INFORMATION TO USERS

WAVELET TRANSFORMS FOR DETECTING

MICROCALCIFICATIONS IN MAMMOGRAPHY

by

Hee Il Hahn

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 9 5

UMI Number: 9622990

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have

read the dissertation prepared by_Hee Il Hahn

entitled Wavelet Transforms for Detecting Microcalcifications

in Mammography.

and recommend that it be accepted as fulfilling the dissertation

requirement for the Degree of Doctor of Philosophy

Dr. Robin N. Strickland   Date   9-11-95

Dr. Lawrence P. Huelsman   Date   9-11-95

Dr. Jeffrey J. Rodriguez   Date   9-11-95

Dr. Jerry Moloney   Date   9-11-95

                                                          Date

Final approval and acceptance of this dissertation is contingent upon
the candidate's submission of the final copy of the dissertation to the
Graduate College.

I hereby certify that I have read this dissertation prepared under my
direction and recommend that it be accepted as fulfilling the dissertation
requirement.

_____   9-11-95
Dissertation Director   Date
Dr. Robin N. Strickland

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: *Neil Hahn*

ACKNOWLEDGEMENTS

To my parents

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF TABLES

ABSTRACT

Clusters of fine, granular microcalcifications in mammograms may be an early sign of disease. A three-stage method based on wavelet transforms for detecting and segmenting calcifications is developed. The first stage consists of preprocessing which is realized by subtracting from the original image a Gaussian low-pass filtered version, followed by a full resolution wavelet transform. Four octaves are computed with three inter-octave voices for finer scale resolution. With a Laplacian of Gaussian wavelet basis the detection of microcalcifications in the relevant size range can be nearly optimized in the details sub-bands. In fact, the separable two-dimensional filters which transform the input image into HH details sub-bands are closely related to prewhitening matched filters for detecting Gaussian objects (idealized microcalcifications) in separable first-order Markov (background) noise. Meanwhile, the addition of LH and HL details sub-bands approximates the prewhitening matched filtered output in nonseparable first-order Markov noise.

Two methods have been proposed to detect candidate microcalcifications. In the first method, the outputs HH and LH + HL from each octave are thresholded at some fixed percentile of the histogram of each component. Then, the detected images from all octaves are logically ORed to yield the binary map of detected pixels. The second method employs a Hotelling observer. The Hotelling discriminant is computed and then thresholded to obtain the binary map.

The second stage is needed to reduce false alarms created by edges and other linear structures of normal breast tissue. False alarms are reduced by analyzing the shapes of the detected pixel regions. The third stage is designed to overcome the limitations of the simplistic Gaussian assumption and provides an accurate segmentation of calcification boundaries. Individual microcalcifications are often greatly enhanced in the output image.

FROC curves are computed from tests using a well-known (and freely distributed) database of digitized mammograms. The algorithm using the Hotelling observer shows the best overall performance in which a true positive fraction of 73% is achieved at 0.7 false positives per image.

CHAPTER 1

Introduction

Breast cancer is one of the most common types of cancers in middle-aged women. Although it can be treated in its early stages, the mortality rate is not decreasing. Breast cancer can be controlled by prevention, early diagnosis, and effective treatment. Unfortunately, primary prevention is very difficult because the cause of this disease is still unknown. However, it can be effectively treated by current methods if detected early. Therefore, treatment of the cancer while it is still in its early stages is the most promising way to reduce the mortality rate. Mammography is known to be the only breast screening technique currently capable of detecting breast cancer at the earliest possible stage.

There are two primary signatures used by the radiologist to discriminate between normal and cancerous tissue [1]. One is related to the mass, density and shape of breast tumors. A benign neoplasm is smoothly marginated whereas a malignancy is characterized by an indistinct border which becomes more spiculated with time. The classic example of malignancy is the stellate lesion which consists of a dense center surrounded by spicules radiating into surrounding tissue. The other is a microcalcification (μCa++) which occurs in intramammary and intradermal ducts. According to the imaging properties of breast lesions in mammography, the morphology of μCa++s

has been classified using five categories by one group [7], with descriptions such as *dustlike, wormlike, crushed stone,* and so on. In this research, the focus is only on μCa++s.

The early detection of small, non-palpable breast cancers is the aim of screening mammography, which is known to be the best method currently available for early detection of breast cancer. It has been reported that there are currently more than 50 million women over the age of 40 at risk of contracting breast cancer in the United States. If 50% of the women at risk are screened by mammographic examination each year, about 30,000 cases of breast cancer per year would be expected to be found by the presence of μCa++s (considering about 20 - 21% of all cancers are found by the presence of μCa++s). An early sign of disease in 30 - 50% of mammographically detected cases is the appearance of clusters of fine, granular μCa++s whose individual grains typically range in size from 0.05 - 1mm in diameter [49]. Individual μCa++s are difficult to detect because of variations in their shape and size and because they are embedded in and camouflaged by varying densities of parenchymal tissue structures. Indeed, recent estimates indicate that between 10 and 30% of breast lesions (of all types) are missed by radiologists during routine screening [56]-[60].

If, as stated earlier, 25 million women are screened each year, the number of mammograms to be processed will be

$$25,000,000 \ patients \times 4 \ views/patient = 100,000,000 \ views/year, \tag{1.1}$$

since 4 views are usually needed for mammographic examination. In addition, if we assume $100\mu m$ resolution on film, the storage requirement will be

$$2,048 \times 2,048 \; pixels/view \times 12 \; bits/pixel \times 100,000,000 \; views$$
$$= 5,033,165 \; Gbits/year. \tag{1.2}$$

So, more than 625 thousand 1 Gbyte-disks will be needed each year.

Since at present, mammogram readings are performed by experts who look for the presence of abnormalities that may be interpretated as cancerous changes, the interpretation of a large volume of images will be a major problem. In addition, it is difficult for radiologists to maintain interest in interpreting large numbers of images in which only a small number show abnormalities. Hence, computer aided diagnosis (CAD) systems are needed to use mammography in mass screening programs. CAD can partially alleviate some of the problems mentioned above.

CAD schemes using digital image processing techniques have the goal of improving the detection performance and throughput of screening mammography. In screening programs, where more than 99% of the images are normal, automated detection of $\mu Ca{+}{+}s$ can improve the radiologist's performance to detect abnormalities. Typically, CAD systems are designed to provide a *second opinion* to aid rather than replace the radiologists.

1.1 Previous Work

Many researchers have tried to detect μCa++s using image processing techniques developed for image enhancement, object location, and pattern recognition. Chan *et al.* [5] use a differencing technique in which a signal-suppressed image is subtracted from a signal-enhanced image to suppress unwanted background texture. Signal enhancement is achieved using an approximation to a matched filter for the μCa++s. Contrast, size, and clustering criteria are used to reduce the false positive rate. Several authors have proposed a number of variants of the matched filter [2]-[4]. In later work, Chan *et al.* [6] include a power spectrum feature for improved discrimination. This work is important in that ROC (receiver operating characteristic) tests showed that CAD can improve the performance of radiologists. The image processing algorithms still produce false positives, although in CAD these are overruled by the radiologist. Fam *et al.* [8] describe a two-part algorithm for detecting clustered calcifications. The first stage identifies individual calcifications using a region-growing procedure which groups neighboring pixels according to similarity in average gray level and the variance of pixels contained in connected structure. Size, gray level, contrast and gradient criteria are applied to the segmented regions and only those falling within selected ranges are positively identified as calcifications. A second stage passes only those potential calcifications which occur in clusters of three or more. One drawback of the first stage is the number of free parameters which must be selected

in order to compensate for the intensity and contrast of the original film. A method based on the same philosophy is proposed by Davies and Dance [9]. Here, local area thresholding is used for segmentation of small, bright, suspicious regions followed by feature analysis and cluster identification. Once again the feature analysis (classification) stage employs specific gray level and contrast values, which are inherently sensitive to imaging conditions. This method tries to emphasize the suspicious regions rather than the individual μCa++s. Strickland [28]-[31] employs a Student's t−test in a curvature feature space and compares its performance with that of pre-whitening matched filter. Karssemeijer [10] also employs local thresholding but uses a Bayesian setting to distinguish elongated positives from faint calcifications. Dhawan [50] has developed an adaptive neighborhood image processing technique to enhance the low contrast of specific features. Nishikawa *et al.* [51] employ an image differencing technique to increase the signal-to-noise ratio of μCa++s. Signal extraction is acomplished by applying a global gray-level thresholding, morphological erosion, and local adaptive gray-level thresholding. Texture analysis, comparison of the size of a μCa++ with its radiation contrast, and spatial distribution of signals within the clusters are included to reduce the false-positive rate. However, small calcifications whose diameter is less than $0.2mm$ can not be detected because the spatial resolution used is $0.4mm$. Dengler *et al.* [52] use a two-stage algorithm for spot detection and shape extraction. After high-pass filtering which is realized by subtracting from the original image a Gausian low-pass filtered image, a weighted difference of Gaussians

filter is applied. Then they try to reproduce the shape of the spots using morphological filtering. The weighted difference of Gaussians can be thought of as a variant of the prewhitening matched filter. The preprocessing operation is basically the same as the weighted difference of Gaussians filtering at different scales. In this respect, the first stage can be modeled as a repeated application of the filter.

Surprisingly few authors address the problem of detecting μCa++s of varying size. An important branch of CAD methods in mammography employs wavelet transforms for feature enhancement. The general approach is : (1) Compute the forward wavelet transform of the image; (2) Nonlinearly transform or adaptively weight the wavelet coefficients; (3) Compute the inverse wavelet transform.

Richardson [14] showed that the details components of the lowest octaves (*i.e.* those containing the finest structure), which are the difference between successive levels in the decomposition, can be useful in enhancing the visibility of μCa++s in mammograms. Laine and Song [15] also use wavelet transforms to enhance μCa++s. The wavelet in this case is the first derivative of Gaussian, whose relative is the well known Canny edge operator [23]. Mammograms are enhanced scale-dependently by selectively weighting and scaling the details images. In later work [61], a dyadic wavelet transform method is shown to be equivalent to unsharp masking at multiple scales. Seaton and Abbott [17] compare multiscale morphological decompositions with Gaussian and Laplacian pyramids. Structuring elements are repeatedly applied

to remove high frequency details in the image. A useful scale-dependent segmentation can be obtained by carefully matching the structuring element to the type of objects being sought. Strickland and Hahn [31] apply an adaptive matched filter to the images formed in a Laplacian pyramid, each of which contains a narrower range of μCa++ scales. Barman *et al.* [32] represent multiresolution images using a Gaussian pyramid created by quadrature filters to produce estimates of orientation, Fourier phase and energy. The outputs from multiscale quadrature filters are processed by specialized object extraction algorithms employing size, orientation, relative brightness and shape features. Clusters of μCa++s are tagged by examining and detecting likely combinations of features in the object list. Qian *et al.* [53] enhance μCa++s by cascading adaptive spatial filtering with multiresolution decomposition and reconstruction. They try to enhance details subimages using tree structured nonlinear filters with fixed parameters and adaptive order statistic filters. Yoshida *et al.* [54] improve the visibility of μCa++s by reconstructing mammograms from selected subbands. Post-detection using morphological operators and thresholding achieves a sensitivity of 85% true clusters detected at 5 false positive clusters per image when tested on a 100μm resolution database of 39 mammograms. The authors conclude that wavelet techniques are especially useful for detecting subtle μCa++s when used to augment existing CAD algorithms.

1.2 Proposed Approach

In mammograms, the μCa++s we want to recognize come in various sizes. The size of the neighborhood where analysis is performed should be adapted to the size of the μCa++s. Hence, it is not possible to define a priori an optimal resolution for analyzing images. We can solve this problem by processing the image at different resolutions. Until now, many multiresolution signal analysis techniques have been developed. Among these, the wavelet transform is known to be the best method which represents a signal at multiple scales. We can interpret the wavelet transform as a signal decomposition onto a set of basis functions called *wavelets*, which are obtained from a single prototype wavelet by dilation and translation. If we select wavelets which are close to the impulse response of a matched filter for detecting μCa++s, then the wavelet transform can be considered to be a bank of multiscale matched filters for detecting μCa++s. A wavelet-based algorithm detecting μCa++s of varying size is proposed. This has evolved from our earlier work on multiscale tumor detection in gamma-ray images [28]-[31]. The idea is to use the wavelet transform to decompose a mammogram into a series of band-pass-filtered images of increasingly coarse detail. The filters which transform the mammogram into the HH and (summed) LH + HL sub-bands are in fact matched filters for detecting idealized μCa++s in separable and nonseparable background noise. Since actual mammogram texture is well modeled by a combination of these noise types the combination of HH and LH + HL is nearly

optimum. Detecting objects in a particular size range is accomplished by processing these sub-bands or *details* images at the appropriate scale. The details images contain μCa++s and other high frequency structures.

Two methods have been proposed to detect the candidate μCa++s. In the first method, the outputs HH and LH + HL from each octave are thresholded at some fixed percentile of the histogram of each component. Then, the detected images from all octaves are logically ORed to yield the binary map of detected pixels. The second method employs a Hotelling observer. Pixels from HH, LH + HL in 7 sub-bands comprise a 7-dimensional feature vectors. Several hundreds of feature vectors extracted from the background and manually segmented μCa++s enable us to compute the Hotelling observer. The Hotelling discriminant is computed and then thresholded to obtain a binary map. In practice, the binary map from either method contains many false alarms as well as real μCa++s because of the inhomogeneous mammogram texture. Object shape is used to reduce the number of false alarms. For the wavelet basis used, the LH + HL sub-band filters are very close to Laplacian of Gaussian, enabling us to extract edges derived from the zero-crossings in those sub-bands. To discriminate betwen μCa++s and false alarms, we first search for a precise closed boundary at each candidate μCa++ site. The resultant objects are referred to as true positives. Although the boundaries of some readily visible μCa++s are closed, many of them have broken or disconnected boundaries. These can be detected by computing a circularity measure such as a Hough transform and a phase distribution.

Thus, highly elongated regions are penalized, and many of the false alarms caused by long strands of connective tissue can be eliminated. Segmentation of μCa++s is realized by weighting the sub-bands at these final detected sites before computing the inverse wavelet transform. Depending on which sub-bands of the forward wavelet transform are included in the inverse operation, the reconstructed image may consist of detected pixels (*i.e.*, μCa++s) only, or detected pixels superimposed on the original mammogram. The basic wavelet transform is implemented at full resolution in each octave level, with additional *voices* inserted between octaves to provide improved scale resolution.

Chapter 2 describes the mammogram images used in this research. The characteristics of μCa++s are investigated and it is shown the texture of mammograms can be modeled as a combination of separable and nonseparable Markov processes. The criteria used for detection of clustered μCa++s are explained although there are no universally accepted criteria. The subject of Chapter 3 is matched filters for detecting μCa++s. First, the power spectra of separable and nonseparable Markov noise are computed. Then, the prewhitening matched filters for detecting μCa++s buried in such backgrounds are derived. Chapter 4 provides an introduction to the wavelet theory. It is shown in Chapter 5 that the matched filter for detecting μCa++s can be implemented using the wavelet transform if the Laplacian of Gaussian wavelet is selected. The detection performance of the wavelet transform is compared with that of the matched filter by computing ROC curves using simulated Gaussian objects

embedded in both separable and nonseparable Markov noise. Chapter 6 describes a nonlinear algorithm for detecting μCa++s. Two methods proposed to detect the candidate μCa++s are explained in detail. Given the candidate μCa++ sites, their shapes are analyzed to reduce false alarms. The intermediate results at each step of the algorithm are also shown. Experimental results from the nonlinear algorithm are described in Chapter 7. Performance resulting from tests on a set of 40 digital mammograms is presented in the form of a free-response receiver operating characteristic (FROC) curve. The FROC curve derived from matched filters is compared with that of the Hotelling observer and the performance gain obtained by applying the shape analysis is depicted. Examples of reconstructed calcifications are also shown.

CHAPTER 2

Data

2.1 Mammogram Database

Our μCa++ detection algorithms are tested on a well known and widely published database of 40 digitized mammograms created by N. Karssemeijer of University Hospital Nijmegen, The Netherlands [10][35]. The mammograms were recorded with a Kodak MIN-R/SO177 screen/film combination. Each image was digitized from film using an Eikonix 1412 CCD camera operating at a spatial resolution of 2048 $\times$ 2048 pixels per image with 12 bits per pixel, using a sampling aperture of 50 μm in diameter and a 100 μm sample spacing. Figure 2.1 shows a typical 2048 $\times$ 2048 mammogram. The 40 images represent 21 different patients. Each image contains one or more μCa++ clusters verified by expert radiologists and histology. The database contains a total of 104 clusters, many of which proved to be malignant. One pair of images, numbered 12o (oblique) and 12c (craniocaudal), contains a total of 28 clusters, which is approximately 27% of the total. The remaining images each contain an average of two clusters. Ground truth for each image consists of the x, y coordinates and the radius of a "truth circle" for each verified cluster. The locations of individual μCa++s are not marked. The visibility of the clusters is highly variable, which is rather typical of clinical cases.

In a preprocessing step, each image was gray-scale transformed using a noise-equalizing look-up table. Noise characteristics of digitized films are complex and depend strongly on the gray level itself. This problem can be solved by mapping the pixel values onto an iso-presion scale on which the absolute error of the pixel values due to film noise and digitization noise is equal over the whole range of pixel values [35]. A constant sensitivity over the whole range of pixel values can be obtained with a parameter setting which is independent of the grey level. The look-up tables are distributed with the database. A representative look-up table is provided in Figure 2.2. Figure 2.3 shows the mammogram after application of the noise-equalizing look-up table to the example in Figure 2.1.

In the original mammogram the breast occupies a relatively small portion of the image area. Hence, we crop the image to a sub-region of 1024 × 1024 pixels, still at full pixel resolution. Figure 2.4 shows a sub-region of Figure 2.3. All images in the database are likewise cropped.

2.2 Cluster Detection Criteria

At present no universally accepted criteria for the detection of clustered μCa++s have been agreed upon, although various research groups are currently attempting to assemble a common database of mammograms with ground truth and a set of standardized detection metrics. We adopt the detection criteria proposed by Karssemeijer [35]; namely that for counting true positives (TPs) a cluster is considered detected if
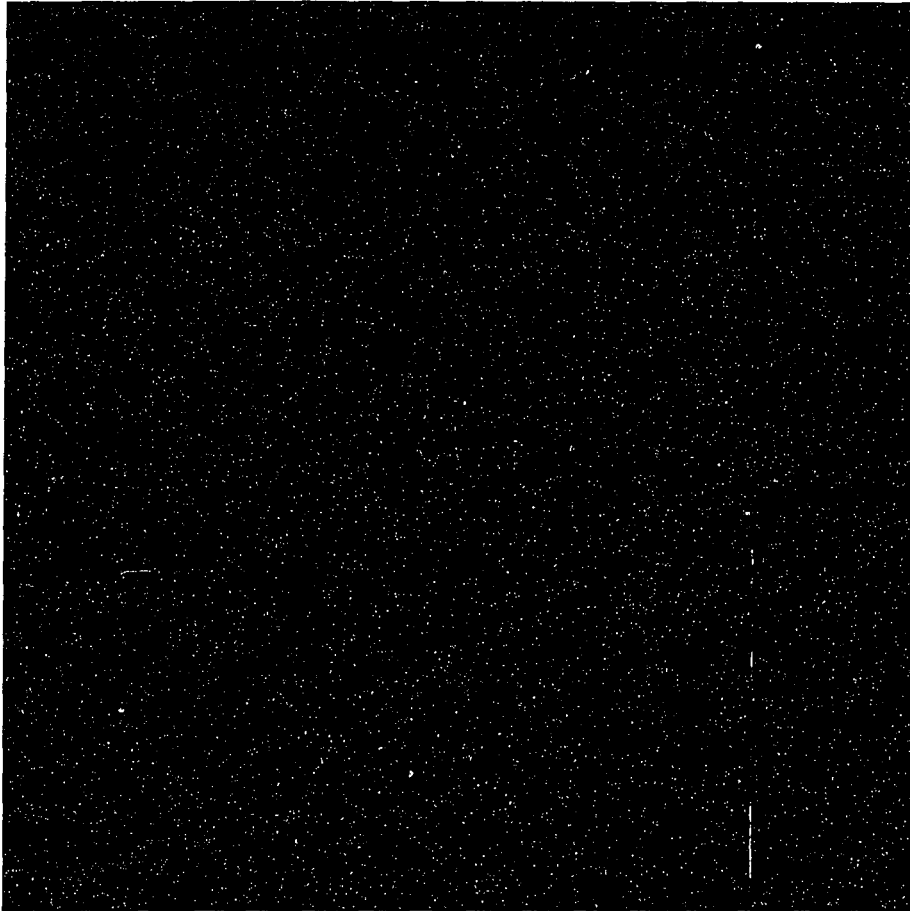
Figure 2.1: A typical 2048 × 2048 mammogram, identified as *c05c*. The breast occupies a relatively small portion of the image area.
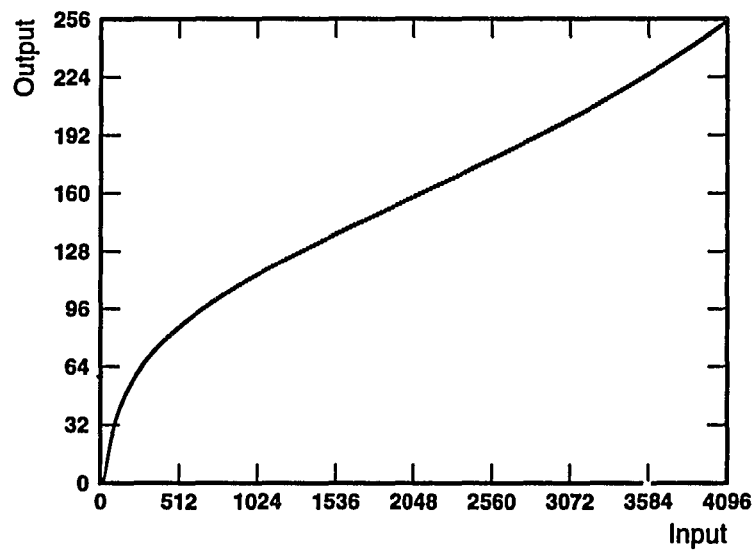
Figure 2.2: Look-up table for image c05c.
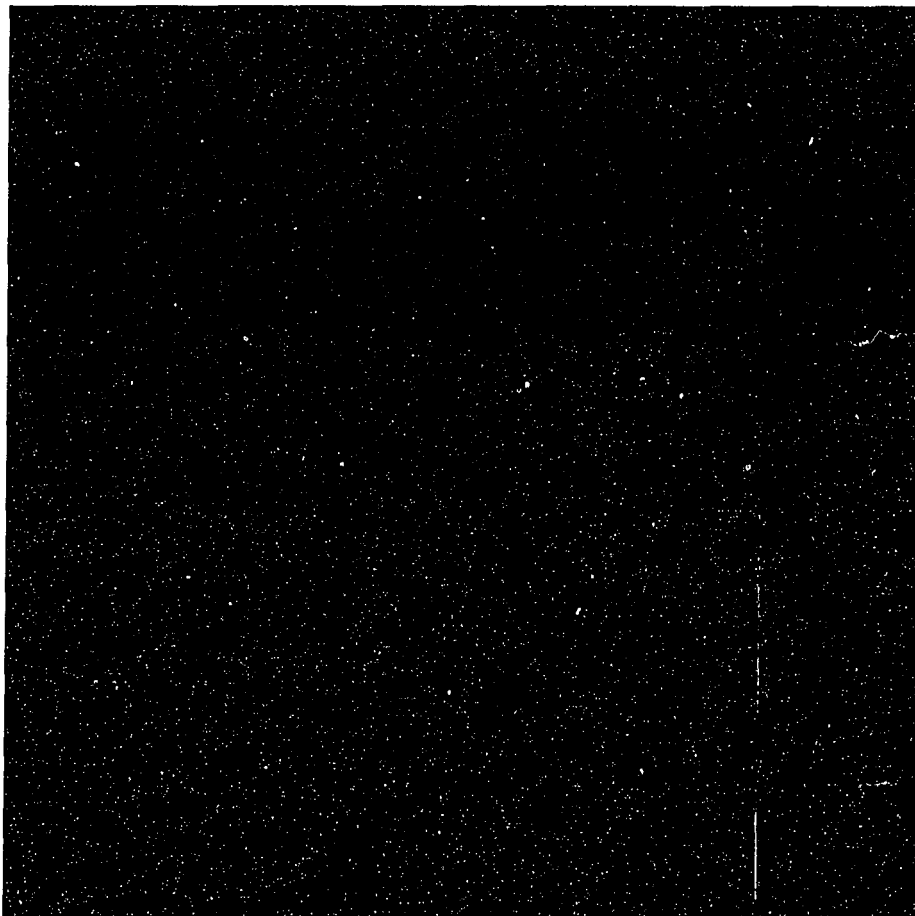
Figure 2.3: This is Figure 2.1 after application of the noise-equalizing look-up table.

two or more μCa++s are found in the region of film (truth circle) identified by the radiologist. A false positive (FP) is counted if two or more erroneous detections are made within an empty, closed region of 0.5 cm in width. One weakness of this metric is that at high sensitivity thresholds the number of false positives may actually drop due to merging of previously separated false clusters. However, in practice this occurs at false positive rates which are too high to be of any clinical value. The useful part of the performance curve lies below one or two false clusters per image.
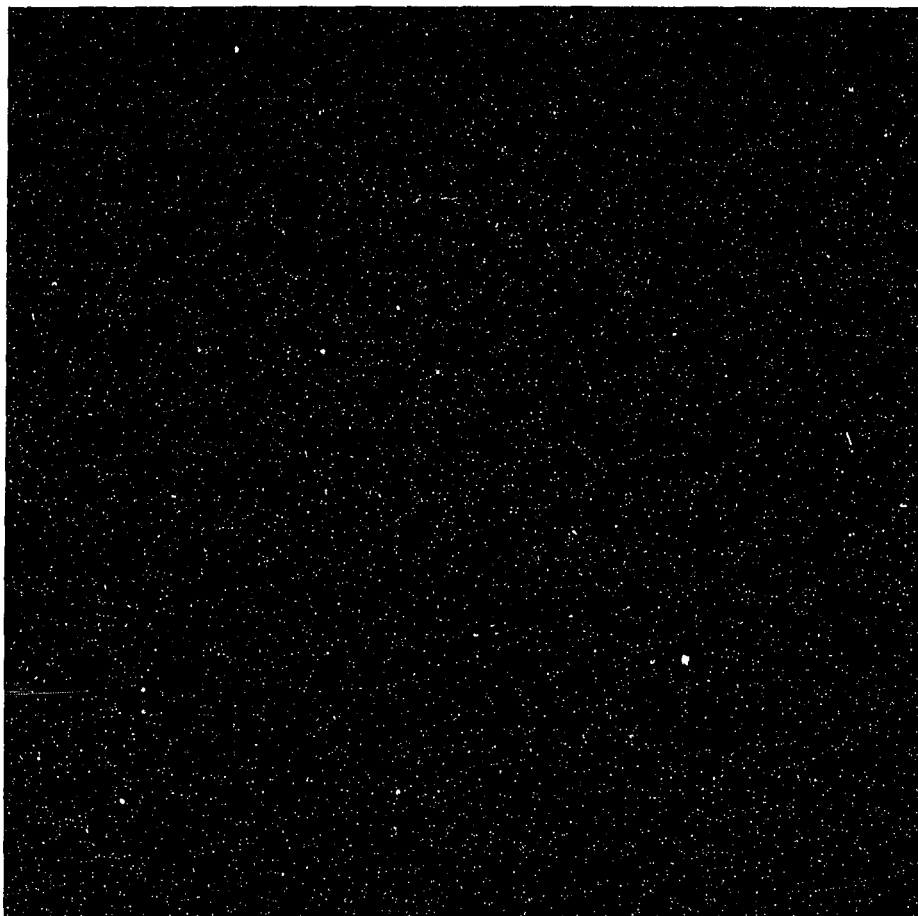
Figure 2.4: 1024 × 1024 Sub-region of Figure 2.2.
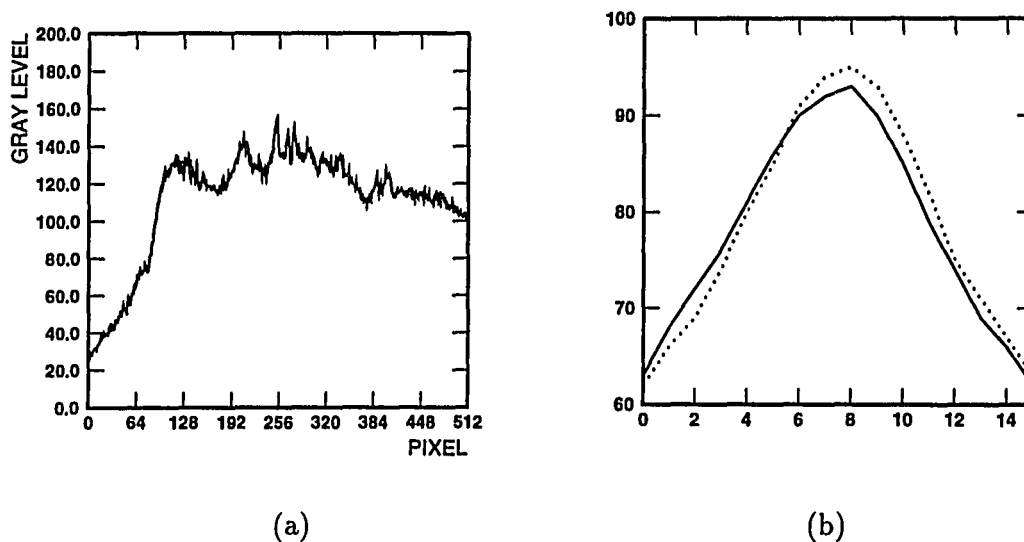
Figure 2.5: (a) Cross section of a prominent μCa++ cluster. (b) Average profile of 80 μCa++s, after normalizing for variations in height and width. The solid line and the dotted line represent the row and column profiles, respectively.

2.3 Characteristics of Microcalcifications

In mammograms, individual μCa++s usually appear as small (typically 0.05 - 1 mm), particulate objects of variable shape (from granular to rod-shaped), and fairly uniform optical density, which look like dusts or crushed stones. Although μCa++s vary in outline and degree of elongation, the *average* form is roughly circular, with a tapered cross-sectional profile. Figure 2.5(a) shows the cross section of a typical cluster with four prominent μCa++s. Figure 2.5(b) shows the average profile computed from 80 μCa++s selected at random from the test images. For a given size, it is not unreasonable to model a μCa++ using a circularly-symmetric Gaussian function (See Figure 2.5(b)). Another merit of the Gaussian model is that leads to separable

detection filters which are implementable via the wavelet transform. Since the Gaussian model represents an *average* profile, individual μCa++s may be modeled as two-dimensional Gaussians with noise. The wavelet transform can smooth high frequency noise to some degree, because it involves a low-pass filtering at each octave. In other words, in the original mammograms, real μCa++s may be very different from two-dimensional Gaussians, but at the higher octaves, the image is convolved with a low-pass filter which resembles a Gaussian and the high frequency components of the image are removed. Thus, μCa++s become smoothed to the two-dimensional Gaussian at higher octaves. For this reason, this model fits even better when the image is smoothed, which occurs when applying the wavelet transform.

Gaussian signals have few features and we have difficulties in detecting them because the mammogram background contains numerous Gaussian shaped features. Because of their simple shape, it is very difficult to find features which distinguish μCa++s from other structures and to discriminate μCa++s from such backgrounds. But, if we differentiate the image (equivalent to high-pass filtering), this has the effect of decorrelating the image and thus backgrounds with low frequency components are suppressed and μCa++s are enhanced. Although μCa++s may not be visible in the original mammogram, they may be apparent in the higher frequency bands when the image is decomposed into sub-bands. Gaussians with small variances have a broad spectrum and will appear in high frequency bands, whereas Gaussians with large variances will appear in low frequency bands. Thus, if the image is decomposed into

many sub-bands, Gaussians of varying size can be detected by examining several sub-bands.

2.4   Model of Mammogram Texture

The visibility of μCa++s is often degraded by the high frequency texture of breast tissue. For designing detectors, a statistical model of this background noise texture is required. A widely used stochastic model of images [47] consists of a nonstationary mean plus a stationary residual component modeled as either a *separable* Markov process with autocorrelation $r_{nn}(k, l) = \sigma_n^2 e^{-\alpha(|k|+|l|)}$, or a *nonseparable* Markov process with autocorrelation $r_{nn}(k, l) = \sigma_n^2 e^{-\alpha\sqrt{k^2+l^2}}$ [12]. The residual component is typically created by subtracting a low-pass filtered version of the image from the original image. The residual image, since it contains the high frequency details of the image, is a logical choice for representing mammogram texture. Figure 2.6 compares simulated textures with the residual texture computed from a sample mammogram. The nonseparable Markov texture is isotropic, which would appear to be better suited for breast texture, with its structures oriented in all directions. Nevertheless, the horizontal and vertical streaking generated by the separable model may be useful for representing the fine vasculature of the breast. In fact, as Figure 2.6 suggests, a combination of separable and nonseparable models may be closest to the truth. This will be explained in more detail in Chapter 3.

After subtracting the spatial mean from each mammogram, the Markov correlation parameter α of the residual process is estimated using [48] $\hat{\alpha} = -ln(\frac{\hat{r}_{nn}(0,1)}{\hat{r}_{nn}(0,0)})$, where $\hat{r}_{nn}(k,l)$ is the estimated autocorrelation of background texture. Values of $\hat{\alpha}$ computed from images in the Nijmegen database fall in the range $0.01 \rightarrow 0.2$.

