
*Feature Extraction in Digital Mammography*¹

R. A. DeVore¹, B. Lucier², and Z. Yang³

¹*Department of Mathematics, University of South Carolina, Columbia, South Carolina*

²*Department of Mathematics, Purdue University, W. Lafayette, Indiana*

³*Institute for Scientific Computation, Texas A&M University, College Station, Texas*

6.1 Introduction

The devastating impact of breast cancer in the United States is well documented [1]. This points to the need for early cancer detection. The number of mammograms generated daily is large and therefore it is very desirable to develop image processing tools which facilitate the handling of mammograms (storage and transmission) and aid the radiologist in diagnosis.

We have previously reported [9, 12] on wavelet-based techniques for compression of mammographic images. The purpose of the present paper is to put forward new wavelet-based techniques for the automated detection (using only a computer) and semi-automated detection (interactive with an operator) of certain early signs of cancer. Signs of breast cancer are typically categorized by radiologist in three categories: clusters of microcalci-

¹This research was supported by the Office of Naval Research Contracts N00014-91-J1076 and N00014-91-J1152, and the National Science Foundation Grant EHR 9108772.

fications, stellate lesions, and circumscribed lesions. We shall be primarily concerned with the automated detection of clusters of microcalcifications. However, the techniques we put forward may also have application to the detection of lesions. Moreover, the detection of microcalcification clusters has several features in common with other problems of feature classification and automated detection.

6.2 Mammograms as Digitized Images

While mammograms usually are taken on film, we assume that they have been digitized (typically by a digital scanner). A new generation of mammography will directly produce digitized images.

A digitized image is an array of pixel values. For mammograms, this array is generally not square and this causes some technical difficulties in wavelet-based image processing algorithms. Also the size of mammograms is large when compared with many other images. However, for the convenience of discussion in this paper, we shall assume that the digitized mammogram is not only square but of the size $2^m \times 2^m$. Typical values are $m = 9, 10$. (All examples of digitized mammograms considered in this paper are of size 512×512 .)

Thus, a digitized mammogram, for the purposes of this paper, will be an array of nonnegative integers

$$p_k, \quad k = (k_1, k_2), \quad k_1, k_2 \in \{0, \dots, 2^m - 1\}.$$

The range of the integer values p_j is related to the scanner and the dynamical range of the film. We shall assume that the p_j take values from $\{0, \dots, 4095\}$ (again for convenience and because all sample images considered will have 12-bit dynamical range). The techniques we develop here are independent of the dynamical range.

The following viewpoint of a digitized image is useful in the analysis that follows. We can view the pixel values as obtained from a bivariate function F defined on the unit square $\Omega := [0, 1]^2$ by taking cell averages

$$p_k = \int_{Q_k} F(x, y) dx dy, \quad 0 \leq k_1, k_2 \leq 2^m - 1$$

where $Q_k := [2^{-m}k_1, 2^{-m}k_1 + 2^{-m}] \times [2^{-m}k_2, 2^{-m}k_2 + 2^{-m}]$ is the square with sidelength 2^{-m} and lower left vertex $2^{-m}k = 2^{-m}(k_1, k_2)$. Thus, we view the pixel values as samples of the underlying function F .

In wavelet-based image processing, we need a representation of the image as a wavelet sum. Let φ be a univariate function which generates multiresolution as described in §1.5 of Chapter 1. Since images are bivariate, we shall need the corresponding bivariate multiresolution and bivariate wavelet bases. Let

$$\phi(x, y) := \varphi(x)\varphi(y)$$

be the bivariate function which is the tensor product of φ with itself. As in the univariate case, we have the bivariate ladder of space V_j , $j \in \mathbf{Z}$, with V_j defined as the $L_2(\mathbf{R}^2)$ -span of the functions²

$$\phi_{j,k} := 2^{-j}\phi(2^{-j}x - k_1, 2^{-j}y - k_2), \quad k := (k_1, k_2) \in \mathbf{Z}^2.$$

Recall that when φ has compact support, the function $\phi_{j,k}$ has support localized near the point $2^j k$. From the pixel values (p_k) , we create an approximation f to F from the space V_{-m} :

$$F \approx f = \sum_j c(k)\phi_{-m,k}. \quad (2.1)$$

We assign the coefficients $c(k)$ so that the portion of the sum (2.1) which is nonzero on Ω is a good approximation to F . This is usually accomplished by defining

$$c(k) := p_k, \quad 0 \leq k_1, k_2 \leq 2^m - 1,$$

and defining $c(k)$ for other values of k by some extension strategy. Typical extensions are either symmetric (cf. Chapter 2) or chosen to preserve constant or linear sequences. We refer the reader to [2, 4, 13] for general discussions of extension strategies.

