. The algorithm has been tested on images collected at two different sites: Maggiore Hospital in Bologna (Italy) and Triemli Hospital in Zurich (Switzerland). Both these hospitals have a Giotto Image MD FFDM systems manufactured by IMS, Italy.

2. Methods

Given a digital image, the main purpose of our algorithm is to reject those parts, which don't contain objects of interest for our analysis. To this end, we perform two kinds of segmentation tasks: an external one, in order to eliminate the part that non represent the breast, and an internal one, for excluding inner breast regions which surely don't contain lesions. In this paper, we focus our attention on the latter procedure. This process consists of a sequence of steps, each one involving the scanning of the entire image by means of several masks. Figure 1 shows a flow chart of the presented algorithm.

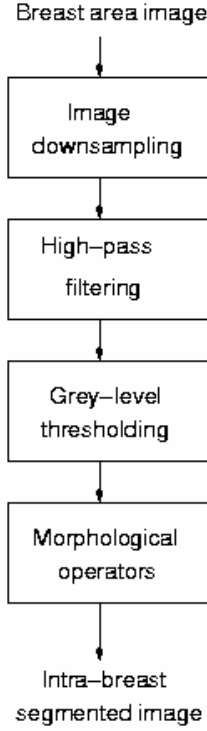


Figure 1. Flow-chart of the segmentation method.

Because of the very high spatial resolution of the mammograms, a downsampling technique is applied for reducing the computation time. To this aim, a Gaussian low-filter implemented with a 3×3 convolution matrix is used. A further advantage of this process is to produce a light smoothing, which allows eliminating little noise imperfections. The filtered image has been reduced in size by a factor 4. As known, masses appear as bright regions in the mammogram, so our next step aims to separate those areas from the structured background due to the breast parenchyma. For this reason, we then applied a High-Boost filter (Gonzalez and Woods 1992), helpful for emphasizing objects with high intensity. After that, we operate a thresholding, in order to identify the bright objects highlighted by the previous filter. Given an image $f(x,y)$ a single threshold is defined as follow:

$$bin(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq T \\ 0 & \text{if } f(x,y) < T \end{cases} \quad (1)$$

where $bin(x,y)$ is the resulting binary image and T the threshold value. The finding of the optimal value for T is often one of the most difficult tasks. As already seen, the main purpose

of the presented algorithm is to reduce the total area of interest, without losing information about lesions. Here, the most critical issue is to determine the best tradeoff between the reduction of the searching area and the loss of mass information. Our threshold algorithm defines a dynamic T value based on mean M and standard deviation $Sigma$, estimated within each considered mask. The entire process consist in the translation of a scanning mask over the image, and in the estimation of the statistical values M and $Sigma$. In order to have a certain degree of overlap between contiguous masks, a scanning step equal to 60% of the mask size is used. At each step, the threshold value T is then calculated as:

$$T = M + \alpha \cdot Sigma + M \cdot k \quad (2)$$

where α and k are two parameters adjustable by the user. In a classic thresholding approach, all the parts inside the breast are considered in the same way (*i.e.* the values of α and k are unchanged over the entire breast area). However, this could lead to some troubles in the regions close to the borders of the breast. In order to reduce those problems, we consider separately the regions close to the pectoral muscle and those within the external border of the breast. In the first case, we notice that the regions of the breast next to the muscle usually have a brightness distribution quite uniform, with a consequent small $Sigma$ value. In this situation, a value for T smaller than threshold values of other parts of breast would give better results. A similar situation is observed on the breast external border. Here, the thickness shrinks very suddenly. As a consequence, a lot of pixels have value over the threshold T . For reducing these problems, a different normalization k value is applied. In particular, the breast is divided into three different parts (external border, border close to the chest and internal breast tissue) processed each one separately. In this way, it is possible to define three different k values to be used inside the breast region, in order to get a more accurate segmentation. Figure 2 shows the difference between the thresholding performed with a fixed k value and with three different k values for the three aforementioned regions. It is possible to

see that the choice of different values allows a more precise segmentation, especially in the external border and in the area close to the chest.