In the next chapter the power spectra of the first-order separable and nonseparable Markov noise images are derived explicitly by Fourier transforming their autocorrelation functions. The estimated power spectrum computed from the 40 mammograms is modeled by combining the power spectrum of the first-order separable Markov noise with that of the first-order nonseparable Markov noise. Then, the matched filters for detecting μCa++s are computed, using the power spectra derived and by modeling μCa++s as two-dimensional Gaussian functions.
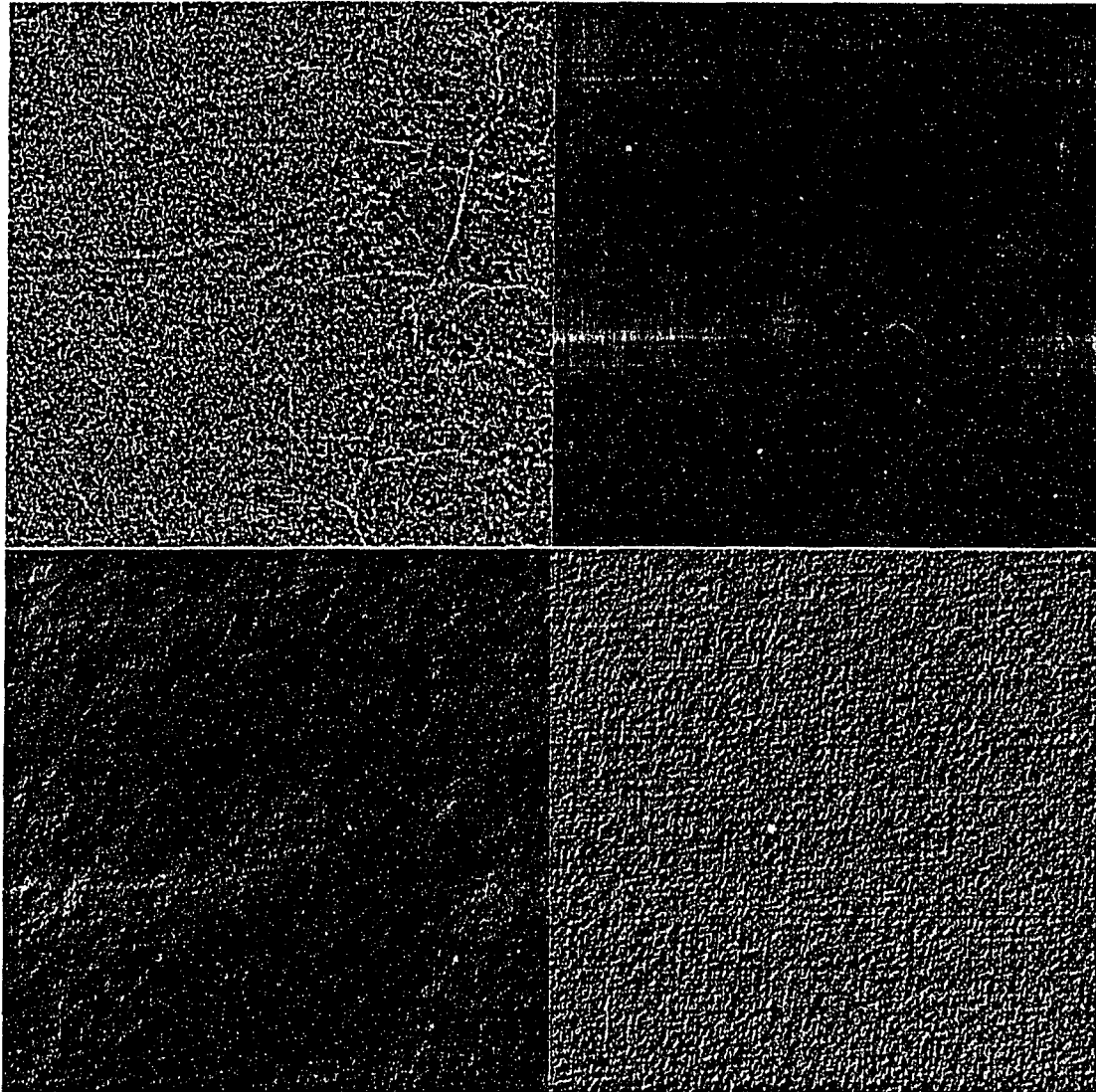
Figure 2.6: Markov model of high frequency mammogram texture. *Upper left* Sample mammogram with spatial mean removed, which we call the "residual" component; *Upper right* Separable Markov model texture with $\alpha = 0.009$. *Lower left* Nonseparable Markov model texture with $\alpha = 0.004$. *Lower right* Weighted average of the two texture types with spatial mean removed. This synthetic texture closely matches the residual mammogram.

CHAPTER 3

Matched Filter for Detecting Microcalcifications

A central problem in stochastic processes is the estimation of a signal in the presence of noise. If the signal to be detected is deterministic and the noise is a wide-sense stationary process with known power spectrum, then a matched filter is known to be the optimum detector in the sense that it maximizes the signal to noise ratio at the detection point or instant. In the following, we estimate the power spectrum of mammograms, discuss the prewhitening matched filter and derive it for detecting μCa++s in mammograms by using the estimated power spectrum.

3.1 Modeling the Power Spectrum of Mammograms

A random process generating an image is said to be stationary in the strict sense if its moments are unaffected by shifts in space. The process is said to be stationary in the wide sense if its mean is constant and its correlation is dependent on the differences in the image coordinates, not on their individual values. In other words for stationary processes, the mean of the image is constant and the image autocorrelation is not a function of position, but a function of difference in image coordinates.

For scenes of man-made objects there is often a large amount of horizontal and vertical image structure, and the spatial separation approximation is quite good.

The autocorrelation function of a random field is called *separable* when it can be expressed as a product of autocorrelation functions of one-dimensional sequences. A separable stationary autocorrelation function often used in image processing is

$$r_{SEP}(m,n) = \sigma_n^2 \rho_1^{|m|} \rho_2^{|n|} \qquad |\rho_1| < 1, \ |\rho_2| < 1, \tag{3.1}$$

which is often expressed in the form

$$r_{SEP}(m,n) = \sigma_n^2 e^{-\alpha_x |m| - \alpha_y |n|}. \tag{3.2}$$

Here σ_n^2 represents the variance of the random field and $\rho_1 = e^{-\alpha_x} = \frac{r_{SEP}(1,0)}{\sigma_n^2}$, $\rho_2 = e^{-\alpha_y} = \frac{r_{SEP}(0,1)}{\sigma_n^2}$ are the one-step correlations in the m and n directions, respectively. We can thus compute the power spectrum of noise $S_{SEP}(w_x, w_y)$ by Fourier transforming $r_{SEP}(m,n)$ in (3.1) as follows:

$$
\begin{aligned}
S_{SEP}(w_x, w_y) &= \mathcal{F}[r_{SEP}(m,n)] \\
&= \sigma_n^2 \frac{(1-\rho_1^2)}{(1-\rho_1 e^{jw_x})(1-\rho_1 e^{-jw_x})} \frac{(1-\rho_2^2)}{(1-\rho_2 e^{jw_y})(1-\rho_2 e^{-jw_y})} \\
&= \frac{\sigma_n^2(1-\rho_1^2)(1-\rho_2^2)}{(1+\rho_1^2-2\rho_1 \cos w_x)(1+\rho_2^2-2\rho_2 \cos w_y)}.
\end{aligned}
\tag{3.3}
$$

In the case of (3.2), $S_{SEP}(w_x, w_y)$ can be computed by replacing ρ_1 and ρ_2 with $e^{-\alpha_x}$, $e^{-\alpha_y}$, respectively. *i.e.*

$$S_{SEP}(w_x, w_y) = \frac{\sigma_n^2(1-2^{-2\alpha_x})(1-2^{-2\alpha_y})}{(1-e^{-(\alpha_x+jw_x)})(1-e^{-(\alpha_x-jw_x)})(1-e^{-(\alpha_y+jw_y)})(1-e^{-(\alpha_y-jw_y)})} \tag{3.4}$$

When α is small, according to the Taylor series, $e^{-(\alpha+jw)}$ and $e^{-(\alpha-jw)}$ can be approximated as $1 - \alpha - jw$, $1 - \alpha + jw$, respectively. Thus, $S_{SEP}(w_x, w_y)$ for a separable process can be simplified as

$$S_{SEP}(w_x, w_y) = \frac{4\alpha_x \alpha_y \sigma_n^2}{(\alpha_x^2 + w_x^2)(\alpha_y^2 + w_y^2)}. \tag{3.5}$$
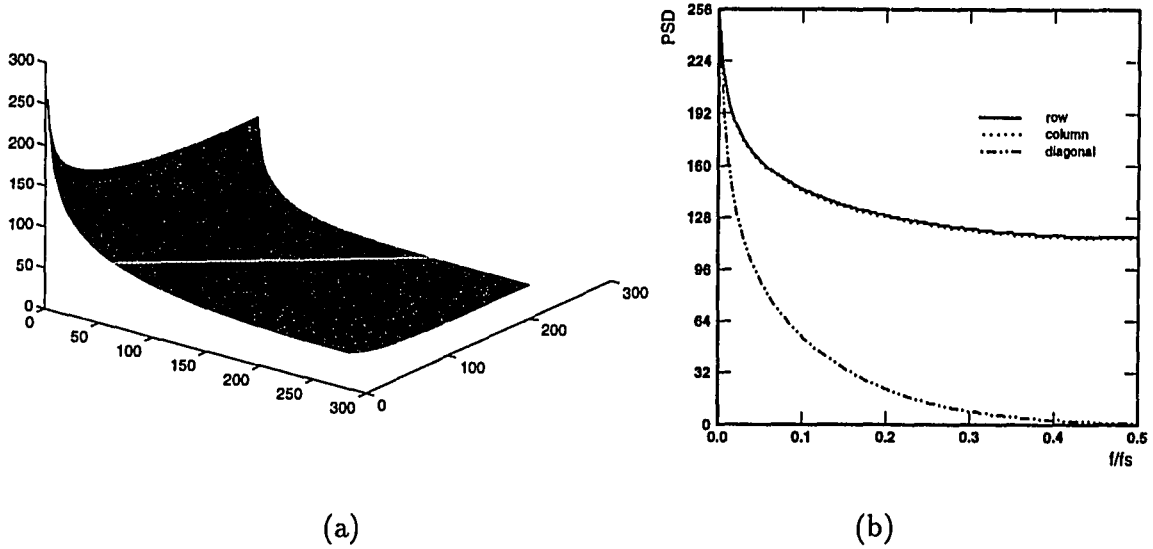
Figure 3.1: (a) Logarithmic power spectrum of a separable Markov process. (b) Its profiles in three directions (row, column and diagonal).

Figure 3.1 shows the power spectrum for a separable Markov process and its profiles in three directions, when $\alpha_x = \alpha_y = 0.009$. Meanwhile, in *natural* scenes, there usually is no preferential direction of correlation; the spatial autocorrelation function tends to be circularly symmetric, and not separable. An image field is often modeled as a sample of a first-order Markov process for which the correlation between points on the image field is proportional to their geometric separation. The *nonseparable* autocorrelation function for the two-dimensional Markov process is

$$r_{NONSEP}(m,n) = \sigma_n^2 e^{-\sqrt{\alpha_x^2 m^2 + \alpha_y^2 n^2}}, \tag{3.6}$$

and the corresponding power spectrum is

$$S_{NONSEP}(w_x, w_y) = \frac{2\sigma_n^2}{\alpha_x \alpha_y \left(1 + \frac{w_x^2}{\alpha_x^2} + \frac{w_y^2}{\alpha_y^2}\right)}. \tag{3.7}$$

(a)　　　　　　　　　　　　　　　(b)
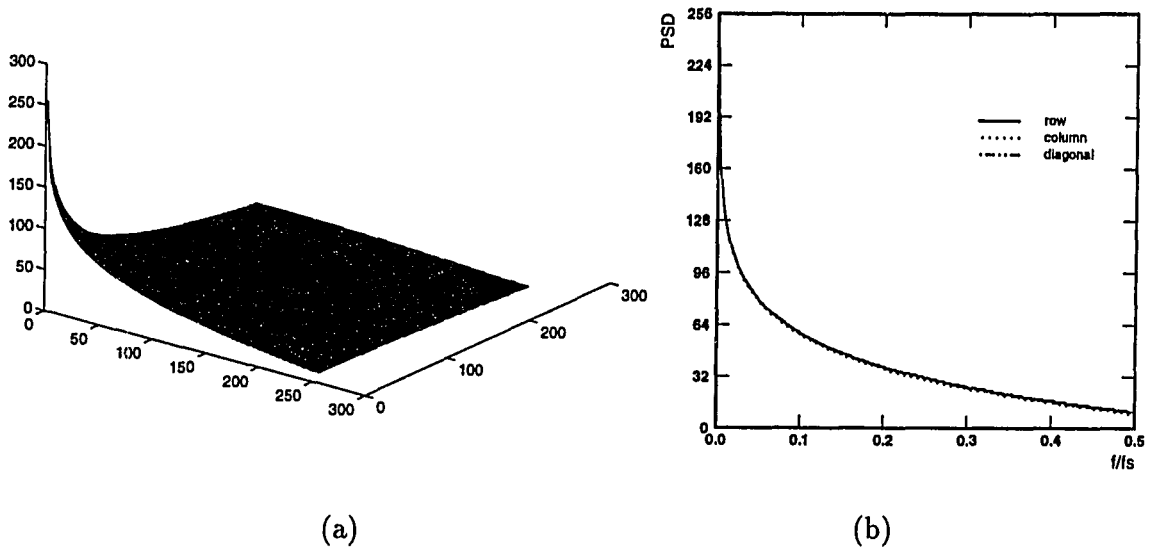
Figure 3.2: (a) Logarithmic power spectrum of a nonseparable Markov process. (b) Its profiles in three directions (row, column and diagonal).

Figure 3.2 shows the power spectrum for a nonseparable Markov process and its profiles in three directions, when $\alpha_x = \alpha_y = 0.004$.

In order to justify the modeling of mammograms as a Markov process, we estimate the autocorrelation function from the preprocessed mammograms. The estimated autocorrelation coefficients are displayed in normalized form, as shown in Figure 3.3. Figure 3.4 shows the estimated power spectrum which is computed by Fourier transforming the estimated autocorrelation coefficients. We assume here that the estimated power spectrum can be modeled as a weighted combination of separable and nonseparable power spectra,

$$\hat{S}(w_x, w_y) = p \cdot S_{SEP}(w_x, w_y) + q \cdot S_{NONSEP}(w_x, w_y). \tag{3.8}$$

(a)   (b)
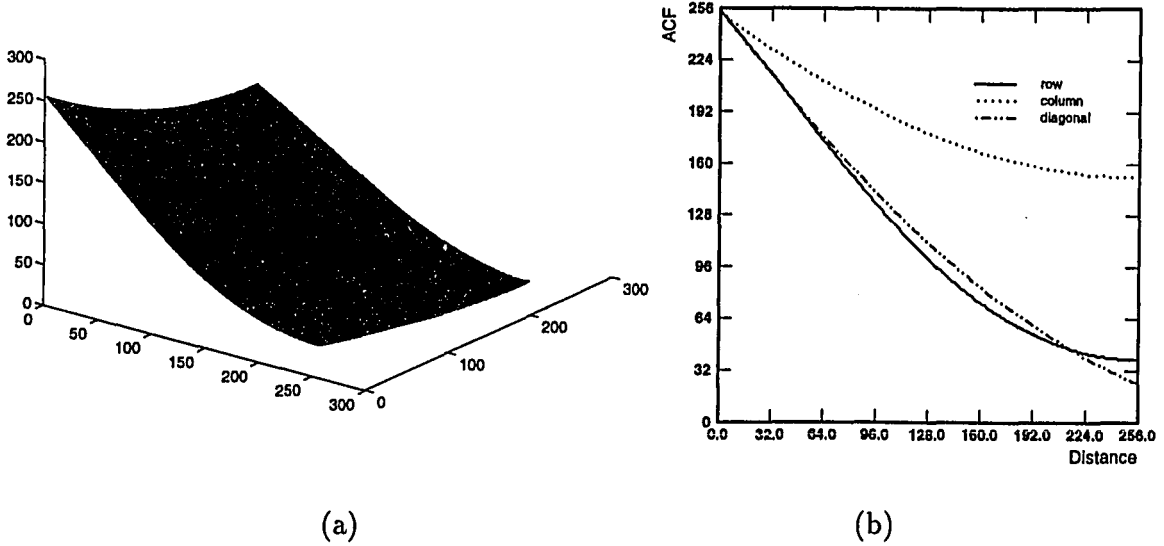
Figure 3.3: (a) Rescaled autocorrelation values computed from 40 mammograms. (b) Its profiles in three directions (row, column and diagonal).

The optimum parameters p, q and α_x, α_y for separable and nonseparable power spectra such that the synthetic power spectrum $\hat{S}(w_x, w_y)$ best matches the estimated power spectrum are obtained in the following way.

First, $\hat{S}(w_x, w_y)$ is computed using (3.5) and (3.7) for each parameter set (p, q, α_x and α_y) according to (3.8). The similarity between $\hat{S}(w_x, w_y)$ and the estimated power spectrum given in Figure 3.4 is investigated by applying a least-squares curve fitting technique to their row, column and diagonal profiles. Curve fitting is performed in the spatial frequency region of $0.1 \leq w_x, w_y \leq 0.5$, since we are only interested in the residual mammograms whose low frequency components are subtracted by applying a high-pass filter.

Secondly, a synthetic texture corresponding to the parameter set is generated, which will be explained in detail in section 5.2. One sample mammogram and the synthetic texture are passed through the same high-pass filter to subtract the local mean. The textural edgeness is computed from the above images. An edge passing through a pixel can be detected using a gradient, which is the sum of the absolute value of the differences between opposite neighboring pixels. Thus a measure of texture for any image can be obtained by computing the gradient image and from it determining the average value of the gradient in the image [66]. The gradient $g(d)$ for every distance d over neighborhood N is defined as

$$g(d) = \Sigma_{(i,j)\in N}\{|I(i,j) - I(i+d,j)| + |I(i,j) - I(i-d,j)|$$
$$+|I(i,j) - I(i,j+d)| + |I(i,j) - I(i,j-d)|\}.$$

(3.9)

This textural edgeness measure $g(d)$ is a feature that characterizes the texture in the spatial domain, while the power spectrum or autocorrelation function basically both relate texture to spatial frequency. Using this measure, the textural edgeness of both the sample mammogram and the synthetic texture are measured and then compared using a least-squares curve fitting technique.

Finally the mammogram texture and the synthetic texture, as shown in Figure 2.6, are visually inspected to ensure that the two images remain close in appearance. The above steps are repeated by changing the parameter set. The optimum parameters are such that their row, column and diagonal profiles are close to each other and the difference between their texture edgeness measures is minimum in the sense of least
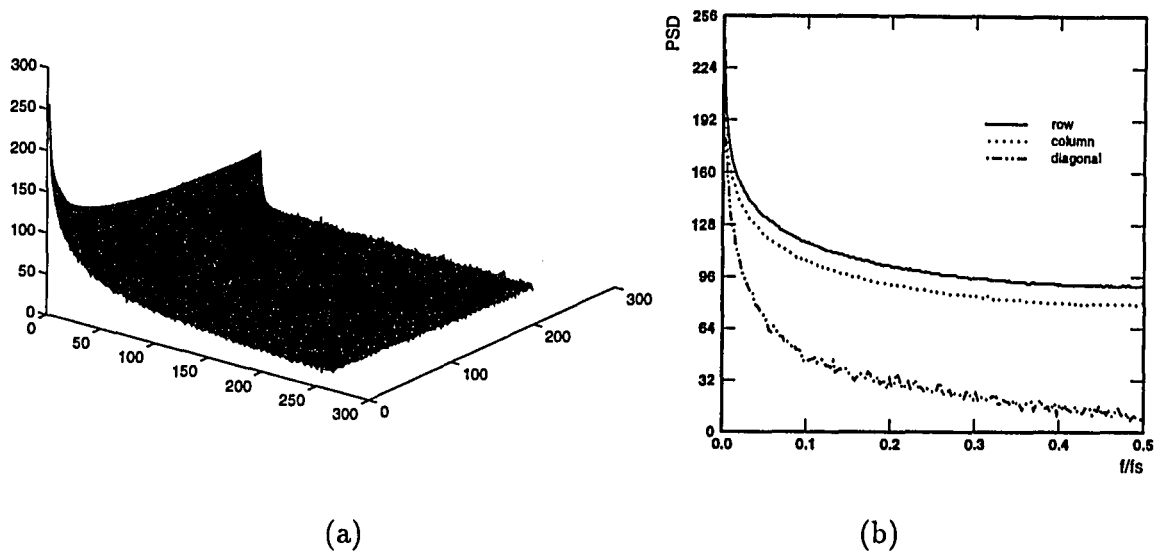
Figure 3.4: (a) Logarithmic power spectrum computed from the 40 mammograms. (b) Its profiles in three directions (row, column and diagonal).

mean square error. The optimal p and q so obtained are 0.375 and 0.625, respectively, and the corresponding α_x ($= \alpha_y$) for separable and nonseparable power spectra are 0.009 and 0.004, respectively.

Figure 3.5 shows the curves of $g(d)$ for the mammogram texture and the texture synthesized by using the above parameters. This figure explains that both textures have the same scales of micro edges because their gradients are very large for distances greater than 3. Thus, from the above analyses it can be asserted that the first order Markov process is a reasonable model for mammograms. Figure 3.6 shows $\hat{S}(w_x, w_y)$ using the above parameters.
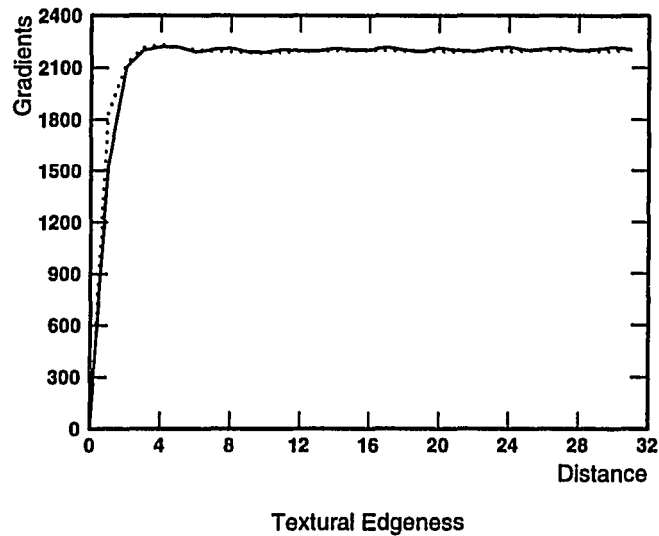
Textural Edgeness

Figure 3.5: Comparison of textural edgeness of a residue of mammogram (*solid*) and synthesized texture (*broken*), shown in upper left and lower left of Figure 2.6, respectively.

From Figure 3.4 and Figure 3.5, we see that the estimated power spectrum can be modeled by combining separable and nonseparable first order Markov processes. This fact is also supported in the spatial domain by Figure 2.6.

3.2 Prewhitening Matched Filter Principle

Suppose a deterministic object $f(x, y)$ (*i.e.* a μCa++) located at (x_0, y_0) is imaged in the presence of some anatomical background noise field $n(x, y)$, giving the observed image

$$i(x, y) = f(x - x_0, y - y_0) + n(x, y). \tag{3.10}$$

(a)   (b)
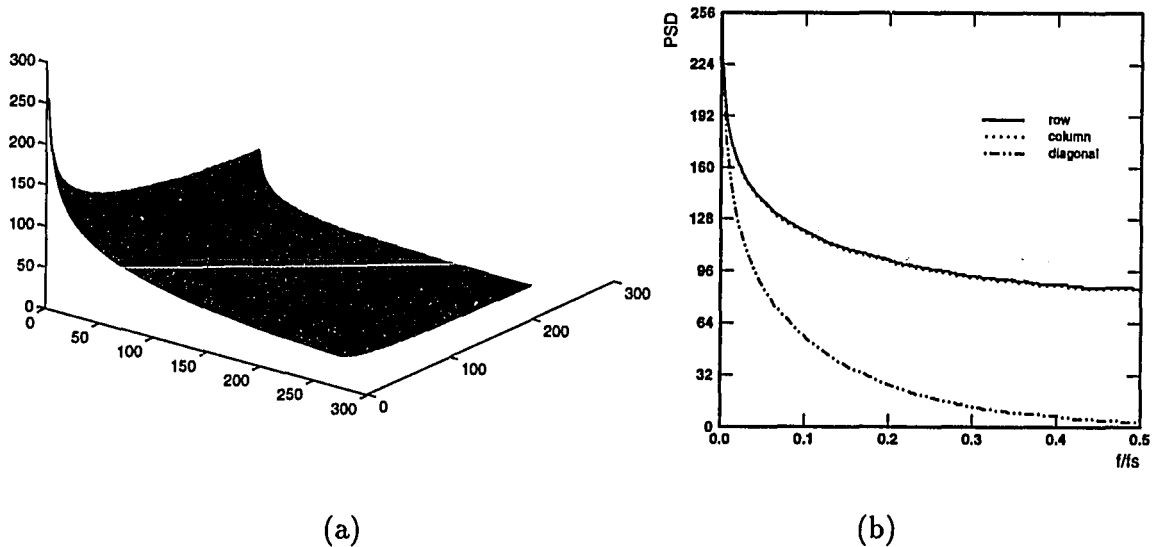
Figure 3.6: (a) Logarithmic power spectrum computed from the addition of separable and nonseparable power spectra. (b) Its profiles in three directions (row, column and diagonal).

If we assume that the noise is stationary with power spectral density $S_{nn}(w_x, w_y)$, then the optimum detector for finding the known $f(x, y)$ is the matched filter with transfer function

$$H_{mf}(w_x, w_y) = \frac{F^*(w_x, w_y)}{S_{nn}(w_x, w_y)} e^{-j(w_x x_0, w_y y_0)}, \qquad (3.11)$$

and impulse response

$$h_{mf}(x, y) = r_{nn}(x, y) * f(-x - x_0, -y - y_0), \qquad (3.12)$$

where $*$ denotes convolution and $r_{nn}(x, y) \equiv \mathcal{F}^{-1} \frac{1}{S_{nn}(w_x, w_y)}$. When the noise is white, $S_{nn}(w_x, w_y)$ is a constant, $r_{nn}(x, y)$ is an impulse, and $h_{mf}(x, y)$ is a coordinate-reversed template of the signal to be detected. Hence, the matched filter performs a

correlation between the input signal and the template. Detection is accomplished by sampling the output at the point of peak correlation. If the output exceeds a threshold the signal is deemed present, or absent if below the threshold. In imaging applications, the location of an object to be detected is unknown, in which case any above-threshold correlation peaks are counted as positive detections. The matched filter is optimum in the sense that the signal-to-noise ratio of the output signal is maximized at the detection instant. Although the matched filter originally assumes knowledge of the shape and location of the signal to be detected, the shape of the signal is more likely to coincide with correlated features in the noise when the background noise is colored. Simply correlating the input data with a template of the desired signal therefore leads to an increase in the false positive rate. In fact, the solution is built into (3.11). It can be seen that $S_{nn}(w_x, w_y)$ acts as a *prewhitening* term by expressing the result of applying the filter to $i(x, y)$ in the form (adopting $x_0 = y_0 = 0$ for simplicity):

$$
\begin{aligned}
G(w_x, w_y) &= H_{mf}(w_x, w_y) I(w_x, w_y) \\
&= \frac{F^*(w_x, w_y)}{\sqrt{S_{nn}(w_x, w_y)}} \frac{I(w_x, w_y)}{\sqrt{S_{nn}(w_x, w_y)}},
\end{aligned}
\tag{3.13}
$$

or, equivalently, in the image domain, and invoking (3.10)

$$
g(x, y) = [p(x, y) * f(-x, -y)] * [p(x, y) * \{f(x, y) + n(x, y)\}],
\tag{3.14}
$$

where $p(x, y) \equiv \mathcal{F}^{-1} \frac{1}{\sqrt{S_{nn}(w_x, w_y)}}$ is the impulse response of a whitening filter. The convolution between $p(x, y)$ and $n(x, y)$ in (3.14) acts to whiten the noise. A side effect of this whitening is object distortion, seen in the convolution between $p(x, y)$

and $f(x, y)$. To compensate, the reversed object template $f(-x, -y)$ is also distorted by $p(x, y)$ in the first convolution term in (3.14). The nature of the matched filter is thereby preserved.

3.3 Derivation of Prewhitening Matched Filter

The matched filter can never be optimum for detecting μCa++s in mammograms since: (a) μCa++s vary in size and shape, thus $f(x, y)$ is not known precisely, and (b) the background texture $n(x, y)$ is nonstationary, meaning that $S_{nn}(w_x, w_y)$ is indeterminate and hence true pre-whitening is unrealizable. Nevertheless a practical algorithm can be implemented with an acceptable drop in performance [30] because of the following two reasons. First, the profile of many μCa++s is approximately Gaussian, and may be modeled as such. Secondly, the nonstationary background can be modeled as a nonstationary mean plus a residual stationary component modeled as a Markov process [47]. Thus, $S_{nn}(w_x, w_y)$ of the residual process [48] can be easily estimated after subtracting the spatial mean from each mammogram. Under these assumptions it is feasible to derive the matched filter in (3.11) for application to the residual component. Hence if we assume that the spatial mean of the background noise is zero and the signal to be detected is the two-dimensional Gaussian of variance σ^2, i.e.,

$$f(x, y; \sigma) = e^{-\frac{(x^2+y^2)}{2\sigma^2}}, \tag{3.15}$$

the prewhitening matched filter for separable Markov process $H_{SEP}(w_x, w_y)$ is

$$H_{SEP}(w_x, w_y) \quad = \quad \frac{F^*(w_x, w_y; \sigma)}{S_{SEP}(w_x, w_y)}$$

$$= \quad \frac{2\pi\sigma^2 e^{-\frac{\sigma^2(w_x^2 + w_y^2)}{2}}}{\frac{4\alpha_x\alpha_y\sigma_n^2}{(\alpha_x^2 + w_x^2)(\alpha_y^2 + w_y^2)}} \tag{3.16}$$

$$= \quad \frac{\pi\sigma^2(\alpha_x^2 + w_x^2)(\alpha_y^2 + w_y^2)e^{-\frac{\sigma^2(w_x^2 + w_y^2)}{2}}}{2\alpha_x\alpha_y\sigma_n^2}.$$

Since $H_{SEP}(w_x, w_y)$ is separable, $H_{SEP}(w_x, w_y)$ can be expressed as

$$H_{SEP}(w_x, w_y) = H_m(w_x)H_m(w_y), \tag{3.17}$$

where for $\alpha_x = \alpha_y = \alpha$,

$$H_m(w) = \frac{\sigma\sqrt{\pi}(\alpha^2 + w^2)}{\sqrt{2}\alpha\sigma_n}e^{-\frac{\sigma^2 w^2}{2}}. \tag{3.18}$$

and for $\alpha \ll 1$,

$$H_m(w) \quad = \quad \frac{\sigma\sqrt{\pi}w^2}{\sqrt{2}\alpha\sigma_n}e^{-\frac{\sigma^2 w^2}{2}}. \tag{3.19}$$

Figure 3.7 shows the impulse response of the separable PWMF given in (3.16) and its profiles. Figure 3.8 shows the shapes of $h_m(x) = \mathcal{F}^{-1}H_m(w)$ as ρ changes.

From this figure, we see that $h_m(x)$ changes from a Gaussian to a Laplacian of Gaussian as ρ increases from 0 to 1. This means that when ρ is 0, i.e., noise is white, the impulse response of the matched filter $h_m(x)$ is just a Gaussian of the same variance, whereas it approaches a Laplacian of Gaussian as ρ increases to 1, i.e., in

Profiles of separable PWMF

(a)   (b)
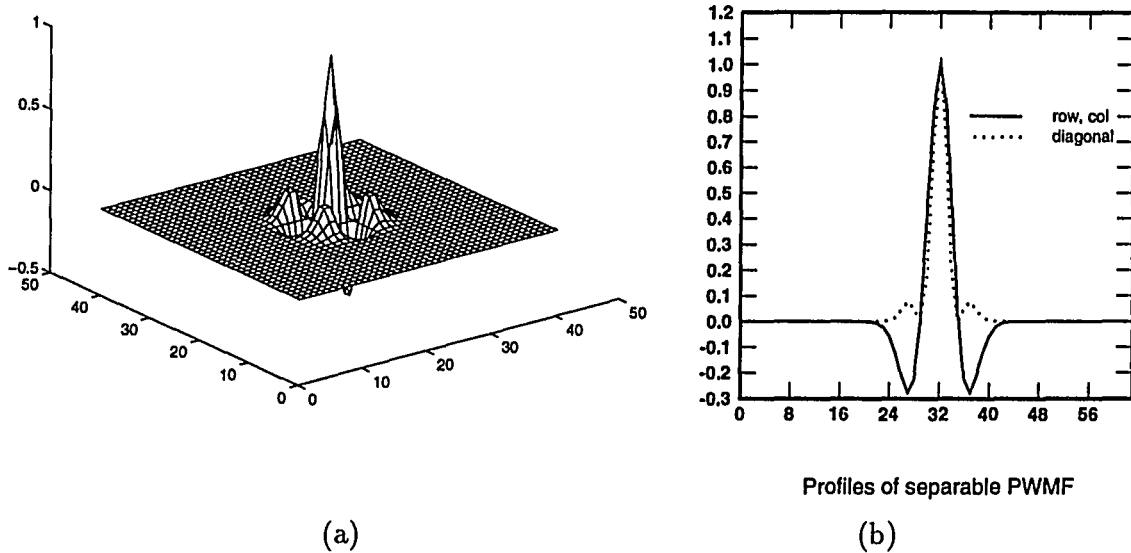
Figure 3.7: (a) Matched filter for detecting Gaussian objects in separable Markov noise. (b) its profiles in the spatial domain. *Solid line* represents row and column profile and the *broken line* represents diagonal profile.