The representation (2.1) is not suitable for most tasks in image processing. We therefore convert (2.1) to the more preferable wavelet basis. For our purposes it will be sufficient to restrict our attention to the cases where the univariate wavelet basis is orthogonal, semi-orthogonal, or bi-orthogonal, as described in §1.5 of Chapter 1. If ψ is the univariate function whose shifts are a basis for the wavelet space W_0 , then the following bivariate functions span the corresponding bivariate wavelet space:

$$\varphi(x)\psi(y) \quad \psi(x)\varphi(y) \quad \psi(x)\psi(y). \quad (2.2)$$

²We have conformed to the notation of this book in the definition of the spaces V_j and the functions $\phi_{j,k}$. Our other work on wavelets utilizes slightly different notation with j replaced by $-j$ in the definition of these spaces and functions.

We let Ψ denote the set consisting of the three functions in (2.2) and for $\eta \in \Psi$, we define

$$\eta_{j,k} := 2^{-j}\eta(2^{-j}x - k_1, 2^{-j}x - k_2), \quad k = (k_1, k_2) \in \mathbf{Z}^2.$$

These functions span the corresponding bivariate wavelet space W_j which encodes the detail between V_j and the finer space V_{j-1} . Namely, we have

$$V_{j-1} = V_j + W_j. \quad (2.3)$$

In the case of orthogonal or semi-orthogonal wavelets, the sum in (2.3) is orthogonal, while in the bi-orthogonal case it is generally oblique.

We can use the fast wavelet transform (FWT) to convert the representation (2.1) to the wavelet representation. In this way, we obtain

$$f = \sum_{k \in \mathbf{Z}^2} c(k)\varphi_{-m,k} = \sum_{k \in \mathbf{Z}^2} c_0(k)\phi_{0,k} + \sum_{j=-1}^{-m+1} \sum_{\eta \in \Psi} \sum_{k \in \mathbf{Z}^2} d_{j,\eta}(k)\eta_{j,k}. \quad (2.4)$$

In each of these sums, the k can be restricted to those values for which $\eta_{j,k}$ is nonzero on Ω . The wavelet coefficients $d_{j,\eta}(k)$ can be computed from the pixel values in $O(N)$ operations with $N = 2^{2m}$ the original number of pixel values. We call (2.4) the wavelet representation (cf. Chapter 2) of the image.

6.2.1 Characteristics of Mammographic Images

It will be useful to note some characteristics of mammographic images since these will motivate our algorithms for microcalcification detection. A more detailed exposition of the characteristics of mammographic images as related to the detection of microcalcifications can be found in [3].

A typical mammogram is shown in Figure 6.1. The microcalcifications which we wish to identify appear as small bright spots in the mammogram. Their diameters are up to .7 mm, with an average size of .3 mm. Thus, the microcalcifications can be identified with certain dyadic levels j in the wavelet decomposition — namely, those for which 2^j is comparable with the size of the microcalcification. In the examples of mammograms that we consider in this paper, the dyadic levels $j = -7, -8, -9$ typically correspond to frequencies where microcalcifications occur. It is important not only to retain all microcalcifications but also their shape since these are important in diagnosis.

The remaining part of the mammogram corresponds to breast tissue and film noise. The breast tissue provides a very inhomogeneous background.

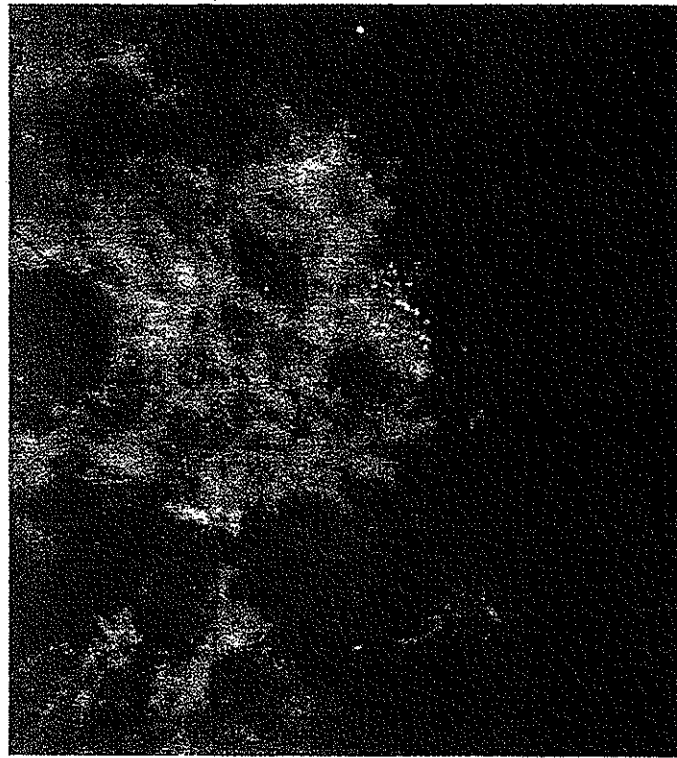


Figure 6.1
A typical mammographic image.

Also, the microcalcifications may be at low contrast to the background and this contrast may vary locally.