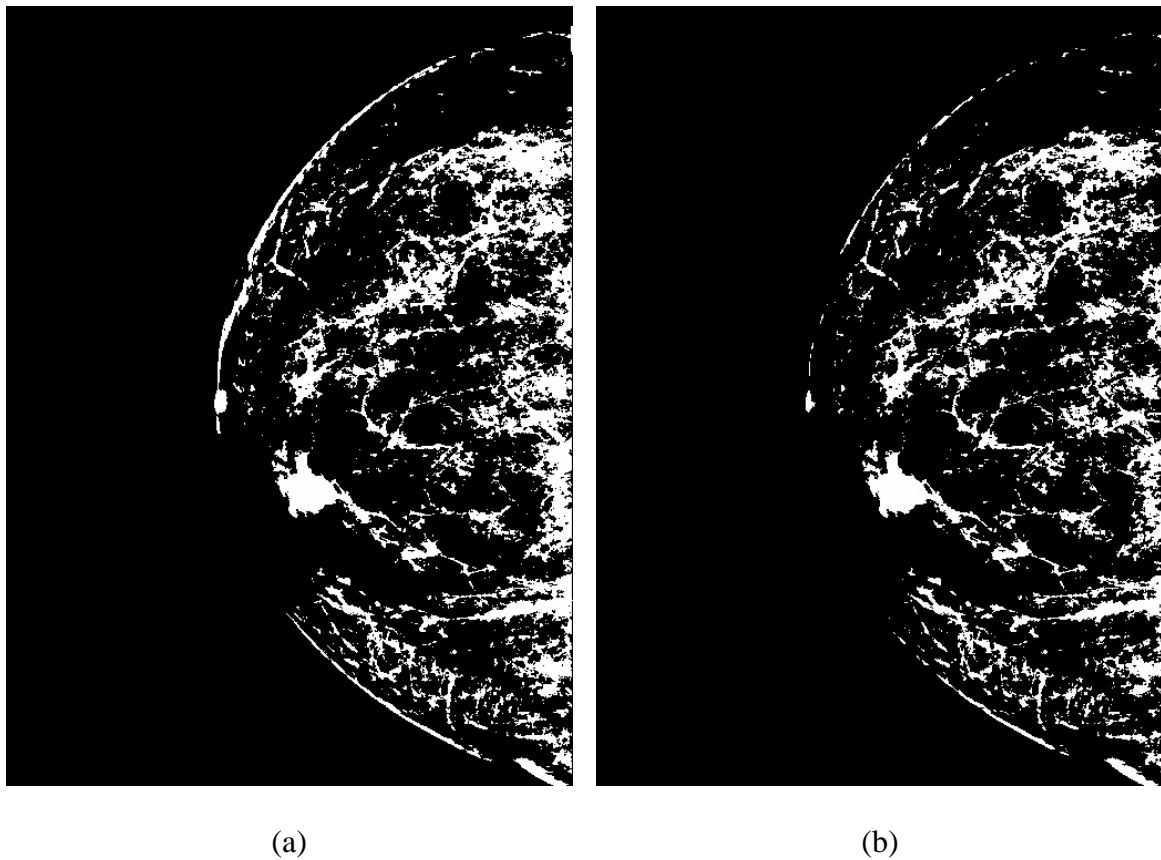


Figure 2. Effect of thresholding with a fixed k value (a), and with three different k values for external border, border close to the chest and internal breast tissue (b).

Algorithms used till this point are all pixel oriented, thus they don't have any knowledge about the characteristics of a mass and its shape. As a consequence, we need to apply further processes to better discover objects of interest. In particular, the lesions we are interested in have an oval shape with linear dimension that ranges between 3 and 40 mm. To this end, we perform some morphological operations as final step. The aim of this phase is to reject linear, very small or very large objects, and to reconstruct the correct shape of the objects altered by the local thresholding. In particular, these operations are combinations of erosions and dilations. Figure 3 shows some images at the different steps of the presented algorithm.