Wavelet Matched filter and PWMFs
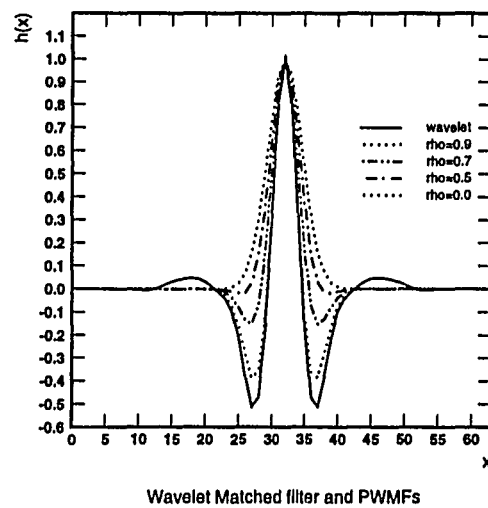
Figure 3.8: Wavelet and Matched filters for a separable Markov model

(a)

Profiles of nonseparable PWMF

(b)
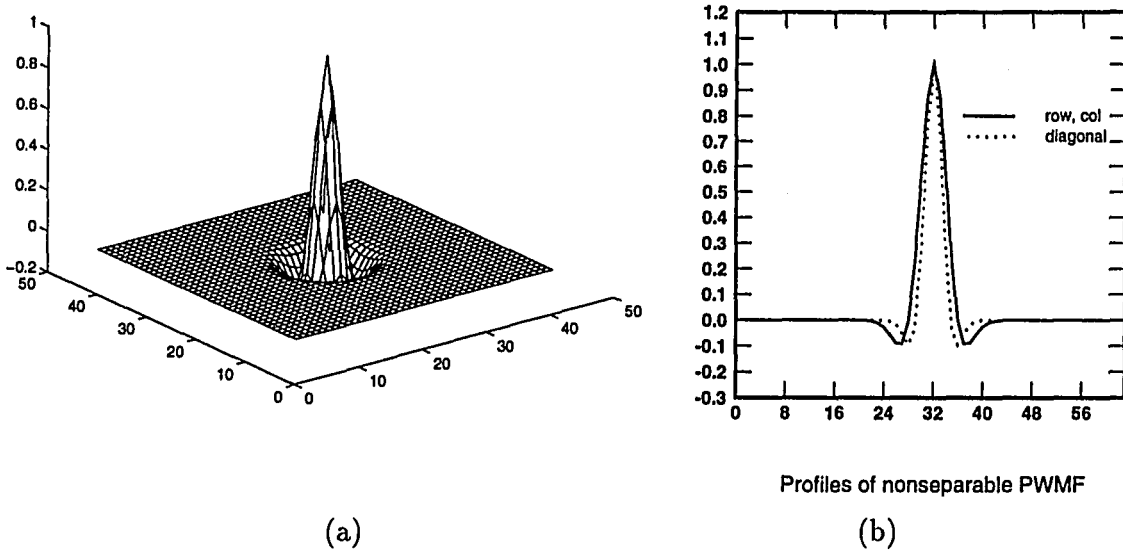
Figure 3.9: (a) Matched filter for detecting Gaussian objects in nonseparable Markov noise. (b) its profiles in the spatial domain.

highly correlated noise. Likewise, for the nonseparable autocorrelation function often considered as more realistic, the prewhitening matched filter $H_{NONSEP}(w_x, w_y)$ is

$$H_{NONSEP}(w_x, w_y) \quad = \quad \frac{F^*(w_x, w_y)}{S_{NONSEP}(w_x, w_y)}$$

$$= \quad \frac{2\pi\sigma^2 e^{-\frac{\sigma^2(w_x^2 + w_y^2)}{2}}}{\frac{2\sigma_n^2}{\alpha_x \alpha_y (1 + \frac{w_x^2}{\alpha_x^2} + \frac{w_y^2}{\alpha_y^2})}} \tag{3.20}$$

$$= \quad \frac{\pi \alpha_x \alpha_y \sigma^2}{\sigma_n^2}(1 + \frac{w_x^2}{\alpha_x^2} + \frac{w_y^2}{\alpha_y^2})e^{-\frac{\sigma^2(w_x^2 + w_y^2)}{2}},$$

where for $\alpha_x = \alpha_y = \alpha \ll 1$,

$$H_{NONSEP}(w_x, w_y) \quad = \quad \frac{\pi\sigma^2}{\sigma_n^2}(w_x^2 + w_y^2)e^{-\frac{\sigma^2(w_x^2 + w_y^2)}{2}}. \tag{3.21}$$

Figure 3.9 shows the impulse response of the nonseparable PWMF given in (3.21) and its profiles. Figure 3.10 shows the profiles of the impulse response of the PWMF computed from the estimated power spectrum. As expected, the nonseparable PWMF is more similar to the estimated one. When $\alpha \to \infty$ the residual component is already white, and the point spread functions $h_{SEP}(x, y)$ and $h_{NONSEP}(x, y)$ are Gaussian. (Note, however, that this condition is not apparent here due to the restriction that $\alpha \ll 1$.) When $\alpha \ll 1$ the increasing inter-pixel correlation of the residual causes both point spread functions to develop increasingly negative side lobes on either side of the positive main lobe, as shown in Figure 3.8. The remaining variable - the size of the μCa++ - may be accommodated by implementing a bank of both forms of matched filter in the form of a wavelet transform.

In the next chapter, wavelet theory is reviewed and the transfer functions of the details images are derived in terms of a scaling function and a wavelet. Finer sampling of scale is also presented.
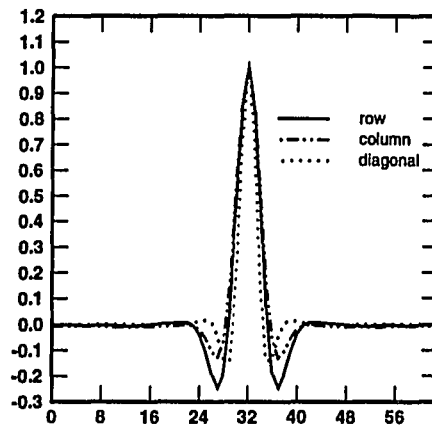
Profiles of estimated PWMF

Figure 3.10: Profiles of the estimated PWMF in the spatial domain.

CHAPTER 4

Wavelet Transform

Multifrequency decomposition schemes were introduced by Crochier, Webber, and Flanagan [40] for sub-band coding of speech signals. The basic idea of a sub-band coding scheme is to decompose an input signal into a number of frequency bands (or sub-bands) using a bank of band-pass filters. Each sub-band is then decimated and encoded appropriately. The encoded sub-bands are decoded, interpolated, and then passed through synthesis filters to reconstruct the signal. Quadrature mirror filters (QMF) were introduced in [41] and it was proved that QMFs can reconstruct the signal without aliasing in the absence of quantization errors. A number of multiresolution signal processing schemes have emerged independently in many different fields. Wavelet theory is recognized as a unified framework for multiresolution signal representations. Wavelets were introduced by Grossmann and Morlet [18] as a family of functions $\psi_{a,b}(x)$, where

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}}\psi(\frac{x-b}{a}). \qquad (4.1)$$

They are derived from translations and dilations of one basic function $\psi(x)$, referred to as the *mother* wavelet. The parameter a in (4.1) is a scale parameter and the constant $\frac{1}{\sqrt{|a|}}$ is used for energy normalization. Depending on the type of application, different families of wavelets may be chosen. There are various continuous and

discrete ways of representing time-scale parameters (b, a), each one yielding a different type of wavelet transform. One can then represent functions $f \in L^2(R)$ by the functions

$$D_{a,b}f = \int_{-\infty}^{\infty} f(x) \cdot \psi_{a,b}^*(x)dx, \qquad (4.2)$$

where the asterisk stands for complex conjugate. $L^2(R)$ denotes the vector space of measurable, square-integrable one dimensional functions. If $\psi(x)$ satisfies the admissibility condition such that

$$\int_{-\infty}^{\infty} |\xi|^{-1}\hat{\psi}(\xi)d\xi < \infty, \qquad (4.3)$$

where $\hat{\psi}$ denotes the Fourier transform of ψ, then, $D_{a,b}$ is an isometry (up to a constant) from $L^2(R)$ into $L^2(R^* \times R)$ [19], where R^* is equal to R except that 0 is excluded. The operator $D_{a,b}$ is called the "Continuous wavelet transform" which was originated by Goupillaud, Grossmann, and Morlet [42]. If ψ has sufficient decay, then the condition (4.3) implies that

$$\int_{-\infty}^{\infty} \psi(x)dx = 0. \qquad (4.4)$$

Typically, the function ψ will therefore have at least some oscillations. The basic idea of the wavelet transform is to represent any arbitrary function f as a superposition of wavelets. Any such superposition decomposes f into different scale levels, where each level is then further decomposed with a resolution adapted to the level. One way to achieve such a decomposition writes f as an integral over a and b of $\psi_{a,b}$

with appropriate weighting coefficients [18] as follows,

$$f(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} D_{a,b} f \psi_{a,b}(x) \frac{dadb}{a^2}. \tag{4.5}$$

We can sample the parameters a and b to represent f as a discrete superposition. If we define $a = a_0^j$, $b = kb_0 a_0^j$ with $j, k \in Z$, and $a_0 > 1$, $b_0 > 0$ fixed, the wavelet decomposition is then

$$f(x) = \sum_{k=-\infty}^{\infty} D_{a_0^j, ka_0^j} f(k) \psi_{a_0^j, kb_0 a_0^j}(x), \tag{4.6}$$

where

$$\psi_{a_0^j, kb_0 a_0^j}(x) = a_0^{-\frac{j}{2}} \psi(a_0^{-j} x - kb_0). \tag{4.7}$$

Decompositions of this type were studied in [43]. For discrete wavelet transforms where $a_0 = 2$, $b_0 = 1$, there exist very special choices of ψ such that the $\psi_{2^j, k2^j}$ constitute an orthonormal basis. Different bases of this nature were constructed by Stromberg [36], Meyer [37], Lemarie [38], Battle [39], and Daubechies [19]. All these examples correspond to a multiresolution analysis, which gives rise to a fast computation algorithm.

4.1 Multiscale Signal Representation using Wavelets

In multiresolution analysis, one really has two functions; the mother wavelet ψ and a *scaling function* ϕ. One introduces dilated and translated versions of the

scaling function as well as the mother wavelet,

$$\phi_{2^j,k2^j}(x) \;=\; 2^{-\frac{j}{2}}\phi(2^{-j}x - k)$$
$$\psi_{2^j,k2^j}(x) \;=\; 2^{-\frac{j}{2}}\psi(2^{-j}x - k). \tag{4.8}$$

For fixed j, $\phi_{2^j,k2^j}$ are orthonormal. We denote by V_j the space spanned by the $\phi_{2^j,k2^j}$; these spaces V_j describe successive approximation spaces, $\cdots V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots$, each with resolution 2^j. For each j, the $\psi_{2^j,k2^j}$ span a space O_j which is exactly the orthogonal complement in V_{j-1} of V_j; the coefficients $< \psi_{2^j,k2^j}, f >$ which are $\int_{-\infty}^{\infty} f(x)\psi_{2^j,k2^j}(x)dx$ therefore, describe the information lost when going from an approximation of f with resolution 2^{j-1} to the coarser approximation with resolution 2^j. All this is translated into the following algorithm for the computation of the $D_{2^j,k2^j}f = < \psi_{2^j,k2^j}, f >$:

$$A_{2^j,k2^j}f(k) \;=\; \sum_{k=-\infty}^{\infty} \tilde{h}(2k - l)A_{2^{j-1},k2^{j-1}}f(k)$$
$$D_{2^j,k2^j}f(k) \;=\; \sum_{k=-\infty}^{\infty} \tilde{g}(2k - l)A_{2^{j-1},k2^{j-1}}f(k). \tag{4.9}$$

$\tilde{h}(\cdot)$ and $\tilde{g}(\cdot)$ are obtained in the following way,

$$h(k) = 2^{\frac{1}{2}} \int_{-\infty}^{\infty} \phi(x - k)\phi(2x)dx, \tag{4.10}$$

$$\tilde{h}(k) \;=\; h(-k)$$
$$g(k) \;=\; (-1)^{1-k}h(-k + 1)$$
$$\tilde{g}(k) \;=\; g(-k). \tag{4.11}$$

If we compute the Fourier transform of (4.10), we get

$$\hat{\phi}(2w) = H(w)\hat{\phi}(w), \tag{4.12}$$

where $\hat{\phi}(w)$ is the Fourier transform of $\phi(x)$. $\tilde{H}(w)$ satisfies the following conditions [22]:

$$|\tilde{H}(0)| = 1; \quad \tilde{h}(n) = o(n^{-2}), n \rightarrow \infty$$
$$|\tilde{H}(w)|^2 + |\tilde{H}(w+\pi)|^2 = 1. \tag{4.13}$$

From (4.12), $\hat{\phi}(w)$ can be expressed as

$$\hat{\phi}(w) = \prod_{p=1}^{\infty} H(2^{-p}w). \tag{4.14}$$

The Fourier transform of (4.11) gives

$$G(w) = e^{-jw}\overline{H(w+\pi)}, \tag{4.15}$$

and the Fourier transform of $\psi(x)$ can be written as

$$\hat{\psi}(w) = G(\frac{w}{2})\hat{\phi}(\frac{w}{2}) \tag{4.16}$$

For details, see [22].

In practice, real signals are represented at finite resolution. For normalization purposes, we suppose that this resolution is equal to 1. The orthonormal projection on V_{2^j} can now be computed by decomposing the signal $f(x)$ on the above orthonormal basis. The approximation of the signal $f(x)$ at the resolution 1, $f_{2^0}(x)$ can be expressed as

$$f(x) = f_{2^0}(x) = \sum_{k=-\infty}^{\infty} A_{2^0}f(k)\phi(x-k), \tag{4.17}$$

where $A_{2^0}f(k)$ is characterized by the set of inner products, which we denote by

$$A_{2^0}f(k) = (< f(x), \phi(x-k) >)_{k \in Z}$$
$$= ((f(x) * \phi(-x))(k))_{k \in Z}. \tag{4.18}$$

In other words, $A_{2^0}f(k)$ corresponds to the sampled version of $f(x) * \phi(-x)$. The Fourier transform of (4.18) gives

$$A_{2^0}F(w) = F(w)\hat{\phi}^*(w), \tag{4.19}$$

where $A_{2^0}F(w)$, $F(w)$, and $\hat{\phi}(w)$ are the Fourier transforms of $A_{2^0}f(k)$, $f(x)$, and $\phi(x)$, respectively. $A_{2^0}f(k)$ is filtered by a low-pass filter $\tilde{H}(w)$ to give the first low frequency sub-band signal $A_{2^1}f(k)$. Using (4.14), $A_{2^1}F(w)$ can be rewritten as

$$
\begin{aligned}
A_{2^1}F(w) &= A_{2^0}F(w)\tilde{H}(w) \\
&= F(w)\hat{\phi}^*(w)H^*(w) \\
&= F(w)\hat{\phi}^*(2w).
\end{aligned}
\tag{4.20}
$$

Likewise, the first high frequency sub-band signal $D_{2^1}f(k)$ is obtained by passing $A_{2^0}f$ through the high-pass filter $\tilde{G}(w)$. Using (4.16), $D_{2^1}F(w)$ can be expressed as

$$
\begin{aligned}
D_{2^1}F(w) &= F(w)\hat{\phi}^*(w)\tilde{G}(w) \\
&= F(w)\hat{\phi}^*(w)G^*(w) \\
&= F(w)\hat{\psi}^*(2w).
\end{aligned}
\tag{4.21}
$$

This can be easily implemented by the filtering process shown in Figure 4.1. If this filtering process is repeated, then

$$
\begin{aligned}
A_{2^j}F(w) &= F(w)\hat{\phi}^*(2^j w) \\
D_{2^j}F(w) &= F(w)\hat{\psi}^*(2^j w).
\end{aligned}
\tag{4.22}
$$

The above equation can be rewritten as

$$
\begin{aligned}
A_{2^j}F(w) &= A_{2^0}F(w)\frac{\hat{\phi}^*(2^j w)}{\hat{\phi}^*(w)} \\
D_{2^j}F(w) &= A_{2^0}F(w)\frac{\hat{\psi}^*(2^j w)}{\hat{\phi}^*(w)}.
\end{aligned}
\tag{4.23}
$$

(a) Analysis   (b) Synthesis

(c)
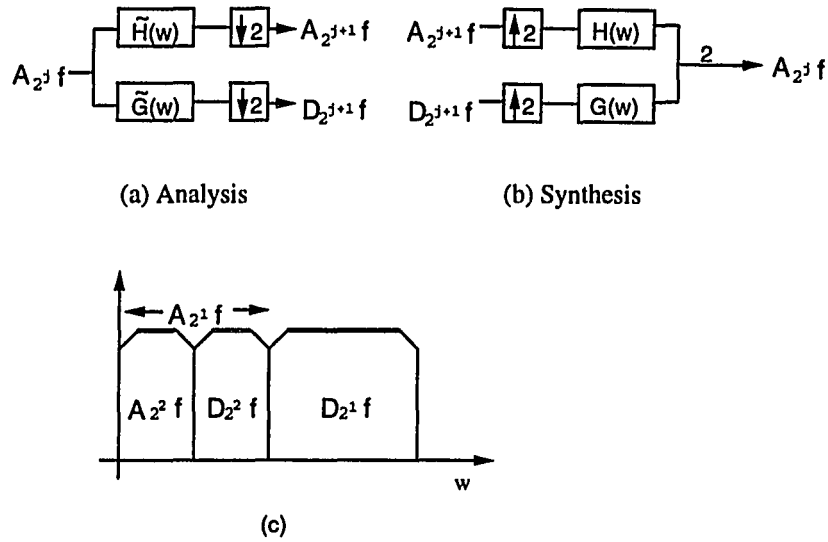
Figure 4.1: One-dimensional wavelet transform

Thus, the approximation at the resolution 2^j, $f_{2^j}(x)$ can be expressed in the spatial domain as

$$f_{2^j}(x) = \sum_{k=-\infty}^{\infty} A_{2^j}f(k)\phi(2^{-j}x - k), \qquad (4.24)$$

where $A_{2^j}f(k)$ is characterized by the set of inner products, which we denote by

$$
\begin{aligned}
A_{2^j}f(k) &= (< f(u), \phi(2^{-j}u - k) >)_{k \in Z} \\
&= ((f(u) * \phi(-2^{-j}u))(2^{-k}))_{k \in Z}.
\end{aligned}
\qquad (4.25)
$$

This set of signals is called an orthogonal wavelet representation, and consists of the reference signal at a coarse resolution $A_{2^J}f$ and the details signals $D_{2^j}f$ at the resolutions 2^j for $1 \leq j \leq J$. The set can be interpreted as a decomposition of the original signal in an orthonormal wavelet basis or as a decomposition of the signal in a set of independent frequency channels as shown in Figure 4.1(c). This model can be easily extended to two dimensions. The signal is now a finite energy continuous

function $f(x,y) \in L^2(R^2)$. $L^2(R^2)$ denotes the vector space of measurable, square-integrable two dimensional functions $f(x,y)$. Let $(V_{2^j})_{j \in Z}$ be such a multiresolution approximation of $L^2(R^2)$. The approximation of a signal $f(x,y)$ at resolution 2^j is equal to its orthogonal projection on the vector space V_{2^j}. One can show that there exists a unique scaling function $\Phi(x,y)$ whose dilation and translation gives an orthonormal basis of each space V_{2^j}. The set $2^{-j}\Phi(2^{-j}x - m, 2^{-j}x - n)_{(m,n) \in Z^2}$ forms an orthonormal basis of V_{2^j}. 2^{-j} is a normalization factor of each function in the $L^2(R^2)$ norm. For the case of separable multiresolution approximations of $L^2(R^2)$, the scaling function $\Phi(x,y)$ can be written as

$$\Phi(x,y) = \phi(x)\phi(y), \qquad (4.26)$$

where $\phi(x)$ is the one dimensional scaling function. The approximation of the signal $f(x,y)$ at the resolution 1, $f_{2^0}(x,y)$ can be expressed as

$$\begin{aligned} f(x,y) &= f_{2^0}(x,y) \\ &= \sum_{m,n=-\infty}^{\infty} A_{2^0}f(m,n)\Phi(x-m,y-n) \qquad (4.27) \\ &= \sum_{m,n=-\infty}^{\infty} A_{2^0}f(m,n)\phi(x-m)\phi(y-n), \end{aligned}$$

where $A_{2^0}f(m,n)$ is characterized by the set of inner products, which we denote by

$$\begin{aligned} A_{2^0}f(m,n) &= (< f(u,v), \phi(u-m)\phi(v-n >)_{m,n \in Z} \\ &= ((f(u,v) \ * \ \phi(-u)\phi(-v))(m,n))_{m,n \in Z}. \end{aligned} \qquad (4.28)$$

In other words, $A_{2^0}f(m,n)$ corresponds to the sampled version of $f(x,y)*\Phi(-x,-y)$. Actually, we are given $A_{2^0}f(m,n)$ rather than $f(x,y)$. We can rewrite the above

equation in the frequency domain

$$A_{2^0}F(w_x, w_y) = F(w_x, w_y)\hat{\phi}^*(w_x)\hat{\phi}^*(w_y), \tag{4.29}$$

where $A_{2^0}F(w_x, w_y)$, $F(w_x, w_y)$, and $\hat{\phi}(w)$ are the Fourier transforms of $A_{2^0}f(m, n)$, $f(x, y)$, and $\phi(x)$, respectively. The above operations can be easily implemented by the filtering process shown in Figure 4.2(a), where $A_{2^0}f$ is filtered to give $A_{2^1}f$, $D_{2^1}^1f$, $D_{2^1}^2f$, $D_{2^1}^3f$. Using (4.14) and (4.16), their Fourier transforms can be written as

$$
\begin{aligned}
A_{2^1}F(w) &= A_{2^0}F(w_x, w_y)\tilde{H}(w_x)\tilde{H}(w_y) \\
&= A_{2^0}F(w_x, w_y)\frac{\hat{\phi}^*(2w_x)}{\hat{\phi}^*(w_x)}\frac{\hat{\phi}^*(2w_y)}{\hat{\phi}^*(w_y)},
\end{aligned} \tag{4.30}
$$

$$
\begin{aligned}
D_{2^1}^1F(w) &= A_{2^0}F(w_x, w_y)\tilde{H}(w_x)\tilde{G}(w_y) \\
&= A_{2^0}F(w_x, w_y)\frac{\hat{\phi}^*(2w_x)}{\hat{\phi}^*(w_x)}\frac{\hat{\psi}^*(2w_y)}{\hat{\psi}^*(w_y)},
\end{aligned} \tag{4.31}
$$

$$
\begin{aligned}
D_{2^1}^2F(w) &= A_{2^0}F(w_x, w_y)\tilde{G}(w_x)\tilde{H}(w_y) \\
&= A_{2^0}F(w_x, w_y)\frac{\hat{\psi}^*(2w_x)}{\hat{\psi}^*(w_x)}\frac{\hat{\phi}^*(2w_y)}{\hat{\phi}^*(w_y)},
\end{aligned} \tag{4.32}
$$

and,

$$
\begin{aligned}
D_{2^1}^3F(w) &= A_{2^0}F(w_x, w_y)\tilde{G}(w_x)\tilde{G}(w_y) \\
&= A_{2^0}F(w_x, w_y)\frac{\hat{\psi}^*(2w_x)}{\hat{\psi}^*(w_x)}\frac{\hat{\psi}^*(2w_y)}{\hat{\psi}^*(w_y)}.
\end{aligned} \tag{4.33}
$$

If this filtering process is repeated, then

$$
\begin{aligned}
A_{2^j}F(w) &= A_{2^0}F(w_x, w_y)\frac{\hat{\phi}^*(2^jw_x)}{\hat{\phi}^*(w_x)}\frac{\hat{\phi}^*(2^jw_y)}{\hat{\phi}^*(w_y)}, \\
D_{2^j}F(w) &= A_{2^0}F(w_x, w_y)\frac{\hat{\phi}^*(2^jw_x)}{\hat{\phi}^*(w_x)}\frac{\hat{\psi}^*(2^jw_y)}{\hat{\psi}^*(w_y)}, \\
D_{2^j}F(w) &= A_{2^0}F(w_x, w_y)\frac{\hat{\psi}^*(2^jw_x)}{\hat{\psi}^*(w_x)}\frac{\hat{\phi}^*(2^jw_y)}{\hat{\phi}^*(w_y)}, \\
D_{2^j}F(w) &= A_{2^0}F(w_x, w_y)\frac{\hat{\psi}^*(2^jw_x)}{\hat{\psi}^*(w_x)}\frac{\hat{\psi}^*(2^jw_y)}{\hat{\psi}^*(w_y)}.
\end{aligned} \tag{4.34}
$$

Thus, the approximation at the resolution 2^j, $f_{2^j}(x, y)$ can be expressed as

$$f_{2^j}(x, y) = \sum_{m,n=-\infty}^{\infty} A_{2^j} f(m, n) \phi(2^{-j} x - m) \phi(2^{-j} y - n), \qquad (4.35)$$

where $A_{2^j} f(m, n)$ is characterized by the set of inner products, which we denote by

$$
\begin{aligned}
A_{2^j} f(m, n) &= \quad (< f(u, v), \phi(2^{-j} u - m) \phi(2^{-j} v - n >)_{m,n \in Z} \\
&= ((f(u, v) * \phi(-2^{-j} u))(2^{-m}) \phi(-2^{-j} v))(2^{-n}))_{m,n \in Z}.
\end{aligned}
\qquad (4.36)
$$

This set of signals is called an orthogonal wavelet representation in two dimensions.

4.2 Implementation of the Wavelet Transform

In fact, $A_{2^j, k2^j} f(k)$ are coefficients characterizing the projection of $f(x)$ onto V_j. If the function f is given in sampled form, then one can take these samples for the highest order resolution approximation coefficients $A_{2^0, k} f$ and (4.9) describes a sub-band coding algorithm on these sampled values, with low-pass filter $\tilde{h}(\cdot)$ and high-pass filter $\tilde{g}(\cdot)$. These filters can give exact reconstruction because they are associated with orthonormal wavelet bases, given as

$$A_{2^{j-1}, k2^{j-1}} f(k) = \sum_{k=-\infty}^{\infty} h(2k - l) A_{2^j, k2^j} f(k) + g(2k - l) D_{2^j, k2^j} f(k). \qquad (4.37)$$

$H(w)$ and $G(w)$ are the complex conjugate of $\tilde{H}(w)$, $\tilde{G}(w)$, respectively, for the orthogonal wavelet transform. In implementing this process, it is desirable that the FIR filters be linear phase, since such filters can be easily cascaded in pyramidal filter structures without phase compensation. Unfortunately, it is known that there

are no nontrivial orthonormal linear phase FIR filters with the exact reconstruction property [44], regardless of any regularity considerations. One can preserve linear phase by using biorthogonal wavelet bases, which have recently been constructed by Cohen, Daubechies and Feauveau [45] and by Herley and Vetterli [46]. In this case, we decompose and reconstruct a signal as in (4.9) and (4.37), respectively. However, the filters $H(w)$ and $G(w)$ may not be the complex conjugates of $\tilde{H}(w)$, $\tilde{G}(w)$, respectively. These filters should satisfy the following conditions for perfect reconstruction,

$$
\begin{aligned}
g(k) &= (-1)^{1-k}\tilde{h}(-k+1) \\
\tilde{g}(k) &= (-1)^{1-k}h(-k+1) \\
\sum_{k=-\infty}^{\infty} h(k)\tilde{h}(k+2n) &= \delta_{n,0}.
\end{aligned}
\tag{4.38}
$$

We define the analysis scaling function $\tilde{\phi}(x)$ and the synthesis function $\phi(x)$ by

$$
\begin{aligned}
\tilde{\phi}(x) &= \sum_{k=-\infty}^{\infty} h(k)\tilde{\phi}(2x-k) \\
\phi(x) &= \sum_{k=-\infty}^{\infty} \tilde{h}(k)\phi(2x-k),
\end{aligned}
\tag{4.39}
$$

with Fourier transforms given as

$$
\begin{aligned}
\hat{\tilde{\phi}}(w) &= \prod_{p=1}^{\infty} H(2^{-p}w) \\
\hat{\phi}(w) &= \prod_{p=1}^{\infty} \tilde{H}(2^{-p}w).
\end{aligned}
\tag{4.40}
$$

Likewise, if we define the analysis wavelet $\tilde{\psi}(x)$ and the synthesis wavelet $\psi(x)$ by

$$
\begin{aligned}
\tilde{\psi}(x) &= \sum_{k=-\infty}^{\infty} g(k)\tilde{\phi}(2x-k) \\
\psi(x) &= \sum_{k=-\infty}^{\infty} \tilde{g}(k)\phi(2x-k),
\end{aligned}
\tag{4.41}
$$

their Fourier transforms are given as

$$
\begin{aligned}
\hat{\psi}(w) &= G(\tfrac{w}{2})\tilde{\phi}(\tfrac{w}{2}) \\
\hat{\psi}(w) &= \tilde{G}(\tfrac{w}{2})\phi(\tfrac{w}{2}).
\end{aligned}
\tag{4.42}
$$

Then, the analysis and synthesis procedures are also the same as shown in (4.9) and (4.37), respectively. Thus, the discrete wavelet transform is equivalent to a sub-band decomposition of one dimensional signals, as shown in Figure 4.1. The input signal $A_{2^0}f$ (for simplicity, we omit time index k) is filtered by a low-pass filter $\tilde{H}(w)$ followed by down-sampling by a factor of two to give the first low-frequency sub-band signal $A_{2^1}f$. The first high-frequency sub-band signal $D_{2^1}f$ is created by passing $A_{2^0}f$ through the high-pass filter $\tilde{G}(w)$, again followed by down-sampling. The structure for applying the filters for computing $A_{2^j}f$ and $D_{2^j}f$ is shown in Figure 4.1(a).

The filters $\tilde{H}(w)$ and $\tilde{G}(w)$ are designed to give an octave decomposition of the frequency domain, as seen in Figure 4.1(c). Together the components $A_{2^1}f$ and $D_{2^1}f$ define the first octave of the wavelet transform. The filtering process is repeated using $A_{2^1}f$ as the input to the next level of the sub-band decomposition. The outcome after N octaves is N details signals $D_{2^1}f$ - $D_{2^N}f$ and one smooth signal $A_{2^N}f$. The details signals are the most interesting as they represent a band-pass decomposition into structure whose scale increases with octave level. Each details sub-band represents a certain scale range of the residual component discussed in the previous section. Reconstruction of $A_{2^0}f$ is achieved by upsampling (inserting zeros between samples of the sub-bands) and filtering using $H(w)$ and $G(w)$ as shown in Figure 4.1(b). This

is called the orthogonal wavelet transform if the synthesis wavelets are the same as the analysis wavelets, otherwise it is called the biorthogonal wavelet transform.

A separable two-dimensional wavelet transform for image analysis can be implemented using the one-dimensional transform (Figure 4.2) if the two-dimensional scaling function $\Phi(x,y)$ and the associated three wavelets $\Psi^1(x,y)$, $\Psi^2(x,y)$ and $\Psi^3(x,y)$ are such that

$$
\begin{aligned}
\Phi(x,y) &= \phi(x)\phi(y) \\
\Psi^1(x,y) &= \phi(x)\psi(y) \\
\Psi^2(x,y) &= \psi(x)\phi(y) \\
\Psi^3(x,y) &= \psi(x)\psi(y).
\end{aligned}
\tag{4.43}
$$

Each row of the input image is separately filtered by the same filters used in the one-dimensional transform. The resulting pair of row-transformed images are likewise filtered in the column direction, yielding four sub-band images at the first octave level. The three details images - termed LH, HL, and HH - correspond to specific bands in the frequency domain as shown in Figure 4.2(c). The LH sub-band is formed by low-pass filtering the rows followed by high-pass filtering the columns, and is therefore sensitive to horizontally oriented features. In the same way the HL sub-band contains vertically oriented details, and the HH sub-band contains primarily diagonal structure. The LL component is a low-pass filtered or *smooth* version of the original image, and is passed through to the next octave for further sub-band decomposition. Thus, the LL band of the frequency domain is segmented
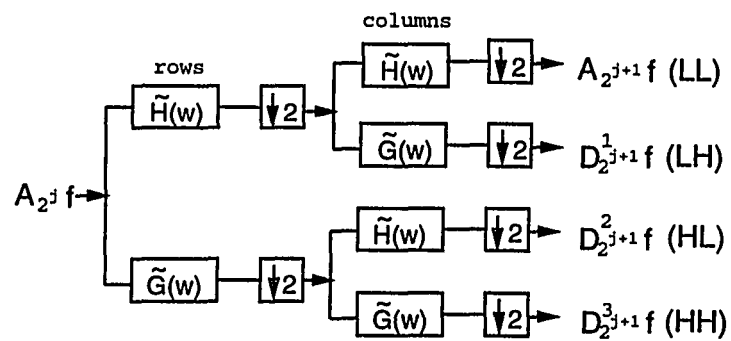
(a) Analysis
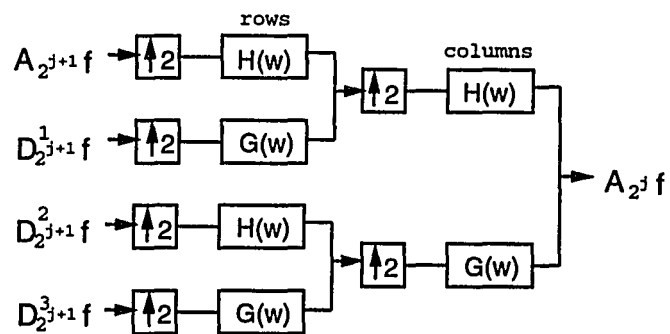
(b) Synthesis

(c)

Figure 4.2: Separable two-dimensional wavelet transform

into four sub-bands at the second octave level, and so on. The two-dimensional wavelet transform is inverted by filtering and combining the details images from all octaves plus the LL component from the highest octave.