6.3 Compression and Noise Removal

While compression and noise removal are important for the storage and transmission of images, in the present paper we are interested in them as a preprocessor for the identification of microcalcification clusters in mammograms. The utilization of compression in this manner has many other related applications such as image registration and object recognition.

We first review briefly the elements of wavelet-based image compression. Figure 6.2 depicts the main steps in wavelet-based compression.

Step I. Computation of wavelet coefficients. We choose a univariate scaling function φ and represent the mammogram as in (2.1). We then use the FWT to change to the wavelet basis representation (2.4). At this stage of the compression algorithm, the file of wavelet coefficients exactly

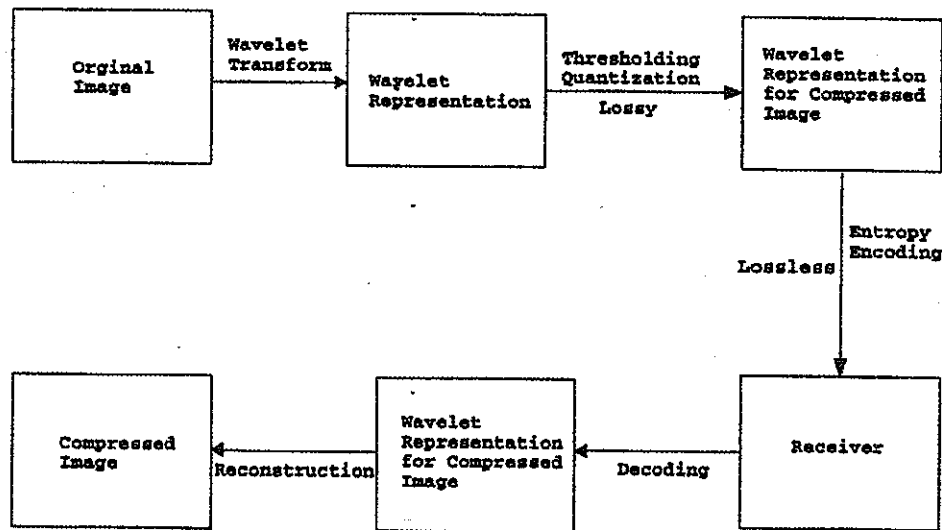


Figure 6.2
A schematic for wavelet-based compression

represents the original image.

Step II. Thresholding/Quantization. We gain lossy compression of the image by reducing the size of the wavelet coefficient file. It will be notationally convenient to speak only about the coefficients $d_{j,\eta}(k)$ of $\eta_{j,k}$. Similar statements apply to the coefficients of $\phi_{0,k}$.

There are two essential (related) methods for compression: thresholding and quantization. Thresholding means that we pick a threshold ϵ_j for each level $j = 0, -1, \dots, -m+1$ and retain only those coefficients whose absolute value exceeds ϵ_j . Thus, thresholding replaces $d_{j,\eta}(k)$ by $\tau_j(d_{j,\eta}(k))$ where the function τ_j is defined by

$$\tau_j(x) := \begin{cases} x, & |x| \geq \epsilon_j \\ 0, & |x| < \epsilon_j. \end{cases}$$

This is called *hard thresholding*; the function τ_j is not continuous. Soft thresholding would replace τ_j by (for example) the Lipschitz continuous function

$$\tau_j^o(x) := \begin{cases} x, & |x| \geq \epsilon_j \\ 2(|x| - \epsilon_j/2) \operatorname{sgn} x, & \epsilon_j/2 \leq |x| < \epsilon_j \\ 0, & |x| < \epsilon_j/2. \end{cases}$$

The graph of τ_j^o is given in Figure 6.3. Soft thresholding is numerically stable; a small change in $d_{j,\eta}(k)$ results in a small change in the output $\tau_j^o(d_{j,\eta}(k))$.

Thresholding will replace small wavelet coefficients by zero. This not only has the desired effect of compression but also removes noise. There

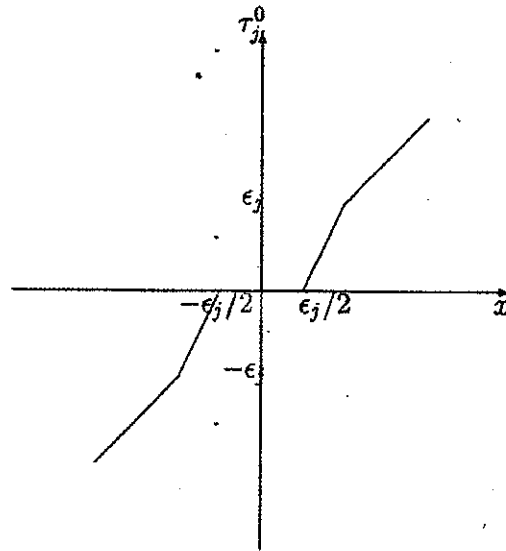


Figure 6.3
The graph of the soft thresholding function τ_j^0 .

are several results which show that (soft) thresholding, with a proper choice of thresholding parameters, gives an optimal (or near optimal) algorithm for removing Gaussian noise. There are two approaches here. One is statistically based as developed by Donoho and Johnstone (see, e.g., [8]) and the other is functional analytic as developed by DeVore and Lucier [7].

