CARS 2002 – H.U. Lemke, M.W. Vannier; K. Inamura, A.G. Farman, K. Doi & J.H.C. Reiber (Editors)

Segmentation of microcalcification in X-ray mammograms using entropy thresholding

Moti Melloul and Leo Joskowicz

School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem 91904, ISRAEL
E-mail: moti@cs.huji.ac.il, josko@cs.huji.ac.il

Abstract

We describe a new algorithm for microcalcification segmentation in mammographic X-ray images. The algorithm detects microcalcifications in two steps. First, it removes background tissue with a multiscale morphological operation. Then, it applies entropy thresholding based on a 3-dimensional co-occurrence matrix. Unlike existing methods, ours is fully automatic, parameter-free, and independent of local statistics. To test its efficacy, we applied it to images from the Mammographic Image Analysis Society database and analyzed the results with the assistance of a clinician. We obtained detection rates of 93.75% of true positives, 6.25% of false positives, and 2% of false negatives.

Keywords: X-ray mammograms, microcalcification segmentation, entropy thresholding.

1. Introduction

Most early breast cancer can be diagnosed by detecting microcalcification clusters in mammographic X-ray images. The clusters appear as groups of small, bright particles with arbitrary shapes. Detecting microcalcifications is difficult because they are embedded in a non-homogeneous background. Many missed radiologist diagnoses can be attributed to human factors such as subjective or varying decision criteria, distraction by other image features, large number of images to be inspected, or simple oversight. Therefore, there is strong motivation to develop reliable and effective methods for automatic microcalcifications detection. While many methods for microcalcification segmentation have been developed in the past ten years, they either require manual thresholds adjustment or depend on local statistics to compute those thresholds. This paper presents a new fully automatic, parameter-free, and local statistics independent algorithm for microcalcification segmentation in mammographic X-ray images. For a detailed description of the method, see [1].

2. Previous work

Strickland and Hahn [2] describe a method that uses multiscale matched filters with wavelet transforms for enhancing and detecting calcifications. Nishikawa *et al* [3] use a

difference technique to enhance microcalcifications. First, it extracts potential microcalcifications with global thresholding based on an erosion operator and local adaptive thresholding. False positives are then eliminated by texture analysis, and the remaining candidates are grouped with a non-linear clustering algorithm. Cheng *et al*. [3] propose a method based on fuzzy logic, which consists of image fuzzification, enhancement, irrelevant structure removal segmentation, and reconstruction. Chan *et al*. [4] investigate a convolution neural network based approach that is effective for reducing false positive detections. Nagel *et al*. [5] compare three feature analysis methods based on rules, artificial neural networks, and a combination of both. They report that the combined method performs best because each filter eliminates different types of false positives. McGarry and Deriche [6] use a hybrid model combining knowledge of mammographic imaging process and anatomical structure and Markov random fields. The drawbacks of these methods are that they either require manual adjustment of thresholds or depend on local statistics to compute the thresholds. This motivated our search for a parameter-free algorithm for threshold estimation.

3. Method

The algorithm detects microcalcifications in two steps. First, it removes background tissue with a multiscale morphological operation. Then, it applies entropy thresholding based on a third-order spatial gray-level dependence matrix.

Background tissue elimination is necessary to enhance the visibility and detectability of microcalcifications. We use a multiscale top hat morphological filtering to remove the slow rate of variation of the image intensity values and to enhance the image contrast. In morphology, filtering is performed using a kernel, and multi-scaling is performed by changing the size of the kernel. The top hat filter γ is a morphological opening operation. For a given image I, the multiscale top hat operator ν removes objects whose size is larger than the given kernel size. The kernel is taken to vary from the smallest to the largest size of individual microcalcifications. The multiscale top hat equation is:

$$\nu_k(I) = I - \gamma_k(I)$$

where k is the kernel size.

The opening operation consists of erosion followed by dilation on a kernel that defines the size of the region over which pixel values are taken. Erosion replaces the pixel value at the center of the kernel by the minimum value of its neighborhood pixels, while dilation replaces it by the maximum value of its neighborhood pixels. The opened image is then subtracted from the original image. We use square kernels whose sizes vary between one and five pixels. For each scale, we obtain different filtered images with candidate microcalcifications.