Down-sampling and up-sampling are normally employed in the forward and inverse transforms, respectively. However, in our work we maintain full resolution throughout the sub-band decomposition, for two reasons. First, from the point of view of a human observer it helps to display the sub-bands at full size during algorithm development simply because features are easier to see at full resolution. Second, it is easier to combine the detected pixels from each sub-band when they are at the same resolution. In addition, a decimation operation can deform the patterns if their sizes are small. So when we try to detect certain patterns in the details images, it is desirable to keep the details images at full resolution. In other words, the wavelet transform without decimation and interpolation is shift invariant. Although the resulting transform is highly redundant from an information theoretic point of view, it is still simple to compute and the same filters are employed for analysis and synthesis.

Filtering with $H(w)$ followed by down-sampling is equivalent to filtering with $H(2w)$ with respect to the spectrum. $H(2w)$ can be obtained by putting a zero between each of the coefficients of $h(n)$. In the case of up-sampling, its operation is the dual of the above process. The discrete wavelet transform without down-sampling and up-sampling can be implemented using the method shown in Figure 4.3. The corresponding two-dimensional wavelet transform is shown in Figure 4.4. Perhaps the

(a) Analysis (b) Synthesis
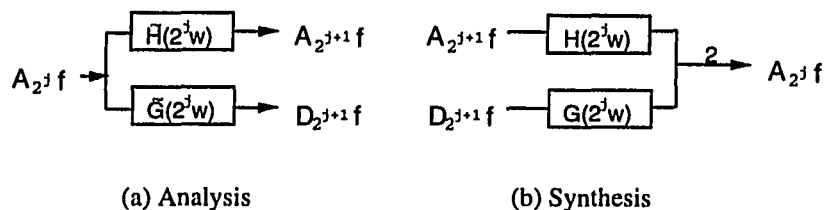
Figure 4.3: One-dimensional wavelet transform at full resolution

simplest way to appreciate the difference between the full resolution and conventional wavelet transforms is to consider the evolution of the smooth component in the one-dimensional case. Let the original signal be sampled at the frequency w_s. Then, as seen in Figure 4.5, the wavelet transform employs down-sampling to fill the available bandwidth in each octave, whereas the full resolution wavelet transform occupies lesss and less of the available spectrum at higher octaves. The same is true of the band-pass details components.

A second modification overcomes the limited dyadic sampling grid of the basic wavelet transform. By passing a Gaussian object of size σ through a sub-band decomposition and recording the peak response in each octave we obtain the scale bandwidth of the first four octaves, and observe significant dips in coverage between octaves 2 and 3, and between octaves 3 and 4, which will be explained in more detail in section 5.1. As proposed by Rioul [34] we can remedy this by computing a separate, three octave wavelet transform using $\times \sqrt{2}$ interpolated versions of $h(x)$ and $g(x)$. Thus, octaves 2 and 3 of the modified transform fill the gaps in coverage as desired. These inserted sub-bands are known as *voices* of the original octave decomposition.
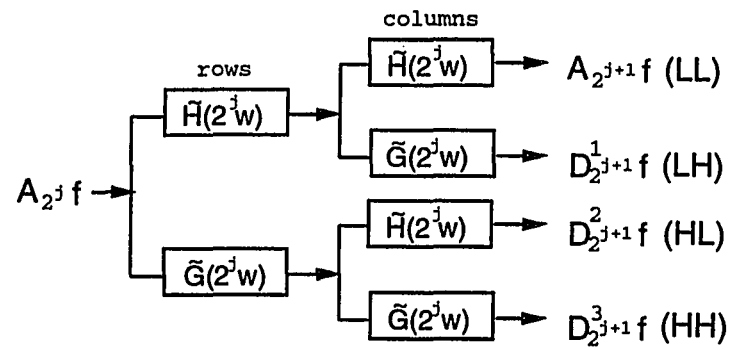
(a) Analysis
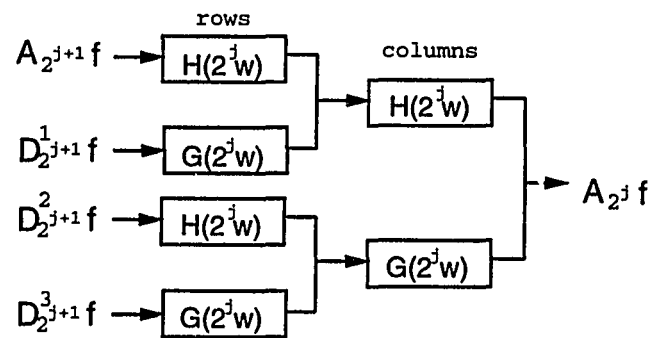
(b) Synthesis

Figure 4.4: Separable two-dimensional wavelet transform at full resolution

Wavelet Transform   Undecimated Wavelet
                     Transform

Octave 1

0              $W_s/4$ 0 $W_s/2$

Octave 2

0              $W_s/8$ 0 $W_s/2$
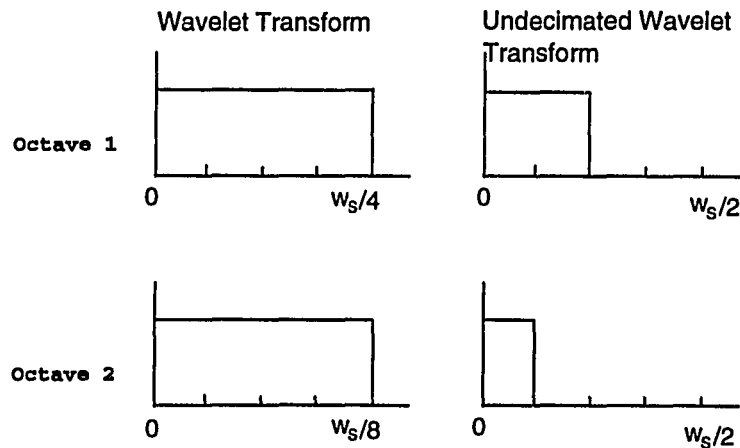
Figure 4.5: Frequency coverage of a 1D octave decomposition with (*left*) and without (*right*) downsampling. The redundancy of the undecimated decomposition is evident. Only the low frequency component is represented.

The scales covered by octaves 1 through 4 (including the voices) occupy the range

$0.4 \leq \sigma \leq 5.2$ which, given that a Gaussian object is visible over $\pm\sigma$ pixels, corresponds to objects of diameter $1 \rightarrow 10$ pixels, or $100 \ \mu m \rightarrow 1 \ mm$. This, as stated earlier, is the relevant scale range for μCa++s.

4.3 Finer Sampling in Scale

The discrete wavelet transform (DWT) is the octave-by-octave computation of the continuous wavelet transform (CWT), and is usually used in the discrete domain to represent signals in multiresolution. But, in signal analysis, the dyadic wavelet transform is generally not enough. Sampling in scale denser than the dyadic grid may be required depending on the application. It is desirable to obtain more wavelet coefficients with finer sampling in scale. In our research, we are interested only in

μCa++s whose variances are less than 10. As shown in Table 5.2, the details at the 1st, 2nd, 3rd and 4th octaves cover μCa++s whose variances are around 0.16, 1.4, 6.8 and 27.0, respectively. Those μCa++s whose variances lie between these octaves are not clear in the dyadic wavelet transform domain. The 1st octave of wavelet coefficients contain the high frequency components, and they are not useful in detecting μCa++s, whereas the 4th or higher octaves of wavelet coefficients cover the frequency range corresponding to μCa++s whose variances are larger than 10. But in almost all mammograms, μCa++s are in the variance range between 1 and 8, so they are rarely used. Thus, the 2nd and 3rd octaves of wavelet coefficients are largely used to detect the μCa++s. This dyadic grid is too sparse to represent μCa++s whose sizes are small. That is why we need to sample the scale parameter more densely between the 2nd and the 3rd octaves. So, we should focus on scales corresponding to actual μCa++s.

We can solve this problem by computing M voices per octave rather than octave by octave computation. This is computed by repeating the dyadic wavelet transform algorithm M times, with the scale $a = 2^j$ replaced by

$$a = 2^{j+m/M}, \quad m = 0, 1, ..., M-1, \tag{4.44}$$

where m is called the *voice* [34]. M should be determined according to the dynamic ranges of μCa++s. For each m, replace $\psi(x)$ by a slightly stretched wavelet $2^{-m/2M}\psi(2^{-m/M}x)$ in the expression $\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k)$. The wavelet basis

functions become

$$2^{-(j+m/M)/2}\psi(2^{-(j+m/M)}(x - k2^j)), \quad j, k \in Z, \quad m = 0, 1, ..., M - 1. \qquad (4.45)$$

Now, the computation of M voices per octave is done by applying the dyadic wavelet transform algorithm M times, with M different prototypes

$$2^{-(m/M)/2}\psi(2^{-(m/M)}x), \quad m = 0, 1, ..., M - 1. \qquad (4.46)$$

In this case, the parameters of each dyadic algorithm should be recomputed for each m. The M voices per octave computation requires about M times the computational load of the DWT. The filter coefficients and their interpolated versions $(M = 2)$ for the biorthogonal spline wavelets proposed by Cohen et al. [45] will be given in Table 5.1 and Table 5.3, respectively.

It will be shown in the next chapter that matched filters for detecting Gaussian signals in separable and nonseparable Markov noise can be implemented using the wavelet transform if the Laplacian of Gaussian wavelet is used. This is verified by computing ROC (receiver operating characteristic) curves using simulated Gaussian objects embedded in such noise.

CHAPTER 5

Matched Filter Implemented using the Wavelet Transform

In chapter 3, it was shown that the prewhitening matched filter for detecting a Gaussian object in highly correlated Markov noise resembled a Laplacian of Gaussian. If we try to detect objects of varying size, we should have as many prewhitening matched filters as different sizes. Thus, σ in (3.15) can be considered a variable, giving the object $f(x, y; \sigma)$. We can detect $f(x, y; \sigma)$ by convolution with the prewhitening matched filters of varying σ in the following way,

$$o(x, y) = h_{mf}(x, y) * f(x, y; \sigma), \tag{5.1}$$

followed by thresholding the output $o(x, y)$.

This can be easily implemented by decomposing the image $f(x, y; \sigma)$ into multiple scales using the wavelet transform. In the following section, it is shown how the separable and nonseparable prewhitening matched filters developed in (3.16) and (3.20) can be implemented via a wavelet transform using Laplacian of Gaussian wavelets. In order to support the above assertion, the detection performance of the wavelet transform is compared with that of the matched filter by computing ROC curves using simulated Gaussian objects embedded in both separable and nonseparable Markov noise.

5.1 Relationship between Matched Filtering and the Wavelet Transform

Consider the 2D wavelet transform with input $i(x, y)$, and sub-band components $A_{2^j}(x, y)$, $D_{2^j}^1(x, y)$, $D_{2^j}^2(x, y)$, $D_{2^j}^3(x, y)$ in octave j. We obtain the four corresponding, separable transfer functions

$$
\begin{aligned}
P_{LL}^j(w_x, w_y) &= H_\phi^j(w_x) H_\phi^j(w_y) \\
P_{LH}^j(w_x, w_y) &= H_\phi^j(w_x) H_\psi^j(w_y) \\
P_{HL}^j(w_x, w_y) &= H_\psi^j(w_x) H_\phi^j(w_y) \\
P_{HH}^j(w_x, w_y) &= H_\psi^j(w_x) H_\psi^j(w_y),
\end{aligned}
\tag{5.2}
$$

where, for example, $P_{LH}^j(w_x, w_y) = D_{2^j}^1(w_x, w_y)/I(w_x, w_y)$, etc., and

$$
H_\phi^j(w) = \frac{\hat{\phi}^*(2^j w)}{\hat{\phi}^*(w)}, \qquad H_\psi^j(w) = \frac{\hat{\psi}^*(2^j w)}{\hat{\phi}^*(w)}.
\tag{5.3}
$$

Here, $\hat{\phi}(w)$ and $\hat{\psi}(w)$ represent the Fourier transforms of the scaling function $\phi(x)$ and the wavelet $\psi(x)$, respectively.

For wavelets close in form to the Laplacian of Gaussian, the details filters can be approximated as point spread functions

$$
\begin{aligned}
p_{LH}(x, y) &= -\frac{\partial^2}{\partial y^2} G(x, y) \\
p_{HL}(x, y) &= -\frac{\partial^2}{\partial x^2} G(x, y) \\
p_{HH}(x, y) &= \frac{\partial^4}{\partial x^2 \partial y^2} G(x, y),
\end{aligned}
\tag{5.4}
$$
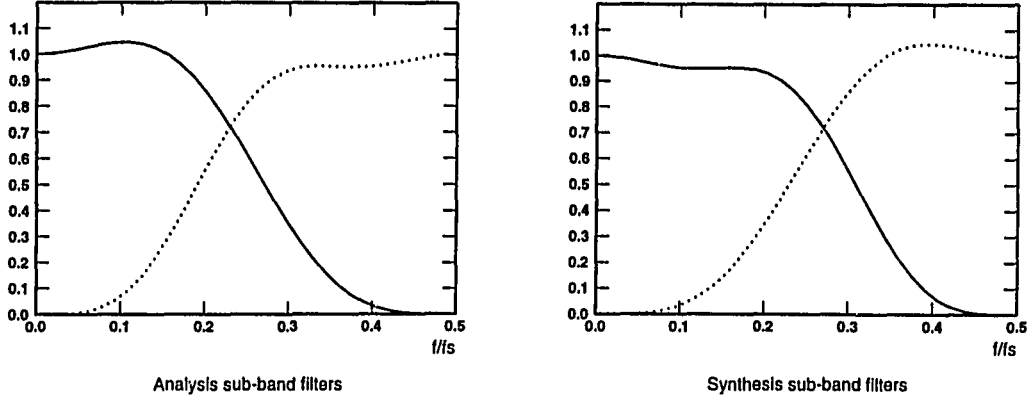
Figure 5.1: Frequency responses of the sub-band filters $\tilde{H}$, H (*solid*) and $\tilde{G}$, G (*broken*). (a) Analysis sub-band filters. (b) Synthesis sub-band filters.

where $G(x,y) = e^{-\frac{(x^2+y^2)}{2\sigma^2}}$ and σ depends on octave j. These point spread functions can be computed explicitly as

$$p_{LH}(x,y) = \frac{1}{\sigma^2}(1 - \frac{y^2}{\sigma^2})e^{-\frac{(x^2+y^2)}{2\sigma^2}}$$

$$p_{HL}(x,y) = \frac{1}{\sigma^2}(1 - \frac{x^2}{\sigma^2})e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{5.5}$$

$$p_{HH}(x,y) = \frac{1}{\sigma^4}(1 - \frac{x^2}{\sigma^2})(1 - \frac{y^2}{\sigma^2})e^{-\frac{(x^2+y^2)}{2\sigma^2}}.$$

The transfer functions corresponding to the components $D_{2j}^3(x,y)$ and $D_{2j}^1(x,y) + D_{2j}^2(x,y)$ compute to be

$$P_{HH}(w_x, w_y) = w_x^2 w_y^2 e^{-\frac{\sigma^2(w_x^2+w_y^2)}{2}}$$

$$P_{LH}(w_x, w_y) + P_{HL}(w_x, w_y) = (w_x^2 + w_y^2)e^{-\frac{\sigma^2(w_x^2+w_y^2)}{2}}. \tag{5.6}$$

We can see that $P_{HH}(w_x, w_y)$ approaches the PWMF for the separable Markov process given in (3.16), while $P_{LH}(w_x, w_y) + P_{HL}(w_x, w_y)$ approach the PWMF for the nonseparable Markov process given in (3.20) if the correlation coefficients ρ_1

and ρ_2 are high (*i.e.* α_x and α_y are small). Thus, the details images $D^3_{2^j}(x,y)$ and

$D^1_{2^j}(x,y) + D^2_{2^j}(x,y)$ can be regarded as the outputs of a bank of PWMFs in two

dimensions for separable and nonseparable Markov noise image, respectively. In our

studies, we selected the biorthogonal spline wavelets proposed by Cohen *et al.*[45],

which are close in form to the Laplacian of Gaussian, and therefore appropriate for

multiscale matched filtering. The filter coefficients corresponding to the wavelets are

shown in Table 5.1. Figure 5.1 shows $\tilde{H}$, H, $\tilde{G}$ and G for the biorthogonal wavelets,

while $\tilde{\phi}$, ϕ, $\tilde{\psi}$ and ψ are depicted in Figure 5.2.

Figure 5.3 compares the separable HH details filter $p^j_{HH}(x,0)$ resulting from the

biorthogonal spline wavelet with the separable matched filter $h_m(x)$ whose transfer

function is given in (3.19), where

$$
\begin{aligned}
p^j_{HH}(x,y) &= \mathcal{F}^{-1} P^j_{HH}(w_x, w_y) \\
&= \mathcal{F}^{-1}\left(\frac{\hat{\psi}^*(2^j w_x)}{\hat{\phi}^*(w_x)} \cdot \frac{\hat{\psi}^*(2^j w_y)}{\hat{\phi}^*(w_y)}\right).
\end{aligned}
\tag{5.7}
$$

Since the filters are separable they can be represented in one dimension and thus easily

compared. Likewise, the nonseparable LH + HL details filter $p^j_{LH}(x,y) + p^j_{HL}(x,y)$

and the nonseparable matched filter $h_{NONSEP}(x,y)$ whose transfer function is given

in (3.20), where

$$
\begin{aligned}
p^j_{LH}(x,y) + p^j_{LH}(x,y) &= \mathcal{F}^{-1}(P^j_{LH}(w_x, w_y) + P^j_{LH}(w_x, w_y)) \\
&= \mathcal{F}^{-1}\left(\frac{\hat{\phi}^*(2^j w_x)}{\hat{\phi}^*(w_x)} \cdot \frac{\hat{\psi}^*(2^j w_y)}{\hat{\phi}^*(w_y)} + \frac{\hat{\psi}^*(2^j w_x)}{\hat{\phi}^*(w_x)} \cdot \frac{\hat{\phi}^*(2^j w_y)}{\hat{\phi}^*(w_y)}\right),
\end{aligned}
\tag{5.8}
$$

are compared in Figure 5.4 using their profiles.
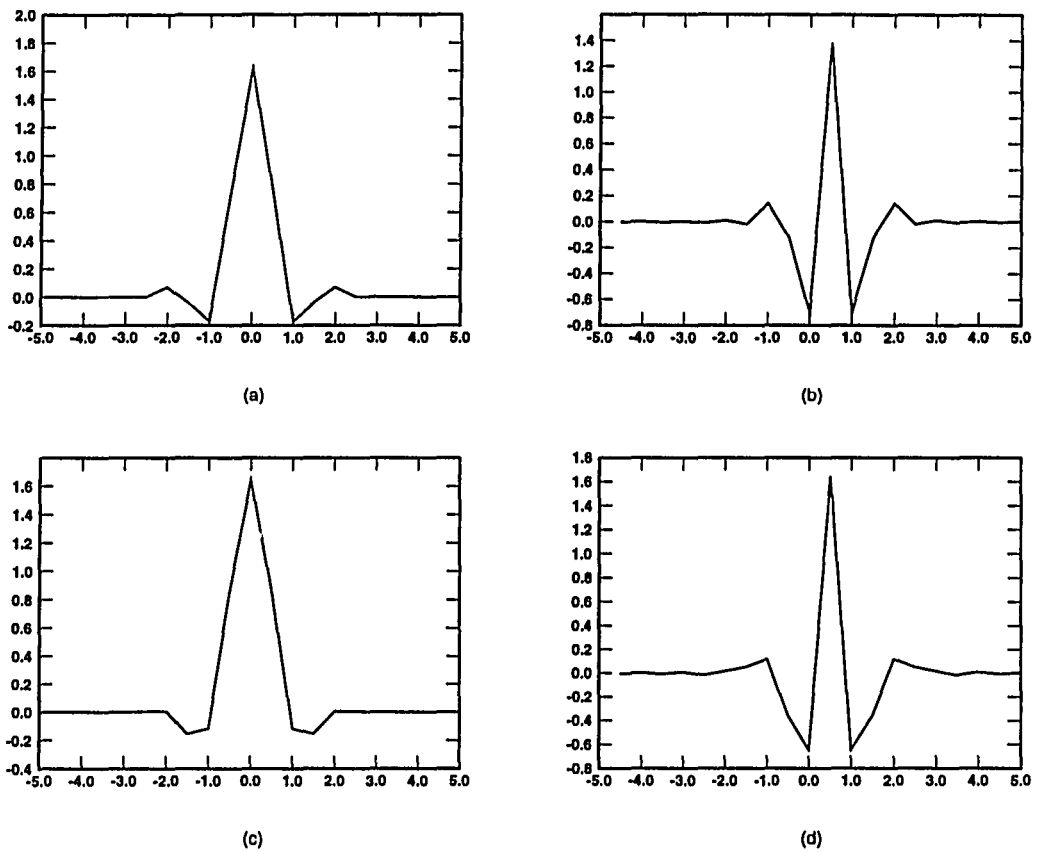
Figure 5.2: Scaling functions and wavelets: (a) Analysis scaling function $\tilde{\phi}$, (b) Analysis wavelet $\tilde{\psi}$, (c) Synthesis scaling function ϕ, (d) Synthesis wavelet ψ.

n	h	$\tilde{h}$
0	0.602949	0.557543
±1	0.266864	0.295636
±2	-0.078223	-0.028772
±3	-0.016864	-0.045636
±4	0.026749	0

Table 5.1: Filter coefficients for the analysis (h) and synthesis $(\tilde{h})$ spline wavelets.

Octave	α	σ
1	0.01	0.4
2	0.01	1.2
3	0.01	2.6
4	0.01	5.2
5	0.01	10.8

Table 5.2: The wavelet transform approximates these values of α and σ in the matched filter.

The details filters are close approximation to the matched filters over four octaves. Table 5.2 shows parameters α and σ of $H_m(w)$ in (3.19) and $H_{NONSEP}(w_x, w_y)$ in (3.20) corresponding to $P^j_{HH}(w, 0)$ and $P^j_{LH}(w_x, w_y) + P^j_{HL}(w_x, w_y)$. The filter shapes change slightly with scale due to the scaling term 2^j only in the numerator. But if the scaling function $\phi(x)$ is sufficiently smooth, then the change in shape is negligible.

The closest fit between the above filter and the matched filter occurs when the Markov noise is highly correlated with $\rho \rightarrow 1$. As ρ gets smaller the negative lobes of the matched filter shrink, the filter approaching a Gaussian in the limit as ρ approaches 0, as shown in Figure 3.8. However the wavelet is constrained by $\int \psi(x)dx = 0$, so this method only works well for highly correlated noise. Fortunately,
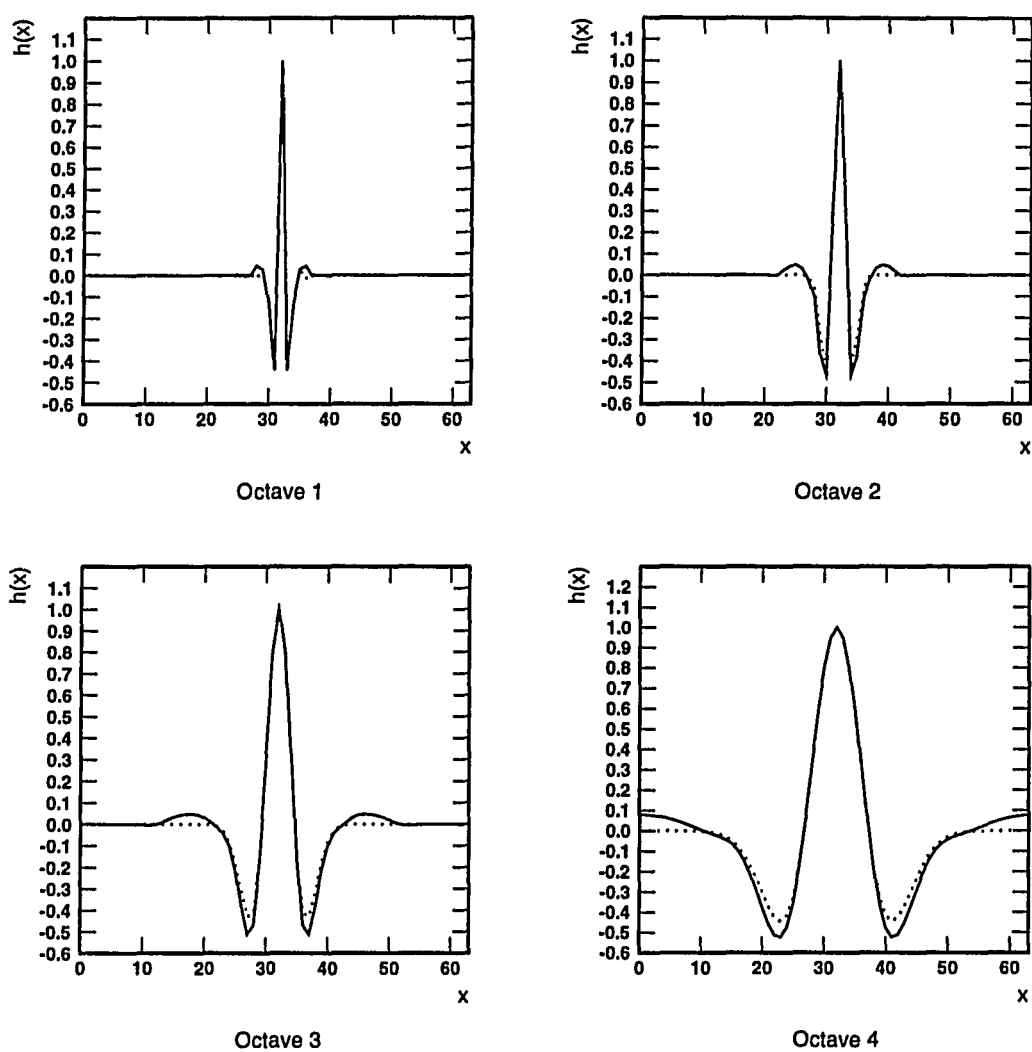
Figure 5.3: Comparison of the point spread functions of the separable details filter $p^j_{HH}(x, 0)$ (*solid*) and the separable matched filter $h_m(x)$ (*broken*) across four octaves, shown from *upper left* to *lower right*.

Markov noise with ρ close to unity is a reasonable model for the large-scale texture in mammograms. We have seen this to be true of the Nijmegen database.

In this noise regime a multiscale image decomposition with the above wavelet creates a sequence of HH and LH + HL details subimages which are equivalent to applying the matched filters in (3.16) and (3.20) at increasing scale in separable and nonseparable Markov noise image, respectively. The smallest μCa++s appear in the low octaves and *vice versa*. Microcalcifications of a certain size give rise to a strong peak response in both HH and LH + HL at the corresponding octave level. Computing a combination of both responses provides an even stronger indication of the presence and location of μCa++s. The combination of separable and nonseparable filter responses is supported by the composite texture model of breast tissue, as shown in Figure 2.6.

We have simulated the wavelet transform using images containing Gaussian objects with varying size and amplitude 200. The scale range each octave covers, shown in Figure 5.5, represents the peak values of the HH details subimages as σ and octave increase. From this figure, we can see that dyadic wavelet transform can cover the scale range from 0.0 to 6.0 by decomposing the image into 4 octaves of details subimages. Given that a Gaussian object is visible over $\pm\sigma$ pixels, this range corresponds to objects of diameter $1 \rightarrow 10$ pixels, or $100 \; \mu m \rightarrow 1 \; mm$. This, as stated earlier, is the relevant scale range for μCa++s. However, we observe significant dips in coverage between octaves 2 and 3, and between octaves 3 and 4. For Gaussian
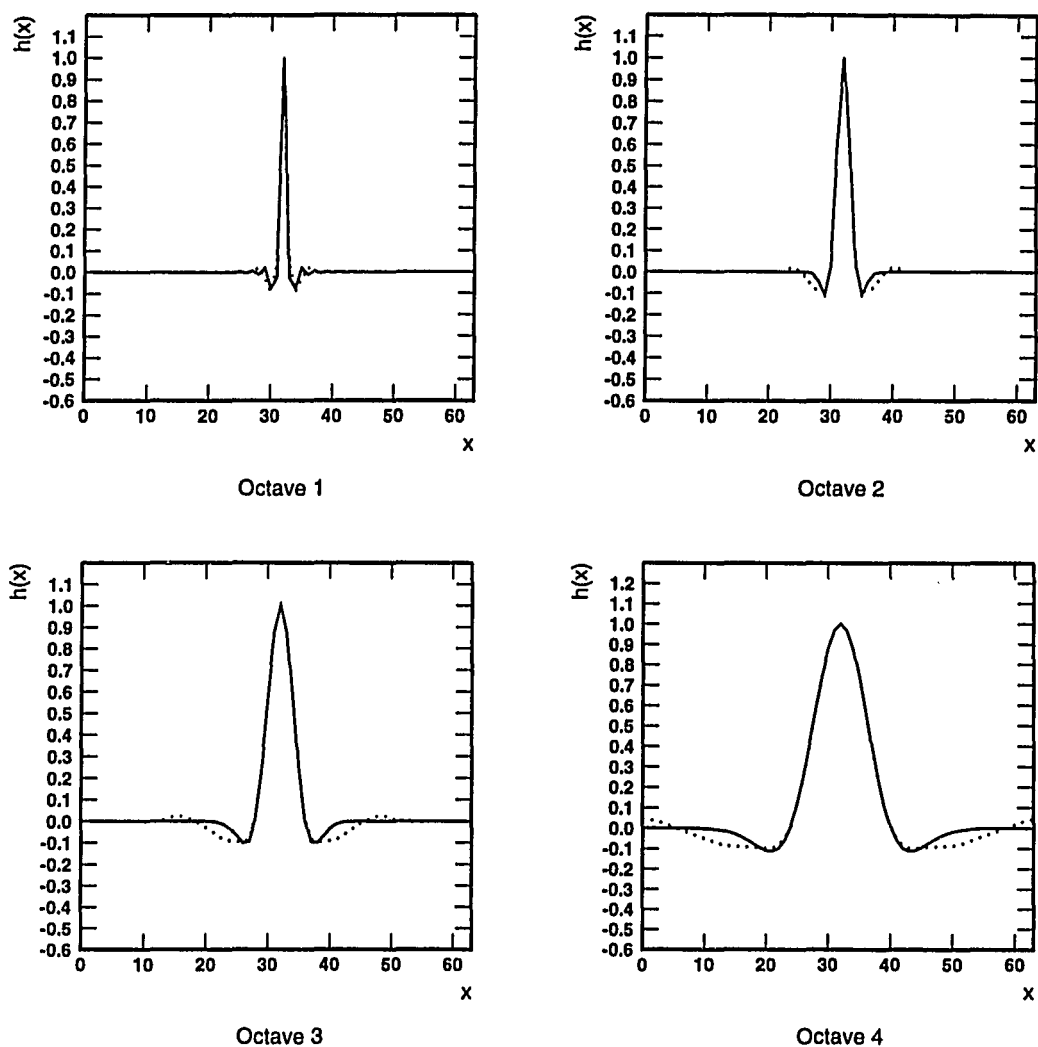
Figure 5.4: The point spread functions of the nonseparable details filter $p^j_{LH}(x,y)$ + $p^j_{HL}(x,y)$ (*solid*) and the nonseparable matched filter $h_{NONSEP}(x,y)$ (*broken*) are compared across four octaves using their row and column profiles, shown from *upper left* to *lower right*.
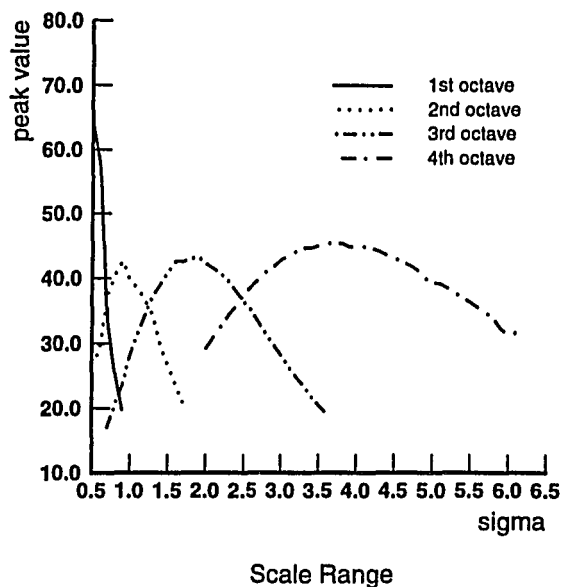
Figure 5.5: Scale range each octave covers.

objects with σ assumed 1.2 or 2.5, the performance is expected to be worst. We can remedy this by computing a separate, three octave wavelet transform using $\times \sqrt{2}$ interpolated versions of $h(x)$ and $g(x)$, mentioned in section 4.3. Table 5.3 lists the interpolated filter coefficients. Thus, octaves 2 and 3 of the modified transform fill the gaps in coverage as desired.

In the next section, the detection performance of the wavelet transform is compared with that of the matched filter by computing ROC curves using simulated Gaussian objects embedded in both separable and nonseparable Markov noise.

n	h	g	$\tilde{h}$	$\tilde{g}$
-6	0.001468	0.004865	0.014129	0.013926
-5	-0.009883	-0.006096	0.017579	0.053966
-4	-0.039354	0.032090	-0.026275	-0.069523
-3	-0.032041	0.028221	-0.061015	-0.220655
-2	0.091054	-0.082874	0.049910	0.076328
-1	0.291691	0.059997	0.292614	0.561895
0	0.394243	-0.481643	0.426351	-0.501956
1	0.291691	0.593791	0.292614	0.079261
2	0.091054	0.107794	0.049910	-0.040136
3	-0.032041	-0.205431	-0.061015	0.041269
4	-0.039354	-0.095943	-0.026275	0.014346
5	-0.009883	0.013006	0.017579	-0.016223
6	0.001468	0.008187	0.014129	0.007383

Table 5.3: Interpolated filter coefficients for the analysis (h and g) and synthesis ($\tilde{h}$ and $\tilde{g}$) spline wavelets.