Quantization is a method for obtaining more compression in the file of wavelet coefficients. The idea is to economically allocate the number of bits to represent the coefficient $d_{j,\eta}(k)$. Namely, the coefficients $d_{j,\eta}(k)$ is replaced by the number $d'_{j,\eta}(k)$ which satisfies

$$|d_{j,\eta}(k) - d'_{j,\eta}(k)| < \epsilon_j$$

and has the fewest number of bits in its binary representation. Note that $\tau_j(d_{j,\eta}(k) - d'_{j,\eta}(k)) = 0$ which is the motivation for quantization.

The thresholding/quantization parameters ϵ_j can be related to the metric in which the approximation takes place. We have normalized the wavelets $\eta_{j,k}$ to have L_2 -norm equal to one. Taking $\epsilon_j = \epsilon$ (i.e., thresholding does not depend on the dyadic level) corresponds to approximation in the L_2 metric. On the other hand, the choice $\epsilon_j := 2^{(1/2-1/p)2j}\epsilon$ corresponds to approximation in L_p , $0 < p \leq \infty^3$. We refer the reader to [5, 6], where a more detailed discussion of the connections between thresholding/quantization and approximation is given.

³The distance between two functions g, h in L_p is given by $(\int |g - h|^p)^{1/p}$.

Thresholding in the metric L_1 produces smoother images while thresholding in L_∞ emphasizes edges. We can emphasize edges even more by going to derivative norms. These correspond to increasing ϵ_j at high frequencies and thereby increasing the likelihood that a coefficient at a high frequency is retained.

Step III. Encoding. For the purposes of storage or transmission of the compressed image (and perhaps for further compression), a lossless encoder is applied to the file of wavelet coefficients. While standard arithmetic and runlength encoding can be applied, the best results are obtained with customized encoders (see, e.g., Shapiro [11]) which take into account the spatial correlation of the wavelet coefficients. The encoded wavelet coefficient file is our compressed representation of the image and can be stored or transmitted. Encoding will not play a role in our microcalcification detection, since in this application we are not interested in the storage or transmission of the mammograms.

The last two steps of the image compression schematic are utilized when we want to display the compressed image.

Step IV. Decoding. The encoded file is decoded to obtain the compressed wavelet coefficient file. This is the same file as at the end of Step II.

Step V. Computation of pixel values. We use the inverse fast wavelet transform to compute the pixel values of the compressed image. This step takes again $O(N)$ operations. These are then the pixel values of the compressed image which is our approximation to the original image.

6.4 Some Issues in Compression Algorithms

There are many important considerations in compression algorithms and we mention only a few that relate to the problems of feature extraction and edge detection.

6.4.1 Choice of Wavelet Basis

The performance of compression algorithms depends on the choice of wavelet basis. It is generally agreed upon that the bi-orthogonal wavelets (when combined with a customized encoder [11]) give the best compression with a fixed wavelet basis of the type (2.2). However, this does not necessarily carry over to feature classification, where the criterion for compression is not the visual quality of the compressed image but its effectiveness in the classification. Also, there are variants to the fixed basis given by (2.2) such

as hyperbolic wavelet bases and adaptive bases (where the basis is allowed to depend on the image (see [4]).

6.4.2 Choice of Metric

We have noted earlier the effect of the choice of the metric on the thresholding/quantization strategies. In feature classification we have the option of tailoring the compression metric to the feature classification problem.

6.4.3 Level of Compression

We can also adjust the level of compression (compression ratio) to the task at hand. In compression for the purposes of storage or transmission, the visual quality of the image is the driving force behind selecting the compression ratio. However, in certain applications such as image registration, one may want to retain very few features (edges, corners) and therefore utilize high compression. In our application, we use compression primarily to remove noise.

6.5 Algorithms for the Detection of Microcalcification Clusters

We shall now propose a general strategy for constructing algorithms for extracting microcalcification clusters. We wish to distinguish between two types of algorithms. The first are autonomous and do not involve a computer operator. The second are interactive and allow decisions to be made to improve parameter settings which depend on the image. We have applied both approaches to a population of eleven mammograms — nine with microcalcification clusters and two without. The numerical results given in this paper are obtained from an autonomous algorithm. We shall point out where interactive algorithms have provided some improvement over the autonomous algorithm.

Both approaches have the same major steps which we now describe.

Step 1. Compression of the digitized image. An algorithm begins by utilizing compression as a preprocessor to the identification of the microcalcification clusters. The purpose of this step is to remove noise and still retain all microcalcifications. Our earlier results on compression of mammograms show that compression up to 50-1 is viable for this purpose. We have found, however, that a moderate compression (on the order of 10 or 15 to 1) has given the best performance on our limited population of examples.