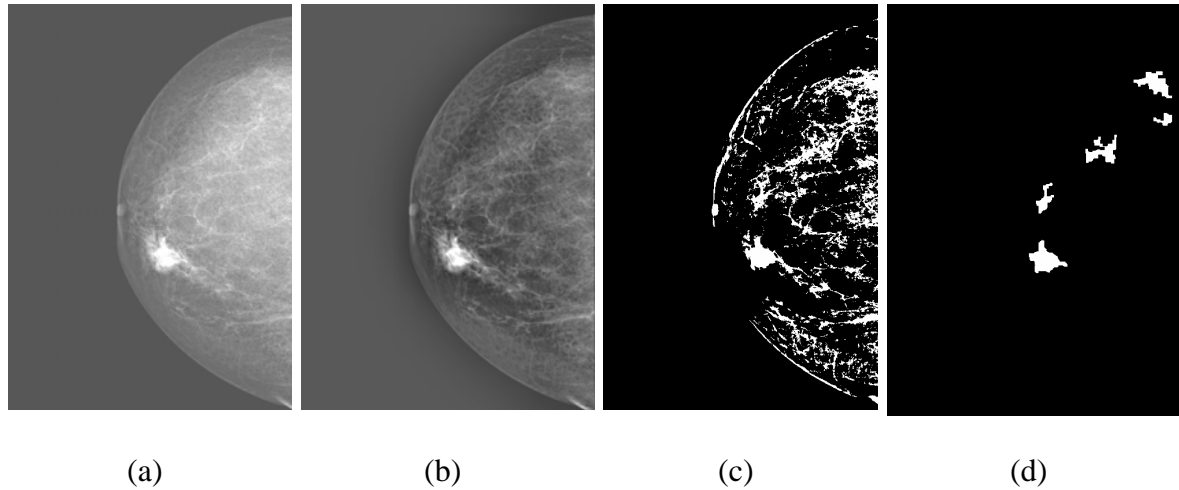


Figure 3. Typical image appearance in the various step of the algorithms: (a) original image; (b) after the high-pass filter; (c) after the gray-level thresholding; and (d) the final image, after the morphological operations.

3. Results

The intra-breast segmentation algorithm has been tested on 317 FFDM images, 89 that present a mass lesion and 228 without lesions. The algorithm reduced the intra-breast area to the 6% of the entire breast area, losing only 1 mass out of 89 (sensitivity equal to 98.9%). As a consequence, the average total time for completing a detection, by using our CAD system drops from about 360 sec per image to about 10 sec per image on a Pentium Xeon 2.8 GHz machine. Table 1 summarizes the results obtained on the FFDM database.

Table 1. Results in terms of reduction in searching area, computation time and sensitivity on the FFDM images.

	Area [%]	Time [s]	Sensitivity [%]
Without segmentation	100	360	100
With segmentation	6	10	98.9

References

- Bazzani A., A. Bevilacqua, D. Bollini, R. Brancaccio, R. Campanini, N. Lanconelli, A. Riccardi, and D. Romani. 2001. An SVM classifier to separate false signals from microcalcifications in digital mammograms. *Phys. Med. Biol.* 46:1651-1663.
- Bischof L., and R. Adams. 1994. Seeded region growing. *IEEE Trans. Pattern Anal. Machine Intell.* 16: 641-647.
- Campanini R., D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli. 2004. A novel featureless approach to mass detection in digital mammograms based on Support Vector Machines. *Phys. Med. Biol.* 49:961-975.
- Gonzalez R.C., and R.E. Woods. 1992. *Digital Image Processing*. Boston, MA, USA, Addison-Wesley Longman Publishing Co.
- Sahiner B., H.P. Chan, M.A. Roubidoux, M.A. Helvie, L.M. Hadjiiski, A. Ramachandran, C. Paramagul, G.L. LeCarpentier, A. Nees, and C. Blane. 2004. Computerized characterization of breast masses on three-dimensional ultrasound volumes. *Med. Phys.* 31: 744-754.