5.2 Simulation of Matched Filter and Wavelet Transform

We showed in the previous section that the wavelet transform can act as a bank of multiscale matched filters for detecting Gaussian objects in Markov noise. The performance of the matched filter compared to that of the wavelet transform is computed using a test image comprising Gaussian objects in Markov noise. First, we create the system function

$$H(w_x, w_y) = \sqrt{S_{oo}(w_x, w_y)}e^{j\theta(w_x,w_y)}, \qquad (5.9)$$

where $S_{oo}(w_x, w_y)$ is the noise power spectrum and $\theta(w_x, w_y)$ is an arbitrary angle. White noise with unit variance is passed through this system yielding an output

whose power spectrum is

$$S_{oo}(w_x, w_y) = S_{ii}(w_x, w_y)|H(w_x, w_y)|^2. \tag{5.10}$$

For convenience, we assume that $S_{oo}(w_x, w_y)$ corresponds to a separable Markov process as given in (3.3). Hence, from (3.3) and (5.10),

$$
\begin{aligned}
H(w_x, w_y)H^*(w_x, w_y) &= S_{oo}(w_x, w_y) \\
&= \frac{(1-\rho_1^2)}{(1-\rho_1 e^{jw_x})(1-\rho_2 e^{-jw_x})} \frac{(1-\rho_1^2)}{(1-\rho_1 e^{-jw_y})(1-\rho_2 e^{-jw_y})}.
\end{aligned} \tag{5.11}
$$

Thus, a first order separable Markov noise image can be generated by passing white noise through the causal system function

$$H(w_x, w_y) = \frac{\sqrt{1-\rho_1^2}}{(1-\rho_1 e^{-jw_x})} \frac{\sqrt{1-\rho_2^2}}{(1-\rho_2 e^{-jw_y})}. \tag{5.12}$$

Likewise, the first order nonseparable Markov noise image can be generated by replacing $S_{oo}(w_x, w_y)$ with $S_{NONSEP}(w_x, w_y)$ in (3.7). Examples of separable and nonseparable first order Markov noise images are located in the upper right and lower left of Figure 2.6. Then, a uniformly-spaced array of simulated Gaussian objects is added. Figure 5.6 shows the simulated Gaussian object images embedded in such noise. Although neither the number nor the spatial distribution of the Gaussians are realistic, the array makes it possible to test multiple object sites in one experiment. Given the homogenity of the background image, this is practically the same as running multiple experiments with single object inserted at random locations.

To justify this combination of matched filter forms, we compare the detection performance of the wavelet method with that of the matched filters by computing
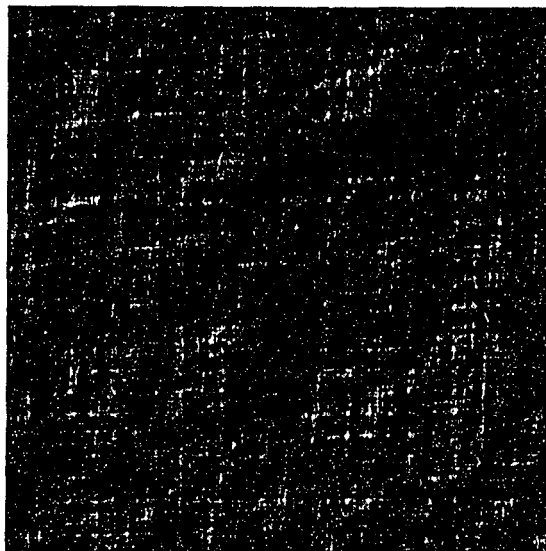
(a)
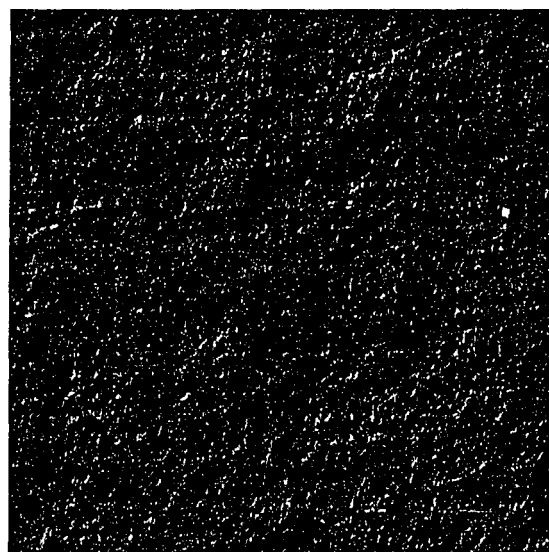
(b)

Figure 5.6: (a) Simulated Gaussian objects embedded in separable Markov noise. (b) Simulated Gaussian objects embedded in nonseparable Markov noise.

ROC (receiver operating characteristic) curves given in Figure 5.7 using simulated Gaussian objects embedded in both separable and nonseparable Markov noise. The ROC curve is explained in more detail in Appendix A. The number of objects is 484, the object size is $\sigma = 2.6$, the SNR used is 0.2 (peak object intensity $\div \sigma_n$) and the Markov correlation parameter $\alpha = 0.01$. Since the objects are of fixed size in the test, only one octave band of the wavelet decomposition is involved in the ROC computation. Figure 5.8 and Figure 5.9 show several output images when the images given in Figure 5.6 are applied to the matched filter and the wavelet transform. The performance of the $P_{HH}^j(w_x, w_y)$ filter is almost identical to that of the matched filter based on separable noise. Likewise, the LH + HL results compare to those of the matched filter based on nonseparable noise.

In the next chapter a nonlinear algorithm for detecting μCa++s is described in detail and the intermediate results at each step of the algorithm are shown.
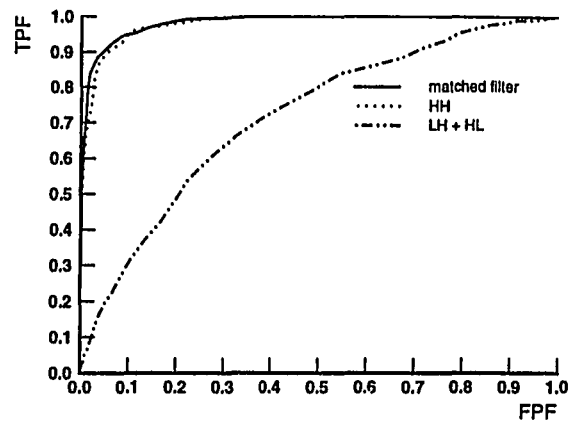
ROC Curve for Separable Model
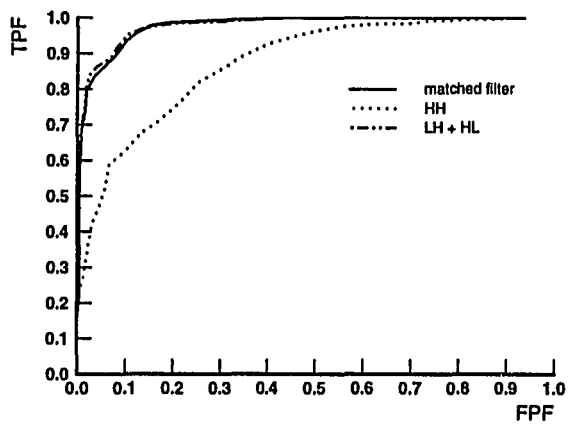
ROC Curve for Nonseparable Model

Figure 5.7: ROC curves of matched filters and details filters of one octave based on Gaussian object in (a) separable Markov noise, (b) nonseparable Markov noise. In (a) the HH filter, and in (b) the LH + HL filters are close to optimum.
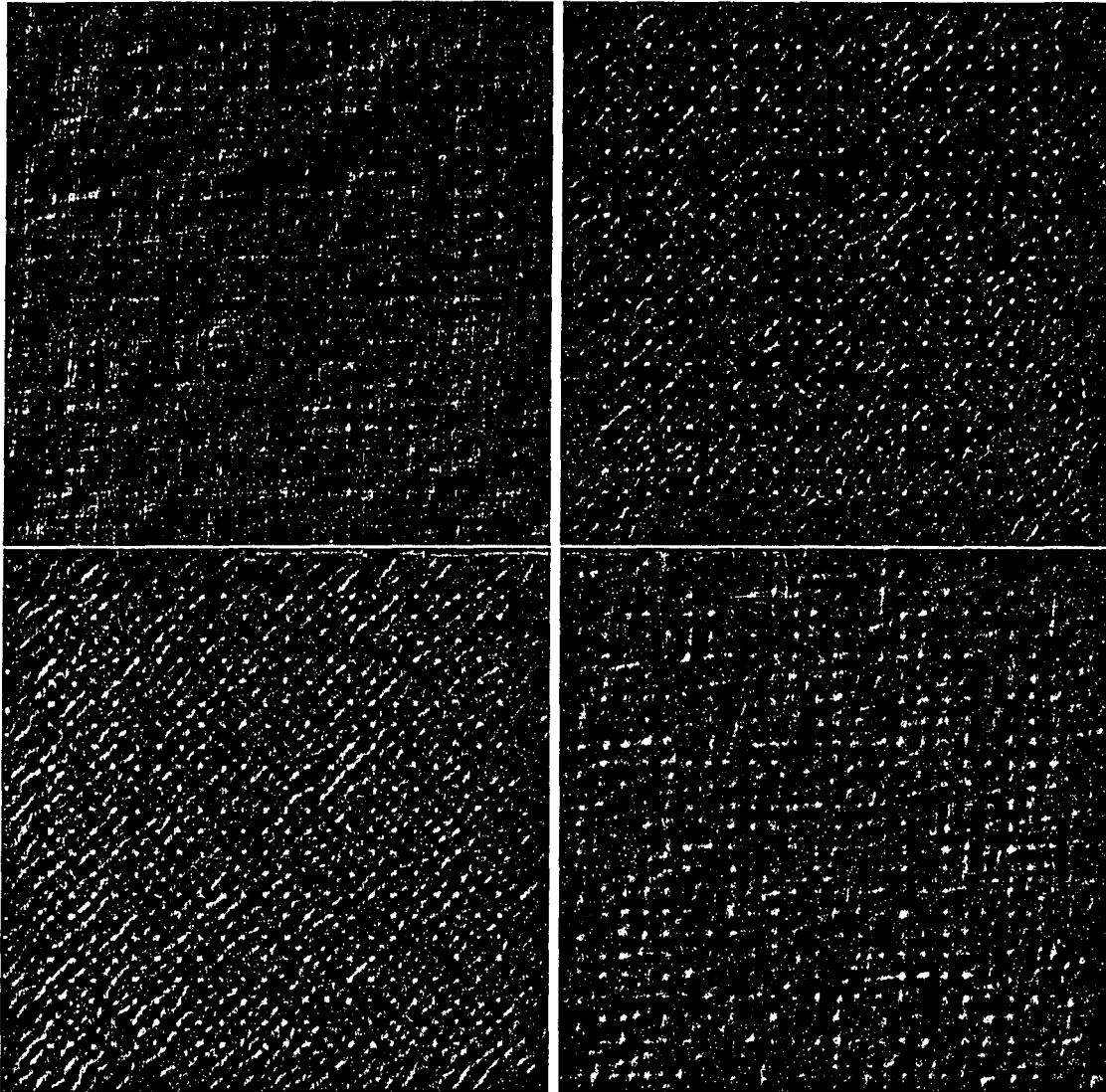
Figure 5.8: Output images of matched filter and wavelet transform. *Upper left* Simulated Gaussian objects embedded in separable Markov noise. *Upper right* Matched filter output. *Lower left* HH image of 3rd octave. *Lower right* LH + HL image of 3rd octave.
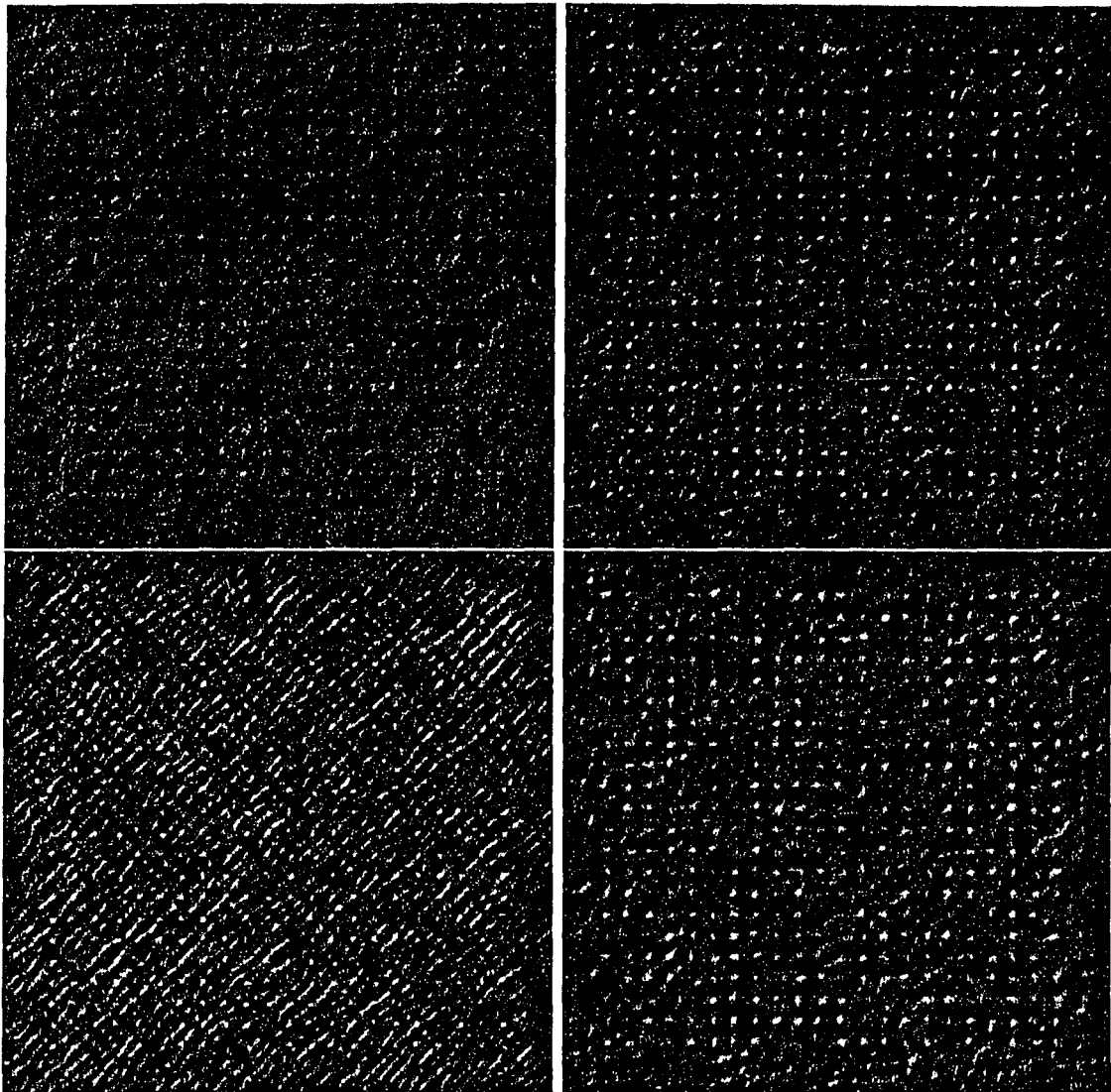
Figure 5.9: Output images of matched filter and wavelet transform. *Upper left* Simulated Gaussian objects embedded in nonseparable Markov noise. *Upper right* Matched filter output. *Lower left* HH image of 3rd octave. *Lower right* LH + HL image of 3rd octave.

CHAPTER 6

Recognition of Microcalcifications

In these studies we have designed a nonlinear algorithm for detecting roughly circular μCa++s ranging in size from 1 - 8 pixels (0.1 - 0.8 mm) in diameter. Although μCa++s vary in outline and degree of elongation, they can be modeled as a circularly-symmetric Gaussian function as in

$$G_s(x,y) = \frac{1}{2\pi s^2}e^{-\frac{x^2+y^2}{2s^2}},\tag{6.1}$$

where s is a scale parameter. This can be supported by Figure 2.5(b). This modeling provides valuable merits. As stated in the previous chapter, the application of the wavelet transform to Gaussian signals, which corresponds to the output of matched filters for detecting them, yields Gaussian related signals in the s direction of scale space although both size and height vary. This property is very important because we want to implement scale independent methods, which immediately implies that changing the resolution should affect the size and perhaps the amplitude but certainly not the structure of the operators.

Figure 6.1 shows the overall recognition process. As explained in Section 2.1, digitized mammograms with 12 bit resolution are first mapped using the noise-equalizing look-up table. Then we employ a four octave wavelet transform, implemented at full
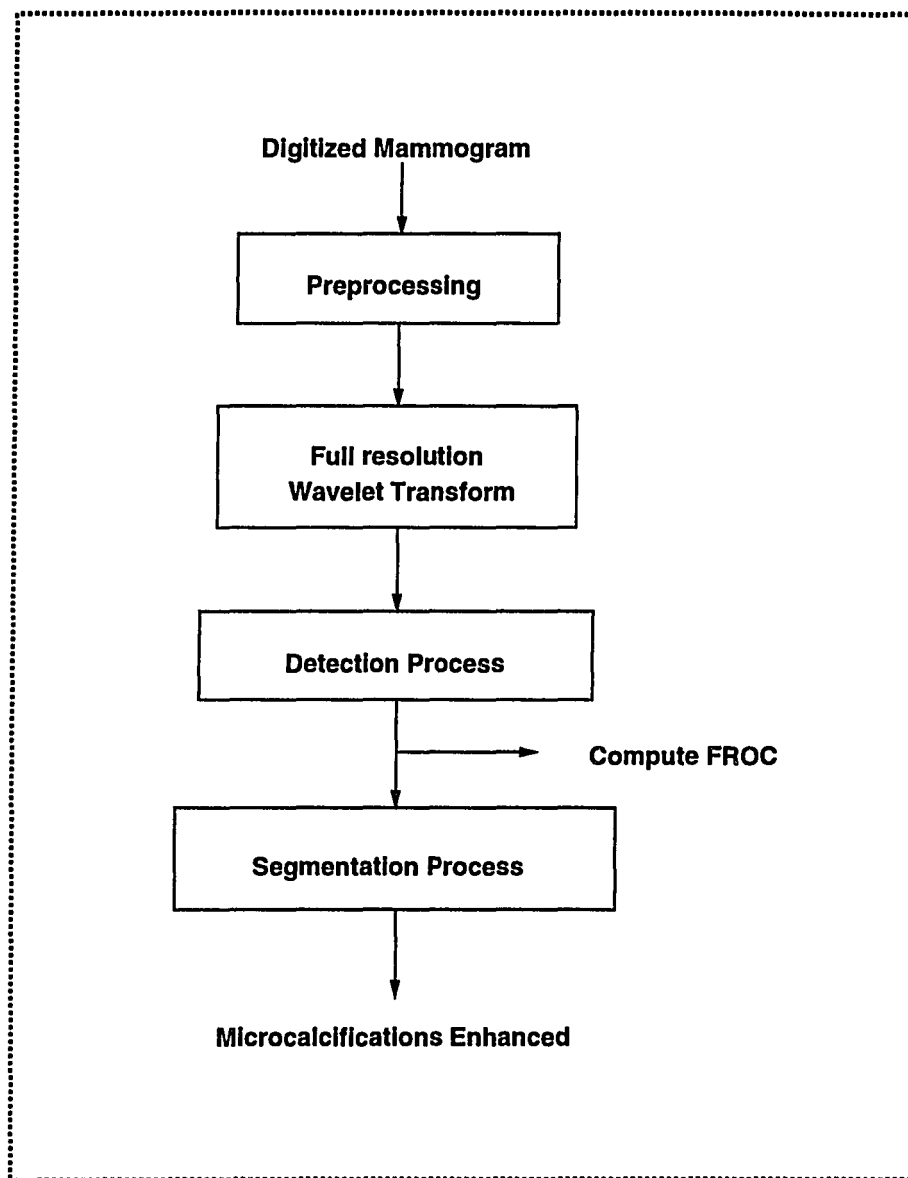
Figure 6.1: Overall recognition process.
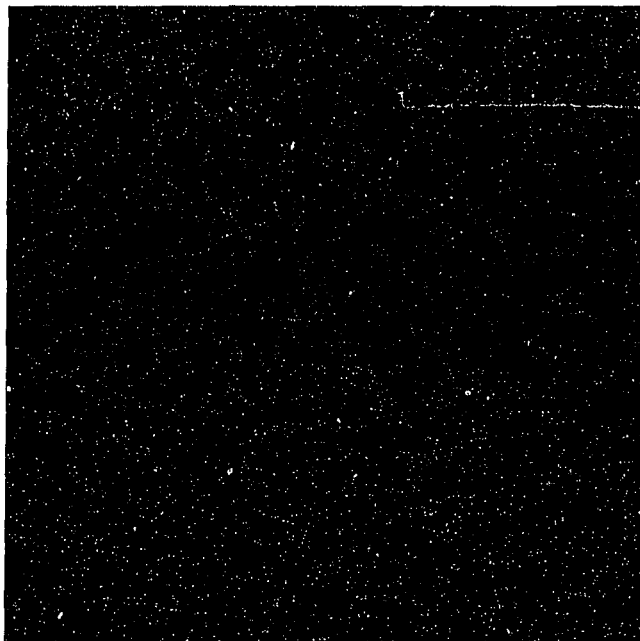
Figure 6.2: 512 × 512 Sub-region of Figure 2.2.

resolution with intermediate voices computed between octaves 1 and 2, between octaves 2 and 3 and between octaves 3 and 4. As a result of using the Laplacian of Gaussian wavelet, the separable 2D filters which transform the input image into the HH (D_{2j}^3) details sub-bands are closely related to the prewhitening matched filters for detecting Gaussian objects in separable Markov noise. The sum of the two sub-band components of the wavelet transform, LH + HL (D_{2j}^1 + D_{2j}^2), which can be approximated as an output of the PWMF in nonseparable Markov noise, helps to preserve the shape of any detected μCa++s. Figure 6.2 shows the 512 × 512 mammogram cropped from Figure 2.2, still at full resolution.

Figures 6.3 and 6.4 show the sub-band images D^3_{2j}, $D^1_{2j} + D^2_{2j}$ of Figure 6.2 at

$1 \leq j \leq 4$, respectively. This analysis phase of the computation is equivalent to a

bank of multiscale matched filters for detecting objects of diameter 100 $\mu m \rightarrow 1\ mm$.

Our recognition algorithm is composed of two major processes: detection and

segmentation. Although the μCa++s can be detected in the detection process, their

individual shapes are not preserved in this stage. That is why the segmentation

process is required.

6.1 Detection process

The major steps in the detection process, as indicated in Figure 6.5, are: (1) find

the candidate μCa++ sites (potential μCa++ locations) within the breast region, (2)

classify the potential μCa++s into true and false positives. False alarms are usually

caused by edges and other linear structures in background texture.

6.1.1 Finding candidate μCa++s

This node is a procedure that finds specific areas which may contain μCa++s.

Candidate μCa++s can be detected by thresholding a combination of the details sub-

bands, which are equivalent to a bank of multiscale matched filters. Two methods

have been proposed to achieve this purpose.

1st octave   2nd octave

3rd octave   4th octave
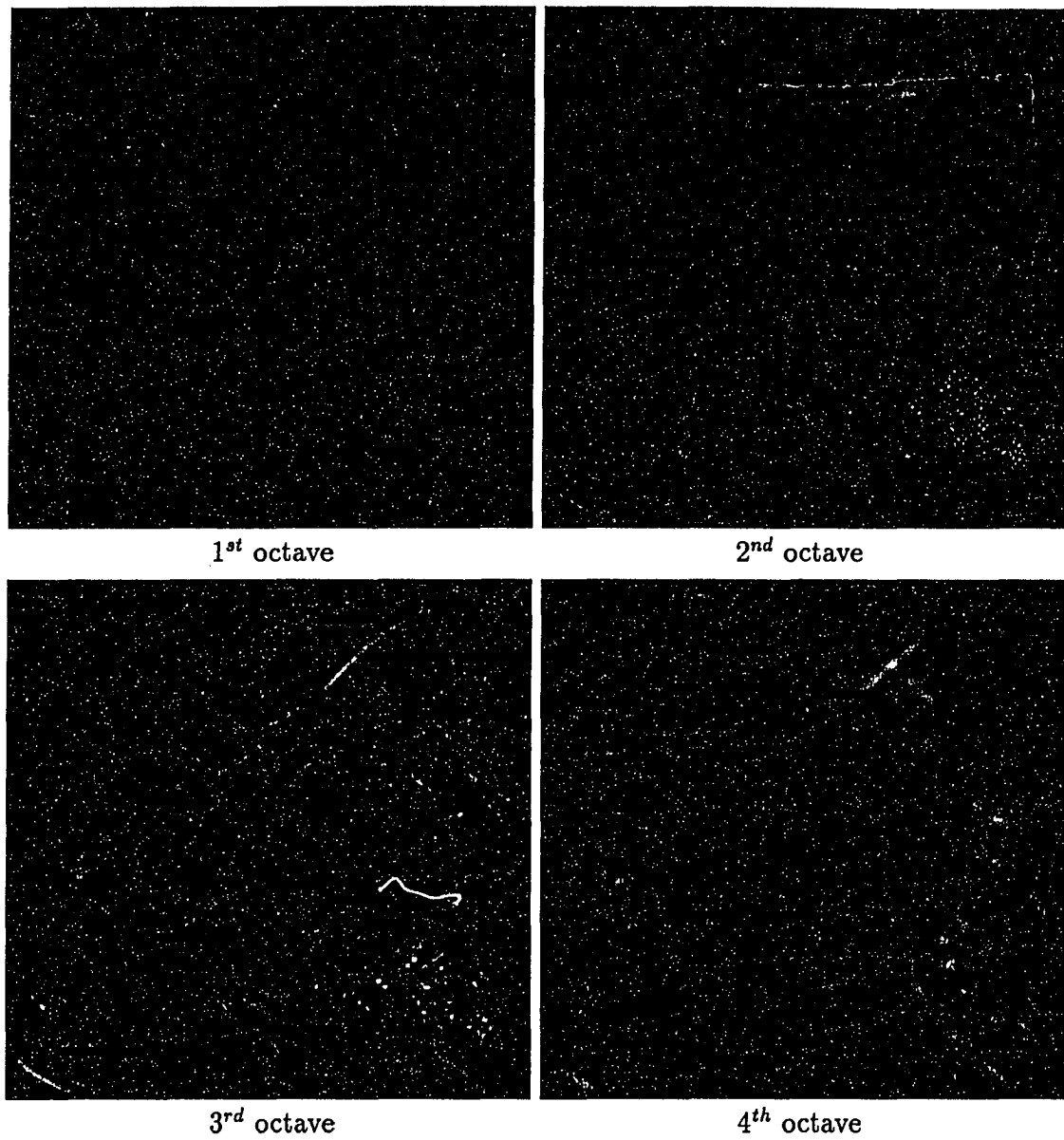
Figure 6.3: Details images $D_{2^j}^3$ (HH) at $1 \leq j \leq 4$.

1st octave · 2nd octave

3rd octave · 4th octave
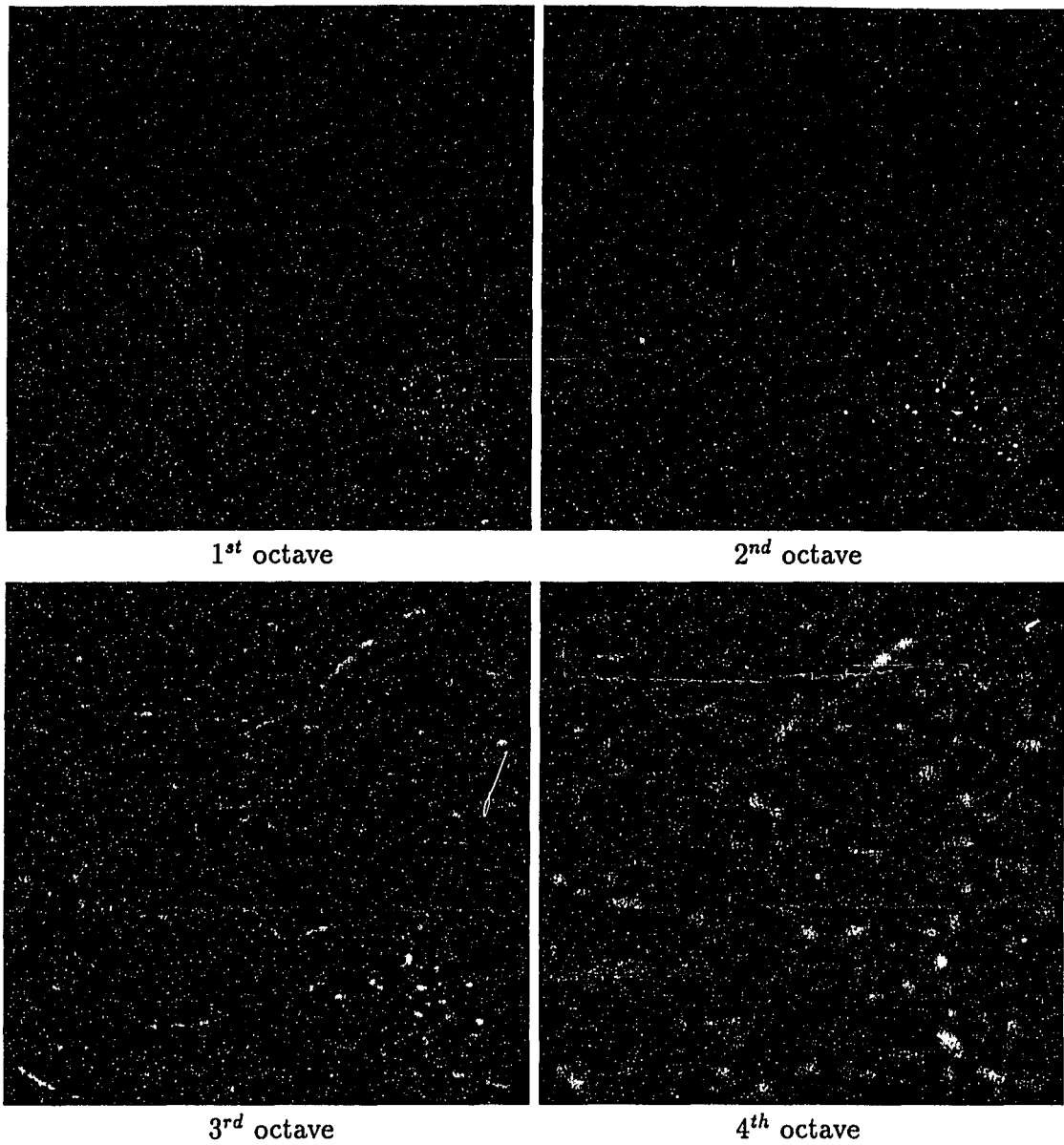
Figure 6.4: Sum of two details images $D^1_{2^j}+D^2_{2^j}$ (LH+HL) at $1 \leq j \leq 4$.
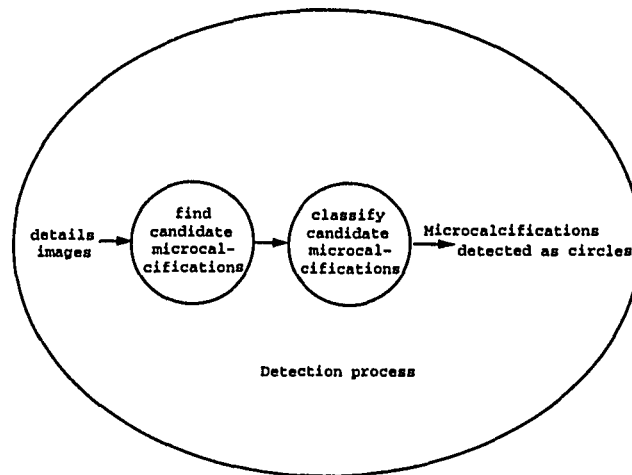
Figure 6.5: Detection process.

6.1.1.1 Method #1

The outputs HH and LH + HL from each octave are thresholded at some fixed percentile of the histogram of each component. (A common threshold for all sub-bands is required for FROC computation.) The detected (binary) images from all octaves are then logically ORed to yield the binary map of detected pixels. Figure 6.6 shows the detailed procedure for computing the binary map, one of which is given in Figure 6.7, where the gray circle represents a true cluster identified by the radiologist.

This detection model can be explained mathematically. The probability of detection for this model $P(detect)$ is a function of the probabilities of detection for each channel (sub-band) $P_i(detect)$, according to

$$P(detect) = 1 - \prod_{i=1}^{L}(1 - P_i(detect)), \qquad (6.2)$$

where L (= 8) is the total number of the channels [74].
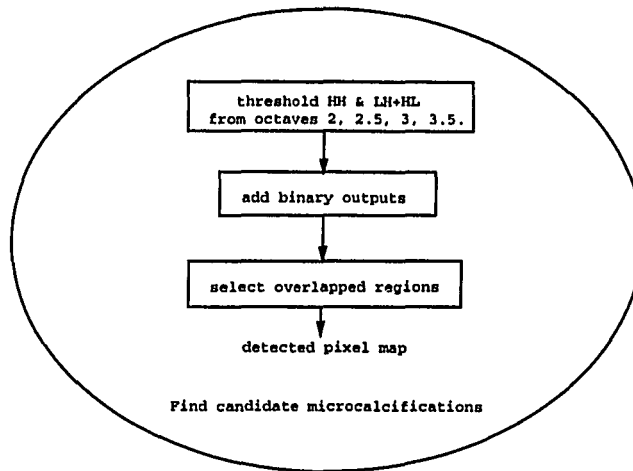
Figure 6.6: A procedure for detecting candidate μCa++s, using Method #1.

It is very difficult to evaluate the performance of this method, because each channel responds differently as the size of the μCa++s vary.