While the choice of compression metric can be used in interactive algorithms to maximize performance, we have found that compression in the metric of L_2 gives quite satisfactory results. We have implemented compression with various wavelet filters. The six-tap Daubechies filter (corresponding to D_3) and certain bi-orthogonal filters give the best performance of fixed wavelet bases in that they retain all microcalcifications and maintain their shapes as well. We have found the use of hyperbolic wavelets and adaptively chosen wavelet bases to be promising and deserving of further analysis.

The autonomous algorithm used for the numerical examples in this paper are all based on the six-tap Daubechies filter with compression of approximately 13-1 in the L_2 metric.

Step 2. Selecting only high frequency terms. The purpose of this step is to retain only those frequency terms that correspond to the size of the microcalcifications. From the terms retained in the compressed image after step one, we delete all terms corresponding to dyadic levels 0, -1, ..., -5. Other possible adaptations of this selection are to retain only those terms at levels $j = -6, -7, -8$ corresponding to one or more of the dilated wavelets corresponding to the functions $\varphi(x)\psi(y)$, $\psi(x)\varphi(y)$, or $\psi(x)\psi(y)$ from Ψ . The first two of these choices emphasize edges in the vertical and horizontal directions while the latter emphasizes point singularities. Since microcalcifications are small, the use of $\psi(x)\psi(y)$ might seem preferred. Our numerical experiments show, however, that it is best to retain terms from all three of these wavelet functions.

Step 3. Reconstructing pixel values. This step constructs the image corresponding to the wavelet decomposition after Step 2. While the microcalcifications are very noticeable after this step, the image still has a nontrivial background which will be removed in the next steps.

Step 4. Nonlinear enhancement. The purpose of this step is to remove background and retain only the microcalcifications. Let us denote by \bar{p}_j the pixel values of the reconstructed image in Step 3. We shall modify these pixel values to obtain the new pixel values p_j^* . The new values p_j^* depend not only on \bar{p}_j but also on p_j :

$$p_j^* := E(p_j, \bar{p}_j)$$

where E is some nonlinear enhancement function.

The properties we want for E are the following. If \bar{p}_j is small then p_j^* should be set to zero. We would also like to enhance \bar{p}_j in a way that depends on the background brightness of the original pixel values. That is, if \bar{p}_j is large and the average intensity of the original pixel values p_ν for ν near j is also large, then we would like p_j^* to be an increase of the value

of \bar{p}_j . Among all steps in this general algorithm, this step and the next benefit most from interaction.

In our autonomous algorithm, we have utilized the following enhancement function

$$E(p_j, \bar{p}_j) := \begin{cases} 0, & p_j \leq 60 \\ \bar{p}_j, & 60 < p_j \leq 300 \\ \bar{p}_j + 3\alpha_j, & p_j > 300. \end{cases}$$

The local intensity α_j is defined as

$$\alpha_j := \max_{\nu \in \Lambda_j} \{p_\nu / 256\},$$

where Λ_j is a set of indices near j . In our autonomous algorithm, we take Λ_j to be a 5×5 square array centered at j .

Step 5. Thresholding pixel values. This step thresholds pixel values to retain only the most intense. In the autonomous algorithm this is obtained by retaining only the 300 largest pixel values p_j^* . We denote by $p_j^\#$ the pixel values after this step has been completed. We call the $p_j^\#$ *hot pixel* values. The hot pixels are indications of microcalcifications. However, further processing is necessary to be sure that the entire microcalcification is retained intact and also to eliminate hot pixels that are not part of microcalcification clusters.

Step 6. Removing isolated pixels. This step is intended to remove any isolated hot pixels. A microcalcification should correspond to several hot pixels. In our autonomous algorithm, we remove isolated hot pixels as follows. For each hot pixel $p_j^\#$, we create a 3×3 square of pixels centered at j . If no pixel in this square other than $p_j^\#$ is hot, then we remove $p_j^\#$ from the hot list.

Step 7. Filling out microcalcifications. The purpose of this step is to fill out microcalcifications. It could happen that the thresholding in Step 5 removed pixels which corresponded to microcalcifications and we want to restore these. In our autonomous algorithm, we do this as follows. If $p_j^\#$ is a hot pixel, we form a 3×3 square of pixels centered at j and we consider a ν from this square. If p_ν (the original pixel value corresponding to this location) is greater than $p_j - 5$ and if $p_\nu^\#$ was not one of the hot pixel values, then we change the value of $p_\nu^\#$ from zero to $p_j^\#$ and add it to our list of hot pixels.