To segment the resulting filtered images, we apply the following entropy-based thresholding method. First, we compute the spatial gray-level dependence matrix. This is a three-dimensional co-occurrence matrix T whose entries are the joint probabilities that pixel triplets' intensities (w_i,w_j,w_k) are in a rectangular region of width s and height h. The

entries of the third -order entropy matrix of the image are then obtained by summing the pixel triplet probability times its logarithm over all regions of size *sxh*. We choose to use the third -order space mean over the more commonly used second order matrix because our experiments indicated that higher order correlations improved the discrimin ation capabilities. The (*i*,*j*,*k*)th entry of the 3D co -occurrence matrix, denoted by T_{ijk}, is defined by:

$$T_{ijk} = \sum_{m=1}^{M} \sum_{n=1}^{N} \delta(m,n)$$

where

$$\delta(m,n) = \begin{cases} 1 \text{ if } I(m,n) = i \text{ and } I(m,n+1) = j \text{ and } I(m+1,n) = k \\ 0 \text{ otherwise} \end{cases}$$

Next, we partition the resulting matrix into two regions: the background information *B*, which appears in the upper left corner of the matrix, and the microcalcification information *O*, which appears in the rest of the matrix. Each region defines a distribution of the of gray-levels transitions. Then, we build for each region the probability *P* based on the distribution of the pixels transitions in the given region of the 3D co -occurrence matrix. By normalizing the total number of transitions in the given region of the co - occurrence matrix, we obtain the desired transition probability *P*.

$$P_{ijk} = \frac{T_{ijk}}{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \sum_{k=0}^{L-1} T_{ijk}}$$

The size of the regions *B* and *O* is adjusted dynamically by changing the position of the boundary between the two regions. We compute the entropy of each region according to the boundary position. The boundary separating background and microca lcifications is the one that gives the maximum sum of the entropies. The optimal threshold is the one that maximizes the sum the entropies of the background and microcalcification regions defined by this boundary

Formally, let *t* be the threshold of the two groups the foreground and the background in the image. The background entropy $H_B(t)$ and the objects entropy $H_O(t)$ are computed on the volumes *B* and *O*. The entropies quantify the background -to-background transitions and objects-to-objects transitions. The image entropy is obtained by

$$H(t) = H_B(t) + H_O(t)$$

The optimal threshold is the value *t* that yields the maximum of the image entropies

$$T_{optimal} = \arg\max_{t}(H(t))$$

Then, we segment the filtered image according the optimal threshold. The fusion of the different scale segmented images produces the final mammogram segmentation.

4. Results

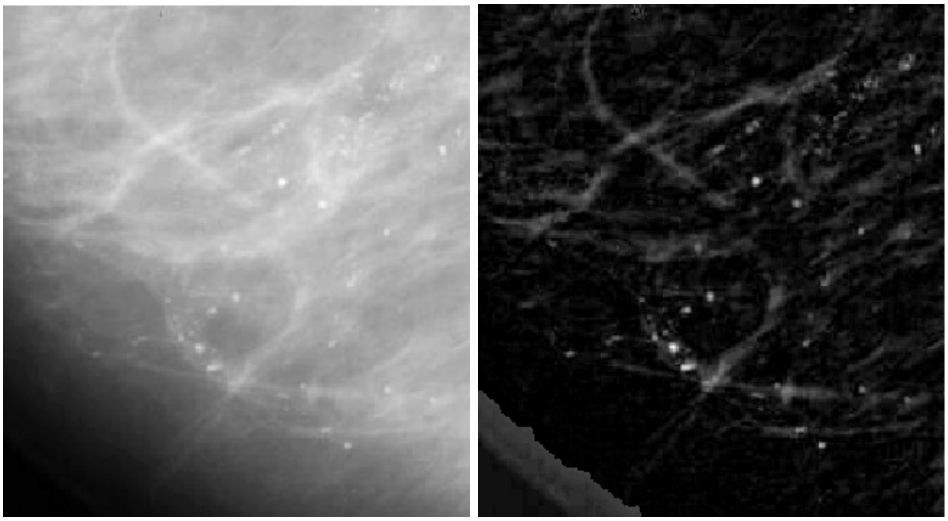
We applied the algorithm to a database consisting of a few dozen images from the Mammographic Image Analysis Society (MIAS) database and our own clinical images. We then performed a quantitative analysis of the results with the assistance of a radiologist. Images are of size *1024x1024* pixels with *8*-bit gray-values. We use a morphological multiscale top hat with 1x1 pixels to 5x5 pixels kernels for background filtering and applied entropy thresholding as described above to detect individual microcalcifications. The algorithm was able to detect subtle microcalcifications and its results were deemed highly accurate. The microcalcifications were then are grouped into clusters based on their proximity using the Cluster Affinity Search Technique. The intermediate steps of the algorithm are illustrated in Fig. 1 on a sample mammogram.

We obtained mean detection rates of *93.75*% of true positives, *6.25*% of false positives, and *2.0*% of false negatives (ranging from *0*% to *3.75*%). The results were evaluated and confirmed by a radiologist. These are considered very low false positive and false negative rates. We noted some variance depending on the size of the region of interest and the texture of the mammogram in the region. Running times on a Pentium III *700* MHz PC with *256*MB of main memory and no code optimization average *20* minutes per image. This was deemed acceptable based on the quality of the results.