6.1.1.2 Method #2

The most elementary situation in detection theory involves discrimination between two hypotheses, H and K, based on a single observation or realization x of a random variable X; that is,

$$H : \text{ hypothesis on the observation } x, \text{ e.g. } X = N$$
$$K : \text{ alternative hypothesis on } x, \text{ e.g. } X = N + s. \tag{6.3}$$

Here we assume that N is a noise random variable, s is a signal to be detected and X is the observation random variable. Given x, a realization of X, we need to decide whether x belongs to H or K, namely, "signal-absent" or "signal-present." In order
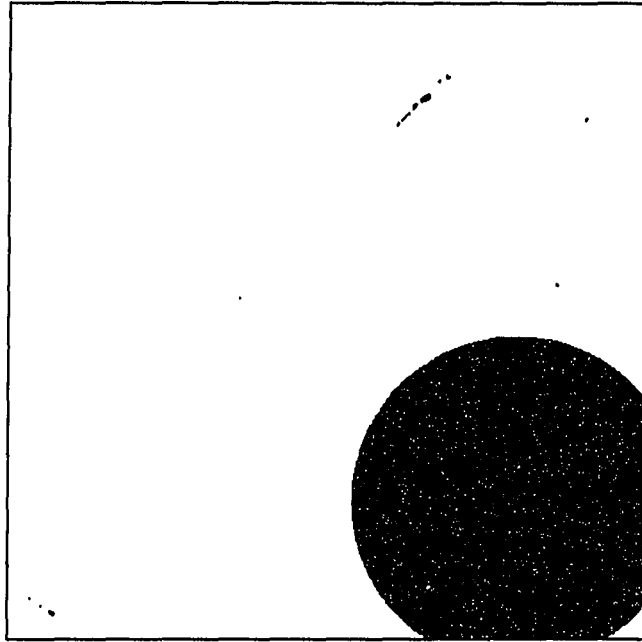
Figure 6.7: This binary map image shows candidate μCa++s detected by Method #1.

to accomplish this task, we need a decision rule $d = d(x)$. The decision rule $d = d(x)$ partitions the X space $\Omega = \{x \in X\}$ into two disjoint regions Ω_H and Ω_K, where Ω_H $\bigcup \Omega_K = \Omega$ and $\Omega_H \bigcap \Omega_K = \phi$.

It is desired to obtain the classifier to separate the two distributions $p(\lambda/H)$ and $p(\lambda/K)$ effectively. A parameter d called "detectability" can be defined as

$$d = \frac{(< \lambda/K > - < \lambda/H >)}{\sqrt{p(H)\sigma_H^2 + p(K)\sigma_K^2}}, \tag{6.4}$$

to measure quantitatively the separability between the two distributions, where $< \lambda/H >$ and σ_H^2 are the mean and variance of the test statistic λ under the hypothesis H. A Hotelling observer is known to be the optimum detector in the sense that

it maximizes the detectability [74]. Thus, a Hotelling observer is employed in this method.

The Hotelling observer, reviewed in more detail in Appendix B, bases its discrimination rule on the first and second order statistics, namely, the mean, variance and covariance of the population data, rather than requiring full statistical knowledge about the population distribution. A Hotelling observer can be defined in terms of either sample statistics estimated from sample data, or ensemble statistics derived from the ensemble properties of the data. In this dissertation, a Hotelling observer is defined by the sample statistics.

In order to apply the Hotelling observer to detect candidate μCa++s, we need to define feature vectors. The feature vector is a set of measurements which is supposed to condense the description of relevant properties of the clusters into a small, Euclidean space. In this method, the feature vector $\mathbf{v}(x,y) = (v_1(x,y), \cdots, v_7(x,y))$ is defined as

$$
\begin{aligned}
v_1(x,y) &= D^1_{2^1}(x,y) + D^2_{2^1}(x,y) + D^3_{2^1}(x,y) \\
v_2(x,y) &= D^1_{2^{1.5}}(x,y) + D^2_{2^{1.5}}(x,y) + D^3_{2^{1.5}}(x,y) \\
v_3(x,y) &= D^1_{2^2}(x,y) + D^2_{2^2}(x,y) + D^3_{2^2}(x,y) \\
v_4(x,y) &= D^1_{2^{2.5}}(x,y) + D^2_{2^{2.5}}(x,y) + D^3_{2^{2.5}}(x,y) \\
v_5(x,y) &= D^1_{2^3}(x,y) + D^2_{2^3}(x,y) + D^3_{2^3}(x,y) \\
v_6(x,y) &= D^1_{2^{3.5}}(x,y) + D^2_{2^{3.5}}(x,y) + D^3_{2^{3.5}}(x,y) \\
v_7(x,y) &= A_{2^3}(x,y)
\end{aligned}
\tag{6.5}
$$

where $A_{2j}(x,y)$, $D_{2j}^1(x,y)$, $D_{2j}^2(x,y)$ and $D_{2j}^3(x,y)$ are wavelet coefficients of the jth octave at (x,y). At first, we get a training data set composed of equal numbers of signal-present vectors $\mathbf{v_1}$ and signal-absent vectors $\mathbf{v_2}$. Signal-present vectors are obtained from the center points where $\mu Ca++s$ are present, whereas signal-absent vectors arise from the background. For each class, 205 samples are randomly extracted from 20 mammograms. Figure 6.9 shows the probability density functions estimated from each component of $\mathbf{v_1}$ and $\mathbf{v_2}$.

The measurement values for a feature should be correlated with its class membership and cluster according to the class from which they are derived. In this respect, measurements v_1, v_2, v_3 and v_4 are good features which separate the different classes, while v_0, v_5 and v_6 are not a good choice of features.

From this training data, a Hotelling observer is computed in the following way. First, the mean vectors of signal-present vectors $\mathbf{v_1}$ and signal-absent vectors $\mathbf{v_2}$, $\bar{\mathbf{v}}_1$, $\bar{\mathbf{v}}_2$ are computed and the difference between two mean vectors $\Delta\mathbf{v}$ is obtained,

$$\Delta\mathbf{v} = \bar{\mathbf{v}}_1 - \bar{\mathbf{v}}_2. \tag{6.6}$$

Then, the scatter matrix $\mathbf{S}$, which corresponds to the estimated covariance matrix of $\mathbf{v_1}$ and $\mathbf{v_2}$ is computed as follows,

$$\mathbf{S} = \tfrac{1}{2} < (\mathbf{v_1} - \bar{\mathbf{v}}_1)(\mathbf{v_1} - \bar{\mathbf{v}}_1)^T > + \tfrac{1}{2} < (\mathbf{v_2} - \bar{\mathbf{v}}_2)(\mathbf{v_2} - \bar{\mathbf{v}}_2)^T > \tag{6.7}$$

where $< >$ represents an average operation. From (6.6) and (6.7), the Hotelling observer weights $(\Delta\mathbf{v})^T \cdot \mathbf{S}^{-1}$ are computed.

(a)

(b)

(c)

(d)

(e)

(f)

(g)
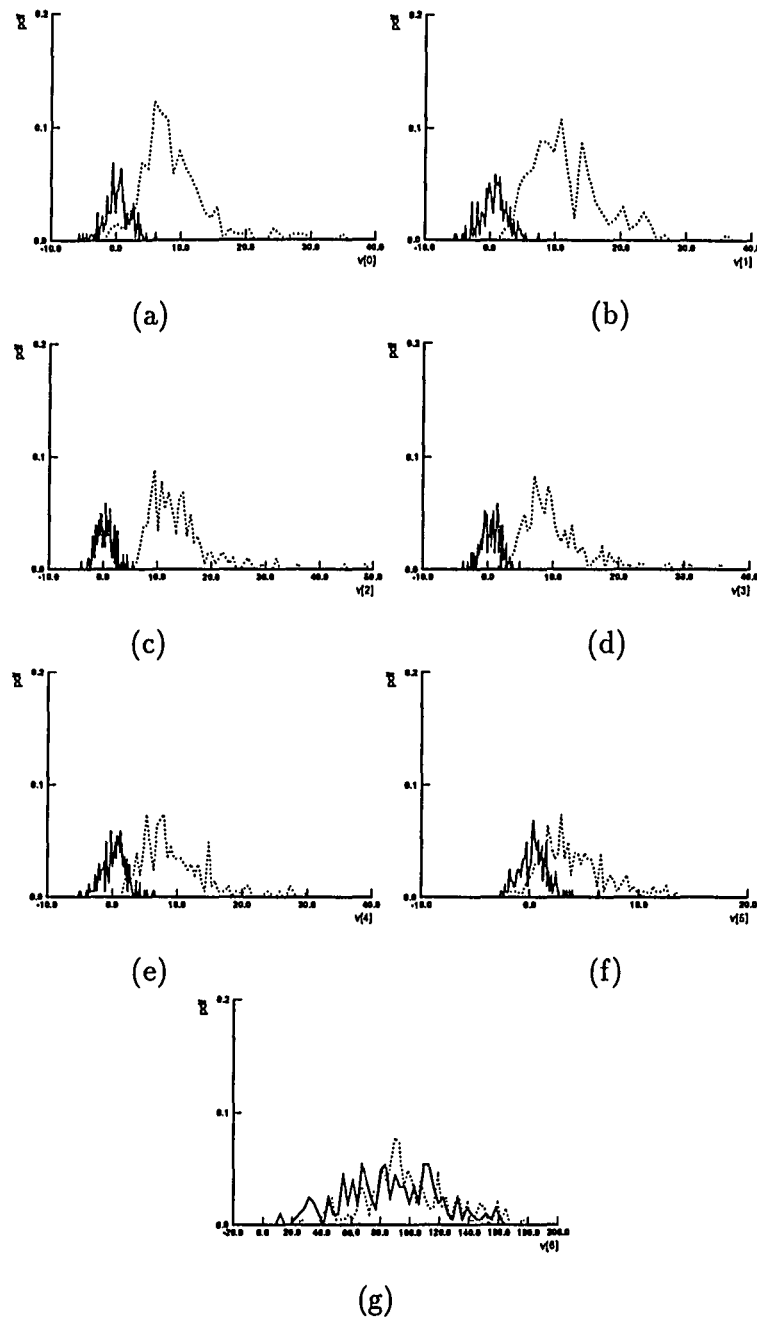
Figure 6.8: Probability density functions of each component of v_1 and v_2. The *solid* line and the *broken* line represent the pdfs of the components of signal-absent and signal-present vectors, respectively.

$$\overline{v}_1 = [\,9.761, \quad 12.897, \quad 14.067, \quad 10.250, \quad 9.474, \quad 4.091, \quad 98.571\,]^{\mathsf{T}}$$

$$\overline{v}_2 = [\,0.242, \quad 0.649, \quad 0.284, \quad 0.527, \quad 0.333, \quad 0.337, \quad 86.658\,]^{\mathsf{T}}$$

$$\Delta v = [\,9.519, \quad 12.248, \quad 13.783, \quad 9.723, \quad 9.141, \quad 3.754, \quad 11.913\,]^{\mathsf{T}}$$

$$S^{-1} = \begin{bmatrix}
0.219 & -0.158 & -0.199 & 0.245 & -0.010 & -0.008 & 0.001 \\
-0.158 & 0.186 & 0.019 & -0.205 & 0.125 & -0.057 & -0.000 \\
-0.199 & 0.019 & 0.629 & -0.439 & -0.266 & 0.264 & -0.002 \\
0.245 & -0.205 & -0.439 & 0.838 & -0.206 & 0.701 & -0.601 \\
-0.010 & 0.125 & -0.266 & -0.206 & 0.701 & -0.601 & 0.002 \\
-0.008 & -0.057 & 0.264 & -0.107 & -0.601 & 0.957 & -0.004 \\
0.001 & -0.000 & -0.002 & 0.002 & 0.002 & -0.004 & 0.001
\end{bmatrix}$$

$$\Delta v^{\mathsf{T}} S^{-1} = [\,-0.318, \quad -0.032, \quad 1.283, \quad -0.341, \quad -0.059, \quad -0.116, \quad 0.015\,]$$

Figure 6.9: $\overline{v}_1$, $\overline{v}_2$, Δv, S^{-1} and the Hotelling observer $(\Delta v)^{T} \cdot S^{-1}$.

$$\text{Output} = (\overline{\Delta \mathbf{v}})^T \mathbf{S}^{-1} \mathbf{v}$$
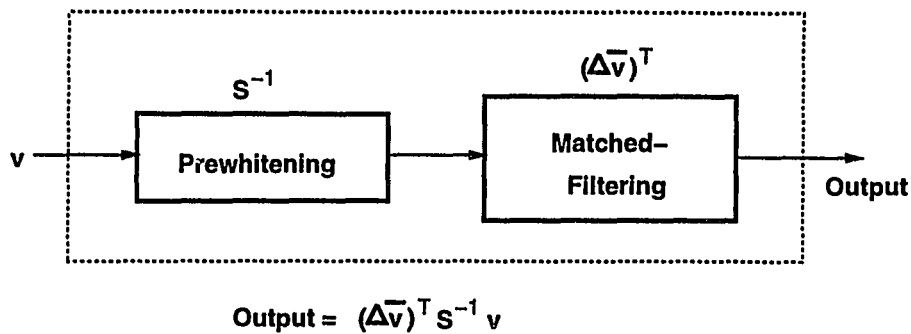
Figure 6.10: A block diagram of the Hotelling observer.

Figure 6.9 shows $\overline{\mathbf{v}_1}$, $\overline{\mathbf{v}_2}$, $\Delta \mathbf{v}$, $\mathbf{S}^{-1}$ and the Hotelling observer $(\Delta \mathbf{v})^T \cdot \mathbf{S}^{-1}$, obtained from the training data. Given the Hotelling observer, the Hotelling discriminant $\lambda(x, y)$ can be obtained as follows,

$$\lambda(x, y) = (\Delta \mathbf{v})^T \cdot \mathbf{S}^{-1} \cdot \mathbf{v}(x, y). \tag{6.8}$$

The inner product of $\mathbf{v}(x, y)$ with $\mathbf{S}^{-1}$ functions as a prewhitening operator, while that of the prewhitened $\mathbf{v}(x, y)$ with $\Delta \mathbf{v}$ is equal to the matched filter. This is similar to the prewhitening matched filtering operation. Figure 6.10 shows a block diagram of the Hotelling observer.

In order to see how the Hotelling observer works, it is applied to the training data. Figure 6.11(a) depicts the probability density functions of the Hotelling discriminant. The corresponding detectability d is 3.182.

For comparison, a nonlinear single layer perceptron [75], as shown in Figure 6.12(a) is trained using the same training data. The single node computes a weighted sum of the feature elements, where w_7 acts as a threshold, and passes the result through
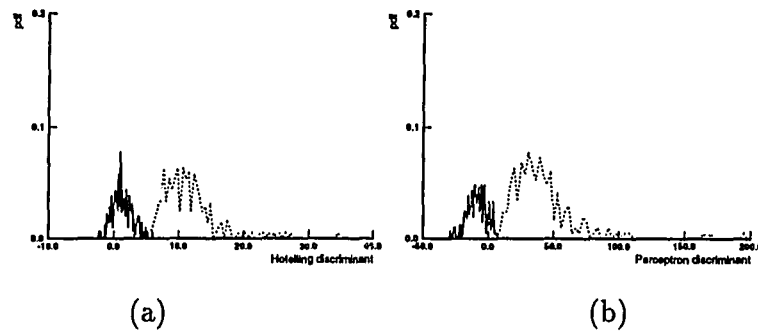
Figure 6.11: Probability density functions of the Hotelling discriminant (a) and the perceptron discriminant (b). The *solid* line and the *broken* line represent the pdfs of the components of signal-absent and signal-present vectors, respectively.

a sigmoid function so that the output y approach either 1 or 0, namely,

$$y = f(\sum_{i=0}^{6} w_i v_i + w_7),$$ (6.9)

where

$$f(x) = \frac{1}{1 + e^{-x}}.$$ (6.10)

The decision rule is to choose class A (signal-present) if the output y is close to 1 and

class B (signal-absent) if the output is close to 0. The perceptron forms two decision

regions separated by a hyperplane. These regions are shown in Figure 6.12(b) when

there are only two inputs and the hyperplane is a line. In this case, inputs above

the boundary line lead to class A responses and inputs below the line lead to class

B responses. As can be seen, the equation of the boundary line depends on the

connection weights and the threshold.

The connection weights after convergence are

$$\mathbf{w} = [-0.520, 0.665, 3.390, 2.194, -1.767, -0.839, -0.118, -7.768].$$ (6.11)
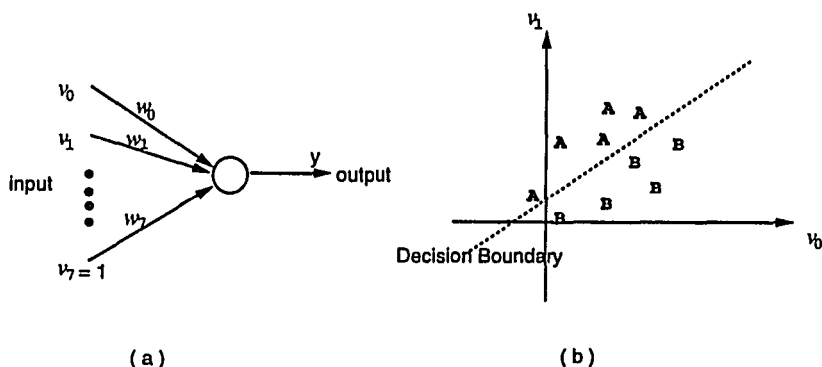
Figure 6.12: A single layer nonlinear perceptron that classifies an input vector into two classes denoted A and B. This perceptron divides the space spanned by the input into two regions separated by a hyperplane or a line in two dimensions.

The probability density functions of the perceptron outputs are depicted in Figure 6.11(b), and the detectability d is 2.483. As expected, the Hotelling observer is better in the sense of detectability although its probability density functions partially overlap.

Figure 6.13 shows the Hotelling discriminant $\lambda(x, y)$ when Figure 6.2 is applied, and the binary map, which is obtained by thresholding $\lambda(x, y)$, is depicted in Figure 6.14. Figure 6.14 shows more true positives and less false positives caused by texture edges, compared to Figure 6.7.

6.1.2 Classifying candidate μCa++s

In practice, the multiscale matched filter detects false alarms created by edges and other linear structures, as well as μCa++s, as shown in Figure 6.7 or Figure 6.14. Many false alarms can be eliminated by shape analysis. Figure 6.15 shows
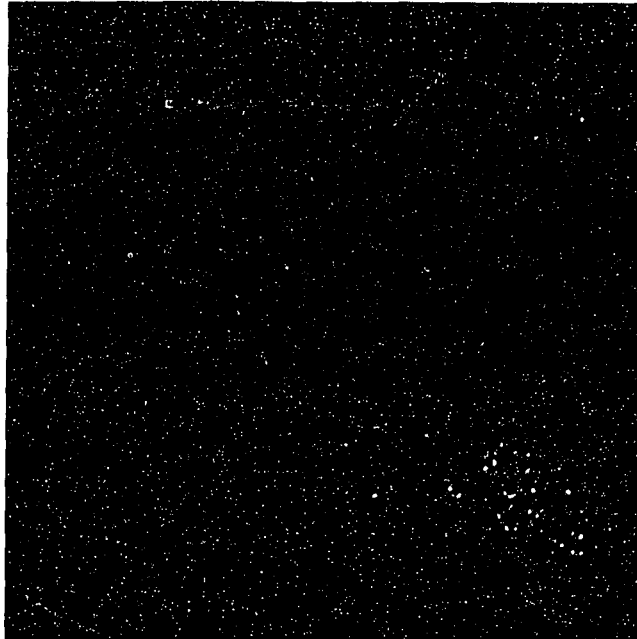
Figure 6.13: This image shows the output when the Hotelling observer is applied to the input vectors $\mathbf{v}(x,y)$.

a procedure for classifying the candidate μCa++s. To reduce the false alarms, we first detect the boundaries of the detected pixel regions. Marr and Hildreth [62] showed that one can obtain the positions of multiscale sharp variation points from the zero-crossings of the signal convolved with the Laplacian of Gaussian.

The first step in zero-crossing edge detection is to smooth an image to remove discontinuities and small fluctuations. Then the second derivative of this smooth function is computed. It will have a zero crossing at the location of the edge. The zero-crossings of a wavelet transform indicate the location of sharp variation points which corresponds to the edges in two dimensions, if the wavelet is the second derivative of a smoothing function, typically a Laplacian of Gaussian [24]. Hence, the boundaries are
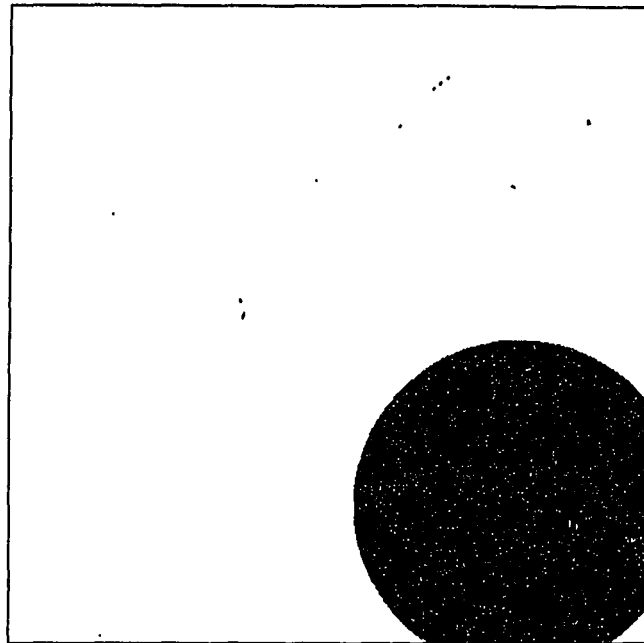
Figure 6.14: This binary map image shows candidate μCa++s detected by Method #2.
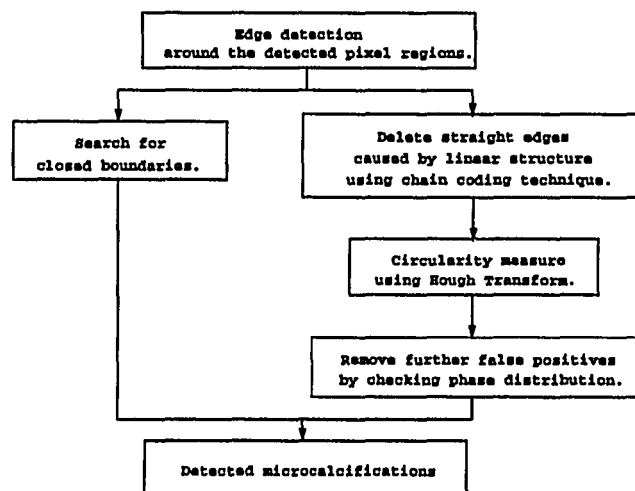
Figure 6.15: A procedure for classifying candidate μCa++s.
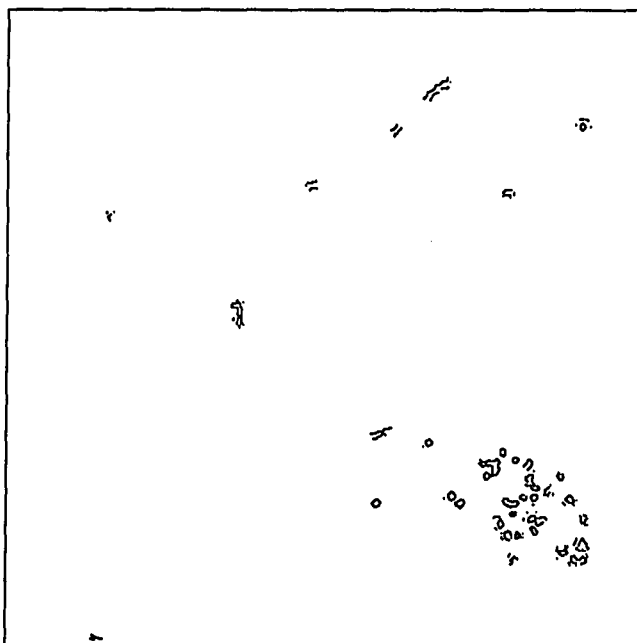
Figure 6.16: Boundaries around the detected pixels.

obtained easily from the zero-crossings locations in the LH + HL sub-bands, which

corresponds to the result of convolution with 2D Laplacian of Gaussian functions.

A pixel is considered a zero-crossing if the pixel value of the LH+HL sub-band is

greater than or equal to zero, and it has at least one 8-connected neighbor with a

value less than zero. The zero-crossing points are detected in LH + HL at each

octave. These zero-crossings, however, may represent spurious edges as well as true

edges. These false edges can be deleted by computing an approximation of the

gradient magnitude at each zero-crossing. In other words, true edges will be more

likely to have large gradient magnitudes, while the gradient magnitude of false edges

is small. To remove false edges, the computed gradient magnitudes are thresholded
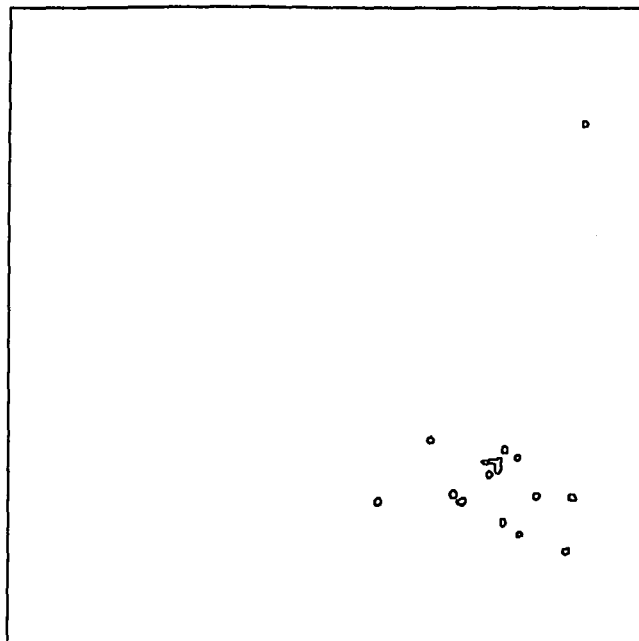
Figure 6.17: The detected regions whose boundaries are closed are classified as true positives.

at some fixed histogram percentile. Figure 6.16 shows the zero-crossing points around the potential μCa++s given in Figure 6.14. μCa++s are usually circular and their boundaries are rounded.

To discriminate between μCa++s and false alarms, we first search for a precise closed boundary of a possible μCa++ at each candidate μCa++ site. When we find such a boundary, we refer to the resultant object as a true positive. Figure 6.17 represents the closed edges detected from the above edge map using a chain coding technique [64]. The remaining edge map without closed edges is given in Figure 6.18.

The straight edges shown in the upper right and lower left of Figure 6.18 should be removed because they are caused by the linear structure of normal breast tissue.

Figure 6.18: The remaining edges without closed boundaries.

At each boundary, both end points are searched. The boundary is regarded as a straight edge and removed if the distance between two end points is long compared to its length. Figure 6.19 represents the edge map in which the straight edges are removed using the chain coding technique.

Then, a circularity measure for each boundary is computed and those detected regions falling below a threshold (determined by training using the database) are discarded. Given the position vector $a = (a_1, a_2)$ and a radius r, the circularity of the boundary for the triplet (a_1, a_2, r) is defined as the number of edge pixels on a circumference at radius r about a. This technique is called the Hough transform [64]. Triplets (a_1, a_2, r) with high circularity values are selected as true positives if

Figure 6.19: The edges caused by long, linear structure are removed.

the radius r is in the range of 1 to 5 pixels. The output image after applying the Hough transform is given in Figure 6.20.

To remove more false positives, a phase distribution feature is calculated from the regions detected by the Hough transform. Figure 6.21 shows the region of support of phase distribution. Phase distribution $p(i,j)$ [67] is depicted as

$$p(i,j) = \frac{\sum_{(k,l)\in R} cos\theta(k,l)}{N},$$
(6.12)

where (i,j) is the location of the center of the region detected by the Hough transform, and N is the number of pixels in the region of support R. The variable $\theta(k,l)$ is defined as the angle between a gradient vector at (k,l) and the straight line connecting pixels at (i,j) and (k,l). The term $cos\theta(k,l)$ is a measure of convergence of the gradient
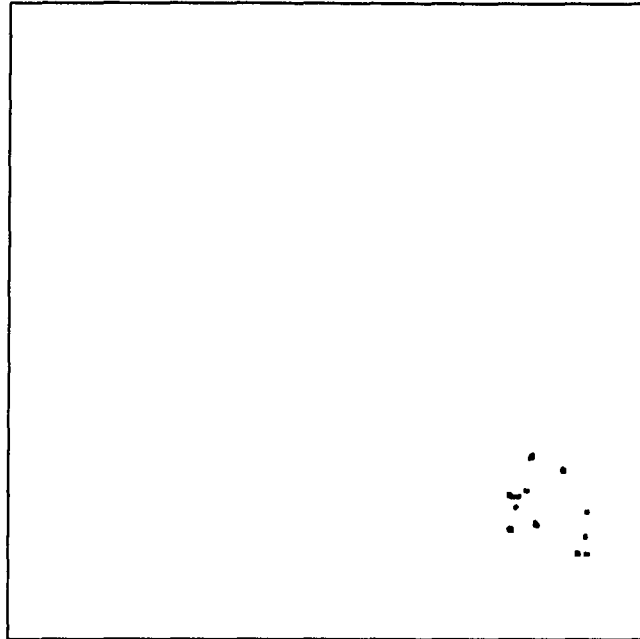
Figure 6.20: A circularity measure is applied. The white regions are detected μCa++s.
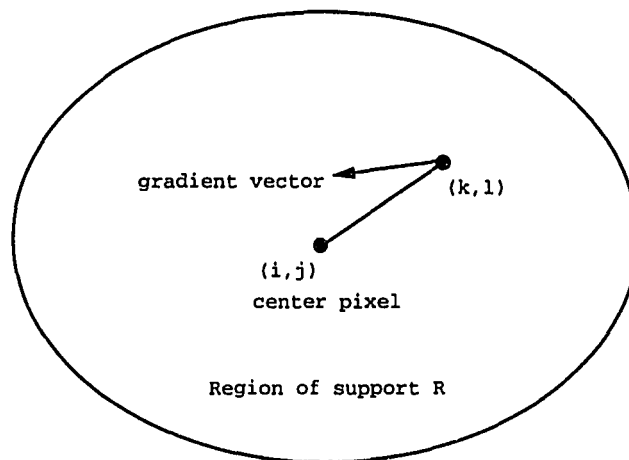
Figure 6.21: The region of support which corresponds to the detected pixel region.

(a) Circular shape.　　　　　(b) Line edge.
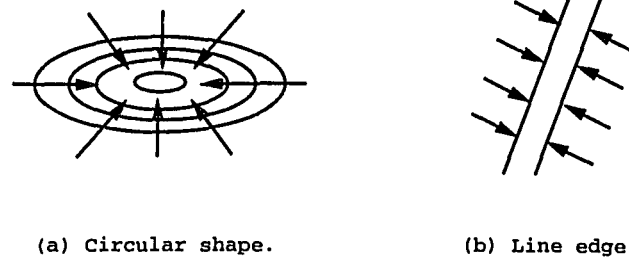
Figure 6.22: Phase distribution patterns of gradient vectors.

vector upon the pixel of interest and $p(i,j)$ is the average convergence of gradient vectors in the region of support. It is bounded between -1 and 1. If it is close to 1.0, it means that almost all gradient vectors in the region of support point to the same point. This can happen near the center of μCa++s, as shown in Figure 6.22(a). On the other hand, the gradient vectors do not converge to a particular point if a local high density area is long and slender as shown in Figure 6.22(b), and $p(i,j)$ for such shapes can not be large. The feature $p(i,j)$ is sensitive not to long and slender shapes but to rounded local high density regions. Among the regions detected by the Hough transform the regions whose $p(i,j)$ are below some threshold are picked out, as shown in Figure 6.23.

The final detected μCa++s, shown in Figure 6.24, are obtained by adding the closed edges shown in Figure 6.17 to the regions detected by the Hough transform and phase distribution shown in Figure 6.23. By comparing Figure 6.24 with Figure 6.14, we see that many false positives are removed after applying the circularity measure technique.
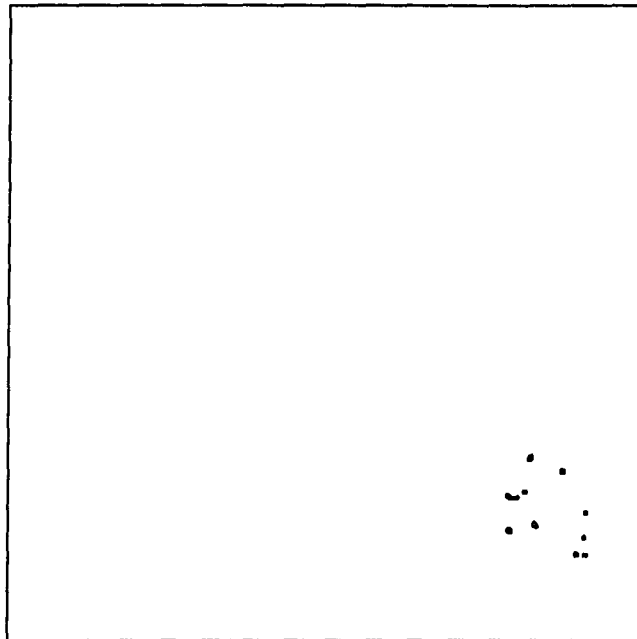
Figure 6.23: Among the regions detected by the Hough transform the regions whose $p(i,j)$ are below some threshold are eliminated.