Step 8. Test for clusters. Finally, we want to test for clusters. Given a hot pixel value $p_j^\#$, we examine all hot pixels in a 61×61 square of pixels centered at j . If $p_\nu^\#$ is a hot pixel from this square and $|p_\nu^\# - p_j^\#| < 10$, then we say $p_\nu^\#$ is connected to $p_j^\#$ and write $p_\nu^\# \sim p_j^\#$. We say that $p_j^\#$ is related to $p_j^\#$ if there is a sequence $p_{\nu_k}^\#$, $k = 0, \dots, m$, of hot pixel values

with $\nu_0 = j$ and $\nu_m = j'$ and $p_{\nu_k}^\# \sim p_{\nu_{k+1}}^\#$, $k = 0, \dots, m-1$. We then choose the smallest rectangle which contains all pixel values $p_{\nu_k}^\#$, which are related to $p_j^\#$. In this way, we obtain boxes of hot pixel values which correspond to boxing our microcalcification-clusters.

6.6 Examples

In Figures 6.4 and 6.5, we give two examples of the autonomous algorithm applied to digitized mammograms which exhibit microcalcification clusters. The upper image in each of these figures is the processed mammogram (after Step 8 has been completed). The lower image is the original mammogram with the clusters boxed.

We have also included an example of the autonomous algorithm applied to a digitized mammogram without microcalcification clusters. Of course, the final image (after Step 8) is completely black (all pixel values $p_j^\# = 0$).

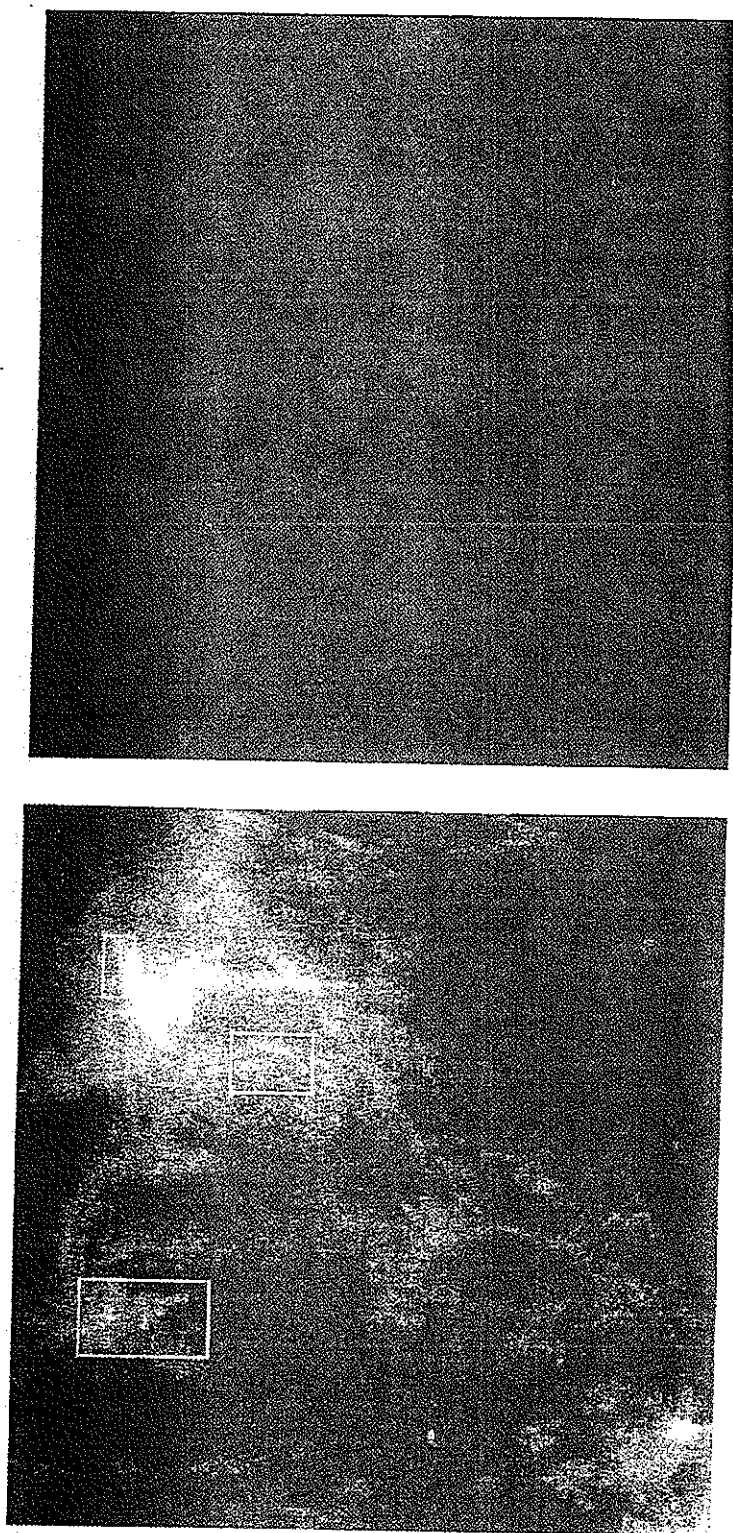


Figure 6.4
The autonomous algorithm applied to a mammogram with microcalcification clusters. Top is the processed image. Bottom is the original image with microcalcification clusters boxed.