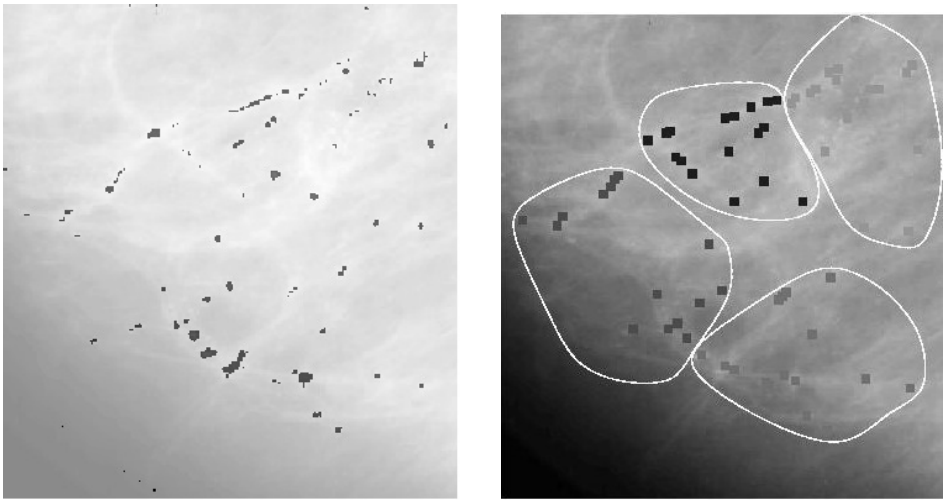
To further evaluate our algorithm, we compared it to three state of the art algorithms. Yu and Guan [7] compute statistical features and report *90*% mean true positives at the cost of *0.5*% false positives per image. On our data set, our implementation of the algorithm produced *92*% mean true positives and *8*% false positives, which is significantly worse than our results. Its drawbacks are that it has many statistical features and parameters to adjust, and is very computation-intensive. Vilarras et al. [8] use morphological operations to remove noise background and report a true positives rate of only *85*% on the MIAS data set. Karssemeijer and Barke [9] apply a statistical method and report a mean of *93*% true positives for a cost of two false positives per image.

4. Conclusion

We have presented a new algorithm for microcalcification segmentation in mammographic X-ray images. The algorithm uses a multiscale morphological operation and entropy thresholding based on a three dimensional co-occurrence matrix. Unlike existing methods, ours is fully automatic, parameter-free, and independent of local statistics. We are currently applying our algorithm to a larger data set and investigating the causes for the small percentage of false positives and false negatives. To improve these results, we plan to perform an analysis of the extracted microcalcifications in the morphological and texture feature spaces.



(a) Original region of interest (b) After background removal



(c) After microcalcification segmentation (d) After clustering into suspicious regions

Fig. 1: Illustration of the algorithm steps on a sample mammogram

Acknowledgments
This research was supported in part by a grant from the Israel Ministry of Trade and Industry under the IZMEL consortium on Image -Guided Therapy. We thank Dr. Miri Sclair for qualifying and quantifying the results of our analysis.

References

[1] M. Melloul, *Segmentation of Microcalcifications in X-ray Mammograms using Entropy Based Thresholding*, Masters Thesis, The Hebrew University of Jerusalem, Dec. 2001. Available in http://www.cs.huji.ac.il/~josko/cas-publications.html

[2] R.N. Strickland and H. I. Hahn, ``Wavelet transforms for detecting microcalcifications in mammograms", *IEEE Transactions on Medical Imaging* , Vol. 15(2), pp. 218-229, April 1993.

[3] H-D. Cheng, Y. M. Lui, and R. I. Freimanis, ``A novel approach to microcalcification detection using fuzzy logic technique", IEEE Transactions on Medical Imaging, Vol. 17(3), pp. 442-450, June 1998.

[4] H.-P. Chan, S.-C. B. Lo, B. Sahiner, K.L. Lam, and M.A. Helvie. ``Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network", *Medical Physics*, Vol. 22(10), pp. 1555-1567, October 1995.

[5] R. H. Nagel, R. M. Nishikawa, J. Papaioannou, and K. Doi, "Analysis of methods for reducing false positives in automated detection of clustered microcalcifications in mammograms", *Medical Physics*, Vol. 25(8), pp. 1502-1506, August 1998.

[6] G.McGarry and M. Deriche, ``Mammographic image segmentation using a tissue-mixture model and Markov random fields", *IEEE International Conference on Image Processing*, ICIP, 2000.

[7] S. Yu and L. Guan , A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films'', *IEEE Transactions on Medical Imaging*, 19(2) 115-126, February 2000.

[8] A. Vilarrasa, V. Gimenez, D. Manrique, J. Rios . A new algorithm for computerized detection of microcalcifications in digital mammograms'', Proceedings of the Int. Conference on Computer-Aided Radiology and Surgery, CARS'98, pp. 224-229.

[9] N. Karssemeijer and G.M. Barke, Recognition of clustered microcalcifications using a random field model'', *Proceedings of SPIE*, Vol. 1905, 1993, pp. 776-786.