To see how well the Hough transform works, it is applied to a subimage containing elongated μCa++s as well as circular μCa++s. Figure 6.25 shows the results at each step. As expected, circular μCa++s can easily be detected by the Hough transform because it is tuned to detect circular shapes. The Hough transform is related to template matching and is used to match the simple analytic form or known shape. Although, as stated earlier, the average μCa++ is assumed to be circular, individual μCa++s may be elongated or curved. So, the Hough transform is inappropriate for detecting μCa++s of arbitrary shape. However, if we assume that elongated or curved shapes can be modeled as a chain of circles, the Hough transform can be applied by modifying the algorithm. In other words, it is regarded as detected, if the
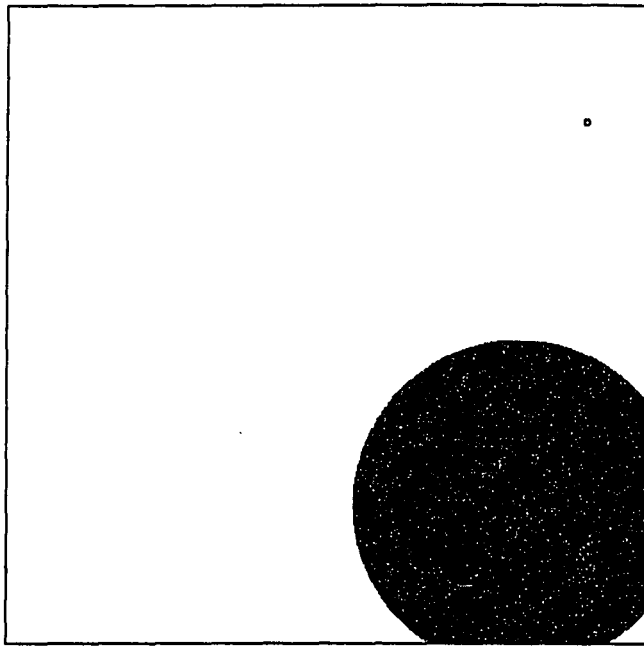
Figure 6.24: The final detected μCa++s obtained by adding the closed edges shown in Figure 6.17 to the regions detected by Hough transform and phase distribution shown in Figure 6.23. The outlined regions correspond to detected sites with closed boundaries. The filled-in regions originally had open boundaries but later passed a shape test.
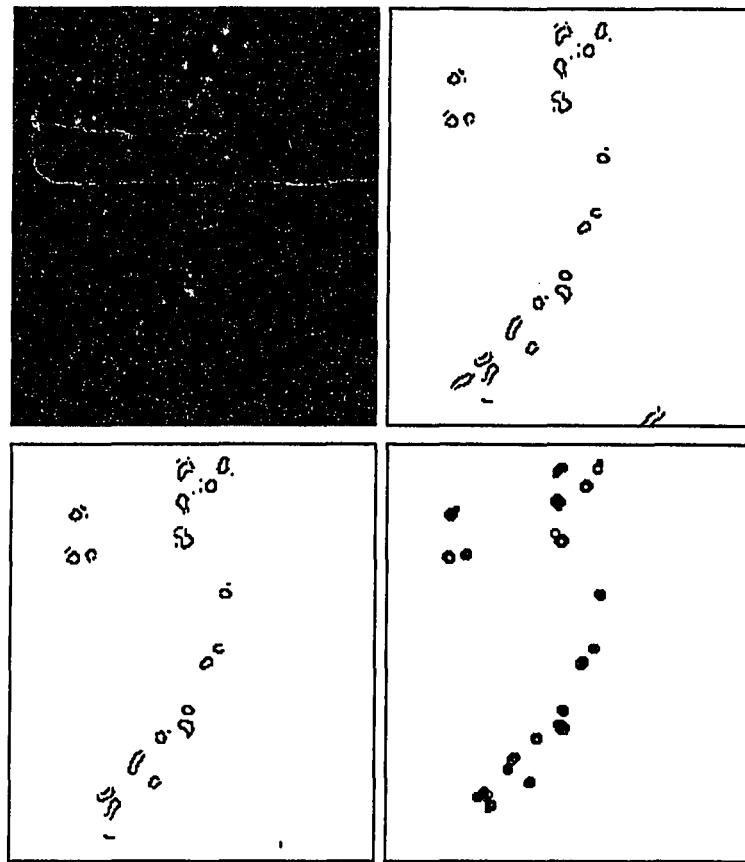
Figure 6.25: The output images at each stage. *Upper left* LH + HL image of octave 2, *Upper right* Zero crossings around the detected pixels, *Lower left* Zero crossings with edges caused by linear structure removed. *Lower right* Hough transform output.

sum of circularity measures of adjacent points is far above a threshold although each of them falls below the threshold. An elongated shape is detected as a superposition of several circles, as shown in the lower right of Figure 6.25.
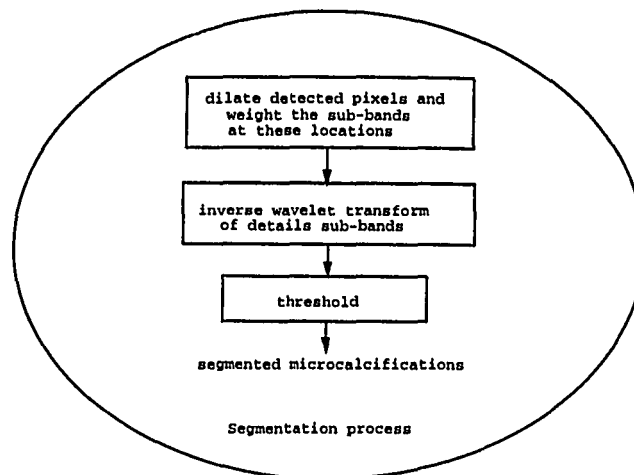
Figure 6.26: Segmentation process.

6.2 Segmentation process

If segmentation of μCa++s is desired, and since matched filter detectors do not preserve μCa++ shape, the following step is required. First, each details sub-band is nonlinearly mapped to achieve multiscale contrast enhancement. The nonlinear mapping function is given as

$$f(x) = a[max \cdot sigm(b(x - c)/max) + min \cdot sigm(b(x + d)/min)] \qquad (6.13)$$

where $sigm(x)$ is defined by

$$sigm(x) = \frac{1}{1 + e^{-x}}, \qquad (6.14)$$

a controls the gain of details images, max and min are the maximum and minimum values of each details sub-band, and c, d and b control the threshold and rate of enhancement, respectively [76].

The wavelet coefficients corresponding to dilated circular regions centered at each detected pixel site are further weighted (typically by 7); then a straightforward inverse wavelet transform of the LL image from octave 4 and the detection-enhanced details subimages from octaves 1 through 4 reconstructs the original mammogram with detected μCa++s visible. A procedure for segmenting the detected μCa++s is indicated in Figure 6.26. Figure 6.27 shows the dilated version of the final map of detected pixels, the reconstructed mammogram with detected μCa++s highlighted and the segmented μCa++s when the mammogram given in Figure 2.4 is applied. This form of output may be useful where further viewing by the radiologist is required, since the context of the detected pixels within the breast will probably be required for diagnosis purposes. Alternatively, the smooth subimage may be omitted in computing the inverse transform, leaving an image containing only suspected μCa++s. The diameter of the circular weighting region increases with octave, and hence object scale.

Experimental results of this nonlinear algorithm are described in the next chapter. Performance resulting from tests on a set of 40 digital mammograms is presented in the form of a free-response receiver operating characteristic (FROC) curve, and examples of reconstructed calcifications are shown.
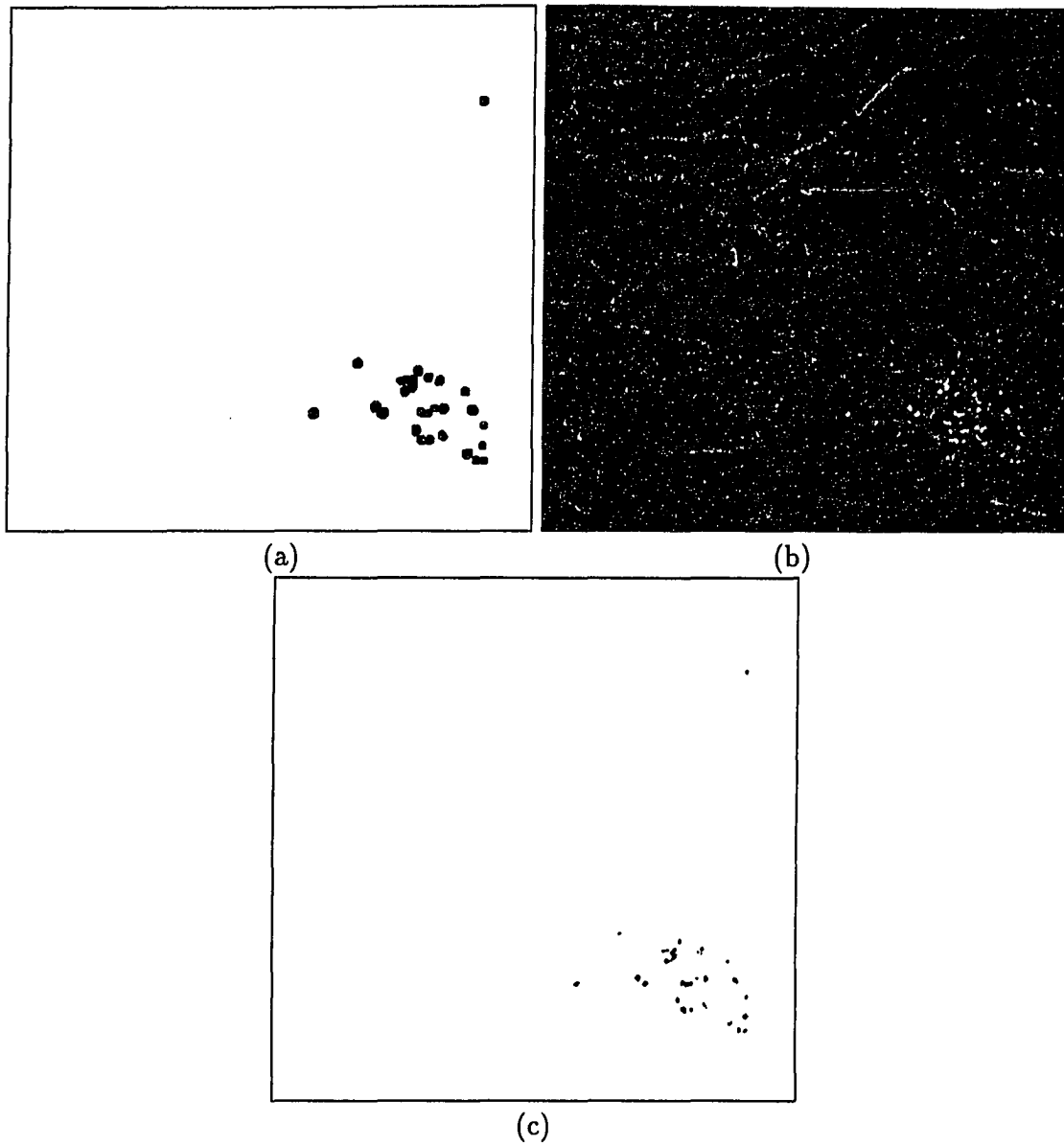
Figure 6.27: (a) Dilated version of the final map image. (b) Reconstructed mammogram with detected μCa++s highlighted. (c) Segmented microcalcifications. (Original image is shown in Figure 6.2.)

CHAPTER 7

Experimental Results

We have tested our detection algorithms on the Nijmegen database of 40 mammograms[10]. Each mammogram contains at least one μCa++s cluster verified by radiologists. Since individual μCa++s are not usually a warning sign in breast cancer, only clusters of μCa++s need to be detected. Although the mammograms were digitized to arrays of 2048 × 2048 pixels, they are reduced to 1024 × 1024 to reduce processing time. This is done by manually cutting out parts of the image that contains at least one μCa++s cluster. Regions of film outside the breast tissue were identified by hand and not processed. The free-response receiver operating characteristic (FROC) curves in Figure 7.1 were generated by counting true positive (TP) and false positive (FP) clusters for each mammogram while varying the threshold applied to the sub-band images. The FROC curve is a graph of the percentage of TP clusters found versus the average number of FP clusters per image detected. So, it is sufficient only to glance at the FROC curve to see what percentage of TP are detected and how many FP clusters there are, if one is viewing an output image and wants to know how accurate it is.

The program code for the FROC testing is obtained from Karssemeijer. It reads in a binary image that contains the location and shape of the clusters. Then each of
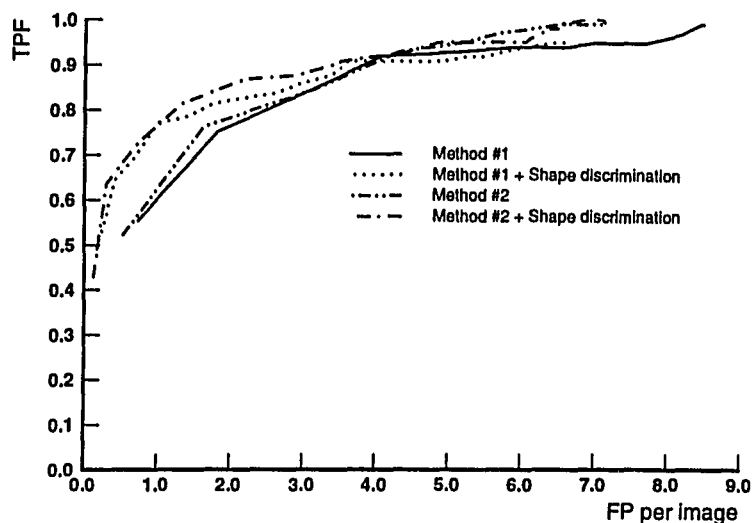
Figure 7.1: Cluster detection performance measured on 40 mammograms.

the μCa++s is thinned to one pixel, then it searches for the circular region known to contain a cluster and counts the number of pixels found in the region. If there are two μCa++s or more in a region, a TP is returned for that cluster. The process is repeated for each cluster. Any μCa++ clusters found outside the true positive regions are FPs.

Figure 7.1 shows that the performance of the method #2 which uses a Hotelling observer is slightly better than that of method #1. The improved FROC curves are obtained by using object shape to reduce false positives caused by linear structure, for both methods. The best performance can be obtained when the method #2 is used to detect candidate μCa++s, followed by shape analysis.
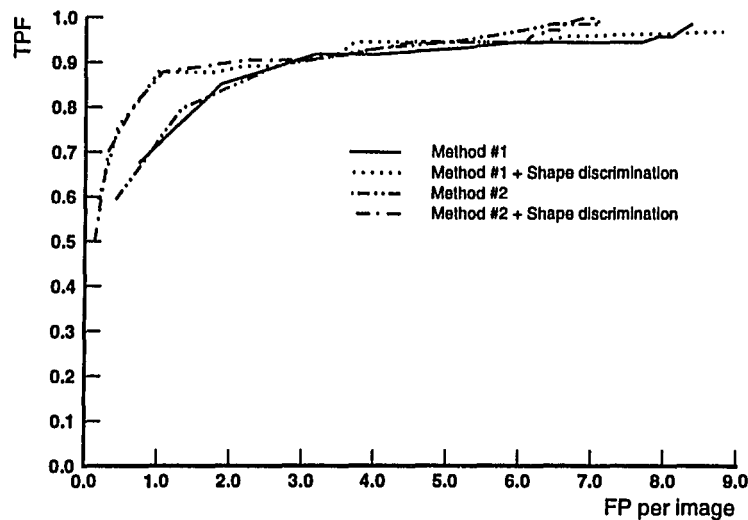
Figure 7.2: Cluster detection performance measured with 12c, 12o omitted.

As stated earlier, each image in the database contains an average of two clusters. However, one pair of images, numbered 12o (oblique) and 12c (craniocaudal), contains a total of 28 clusters, which is approximately 27% of the total. Furthermore, many of these clusters are closely grouped or partly overlapping, and many μCa++s fall outside the truth circles marked by the radiologists. These μCa++s are often detected, resulting in a higher false positive rate, and an overall drop in FROC performance. To demonstrate the impact of the images numbered 12c and 12o, Figure 7.2 also shows an FROC curve computed when these images were omitted from the test. Examples showing μCa++s reconstructed by the second, segmentation step are given in Figure 7.3.

Notice that, although the matched filter is predicated on circular μCa++s, the segmentation process is able to reconstruct μCa++s of arbitrary shape. This can be explained as follows. The matched filter, although designed for detecting objects of Gaussian shape, nevertheless produces a significant peak response to objects of similar shape and of the same size as the Gaussian. However, the shape of the matched filter response is not an estimate of the object; it is the location and height of the peak value that is significant. Hence, when thresholded, the matched filter output is typically smaller than the object to be detected. By dilating each detected pixel, we effectively "cast a net" encircling that location to find the object responsible for the output peak at that point. This process of weighting circular regions centered on detected pixels is analogous to filtering in the Fourier domain, except that the wavelet transform occupies a domain whose coordinates are space and scale. One example shown here (image 7c) contains elongated calcifications. That these are recovered by our approach can be explained by noting that fine, linear structures may in some cases be modeled by superposing Gaussian functions of appropriate scale along a line or curve. To justify the inverse wavelet transform stage we present a close-up example of segmented μCa++s in Figure 7.4. The calcification boundaries generated by the second, segmentation step are visibly more accurate than those present in the detected pixel map created by the first stage.

(a)

(b)

(c)

(d)
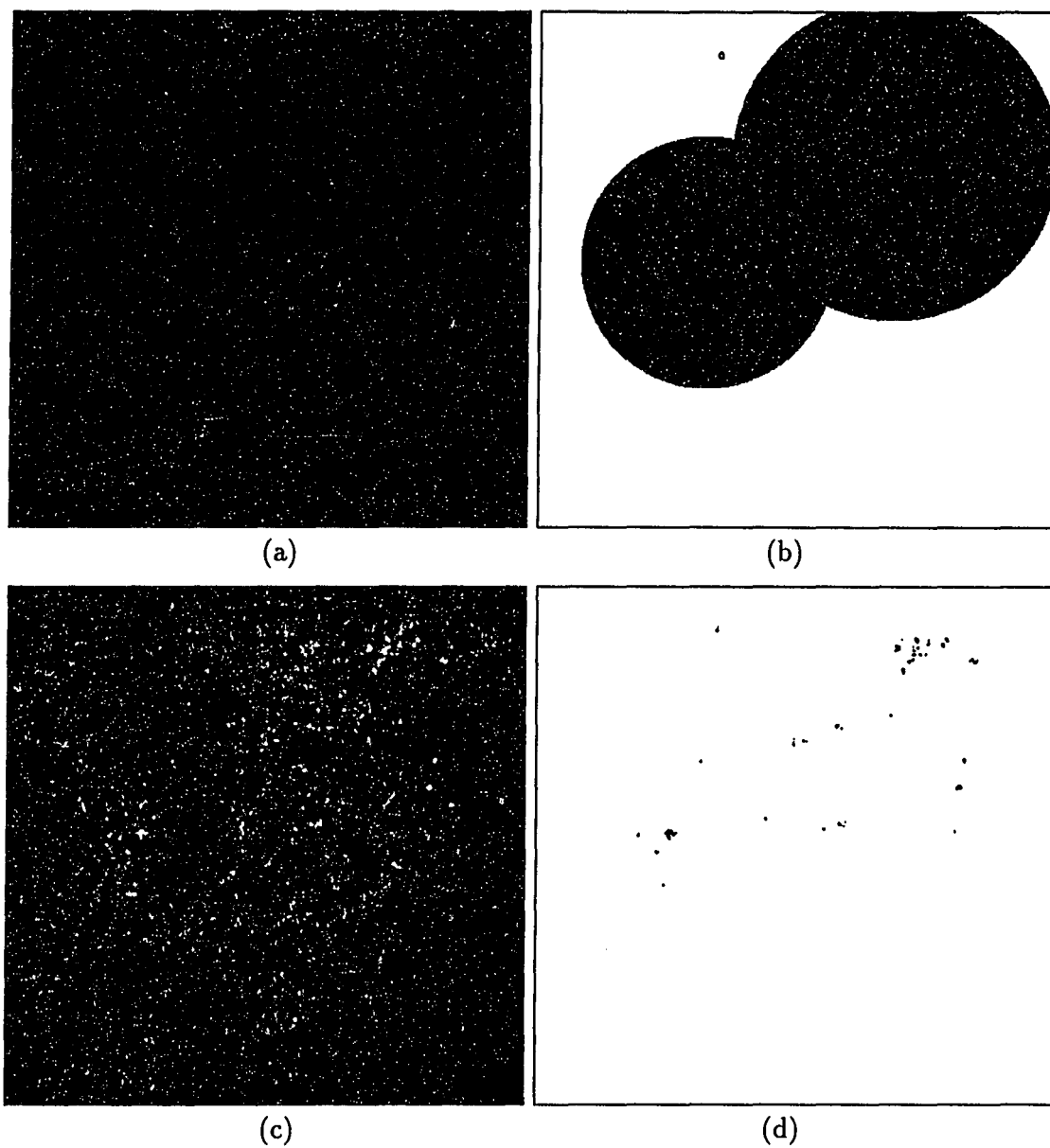
04c

Figure 7.3: Four examples of microcalcification clusters reconstructed using the complete algorithm shown in Figure 6.1. (a) Original mammogram. (b) Hough transformed output. (c) Reconstructed mammogram with detected microcalcifications highlighted. (d) Segmented microcalcifications.

(a)
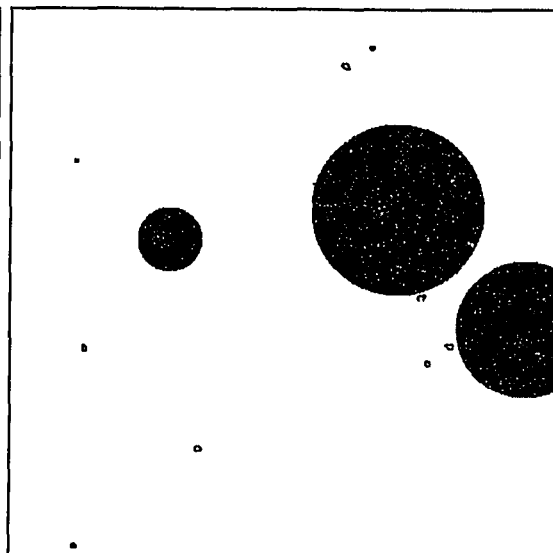
(b)
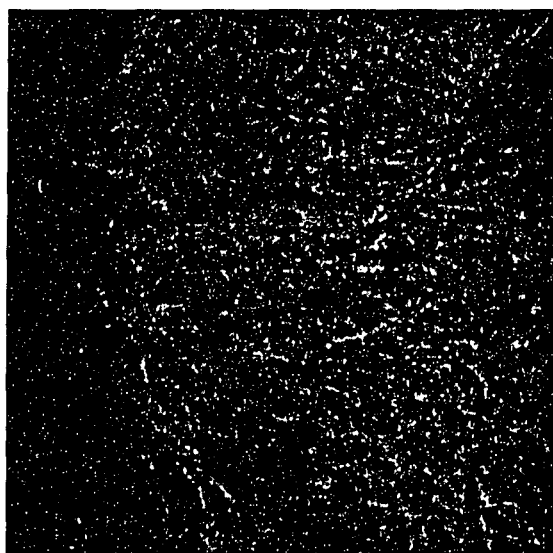
(c)

(d)

06c

(a)

(b)

(c)

(d)

16c

(a)

(b)

(c)

(d)

210

(a)   (b)

(c)
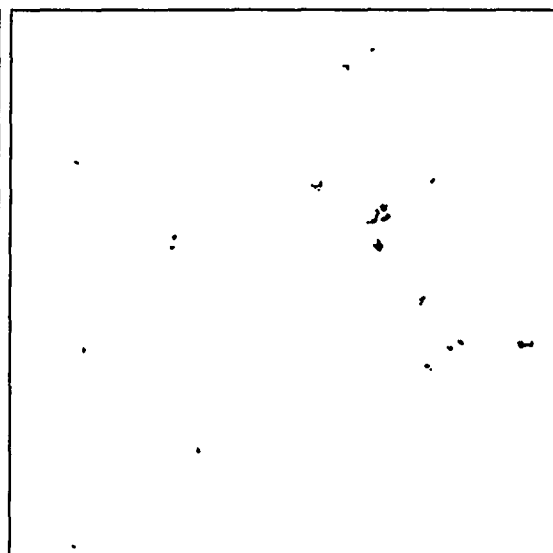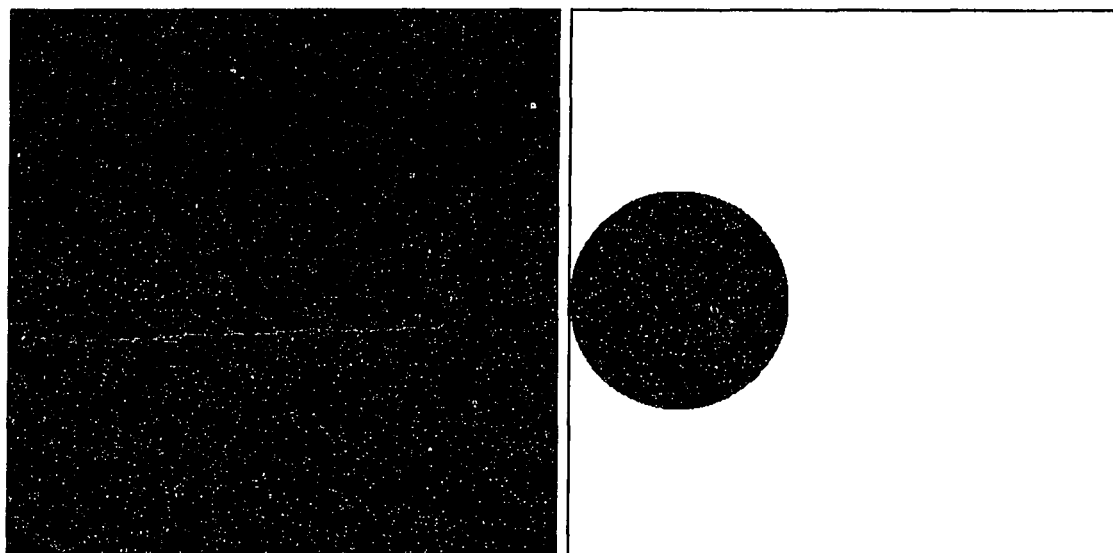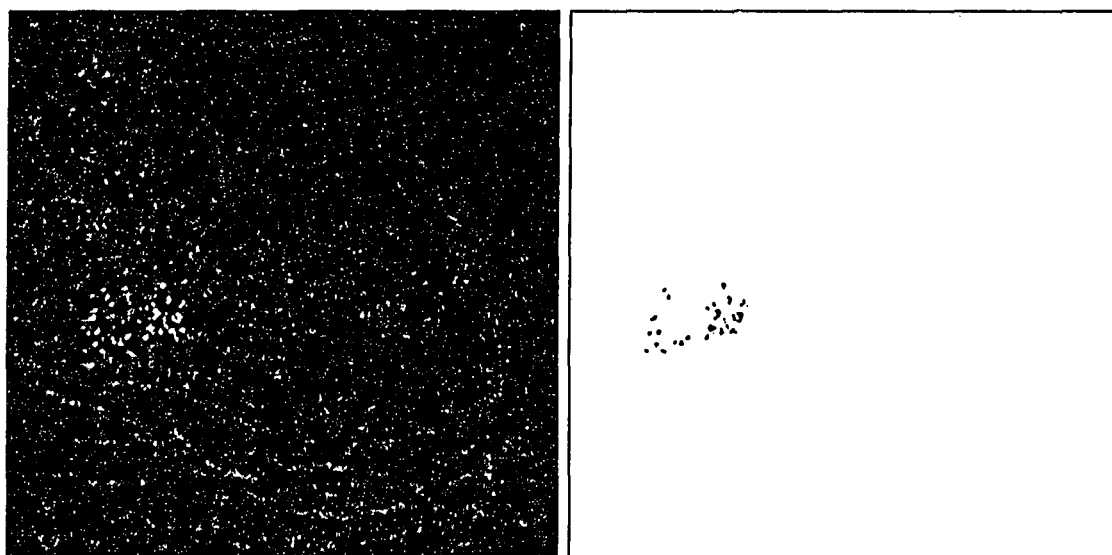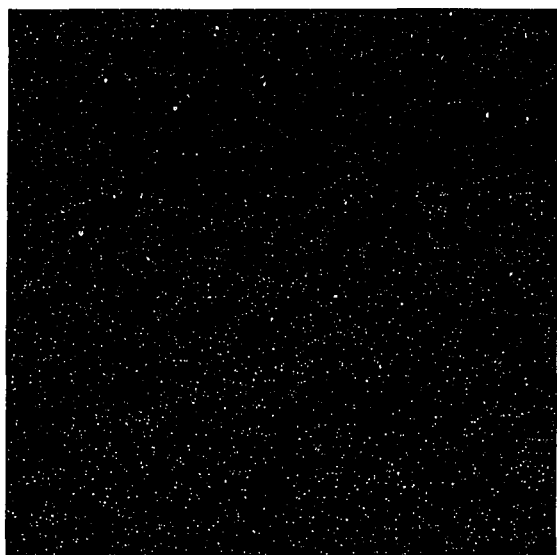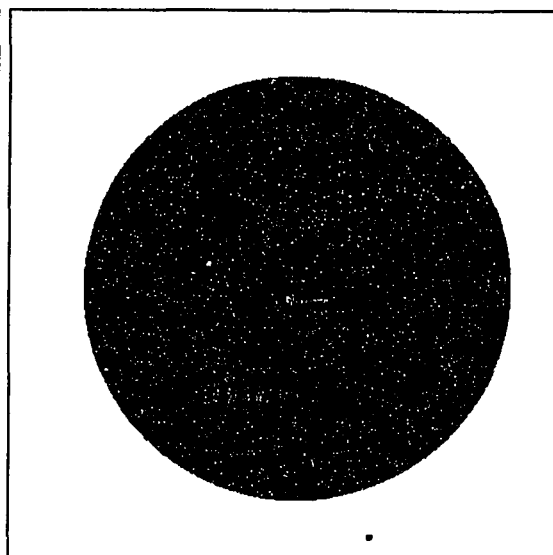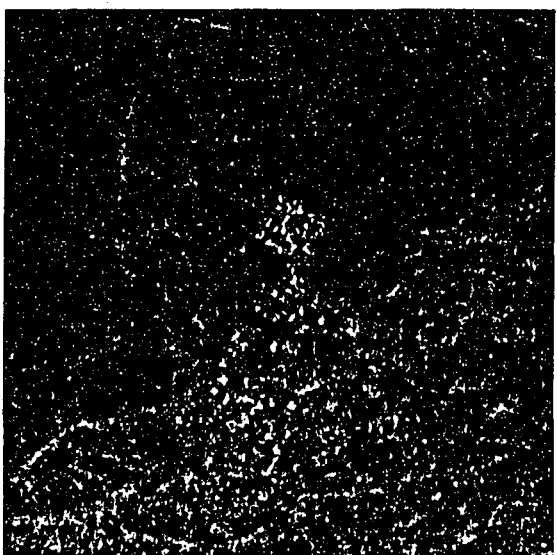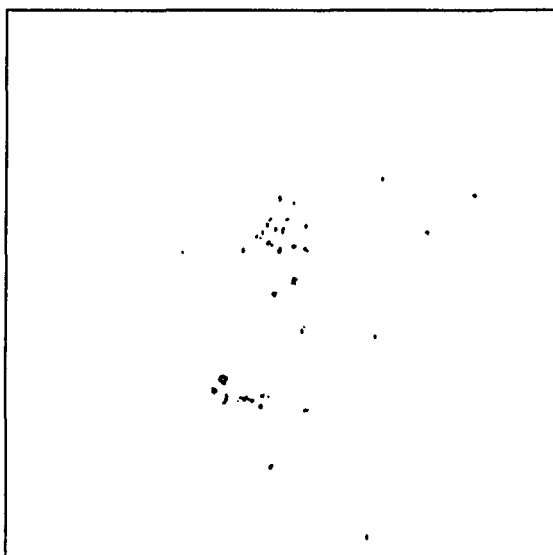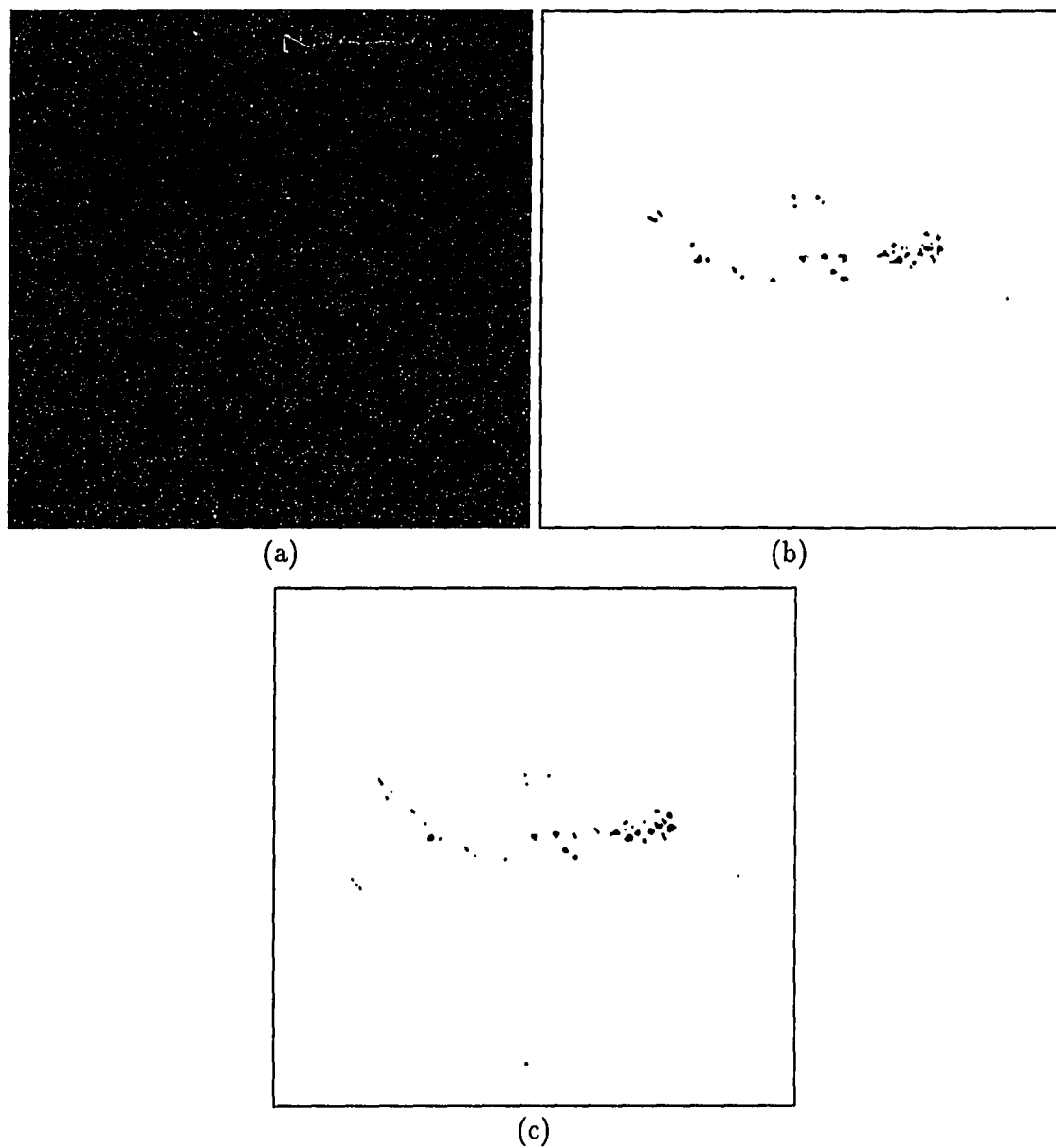
Figure 7.4: Close-up view of a cluster showing that microcalcifications are more accurately segmented when the second stage processing is implemented (weight/inverse wavelet transform/threshold) (b), compared to the detected pixel map (c) generated by the first stage. The original mammogram section is in (a).