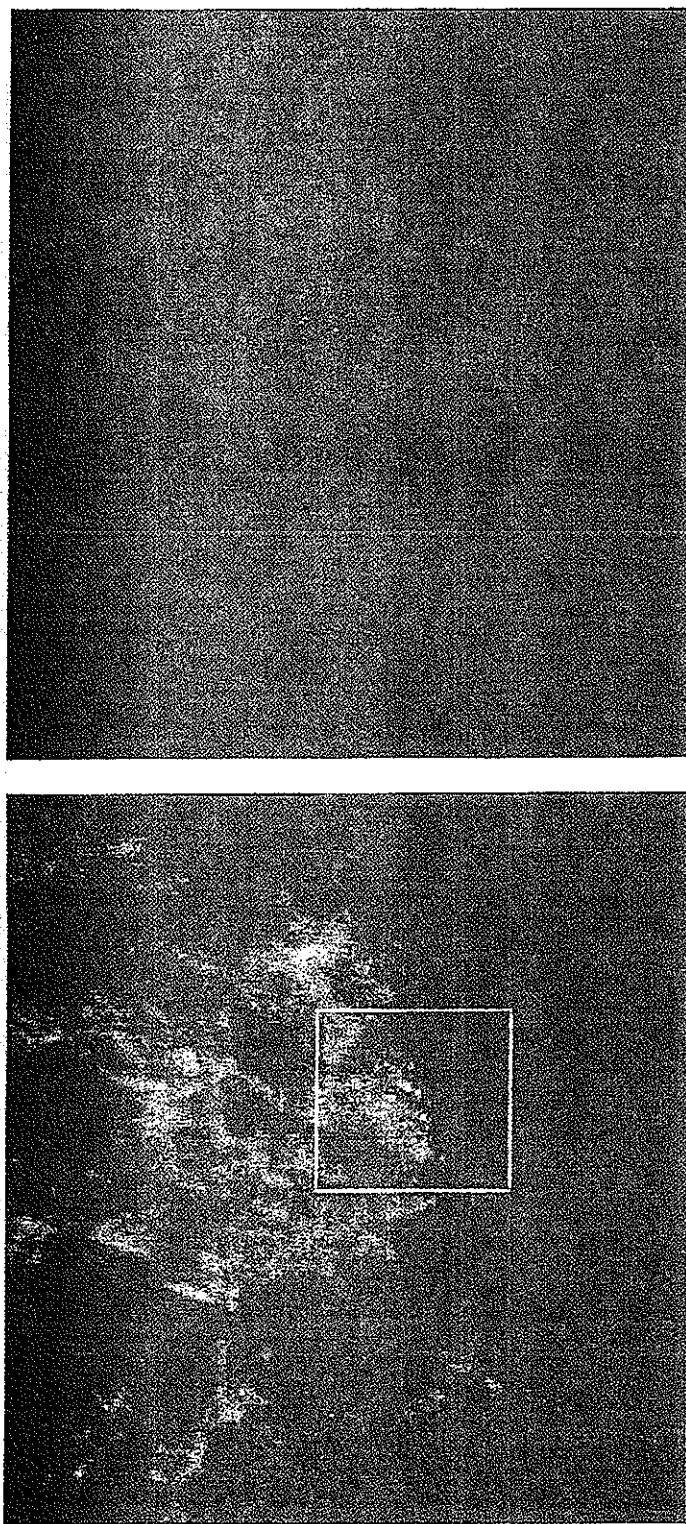


Figure 6.5

The autonomous algorithm applied to a mammogram with microcalcification clusters. Top is the processed image. Bottom is the original image with microcalcification clusters boxed.