CHAPTER 8

Summary

We have developed wavelet-based algorithms for detecting μCa++s of varying size in mammograms. Our work has been reported in [31], [68] - [72]. With an appropriate choice of wavelet basis the undecimated wavelet transform with inter-octave voicing can be a useful tool for detecting μCa++s in digitized mammograms. The method uses multiscale matched filters designed for detecting Gaussian objects in correlated Markov noise. We assume that individual μCa++s possess similar rounded profiles. We further simplify their profiles as a Gaussian signal.

Under the assumption of a known object in stationary background noise, an optimum detector can be derived, namely, the pre-whitening matched filter(PWMF)[13]. Even when noise is nonstationary, a spatially-adaptive filter may be used [28]-[31] to create a significantly decorrelated and more homogeneous background. In these circumstances we have shown that the PWMF is useful, although non-optimum, when applied to pre-processed gamma-ray images. The shape of the PWMF for a simple Gaussian object in stationary noise is similar to that of the second derivative of Gaussian with its characteristic positive center lobe surrounded by a smaller negative lobe, as shown in section 3.3. Thus, we expect that wavelets of form akin to the second

derivative of a Gaussian approximate multi-scale, pre-whitening matched filters as they compute the details sub-bands of the various octaves.

Our recognition algorithm is composed of two major process: detection and segmentation. The purpose of the detection process is to find the candidate μCa++ locations within the breast region, and then classify the potential μCa++s into true and false positives. Two methods have been proposed to find the potential μCa++s within the breast region. One method uses the details sub-bands HH and LH + HL from each octave, which are thresholded at some fixed percentile of the histogram of each component. The detected binary images from all octaves are then logically ORed to yield the binary map of detected pixels. The other method employs a Hotelling observer. The input vectors needed to apply the Hotelling observer to detect candidate μCa++s are defined using the wavelet coefficients from each octave. Since the procedure to compute the Hotelling discriminant is similar to the prewhitening matched filtering, this method can be modeled as concatenating two matched filters. Better performance is obtained from this method. The detected potential μCa++s are classified to reduce the false alarms by analyzing their shape.

The classical matched filter used in communications receivers makes no attempt to recover the shape of the incoming signal. Likewise, while the first (detection) stage of our method detects the presence of μCa++s, it does not accurately segment them. However, by weighting the details sub-bands at the sites of detected pixels followed by computing the inverse wavelet transform we achieve a reasonable segmentation

of μCa++s outlines. The effectiveness of the technique is constrained by certain approximations and assumptions. For example, adopting the Gaussian object model prevents us from optimally detecting anything but an average μCa++s at any given scale. The detection performance of the method has been tested by computing FROC curves using the Nijmegen database of 40 mammograms. Several qualitative examples of segmentated clusters are presented as evidence that μCa++s of varied shape can be recovered.

The method requires computing one 4 octave wavelet transform, one 3-octave wavelet transform (for the inter-octave voices), and one inverse wavelet transform if accurate μCa++s segmentation (as opposed to just detection) is desired. Efficient wavelet transform techniques are employed throughout. In practice we notice little drop in detection results when we omit the contributions from HH and LH+HL in octaves 1 and 4. Hence octave 4 need not be computed, whereas octave 1 is required in the sub-band decomposition. Octaves 2, 2.5, 3 and 3.5 are required to obtain best performance at the working resolution of the Nijmegen database. Further work is needed to ascertain the optimum sampling of this scale range.

The weighting factor applied during the reconstruction stage influences the contrast of the segmented μCa++s relative to the background. We currently use a factor of 7. The smaller the value used, the more noticeable the background breast texture appears, and *vice versa*. As stated earlier, we usually omit the LL component in the

final reconstruction, unless the context of the large scale breast structure is deemed to be valuable in a given situation, such as for viewing by a radiologist.

The proposed method is founded on the principles of optimum detectors (matched filters); furthermore, the analysis/synthesis paradigm of the wavelet transform is elegant, simple to implement. In addition, the wavelet basis used here, being close to the Laplacian of Gaussian, provides useful zero-crossing information in the details LH + HL sub-bands which could be applied to the suppression of false edges and lines. Many false positives are removed by analyzing the shape of the detected pixels through the edges.

The basic matched filtering / probability summation method attains a TPF of 55% at a false positive rate of 0.7 clusters per image. Introducing shape discrimination improves the TPF to 70% at the same false positive rate as above. The Hotelling observer method has slightly better performance, achieving a detection rate of 73%. When the images 12c and 12o are omitted, even better performance of 81% is obtained.

8.1 Future Directions

This method affords the potential for future enhancements of the basic approach. For example, the sub-bands could also be scanned for, and weighting applied to, clusters of detected pixels, prior to the inverse wavelet transform step. (The current method applies equal weight to all detected pixels, whether isolated or clustered.)

We will be able to improve our method further by: (1) using a neighborhood interaction mechanism which disallows detected pixels from regions containing both calcifications and linear structure, and (2) rewarding detected pixels which occur in local clusters. Action (1) will reduce the number of false positives due to long strands of connective tissue, and action (2) will tend to eliminate many isolated false detections.

As well, future work will address the potential of using multiple wavelet bases for computing multiple forward wavelet transforms. The idea here is that different wavelets may target different object classes. For example, a given wavelet may be more appropriate for enhancing the background texture in mammograms. By combining sub-bands from the multiple transforms it is conceivable that background structures could be favorably suppressed. Significantly, only one set of sub-bands needs to be involved in the reconstruction (inverse wavelet transform) stage.

Appendix A

Receiver Operating Characteristic Curve

This appendix provides the details of the ROC curves used for the performance evaluation of the wavelet transform and the matched filters in Chapter 5.

The most elementary situation in the detection theory involves discrimination between two hypotheses, H and K, based on a single observation or realization x of a random variable X; that is,

$$H : \text{ hypothesis on the observation } x, \text{ e.g. } X = N$$
$$K : \text{ alternative hypothesis on } x, \text{ e.g. } X = N + s. \tag{A.1}$$

Here we assume that N is a noise random variable, s is a signal to be detected and X is the observation random variable. Given x, a realization of X, we need to decide whether x belongs to H or K, namely, "signal absent" or "signal present". In order to accomplish this task, we need a decision rule $d = d(x)$. The decision rule $d = d(x)$ partitions the X space $\Omega = \{x \in X\}$ into two disjoint regions Ω_H and Ω_K, where $\Omega_H \cup \Omega_K = \Omega$ and $\Omega_H \cap \Omega_K = \phi$.

Then we can commit such errors as deciding for K when H is true. These errors are often called *false alarms*. The probability of false alarms can be written as

$$\alpha = P_{FA} = P(d_K/H) = \int_{\Omega_K} p_H(x)dx = 1 - \int_{\Omega_H} p_H(x)dx. \tag{A.2}$$
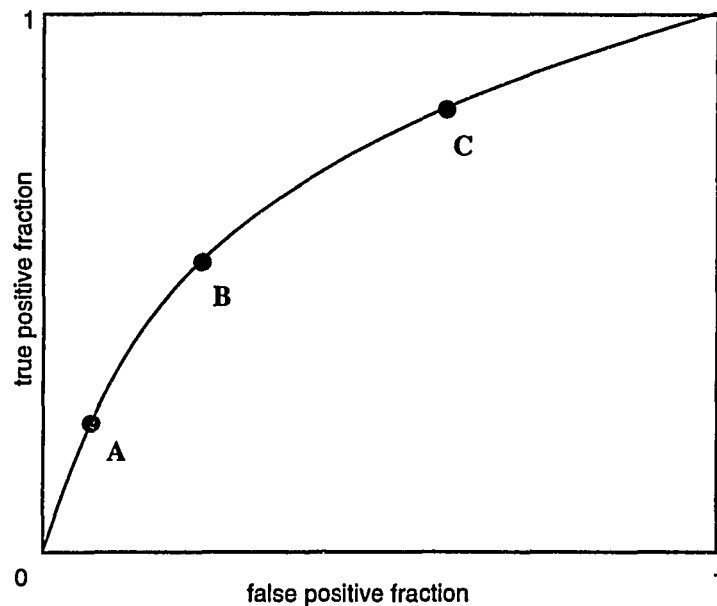
Figure A.1: Receiver operating characteristic curve for some particular detector. Points A, B and C correspond to different decision thresholds.

The probability of the correct detection of K is

$$\beta = P(d_K/K) = \int_{\Omega_K} p_K(x)dx. \tag{A.3}$$

where d_K represents a decision for K and $p_H(x)$, $p_K(x)$ are *a posteriori* conditional densities of the random variable X under H and K, respectively.

Then the detector performance can be measured by examining β as a function of the threshold, α as a function of the threshold, and β as a function of α. This is called a *receiver operating characteristic* or ROC curve, as illustrated in Figure A.1.

In some cases where $p_H(x)$ and $p_K(x)$ are not known, it is very difficult to compute α, β and thus, the ROC curve. Under this situation, we can estimate α and β in the following way.

We first define some terminologies, namely, *true positives* (TP), when we correctly decide for K when K is true, *true negatives* (TN), when we correctly decide for H when H is true, *false positives* (FP), when we incorrectly decide for K when H is true, and *false negatives* (FN), when we incorrectly decide for H when K is true. Suppose that N observations are made, with N_{TP} of them being true positives, N_{TN} true negatives, N_{FP} false positives, and N_{FN} false negatives. Then α can be estimated as

$$
\begin{aligned}
\alpha \quad &\simeq \quad \textit{False positive fraction}(FPF) \\
&= \quad \frac{(Number\ of\ false\ positive\ decisions)}{(Number\ of\ actually\ negative\ cases)} \\
&= \quad \frac{N_{FP}}{N_{TN} + N_{FP}}.
\end{aligned}
\tag{A.4}
$$

Likewise, β can be obtained by

$$
\begin{aligned}
\beta \quad &\simeq \quad \textit{True positive fraction}(TPF) \\
&= \quad \frac{(Number\ of\ true\ positive\ decisions)}{(Number\ of\ actually\ positive\ cases)} \\
&= \quad \frac{N_{TP}}{N_{TP} + N_{FN}}.
\end{aligned}
\tag{A.5}
$$

We can compute ROC curve by plotting TPF *vs* FPF, with various points along the curve obtained simply by varying the decision threshold. The imaging system, the input object, and the viewing conditions should be held constant for a particular curve.

ROC curves are very important measures to analyze the performances of the different detectors as shown in Figure A.2. An ideal detector would give no false positives until all true positives are detected. Its ROC curve would therefore hug the left and top edges of the graph, something like curve IV in Figure A.2. On the other hand, if the ROC curve would be the diagonal line from lower left to upper right, this detector simply guesses whether the object is present, not using any information. Thus the amount by which the ROC curve bows away from the diagonal and towards the upper left-hand corner is a measure of the performance of the detector, at least for the simple binary detection experiment for which the ROC curve was computed [65].
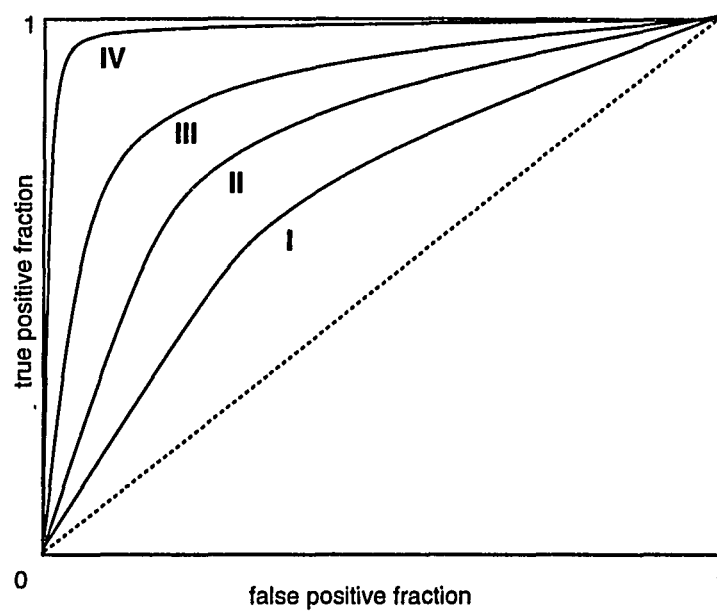
Figure A.2: ROC curves for four different detectors. System IV is the best, system I is the worst for this detection task.

Appendix B

Hotelling Observer

This appendix provides the details of a Hotelling observer whose strategy is based on the discrimination theory proposed in the early 30's by three eminent statisticians Hotelling, Fisher and Mahalanobis.

Starting with different practical problems, their theories more or less converge to answer a common question: How to quantify the discrimination of two population distributions? Hotelling and Mahalanobis each suggested a figure of merit, later known as the Hotelling Trace and Mahalanobis Distance, respectively, to measure statistical distance between the two distributions. Equivalently, Fisher derived an optimum linear discriminant that maximized the statistical distance [73][74].

In order to separate several clusters well, the difference between the center points of the clusters should be large compared to the standard deviation for each cluster.

Hotelling proposed a quantity called "Hotelling Trace" which is a scalar that measures the separability of two or more clusters in the sample space. The Hotelling Trace is defined as

$$J = tr(\mathbf{S}_W^{-1}\mathbf{S}_B), \tag{B.1}$$

where $\mathbf{S}_W$ and $\mathbf{S}_B$ are matrices that describes the scatter characteristics of these clusters.

If $\mathbf{m}_i$ is the sample mean of ith cluster given by

$$\mathbf{m}_i = \frac{1}{N_i} \sum_{\mathbf{x} \in \mathcal{X}_i} \mathbf{x}, \qquad (B.2)$$

then the "within-class" scatter matrix $\mathbf{S}_W$ is defined as

$$\mathbf{S}_W = \sum_{i=1}^{N} p_i < (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T >, \qquad (B.3)$$

where N is the number of clusters, p_i represents *a priori* probability of occurrence of the ith cluster and $< >$ means an average operation. $\mathbf{S}_W$, which is proportional to the covariance matrices of the clusters, indicates the average scattering of the clusters. On the other hand, the "between-class" scatter matrix $\mathbf{S}_B$ is defined as

$$\mathbf{S}_B = \sum_{i=1}^{N} p_i (\mathbf{m}_i - \mathbf{m}_c)(\mathbf{m}_i - \mathbf{m}_c)^T, \qquad (B.4)$$

where $\mathbf{m}_c$ represents the average over the whole clusters given by

$$\mathbf{m}_c = \sum_{i=1}^{N} p_i \mathbf{m}_i. \qquad (B.5)$$

$\mathbf{S}_B$, an average distance between each cluster center $\mathbf{m}_i$ and the center of whole clusters, represents the relative distance between the clusters.

The Hotelling Trace, expressed by two scatter matrices, measures the separability between the clusters. This is proportional to the distance between the clusters. The feature operator $\mathbf{w}$ is obtained so that the Hotelling Trace J has a maximum. The "Hotelling Discriminant" is then defined as the linear function $\mathbf{w}^T\mathbf{x}$.

If we assume that there are two clusters of samples with equal *a priori* probability $(p_1 = p_2 = \frac{1}{2})$, then the scatter matrix S_B reduces to

$$S_B = \tfrac{1}{4}(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T, \qquad (B.6)$$

and the Hotelling Trace J can be written as

$$J = tr(\mathbf{S}_W^{-1}\mathbf{S}_B) = \tfrac{1}{4}tr((\mathbf{m}_1 - \mathbf{m}_2)\mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)^T). \qquad (B.7)$$

Since $(\mathbf{m}_1 - \mathbf{m}_2)\mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)^T$ is a scalar, the above equation reduces to

$$J = \tfrac{1}{4}(\mathbf{m}_1 - \mathbf{m}_2)\mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)^T. \qquad (B.8)$$

This Hotelling Trace J satisfies the eigen value equation of the matrix $\mathbf{S}_W^{-1}\mathbf{S}_B$. In other words, it can be expressed as

$$\mathbf{S}_W^{-1}\mathbf{S}_B\mathbf{w} = \mathbf{w}J, \qquad (B.9)$$

which is a generalized eigenvalue problem. By substituting the expression for J given in (B.8) into (B.9), the above equation can be written as

$$\tfrac{1}{4}\mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T\mathbf{w} = \tfrac{1}{4}\mathbf{w}(\mathbf{m}_1 - \mathbf{m}_2)\mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)^T, \qquad (B.10)$$

and the eigenvector satisfying the above equation has the form

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \qquad (B.11)$$

Thus, the Hotelling discriminant λ can be computed in the following way,

$$\begin{aligned} \lambda &= \mathbf{u}^T \cdot \mathbf{x} \\ &= (\mathbf{m}_1 - \mathbf{m}_2)\mathbf{S}_W^{-1} \cdot \mathbf{x}. \end{aligned} \qquad (B.12)$$

The inner product of $\mathbf{x}$ with $\mathbf{S}_W^{-1}$ functions as a prewhitening operator, while that of the prewhitened $\mathbf{x}$ with $\mathbf{m}_1 - \mathbf{m}_2$ is equal to the matched filter. This is similar to the prewhitening matched filtering operation, as shown in Figure 6.9.

REFERENCES

[1] S. L. Olson, B. W. Fam, P. F. Winter, F. J. Scholz, A. K. Lee S. E. Gordon. Breast Calcifications: Analysis of Imaging Properties. *Radiology*, Vol. 169, pp. 329-332, 1988.

[2] S. J. Riederer. Temporal integration processing techniques. *Digital Radiography:* Selected Topics, J. G. Kereiakes, S. R. Thomas, C. G. Orton, Eds., Plenum Press, New York, pp. 35-53, 1986.

[3] M. L. Giger, K. Boi, H. MacMahon, F. F. Yin. Image Processing Techniques used in the computer-aided detection of radiographis lesions in anatomic background. *Medical Imaging II, Proc. SPIE 914*, pp. 635-637, 1988.

[4] S. M. Lai, X. Li and W. F. Bischof. On techniques for detecting circumscribed masses in mammograms. *IEEE Trans. on Med. Imaging, Vol. 8, No. 4*, pp.377-386, 1989.

[5] H. P. Chan, K. Doi, S. Galhotra, C. J. Vyborny, H. MacMahon, and P. M. Jokich. Image feature analysis and computer-aided diagnosis in digital radiography: 1. Automated detection of microcalcifications in mammography. *Med. Phys, vol. 14, No. 4*, pp. 538-548, 1987.

[6] H. P. Chan, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, K. L. Lam, T. Ogura and H. MacMahon. Improvement in radiologists' detection of clustered microcalcifications in mammograms. *Investigative Radiology, vol. 25*, pp. 1102-1110, October 1990.

[7] M. LeGal, J. C. Durand, and M. Laurent. Conduite a tenir devant une mammographie revelatrice de microcalcifications groupees decouvertes par mammographie. *Nouv Presse Med, vol. 5*, pp. 1623-1627, 1976.

[8] B. W. Fam, S. L. Olson, P. F. Winter and F. J. Scholz. Algorithm for the detection of fine clustered calcifications on film mammograms. *Radiology, Vol. 169*, pp. 333-337, 1988.

[9] D. H. Davies and D. R. Dance. Automatic computer detection of clustered calcifications in digital mammograms. *Phys. Med. Bio., Vol. 35, No. 8*, pp. 1111-1118, 1990.

[10] N. Karssemeijer. A stochastic model for automated detection of calcifications in digital mammograms. *12th Int. Conf. on Information Processing in Medical Imaging, Wye, U.K.*, pp. 227-238, July 1991.

[11] D. Brzakovic, X. M. Luo, and P. Brzakovic. An approach to automated detection of tumors in mammograms. *IEEE Trans. on Med. Imaging, Vol. 9, No. 3*, pp.233-241, 1990.

[12] A. K. Jain. *Fundamentals of digital image processing* Prentice Hall, Eaglewood Cliffs, 1989, p. 403.

[13] W. K. Pratt. *Digital Image Processing* Wiley-Interscience, New York, p. 666, 1991.

[14] W. B. Richardson. Nonlinear filtering and multiscale texture discrimination for mammograms. *Mathematical Methods in Medical Imaging, David C. Wilson, Joseph N. Wilson, Editors, Proc. SPIE 1768*, pp. 293-305, 1992.

[15] A. F. Laine and S. Song. Multiscale wavelet representations for mammographic feature analysis. *Mathematical Methods in Medical Imaging, David C. Wilson, Joseph N. Wilson, Editors, Proc. SPIE 1768*, pp. 306-316, 1992.

[16] P. J. Burt and E. H. Adelson. The Laplacian Pyramid as a Compact Image Code. *IEEE Trans. on Communications, Vol. COM-31, No. 4*, pp.532-540, 1983.

[17] J. Seaton and A. L. Abbott. A comparison of multiresolution morphological and Laplacian techniques for automated inspection. *Applications of Artificial Intelligence X: Machine Vision and Robotics, Proc. SPIE 1708*, pp. 789-800, 1992.

[18] A.Grossmann and J.Morlet. Decomposition of Hardy functions into square interable wavelets of constant shape. *SIAM J. Math., vol.15*, pp.723-736, 1984.

[19] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. in Pure Applied Math., vol. 41, no. 7*, pp. 909-996, 1988.

[20] I. Daubechies. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inform. Theory, vol.36*, pp. 961-1005, Sept. 1990.

[21] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Mag.*, Oct. 1991.

[22] S. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. PAMI, vol. 11, no. 7*, pp. 674-693, Jul. 1989.

[23] J. Canny. A computational approach to edge detection. *IEEE Trans. PAMI, vol. 8, no. 7*, pp. 679-698, Jul. 1986.

[24] S. Mallat. Zero-crossings of a wavelet transform. *IEEE Trans. Inform. Theory, vol. 37*, pp. 1019-1033, Jul. 1991.

[25] S. Mallat and W. L. Hwang. Singularity detection and processing with wavelets. *IEEE Trans. Inform. Theory, vol. 38, no2*, Mar. 1992.

[26] S. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. Acoustic Speech Signal Processing, vol. 37, no. 12*, pp. 2091-2110, Dec. 1989.

[27] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. PAMI, vol. 14, no. 7*, pp. 710-732, Jul. 1992.

[28] R. N. Strickland. Tumor detection in gamma-ray images. *Mathematical Methods in Medical Imaging. DC Wilson & JN Wilson (Eds), Proc. SPIE 1768*, pp. 142-152, 1992.

[29] R. N. Strickland. Detecting small tumors in medical images. *Proc. 8th Scandinavian Conference on Image Analysis, Tromso, Norway*, pp. 465-472, 1993.

[30] R. N. Strickland. Tumor detection in nonstationary backgrounds. *IEEE Trans. Medical Imaging, Vol. 13, No. 3*, pp. 491-499, September 1994.

[31] R. N. Strickland and H. I. Hahn. Tumor detection at multiple scales. *Mathematical Methods in Medical Imaging II. DC Wilson & JN Wilson (Eds), Proc. SPIE 2035*, pp. 180-191, 1993.

[32] H. Barman and Gosta Granlund. Using simple local Fourier domain models for computer-aided analysis of mammograms. *Proc. 8th Scandinavian Conference on Image Analysis, Tromso, Norway*, pp. 479-486, 1993.

[33] J. J. Koenderink and A. J. van Doorn. Generic Neighborhood Operators. *IEEE Trans. PAMI, vol. 14, no. 6*, pp. 597-605, Jun. 1992.

[34] O. Rioul and P. Duhamel. Fast algorithms for discrete and continuous wavelet trasnforms. *IEEE Trans. Inform. Theory, vol. 38, no2*, pp. 569-586, Mar. 1992.

[35] N. Karssemeijer. Adaptive noise equalization and image analysis in mammography. *13th Int. Conf. on Information Processing in Medical Imaging, Flagstaff, AZ*, pp. 472-486, 1993.

[36] J. O. Stromberg. A modified haar system and higher order spline systems. *Conf. in Harmonic Analysis in Honor of Antoni Zygmund, Vol. II*, pp. 475-493.

[37] Y. Meyer. Principe d'incertitude, bases hilbertiennes et algebres d'operateurs. *Seminaire Bourbaki, no. 662*, 1985-1986.

[38] P. G. Lemarie. Une nouvelle base d'ondelettes de $L^2(R)$. *J. Math. Pures et Appl., vol. 67*, pp. 227-238, 1988.

[39] G. Battle. A block spin construction of wavelets. Part I Lemarie functions. *Comm. Math. Phys., vol. 110*, pp. 601-615, 1987.

[40] R. E. Crochiere, S. A. Webber, and J. L. Flanagan. *Digital coding of speech in subbands*. Bell Syst. Tech. J., vol. 55, pp. 1069-1085, Oct. 1976.

[41] A. Croisier, D. Esteban, and C. Galand. Perfect channel splitting by use of interpolation/decimation/tree decomposition techniques. *Conf. Proceedings, 1976 IEEE Int. Conf. on Inform. Sci. Syst., Patras, Greece*, May. 1976.

[42] P. Goupillaud, A. Grossmann, and J. Morlet. Cycle-octave and related transforms in seismic signal analysis. *Geoexploration, vol. 23*, pp. 85-102, 1984/85.

[43] I. Daubechies, A. Grossmann, and Y. Meyer. Painless nonorthogonal expansions. *J. Math. Phys., vol. 27*, pp. 1271-1283, 1986.

[44] M. J. Smith and D. P. Barnwell. Exact reconstruction for tree-structured sub-band coders. *IEEE Trans. Acoust., Speech Proc., vol. ASSP-34*, pp. 434-441, 1986.

[45] A. Cohen, I. Daubechies, and J. C. Feauveau. Biorthogonal bases of compactly supported wavelets. *AT&T Bell Lab., Tech. Rep., TM 11217-900529-07*, 1990.

[46] M. Vetterli and C. Herley. Wavelets and filter banks: Relationships and new results. *Proc. IEEE ICASSP, Albuquerque*, Apr. 1990.

[47] B. R. Hunt and T. M. Cannon. Nonstationary assumptions for Gaussian Models of images. *IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-6*, pp. 876-882, Dec. 1976.

[48] R. N. Strickland. Transforming images into block stationary behavior. *Applied Optics, Vol. 22, No. 10*, pp. 1462-1473, May 1983.

[49] L. W. Basset. Mammographic analysis of calcifications. *Radiologic Clinics of North America, Vol. 30*, pp. 93-105, 1992.

[50] A. P. Dhawan, G. Buelloni, and R. Gordon. Enhancement of mammographic features by optimal adaptive neighborhood image processing. *IEEE Trans. Medical Imaging, vol. 5, no. 1*, pp. 8-15, 1986.

[51] R. M. Nishikawa, M. L. Giger, K. Doi, C. J. Vyborny, R. A. Schmidt, C. E. Metz, Y. Wu, F. F. Yin, Y. Jiang, Z. Huo, P. Lu, W. Zhang, T. Ema, U. Bick, J. Papaioannou, R. H. Nagel. Computer-aided detection and diagnosis of masses and clustered microcalcifications from digital mammograms. *Proc. SPIE 1905*, pp. 422-427, 1993.

[52] J. Dengler, S. Behrens, J. F. Desaga. Segmentation of Microcalcifications in Mammograms. *IEEE Trans. Medical Imaging, vol. 12, no. 4*, pp. 634-642, Dec. 1993.

[53] W. Qian, P. Clarke, M. Kallergi, H.D. Li, R. Velthuizen, R. A. Clark, and M. L. Silbiger. Tree-structured nonlinear filter and wavelet transform for micro-calcification segmentation in mammography. *Biomedical Image Processing and Biomedical Visualization, Raj S. Acharya, Dmitry B. Goldgof, Editors, Proc. SPIE 1905*, pp. 509-520, 1993.

[54] Y. Yoshida, K. Doi, R. M. Nishikawa, T. Ema, and W. Zhang. Automated detection of clustered microcalcifications in digital mammograms using wavelet processing techniques. *Medical Imaging, Proc. SPIE 2167*, 1994.

[55] H. P. Chan, K. Doi, C. J. Vyborny, K. L. Lam and R. A. Schmidt. Computer-aided detection of microcalcifications in mammograms: methodology and pre-liminary clinical study. *Investigative Radiology, Vol. 23*, pp. 664-671, 1988.

[56] C. J. Baines, A. B. Miller, C. Wall, D. V. McFarlane, I. S. Simor, R. Jong, B. J. Shapiro, L. Audet, M. Petitclerc, D. Ouimet-Oliva, J. Ladouceur, G. Hebert, T. Minuk, G. Hardy, and H. K. Standing. Sensitivity and specificity of first screen mammography in the Canadian national breast screening study. *Radiology, Vol. 160*, pp. 295-298, 1986.

[57] C. J. Baines, D. V. McFarlane and A. B. Miller. The role of the reference radiologist: estimates of interobserver agreement and potential delay in cancer detection in the national screening study. *Investigative Radiology, Vol. 25*, pp. 971-976, 1990.

[58] L. W. Basset, D. H. Bunnel, R. Jahanshahi, R. H. Gold, R. D. Arndt, and J. Linsman. Breast cancer detection: one versus two views. *Radiology, Vol. 165*, pp. 95-97, 1987.

[59] P. H. Peeters, A. L. Verbeek, J. H. Hendriks, R. Holland, M. Mravunac, and G. P. Vooijis. The occurrence of interval cancers in the Nijmegen screening programme. *British Journal of Cancer, Vol. 69*, pp. 929-932, 1989.

[60] E. Silverberg, C. C. Boring, and T. S. Squires. Cancer statistics, 1990. *Ca-A Journal for Clinicians, Vol. 40*, pp. 9-26, 1990.

[61] A. Laine, J. Fan, and S. Schuler. A framework for contrast enhancement by dyadic wavelet analysis. *Proc. Second Int. Workshop on Digital Mammography (Elsevier, Amsterdam), A. G. Gale et al., Editors, York, UK*, pp. 91-100, 1994.

[62] D. Marr and E. Hildreth. Theory of edge detection. *Proc. Roy. Soc. Lon., vol. 207*, pp. 187-217, 1980.

[63] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. *IEEE Trans. on Image Processing, Vol. 1, No. 2*, pp.205-220, 1992.

[64] D. H. Ballard, C. M. Brown. *Computer Vision.* Prentice Hall, Eaglewood Cliffs, p. 235, 1982.

[65] H. H. Barrett, W. Swindell. *RADIOLOGICAL IMAGING - The Theory of Image Formation, Detection, and Processing, Volume 2.* Academic Press, p. 622, 1981.

[66] R. M. Haralick, L. G. Shapiro. *Computer and Robot Vision.* Addison-Wesley Publishing Company, Inc., pp. 469-470, 1992.

[67] H. Kobatake, Y. Yoshinaga, and M. Murakami. Automatic detection of malignant tumors on mammogram. *IEEE International Conference on Image Processing, Proc. ICIP-94*,pp.205-220, Austin, Texas, 1994.

[68] R. N. Strickland and H. I. Hahn. Detection of microcalcifications in mammograms using wavelets. *Wavelet Applications in Signal and Image Processing II, Proc. SPIE 2303*, pp. 430-441, Jul., 1994.

[69] R. N. Strickland and H. I. Hahn. Detection of microcalcifications using wavelets. *Proceedings of the second International Workshop on Digital Mammography*, York, England, pp. 79-88, Jul., 1994.

[70] R. N. Strickland and H. I. Hahn. Wavelet transforms for detecting microcalcifications in mammography. *IEEE International Conference on Image Processing, Proc. ICIP-94*,pp.402-406, Austin, Texas, 1994.

[71] R. N. Strickland and H. I. Hahn. Wavelet transforms for detecting microcalcifications in mammograms. *submitted to IEEE Transactions on Medical Imaging.*, Dec., 1994.

[72] R. N. Strickland and H. I. Hahn. Wavelet transform matched filters for the detection and classification of microcalcifications in mammography. *IEEE International Conference on Image Processing, Proc. ICIP-95*, Washington, D.C., Oct., 1995.

[73] R. O. Duda, P. E. Hart. *Pattern Classification and Scene Analysis.* A Wiley-Interscience Publicatiobn., pp. 114-118, 1973.

[74] Jie Yao. Model observers for predicting human performance on signal-detection tasks. Ph.D. Dissertation, Univ. of Arizona (1994).

[75] R. P. Lippmann. An introduction to computing with neural nets. *IEEE ASSP MAGAZINE*, pp.4-22, 1987.

[76] A. Laine, S. Schuler, J. Fan, and W. Huda. Mammographic feature enhancement by multiscale analysis. *IEEE Trans. Medical Imaging, vol. 13, no. 4*, pp. 725-740, Dec. 1 994.