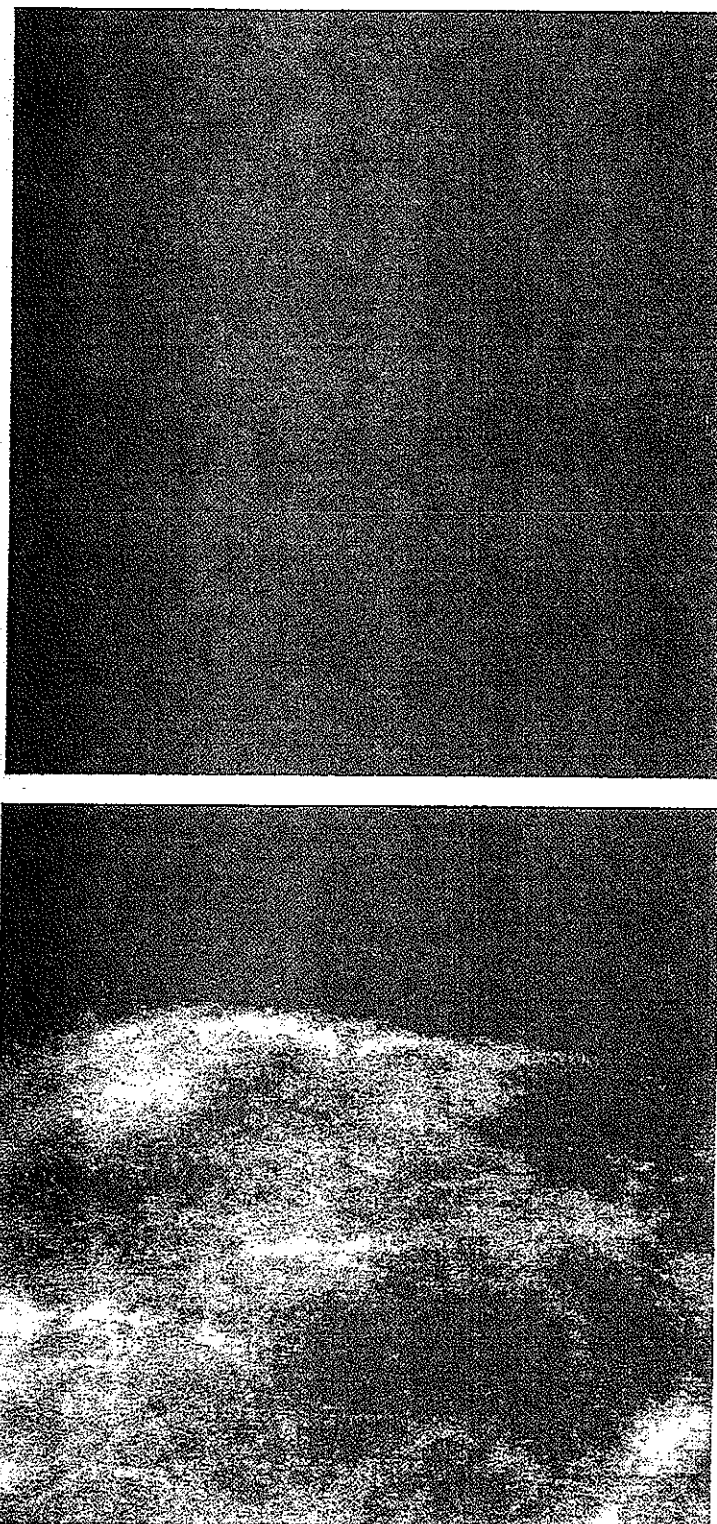


Figure 6.6
The autonomous algorithm applied to a mammogram without microcalcification clusters. Top is the processed image. Bottom is the original image.

References

- [1] "Cancer facts & figures", Technical report, American Cancer Society, 1991.
- [2] A. Cohen, I. Daubechies, and J. C. Feauveau, Biorthogonal bases of compactly supported wavelet, *Commun. Pure Appl. Math.* 45 (1992), 485-560.
- [3] J. Dengler, S. Behrens, and J. F. Desaga, Segmentation of microcalcifications in mammograms, *IEEE Trans. Med. Imaging* 12 (1993), 634-642.
- [4] R. DeVore, Adaptive wavelet bases for image compression, *Curves and Surfaces, II*, P. J. Laurent, A. Le Méhauté, and L. L. Schumaker, eds., A. K. Peters, Boston, 1994.
- [5] R. DeVore, B. Jawerth, and V. Popov, Compression of wavelet decompositions, *Am. J. Math.* 114 (1992), 737-785.
- [6] R. DeVore, B. Jawerth, and B. Lucier, Image compression through transform coding, *IEEE Proc. Inf. Theory* 38 (1992), 719-746.
- [7] R. DeVore and B. Lucier, Fast wavelet techniques for near-optimal image processing, *1992 IEEE Military Communications Conference*, IEEE Communications Society, 1992, 1129-1135.
- [8] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, *Biometrika* 81 (1994), 425-455.
- [9] B. Lucier, M. Kallergi, Wei Qian, R. DeVore, R. Clark, E. Saff, and L. P. Clarke, Wavelet compression and segmentation of mammographic images, *J. Digital Imaging* 7 (1994) 27-38.
- [10] Laszlo Taber and Peter B. Dean, *Teaching Atlas of Mammography*, Georg Thieme Verlag, Stuttgart, 1985.
- [11] Jerome Shapiro, An embedded hierarchical image coder using zero-trees of wavelet coefficients, *Data Compression Conference*, J. A. Storer and M. Cohn (eds.), IEEE Computer Society Press, Los Alamitos, CA, 1993, 214-223.
- [12] Z. Yang, M. Kallergi, R. DeVore, B. Lucier, W. Qian, R. A. Clark, and L. P. Clarke, The effect of wavelet bases on compression of digital mammograms, *IEEE Trans. Med. Imaging* (to appear).

- [13] Z. Yang, Wavelets and Image Compression, *Ph.D. thesis*, the University of South Carolina, Columbia, 1995.